Machine Learning Architectures for Modelling International Roughness in Cold Region Pavements

Eyad Abu Rish

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

GRADUATE PROGRAM IN CIVIL ENGINEERING

YORK UNIVERSITY

TORONTO, ONTARIO

November 2022

Abstract

Governments often monitor pavement conditions and perform the required maintenance and rehabilitation using Pavement Management Systems (PMS). A PMS is an intricate system for evaluating pavement performance and maintaining road networks. An integral part of the PMS is the pavement performance prediction models. One of the most commonly used pavement performance indicators is the International Roughness Index (*IRI*). Currently used *IRI* models are often developed using regression analysis. Climate change is expected to affect the evolution of *IRI* over the life cycle of the pavement. Climate parameters are part of some of the *IRI* models found in the literature; however, their effects on the models' output were insignificant. Machine Learning (ML) has recently gained traction in the development of regression models. Recent studies have started using ML for *IRI* model development; however, the scope of the studies is limited and is often restricted to algorithms such as neural networks. Additionally, a systematic comparison between different ML algorithms in modelling *IRI* cannot be found in the literature. In this study, data provided by Alberta Transportation (AT) and the Federal Highway Administration (FHWA) are used to develop *IRI* models using regression analysis and ML methods. This research also examines several ML algorithms, including emerging algorithms that have yet to be employed for *IRI* model development. A systematic comparison between models developed using different methodologies is also part of this research. The issues of overfitting and model complexity are also addressed within the context of regression analysis and ML. Life Cycle Analysis (LCA) and a Life Cycle Cost Analysis (LCCA) are conducted to examine and highlight the relative financial advantage of utilizing site-specific models (developed in this research) over the widely used Mechanistic-Empirical Pavement Design Guide (MEPDG) *IRI* models. The later part of this research examines the implications of climate change on pavement roughness for pavements with different subgrade soil types. This is accomplished by developing *IRI* models for the pavement's center lane, which is expected to be affected predominately by climate due to limited exposure to traffic. The use of ML methods showed great potential for *IRI* model development. The best-performing algorithms for this research's application were the gradient-boosted

ensemble ML algorithms, specifically, the XGBoost and CatBoost algorithm. This research also supports the use of ML in developing *IRI* models as they have superior predicting capabilities and can provide much more value than traditional regression methods, such as regression analysis. In some instances, ML was found to produce meaningful results when regression analysis failed to do so. The results of the conducted LCA and LCCA found a relative financial advantage and more accurate environmental impact estimates from site-specific models developed in this research over the MEPDG *IRI* models. This research also examined the implications of climate change on pavement roughness for pavements with different subgrade soil types. The analysis of the LTPP data found that climate and subgrade soil type parameters are better at describing variations in the center lane *IRI* than the wheel-path *IRI* due to the absence of traffic. Furthermore, soils with higher plasticity are found to be more susceptible to freeze-thaw damage than soils with lower soil plasticity. Additionally, finer-grained subgrade soils were found to have a greater change in pavement roughness due to climate than coarse-grained subgrade pavements.

Keywords: International Roughness Index (IRI); Regression Analysis; Artificial Intelligence; Machine Learning; Supervised Learning; Alberta Transportation; Long-Term Pavement Performance (LTPP); Climate Change; Pavement Roughness; Mechanistic-empirical Pavement Design Guide (MEPDG).

Acknowledgments

I would like to express my sincere gratitude to my supervisor Dr. Rashid Bashir and committee member Dr. Mohammad Shafiee. Their constant encouragement and mentoring have assisted me throughout the completion of this research, especially amid a global pandemic. They provided me with guidance whenever it was required.

Thank you to Dr. Wei He of Alberta Transportation (AT) for generously supplying us with AT's data used in this study.

Also, thank you to Mr. Larry Wiser of the long-term infrastructure performance team of the United States Department of Transportation Federal Highway Administration for assisting with the LTPP data collection and interpretation

Last but not least, I am genuinely thankful for the constant support of my family and friends. This accomplishment would not have been possible without them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AADT	Average Annual Daily Traffic
AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
AI	Artificial Intelligence
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
AT	Alberta Transportation
BPR	Bureau of Public Roads
CC	Construction Cost
CCOLL	Construction Cost of the Last Overlay
CH	Organic clays of high plasticity
CI	Clays of medium plasticity, gravelly clays, sandy clays, silty clays
CI-CH	Clays of medium plasticity, gravelly clays, sandy clays, silty clays - Organic clays of high plasticity
CL	Clay of low plasticity, lean clay
CL-CI	Clay of low plasticity, lean clay - Clays of medium plasticity, gravelly clays, sandy clays, silty clays
ESAL	Equivalent Single Axle Loads traffic per day
FHWA	Federal Highway Administration
GBC	Granular Base Course
GC	Clayey gravels, gravel-sand-clay mixtures
GM	General Motors
GM	silty gravel
GP	poorly graded grave

GPS	General Pavement Studies
IPCC	Intergovernmental Panel on Climate Change
IRI	International Roughness Index
IRRE	International Road Roughness Experiment
KNN	K-Nearest-Neighbor
LCA	Life Cycle Analysis
LCCA	Life-Cycle Cost Analysis
LightGBM	Light Gradient Boosting Machine
LTPP	Long Term Pavement Performance
LWPc	Longitudinal Wheel path cracking as a percentage of length
M&R	Maintenance and Rehabilitation
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MEPDG	Mechanistic-Empirical Pavement Design Guide
ML	Machine Learning
ML	Inorganic silts and sandy silts of slight plasticity
MSD	Mean Square Deviation
MSE	Mean Square Error
NCHRP	National Cooperative Highway Research Program
NPV	Net Present Value
OtherCAr	Other Cracking as a percentage of Area
P_{200}	Percent passing 200
PI	Plasticity Index
PMS	Pavement Management System
PPDB	Pavement Performance Database

PPT	Annual precipitation
PQI	Pavement Quality Index
R^2	Coefficient of determination
RMC	routine maintenance cost
RMSD	Root Mean Square Deviation
RMSE	Root Mean Square Error
RUT	Is from automated data collection and represents the 80th percentile rut depth for the section (i.e.: 80% are less than & 20% are greater than the value), in mm.
SAI	Structural Adequacy Index
SC	Clayey sands, sand-clay mixtures
SC-CL	Clayey sands, sand-clay mixtures - Inorganic clays of low plasticity, gravelly clays, sandy clays, silty clays, lean clays
SDI	Surface Distress Index
SM	silty sand
SP	Poorly graded sands, little or no fines
SPS	Specific Pavement Studies
Surfthickness	Total surface thickness
SV	Salvage Value
SVM	Support Vector Machines
SVR	Support Vector Regression
TEC	Total Expected Cost
TELOL	Total Expected Life of the last Overlay
TrcAr	Transverse Cracking as a percentage of Area

ULLOL	Usable Life in the Last Overlay

Chapter 1 : Introduction

Governments and transport agencies spend a significant portion of their annual expenses on road networks. The expenditure includes both new road projects as well as road rehabilitation to improve road quality and extend their longevity (AASHTO 2009, ARA Inc. 2020). Generally, transportation agencies maintain detailed records of their network's road conditions and activities in a Pavement Management System (PMS) database. The system consists of varied aspects of pavement design, planning, construction, and maintenance for the road network. The structure of the PMS is heavily dependent on the adopting agency; generally, the PMS provides a set of tools to assist in pavement-related decision-making and relies heavily on pavement performance prediction models. Predictive pavement performance models are crucial for forecasting pavement deterioration over a desired period of time.

