Safety Performance of Pavement Marking Retroreflectivity in Alabama

Final Report

Prepared By:

Huaguo Zhou, Ph.D. & P.E.
Md Mahmud Hossain, Ph.D.

Highway Research Center
Department of Civil and Environmental Engineering
238 Harbert Engineering Center

Auburn University
Auburn, AL 36849

Sponsoring Agency:

Alabama Department of Transportation
1409 Coliseum Blvd
Montgomery, AL 36110

April 2025

1. Report No.	2. Government Accession No.	3. Recipient Catalog No.
4. Title and Subtitle Safety Performance of Pavement Marking Retroreflectivity in Alabama		**5. Report Date** April 2025 **6. Performing Organization Code**
7. Author(s) Huaguo Zhou and Md Mahmud Hossain		**8. Performing Organization Report No.** ALDOT 931-078
9. Performing Organization Name and Address Highway Research Center Department of Civil and Environmental Engineering 238 Harbert Engineering Center Auburn, AL 36849		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** **FHWA/ALDOT 931-078**
12. Sponsoring Agency Name and Address Alabama Department of Transportation 1409 Coliseum Boulevard Montgomery, AL 36130		**13. Type of Report and Period Covered:** Technical Report **14. Sponsoring Agency Code**

15. Supplementary Notes

Project performed in collaboration with the Alabama Department of Transportation.

16. Abstract

The evaluation of pavement marking materials primarily relies on their retroreflectivity (RL) — a measure of how effectively they reflect vehicle headlights back toward the driver. State DOTs typically contract private firms to measure RL, often without accurate records of when markings were installed or restriped. Additionally, establishing the statistical relationship between RL and roadway safety demands rigorous data preparation, roadway segmentation, and reasonable assumptions to predict and assign RL values to individual road segments. Therefore, to promote the broader application of RL as a crucial indicator of road safety, it is essential to develop a robust approach that integrates RL prediction models to analyze its impact on crashes.

This study focuses on rural roadways with thermoplastic markings to assess the effect of RL on highway safety. In the first phase, regression models were developed for yellow centerlines (YCL) and white right-edge lines (WREL) on rural two-lane and multilane roads. These models were then used in the second phase to predict RL values for prior years, increasing the number of segments with usable RL data. This expanded dataset was employed to examine the statistical relationship between RL and road crashes. RL data from 2020 and 2021 were obtained for the Montgomery area in Alabama. Corresponding traffic flow, roadway characteristics, location, and crash data were compiled from six different sources. Multiple linear regression and binary logit models were applied to uncover statistically significant patterns. This method proved effective by increasing sample size for more robust crash analysis. Results showed that WREL levels below 250 mcd/m²/lux significantly increased the likelihood of single-vehicle run-off-road crashes, especially on curve segments and under dark conditions.

The developed RL prediction models serve as a valuable tool for local agencies, allowing them to forecast RL using just one year of RL and traffic data. The study's findings reinforce the importance of maintaining adequate RL levels and align with the Safe System Approach—specifically the principle of "Making Our Roads Safer.

17. Key Words: Retroreflectivity Degradation, Regression Modeling, Crash Data Analysis, Road Alignment		**18. Distribution Statement** No restrictions	
19. Security Classification (of this report) Unclassified	**20. Security Classification (of this page)** Unclassified	**21. No. of Pages** 72	**22. Price** None.

Acknowledgment

The authors would like to thank the project panel members for their guidance throughout the project, especially D. Jason Daniels, P.E., State Traffic Engineer, for helping collect data. A special thanks to Md Roknuzzaman for editing the final report.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

Longitudinal pavement markings serve as crucial sources of conveying regulations, guidance, or warnings to motorists, offering insight into road geometry (Rasdorf et al., 2009). They delineate boundaries between lanes to assist drivers in choosing the correct travel path and keeping their vehicles within the designated lane (Fu et al., 2013). Undoubtedly, pavement markings enhance visibility for driving both day and night by improving the driver's ability to make informed decisions (Hussein et al., 2020; Zwahlen et al., 1997). The specifics of longitudinal pavement markings can differ depending on the number of highway lanes. For instance, on two-lane roads, as viewed from the direction of travel, they encompass a yellow centerline (YCL) and white right edgeline (WREL), while on multilane roads, they include a yellow left edge line (YLEL), white laneline(s) (WLL), and WREL (**Figure 1**).

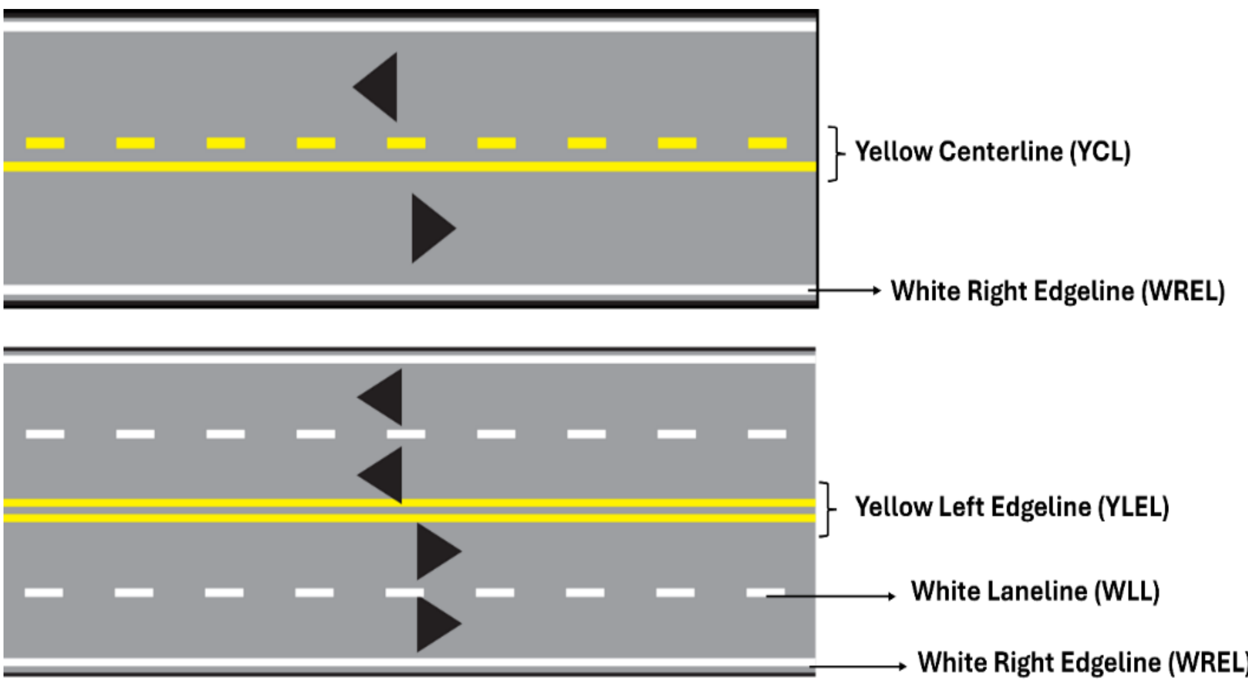

Figure 1 Pavement marking line type by lane configurations

The service life of pavement markings is significantly influenced by the materials used and traffic volumes. Across the nation, the predominant materials for longitudinal striping include thermoplastics, solvent-borne paints, and water-based paints (Idris et al., 2024). Typically, agencies regard restriping intervals ranging from 6 to 12 months for paints and 2 to 6 years for thermoplastics (Thomas et al., 2001). Thermoplastic pavement markings can last up to 8 years in low-traffic flow conditions. Thermoplastics' higher durability than paints is primarily attributed to their greater thickness and highly adhesive coating, resulting in a stronger thermal bond with the pavement surface, making the markings more resistant to wear and tear (Abboud et al., 2002).

Deciding when the pavement marking material needs to be restriped can be challenging. The evaluation of pavement marking materials primarily relies on their retroreflectivity (RL), which refers to how well they bounce light from vehicle headlights back toward the driver (Thomas et al., 2001). This reflective quality is what makes the markings appear bright at night and serves as an indicator of their visibility in darkness (Abboud et al., 2002). RL is usually measured in units of millicandelas per square meter per lux (mcd/m²/lux). Quicker identification of pavement markings is linked to their RL levels, which vary in brightness (Aktan et al., 2004). For example, markings with RL of 228 mcd/m²/lux have a longer detection distance than those with RL of 118 mcd/m²/lux (**Figure 2**).

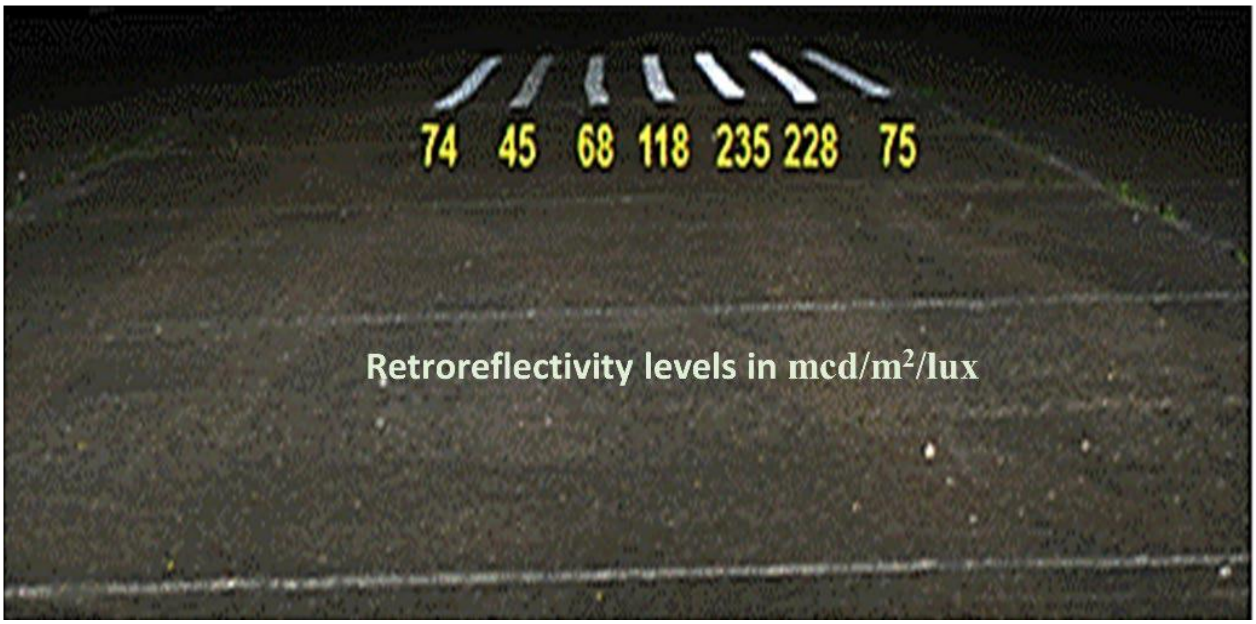

Figure 2 Pavement markings with different RL levels at nighttime (FHWA, 2017)

In recent years, pavement marking RL has been documented as a crucial characteristic for enhancing infrastructure compatibility with advanced driver assistance systems (ADAS) and automated driving systems (ADS) (Pike et al., 2018). Moreover, ensuring that markings maintain certain levels of RL is perceived to improve road safety. Maintaining an adequate level of RL could significantly reduce nighttime crashes and fatalities in dark and adverse weather conditions (Hussein et al., 2020). Enhancing safety on rural roads is the primary incentive for exploring the correlation between RL and crashes. Lower RL values may increase the likelihood of single-vehicle run-off-road (ROR) crashes, which are more frequent in rural areas (Smadi et al., 2008). Based on RL data experiments, the qualitative states of thermoplastic marking can be characterized from prior studies (Ortiz-García et al., 2006; Porras-Alvarado et al., 2014), as depicted in **Table 1.** Assuming a failure threshold of 100 mcd/m²/lux for yellow markings, these studies observed a similar degradation nature in white markings at 150 mcd/m²/lux. The qualitative states were determined by analyzing the patterns of RL degradations using a transition probability matrix.

Table 1 RL states of thermoplastic markings (Ortiz-García et al., 2006; Porras-Alvarado et al., 2014)

Qualitative state	RL interval (mcd/m²/lux)	
	White thermoplastic	Yellow thermoplastic
Excellent	≥ 450	≥ 400
Good	350-449	300-399
Fair	250-349	200-299
Poor	150-249	100-199
Failure	≤ 150	≤ 100

The degradation rate of RL, a crucial factor determining the service life of pavement markings, can be influenced by various factors, including the type of material, traffic volume, percentage of heavy vehicles, geographic location, climate conditions, road surface type, material thickness, type and quality of glass beads or other reflective elements used, and the initial value of installation (Idris et al., 2024). Transportation agencies such as the American Association of State Highway and Transportation Officials (AASHTO), Federal Highway Administration (FHWA), and State Departments of Transportation (DOTs) have made significant efforts to prepare comprehensive standard specifications, manuals, and procedures to ensure the appropriate application and maintenance of pavement markings (Sasidharan et al., 2009). In this perspective, FHWA aims to enhance safety by setting minimum RL levels in the Manual on Uniform Traffic Control Devices (MUTCD), thereby meeting the nighttime visibility requirements of drivers on public roadways. On August 5, 2022, FHWA published a new regulation establishing national standards for minimum RL levels of longitudinal marking lines on all roads, varying according to posted speed limits (2009 MUTCD Revision 3) (FHWA, 2022). This regulation mandates that all state agencies establish a method by 2027 to maintain the designated minimum RL level and integrate RL considerations into future restriping endeavors. As per the regulation, roads with posted speed limits equal to or greater than 35 mph must maintain a minimum RL of 50 mcd/m²/lux, while no minimum threshold is set for speed limits below 35 mph. Furthermore, it is recommended to maintain a minimum RL of 100 mcd/m²/lux for freeways with posted speed limits of 70 mph and above. These regulations and guidelines apply under dry conditions and exclude conditions such as ambient lighting, roads with less than 6,000 annual average daily traffic (AADT), dashed extension lines, curb markings, parking spaces, and shared-use paths. The changes regarding RL in the 2009 MUTCD revisions are exhibited in **Table 2**.

Table 2 Minimum RL values in MUTCD revisions

Revision 3 (Effective September 6, 2022)				Revision 1 & 2 (Effective June 13, 2012)			
	Standard		Guidance	Two-lane with centerline only		All other roads	
Speed limit (mph)	< 35	≥ 35	≥ 70	35-50	≥ 55	35-50	≥ 55
RL (mcd/m²/lux)	n/a	50	100	100	250	50	100

Notably, in the latest 11th edition of the MUTCD, effective in December 2023, the minimum RL requirements have remained consistent with the previous version, 2009 MUTCD Revision 3. In recent years, the Alabama Department of Transportation (ALDOT) has hired a consulting firm to gather RL data on Alabama highways. According to the ALDOT Standard Specifications for Highway Construction Special Provisions 2022, the initial RL of standard thermoplastic in Alabama should be at least 450 mcd/m²/lux for white markings and 300 mcd/m²/lux for yellow markings in Alabama. (ALDOT, 2022).

This project aims to evaluate the impact of pavement marking RL on highway safety in Alabama. RL measurements require mobilizing equipment and personnel, making it challenging to conduct quarterly/monthly readings of selected road segments within the constraints of limited budgets and manpower. Therefore, agencies often prioritize restricting RL assessments to one or two readings per segment yearly. Additionally, state DOTs commonly enlist private consulting firms to measure RL in specific road segments without precise records indicating the timing of pavement marking installation or restriping. To address this challenge, this project develops RL prediction models capable of forecasting RL without requiring initial RL readings or marking age data. These models ensure adequate RL (measured and predicted) data that matches the historical crash information for identifying statistical relationships between RL and crash frequency. The findings of this project can help ALDOT, and local transportation agencies assess the service life of pavement markings effectively from a more scientific perspective. Also, the knowledge gained could be beneficial in strengthening the existing pavement management programs aimed at reducing the crashes associated with lower RL levels.

1.2 Research Objectives

1.2.1 Develop Predictive Models to Estimate RL Levels Over Time

In the first phase, this project utilizes a multiple linear regression approach to develop separate regression models for YCL and WREL on rural two-lane and multilane roads. Predictor variables are selected based on an extensive literature review, the availability of relevant variables within the RL and traffic flow database obtained from ALDOT, and their statistical significance. The newly developed models will be compared to the calibrated models from a previous ALDOT study (Ozelim et al., 2014).

1.2.2 Comprehend the Relationship between RL and Crash Frequency

This project investigates the statistical relationship between RL and road crashes in the second phase. Unlike previous studies that calculated average RL over 1-mile segments, this research uses shorter segments, specifically 0.5 miles, for more granular analysis. Practical assumptions are made based on the available data and the limitations of the predictive models. Two widely used regression models, the binary logit model, and the Negative Binomial (NB) model, are applied to data analysis.

1.3 Summary

This project is carried out in two phases to analyze the relationship between pavement marking RL and road safety on rural roads. First, multiple linear regression models are developed for different marking line types to predict RL at a given timeframe. Later, the study explores the statistical correlation between RL and road crashes, enhancing the dataset by applying RL prediction models to estimate RL values for more road segments.

The rest of the chapter of this report is structured as follows: Chapter 2 presents a comprehensive literature review on (i) synthesizing existing RL prediction models and (ii) exploring established correlations between RL and road safety. Chapter 3 provides an overview of the datasets collected from multiple sources and the variables extracted for analyses, featuring RL, traffic flow, road, and crash characteristics. Chapter 4 introduces multiple regression models developed to predict RL for WREL and YCL, considering different lane configurations. Chapter 5 analyzes crash data to establish statistical correlations between RL and crash frequency, emphasizing dark lighting conditions and single-vehicle run-off-road (ROR) collisions. Finally, Chapter 6 highlights the research contributions, discusses key findings and recommendations, addresses study limitations, and suggests potential directions for future research.

2 LITERATURE REVIEW

This section summarizes past studies’ findings to understand the diverse methodologies used to develop RL predictive models and investigate the correlation between RL levels and road crashes. The discussion covers a range of factors, including road conditions, traffic patterns, and pavement characteristics, that contribute to RL degradation. Overall, this literature review serves two primary purposes: i) identifying potential variables for developing regression models to predict RL degradation and ii) facilitating the selection of relevant crash types to explore the statistical relationship between crash frequency and RL levels.

2.1 RL Degradation Modeling

During the late 1990s and early 2000s, the prevalent approach to predicting RL relied on simplistic linear and nonlinear regression models centered on a single variable, such as marking age or traffic volume (Abboud et al., 2002; Andrady et al., 1997; J. T. Lee et al., 1999; Migletz et al., 2001). However, the low goodness of fit of these models in capturing the complexity of RL led to their gradual replacement with multiple linear regression models. These regression models have offered considerably higher goodness of fit by incorporating more variables such as initial RL (within 30 days after installation), marking age, geographic location, traffic characteristics, marking color, and position (Fu et al., 2013; Malyuta, 2015; Mull et al., 2012; Ozelim et al., 2014; Robertson et al., 2013; Sasidharan et al., 2009; Sitzabee et al., 2009; Wang et al., 2016). However, considering highly correlated variables can lead to insignificant model outcomes; for instance, cumulative traffic passage (CTP) may not be included in the model if AADT and time are already taken into account (Zhang et al., 2010). Additionally, in larger datasets, increased heterogeneity can significantly diminish the model's performance by introducing greater variability that is difficult to account for (Zuur et al., 2009). Among the range of material types studied, paint and thermoplastic emerged as the most commonly used. However, with growing interest in alternative durable materials like epoxy, polyurea, preformed tape, and methyl methacrylate, researchers have developed predictive models tailored to these materials (Fu et al., 2013; Idris et al., 2024; Kopf, 2004; Migletz et al., 2001; Mull et al., 2012; Pike et al., 2015; Sasidharan et al., 2009; Wang et al., 2016). Regarding marking color, it is observed that yellow paint and thermoplastic generally exhibit slower RL degradation compared to white markings of the same type. However, when considering the position of markings, the influence of color on RL models is automatically regulated (with possible variations by location), as specific marking line types correspond to distinct colors: solid white for the right edgeline, broken/solid yellow for centerline, broken white for laneline, and solid yellow for left edgeline. Typically, RL models are significantly associated with initial RL readings, traffic flow attributes, and marking age. It is worth noting that a few studies have revealed the non-linear degradation trend of RL, taking on an inverted parabolic shape, where readings initially rise above the initial values before gradually declining thereafter (Chimba et al., 2018; Y.-J. Lee, 2009; Sitzabee et al., 2009).

In Alabama, over twenty years ago, Abboud et al. (2002)conducted a case study aimed at determining the service life of paint and white thermoplastic markings. They utilized field data from 520 miles of longitudinal pavement markings across nine counties in Alabama, setting a minimum RL standard of 150 mcd/m²/lux for WREL. The study revealed a logarithmic relationship between RL and traffic exposure, a function of AADT and marking age. Building upon this groundwork, Ozelim et al. (2014) embarked on a study in the last decade to assess the applicability of established RL models to Alabama's road conditions, alongside testing their newly developed models. They used four years of RL data, ensuring a comprehensive representation of road segments across the state, encompassing both two-lane and four-lane highways. While some conventional models showed promising performance on actual RL data, the service life estimates derived from the best-fitted models for thermoplastic materials seemed impractical. Consequently, the study pivoted towards developing new regression models to forecast RL for each marking color attribute. The results underscored a strong correlation between RL and factors such as marking age and AADT, with initial RL value exhibiting minimal influence on model accuracies. The unit of AADT is vehicles per day (vpd). The final regression equations from both Alabama studies are outlined in **Table 3**.

Table 3 Regression equations for RL prediction in Alabama

Reference	**Equations**
Abboud et al. (2002)	$R_L = -19.457 * \ln(VE) + 26.27$ (for paint) $R_L = -70.806 * \ln(VE) + 639.66$ (for thermoplastic) $VE = ADT_{ln} * PM_{age} * 20.4 * 10^{-3}$ R_L = retroreflectivity (mcd/m²/lux), VE = vehicle exposure (thousands of vehicles), ADT_{ln} = ADT per lane, PM_{age} = marking age (months).
Ozelim et al. (2014)	$R_L = 619.4 - 5.13t - 0.00699 * AADT$ (for white marking) $R_L = 407.3 - 4.969t - 0.00217 * AADT$ (for yellow marking) R_L = retroreflectivity (mcd/m²/lux), t = marking age (months), $AADT$ = Annual average daily traffic (vpd)

In addition to utilizing RL data collected from specific states, numerous studies have incorporated data from the National Transportation Product Evaluation Program (NTPEP) to develop RL prediction models (Idris et al., 2024; Mousa et al., 2021; Wang et al., 2016). NTPEP selects test decks from various locations across the United States (Minnesota, Pennsylvania, Florida, and Wisconsin), encompassing diverse traffic patterns and geographical conditions, and installs marking products in a transverse direction. The degradation of RL in transverse skip lines aligns with longitudinal pavement markings, providing a strong representation of RL degradation. RL readings are collected at 12 different time intervals (0, 1, 2, 3, 11, 12, 15, 21, 24, 27, 33, and 36 months) over a period of up to three years. Consequently, there exists a substantial volume of data suitable for the implementation of comprehensive machine learning algorithms. A few researchers have taken an intriguing approach by categorizing RL levels into bins (Sitzabee et al., 2013; Xu et al., 2021). However, this method overlooks the fact that linearly deriving the thresholds is inherently limited, as RL performance has consistently demonstrated better modeling results when approached logarithmically. These researchers explore various bin intervals, ensuring that all bins maintain RL levels above previously recommended minimum thresholds. **Table 4** outlines a compilation of literature reviews in the U.S. proposing models for RL degradation.