One of the most commonly used pavement performance predictive models includes the International Roughness Index (*IRI*). The *IRI* is a key indicator of pavement roughness and is used to assess the health and serviceability of roads. The *IRI* can be expressed as a mathematical equation that represents pavement roughness conditions over a desired road length and can be written out as follows:

$$IRI = \frac{1}{L}\int_0^{L/V} |\dot{z}_s - \dot{z}_u|\, dt \tag{1.1}$$

The equation focuses on the simulated motion between the sprung and un-sprung masses in a quarter-car model normalized by Length L in metres, with t representing time in seconds, V being the speed of the quarter-car in metres per second, $\dot{z}_s$ is the vertical speed of the sprung mass in metres per second, and $\dot{z}_u$ is the vertical speed of the un-sprung mass in metres per second (Sayers 1995a). The index itself can quantify the pavement's deterioration over a selected period of time due to age, climate, distresses, and traffic loads. These variables are utilized to develop the *IRI* models by empirically assigning different weights to each variable. The *IRI* assists with various stages of pavement planning, whether it is in assessing the road network's performance or estimating and predicting the future condition of a pavement section. In the past,

various attempts have been made to develop an *IRI* with high accuracy, regardless of the road network's region. However, as each region exhibits a distinctive blend of traffic, climate, and distress patterns, locally developed models tend to be more representative of the pavement conditions. The *IRI* models developed using data from transport agencies produce higher accuracy for the agencies' road network as the data used to develop the models are from the same road network. Global-scale pavement performance models, such as the Mechanistic-Empirical Pavement Design Guide (MEPDG) *IRI* model, are created with data retrieved from the Long-Term Pavement Performance (LTPP) database. The MEPDG *IRI* model utilized data from the LTPP database, which contains measurements from various locations across North America. The MEPDG *IRI* model is developed using regression analysis and not making use of emerging technologies and techniques, such as Machine Learning (ML) modelling. The reliance of the MEPDG *IRI* model on regression analysis limits the model in terms of accuracy and predictive capabilities. The performance of the MEPDG *IRI* models varies across North America; thus, local models and calibrations are preferred for better prediction of pavement *IRI*. An *IRI* model with higher accuracy would assist with better Maintenance and Rehabilitation (M&R) decisions, resulting in more realistic expenditure estimates by the transport agencies and securing funding when most needed. Additionally, the *IRI* models can help identify road sections that need urgent maintenance and repairs from a road user safety perspective.

1.1 Thesis Objectives

The objective of this study is to develop more accurate *IRI* models using regression analysis and ML methods. The utilization of ML methods is due to the recognition of the limitations of modelling *IRI* using regression analysis. Additionally, the effectiveness of using regression analysis and ML in modelling *IRI* is examined and analyzed. This study also examines ML performance over regression analysis, and any additional insight ML provides into the relationships between the independent variables and the target variable. The objectives are achieved through the use of two datasets of collected field data. The framework of this research is presented below in Figure 1.1.

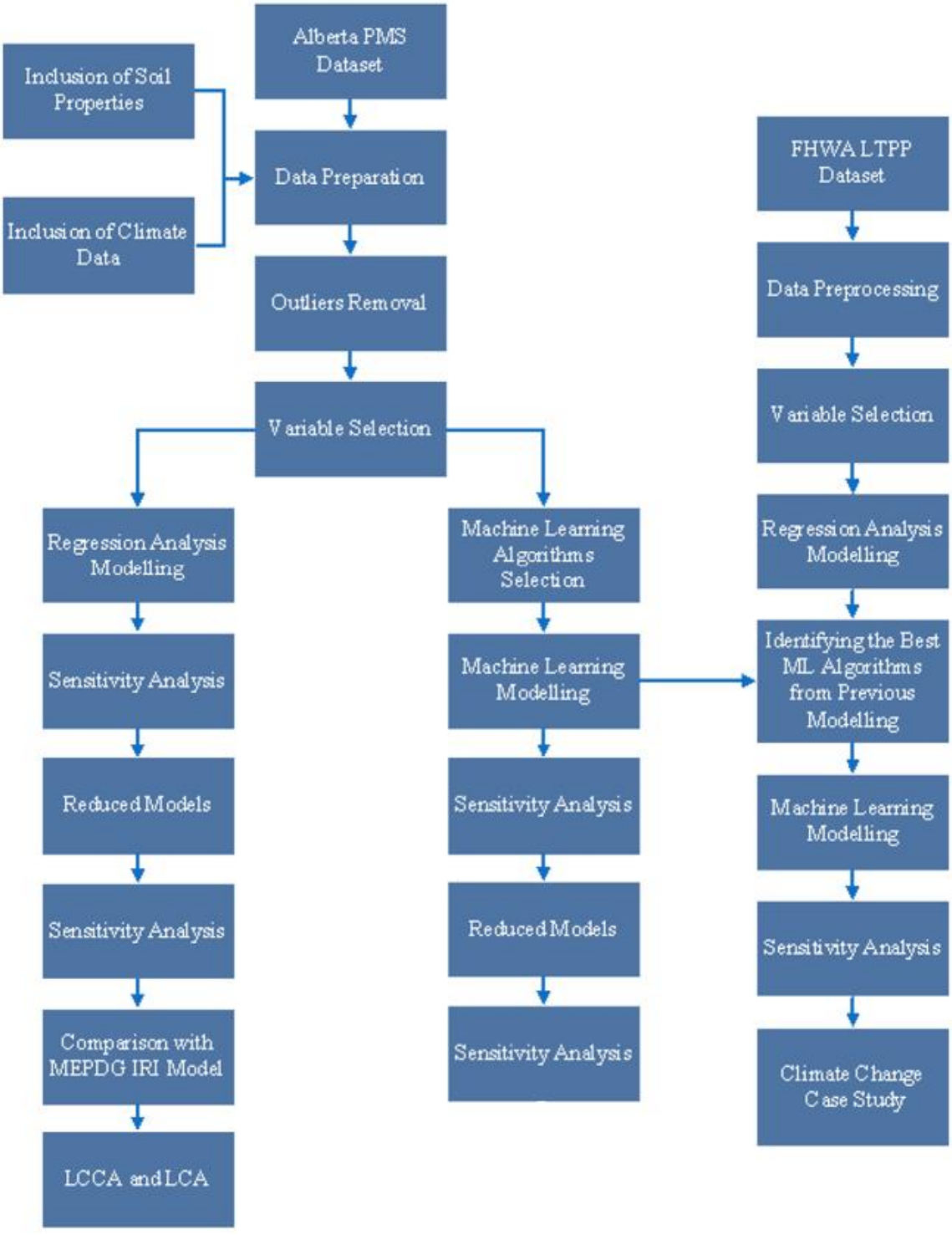

Figure 1.1 - Research Framework

The datasets used to achieve this study's objectives were extracted from a representative PMS database for the province of Alberta and the LTPP database containing *IRI* data from various locations across the contiguous United States. Alberta's PMS dataset is used to develop ML and regression analysis models to assess the performance gain in using ML over regression analysis in modelling *IRI*. A comprehensive literature review has also been conducted to review research efforts in developing *IRI* indices. There has been little attention given to emerging ML algorithms in modelling *IRI*. A systematic comparison is conducted to test the emerging ML algorithms and their performance compared to regression analysis and other widely used ML algorithms. Sensitivity analysis studies are also conducted for the developed ML and regression analysis *IRI* models to examine if the ML models provide a comprehensive view of the variables' effect on pavement roughness. The regression analysis *IRI* models are also used to conduct a case study using a Life Cycle Analysis (LCA) and a Life Cycle Cost Analysis (LCCA) to examine the environmental and economic impacts, respectively. The case study also incorporates the MEPDG *IRI* model to be compared to the models developed in this study using Alberta's PMS dataset. The dataset extracted from the LTPP is used to develop *IRI* models for the wheel paths and center lane to evaluate the changes in *IRI* over time due to environmental and subgrade parameters using both regression analysis and ML. The dataset is also used to confirm the improvement in predictive capabilities in modelling *IRI* using ML over regression analysis. Additionally, this research conducts a case study into the effects of climate change by first understanding the effect of climate on pavement roughness in the absence of traffic and then assessing the impact of the regional climate change in the contiguous United States on pavement roughness over time for different subgrade soil type pavements.

1.2 Thesis Outline

Chapter 1 presents a general introduction of the thesis's topic accompanied by the thesis outline and objectives. Chapter 1 lists the outlines of the following five chapters in chronological order. Chapter 2 begins with a background on PMS and pavement distresses, followed by a background on the *IRI*; the background includes its history, significance, and existing *IRI* models reported in the literature.