Table 4 A summary of studies in the U.S. on RL degradation modeling

Reference	**Study area**	**Material(s)**	**Input(s)**	**Method(s)**	R^2
Andrady et al. (1997)	TN, KY, CO, and OH	Thermoplastics	Marking age, initial RL	Simple linear regression	0.50-0.71
J. T. Lee et al. (1999)	MI	Waterborne paint, thermoplastic, polyester, tape	Marking age	Simple linear regression	0.14-0.18
Migletz et al. (2001)	AZ, AR, CA, CO, FL, GA, IA, KS, LA, MN, MO, NH, NC, OK, UT, VI, WA, WV, WI	Waterborne paint, thermoplastics, polyurea material, epoxy paint	Traffic volume, initial RL	Linear and exponential regression	----
Abboud et al. (2002)	AL	Waterborne paint, thermoplastic	Marking age, traffic volume	Exponential regression	0.31-0.58
Sarasua et al. (2003)	SC	Thermoplastic, epoxy	Marking age	Simple linear regression	0.21-0.47
Thamizharasan et al. (2003)	SC	Thermoplastic and epoxy	Marking age, traffic volume	Multiple linear regression	0.21-0.78
Lindly et al. (2003)	AL	Thermoplastics and profiled	Traffic volume	Linear and exponential regression	0.53-0.67
Kopf (2004)	WA	Waterborne and solvent-borne paint	Marking color, AADT, marking age	Linear and exponential regression	0.33-0.73
Bahar et al. (2006)	AL, CA, MN, MO, PA, TX, UT, WI	Waterborne paint, thermoplastic	Marking age	Inverse polynomial model	---

Reference	**Study area**	**Material(s)**	**Input(s)**	**Method(s)**	R^2
Hollingsworth (2012)	NC	Waterborne paint, thermoplastic	Marking age, traffic volume,	Logarithmic	0.53
Sitzabee et al. (2009)	NC	Thermoplastic, waterborne paint	Marking age, initial RL, traffic volume, marking location, marking color	Multiple linear regression	0.60-0.75
Sasidharan et al. (2009)	PA	Epoxy and waterborne paint	Marking age, ADT, line type, Surface type	Multiple linear regression	0.27-0.44
Hummer et al. (2011)	NC	Solvent-borne paint	Initial RL, marking age	Linear mixed-effect model	0.67
Karwa et al. (2011)	NC	Thermoplastic	Initial RL, marking age, traffic volume, marking type, marking location	Artificial Neural Network (ANN)	---
Mull et al. (2012)	NC and OH	Solvent-borne paint	Marking age, AADT, initial RL, winter road maintenance activities	Multiple linear regression	0.76
Fu et al. (2013)	LA	Thermoplastic, tape, and inverted profile thermoplastic	Marking age, traffic volume	Multiple linear regression	0.18-0.89
Robertson et al. (2013)	SC	Waterborne paint	Marking age, traffic volume, lane width, shoulder width	Multiple linear regression	0.24-0.34
Ozelim et al. (2014)	AL	Thermoplastic	Marking age, traffic volume, initial RL	Multiple linear regression	0.46-0.50
Malyuta (2015)	TN	Waterborne paint, thermoplastic	Marking age, traffic volume	Multiple linear regression	0.33-0.46

Reference	**Study area**	**Material(s)**	**Input(s)**	**Method(s)**	$\boldsymbol{R^2}$
Pike et al. (2015)	TX	Waterborne and epoxy paint, polyurea material, thermoplastics	Marking age, initial RL	Exponential regression	0.64-0.98
Wang et al. (2016)	FL, PA, MN	Permanent polymeric tape and methyl methacrylate	Marking age, maximum RL value, and traffic volume	Multiple linear regression	0.56-0.68
Chimba et al. (2018)	TN	Thermoplastic	Marking age	Markov Chain Model	----
Mohamed et al. (2019)	ID	Waterborne	Marking age	Simple linear regression	0.80-0.97
Mousa et al. (2021)	FL, PA, MN, MS	Waterborne paint	Initial RL, manufacturer, surface type, marking color, thickness, bead types, marking age, air temperature, rainfall, snowfall, traffic volume, surface age	Categorical Boosting	0.83-0.98
Idris et al. (2024)	MN, PA, Fl, WI	Waterborne paint, thermoplastic, preformed thermoplastic, permanent polymeric tape, epoxy, polyurea, and methyl methacrylate	Marking age, traffic volume, snowfall, thickness, surface type, marking materials, marking color, manufacturer, bead type of the first drop, bead type of the second drop	Decision Tree (DT) and ANN	0.55-0.96

2.2 Relationship Between RL and Road Safety

Because of the higher cost involved, RL data is typically collected during summers with one to three intervals per year to minimize the winter effects. This challenges accurately determining the RL levels of pavement markings at the precise time and location of each crash. Despite the availability of crash data, researchers have often had to make assumptions regarding RL levels for their analysis. Earlier, a few researchers utilized measured data to model RL for predictive purposes, while others relied on assumptions without direct measurements. Furthermore, in most states, roadways with high traffic volume are restriped every 2 to 4 years with thermoplastic markings. Consequently, there are very few roadways with RL levels below 50 mcd/m^2/lx.

A few studies have used RL measurements to assess their impact on safety and crashes. Earlier studies primarily established minimum RL thresholds based on crash rates. For instance, a study from New Zealand conducted a before-and-after comparison of crash rates to evaluate the effects of a policy implemented in 1997, setting a minimum RL of 70 mcd/m^2/lx for pavement markings (Dravitzki et al., 2006). This study was performed on the assumption of increased brightness in the post-implementation period. Subsequent studies have utilized measured RL data, which has been available for specific road segments across various settings over multiple years. Retroreflective markings reflect the light from headlights back toward the driver, enhancing the visibility of road lanes, edges, and centerlines in low-light or dark conditions. As a result, most studies have concentrated on examining the impact of RL on nighttime crashes and crashes in dark conditions, given its critical role in improving driver guidance and safety during these conditions. However, very few have successfully established a statistical relationship between crashes and RL. One key challenge lies in the scarcity of target crashes within segments, making it difficult to select appropriate statistical techniques. Moreover, the distribution of RL is highly skewed, necessitating a logarithmic description of the data rather than a linear one. **Table 5** summarizes studies that have endeavored to establish the correlation between crashes and RL.

Table 5 Overview of studies investigating the association between target crashes and RL

References	**Study area**	**Method(s)**	**Target crashes**	**Key findings**
Abboud et al. (2002)	AL	Descriptive analysis	Nighttime	Thermoplastic long lines offered a safer traffic operation compared to painted highways.
Dravitzki et al. (2006)	New Zealand	Descriptive analysis	Dark conditions, nighttime, curve, curve during dark	No significant changes observed in crash trends following the maintenance of a minimum RL of 70 mcd/m²/lx.
Masliah et al. (2007)	CA	Innovative time series analysis	Nighttime and non-intersection	Minimal effect of RL on crash modification factors.
Smadi et al. (2008)	IA	Logistic regression	Nighttime and lane departure not attributed to collision with animals or objects, vehicle-to-vehicle collisions, maneuvers to avoid collisions with other vehicles, or equipment malfunctions.	Lower RL associated with higher crash probability when segments with RL $\leq$ 70 mcd/m²/lx analyzed separately.
Donnell et al. (2009)	NC	ANN and generalized estimating equations with negative binomial (NB) distribution	Nighttime, sideswipe collisions, and ROR collisions involving fixed objects in non-work zones, without alcohol involvement, and in dry weather.	Increasing RL on two-lane highways may correspond to reduced frequencies of target crashes. The related outcomes for YCL showed marginal significance, whereas the WREL results were not significant.
Carlson et al. (2013)	MI	NB regression	Nighttime and single-vehicle nighttime crashes at non-intersection during the non-winter months	The expected crash frequency decreases when RL increases; it varies in datasets within the range of $\leq$ 150-200 mcd/m²/lx across marketing positions.

References	Study area	Method(s)	Target crashes	Key findings
Avelar et al. (2014)	MI	Generalized linear mixed-effects models	Single-vehicle, sideswipe, and ROR crashes at night	Segments with higher RL were observed to have fewer crashes compared to segments with lower retroreflectivity, for both WREL and YCL
Bektas et al. (2016)	IA	NB regression	Nighttime, ROR crashes, and nighttime ROR crashes	For 1-mile four-lane road segments, the expected number of crashes decreased significantly as the increase of RL of WREL

2.3 Existing Research Gaps

RL measurements require mobilizing equipment and personnel, making it challenging to conduct quarterly/monthly readings of selected road segments within the constraints of limited budgets and manpower. Therefore, agencies often prioritize restricting RL assessments to one or two readings per segment per year. Additionally, state DOTs commonly enlist private consulting firms to measure RL in specific road segments without precise records indicating the timing of pavement marking installation or restriping. Thus, to expand the practice and implementation of pavement marking RL measurement as a crucial aspect of road service life and safety, it becomes imperative to develop RL models that do not rely on input regarding the marking age and initial RL readings. These regression models can help ALDOT's pavement management system in identifying locations where RL levels are projected to drop below safe thresholds in the coming years. In previous studies examining factors influencing RL degradation or developing RL prediction models, researchers typically focused on calculating the average RL for each 1-mile road segment. While this approach provides a general view, a more effective and practical method would involve calculating the average RL for road segments with similar geometric and traffic characteristics. This approach can yield more precise insights into RL degradation patterns by grouping segments with comparable factors—such as traffic volume, posted speed limit, functional class, lane configuration, and horizontal alignment. In addition, this method can reduce heterogeneity within the dataset, leading to more accurate and reliable predictions of RL degradation. Moreover, most studies have combined WREL data for rural two-lane and multi-lane roads in their analyses, even though degradation patterns may vary depending on the number of lanes.

Establishing the statistical correlation between RL and road safety relies significantly on meticulous data preparation, segmentation, and practical assumptions regarding the assignment of retroreflectivity values for each segment throughout the specified analytical period. Hence, there is a need for further research to investigate the statistical relationship between RL and crashes thoroughly. This entails organizing and segmenting the dataset using multiple criteria/assumptions and adopting appropriate statistical techniques to uncover meaningful insights into this correlation. The findings can help determine the RL threshold below which the probability of crashes increases significantly.

3 DATA COLLECTION

This section provides an overview of the datasets collected from multiple sources to extract the variables required to fulfill the research objectives. The primary datasets encompass RL data and crash data acquired from ALDOT, traffic flow data sourced from ALDOT's traffic data website, road inventory data obtained from Google Earth Pro, crash data, and the Highway Performance Monitoring System (HPMS). This section discusses the accessibility, limitations, and constraints associated with each data source, as such factors play a significant role in selecting the preliminary contributing factors and the choice of analytical methodologies.

3.1 RL data

Thermoplastic is a widely used material for pavement markings due to its moderate cost and exceptional durability, making it the predominant choice in Alabama. When reapplying thermoplastics over existing markings, there is no need to remove the old markings. For this study, RL data were collected from ALDOT for the Montgomery area in 2020 and 2021. All RL measurements were conducted under dry weather and daylight conditions. A thorough analysis of the RL data revealed that 575 miles of road segments were covered in 2020, with approximately 75% located in rural areas. In 2021, coverage increased to 1,525 miles, adding 950 miles measured only that year. Of these additional 950 miles, around 90% were in rural areas. Notably, the original dataset lacked area setting information, which was subsequently extracted from Topologically Integrated Geographic Encoding and Referencing (TIGER) database using latitude and longitude information. The United States Census Bureau created and maintained this database, which offers detailed road networks (highways, local roads, and railroads) along with their classifications and area contexts. RL measurements were typically taken every 1/10th mile along most road segments. The datasets and relevant documents did not specify which RL measurement method was used. However, previous studies have indicated that LaserLux vans were usually used to collect RL data on freeways and four-lane roads, while handheld Retroreflectometer LTL-X devices were used on other roads (Ozelim et al., 2014; Smadi et al., 2008).

The original RL datasets included route identification numbers, mile markers, measured RL values, marking line types, measurement direction, date of RL measurements, geocoordinates, and subjective ratings. A sample of the original RL datasets is shown in **Figure 3**. In 2020, subjective ratings ranged from 1 to 4; in 2021, the scale was expanded from 1 to 5. Within urban road segments, a majority had no RL values on WREL due to the frequent minor road interactions. Consequently, this research focuses on only the rural road segments, as they are more pertinent to the type of collisions identified in the literature to be associated with RL, such as single-vehicle ROR crashes. **Figure 4** illustrates the rural sites where RL data were collected for both years and exclusively for 2021. In the original datasets, RL of each 0.1-mile segment was measured in September 2020 and November 2021.

ROUTE_ID	MILE_MARKER	RL	LINE	DIRECTION	DATE_COLLECTED	GPS_LAT	GPS_LONG	SUB_RATE
AL0000030000	178.6	168	LEL	NB	9/12/2020	32.32923667	-86.34622	3
AL0000030000	178.7	283	LEL	NB	9/12/2020	32.32957333	-86.34791833	3
AL0000030000	178.8	246	LEL	NB	9/12/2020	32.329905	-86.34959667	3
AL0000030000	178.9	199	LEL	NB	9/12/2020	32.33024333	-86.35122833	3
AL0000030000	179	195	LEL	NB	9/12/2020	32.330585	-86.35289	3
AL0000030000	179.1	167	LEL	NB	9/12/2020	32.33098	-86.35456167	3
AL0000030000	179.2	0	LEL	NB	9/12/2020	32.33142333	-86.35619167	3
AL0000030000	179.3	142	LEL	NB	9/12/2020	32.33195667	-86.35776833	3
AL0000030000	179.4	125	LEL	NB	9/12/2020	32.33258	-86.35931167	3
AL0000030000	179.5	128	LEL	NB	9/12/2020	32.33333833	-86.360865	3
AL0000030000	179.6	134	LEL	NB	9/12/2020	32.33411833	-86.362255	3
AL0000030000	179.7	135	LEL	NB	9/12/2020	32.335005	-86.36360833	3

Figure 3 A sample of original RL datasets collected from ALDOT

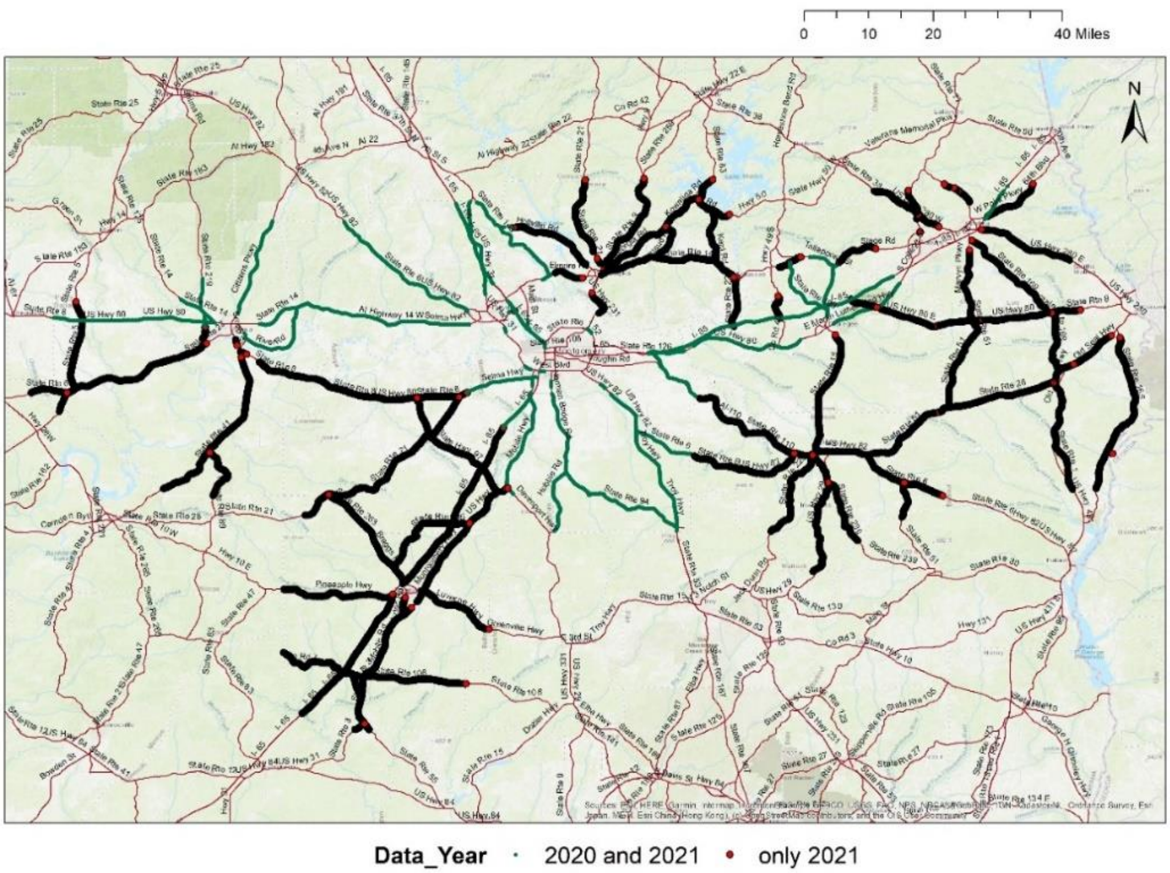

Figure 4 Rural locations of RL measurement in Montgomery

3.2 Crash data

The Critical Analysis Reporting Environment (CARE) system, developed and maintained by the Center for Advanced Public Safety (CAPS) at the University of Alabama, serves as a comprehensive repository documenting all traffic crashes within Alabama. This system provides extensive data, including location, road, environment, driver, and vehicle characteristics, among other details. In the context of this study, a few key road inventory variables available in the crash datasets include posted speed limit, location classification, functional class, road surface type, and number of lanes. This research obtained two years of crash records from 2020 to 2021. The geocoordinate of locations, route identification numbers, and mile marker information associated with crashes are vital for linking crash records to RL data.

3.3 Traffic Flow and Road Inventory Data

AADT was obtained from the ALDOT's traffic data website for the entire state of Alabama for 2020 and 2021 (ALDOT, 2023). The original dataset included route identification numbers, milepoints, and county details, which served as key reference points for accurately assigning AADT for each 0.1-mile road segment. Notably, the traffic flow datasets include information on the percentage of truck ADT, allowing for the calculation of truck AADT. However, including passenger vehicle AADT and truck AADT as separate predictors may influence the model due to their derivation from the same overall AADT variable. This setup often results in truck AADT displaying a negative correlation with RL degradation, given that passenger vehicle AADT typically shows a strong positive correlation with RL degradation. Consequently, researchers usually prefer to incorporate the percentage of trucks rather than the percentage of trucks in ADT. The percentage of trucks indicates the proportion of trucks within the total vehicle count over a specific period, while the percentage of trucks in ADT reflects the proportion of trucks within the average daily traffic. For these reasons, truck AADT was excluded to avoid multicollinearity. A sample of original traffic flow datasets collected from ALDOT's traffic data website is provided in **Figure 5**.

A	B	C	D	E	F	G	H	I	J
FID	TRNumbe	UCountyI	Station	RouteID	FromMeasur	ToMeasure	AADT	KFactor	DFactor
0	0	Randolph	527	AL000022	153.650857	156.8288534	1985	10	55
1	0	Randolph	526	AL000022	149.0878623	153.650857	1985	10	56
2	0	Washingto	809	AL000017	40.5991724	42.9983769	4721	10	53
3	0	Henry	533	AL000134	73.729	75.422107	648	10	52
4	0	Henry	530	AL000134	63.709	64.829	1575	11	67
5	0	Henry	138	AL000173	1.0907556	1.508	2748	10	68
6	0	Henry	136	AL000001	25.72405	26.241515	15733	9	54
7	0	Henry	143	IV0003200	0	0.3125047	2614	15	51
8	0	Henry	451	AL000134	58.885	59.22	3076	9	63

Figure 5 A sample of the original traffic flow datasets

Currently, there is no comprehensive documentation of road inventory information for Alabama. Relevant horizontal alignment information—such as the presence of curves, curve length, and chord length—was extracted from Google Earth Pro using the geocoordinates for each 0.1-mile road segment. This data was then processed in ArcGIS 'COGO' package to obtain additional metrics, including curve angle and radius. **Figure 6** illustrates a basic example of curve data entry and relevant outputs in ArcGIS.

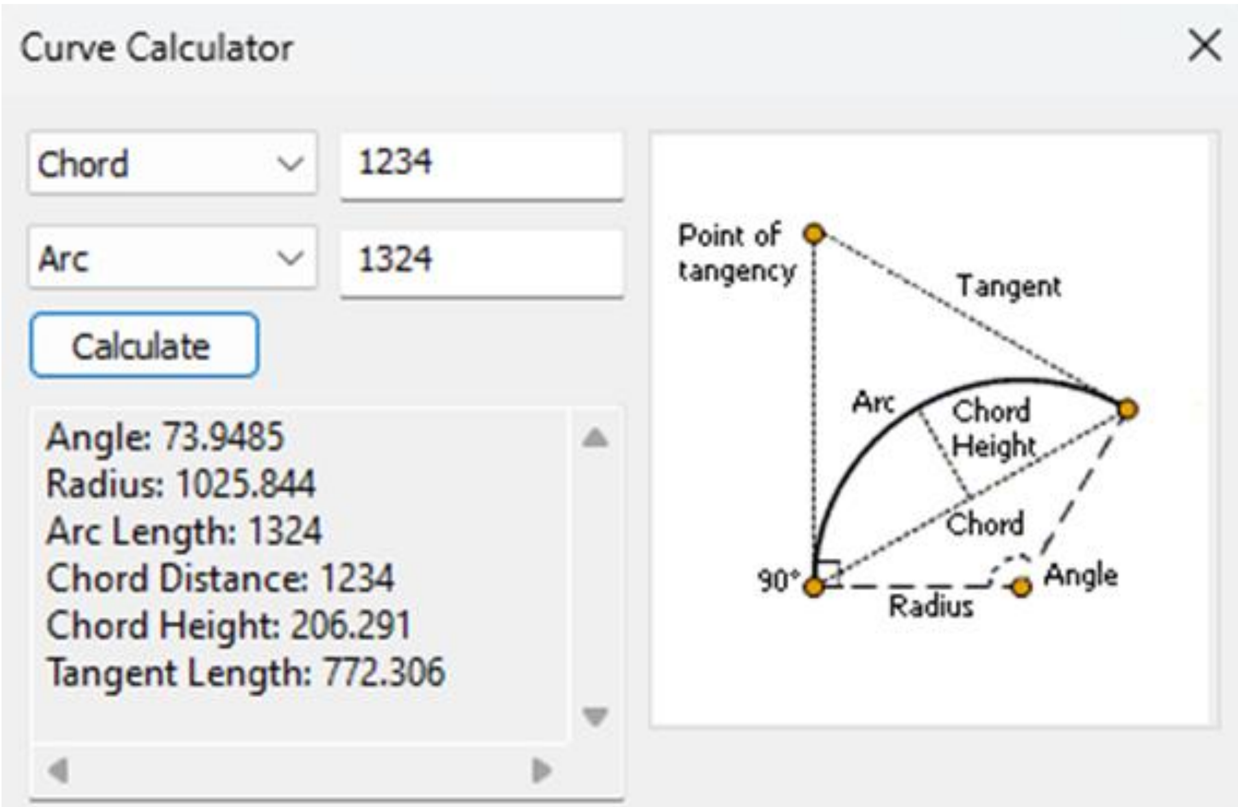

Figure 6 'COGO' package of ArcGIS for obtaining horizontal curve attributes

The Highway Performance Monitoring System (HPMS) is a nationwide program aimed at providing an inventory of all public roads in the U.S. (Federal Highway Administration (FHWA), 2023). HPMS encompasses various roadway feature data such as the number of lanes, functional class, speed limit, lane width, shoulder width, vertical alignment, and more. The database includes route identification numbers and mile marker information, which served as reference points for extracting the posted speed limits of each 0.1-mile segment. While assigning speed limits, it was assumed that the same posted speed limit was applicable across an entire 1-mile segment unless the AADT varied within that 1-mile segment. Posted speed limit data for 25% of the roadway segments was sourced from HPMS, and 45% was obtained from crash data. This information was further manually verified using Google Earth Pro. The remaining 30% of segments posted speed limit information were manually extracted from Google Earth Pro.

The crash database contains route identification numbers and mile marker information, which served as reference points for extracting the location class information for each 0.1-mile segment. It is important to mention that the location classes were categorized into two groups: open country and residential/business/mixed areas. For consistency, it was assumed that the same location class was applicable across an entire 1-mile segment unless AADT and posted speed limits varied within that segment. Location class data for 60% of the roadway segments was obtained from the crash datasets and subsequently cross-verified using Google Earth Pro. The remaining 40% of location class information was manually extracted from Google Earth Pro.

Road functional class information was extracted from HPMS and crash databases. Given HPMS's comprehensive details on road functional classes, there were a few instances with missing information. Additionally, the number of lane information was obtained from HPMS and Google Earth Pro. Notably, marking line type can also indicate the number of lanes, as two-lane roads typically have two marking lines (YCL and WREL), while multilane roads have three (YLEL, WLL, WREL). When extracting road surface type information from the crash datasets, it was found that over 95% of the segments were associated with asphalt surfaces, indicating that this was the predominant surface type. Therefore, road surface type was excluded from further analyses. **Figure 7** illustrates the extracted variables from each of the above-mentioned data sources.

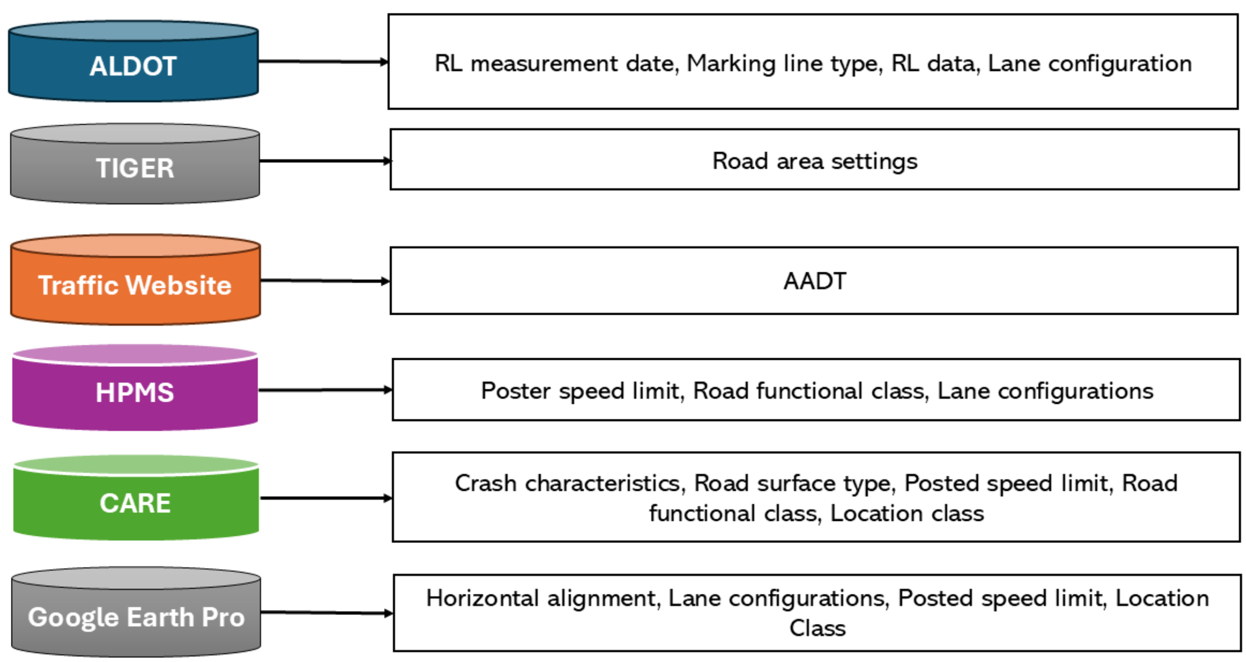

Figure 7 Extracted variables from different data sources

4 RL PREDICTION MODEL

This chapter focuses on developing regression models to predict RL for YCL and WREL. Initially, descriptive analyses are conducted to understand the distribution and characteristics of the datasets. Before constructing new models, an existing model is fitted to the datasets to evaluate its goodness of fit. Subsequently, new regression models are developed separately for YCL and WREL, considering different lane configurations. The practical applicability of these new models is then assessed, identifying the limitations in their implementation and potential improvements.

4.1 Data Preprocessing

Considerable effort has been dedicated to preprocessing the RL data before conducting the analysis. As mentioned earlier, this study exclusively focuses on rural roads. The initial step in processing the RL dataset obtained from ALDOT involves assigning the appropriate marking line for each measured RL. By manually reviewing lane configurations of each measured RL location using Google Earth Pro, the rural roads were categorized into two configurations: two-lane and multilane. For rural two-lane roads, the markings consist of YCL and WREL, while for multilane roads, the markings comprise YLEL, WLL, and WREL. While reviewing the original RL data, it was observed that the marking line types were inaccurately reported in a few instances, such as the transition from two-lane roads to four-lane roads or vice versa. For example, during a transition from two-lane to four-lane roads, the measured RL for centerlines was inaccurately reported as YCL when it should have been recorded as YLEL. Similarly, during transitions from four-lane to two-lane roads, the reverse misclassification was observed, with YLEL being reported as YCL. Approximately 15% of 0.1-mile segments contained these inaccuracies, as the errors persisted across multiple continuous segments. Such errors were corrected to ensure the accuracy of further analysis.

If the RL of a specific marking on a segment was lower in 2020 than 2021, the segment most likely was restriped during that period. Consequently, such segments were excluded from further analysis. In total, 175 and 112 one-mile segments were identified for rural two-lane and multilane respectively where no restriping activities had been performed. Adjusted AADT (Adj AADT) was computed for each 0.1-mile road segment based on the date of RL measurement. For instance, if RL was measured on September 12, 2020, and November 20, 2021, and the AADT for that road segment in 2020 was 2,455 vpd, and in 2021 was 2,800 vpd, the adj AADT would be calculated as follows:

$$Adjusted\ AADT\ (Adj\ AADT) = \frac{110 * 2{,}455 + 323 * 2{,}800}{110 + 323} = 2{,}713\ vpd$$

Where, 110 represents the number of days between September 12, 2020, and December 31, 2020, while 323 represents the number of days from January 1, 2021, to November 20, 2021. Most previous studies have calculated the average RL for each 1-mile segment to develop datasets for degradation modeling. However, this approach does not account for potential variations in road and traffic characteristics within a 1-mile segment. Recognizing this gap in data preparation, the current study has developed a more refined dataset for RL modeling, ensuring that variations in road, traffic flow, location, and marking characteristics are adequately considered. The detailed steps followed in preparing the dataset for this study are presented below:

- **Step 1**
  Identify 0.1-mile road segments with similar road, traffic flow, and location characteristics. This includes ensuring consistency in factors such as the number of lanes, posted speed limits, location classification, functional classification, horizontal alignment, and Adj AADT.
- **Step 2**
  Arrange each group of 0.1-mile segments based on their measured RL values in 2020.
- **Step 3**
  Calculate the moving average of the RL degradation proportion for each group of 0.1-mile segments by measured RL values in 2020 (RL_{2020}), following the qualitative states. During this process, it was observed that the degradation proportions exhibited considerable fluctuations in a few instances, even with minimal differences in RL_{2020}. For example, within the same group, a particular 0.1-mile segment exhibited RL degradation proportion equal to or more than twice, despite having minimal differences in the measured RL_{2020}. RL data points showing these unexpected changes were excluded to minimize heterogeneity within the final datasets. A threshold of a twofold or greater increase in RL degradation proportion was used for exclusion, based on analytical judgments. It is important to note that less than 5% of the total 0.1-mile segments belonged to this category.
  The above steps were applied to prepare the datasets for modeling RL degradation for YCL and WREL for rural two-lane and multi-lane road segments.

4.2 Descriptive Statistics

Table 6 presents the total number of observations in the final datasets by marking line type and the count of unique 1-mile road segments. Across all datasets, on average, there is more than one observation per 1-mile road segment. This indicates that, in most cases, each observation represents a segment smaller than 1 mile.

Table 6 Observation counts in final datasets for regression analysis

Marking line type	Number of observations	Count of different 1-mile segments
YCL	209	129
WREL for rural two-lane	347	167
WREL for rural multilane	235	108

The following observations provide insights based on **Table 7**, presenting an overview of variables used in modeling RL degradation proportion:

- In rural two-lane roads, the mean RL_{2020} values are higher for WREL (338 mcd/m²/lx) than for YCL (224 mcd/m²/lx). This is attributed to the relatively higher installation and restriping standards for white pavement markings, resulting in higher RL measurements.
- The data distribution indicates an expected pattern, with adj AADT being higher on multilane roads compared to two-lane roads.
- In the current version of MUTCD, the minimum RL requirement is based on speed limits of <35 mph or ≥35 mph. The earlier MUTCD version categorized speed limits into 35-50 mph and ≥55 mph for lane configurations. The RL dataset collected from ALDOT had very few segments with a posted speed limit below 35 mph; therefore, for this study, posted speed limit was categorized as <55 mph and ⩾55 mph for two-lane roads and <65 mph and ≥65 mph for multilane roads. Given that posted speed limit and road functional class are highly correlated (FHWA, 2000), only one of these variables was used in the regression modeling to satisfy associated assumptions.
- The diversity of data in residential, business, or mixed-use areas is greater for multilane roads (19.41%) compared to two-lane roads (14.57%). Conversely, data diversity on curved road segments is higher for two-lane roads (16.19%) than for multilane roads (10.84%).

Table 7 Overview of variables used in degradation proportion modeling by marking line type

Variable attribute	**Rural two-lane**		**Rural multi-lane**
	YCL (209)	**WREL (347)**	**WREL (235)**
RL_{2021} (mcd/m²/lux)	Min: 51, Max: 314, Median: 158, Mean: 163	Min: 73, Max: 447, Median: 240, Mean: 243	Min: 76, Max: 537, Median: 254, Mean: 255
RL_{2020} (mcd/m²/lux)	Min: 91, Max: 451, Median: 216, Mean: 224	Min: 94, Max: 711, Median: 331, Mean: 338	Min: 90, Max: 742, Median: 375, Mean: 379
Adj AADT (vpd)	Min: 546, Max: 11,220, Median: 3,556, Mean: 3,447	Min: 546, Max: 13,560, Median: 2,874, Mean: 3,443	Min: 3,610, Max: 85,257, Median: 16,258, Mean: 22,849
Posted speed limit			
<55mph	17 (8.13%)	46 (13.26%)	-----
≥55mph	192 (91.87%)	301 (86.74%)	-----
<65mph	-----	-----	74 (31.49%)
≥ 65mph	-----	-----	161 (68.51%)
Location class			
open country	178 (85.17%)	297 (85.59%)	184 (78.30%)
residential/business/mixed	31 (14.83%)	50 (14.41%)	51 (21.70%)
Horizontal alignment			
straight	188 (89.95%)	278 (80.12%)	210 (89.36%)
curve	21 (10.05%)	69 (19.88%)	25 (10.64%)

It is important to note that only two RL measurements of each location are available: one taken in September 2020 (RL_{2020}) and the other in November 2021 (RL_{2021}), providing a 14-month interval. It is noted that RL_{2020} may be a significant predictor of RL_{2021}, meaning the distribution of RL_{2020} can substantially influence model performance. **Table 8** provides the distribution of RL_{2020} for both YCL and WREL, with RL categorization based on the qualitative standards outlined in **Table 1**.

Table 8 Distribution of datasets by RL_{2020}

Marking line type	**Count of observations (Percentage of total observations)**		
	< 150 mcd/m²/lux	**150 - 249 mcd/m²/lux**	**>249 mcd/m²/lux**
WREL for rural two-lane	6 (1.73%)	56 (16.19%)	285 (82.13%)
WREL for rural multilane	8 (3.40%)	32 (13.62%)	195 (82.98%)
	< 100 mcd/m²/lux	**100 - 199 mcd/m²/lux**	**>199 mcd/m²/lux**
YCL for rural two-lane	5 (2.29%)	84 (40.19%)	120 (57.52%)

The distribution reveals that the datasets for YCL and WREL contain very few observations where the measured RL_{2020} values were low based on the qualitative states. Out of the total segments where RL data was collected in 2020, only a small proportion had RL_{2020} values below 150 mcd/m²/lux for WREL and below 100 mcd/m²/lux for YCL. This may be because segments are typically restriped before their RL values drop below these thresholds, which several earlier studies have identified as the failure point for pavement markings in terms of RL. For WREL, only 1.73% of observations on two-lane roads and 3.40% on multilane roads fall below 150 mcd/m²/lux. Similarly, only 2.29% of YCL observations on two-lane roads are below 100 mcd/m²/lux. Compared to WREL, YCL are more evenly distributed with respect to RL_{2020} values 100 mcd/m²/lux and above, with 40.191% of observations ranging between 100 - 199 mcd/m²/lux and 57.517% above 199 mcd/m²/lux.

4.3 Methodology

Earlier RL models utilized multiple linear regression (MNL) models as they provided a good fit for the datasets. The standard format for the linear prediction model can be stated as follows (Marill, 2004):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

Where, Y is the dependent variable; β_0 means intercept representing the expected value of Y when all X variables are zero; β_1, β_2, …..,and β_n are the coefficients for each independent variable; X_1, X_2, …., X_n are the independent variables; ϵ is the error term capturing the variability in Y not explained by the independent variables.

R^2 or the coefficient of determination is a measure used to assess the goodness-of-fit of an MNL model. It represents the proportion of the variance in the dependent variable that is predictable from the independent variables (Quinino et al., 2013).

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}}$$

Where, $SS_{residual}$ is the sum of squared residuals (differences between observed and predicted values); SS_{total} is the sum of squares (variance of observed values from the mean). An R^2 closer to 1 indicates a better fit, meaning the model explains a larger portion of the variance in Y.

4.4 Results and Discussions

Initially, the MNL models were applied to the existing model by Ozelim et al. (2014) to evaluate how well the previously developed equation fits the current dataset. The transformed equations from the earlier models for predicting R_{2021} for white and yellow markings are provided below:

White marking: $TM_w = R_{2020} - 71.82 - 0.00699 * (AADT_{2021} - AADT_{2020})$

Yellow marking: $TM_y = R_{2020} - 69.57 - 0.00217 * (AADT_{2021} - AADT_{2020})$

The calibrated model predicts the measured RL in 2021 based on the measured RL in 2020 and the AADT difference between 2021 and 2020. Subsequently, new models were developed specifically for YCL on two-lane segments, WREL on two-lane segments, and WREL on multilane segments. It is important to note that, while constructing these prediction models, only variables identified as the most relevant through Stepwise Akaike Information Criterion (stepAIC) were included to ensure relevance to each line type. The validity and limitations of these models were then discussed using practical examples to illustrate their applicability and potential constraints. It is important to note that since only two RL measurements were available for each 0.1-mile segment, the time interval for the RL prediction models was fixed at 14 months. This indicates that the predicted R_{2021} represents the RL of 14 months after the previous measurement, R_{2020}.

4.4.1 RL Prediction Model for YCL

Table 9 presents the linear regression results for YCL following the previous RL prediction model.

Table 9 Linear regression results for YCL using the previous RL prediction model

	Coef.	SE	t-value	p-value
(intercept)	74.6942	4.6589	16.032	<0.001
TM_y	0.5957	0.0272	20.773	<0.001
R^2 = 0.6752				

Note: Coefficient (Coef.) and Standard Error (SE)

The RL prediction model for yellow markings developed by Ozelim et al. (2014) shows a good fit with the dataset, achieving an R^2 value of 0.6752, which explains 67.52% of the overall variability.

Table 10 presents the MNL results for YCL on rural two-lane roads, aiming to develop a new RL prediction model without initial RL and marking age. Including only R_{2020} in a simple linear regression model accounts for 65.61% of the variability. Consistent with the degradation proportion model for YCL (**Table 9**), all selected variables are found to be statistically significant. A one-unit increase in R_{2020} (mcd/m²/lux) leads to a 0.57 mcd/m²/lux increase in R_{2021} when other predictors are constant. Additionally, 1,000 units increase in AADT decreases R_{2021} by 1.20 mcd/m²/lux, indicating a statistically significant correlation between AADT and RL degradation. For roads in residential, business, or mixed-use areas, R_{2021} decreases by 14.03 mcd/m²/lux, when all other variables remain constant. It is worth noting that, although residential/business/mixed areas are found to be statistically significant in the model, their inclusion marginally improves model performance, as R^2 increases slightly. Following this, the final RL prediction model is provided below:

$$R_{2021} = 39.35 + 0.578 * R_{2020} - 0.0011 * \text{Adj AADT}$$

Table 10 MNL results for YCL

	Coef.	**SE**	**t-value**	**p-value**
(intercept)	44.5800	7.8472	5.680	<0.001
R_{2020} (mcd/m²/lux)	0.5684	0.0284	19.578	<0.001
Adj AADT (vpd)	-0.0012	0.0013	-5.401	<0.001
Location class (ref. open country)				
residential/business/mixed	-14.0312	5.8432	-2.036	0.043
R^2 = 0.6824				
Equation				$\boldsymbol{R^2}$
$R_{2021} = 44.58 + 0.568 * R_{2020} - 0.0012$ * Adj AADT – 14.03 * (location class @ residential/business/mixed)				0.6824
$R_{2021} = 39.35 + 0.578 * R_{2020} - 0.0011$ * Adj AADT				0.6785

Note: Coefficient (Coef.) and Standard Error (SE)

4.4.2 RL Prediction Model for WREL (Rural Two-lane)

Table 11 shows the linear regression results for WREL on rural two-lane following the previous RL prediction model. The RL prediction model for white markings developed by Ozelim et al. (2014) shows a good fit with the dataset, explaining 59.34% of the overall variability.

Table 11 Linear regression results for WREL (rural two-lane) using the earlier RL prediction model

	Coef.	SE	t-value	p-value
(intercept)	83.8771	7.5372	11.131	<0.001
TM_w	0.5640	0.0273	22.446	<0.001
R^2 = 0.5934				

Note: Coefficient (Coef.) and Standard Error (SE)

Table 12 presents the MNL results for WREL on rural two-lane roads, aiming to develop a new RL prediction model for WREL without initial RL and marking age.

Table 12 MNL results for WREL (rural two-lane)

	Coef.	SE	t-value	p-value
(intercept)	63.4151	13.7742	5.578	<0.001
R_{2020} (mcd/m²/lux)	0.5613	0.0277	21.303	<0.001
Adj AADT (vpd)	-0.0049	0.0016	-4.673	<0.001
R^2 = 0.6153				
Equation				$\boldsymbol{R^2}$
$R_{2021} = 63.42 + 0.561 * R_{2020} - 0.0049$ * Adj AADT				0.6153

Note: Coefficient (Coef.) and Standard Error (SE)

A simple linear regression model that includes only R_{2020} explains 58.55% of the variability. Both R_{2020} and Adj AADT are found to be statistically significant predictors. A one mcd/m²/lux increase in R_{2020} results in a 0.56 mcd/m²/lux increase in R_{2021} when other predictors remain constant. Moreover, an increase of 1,000 units in AADT decreases R_{2021} by 4.90 mcd/m²/lux, indicating a statistically significant association between AADT and RL degradation. Notably, the absolute magnitudes of influence regarding R_{2020} and Adj AADT are higher for WREL than for YCL. Consistent with earlier studies, this highlights that yellow markings experience less RL degradation over time compared to white markings. The final RL prediction model for WREL on rural two-lane road segments is provided below:

$$R_{2021} = 63.42 + 0.561 * R_{2020} - 0.0049 * \text{Adj AADT}$$

4.4.3 RL Prediction Model for WREL (Rural Multilane)

Table 13 represents the linear regression results for WREL on rural multilane using the previous RL prediction model for white marking (Ozelim et al., 2014). The earlier RL prediction model shows a good fit with the dataset, explaining 58.92% of the overall variability. The R^2 value for multilane roads is slightly lower than that for two-lane roads, likely due to the limited number of segments on multilane in the previously developed model.

Table 13 Linear regression results for WREL (rural multilane) using the previous RL prediction model

	Coef.	**SE**	**t-value**	**p-value**
(intercept)	88.4553	9.5126	10.035	<0.001
TM_w	0.4782	0.0293	18.149	<0.001
R^2 = 0.5892				

Note: Coefficient (Coef.) and Standard Error (SE)

Table 14 shows the MNL results for WREL on rural multilane roads, aiming to predict RL without initial RL and marking age.

Table 14 MNL results for WREL (rural multilane)

	Coef.	**SE**	**t-value**	**p-value**
(intercept)	66.7492	13.5372	3.314	0.001
R_{2020} (mcd/m²/lux)	0.4642	0.0288	18.383	<0.001
Adj AADT (vpd)	-0.0007	0.0003	-3.455	<0.001
R^2 = 0.6034				
Equation				$\boldsymbol{R^2}$
$R_{2021} = 66.75 + 0.464 * R_{2020} - 0.0007$ * Adj AADT				0.6034

Note: Coefficient (Coef.) and Standard Error (SE)

A simple linear regression model that includes only R_{2020} explains 58.22% of the variability. Both R_{2020} and Adj AADT are found to be statistically significant predictors. A one mcd/m²/lux increase in R_{2020} results in a 0.46 mcd/m²/lux increase in R_{2021} when other predictors remain constant. Moreover, an increase of 1,000 units in AADT decreases R_{2021} by 0.7 mcd/m²/lux, indicating a statistically significant positive correlation between AADT and RL degradation. It is noteworthy that the impact of AADT in RL degradation for WREL is lower on rural multilane compared to rural two-lane. This is because, on multilane roads, the total vehicle volume is distributed across multiple lanes, which spreads out the wear and reduces pressure on individual markings. The final RL prediction model for WREL on rural multilane road segments is provided below:

$$R_{2021} = 66.75 + 0.464 * R_{2020} - 0.0007 * \text{Adj AADT}$$

4.4.4 Model Limitations

Multiple validation steps are required before finalizing the models to enhance their practical applicability. An established finding in pavement marking research is that yellow markings generally degrade slower compared to white markings. To assess whether the developed models align with this basic concept, the following estimations have been conducted. At both high and low RL levels, yellow markings exhibit lower degradation compared to white markings, which aligns with findings from previous studies.

i) Assume, measured RL of any given time (RL_t) of YCL = 300 mcd/m²/lux, AADT = 10,000 vpd, and road setting = rural two-lane
Predicted RL after 14 months (RL_{t+14}) $= 39.35 + 0.578 * 300 - 0.0011 * 10{,}000 = 201.75 \approx 202$ mcd/m²/lux.

ii) Assume, measured RL of any given time (RL_t) of YCL = 100 mcd/m²/lux, AADT = 10,000 vpd, and road setting = rural two-lane
Predicted RL after 14 months (RL_{t+14}) $= 39.35 + 0.578 * 100 - 0.0011 * 10{,}000 = 86.15 \approx 87$ mcd/m²/lux.

iii) Assume, measured RL of any given time (RL_t) of WREL = 300 mcd/m²/lux, AADT = 10,000 vpd, and road setting = rural two-lane
Predicted RL after 14 months (RL_{t+14}) $= 63.42 + 0.561 * 300 - 0.0049 * 10{,}000 = 182.72 \approx 183$ mcd/m²/lux.

iv) Assume, measured RL of any given time (RL_t) of WREL = 100 mcd/m²/lux, AADT = 10,000 vpd, and road setting = rural two-lane
Predicted RL after 14 months (RL_{t+14}) $= 63.42 + 0.561 * 100 - 0.0049 * 10{,}000 = 70.52 \approx 71$ mcd/m²/lux.