Furthermore, different types of regression analysis, model selection methods, and model evaluations are also discussed in depth. Additionally, this chapter also discusses Artificial Intelligence (AI), ML, and the various algorithms that are used in this study. Chapter 3 presents the methodology and development of regression analysis models. The chapter also includes descriptions of the measured *IRI* data. Additionally, the provided dataset by AT did not include climate or soil properties data; thus, in this chapter, the climate data and soil properties are compiled and added to the dataset. Chapter 3 also includes the identification and removal of data outliers, model selection methods, model training using regression analysis, and the validation for the *IRI* regression models. After developing an *IRI* model through regression, it is then compared with the commonly used MEPDG *IRI* model as a benchmark. An LCA and an LCCA are then conducted to compare the regression developed and the MEPDG *IRI* models to examine the impacts caused by differences in predictive capabilities. Chapter 4 presents ML modelling methodologies and their effectiveness in modelling *IRI*. Various algorithms families are examined in this chapter, including decision trees, ensemble methods, neural networks, nearest neighbours, and support vector machines. Emerging ML algorithms under supervised learning such as XGBoost, CatBoost, LightGBM, and Random-Forest are also discussed and used in this chapter in addition to decision trees, Ridge, lasso, Artificial Neural Networks, Elastic-Net, support vector machines, and K-nearest-neighbor. The models developed through these methods are then assessed and compared to the *IRI* models developed through regression analysis and the MEPDG *IRI* model. Chapter 5 examines the effectiveness of the best-performing ML algorithms identified in Chapter 4 in modelling *IRI* where regression analysis fails to provide meaningful correlations. This chapter is an attempt to use ML algorithms to identify and correlate variables that contribute to *IRI* in instances where traditional regression analyses are ineffective. The chapter uses data from the FHWA under the LTPP, to model *IRI* due to climate change for different subgrade types of pavements in the contiguous United States. This chapter also presents the correlation between climate for different soil types with the changes in *IRI* both at the center of the lane and along the wheel path. Climate and subgrade soil parameters' impact on *IRI* is also examined through sensitivity analyses. Additionally, a climate change case study is conducted for two states in the United States to

identify the impact of climate change on pavement roughness. Chapter 6 presents the summary and conclusion of this research. The chapter also outlines recommendations for future research.

Chapter 2 : Background Information

2.1 Introduction

This research reviewed several publications on pavement management, pavement performance indicators, regression analysis, and Artificial Intelligence (AI). The literature review first examines pavement management and pavement management systems, including their development and utilization. The Pavement Management System (PMS) database is also described in this chapter, including pavement distress definitions. The development history and significance of the *IRI*, along with previously developed models, are all described in detail in this chapter. Additionally, regression analysis is also discussed and includes a discussion on linear regression models and model selection methods. Afterwards, an extensive literature review on AI and Machine Learning (ML) is presented. Various categories of ML algorithms, such as unsupervised, supervised, and reinforcement learning are discussed. Following this, supervised learning algorithms are described and discussed, including the fundamentals associated with each algorithm. Model evaluation and validation methods are then described, along with common statistics used to estimate the model's predictive capabilities and accuracy. The primary aim of this literature review is to provide an overview of the existing research and highlight the importance of this research.

2.2 Pavement Distresses

Pavement distresses are key contributors to pavement roughness. Pavement distress conditions aid with rating pavements through pavement performance indicators such as the *IRI*. Pavements can be maintained to ensure safe and serviceable roads by using a pavement rating system. Different agencies define the impact of different distress types on road networks and their significance towards the rating system. The measurements of distresses are done through manual or automated collection systems. An example of an automated collection system is the high-speed inertial profiler that is often utilized to collect data such as the pavements' rut depth (Huang 2003). The profiler collects a variety of measurements through the use of sensors such as gyroscopes, ultrasonics, lasers, GPS, HD camera, and

accelerometers along road paths (Huang 2003, City of Edmonton 2021). An example of a high-speed inertial profiler is shown in Figure 2.1.

Figure 2.1 - Automatic Road Analyzer (ARAN) (City of Edmonton 2021)

A systemized approach is adopted to record and measure distresses. The roads are divided into segments to track the pavement segment's performance and maintenance costs over time. The road representative segments are set as permanent segment boundaries, given that the road segments are not altered. The transport agency's raters assign the segments' representative gauging length, and the distresses measurements are taken over this defined road segment, as shown in Figure 2.2 (Huang 2003, Haas et al. 2015).

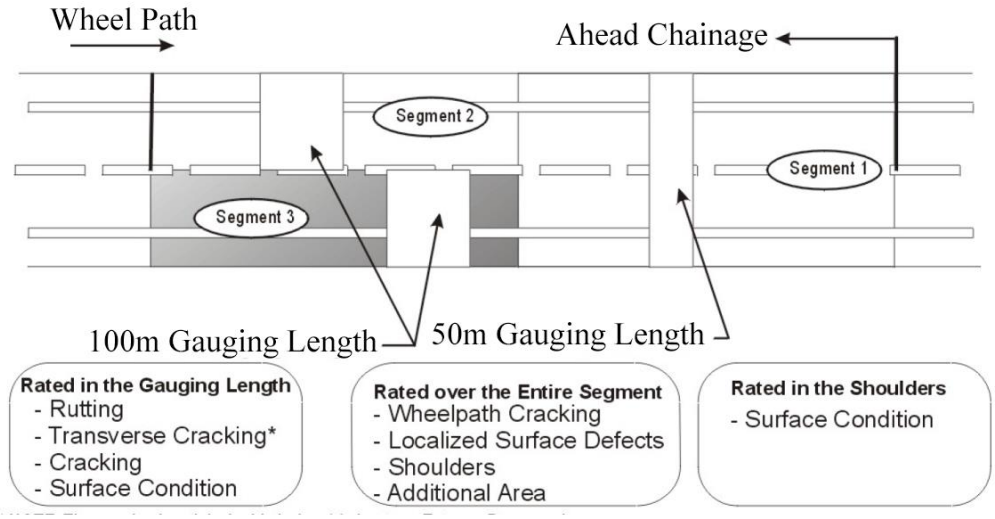

Figure 2.2 - Road segments over Gauging Length (modified from Alberta Transportation 2003)

The pavement distresses collected include but are not limited to rutting, transverse cracking, longitudinal cracking, and other cracking. Rutting is the longitudinal depression along the pavements' wheel path caused by repetitive traffic loads. A possible cause of rutting is the densification of the subgrade and pavement layers under the traffic loads. Rutting depressions can be present in either a single or a double form along the wheel path (Huang 2003, Haas et al. 2015). A representative cross-section affected with rutting is presented in Figure 2.3.