Figure 8, Figure 9, and Figure 10 show the normal QQ plots and residual vs. fitted plots for the three RL prediction models: YCL, WREL on two-lane roadways, and WREL on multilane roadways, respectively. These plots provide insights into the distribution of residuals and the fit of each model, allowing for a visual assessment of model assumptions. In each case, a few outliers are present in the datasets. Most residuals (represented by the red lines in the residual plots) are close to the dashed line, indicating a high degree of linearity in the prediction model. However, the residual variance changes across the range of fitted values, a phenomenon known as heteroscedasticity. To evaluate the predictive performance of each model in estimating the service life of pavement marking RL, each model was tested across various AADT ranges.

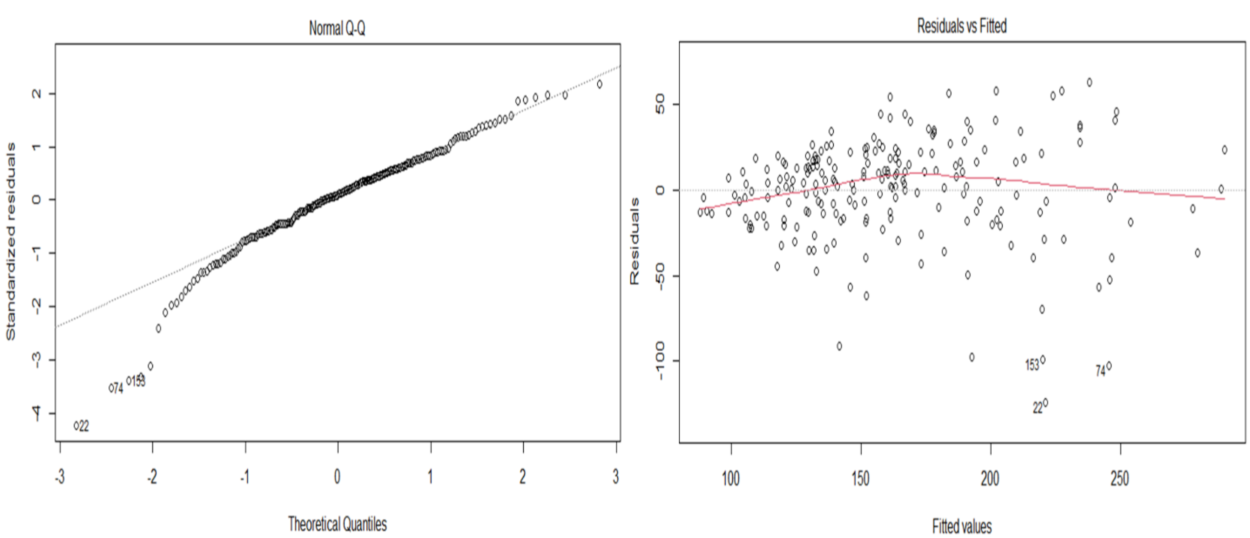

Figure 8 QQ and Residual plots for YCL model (rural two-lane)

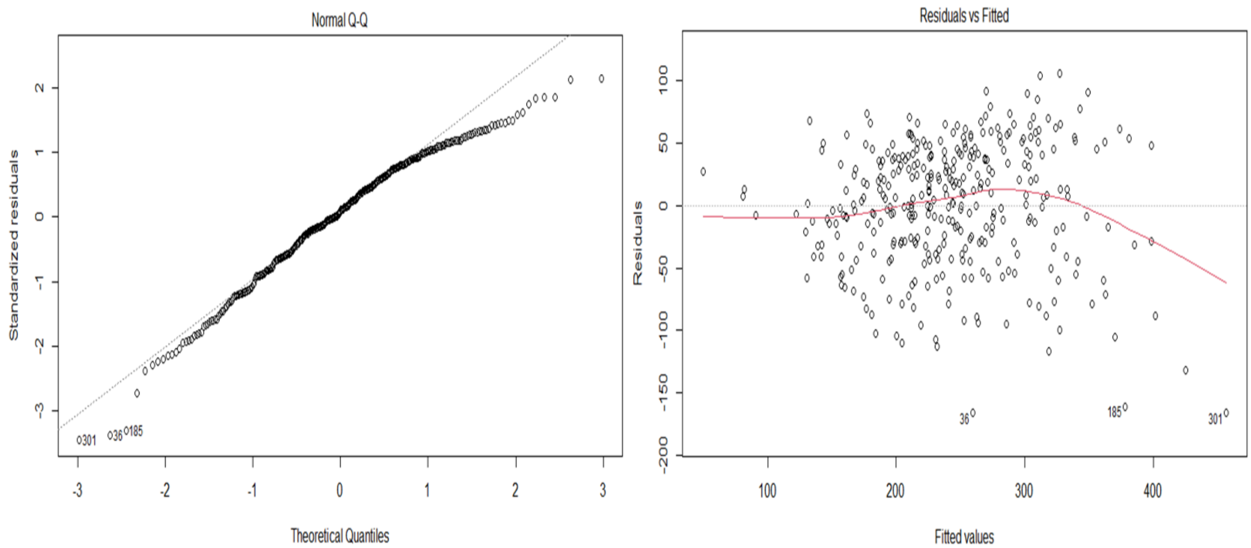

Figure 9 QQ and Residual plots for WREL model (rural two-lane)

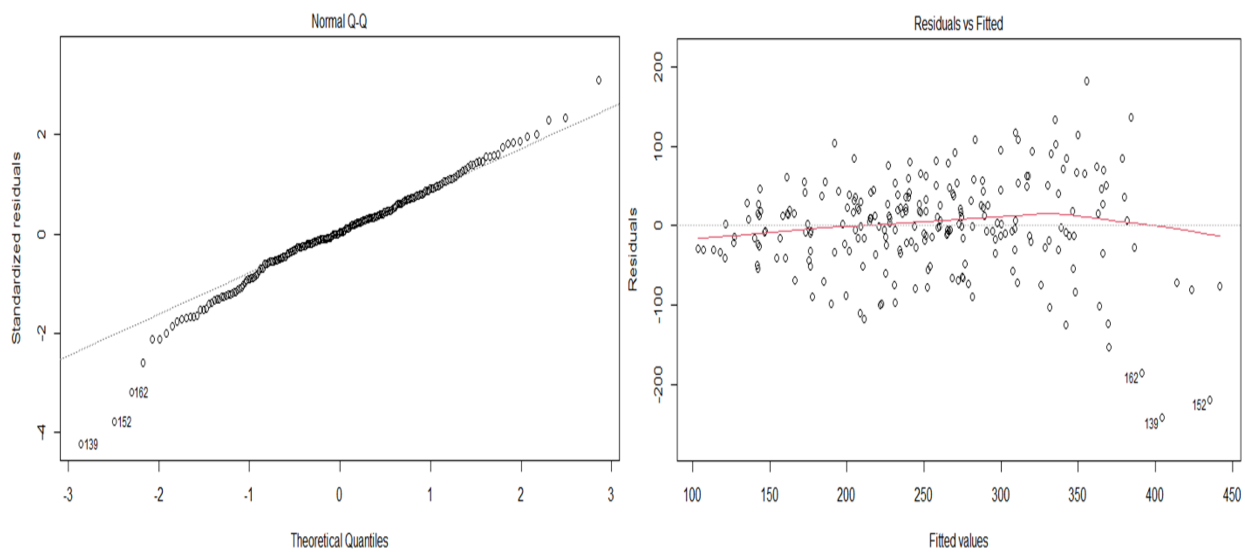

Figure 10 QQ and Residual plots for WREL model (rural multilane)

In RL predictive modeling, estimating service life can help to reveal the limitations of regression models. Since the MUTCD has lowered the minimum RL requirement to 50 mcd/m²/lux, it is crucial to have enough RL data observations within the 50-100 mcd/m²/lux range across diverse traffic volumes to develop a model capable of better predicting RL below this threshold. However, in the RL datasets, very few observations fall below 100 mcd/m²/lux for YCL (below 3%) and below 150 mcd/m²/lux for WREL (less than 4%). As a result, it is anticipated that the models developed in this research may not perform well when the RL_t levels fall within such ranges. Ozelim et al. (2014) also highlighted such limitations in their study, noting unexpected results while calculating service life using the regression equations of earlier studies. In this study,

Figure 11-Figure 16 illustrate the predicted RL values for YCL and WREL based on minimum RL installation standards and diverse AADT ranges. The two primary issues are: i) RL degradation becomes zero at a certain point, resulting in no further decline, and ii) percentage of degradation is not stable after a certain RL, specifically when RL_t below 150 mcd/m²/lux for WREL and below 100 mcd/m²/lux for YCL. For yellow markings, the RL_{t+14} predictions for YCL are less reliable when RL_t values fall below 100 mcd/m²/lux. Similarly, for white markings, the RL_{t+14} predictions for WREL are less reliable when RL_t values fall below 150 mcd/m²/lux, for both two-lane and multilane road segments. One primary reason for such limitations is the scarcity of road segments at these low RL levels across a range of traffic volumes, leading to limited data distribution. However, this limitation may have minimal practical impact, as segments are often restriped before reaching 100 mcd/m²/lux for yellow markings and 150 mcd/m²/lux for white markings. With recent changes in the MUTCD, it is increasingly important to develop models that can more accurately predict future RL levels below 100 mcd/m²/lux, as the minimum maintained RL threshold has been reduced from 100 mcd/m²/lux to 50 mcd/m²/lux for certain road segments.

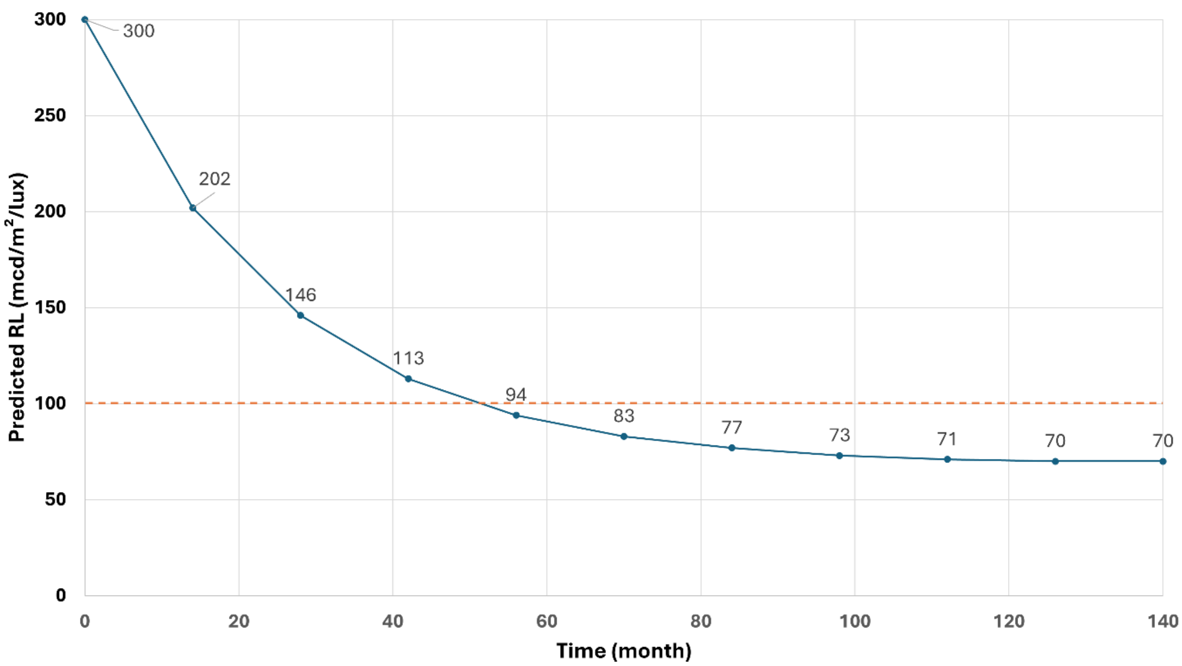

Figure 11 Predicted RL of YCL by month ($\boldsymbol{RL_t}$ = 300 mcd/m²/lux and AADT = 10,000 vpd)

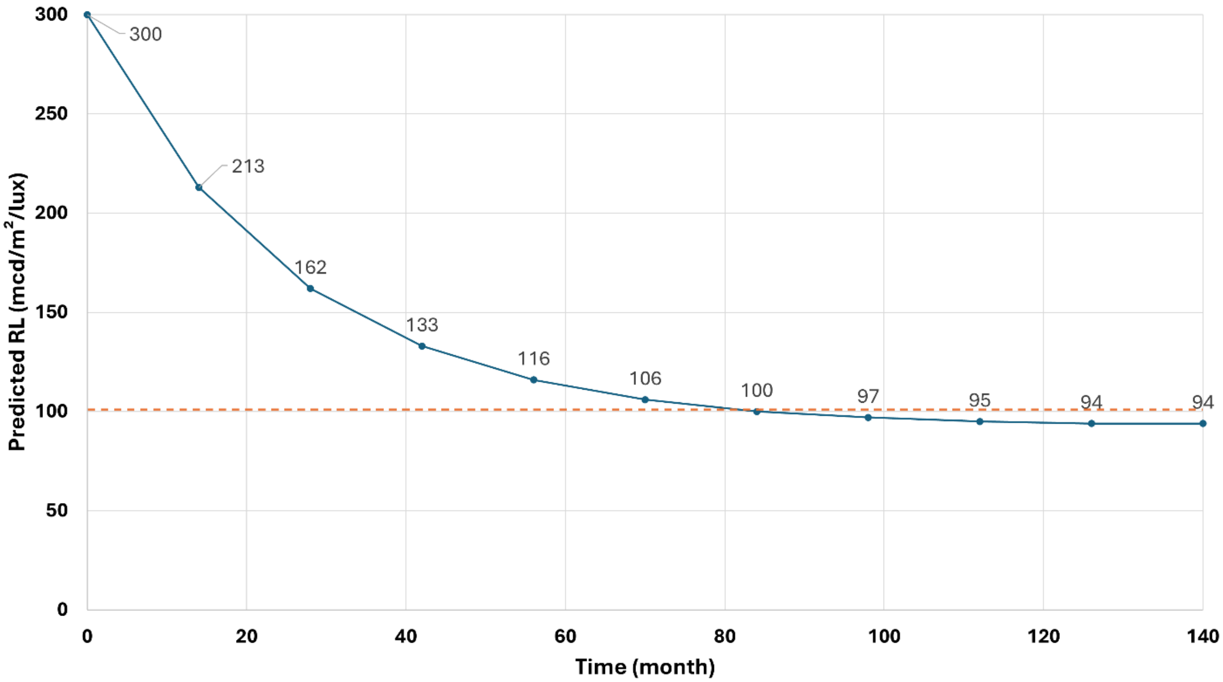

Figure 12 Predicted RL of YCL by month ($\boldsymbol{RL_t}$ = 300 mcd/m²/lux and AADT = 546 vpd)

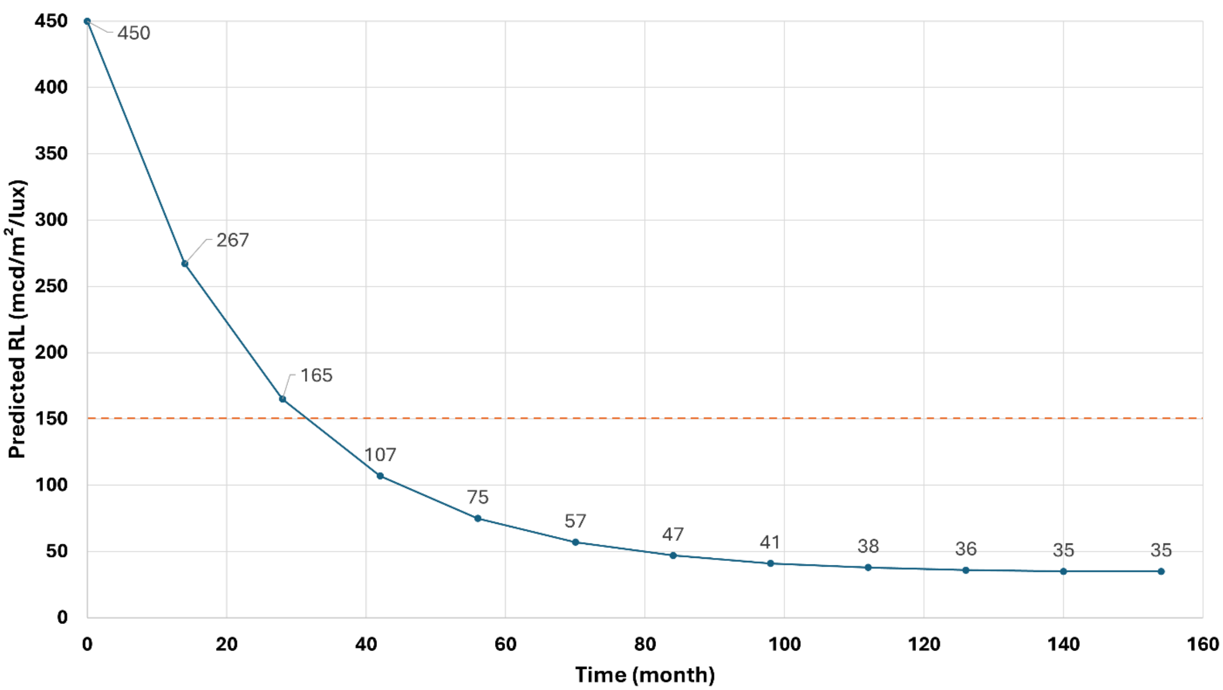

Figure 13 Predicted RL of WREL (rural two-lane) by month ($\boldsymbol{RL_t}$ = 450 mcd/m²/lux and AADT = 10,000 vpd)

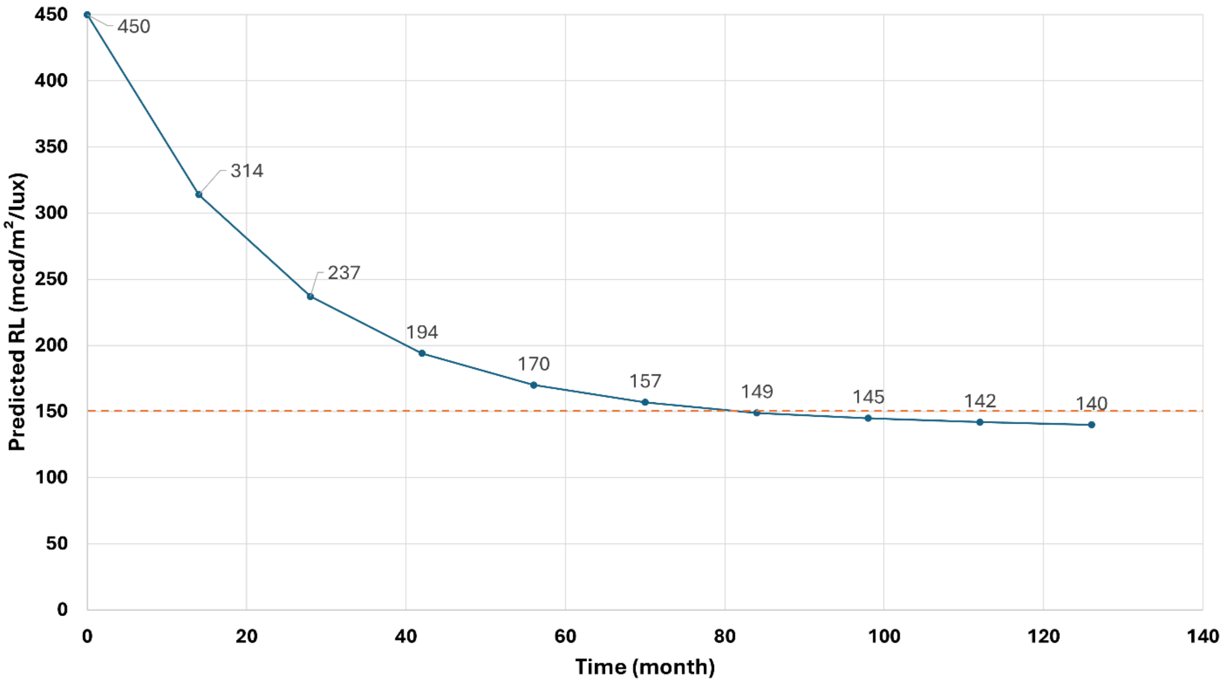

Figure 14 Predicted RL of WREL (rural two-lane) by month ($\boldsymbol{RL_t}$ = 450 mcd/m²/lux and AADT = 546 vpd)

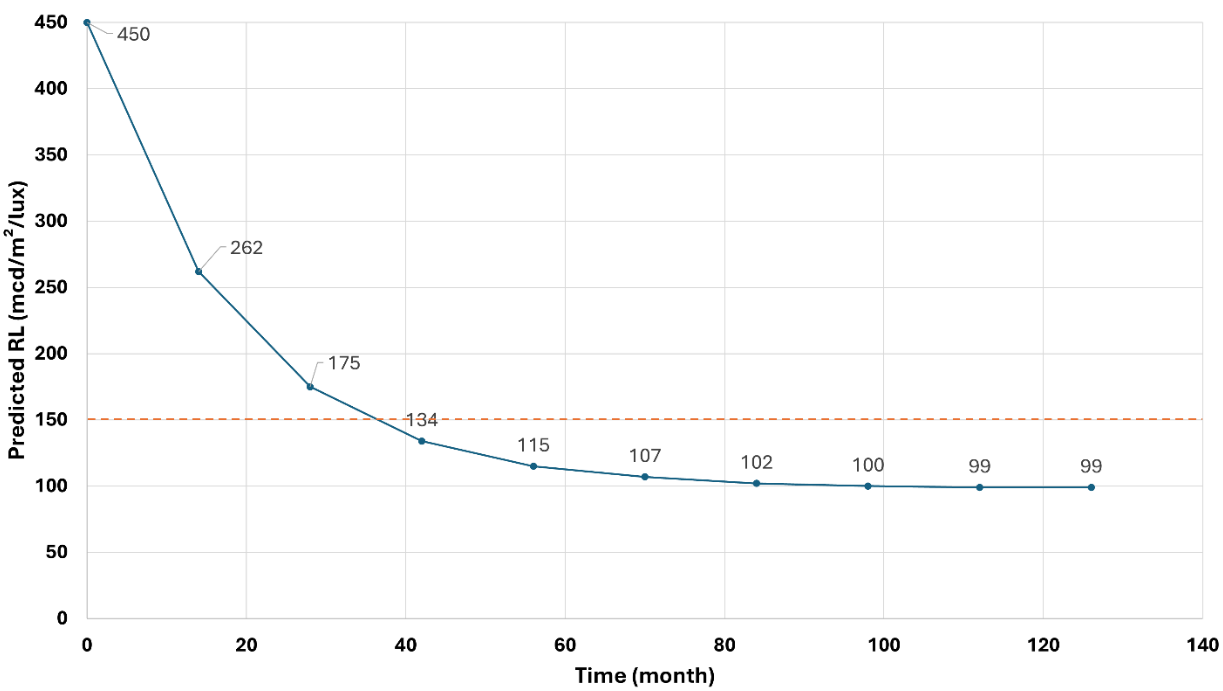

Figure 15 Predicted RL of WREL (rural multilane) by month ($\boldsymbol{RL_t}$ = 450 mcd/m²/lux and AADT = 20,000 vpd)

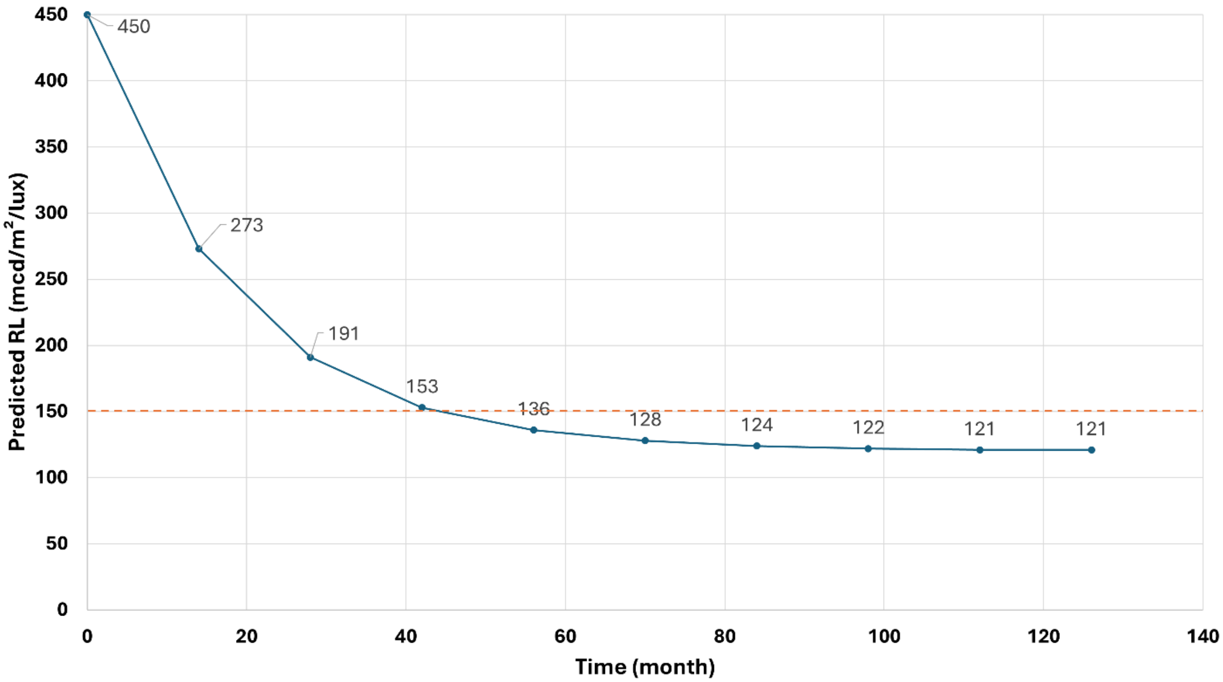

Figure 16 Predicted RL of WREL (rural multilane) by month ($\boldsymbol{RL_t}$ = 450 mcd/m²/lux and AADT = 3,610 vpd)

4.5 Key Findings

This chapter focuses on developing RL prediction models for YCL and WREL without requiring initial RL and pavement marking age information. Using a multiple linear regression approach, this study addresses a gap in the literature by providing separate regression models for WREL on different rural road lane configurations. The newly developed models demonstrate improved goodness-of-fit compared to the calibrated models from a previous Alabama study. However, a limitation arises due to the limited sample size, particularly in lower RL levels, which constrains the accuracy of predicting the service life of pavement markings. The resulting RL prediction equations, along with guidance on their application, are presented below:

YCL (rural two-lane): $RL_{t+14} = 39.35 + 0.578 * RL_t - 0.0011 * AADT$

WREL (rural two-lane): $RL_{t+14} = 63.42 + 0.561 * RL_t - 0.0049 * AADT$

WREL (rural multilane): $RL_{t+14} = 66.75 + 0.464 * RL_t - 0.0007 * AADT$

Where, RL_t is the measured RL at any given time, RL_{t+14} is the predicted RL after 14 months of *t*. For better accuracy in RL measurements, it is recommended to use these models when $RL_t \geq 100$ mcd/m²/lux for yellow markings and $RL_t \geq 150$ mcd/m²/lux for white markings. Additionally, it is recommended that adj AADT be calculated if AADT values vary between years.

5 RELATIONSHIP BETWEEN RL AND SAFETY PERFORMANCE

This section aims to examine the statistical relationship between RL and specific types of crashes, including single-vehicle ROR collisions, crashes occurring in dark conditions, and crashes at night with no street lighting. Since RL measurements for 856 miles of rural road segments were only taken in 2021, prediction models for YCL and WREL have been used to estimate RL for 2020. Within the majority of 1-mile segments, significant variability in measured RL along with road and traffic flow characteristics were observed, highlighting the importance of analyzing shorter segments to more accurately assess safety benefits. Consequently, average RL values are calculated for each 0.5-mile segment, and crash counts of those segments are compiled over a defined period. Finally, regression models are applied to investigate any statistically significant relationship between RL and relevant crash types.

5.1 Data Preprocessing

Investigating the safety implications of pavement marking RL reveals a complexity that goes beyond initial assumptions. RL fluctuates notably due to diverse factors such as location and environmental conditions, complicating safety assessments. To address these complexities, this study focuses on the RL values of WREL and YCL markings on rural two-lane and multilane roads. The objective is to evaluate the statistical correlation between these RL values and road crashes, providing a clearer understanding of their impact on road safety. There were 856 1-mile rural road segments with RL measurements only available in November 2021. To estimate RL values for 2020 on these segments, prediction models for WREL on two-lane and multilane roads and YCL on two-lane roads, developed in Chapter 4, were utilized. Since road segments are more frequently restriped during summer, and to accurately assess the statistical relationship between crashes and RL without the seasonal effects of winter, this analysis focused on a period between March 2021 and November 2021 (mid-spring to late fall). The following steps are performed to assign average RL values to the road segments and select the appropriate road segments for further analysis.

- *Segments with measured RL data in September 2020 and November 2021:* Assuming a linear degradation of RL, the RL were calculated for March 2021. Restriped 0.1-mile segments were removed while comparing two years of measurements. Each 0.1-mile segment was then assigned an average RL value calculated using the following formula:

$$Average\ RL = \frac{RL\ in\ March\ 2021\ + RL\ in\ November\ 2021}{2}$$

An example is provided to enhance clarity. For instance, consider a specific 0.1-mile segment where the measured RL in September 2020 was 240 mcd/m²/lx, and in November 2021, it was 140 mcd/m²/lx. Using this information, the RL in March 2021, as well as the average RL between March 2021 and November 2021, can be calculated as follows:

$$RL\ in\ March\ 2021 = 240 - \left(\frac{100}{14}\right) * 6 = 197\ mcd/m^2/lx$$

$$Average\ RL = \frac{197\ +140}{2} = 169\ mcd/m^2/lx$$

- *Segments with measured RL only in November 2021*: The developed RL prediction models for WREL and YCL were used to estimate the RL values in September 2020 as well as March 2021. To ensure the exclusion of restriped segments, thresholds were established beyond which segments are considered restriped during the summer. In Alabama, the minimum RL values required for white and yellow thermoplastic pavement marking for installation and restriping are 450 mcd/m²/lx and 300 mcd/m²/lx, respectively. Hence, for added precaution, only segments with RL values ≤ 425 mcd/m²/lx for WREL and ≤ 275 mcd/m²/lx for YCL in March 2021 were considered for further analyses. Subsequently, each segment was assigned an average RL value between March 2021 and November 2021 using the methodology outlined earlier.

For each 0.1-mile segment, relevant data—including AADT, location classification, road alignment, functional classification, and posted speed limit—were assigned following the procedure outlined in the section 3. To address variability in road, traffic flow, and location characteristics more accurately, RL values were assigned to shorter segments. Therefore, 0.5-mile segments were chosen to improve analytical precision. 0.1-mile segments were compiled into a 0.5-mile segment, with the calculated average RL. Since a single curve can span across two 0.5-mile segments, a variable was created to indicate the presence of a curve within each 0.5-mile segment. Although RL was measured in one specific direction of traffic flow, both directions were assumed to have the same RL value to maximize the number of crashes attributed to each 0.5-mile segment.

Table 15 presents the number of 0.5-mile segments incorporated into the safety analysis using RL prediction models.

Table 15 Summary of 0.5-mile segments for safety analyses in relation to WREL

Lane configuration	Total segment	Segments attained from actual RL measurements	Segments attained from predictive modeling
Rural two-lane	1,411	288	1,123
Rural multilane	517	152	365

More than three times 0.5-mile segments from rural two-lane roads and more than two times 0.5-mile segments from rural multilane roads were included using RL prediction models. **Figure 17** illustrates the distribution of 0.5-mile segments on rural two-lane and multilane roads by RL of WREL: < 250 mcd/m²/lux and ≥ 250 mcd/m²/lux. In both road settings, a higher percentage of segments fall into the ≥ 250 mcd/m²/lux category, with 65.13% on rural two-lane roads and 56.78% on rural multilane roads.

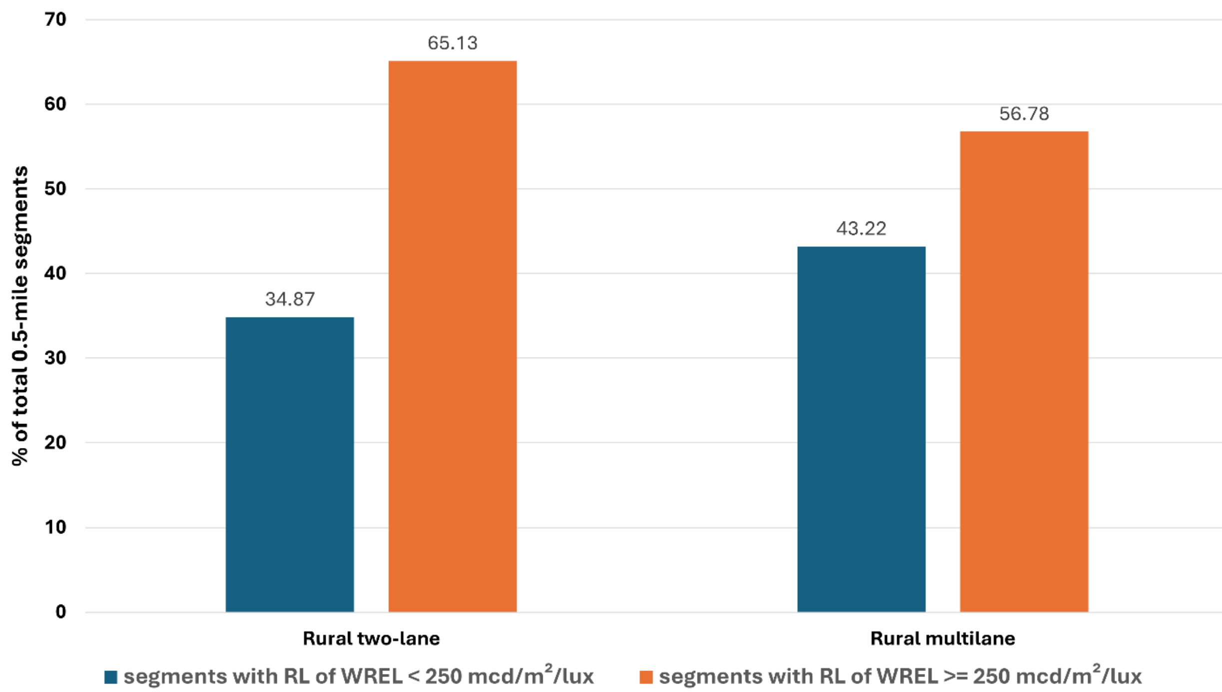

Figure 17 Distribution of 0.5-mile segments by RL of WREL

While extracting crash data, only crashes occurring between March 2021 and November 2021—covering the Mid Spring to Late Fall seasons—were included. Each measured RL had geocoordinate information, so crashes occurring within 250 feet of a specific 0.1-mile segment were considered to belong to that segment. Crashes involving alcohol impairment were excluded, aligning with previous studies (Donnell et al., 2009; Smadi et al., 2008). Alcohol impairment significantly reduces reaction time, judgment, and attentiveness, resulting in crashes more directly attributed to human error rather than external visibility factors like RL. Both driver condition and primary contributing factors were used to identify and exclude alcohol-impaired crashes. Additionally, intersection crashes were excluded due to the variety of complex factors at intersections—such as traffic signals, signage, pedestrian crossings, and multidirectional vehicle movements—that go beyond pavement marking visibility (Carlson et al., 2013; Masliah et al., 2007). Furthermore, when assigning intersection crashes to the corresponding 0.1-mile segments, misassignments can occur due to the high likelihood of including crashes from minor roads within a 250-foot radius of the RL measurement point on the major road. An example is provided in **Figure 18**.

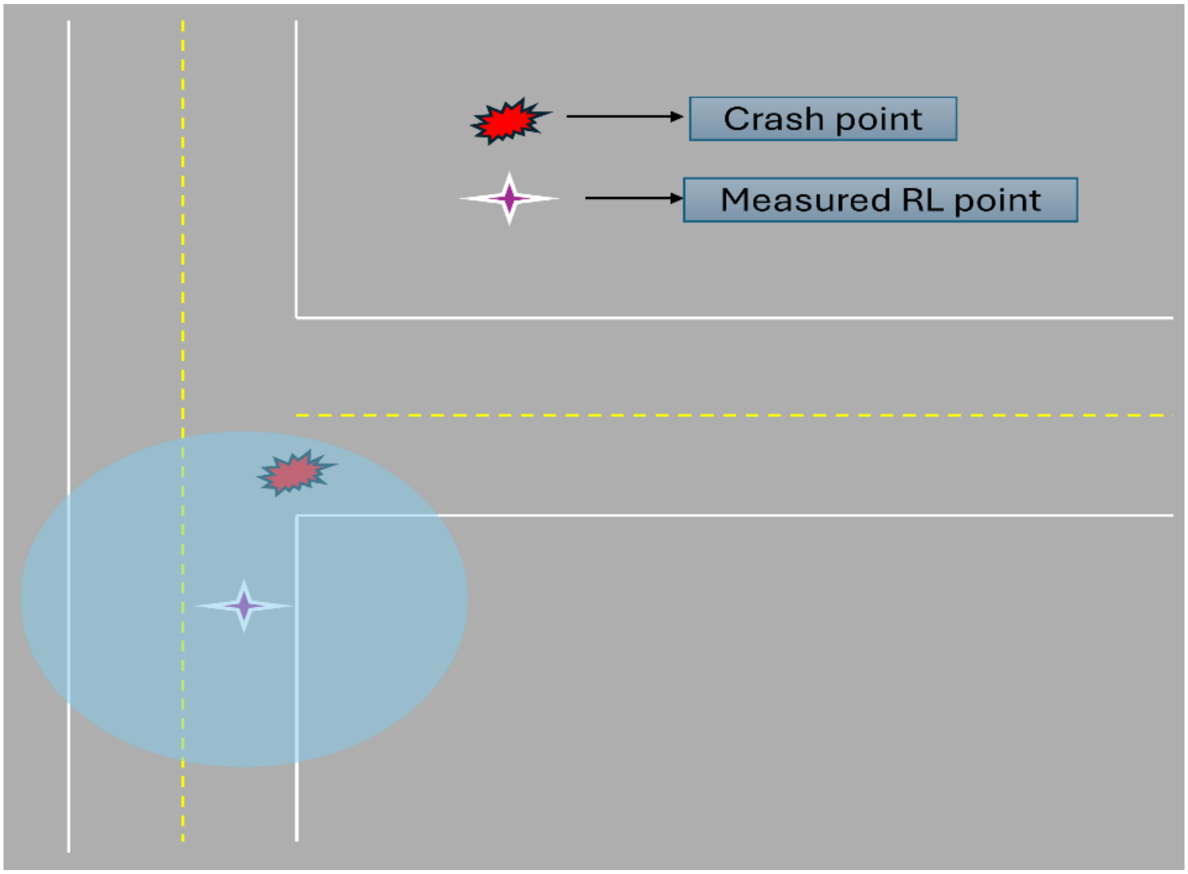

Figure 18 Potential complexities of including intersection crashes

This study focuses specifically on single-vehicle ROR crashes, which are more likely to be influenced by low RL of WREL. The criteria for selecting single-vehicle ROR crashes are detailed below.

- Exclude from first/most harmful events: crashes with other vehicles, crashes with parked vehicles or pedestrians, crashes with animals, cross-centerline crashes, equipment malfunctions, and other crashes that potentially happened on the roadway.
- Number of vehicle: One (also cross-checked with collision type- single vehicle/non-collision)

The combined effect of WREL and YCL might influence nighttime crashes with no street lighting conditions. Therefore, in addition to single-vehicle ROR crashes, crashes occurring in dark conditions with no streetlighting were also considered to assess the impact of interactions between RL levels of YCL and WREL. In such cases, crashes involving parked vehicles or pedestrians, collisions with animals, and equipment malfunctions were excluded; however, cross-centerline crashes and collisions with other vehicles were included. In total, five types of crashes were considered for further analysis, as listed below:

- Single vehicle ROR crashes (rural two-lane and multilane separately)
- Single vehicle ROR crashes in dark, dusk, and dawn conditions (rural two-lane and multilane separately)

- Single vehicle ROR crashes on dark no streetlighting conditions (rural two-lane and multilane separately)
- Crashes in dark no streetlighting conditions (rural two-lane only)
- Single vehicle ROR crashes on curve segments (merging rural two-lane and multilane)

Notably, for single-vehicle ROR crashes occurring on curve segments, the average RL and crash count for each 0.1-mile segment were analyzed, incorporating curve radius and angle as additional covariates. This approach allows for a more comprehensive understanding of the combined impact of curve geometry and RL on crash frequency and the likelihood of crashes.

5.2 Descriptive Analysis

Given the limitations of RL prediction models in estimating lower RL levels, categorizing average RL values into multiple groups offers an effective approach for RL safety analysis. On rural two-lane roads, a small fraction of 0.5-mile segments had RL of WREL below 150 mcd/m²/lux. When factoring in the presence of crashes, the number of such segments was even smaller. Therefore, for WREL on rural two-lanes, RL values are grouped into <250 mcd/m²/lux and ≥250 mcd/m²/lux. For multilane roads, a more detailed categorization is feasible due to a relatively higher proportion of crashes in segments with lower RL values. Categories for multilane roads include <250 mcd/m²/lux and ≥250 mcd/m²/lux, as well as broader categories of <150 mcd/m²/lux, 150-249 mcd/m²/lux, and ≥250 mcd/m²/lux. On rural two-lanes, while analyzing crashes at dark with no streetlighting conditions, the RL of YCL is categorized as <200 mcd/m²/lux and ≥200 mcd/m²/lux. Additionally, a separate analysis considers the interaction between RL of WREL and YCL, with the following interaction categories:

- WREL ≥ 250 mcd/m²/lux and YCL ≥ 200 mcd/m²/lux
- WREL ≥ 250 mcd/m²/lux and YCL < 200 mcd/m²/lux
- YCL ≥ 200 mcd/m²/lux and WREL < 250 mcd/m²/lux
- YCL ≥ 200 mcd/m²/lux and WREL ≥ 250 mcd/m²/lux

Table 16 to **Table 18**

Table 18 and **Figure 19** to **Figure 23** present an overview of the crash datasets prepared for RL safety analysis. Key observations from these datasets are outlined below:

- The majority of rural two-lane segments are located in open country areas (92.63%) and have posted speed limits of 55 mph or higher (90.645%). Compared to segments with actual RL measurements, the segments included in predictive modeling contain a higher proportion of curves, resulting in a good percentage of curve segments (28.99%) in the overall dataset. On rural two-lanes, the total crash counts by all targeted crash types are equal to or slightly higher than the number of 0.5-mile segments with at least one crash occurrence, indicating that only a few segments experienced multiple crashes (0% to 4.79%). Specifically, there are only 71 segments with RL levels below 150 mcd/m²/lux, and further analysis exhibits that only six crashes occurred within these segments.

- In comparison to rural two-lane roadways, the total crash counts by all targeted crash types are considerably higher than the number of 0.5-mile segments with at least one crash occurrence on multilane roads (23.88% to 38.81%), indicating a more dispersed crash distribution. The proportion of segments with crashes is also higher on multilane roads (12.98% to 32.36%) than on two-lane roads (3.26% to 10.35%), providing a relatively sufficient number of segments with RL levels below 150 mcd/m²/lux for modeling. Although segments with curves are less common on multilane roads compared to two-lane roads, the proportion of curved segments with at least one crash occurrence is higher on multilane roadways.
- Among rural two-lane segments where RL measurements are available for both YCL and WREL, a substantial proportion have RL of YCL below 200 mcd/m²/lux (68.26%). This may indicate a lower frequency of restriping YCL, potentially due to its slower degradation rate compared to WREL. Rural two-lane segments with low RL levels of YCL and WREL are particularly susceptible to road crashes. For example, there are 42.11% of segments with at least one crash in dark no streetlighting conditions have RL of below 200 mcd/m²/lux and below 250 mcd/m²/lux for YCL and WREL, respectively. This percentage is disproportionately high compared to the overall percentage of 0.5-mile segments with low YCL and WREL RL levels (31.42%).
- When focusing on single-vehicle ROR crashes, the proportion of crashes on curve segments is relatively higher than on straight segments, suggesting that single-vehicle ROR crashes are strongly associated with curve segments. It is important to note that since data for both two-lane and multilane roads have been combined for 0.1-mile curved segments, functional class is used as a covariate instead of a posted speed limit.