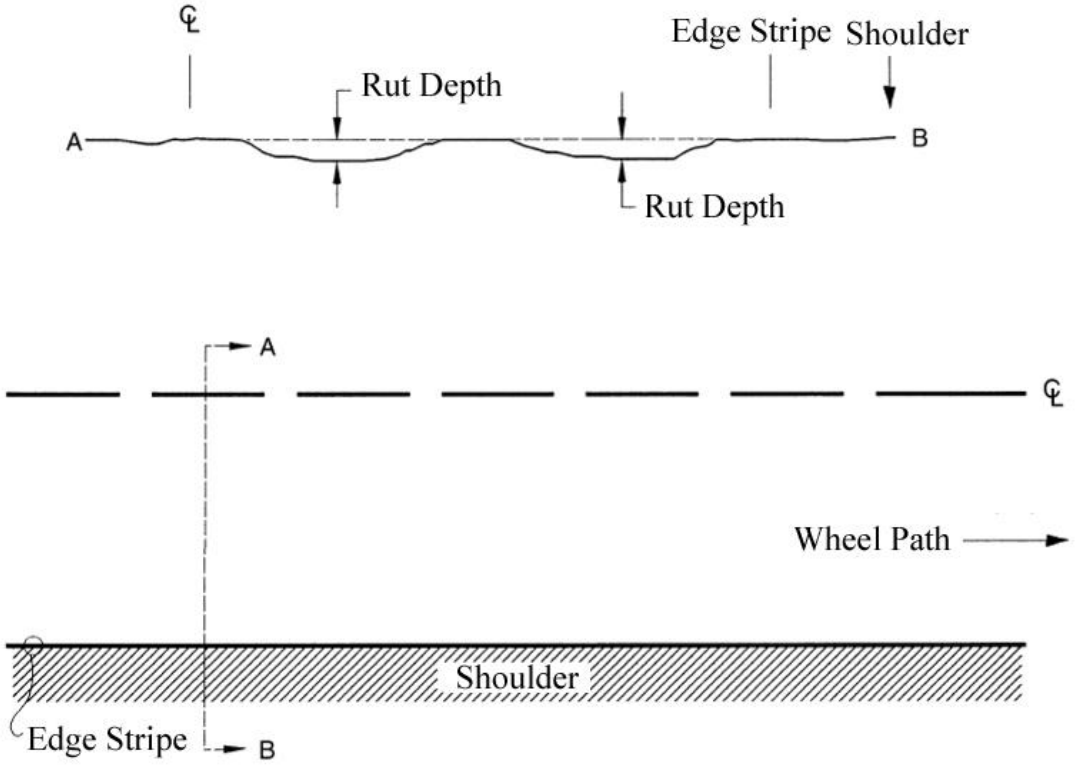

Figure 2.3 - Rutting representative cross-section (modified from Miller and Bellinger 2014)

The measurements of rut depth are taken at multiple sections along the gauging length. Eight rutting measurements are taken within the gauging length in both the inner and outer wheel paths, as seen in Figure 2.4 (Huang 2003, Haas et al. 2015). The categorization of rutting depth severity varies across agencies (Miller and Bellinger 2014). Rutting depth of high severity varies in definition between

agencies. Generally, the high severity threshold for rutting depth varies between 12 and 13 millimetres; any rutting depths greater than the given threshold are considered high severity (Alberta Transportation 2003, Miller and Bellinger 2014).

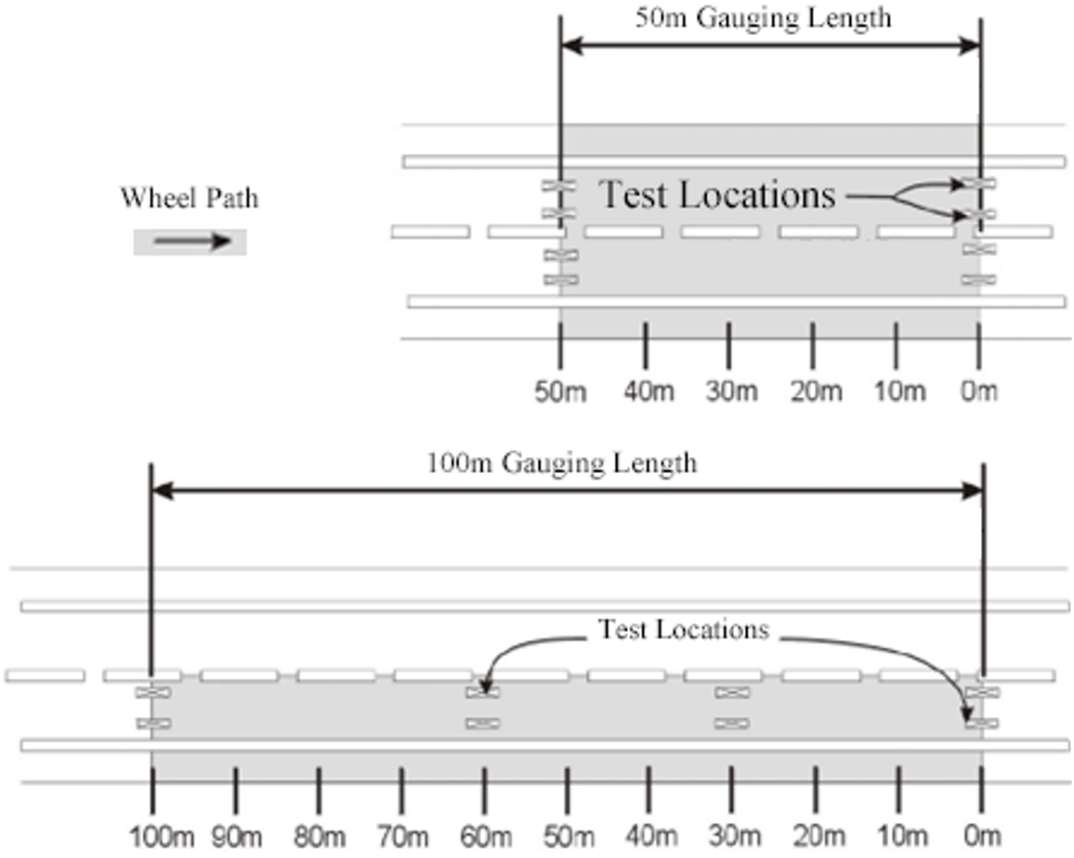

Figure 2.4 - Rutting measurement locations (modified from Alberta Transportation 2003)

Transverse cracking is a non-wheel load-related cracking that is caused by low temperature or thermal cycling (ARA Inc. 2020). Transverse cracks are approximately perpendicular to the pavement centerline and tend to be regularly spaced along the road's length (Miller and Bellinger 2014). An example of transverse cracking is shown in Figure 2.5. The representative roadway section presents several transverse cracking in the road segment. As observed in the figure, the transverse cracking is approximately perpendicular to the road's width.

Figure 2.5 - Transverse Cracking field example (National Cooperative Highway Research Program 1997)

The measurement of transverse cracking is carried out using different techniques. Generally, transverse cracks are counted along the width of a road section (Miller and Bellinger 2014). A representative cross-section affected with transverse cracking, along with some transverse cracking measurements, is presented in Figure 2.6.

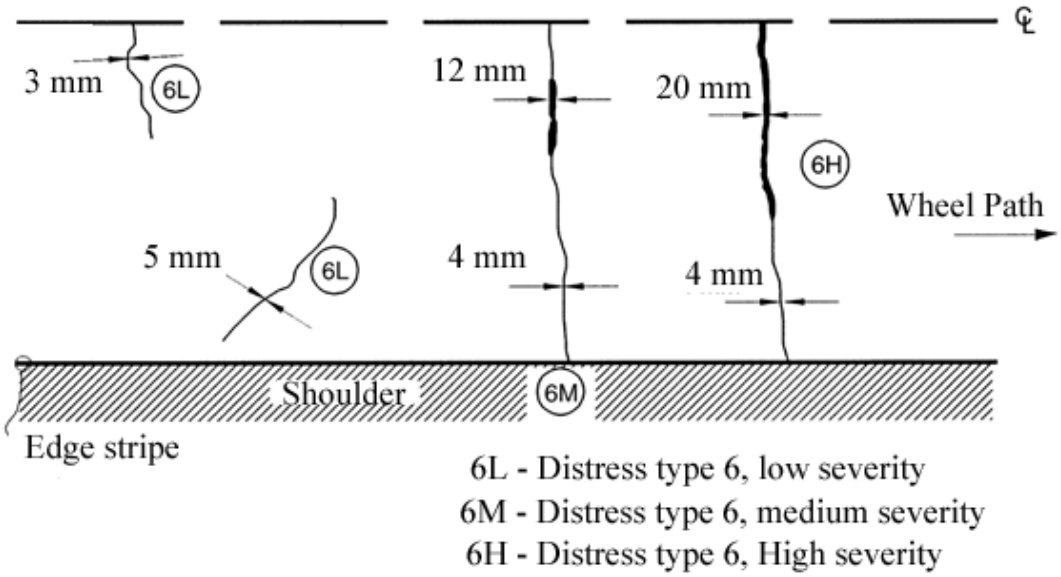

Figure 2.6 - Transverse cracking cross-section (modified from Miller and Bellinger 2014)

Longitudinal cracking is another type of cracking that is distinguished in PMS databases. A longitudinal crack is a crack that extends parallel to the pavement's centerline along the traffic's wheel path direction (City of Edmonton 2021). An example of longitudinal cracking in a roadway segment is presented in Figure 2.7.