Table 16 Summary of crash datasets on rural two-lane segments

Variable	**Rural two-lane (Number of 0.5-mile segments = 1,411)**		
	Single vehicle ROR	**Single vehicle ROR at dark, dusk, and dawn**	**Single vehicle ROR at dark no streetlighting**
Crash count	Min: 0, Max: 2, Median: 0, Mean: 0.11 (Number of crashes = 153)	Min: 0, Max: 2, Median: 0, Mean: 0.04 (Number of crashes = 60)	Min: 0, Max: 1, Median: 0, Mean: 0.03 (Number of crashes = 46)
AADT (vpd)	Min: 229, Max: 13,731, Median: 2,323, Mean: 2,822		
Crash occurrence			
yes	146 (10.35%)	58 (4.11%)	46 (3.26%)
no	1,265 (89.65%)	1,353 (95.89%)	1,365 (96.74%)
Location class			
open country	1,307 (92.63%)		
residential/business/mixed	104 (7.37%)		
Posted speed limit			
< 55mph	132 (9.36%)		
≥ 55mph	1,279 (90.65%)		
Presence of curve			
no	1,002 (71.01%)		
yes	409 (28.99%)		
RL by category (mcd/m²/lux)			
<250	492 (34.87%)		
≥ 250	919 (65.13%)		

Table 17 Summary of crash datasets on rural multilane segments

Variable	**Rural multilane (Number of 0.5-mile segments = 516)**		
	Single vehicle ROR	**Single vehicle ROR at dark, dusk, and dawn**	**Single vehicle ROR at dark no streetlighting**
Crash count	Min: 0, Max: 6, Median: 0, Mean: 0.52 (Number of crashes = 267)	Min: 0, Max: 4, Median: 0, Mean: 0.20 (Number of crashes = 101)	Min: 0, Max: 4, Median: 0, Mean: 0.17 (Number of crashes = 86)
AADT (vpd)	Min: 3,675, Max: 86,009, Median: 19,075, Mean: 22,886		
Crash occurrence			
yes	167 (32.36%)	80 (15.50%)	67 (12.98%)
no	349 (67.64%)	436 (84.50%)	449 (87.02%)
Location class			
open country	464 (89.92%)		
residential/business/mixed	52 (10.08%)		
Posted speed limit			
<65mph	112 (21.71%)		
≥ 65mph	404 (78.29%)		
Presence of curve			
no	443 (85.85%)		
yes	73 (14.15%)		
RL by category (mcd/m²/lux)			
<150	83 (16.09%)		
150-249	140 (27.13%)		
<250	223 (43.22%)		
≥ 250	293 (56.78%)		

Table 18 Summary of crash datasets on curve segments and at dark no streetlighting conditions

Variable	**Dark no streetlighting – rural two-lane only (Number of 0.5-mile segments = 1,235)**	**Single-vehicle ROR crashes on curve segments (Number of 0.1-mile segments = 871)**
Crash count	Min: 0, Max: 2, Median: 0, Mean: 0.09 (Number of crashes = 106)	Min: 0, Max: 5, Median: 0, Mean: 0.18 (Number of crashes = 158)
AADT (vpd)	Min: 229, Max: 13,731, Median: 2,404, Mean: 2,851	Min: 229, Max: 140,466, Median: 2,955, Mean: 5,256
Curve angle (degree)	----	Min: 8.96, Max: 84.77, Median: 35.04, Mean: 37.19
Curve radius (ft)	----	Min: 317, Max: 7,968, Median: 2,135, Mean: 2,623
Crash occurrence		
yes	95 (7.69%)	130 (14.93%)
no	1,140 (92.31%)	741 (85.07%)
Location class		
open country	1,147 (92.87%)	777 (89.21%)
residential/business/mixed	88 (7.13%)	94 (10.79%)
Posted speed limit		----
< 55mph	117 (9.47%)	----
≥ 55mph	1,118 (90.53%)	----
Presence of curve		----
yes	367 (29.72%)	----
no	868 (70.28%)	----
Number of lanes		
two	----	718 (82.43%)
>two	----	153 (17.57%)

Variable	**Dark no streetlighting – rural two-lane only (Number of 0.5-mile segments = 1,235)**	**Single-vehicle ROR crashes on curve segments (Number of 0.1-mile segments = 871)**
RL of WREL by category (mcd/m²/lux)		
<250	447 (36.19%)	326 (37.43%)
≥ 250	788 (63.81%)	545 (62.57%)
RL of YCL by category (mcd/m²/lux)		----
<200	843 (68.26%)	----
≥ 200	392 (31.74%)	----
Interaction of RL (*(mcd/m²/lux)*		----
YCL<200 and WREL<250	388 (31.42%)	----
YCL≥200 and WREL<250	59 (4.78%)	----
YCL<200 and WREL≥250	455 (36.84%)	----
YCL≥200and WREL≥250	333 (26.96%)	----
Functional class		
major collector	----	265 (30.43%)
arterial	----	591 (67.85%)
interstate	----	15 (1.72%)

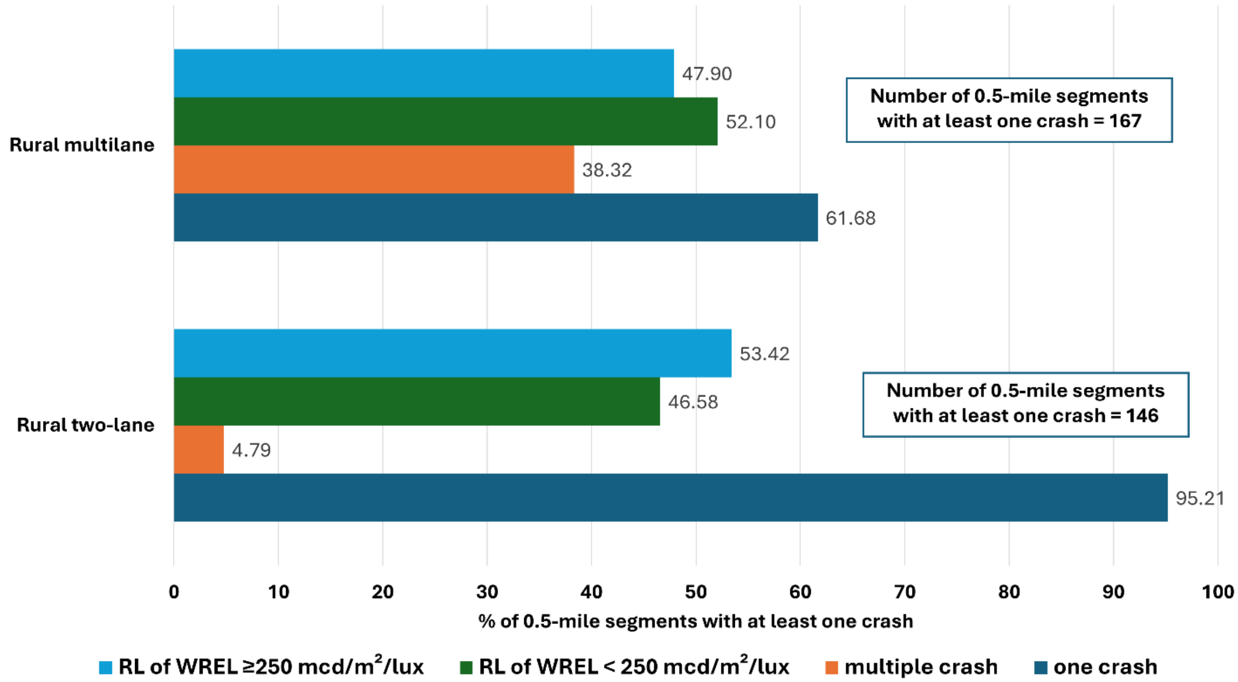

Figure 19 Distribution of 0.5-mile segments with at least one single-vehicle ROR crash

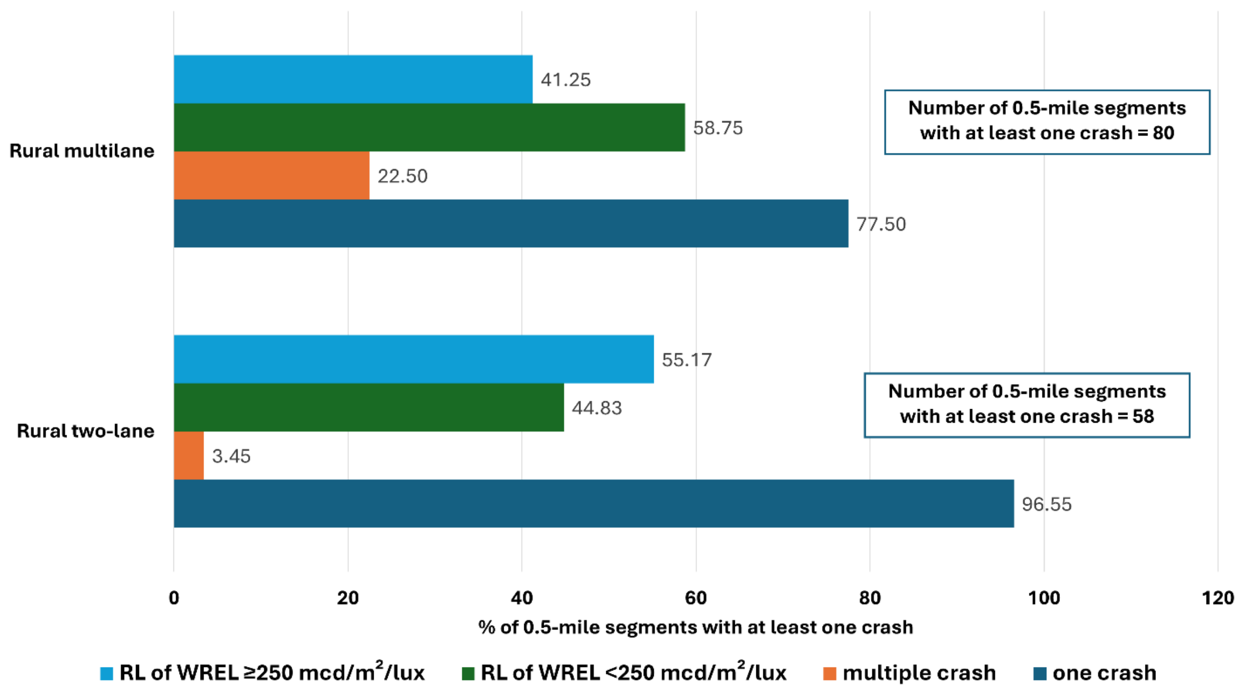

Figure 20 Distribution of 0.5-mile segments with at least one single-vehicle ROR crash at dark, dusk, and dawn

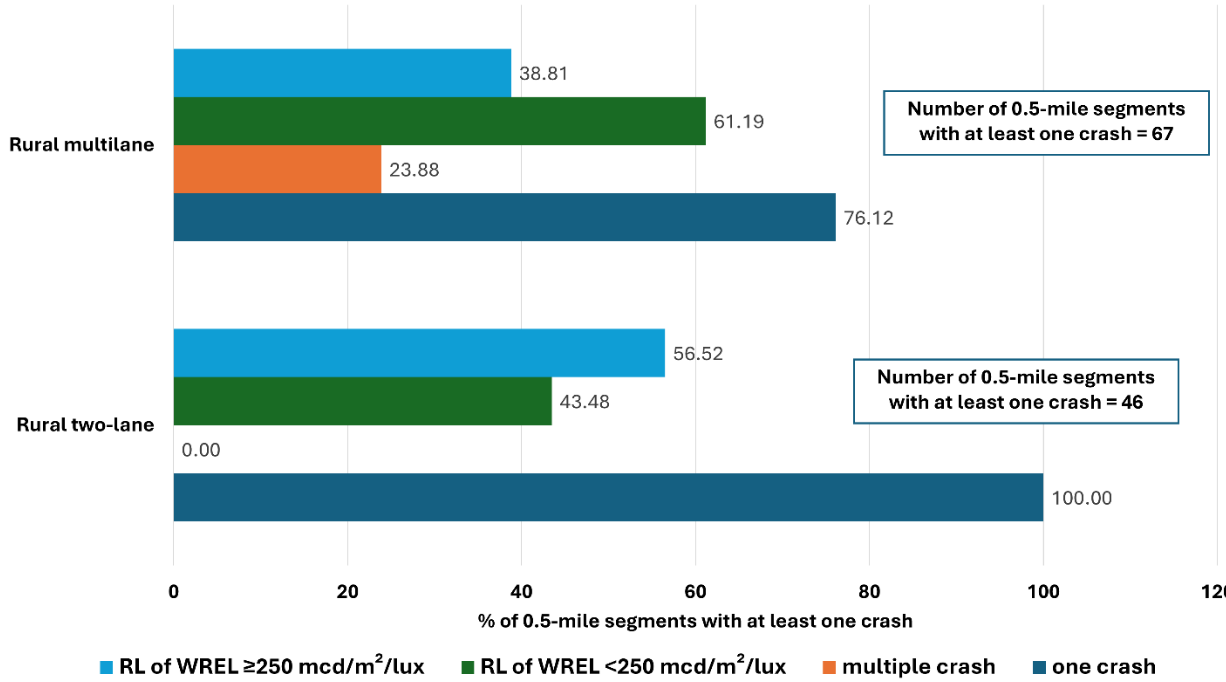

Figure 21 Distribution of 0.5-mile segments with at least one single-vehicle ROR crash at dark with no streetlighting

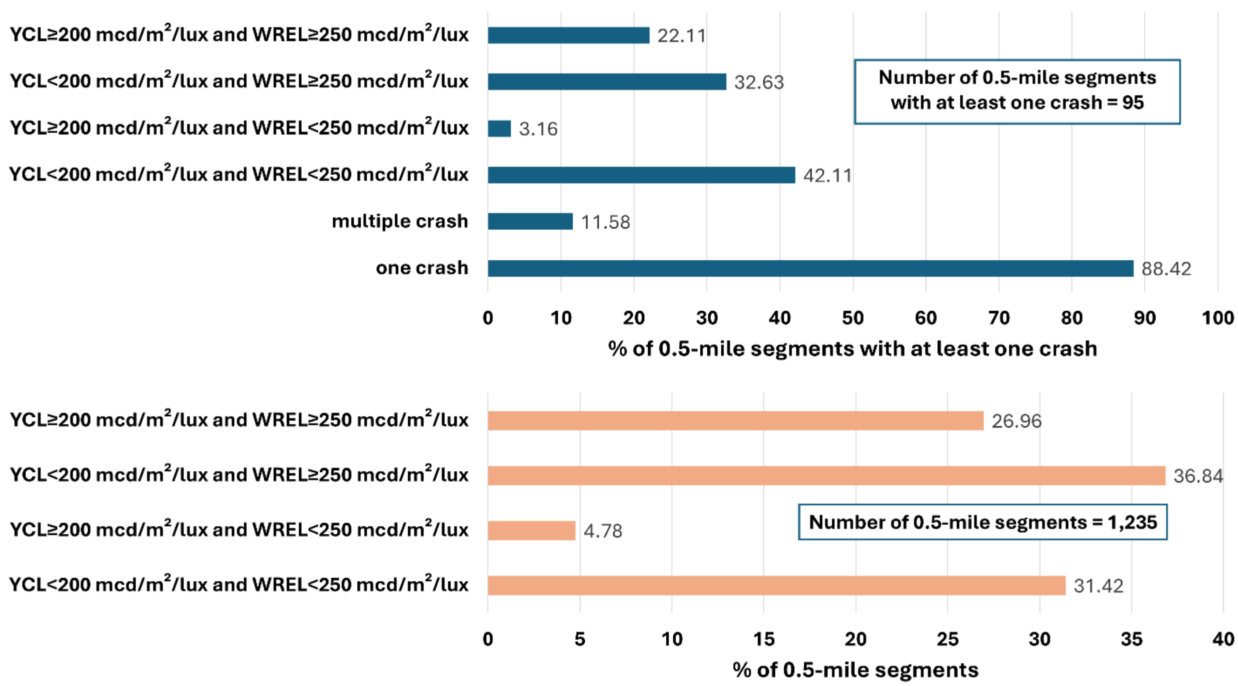

Figure 22 Distribution of 0.5-mile segments with at least one crash at dark with no streetlighting (rural two-lane only)

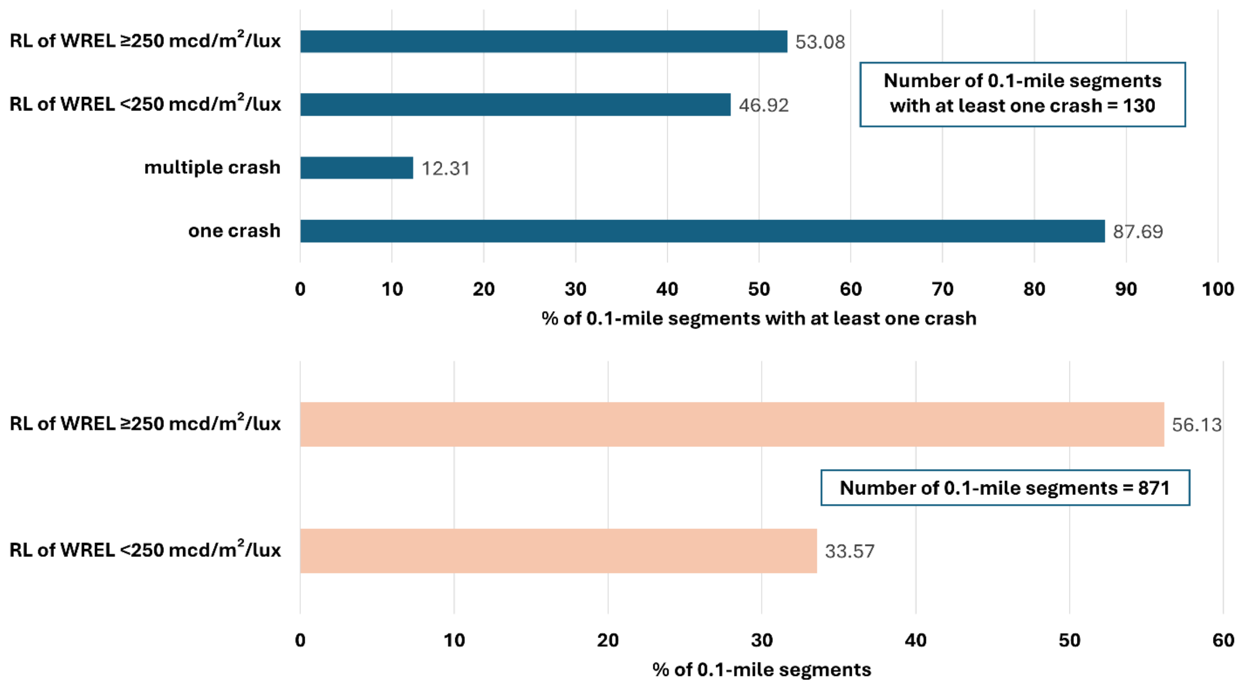

Figure 23 Distribution of 0.1-mile segments with at least one crash on curve segments

5.3 Methodology

In this research, both binary logit regression and negative binomial (NB) regression are applied to examine the statistical relationship between RL and road crashes. In the binary logit regression model, the dependent variable indicates whether a crash occurred within a segment, providing the probability of crash occurrence. In contrast, the NB regression model utilizes crash frequency as the dependent variable, allowing for the assessment of the percentage increase or decrease in expected crash frequency.

5.3.1 Binary Logit Regression

Binary logit regression has been employed in traffic safety research to examine the association between binary outcome variables and explanatory factors (Rahman et al., 2021). In this study, the dichotomous outcome variable y_{in} is defined as:

$$y_{in} = \begin{cases} 1, if\ presence\ of\ crashes \\ 0, if\ no\ presence\ of\ crashes \end{cases}$$

If the probability of a crash occurring is represented by $P(y_{in} = 1)$, then the logistic function can be defined as follows (McFadden, 1981):

$$P(y_{in} = 1) = \frac{\exp\,(\beta X_{in})}{1 + \exp\,(\beta X_{in})}$$

In this context, X_{in} refers to the independent variable vector, while β represents the estimable coefficient vector. The estimation of the coefficients can be achieved by maximizing the following log-likelihood function (Minka, 2001):

$$LL(\beta) = \sum_{i=1}^{n} \{(y_{in}\ln(P(y_{in})) + (1 - y_{in})\ln(1 - P(y_{in})))\}$$

The Akaike Information Criterion (AIC) is a statistical measure used to evaluate the relative quality of regression models. It is rooted in information theory and provides a means of comparing multiple models to determine which is the most suitable for explaining the observed data (Bozdogan, 1987; Cavanaugh et al., 2019). The AIC assesses both the goodness-of-fit and the complexity of the model, aiming to strike a balance between them to avoid overfitting or underfitting (Van Petegem et al., 2014) .

$$AIC = 2k - 2\ln(L)$$

Where, k is the number of parameters in the model and ln(L) is the natural logarithm of the likelihood of the model.

5.3.2 Negative Binomial (NB) Regression

The negative binomial model, also regarded as the Poisson-Gamma model, is an extension of the conventional Poisson model to overcome potential over-dispersion in the data. The model hypothesizes that the Poisson parameter supports a gamma likelihood distribution (Poch et al., 1996). The NB assumes that crash frequencies are independent for an entity for any month. The model outcomes in a closed-form equation and the computation to manipulate the association between the mean and the variance is relatively straightforward (Lord et al., 2010). The equations of the NB regression model are specified below:

$$P(z_i) = \frac{\Gamma\left(\left(\frac{1}{\beta}\right) + z_i\right)}{\Gamma\left(\frac{1}{\beta}\right) z_i!} \left(\frac{\frac{1}{\beta}}{\left(\frac{1}{\beta}\right) + \mu_i}\right)^{\frac{1}{\beta}} \left(\frac{\mu_i}{\left(\frac{1}{\beta}\right) + \mu_i}\right)^{z_i}$$

$$\mu_i = exp(\gamma Y_i)$$

$$L(\gamma) = \prod_i \frac{\Gamma\left(\left(\frac{1}{\beta}\right) + z_i\right)}{\Gamma\left(\frac{1}{\beta}\right) z_i!} \left(\frac{\frac{1}{\beta}}{\left(\frac{1}{\beta}\right) + exp(\gamma Y_i)}\right)^{\frac{1}{\beta}} \left(\frac{exp(\gamma Y_i)}{\left(\frac{1}{\beta}\right) + exp(\gamma Y_i)}\right)^{z_i}$$

$$LL(\gamma) = \sum_{i=1}^{n} \left(ln\left(\frac{\Gamma\left(\left(\frac{1}{\beta}\right) + z_i\right)}{\Gamma\left(\frac{1}{\beta}\right) z_i!}\right) + \left(\frac{1}{\beta}\right) ln\left(\frac{\frac{1}{\beta}}{\left(\frac{1}{\beta}\right) + exp(\gamma Y_i)}\right) + z_i \left[\gamma Y_i - ln\left(\left(\frac{1}{\beta}\right) + exp(\gamma Y_i)\right)\right]\right)$$

Where, $P(z_i)$ = the likelihood of segment i having z_i crashes; μ_i = the expectation of z_i conditional on Y_i; β = over-dispersion parameter; Y_i = a vector of explanatory variables; γ = a vector of unknown parameter; $L(.)$ = likelihood function; $LL(.)$ = logarithm of likelihood function. Compared to the Poisson distribution, the NB distribution can allow for over-dispersion. If $\beta \rightarrow 0$, the NB model converges to the Poisson model as the variance equals the mean (Yang et al., 2021).

Given that crashes are typically infrequent and occur randomly, the zero-inflated negative binomial (ZINB) model can be employed to address segments with zero crashes, which cannot be adequately explained by NB model alone. The probability distribution of ZINB for random variable y_k is provided below (Garay et al., 2011):

$$P(y_k = j) = \begin{cases} \pi_k + (1 - \pi_k) g(\pi_k = 0), if\ j = 0 \\ (1 - \pi_k) g(y_k), if\ j > 0 \end{cases}$$

$$g(y_k) = P(Y = y_k | \mu_k, \alpha) = \frac{\Gamma(y_k + \alpha^{-1})}{\Gamma(\alpha^{-1})\, \Gamma(y_k + 1)} (\frac{1}{1 + \alpha\mu_k})^{\alpha^{-1}} (\frac{\alpha\mu_k}{1 + \alpha\mu_k})^{y_k}$$

$$\pi_k = \frac{\vartheta_k}{1 + \vartheta_k}$$

Where, $g(y_k)$ represents the NB distribution and π_k denotes the logistic link function. μ_k represents the average crash frequency, α denotes the over-dispersion parameter, ϑ_k represents the probability of segment k being completely safe, i.e., the probability of a true absence of crashes at segment k.

5.4 Analysis and Discussions

As discussed, both binary logistic regression and NB regression were applied to crash datasets, covering various types of targeted crashes along with the corresponding RL values of WREL and YCL for specific lengths of segments. The analyses were done at a 95% confidence interval (α = 0.05). The results indicate that both AIC and Pseudo R^2 values are higher in binary logistic regression compared to NB regression. A high dispersion value was observed in NB models, particularly for rural two-lane segments, suggesting significant overdispersion where the variance is much greater than the mean. In rural two-lane segments, only a small fraction of segments had more than one crash, therefore, ZINB models were also applied. However, these models did not yield any significant improvement in performance or changes in coefficient values. Comparing logistic and NB regression, the coefficients and variable significance were nearly identical, with logistic regression consistently performing better across all crash datasets. In all targeted crashes, the binary logit model performs well compared to NB and ZINB in terms of AIC and Pseudo R^2 values. Previously, several highway safety studies have selected the best model outputs by comparing AIC values across multiple regression models (Chen et al., 2000; Li et al., 2019; Montella et al., 2008). Consequently, the results from logistic regression were selected for further discussion. It is important to note that the F-statistics is used to determine whether adding new predictors significantly enhances a model's explanatory power, rather than to identify the best model by comparing multiple regression models.

5.4.1 Regression Results for Rural Two-Lane Segments

Table 19 to

Table 23 provide a summary of regression results for rural two-lane segments with respect to various targeted crashes, including single-vehicle ROR crashes, single-vehicle ROR crashes at dark, dusk, and dawn conditions, single-vehicle ROR crashes at dark not-lighted, and crashes at dark with no street lighting. In all crash types, AADT is found to be a significant factor, indicating that higher AADT is associated with increased odds of a crash occurrence.