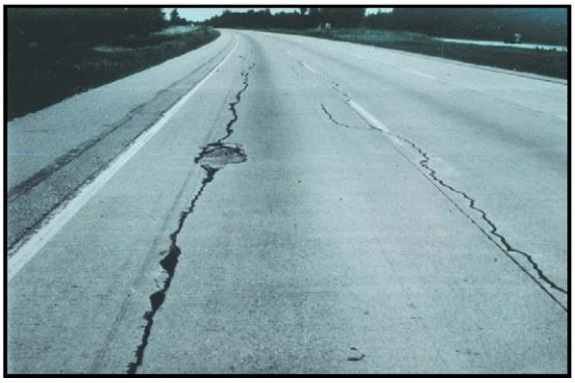

Figure 2.7 - Longitudinal cracking example (National Cooperative Highway Research Program 1997)

Longitudinal cracking is often broken down into wheel path and non-wheel path longitudinal cracking (Miller and Bellinger 2014, Haas et al. 2015). A representative cross-section affected by both wheel path and non- wheel path longitudinal cracking is presented in Figure 2.8.

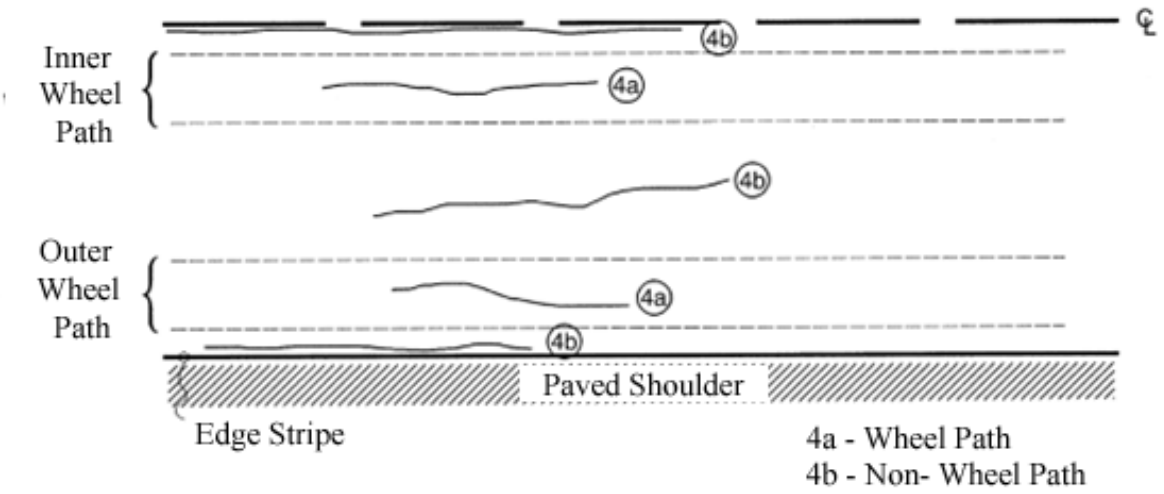

Figure 2.8 - Longitudinal cracking cross-section (modified from Miller and Bellinger 2014)

Other forms of cracking can be described as “Other Cracking Area” and include most pavement cracking types except the aforementioned distresses. This additional cracking includes centerline cracking, block or alligator cracking, braided or tree branch cracking, shoulder line cracking, and short transverse cracking that does not extend across the lane. Centerline cracking is simply cracking that occurs along the centerline. Moreover, block or alligator cracking is cracking that is in the form of a pattern that divides the pavement into rectangular blocks. Braided or tree branch cracking is cracking that occurs in a

branching pattern similar to a tree branch. Similar to centerline cracking, shoulder line cracking occurs along the shoulder line. Finally, short transverse cracking is cracking perpendicular to the wheel path that does not extend across the width of the lane (Alberta Transportation 2003, Huang 2003, Miller and Bellinger 2014, Haas et al. 2015). An example of a pavement with multiple cracking distresses is presented in Figure 2.9.

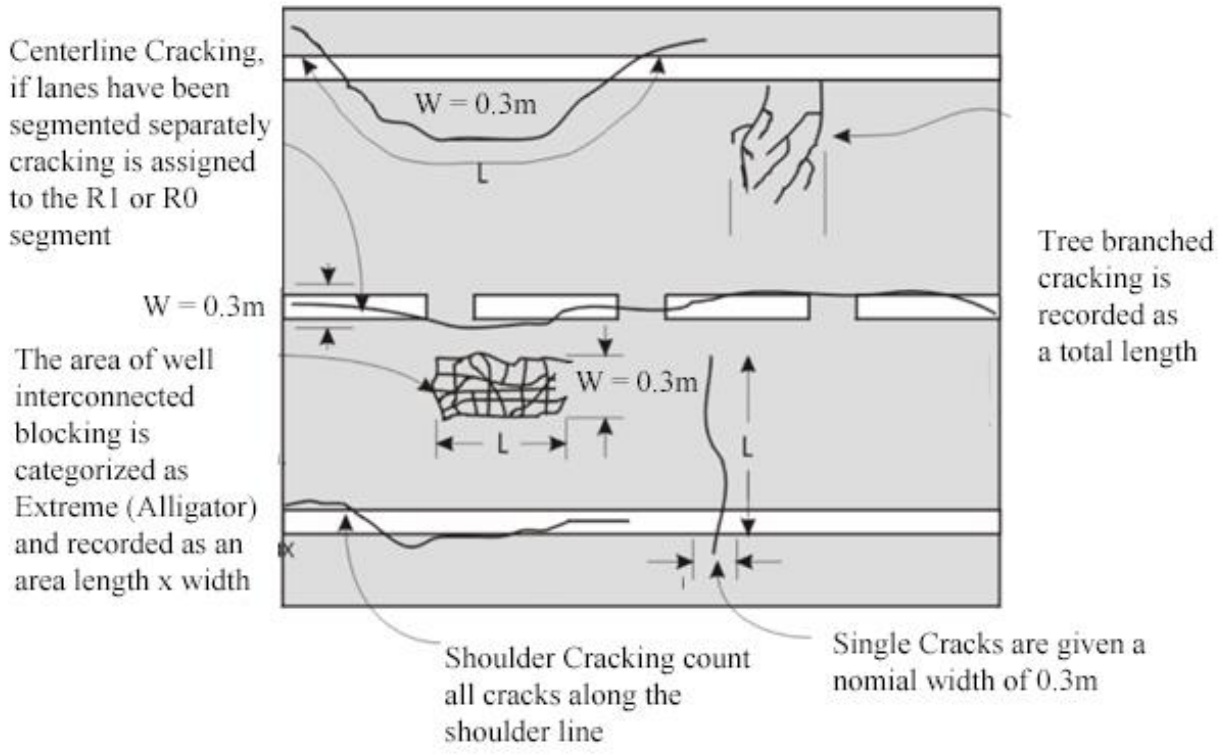

Figure 2.9 - Pavement cross-section with mixed cracking (modified from Alberta Transportation 2003)

2.3 Pavement Management System

Pavements are a critical component of transportation infrastructure systems. The deterioration of pavement occurs as a result of different environmental factors and distresses. Transport agencies are responsible for providing a network of safe and serviceable pavements for road users. Transport agencies face continuous challenges in maintaining and managing pavements within their pavement network system. With deteriorating road infrastructure and limited resources, agencies turn to decision-making tools to make the best use of the available resources. The management of pavements includes all activities

related to planning, design, construction, budgeting, monitoring, maintenance, and rehabilitation (Huang 2003). The development and maintenance of the PMS and its components were often challenging for transportation agencies; thus, through various studies, a standardized PMS was developed (Finn 1998, ARA Inc. 2001a).