Table 19 Summary of regression results for single-vehicle ROR crashes on rural two-lane segments

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-3.102	0.178	-17.461	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<250 mcd/m²/lux	0.373	0.185	2.021	0.043
AADT (in 1,000 vpd)	0.211	0.031	6.050	<0.001
Location class (ref. open country)				
residential/business/mixed	-0.347	0.359	-0.965	0.335
Posted speed limit (ref. ≥55mph)				
<55mph	-0.186	0.314	-0.594	0.553
Presence of curve (ref. no)				
yes	0.477	0.188	2.538	0.011
	AIC: 901.64, Pseudo R^2: 0.052			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 20 Summary of regression results for single-vehicle ROR crashes on rural two-lane segments at dark, dusk, and dawn conditions

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-3.938	0.265	-14.883	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<250 mcd/m²/lux	0.296	0.281	1.051	0.293
AADT (in 1,000 vpd)	0.173	0.052	3.416	0.001
Location class (ref. open country)				
residential/business/mixed	-0.762	0.629	-1.211	0.226
Posted speed limit (ref. ≥55mph)				
<55mph	-0.021	0.455	-0.047	0.963
Presence of curve (ref. no)				
yes	0.474	0.282	1.682	0.093
	AIC: 479.79, Pseudo R^2: 0.039			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 21 Summary of regression results for single-vehicle ROR crashes on rural two-lane segments at dark not-lighted conditions

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-4.301	0.306	-14.044	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<250 mcd/m²/lux	0.216	0.317	0.682	0.496
AADT (in 1,000 vpd)	0.213	0.057	3.762	<0.001
Location class (ref. open country)				
residential/business/mixed	-1.734	1.047	-1.646	0.099
Posted speed limit (ref. ≥55mph)				
<55mph	-0.547	0.617	-0.887	0.374
Presence of curve (ref. no)				
yes	0.678	0.311	2.179	0.029
	AIC: 396.71, Pseudo R^2: 0.051			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 22 Summary of regression results for crashes on rural two-lane segments at dark not-lighted conditions (interaction of line type as a covariate)

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-3.240	0.265	-12.221	<0.001
RL interaction (ref. YCL≥200 mcd/m²/lux and WREL≥250 mcd/m²/lux)				
YCL<200 *mcd/m²/lux* and WREL<250 *mcd/m²/lux*	0.331	0.291	1.137	0.255
YCL≥200 *mcd/m²/lux* and WREL<250 *mcd/m²/lux*	-0.409	0.651	-0.629	0.529
YCL<200 *mcd/m²/lux* and WREL≥250 *mcd/m²/lux*	0.048	0.295	0.163	0.870
AADT (in 1,000 vpd)	0. 201	0.040	4.994	<0.001
Location class (ref. open country)				
residential/business/mixed	-0.021	0.411	-0.050	0.959
Posted speed limit (ref. ≥55mph)				
<55mph	-0.916	0.488	-1.878	0.060
Presence of curve (ref. no)				
yes	0.121	0.238	0.510	0.610
	AIC: 654.40, Pseudo R^2: 0.048			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 23 Summary of regression results for crashes on rural two-lane segments at dark not-lighted conditions

Attribute	Coef.	SE	z-value	p-value
(Intercept)	-3.325	0.257	-12.917	<0.001
YCL (ref. ≥200 mcd/m²/lux)				
<200 *mcd/m²/lux*	0.195	0.260	0.749	0.454
WREL *(ref. ≥250 mcd/m²/lux)*				
<250 *mcd/m²/lux*	0.177	0.235	0.753	0.452
AADT (in 1,000 vpd)	0.208	0.040	4.988	<0.001
Location class (ref. open country)				
residential/business/mixed	-0.033	0.412	-0.080	0.936
Posted speed limit (ref. ≥55mph)				
<55mph	-0.909	0.486	-1.871	0.061
Presence of curve (ref. no)				
yes	0.114	0.237	0.481	0.630
	AIC: 653.49, Pseudo R^2: 0.045			

Note: Coefficient (Coef.) and Standard Error (SE)

Both single-vehicle ROR crashes and single-vehicle ROR crashes at dark not-lighted conditions show a statistically significant positive correlation with curve segments at a 95% confidence level, aligning with prior studies that highlight a higher likelihood of single-vehicle ROR crashes on curves (Duddu et al., 2020), especially in rural roadways (Gong et al., 2017). At a 90% confidence level (α = 0.1), the presence of curves is also found to be statistically significant for single-vehicle ROR crashes at dark, dusk, and dawn conditions. Interestingly, at a 90% confidence level, the odds of crashes on rural two-lane road segments in dark, not-lighted conditions decrease by 17.66% in residential, business, or mixed-use areas. This reduction may be due to drivers' tendency to lower their speed in these areas because of the increased presence of pedestrians and parked vehicles, particularly during nighttime hours. On rural two-lane segments, RL of WREL below 250 mcd/m²/lux increases the odds of single-vehicle ROR crashes by 45.21%. In other crash types, however, lower RL values of WREL are not found to be statistically significant, though they are positively correlated with crash occurrences. This lack of statistical significance could be attributed to the limited observations of rural two-lane segments with crashes. When analyzing the combined influence of both YCL and WREL on crashes at dark with no street lighting, neither the individual nor the interaction of RL levels is found to be statistically significant. However, lower RL levels are positively associated with crashes in dark not-lighted conditions.

5.4.2 Regression Results for Rural Multilane Segments

Table 24 to **Table 29** exhibit a summary of regression results for rural multilane segments for various targeted crash types, including single-vehicle ROR crashes, single-vehicle ROR crashes at dark, dusk, and dawn conditions, and single-vehicle ROR crashes at dark not-lighted, categorized by different RL levels of WREL.

Table 24 Summary of regression results for single-vehicle ROR crashes on rural multilane segments (RL category 1)

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-1.975	0.258	-7.652	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<150 mcd/m²/lux	0.226	0.277	0.817	0.414
150-249 mcd/m²/lux	0.045	0.242	0.187	0.852
AADT (in 1,000 vpd)	0.051	0.013	6.451	<0.001
Location class (ref. open country)				
residential/business/mixed	-1.002	0.445	-2.250	0.025
Posted speed limit (ref. ≥65mph)				
<65mph	-0.087	0.329	-0.265	0.791
Presence of curve (ref. no)				
yes	0.123	0.318	0.386	0.700
	AIC: 593.64, Pseudo R^2: 0.108			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 25 Summary of regression results for single-vehicle ROR crashes on rural multilane segments (RL category 2)

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-1.972	0.258	-7.654	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<250 mcd/m²/lux	0.116	0.212	0.546	0.585
AADT (in 1,000 vpd)	0.058	0.017	6.455	<0.001
Location class (ref. open country)				
residential/business/mixed	-1.006	0.445	-2.261	0.024
Posted speed limit (ref. ≥65mph)				
<65mph	-0.089	0.329	-0.272	0.786
Presence of curve (ref. no)				
yes	0.126	0.318	0.395	0.693
	AIC: 592.02, Pseudo R^2: 0.107			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 26 Summary of regression results for single-vehicle ROR crashes on rural multilane segments at dark, dusk, and dawn conditions (RL category 1)

Attribute	Coef.	SE	z-value	p-value
(Intercept)	-2.815	0.342	-8.230	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<150 mcd/m²/lux	0.570	0.332	1.719	0.086
150-249 mcd/m²/lux	0.417	0.294	1.417	0.156
AADT (in 1,000 vpd)	0.043	0.011	4.176	<0.001
Location class (ref. open country)				
residential/business/mixed	-0.912	0.609	-1.497	0.135
Posted speed limit (ref. ≥65mph)				
<65mph	-0.182	0.462	-0.394	0.693
Presence of curve (ref. no)				
yes	0.099	0.424	0.234	0.815
	AIC: 425.44, Pseudo R^2: 0.076			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 27 Summary of regression results for single-vehicle ROR crashes on rural multilane segments at dark, dusk, and dawn conditions (RL category 2)

Attribute	Coef.	SE	z-value	p-value
(Intercept)	-2.812	0.341	-8.236	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<250 mcd/m²/lux	0.477	0.261	1.829	0.067
AADT (in 1,000 vpd)	0.042	0.015	4.175	<0.001
Location class (ref. open country)				
residential/business/mixed	-0.918	0.609	-1.507	0.132
Posted speed limit (ref. ≥65mph)				
<65mph	-0.185	0.462	-0.399	0.690
Presence of curve (ref. no)				
yes	0.103	0.424	0.242	0.809
	AIC: 423.64, Pseudo R^2: 0.075			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 28 Summary of regression results for single-vehicle ROR crashes on rural multilane segments at dark not-lighted conditions (RL category 1)

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-2.941	0.367	-8.016	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<150 mcd/m²/lux	0.771	0.347	2.223	0.026
150-249 mcd/m²/lux	0.445	0.319	1.394	0.163
AADT (in 1,000 vpd)	0.033	0.009	3.484	<0.001
Location class (ref. open country)				
residential/business/mixed	-0.901	0.679	-1.326	0.185
Posted speed limit (ref. ≥65mph)				
<65mph	-0.288	0.514	-0.559	0.576
Presence of curve (ref. no)				
yes	-0.034	0.475	-0.071	0.943
	AIC: 383.22, Pseudo R^2: 0.073			

Note: Coefficient (Coef.) and Standard Error (SE)

Table 29 Summary of regression results for single-vehicle ROR crashes on rural multilane segments at dark, not-lighted conditions (RL category 2)

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-2.934	0.365	-8.031	<0.001
RL category (ref. ≥250 mcd/m²/lux)				
<250 mcd/m²/lux	0.576	0.280	2.055	0.039
AADT (in 1,000 vpd)	0.032	0.009	3.478	<0.001
Location class (ref. open country)				
residential/business/mixed	-0.918	0.681	-1.349	0.177
Posted speed limit (ref. ≥65mph)				
<65mph	-0.292	0.515	-0.566	0.571
Presence of curve (ref. no)				
yes	-0.024	0.475	-0.051	0.959
	AIC: 382.05, Pseudo R^2: 0.071			

Note: Coefficient (Coef.) and Standard Error (SE)

Consistent with findings for rural two-lane segments, AADT shows a statistically significant correlation with crash occurrences in relation to all targeted crash types. At a 95% confidence interval, the odds of single-vehicle ROR crashes on rural multilane segments decrease by over 30% in residential, business, or mixed-use areas. Similar statistically significant correlations are observed for single-vehicle ROR crashes at dark, dusk, and dawn conditions at an

85% confidence interval (α = 0.15), and for single-vehicle ROR crashes at dark not-lighted conditions at an 80% confidence interval (α = 0.2).

Under dark, dusk, and dawn conditions, rural multilane segments with RL of WREL below 150 mcd/m²/lux experience a 76.84% increase in the odds of single-vehicle ROR crashes at a 90% confidence interval. In similar lighting conditions, if RL values of WREL fall between 150 and 249 mcd/m²/lux, the odds of single-vehicle ROR crashes increase by 51.77% at an 80% confidence interval. When analyzing only two RL categories, ≥250 mcd/m²/lux and <250 mcd/m²/lux, the odds of single-vehicle ROR crashes at dark, dusk, and dawn conditions increases by 61.08% for segments with RL below 250 mcd/m²/lux at a 90% confidence interval. In dark conditions with no street lighting, rural multilane segments with RL levels below 150 mcd/m²/lux showed a 116.19% increase in the odds of single-vehicle ROR crashes at a 95% confidence interval. When examining only two RL categories, ≥250 mcd/m²/lux and <250 mcd/m²/lux, the odds of single-vehicle ROR crash occurrences at dark not-lighted conditions rise by 77.89% on segments with RL below 250 mcd/m²/lux at a 90% confidence interval.

5.4.3 Regression Results for Curve Segments

Table 30 summarizes the regression results for curve segments in relation to single-vehicle ROR crashes. For this analysis, the dataset was prepared using average values for each 0.1-mile segment, and both two-lane and multilane roads were merged, including a 'number of lanes' variable. Like previous segments, AADT is found to be a significant factor. The odds of single-vehicle ROR crashes on curve segments increase with higher road functional classifications, rising by 144% for arterial roads and 373.04% for interstates, which may be associated with higher posted speed limits. No significant correlation has been observed with respect to curve angle and radius. On curve segments, RL values of WREL below 250 mcd/m²/lux are associated with a 59.89% increase in the odds of single-vehicle ROR crashes. Low RL levels on curves can reduce drivers' ability to perceive lane boundaries, making it more challenging to navigate curves safely and increasing the risk of vehicles veering off the roadway.

Table 30 Summary of regression results for single-vehicle ROR crashes on curve segments

Attribute	**Coef.**	**SE**	**z-value**	**p-value**
(Intercept)	-3.240	0.573	-5.655	<0.001
Number of lanes (ref. >two)				
two	0.262	0.374	0.703	0.482
RL category (ref. ≥250 mcd/m²/lux)				
<250 mcd/m²/lux	0.469	0.201	2.337	0.019
AADT (in 1,000 vpd)	0.052	0.023	2.567	0.010
Location class (ref. open country)				
business/residential/mixed	-0.037	0.312	-0.118	0.906
Functional class (ref. major collector)				
arterial	0.892	0.282	3.165	0.002

Attribute	Coef.	SE	z-value	p-value
interstate	1.554	0.779	1.994	0.046
Curve angle	0.006	0.007	0.920	0.358
Curve radius (in 1,000)	-0.065	0.070	-0.950	0.342
	AIC:703.28, Pseudo R^2: 0.079			

Note: Coefficient (Coef.) and Standard Error (SE)

5.5 Key Findings

This chapter aims to investigate the statistical relationship between RL and road crashes. To expand the dataset, the RL prediction models developed in Chapter 6 were applied to estimate RL of WREL and YCL for additional segments, increasing the sample size significantly. Unlike previous studies that calculated average RL over 1-mile segments, this study used shorter segments, specifically 0.5-mile, for more granular analysis. Practical assumptions were made based on the available data and the limitations of the predictive models. Several covariates were considered to improve the validity of the model results by considering the effect of other factors, such as AADT, posted speed limit, location class, road functional class, and horizontal alignment. Two widely used regression models, binary logit model and NB model, were applied to all prepared crash datasets. Binary logit model demonstrated better performance in terms of AIC and Pseudo R^2, particularly given the zero-inflated nature of the datasets. The key findings from these analyses are summarized below:

- RL level of WREL below 250 mcd/m²/lux statistically increases the likelihood of single-vehicle ROR crashes on rural two-lane road segments.
- On rural multilane segments in dark, dusk, and dawn conditions, both RL levels of WREL below 150 mcd/m²/lux and 250 mcd/m²/lux increase the likelihood of single-vehicle ROR crashes.
- Lower RL of WREL, specifically below 250 mcd/m²/lux, plays a significant role in increasing the probability of single-vehicle ROR crashes on rural multilane segments in dark, not-lighted conditions.
- On curve segments, RL of WREL below 250 mcd/m²/lux is associated with a statistically significant increase in the likelihood of single-vehicle ROR crashes.

6 CONCLUSIONS

This chapter aims to highlight the contributions of this research by providing in-depth aspects of RL prediction modeling and the statistical relationship between RL and crash frequency. The key findings are also presented in alignment with the project objectives, contextualizing the contributions of this study. This discussion further provides practical implementation ideas and recommendations, highlighting the significance of these findings. Additionally, study limitations are addressed, offering insights and potential directions for future research.

6.1 Research Contributions

Local transportation agencies often do not have initial RL measurements and marking age information when assessing the RL of pavement markings on specific segments. This gap creates a need for models that can predict RL without relying on initial RL measurements taken after installation or restriping and knowing when the markings were installed or restriped. Prediction models must remain practical and usable, incorporating only the key variables that contribute to RL degradation to avoid unnecessary complexity. This study is the first in the existing literature to take such an approach in developing regression models for predicting the RL of YCL and WREL, considering lane configurations of rural roads. While the model showed some limitations, particularly in predicting future RL at lower levels, it provides valuable contributions. The latest MUTCD version lowers the minimum RL standard to 50 mcd/m²/lux, meaning transportation agencies must prioritize collecting more data on segments with RL levels between 50 and 100 mcd/m²/lux. This data can be vital for developing predictive models to better estimate the service life of marking up to 50 mcd/m²/lux. Despite limitations in predicting RL at lower levels, this research's developed RL prediction models offer agencies a useful tool for forecasting RL on segments that previously showed high RL values. A practical application of the developed RL prediction models is demonstrated in the Appendix Section (**Appendix A1** and **Appendix A2)**. By utilizing one year of RL data, agencies can predict future RL values and rank road segments accordingly, from lowest to highest predicted RL. This ranking enables ALDOT to efficiently identify and prioritize segments with lower RL levels, ensuring timely restriping and maintaining road safety standards.

This research demonstrates an effective approach for examining the statistical relationship between RL and road crashes by increasing the sample size by predicting RL following a single measured RL value. Employing an appropriate statistical modeling approach, which accounts for the logarithmic relationship of crashes and strategic RL categorization, has strengthened the validity of this research findings. Unlike previous studies that calculated average RL over 1-mile or 2-mile segments, this research considered smaller segment sizes, such as 0.5-mile segments in general and 0.1-mile segments specifically for curves. Moreover, multiple covariates, such as posted speed limit, AADT, location classification, and curve presence, were incorporated to validate the correlations. The results revealed a statistically significant relationship between lower RL of WREL and single-vehicle ROR crashes, confirming that this refined approach effectively highlights the significance of maintaining RL regarding road safety. Furthermore, the findings highlighted the vulnerability of lower RL levels on curve segments and low lighting conditions, underscoring their combined impact on road safety. Overall, the findings of this research align with one of the core principles of the Safe System Approach: Making Our Roads Safer' (Doctor et al., 2022).

6.2 Recommendations

This study discusses recommendations from three key perspectives: i) the need for additional and expanded datasets to enable more granular analysis, ii) alignment with current MUTCD guidelines, and iii) the key findings related to this research objectives. A summary of the key recommendations is provided in the following:

- This study identified that only a limited number of segments had RL levels below 100 mcd/m²/lux for rural two-lane and multilane roadways. Additional RL data from segments with lower RL levels is required to develop a more robust and comprehensive model capable of accurately predicting the service life of thermoplastic pavement markings. This need is especially important given the recent changes to the MUTCD guidelines, which have reduced the minimum RL requirement to 50 mcd/m²/lux. Therefore, ALDOT needs to implement a more systematic approach to RL data collection. One potential strategy is to conduct an initial subjective rating of specific segments, identify those likely to have low RL levels, and then prioritize those segments for precise RL measurement.
- To create a comprehensive RL degradation dataset for modeling, it is essential to include detailed road characteristics, such as lane width, shoulder width, and other relevant features. These attributes can vary within a one-mile segment, making precise calculation through tools like Google Earth Pro difficult. Most states maintain road inventory databases that contain such information, particularly for major roadways. It is recommended that ALDOT improve accessibility to Alabama road inventory data to ensure that critical road characteristics are readily available for further research.
- While analyzing the measured RL data, it was observed that RL of YCL and YLEL were often misassigned when lane transitions occurred, such as when transitioning from a rural two-lane to a three-lane configuration or vice versa. To address this issue, closer monitoring and quality control are required to ensure the correct assignment of marking line type to each of the measured RL.
- The current version of MUTCD reduces the minimum RL requirement to 50 mcd/m²/lux. This research found that lower RL levels, particularly below 250 mcd/m²/lux for WREL, significantly increase the likelihood of crash occurrences. Additionally, vehicles equipped with ADAS rely on adequate RL levels for optimal sensor performance, and a few studies have concluded that RL levels below 88 mcd/m²/lux can significantly reduce sensor accuracy. Therefore, continuous monitoring is essential to ensure this lower RL requirement does not compromise road safety. While high vehicle automation is still in its early stages with limited penetration, balancing these considerations without overemphasizing automation concerns and benefits is important.

- The current MUTCD guidelines for minimum RL standards do not consider low lighting conditions. However, higher RL levels are particularly beneficial in nighttime driving, as they help guide drivers to stay within the correct travel lane. Given this, developing a standardized method for determining RL under adverse and low lighting conditions, such as bad weather or dark with no street lighting, would be more practical.
- This study applied a segment-based approach to examine the statistical relationship between RL and crashes. However, each crash is associated with a combination of multiple factors. As a result, investigating individual crashes with measured RL values, rather than focusing solely on segments, could reveal hidden patterns among variables. It is recommended that RL values be measured at specific hotspot locations over a period of time to enable a more in-depth analysis of the relationship between RL and other crash contributing factors.

6.3 Study Limitations and Future Scope

This research has several limitations that can be addressed in future studies. The study utilized RL data exclusively from the Montgomery region, limiting the sample size. Expanding the dataset to include additional regions would increase the sample size, helping to extend the minimum thresholds for both marking colors. Typically, RL degrades more rapidly after installation until it reaches a certain level, after which it declines at a more stable rate. This degradation pattern could be more effectively captured through machine learning algorithms. However, due to the limited number of observations, machine-learning techniques were not applied in this study (Brodley et al., 2012). In the future, machine learning algorithms, such as random forests, could be applied to larger datasets to develop more practical RL prediction models. Additionally, factors like weather conditions, material composition, and quality of installation may influence RL degradation. Future research could explore these variables to understand better their impact on the transition of RL between qualitative states. The research findings and conclusions from rural Alabama roads warrant further investigation to ascertain their applicability to other states sharing similar climatic conditions. Future research could delve into more material-specific modeling to capture the distinct effects of various independent variables on RL degradation. Additionally, conducting a human-factors study on the correlation between pavement marking RL and speed could shed light on whether drivers perceive increased comfort with higher RL values, potentially leading to higher speeds. In this research, the safety analyses did not account for temporal variability in AADT. The findings highlighted significant differences in how various factors contribute to RL degradation between WLL and WREL. Consequently, further research is needed to explore these variations in greater depth and compare RL's safety impacts for WLL and WREL.

REFERENCES

Abboud, N., & Bowman, B. L. (2002). Cost- and longevity-based scheduling of paint and thermoplastic striping. *Transportation Research Record*, *1794*(1794), 55–64. doi: 10.3141/1794-07

Aktan, F., & Schnell, T. (2004). Performance evaluation of pavement markings under dry, wet, and rainy conditions in the field. *Transportation Research Record*, *1877*, 38–49. doi: 10.3141/1877-05

ALDOT. (2022). Standard Specifications for Highway Construction. In Alabama Department of Transportation.

ALDOT. (2023). *TDM Public: Alabama Traffic Data*. Retrieved from https://aldotgis.dot.state.al.us/TDMPublic/

Andrady, A. L. (Anthony L.), National Cooperative Highway Research Program., & National Research Council (U.S.). Transportation Research Board. (1997). *Pavement marking materials : assessing environment friendly performance* (Vol. 392). Transportation Research Board.

Avelar, R. E., & Carlson, P. J. (2014). Link between pavement marking retroreflectivity and night crashes on Michigan two-lane highways. *Transportation Research Record*, *2404*(1), 59–67. doi: 10.3141/2404-07

Bahar, G., Masliah, M., Erwin, T., Tan, E., & Hauer, E. (2006). Pavement marking materials and markers: Real-world relationship between retroreflectivity and safety over time. In National Academies of Sciences, Engineering, and Medicine. doi: 10.17226/23255

Bektas, B. A., Gkritza, K., & Smadi, O. (2016). Pavement marking retroreflectivity and crash frequency: segmentation, line type, and imputation effects. *Journal of Transportation Engineering*, *142*(8), 4016030. doi: 10.1061/(ASCE)TE.1943-5436.0000863

Bozdogan, H. (1987). Model Selection and Akaike’s Information Criterion (AIC): The General Theory and its Analytical Extensions. *Psychometrika*, *52*(3), 345–370.

Brodley, C. E., Rebbapragada, U., Small, K., & Wallace, B. (2012). Challenges and opportunities in applied machine learning. *Ai Magazine*, *33*(1), 11–24.

Carlson, P. J., Park, E. S., & Kang, D. H. (2013). Investigation of longitudinal pavement marking retroreflectivity and safety. *Transportation Research Record*, *2337*(1), 59–66. doi: 10.3141/2337-08

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, *11*(3), e1460.

Chen, W. H., & Jovanis, P. P. (2000). Method for identifying factors contributing to driver-injury severity in traffic crashes. *Transportation Research Record*, *1717*(1), 1–9.

Chimba, D., Kidando, E., & Onyango, M. (2018). Evaluating the service life of thermoplastic pavement markings: Stochastic approach. *Journal of Transportation Engineering, Part B: Pavements*, *144*(3), 4018029. doi: 10.1061/jpeodx.0000055

Doctor, M., & Ngo, C. (2022). MAKING OUR ROADS SAFER: Through a Safe System Approach. In Public Roads (Vol. 85, Issue 4).