The PMS provides a documented methodological process for coordinating and documenting pavement activities (Peterson 1987). The PMS is a multivariate tool that provides a systematic approach to making pavement-related decisions more consistent, cost-effective, and transparent (Huang 2003). The PMS also provides the required information to support fund requests and justify Maintenance and Rehabilitation (M&R) programs (Huang 2003). The PMS does not provide a final decision but rather enables the decision-makers to understand the alternatives and their impact. The American Association of State Highway and Transportation Officials (AASHTO) defines the PMS as "a set of tools or methods that assist decision-makers in finding optimum strategies for providing, evaluating, and maintaining pavements in a serviceable condition over a period of time" (AASHTO 2012). The utilization of the PMS varies from one agency to another but generally aids with the following activities (AASHTO 2012):

- Assessing the current pavement conditions;
- Predicting the pavement conditions:
- Estimating the funds needed to achieve the targeted pavement conditions;
- Identify the pavement M&R needs;
- Optimize the use of allocated funds for pavement activities;
- Highlight the consequences of alternative pavement management strategies; and
- Evaluate the impact on pavements resulting from changes in pavements' material properties, design, construction strategy, M&R activities and other related pavement activities.

Through their publications, AASHTO provides examples of guidelines for the PMS (AASHTO 1990). The PMS varies from one agency to another depending on the utilizing agency; however, the PMS often

follows a list of standard components. The standard components of a PMS can be described as follows (AASHTO 1990):

- An inventory of all of the pavements within the agency's road network categorized according to pavement type, location, functional classification, length, and distress level;
- A detailed database of pavement conditions, observed traffic patterns, construction practices, maintenance and rehabilitation, and any other relevant information:
- The pavement network "health" condition based on systematic thresholds such as pavement threshold identifiers;
- The forecasted pavement "health" condition through pavement performance predictive models;
- The required budget needed for M&R activities to achieve acceptable road conditions;
- Forecasted budget over a multi-year period;
- Pavement plans including design, construction, and maintenance for a single or multi-year period;
- Strategies for prioritizing expenditure when fundings are less than the required budget;
- A methodology for communication between different groups within the agencies, such as design, planning, construction, and maintenance groups;
- Communication strategy with groups outside an agency, such as, local governments, legislature, media, and public interest groups; and
- Strategies for comparing alternative preservation decisions for M&R activities and reconstruction of pavements in the network.

The PMS can be applied in various areas of planning, designing, budgeting, scheduling, and performance evaluation of pavements. The PMS can be used to prioritize funding and outline strategies to meet safety and comfort standards for road conditions. Unique requirements are set within the PMS by the utilizing agency, and they can be updated regularly to enhance the efficiency of the PMS and, in turn, further enhance the efficiency of the utilizing agency. These updates are the result of a variety of outputs, including but not limited to database development (including a larger collected dataset), research, and

technological advancements (AASHTO 1990). Technological advancements include increased computer hardware performance and capabilities, advances in software, the implementation of geographic information systems, and the development of more accurate pavement prediction models (AASHTO 1990). In general, the PMS includes several main components, including analysis and feedback. A schematic representation of the PMS modules is presented in Figure 2.10. This figure outlines the link between the database and the feedback process in the PMS. The PMS database is used for analysis and can be broken down into pavement condition analysis, priority assessment models, and network optimization models. The analyses aid in identifying key areas on which the PMS database can be improved and are done through the feedback process.

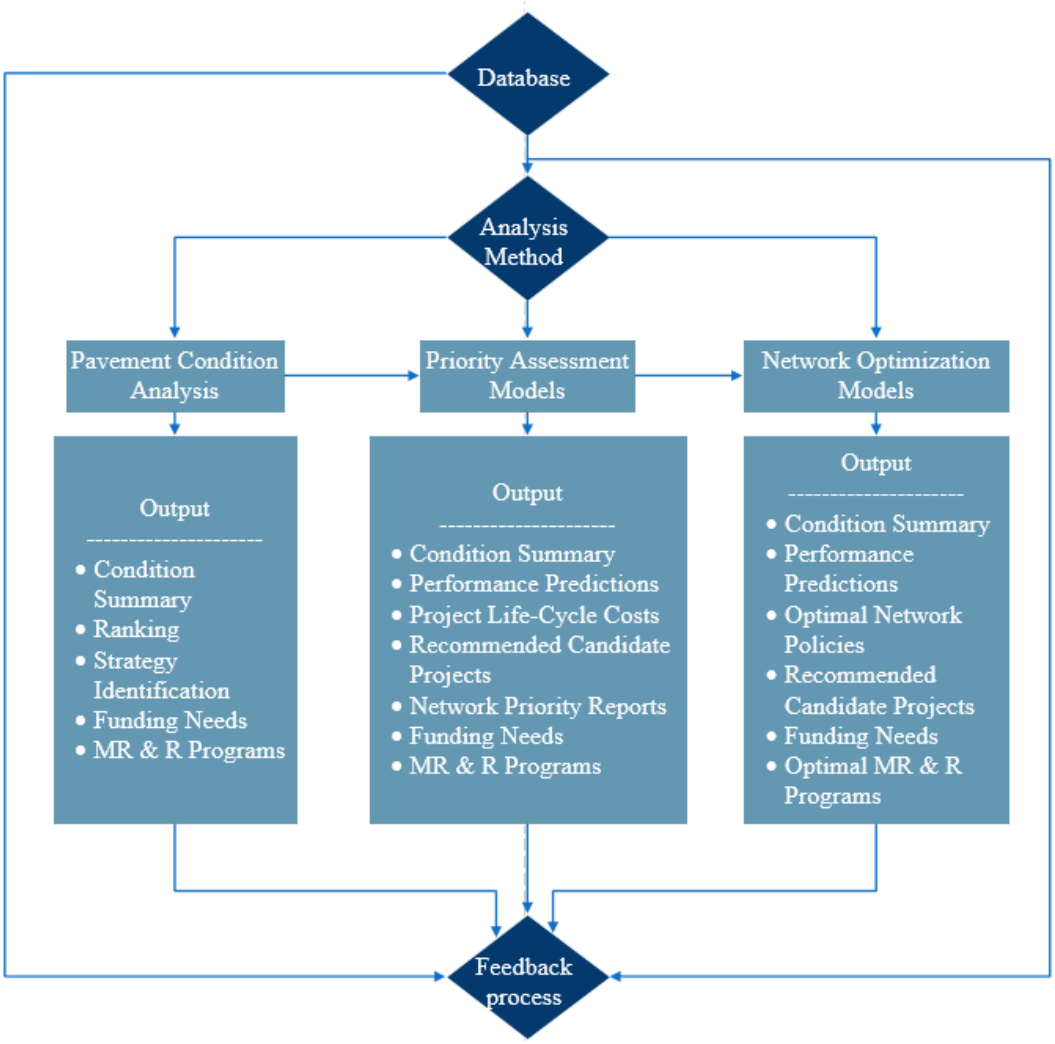

Figure 2.10 - Schematic Representation of PMS Modules (modified from AASHTO 1990)

2.4 International Roughness Index (IRI)

Pavement roughness is a critical factor in determining the health condition of a road network and is considered the primary indicator of pavement serviceability (Haas et al. 2015). Pavement roughness is characterized by distortions in the pavement surface that leads to an undesirable and uncomfortable ride for road users (Haas et al. 2015). The distortions contribute to a vertical acceleration of the riders, which is perceived as uncomfortable by the users (Haas et al. 2015). Thus, it is critical for transport agencies to monitor the road network roughness to maintain the roads' serviceability and to perform the required maintenance for acceptable road conditions. There are three components needed to evaluate pavement roughness: accurate profile measurements, a mathematical model, and roughness statistic interpretation (Huang 2003). Pavement roughness indices, such as the *IRI*, aid in establishing standards and thresholds across the transportation agency's road network. The *IRI* provides a standardized method to measure road roughness and can identify and predict roughness trends (Huang 2003). Theoretically, a perfectly smooth pavement should have an *IRI* value of 0, but realistically, even immediately after construction, the pavement roughness will have some roughness. The *IRI* value increases proportionately with roughness and is affected by factors such as traffic and environmental elements. The index is typically calculated using field measurements that are achieved using wavelengths and amplitudes measurements on the pavement's surface. Deformed or rougher road surfaces propagate back unique wavelength combinations that can be translated to road roughness (Paterson 1986). The translated road roughness values are then used to assess pavement health and serviceability.

2.4.1 IRI History

The origins of the *IRI* date back to the 20th century. The Bureau of Public Roads (BPR) attempted to standardize pavement performance measurements by introducing the Roughometer in 1941 (Gillespie 2001). The Roughometer is a single-wheel trailer whose dimensions, tire, mass, and suspension properties have been standardized to achieve comparable pavement performance measurements (Gillespie 2001). The BPR Roughometer is presented in Figure 2.11.

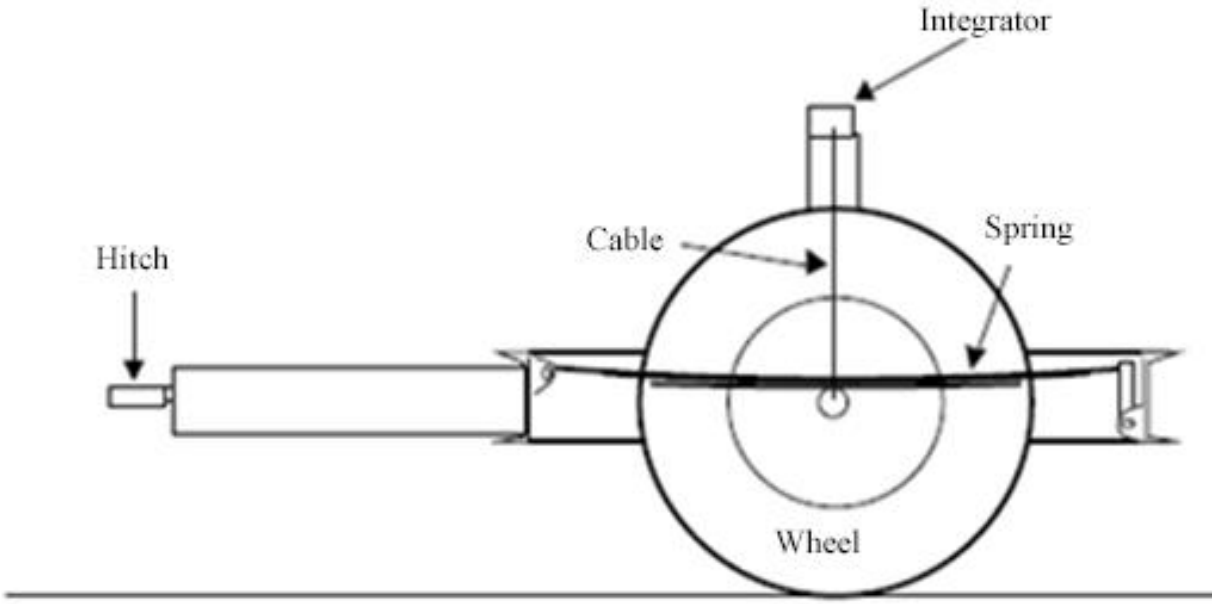

Figure 2.11 - BPR Roughometer (modified from Gillespie 2001)

Around the 1950s, the American Association of State Highway Officials (AASHO) began examining pavement roughness measurements in terms of pavement serviceability (Carey Jr and Irick 1960). By the 1960s, General Motors (GM) developed high-speed road profilers that were capable of measuring road profiles through vehicle vibrations (Spangler and Kelly 1964). Soon thereafter, the application of this system was combined with the quarter-car model (a special purpose analog computer) that replicated the BPR Roughometer (Darlington 1970, Sayers et al. 1986). Two versions of the quarter-car were available by the late 1960s, a BPR Roughometer and a 1968 Chevrolet Impala (Sayers et al. 1986). In the late 1970s, the foundational research of *IRI* took place under the National Cooperative Highway Research Program (NCHRP) to develop calibration methods for systems such as the BPR Roughometer (Gillespie et al. 1980). Shortly after, the World Bank expressed interest when the need for standardized and comparable roughness measurements became apparent for funding purposes (Sayers et al. 1986). The World Bank initiated the International Road Roughness Experiment (IRRE) experiment in Brazil in 1982 to develop a standardized scale (Sayers et al. 1986). The result of the study was the development of the *IRI* using the quarter-car simulation operating at a standard speed of 80km/hr (Sayers et al. 1986).

2.4.2 IRI Significance

The *IRI* is the most commonly used pavement performance index in representing road networks' conditions (AASHTO 2012). Road conditions can be classified through *IRI* thresholds; these thresholds are used to make M&R decisions as a part of the PMS. The *IRI* thresholds can be used to determine riding quality and help maintain targeted greenhouse gas emission levels due to fuel inefficiency. Greenhouse gas emissions are directly linked to pavement roughness, as rougher pavements contribute to an increase in fuel consumption (AASHTO 2009). Maintaining roads within the road network and keeping track of the networks' pavement roughness is in the interest of road users. An example of the indirect cost of poor road conditions is given in this quote by AASHTO: "The American public pays for poor road conditions twice—first through additional vehicle operating costs and then in higher repair and reconstruction costs" (AASHTO 2009).

2.4.3 Existing Models

Several attempts have been made to develop more accurate *IRI* models by utilizing various datasets. The most frequently cited *IRI* models are based on LTPP and local transport agencies' datasets worldwide. A summary of the most cited *IRI* predictive models is presented in Table 2.1. Models' information presented in Table 2.1 includes the coefficient of determination (R^2) and the number of data points (N) used to develop the model, where available. The models' R^2 presented from the literature ranges from 0.35 (George 2000) to 0.99 (Kargah-Ostadi et al. 2010, Mazari and Rodriguez 2016). Generally, a model's predictive capability is better for the data used to develop the model. The predictive *IRI* models are developed using various techniques, including regression analysis and ML. The variables used in the modelling *IRI* varied significantly between the models. Some of the more common variables in modelling *IRI* are *Age*, *ESAL*, *AADT*, structural number, transverse cracking, precipitation, pavement thickness, and rutting.

Table 2.1 - *IRI* models from the literature

Model	***Model Variables***	R^2	***N***
(Abdelaziz et al. 2020)	Age, fatigue %area, IRI_0, Rut Depth, and transverse cracking length	0.75	2439
(Albuquerque and Núñez 2011)	ESAL, mean precipitation, potential evapotranspiration, and SN	0.87-0.94	20
(Al-Suleiman (Obaidat) and Shiyab 2003)	Age	0.61-0.80	440
(Choi et al. 2004)	AC, ESAL, P_{200}, SN, and TO	0.71	117
(Choi and Do 2019)	AADT, Avg. max. temp., Avg. min. temp., Avg. temp, Deciding agent, ESAL, and Total rainfall	0.87	1880
(Chou and Pellinen 2005a)	Age, ESAL, FI, IRI_0, NoFT, and Precipitation	0.98	90
(George 2000) (1)	Age, ESAL, MSN, RES, and TO	0.48	4109
(George 2000) (2)	Age, ESAL, and MSN	0.35	690
(Georgiou et al. 2018)	IRI_{t-6}, IRI_{t-5}, IRI_{t-4}, IRI_{t-3}, IRI_{t-2}, and IRI_{t-1}	0.93-0.94	-
(Gong et al. 2018b)	Age, Block, Edge, Fatigue, FI, pavement thickness, IRI_0, ESAL, longitudinal cracking, patch, polish, potholes, precipitation, Ravel, Rut, and Shove	0.97	2343
(Hossain et al. 2019)	AADT, AADTT, AAMaH, AAMiH, AAP, annual average freezing index, and annual average temperature	-	-
(Joni et al. 2020a)	high and medium severity potholes, high severity ravelling and corrugation, medium severity alligator cracking, medium severity patching, and polished aggregate	0.78	395
(Kaya et al. 2020)	Age, previous IRI, longitudinal cracking, rut, traffic, and transverse cracking	0.87-0.99	360
(Kargah-Ostadi et al. 2010)	Age, FI, Milling depth, P_{200}, previous IRI, surface layer thickness, TO, and Time since last reading,	0.96	214
(Khattak et al. 2014)	Age, CTI, ESAL, FN, IRI_0, PI, and TO	0.47	623
(Lin et al. 2003)	Alligator Cracking, bleeding, corrugation, manholes, patches, potholes, Rutting, and stripping	0.94	125
(Mactutis et al. 2000)	fatigue %area, IRI_0, and Rut Depth	0.71	317
(Mazari and Rodriguez 2016)	Age, ESAL, IRI_0, and SN	0.99	2500
(ARA Inc. 2020) / (MEPDG)	Average rut depth, fatigue %area, SF, and transverse cracking length	0.56	1926
(Owolabi et al. 2012)	patches, severity level of rut, and severity of longitudinal crack	0.78	-
(Ziari et al. 2016a)	AADT, AADTT, Age, average temperature, ESAL, freezing index, pavement thickness, precipitation, and surface thickness	0.84	-
(Ziari et al. 2016b)	AADT, AADTT, Average precipitation, FI, pavement thickness, and Surface thickness	0.94	-