Donnell, E. T., Karwa, V., & Sathyanarayanan, S. (2009). Analysis of effects of pavement marking retroreflectivity on traffic crash frequency on highways in North Carolina. *Transportation Research Record*, *2103*, 50–60. doi: 10.3141/2103-07

Dravitzki, V. K., Wilkie, S. M., & Lester, T. J. (2006). *The safety benefits of brighter roadmarkings* (Issue 310).

Duddu, V. R., Pulugurtha, S. S., & Kukkapalli, V. M. (2020). Variable categories influencing single-vehicle run-off-road crashes and their severity. *Transportation Engineering*, *2*, 100038.

Federal Highway Administration (FHWA). (2023). *Highway Performance Monitoring System (HPMS)*.

FHWA. (2017). *Nighttime Visibility*.

FHWA. (2022). National Standards for Traffic Control Devices; the Manual on Uniform Traffic Control Devices for Streets and Highways; Maintaining Pavement Marking Retroreflectivity. *Federal Register*, *87*(150).

Fu, H., & Wilmot, C. G. (2013). Evaluating alternative pavement marking materials. *Public Works Management & Policy*, *18*(3), 279–297. doi: 10.1177/1087724X12451844

Garay, A. M., Hashimoto, E. M., Ortega, E. M. M., & Lachos, V. H. (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics & Data Analysis*, *55*(3), 1304–1318.

Gong, L., & Fan, W. (David). (2017). Modeling single-vehicle run-off-road crash severity in rural areas: Accounting for unobserved heterogeneity and age difference. *Accident Analysis and Prevention*, *101*, 124–134. doi: 10.1016/j.aap.2017.02.014

Hollingsworth, J. D. (2012). *Understanding the impact of bead type on paint and thermoplastic pavement markings*. *March*.

Hummer, J. E., Rasdorf, W., & Zhang, G. (2011). Linear mixed-effects models for paint pavement-marking retroreflectivity data. *Journal of Transportation Engineering*, *137*(10), 705–716.

Hussein, M., Sayed, T., El-Basyouny, K., & de Leur, P. (2020). Investigating safety effects of wider longitudinal pavement markings. *Accident Analysis & Prevention*, *142*, 105527. doi: 10.1016/j.aap.2020.105527

Idris, I. I., Mousa, M., & Hassan, M. (2024). Modeling retroreflectivity degradation of pavement markings across the US with advanced machine learning algorithms. *Journal of Infrastructure Preservation and Resilience*, *5*(1), 3.

Karwa, V., & Donnell, E. T. (2011). Predicting pavement marking retroreflectivity using artificial neural networks: Exploratory analysis. *Journal of Transportation Engineering*, *137*(2), 91–103. doi: 10.1061/(ASCE)TE.1943-5436.0000194

Kopf, J. (2004). *Reflectivity of pavement markings: Analysis of retroreflectivity degradation curves*. Washington State Department of Transportation Olympia.

Lee, J. T., Maleck, T. L., & Taylor, W. C. (1999). Pavement marking material evaluation study in Michigan. *ITE Journal (Institute of Transportation Engineers)*, *69*(7).

Lee, Y.-J. (2009). *Life Cycle and Econimic Efficiency Analysis: Durable Pavement Markings*.

Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., Qian, Z. S., Prevedouros, P. D., & Ma, D. T. (2019). Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. *Accident Analysis & Prevention*, *124*, 219–229.

Lindly, J. K., & Wijesundera, R. K. (2003). Evaluation of profiled pavement markings. In Security (Issue 01465). University Transportation Center for Alabama.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, *44*(5), 291–305.

Malyuta, D. A. (2015). *Analysis of factors affecting pavement markings and pavement marking retroreflectivity in Tennessee Highways* (Vol. 53, Issue 9). University of Tennessee at Chattanooga.

Marill, K. A. (2004). Advanced statistics: linear regression, part II: multiple linear regression. *Academic Emergency Medicine*, *11*(1), 94–102.

Masliah, M., Bahar, G., & Hauer, E. (2007). Application of innovative time series methodology to relationship between retroreflectivity of pavement markings and crashes. *Transportation Research Record*, *2019*(1), 119–126. doi: 10.3141/2019-15

McFadden, D. (1981). Econometric models of probabilistic choice, Structural Analysis of Discrete Data with Econometric Applications. *MIT Press, Cambridge, MA.*

Migletz, J., Graham, J. L., Harwood, D. W., & Bauer, K. M. (2001). Service life of durable pavement markings. *Transportation Research Record*, *1749*(1), 13–21. doi: 10.3141/1749-03

Minka, T. P. (2001). Algorithms for maximum-likelihood logistic regression. *October*, *4*, 1–15.

Mohamed, M., Skinner, A., Abdel-Rahim, A., Kassem, E., & Chang, K. (2019). Deterioration characteristics of waterborne pavement markings subjected to different operating conditions. *Journal of Transportation Engineering, Part B: Pavements*, *145*(2), 4019003. doi: 10.1061/jpeodx.0000101

Montella, A., Colantuoni, L., & Lamberti, R. (2008). Crash prediction models for rural motorways. *Transportation Research Record*, *2083*(1), 180–189.

Mousa, M. R., Mousa, S. R., Hassan, M., Carlson, P., & Elnaml, I. A. (2021). Predicting the retroreflectivity degradation of waterborne paint pavement markings using advanced machine learning techniques. *Transportation Research Record*, *2675*(9), 483–494.

Mull, D. M., & Sitzabee, W. E. (2012). Paint pavement marking performance prediction model. *Journal of Transportation Engineering*, *138*(5), 618–624.

Ortiz-García, J. J., Costello, S. B., & Snaith, M. S. (2006). Derivation of transition probability matrices for pavement deterioration modeling. *Journal of Transportation Engineering*, *132*(2), 141–161.

Ozelim, L., & Turochy, R. E. (2014). Modeling retroreflectivity performance of thermoplastic pavement markings in Alabama. *Journal of Transportation Engineering*, *140*(6), 5014001.

Pike, A. M., Barrette, T. P., & Carlson, P. J. (2018). Evaluation of the Effects of Pavement Marking Characteristics on Detectability by ADAS Machine Vision. *National Cooperative Highway Research Program (NCHRP): Washington, DC, USA.*

Pike, A. M., & Songchitruksa, P. (2015). Predicting pavement marking service life with transverse test deck data. *Transportation Research Record*, *2482*(1), 16–22.

Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, *122*(2), 105–113.

Porras-Alvarado, J. D., Zhang, Z., & Salazar, L. G. L. (2014). Probabilistic approach to modeling pavement performance using IRI data. In Transportation Research Board 93rd Annual Meeting.

Quinino, R. C., Reis, E. A., & Bessegato, L. F. (2013). Using the coefficient of determination R2 to test the significance of multiple linear regression. *Teaching Statistics*, *35*(2), 84–88.

Rahman, M. A., Sun, X., Das, S., & Khanal, S. (2021). Exploring the influential factors of roadway departure crashes on rural two-lane highways with logit model and association rules mining. *International Journal of Transportation Science and Technology*, *10*(2), 167–183.

Rasdorf, W. J., Hummer, J. E., Zhang, G., & Sitzabee, W. (2009). *Pavement marking performance analysis*.

Robertson, J., Sarasua, W., Johnson, J., & Davis, W. (2013). A methodology for estimating and comparing the lifecycles of high-build and conventional waterborne pavement markings on primary and secondary roads in South Carolina. *Public Works Management & Policy*, *18*(4), 360–378. doi: 10.1177/1087724X12451441

Sarasua, W. A., Clarke, D. B., & Davis, W. J. (2003). Evaluation of interstate pavement marking retroreflectivity. In Report to South Carolina Department of Transportation, Report No. FHWA-SC-03-01 (2003).

Sasidharan, L., Karwa, V., & Donnell, E. T. (2009). Use of pavement marking degradation models to develop a pavement marking management system. *Public Works Management & Policy*, *14*(2), 148–173. doi: 10.1177/1087724X09349513

Sitzabee, W. E., Hummer, J. E., & Rasdorf, W. (2009). Pavement marking degradation modeling and analysis. *Journal of Infrastructure Systems*, *15*(3), 190–199.

Sitzabee, W. E., White, E. D., & Dowling, A. W. (2013). Degradation Modeling of Polyurea Pavement Markings. *Public Works Management and Policy*, *18*(2), 185–199. doi: 10.1177/1087724X12462831

Smadi, O., Souleyrette, R. R., Ormand, D. J., & Hawkins, N. (2008). Pavement marking retroreflectivity analysis of safety effectiveness. *Transportation Research Record*, *2056*(2056), 17–24. doi: 10.3141/2056-03

Thamizharasan, A., Sarasua, W. A., Clarke, D. B., & Davis, W. J. (2003). A methodology for estimating the lifecycle of interstate highway pavement marking retroreflectivity. *82nd Annual Meeting of the Transportation Research Board of the National Academies, Washington, DC*, *843*.

Thomas, G. B., & Schloz, C. (2001). *Durable, cost-effective pavement markings. Phase I: synthesis of current research*. Iowa State University. Center for Transportation Research and Education.

Van Petegem, J. W. H., & Wegman, F. (2014). Analyzing road design risk factors for run-off-road crashes in the Netherlands with crash prediction models. *Journal of Safety Research*, *49*, 121.e1-127. doi: 10.1016/j.jsr.2014.03.003

Wang, C., Wang, Z., & Tsai, Y.-C. C. (2016). Piecewise multiple linear models for pavement marking retroreflectivity prediction under effect of winter weather events. *Transportation Research Record*, *2551*(1), 52–61. doi: 10.3141/2551-07

Xu, L., Chen, Z., Li, X., & Xiao, F. (2021). Performance, environmental impact and cost analysis of marking materials in pavement engineering, the-state-of-art. *Journal of Cleaner Production*, *294*, 126302.

Yang, J., Yamamoto, T., & Ando, R. (2021). The impact of mandating a driving lesson for elderly drivers in Japan using count data models: Case study of Toyota City. *Accident Analysis and Prevention, 153*. doi: 10.1016/j.aap.2021.106015

Zhang, Y., & Wu, D. (2010). Methodologies to Predict Service Lives of Pavement Marking Materials. *Journal of the Transportation Research Forum*. doi: 10.5399/osu/jtrf.45.3.601

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Dealing with Heterogeneity*. 71–100. doi: 10.1007/978-0-387-87458-6_4

Zwahlen, H. T., & Schnell, T. (1997). Visibility of new centerline and edge line pavement markings. *Transportation Research Record*, *1605*, 49–61. doi: 10.3141/1605-07

APPENDIX

Appendix A1

Ranking of rural two-lane 0.5-mile road segments with average RL of WREL ≥450 mcd/m²/lux in November 2021, sorted by predicted RL values for December 2024 (from lowest to highest)

Rank	Route Identification Number	Starting point of 0.5-mile segment		Direction	RL of WREL (mcd/m²/lux)	
		Latitude	Longitude		Nov-21	Dec-24
1	AL0000030000	32.52648167	-86.45536833	SB	496	150
2	AL0002290000	32.42678167	-85.898285	SB	470	156
3	AL0002290000	32.523365	-85.89472	SB	535	159
4	AL0002290000	32.51616667	-85.89552833	SB	544	161
5	AL0002290000	32.50171667	-85.89380667	SB	485	162
6	AL0002290000	32.49449	-85.89281667	SB	503	166
7	AL0002290000	32.465725	-85.890285	SB	509	167
8	AL0002290000	32.50892	-85.89485167	SB	549	168
9	AL0000150000	32.73247833	-85.23268667	NB	479	168
10	AL0002290000	32.48732667	-85.89183833	SB	521	170
11	AL0002290000	32.480115	-85.89086167	SB	526	172
12	AL0002290000	32.47289667	-85.88985167	SB	537	174
13	AL0000030000	32.21785167	-86.348755	SB	465	175
14	AL0000150000	32.68367833	-85.309975	NB	501	176
15	AL0002290000	32.45893667	-85.89340833	SB	566	179
16	AL0002290000	32.58672333	-85.93828333	SB	458	180
17	AL0000150000	32.67988333	-85.31719167	NB	522	181
18	AL0000030000	32.51213	-86.45646333	SB	625	182
19	AL0000210000	32.25147833	-86.52353	SB	450	183
20	AL0000030000	32.51936833	-86.45704167	SB	631	183
21	AL0000090000	32.74713667	-86.10081667	NB	470	184
22	AL0002290000	32.45223667	-85.89656667	SB	621	185
23	AL0000140000	32.580055	-85.71172167	EB	460	187
24	AL0000030000	32.505	-86.45648	SB	652	188
25	AL0000140000	32.581775	-85.7199	EB	466	188
26	AL0000090000	32.71542167	-86.10802833	NB	493	189
27	AL0000210000	32.22625167	-86.54039333	SB	485	191
28	AL0000140000	32.55879	-85.79491167	EB	533	192
29	AL0002290000	32.69239333	-85.96013667	SB	478	194
30	AL0000210000	32.257995	-86.51976667	SB	498	194
31	AL0000060000	32.113505	-85.58548333	WB	457	194

Rank	Route Identification Number	Starting point of 0.5-mile segment		Direction	RL of WREL (mcd/m²/lux)	
		Latitude	Longitude		Nov-21	Dec-24
32	AL0002290000	32.60675833	-85.94304333	SB	512	195
33	AL0000140000	32.57779	-85.69498167	EB	479	195
34	AL0002290000	32.60089333	-85.93858833	SB	516	196
35	AL0000090000	32.71076167	-86.10147333	NB	525	196
36	AL0000090000	32.72896667	-86.112645	NB	523	196
37	AL0000030000	32.66191833	-86.489725	SB	451	196
38	AL0000030000	32.67638333	-86.49105333	SB	451	196
39	AL0000210000	32.23238167	-86.53589833	SB	512	197
40	AL0000150000	32.17816667	-85.70785333	SB	461	197
41	AL0000140000	32.58069333	-85.751615	EB	489	198
42	AL0000060000	32.12004667	-85.58910667	WB	512	199
43	AL0002290000	32.69623167	-85.96742833	SB	508	200
44	AL0000210000	32.24493333	-86.52730667	SB	527	200
45	AL0000090000	32.66703333	-86.10888667	NB	549	200
46	AL0000030000	32.497845	-86.45509833	SB	684	200
47	AL0002290000	32.593655	-85.93888333	SB	541	201
48	AL0002290000	32.660535	-85.94316667	SB	484	201
49	AL0000150000	32.19936833	-85.70809167	SB	475	201
50	AL0001650000	32.13932667	-85.06619	NB	450	202
51	AL0002290000	32.67470333	-85.94450667	SB	485	202
52	AL0002290000	32.70121667	-85.97352333	SB	517	202
53	AL0000150000	32.21391667	-85.70868167	SB	481	202
54	AL0000030000	32.64763	-86.48727167	SB	479	202
55	AL0001990000	32.47929	-85.72642667	NB	450	203
56	AL0000030000	32.127095	-86.39189	SB	532	203
57	AL0002230000	32.07999167	-85.77110167	SB	452	204
58	AL0000210000	32.06235167	-86.71920667	SB	456	204
59	AL0000030000	32.66911833	-86.49094333	SB	489	204
60	AL0001990000	32.486015	-85.73956833	NB	467	206
61	AL0000970000	32.079415	-86.44395667	NB	460	206
62	AL0000210000	32.05117	-86.80467667	SB	460	206
63	AL0000150000	32.206675	-85.70840167	SB	501	206
64	AL0000140000	32.579535	-85.70320833	EB	528	206
65	AL0002290000	32.44682333	-85.899855	SB	717	207
66	AL0000210000	32.03850333	-86.824805	SB	459	207
67	AL0000150000	31.87721833	-85.72134833	SB	455	207

Rank	Route Identification Number	Starting point of 0.5-mile segment		Direction	RL of WREL (mcd/m²/lux)	
		Latitude	Longitude		Nov-21	Dec-24
68	AL0000140000	32.561445	-85.78701167	EB	528	207
69	AL0000090000	32.72178833	-86.11206333	NB	577	208
70	AL0000140000	32.57251	-85.68919333	EB	537	208
71	AL0002290000	32.57318833	-85.93954167	SB	585	209
72	AL0002230000	32.03085167	-85.77407833	SB	464	209
73	AL0000030000	32.69084833	-86.49180167	SB	493	209
74	AL0002930000	32.34752667	-86.06412	SB	455	210
75	AL0000940000	32.01664167	-86.07472667	EB	457	210
76	AL0000100000	31.83578833	-86.87680833	EB	468	210
77	AL0002630000	31.994745	-86.74017	SB	470	211
78	AL0002630000	32.00597	-86.75097167	SB	469	211
79	AL0000210000	32.04841833	-86.81227667	SB	474	211
80	AL0001650000	32.13257	-85.06937667	NB	492	212
81	AL0002630000	32.01752	-86.761175	SB	474	212
82	AL0000210000	32.06568	-86.72674667	SB	490	212
83	AL0002630000	31.90277833	-86.678015	SB	477	213
84	AL0002630000	31.93408833	-86.69787167	SB	482	213
85	AL0002630000	31.94739167	-86.70493667	SB	477	213
86	AL0000210000	32.04345667	-86.81853667	SB	483	213
87	AL0000140000	32.56959333	-85.77314333	EB	555	213
88	AL0002930000	32.35417333	-86.06757667	SB	476	214
89	AL0002230000	31.995945	-85.77464167	SB	480	214
90	AL0002230000	32.016885	-85.771975	SB	481	214
91	AL0002630000	31.90784333	-86.68404	SB	488	215
92	AL0000140000	32.564575	-85.77931667	EB	571	216
93	AL0002630000	31.94075667	-86.70148	SB	496	217
94	AL0002630000	31.96145	-86.70880667	SB	494	217
95	AL0002630000	31.978145	-86.72367333	SB	496	217
96	AL0000970000	32.06065667	-86.43154667	NB	508	217
97	AL0002630000	32.040635	-86.79321167	SB	491	217
98	AL0002290000	32.633485	-85.95402833	SB	553	217
99	AL0000030000	32.19192	-86.35847833	SB	601	217
100	AL0002230000	31.98417167	-85.79548667	SB	500	218
101	AL0002230000	32.009675	-85.77097167	SB	499	218
102	AL0000210000	32.007515	-86.844435	SB	504	218
103	AL0002290000	32.6534	-85.944935	SB	563	219

Rank	Route Identification Number	Starting point of 0.5-mile segment		Direction	RL of WREL (mcd/m²/lux)	
		Latitude	Longitude		Nov-21	Dec-24
104	AL0000210000	32.012655	-86.83854833	SB	508	219
105	AL0002630000	32.04435667	-86.80054	SB	505	220
106	AL0000140000	32.57756167	-85.75897833	EB	585	220
107	AL0002630000	31.92077	-86.691375	SB	517	221
108	AL0002630000	31.927585	-86.69438833	SB	514	221
109	AL0002630000	31.98921167	-86.73471667	SB	515	221
110	AL0002290000	32.62693833	-85.95028167	SB	572	221
111	AL0000210000	32.02576167	-86.83122667	SB	519	221
112	AL0002290000	32.64073833	-85.95324833	SB	575	222
113	AL0002630000	31.88690833	-86.66061667	SB	550	224
114	AL0002290000	32.620215	-85.94791167	SB	585	224
115	AL0000140000	32.57298333	-85.765575	EB	603	224
116	AL0002630000	32.03321167	-86.77860333	SB	529	225
117	AL0000210000	32.00430167	-86.85191167	SB	538	225
118	AL0000210000	32.0192	-86.83486667	SB	538	225
119	AL0002630000	31.88104833	-86.65593333	SB	564	226
120	AL0002630000	31.897685	-86.67194667	SB	539	226
121	AL0002230000	32.00250333	-85.77154833	SB	531	226
122	AL0000210000	32.03293333	-86.829975	SB	541	226
123	AL0000150000	32.22115333	-85.709005	SB	601	229
124	AL0002290000	32.66773833	-85.94238333	SB	614	230
125	AL0002630000	32.023365	-86.76635333	SB	553	231
126	AL0002290000	32.64706333	-85.94905167	SB	626	233
127	AL0002630000	31.913995	-86.68842667	SB	601	240
128	AL0002630000	32.03690833	-86.78587167	SB	604	243
129	AL0002630000	31.89262	-86.66593833	SB	645	250

Appendix A2

Ranking of rural multilane 0.5-mile road segments with average RL of WREL ≥450 mcd/m²/lux in November 2021, sorted by predicted RL values for December 2024 (from lowest to highest)

Rank	Route Identification Number	Starting point of 0.5-mile segment		Direction	RL of WREL (mcd/m²/lux)	
		Latitude	Longitude		Nov-21	Dec-24
1	IN0000650000	32.39204667	-86.32639667	SB	451	81
2	IN0000650000	32.37867833	-86.32205667	SB	573	84
3	IN0000650000	32.407365	-86.34465167	SB	477	84
4	IN0000650000	32.40229167	-86.33859833	SB	537	93
5	IN0000650000	32.422925	-86.36274667	SB	467	96
6	IN0000650000	32.42950833	-86.36844	SB	492	99
7	IN0000650000	32.461355	-86.388885	SB	498	112
8	IN0000850000	32.37596333	-86.02723167	NB	450	124
9	IN0000850000	32.71716	-85.30358	NB	469	129
10	IN0000850000	32.71063667	-85.30725667	NB	485	131
11	IN0000850000	32.68462667	-85.32244333	NB	530	137
12	IN0000850000	32.536215	-85.54773167	NB	516	141
13	IN0000850000	32.532865	-85.55535167	NB	527	143
14	AL0000380000	32.68786667	-85.49692833	WB	457	156
15	AL0000030000	32.54097167	-86.45686	SB	452	159
16	IN0000650000	32.41759333	-86.356855	SB	940	160
17	AL0000030000	32.53378167	-86.45575167	SB	456	160
18	AL0000080000	32.26218833	-86.67040667	WB	465	163
19	AL0000060000	32.29251833	-86.19132	EB	554	165
20	AL0000080000	32.44389	-87.193655	WB	456	166
21	AL0000080000	32.43920833	-87.38155	WB	460	168
22	AL0000080000	32.44105833	-87.38979	WB	463	169
23	AL0000080000	32.43974667	-87.31307	WB	469	169
24	AL0000080000	32.44336833	-87.21938167	WB	467	169
25	AL0000080000	32.44349333	-87.21083667	WB	478	169
26	AL0000080000	32.44008667	-87.27023667	WB	486	172
27	AL0000080000	32.44026	-87.25310667	WB	492	173
28	AL0000090000	32.15447833	-86.27141333	SB	498	173
29	AL0000080000	32.44466167	-87.40626833	WB	509	175
30	AL0000080000	32.44344167	-87.39788167	WB	505	175
31	AL0000080000	32.43984333	-87.29596667	WB	506	175
32	AL0000080000	32.43991	-87.28734833	WB	508	175

Rank	Route Identification Number	Starting point of 0.5-mile segment		Direction	RL of WREL (mcd/m²/lux)	
		Latitude	Longitude		Nov-21	Dec-24
33	AL0000080000	32.44017667	-87.261675	WB	512	175
34	AL0000380000	32.68201	-85.49179	WB	610	175
35	AL0000410000	32.39862833	-86.99261667	SB	543	175
36	AL0000550000	31.63597833	-86.74899833	NB	527	176
37	AL0000080000	32.43955667	-87.33867	WB	516	177
38	AL0000080000	32.43981167	-87.30447167	WB	520	177
39	AL0000550000	31.62195167	-86.744725	NB	548	179
40	AL0000550000	31.62896	-86.746885	NB	544	179
41	AL0000380000	32.67658	-85.48615333	WB	661	180
42	AL0000080000	32.26221167	-86.67892333	WB	592	181
43	AL0000550000	31.60821667	-86.739155	NB	575	182
44	AL0000080000	32.31661	-86.88712167	WB	590	182
45	AL0000080000	32.26780833	-86.70340333	WB	612	183
46	AL0000550000	31.64300333	-86.75114667	NB	589	184
47	AL0000550000	31.61501667	-86.74210667	NB	592	185
48	AL0000080000	32.26454	-86.68701167	WB	635	187
49	AL0000080000	32.267165	-86.69494	WB	790	208