AC: Asphalt Content, AADT: Annual Average Daily Traffic, AADTT: Annual Average Daily Truck Traffic, AAMaH: Annual Average Maximum Humidity, AAMIH: Annual Average Minimum Humidity, AAP: Annual Average Precipitation, ESAL: Equivalent Single Axle Load, FI: Freezing Index, FN: Functional Classification, MSN: Modified SN, NoFT: Number of Freeze and Thaw cycles, P_{200}: % passing 200 sieve, PI: Precipitation Index, RES: Resurfacing type, SF: Site Factor, SN: Structural Number, and TO: Thickness Overlay.

In literature, the most frequently used datasets for developing models are those from the Long-Term Pavement Performance (LTPP) database maintained by the Federal Highway Administration (FHWA) (Mactutis et al. 2000, George 2000, Choi et al. 2004, Chou and Pellinen 2005a, Kargah-Ostadi et al. 2010, Khattak et al. 2014, Ziari et al. 2016a, Mazari and Rodriguez 2016, Ziari et al. 2016b, Gong et al. 2018b, Hossain et al. 2019, ARA Inc. 2020, Abdelaziz et al. 2020). The number of data points used, where reported, varied significantly between the models. The number of reported data points for models' development varied between 20 and 4109. Additionally, the number of variables used in the models ranged between 1 and 16. From the review of the *IRI* models, the variables used, the number of variables, and the number of data points varied significantly for model development. The review suggests that a variety of variables with sufficient data points should be considered for modelling *IRI*.

2.5 Regression Analysis

Regression analysis is one of the most common techniques to develop a relationship between a variable(s) and the variable of interest. There are various forms of regression analysis; some of the simplest regression analysis forms are linear regression and polynomial regression models. In regression analysis, the model development can be done through various variable selection methods to develop the best possible model. The developed models can be evaluated through statistical measures such as R^2 and the Root Mean Square Error (*RMSE*), among other statistical measures later discussed in this section.

2.5.1 Linear Regression Models

One of the most commonly used regression analysis models is the linear regression model. The simplest form of regression modelling is the simple linear regression model, which includes only two variables. The simple linear regression model can be mathematically expressed as follows:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.1)$$

Where β_0 is the model's intercept (also referred to as the model's constant), β_1 is the regression independent variable's coefficient (also known as the gradient), X represents the independent variable, Y

represents the dependent variable, and ε is the unknown error term (also known as the residual). The unknown error term is the difference between the observed and the predicted *Y* values.

When there are more than two variables to be used in linear regression modelling, multiple linear regression is used. The multiple linear regression model equation could be written as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (2.2)$$

The main difference between the simple and multiple linear regression models is the number of independent variables that are used to predict the dependent variable. Other lesser common regression analysis modelling techniques include polynomial and logistic regression models.

2.5.2 Variable Selection Methods

Variable selection methods play a crucial role in modelling linear regression models. Selecting the variables with the most significance to the target variable aids in creating more accurate models. The most frequently used selection methods for regression applications include forward selection, backward elimination (also known as the backward deletion method), and stepwise selection methods. The forward selection method develops the multiple linear regression models by adding independent variables to the equation one at a time. The independent variable with the strongest correlation to the dependent variable is added using p-value and t-statistics. The process repeats itself until no more independent variables with the required tolerance level can be added to the model (Sweet and Grace-Martin 2010, NCSS 2021).

The backward elimination method, or the backward deletion method, works similarly to the reverse of the forward selection method. The multiple linear regression model is first generated with all the independent variables, and the variables with the weakest correlation to the dependent variable are removed one at a time. The process repeats itself until all the remaining independent variables are highly correlated to the dependent variable (Sweet and Grace-Martin 2010, NCSS 2021).

Finally, the stepwise selection method combines the two selection methods mentioned above. The stepwise selection method is a modified version of the forward selection method. A correlation test is performed

with the addition of each independent to test whether the other independent variables' correlation with the dependent variable has been reduced below a tolerance level. At each step, if a variable falls below the tolerance level, they are removed from the multiple linear regression model (Sweet and Grace-Martin 2010, NCSS 2021).

2.6 Artificial Intelligence

Through the utilization of training data, AI has numerous applications related to classification and regression. The technology attempts to mimic advanced human skills and decision-making is used across various industries and disciplines. A famous application of AI involves the famous chess-playing program that has beaten some of the world's best chess players. However, applications of AI extend far beyond board game applications. Some of the most common applications include image recognition, video classification, speech-to-text, natural language processing, tabular and time-series data applications, and recommendation systems, as well as other predictive modelling functions (Mary 2020).

2.6.1 Machine Learning

As a subset of AI, ML focuses on learning from a given set of data. The main principle of ML is to train the computer to analyze data and identify patterns using statistical learning and optimization methods. The ML algorithms generally have three components: a decision process, an error function, and an optimization process (EL Naqa and Murphy 2015). Generally, ML can be divided into three categories: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, algorithms are trained to classify or predict outcomes based on labelled datasets. Unsupervised learning identifies patterns in unlabeled datasets and classifies the data accordingly. The reinforcement learning process is similar to supervised learning, with the exception of how the model is trained. The model is trained by trial and error in a reward/punishment feedback system in an attempt to offer the best solution for a given problem (IBM Cloud Education 2020).

2.6.2 Supervised Machine Learning Algorithms

Supervised ML is performed through a set of defined rules and methods in the form of algorithms. Many supervised ML algorithms exist; however, none of the algorithms are capable of providing suitable results for all applications. This poses the need to test various ML algorithms to identify the best algorithms for a given application. Commonly used supervised ML algorithms fall into the categories of decision trees, ensembles, nearest neighbours, regularized linear regression, Artificial Neural Networks (ANN), and Support Vector Machines (SVM). The supervised ML algorithms considered in this study are the ANN, CatBoost, decision trees, Elastic-Net, K-Nearest-Neighbor (KNN), Lasso, LightGBM, Random-Forest, Ridge, SVM, and XGBoost regression algorithms.

2.6.2.1 Decision Trees

Decision trees can be built in different configurations; the main characteristic of decision trees is utilizing a training dataset as input to recursively create a flow chart with binary filters at the intermediate or decision nodes. The model's top-down structure starts with the general input at the top and breaks down through binary filters until the process reaches a leaf predicting an output (de Ville 2013). The binary filters used in decision trees' flow charts are simply to guide the model to a leaf or terminal node to produce a prediction. An example of a binary filter could be as simple as filtering if a value is greater than or less than a specified value. Decision trees can be applied to regression applications where the algorithm infers binary filters based on the training data to create leaves endpoints where an output could be predicted. A simplified example of such a structure is presented in Figure 2.12 to demonstrate the flowchart-like structure of decision trees. The figure starts with a root node and gets split into decision nodes using a binary filter. At each decision node, a splitting occurs using a binary filter until a leaf or terminal node is reached and a prediction is generated. In the figure, a simple decision tree is presented to predict if an object is a tree. The Root node starts by deciding if the object has a root; if the answer is no, a terminal node is reached, and a prediction is generated; if the answer is yes, another decision node is

reached, and another question is raised. The process repeats itself until a terminal node is reached and a prediction is generated.

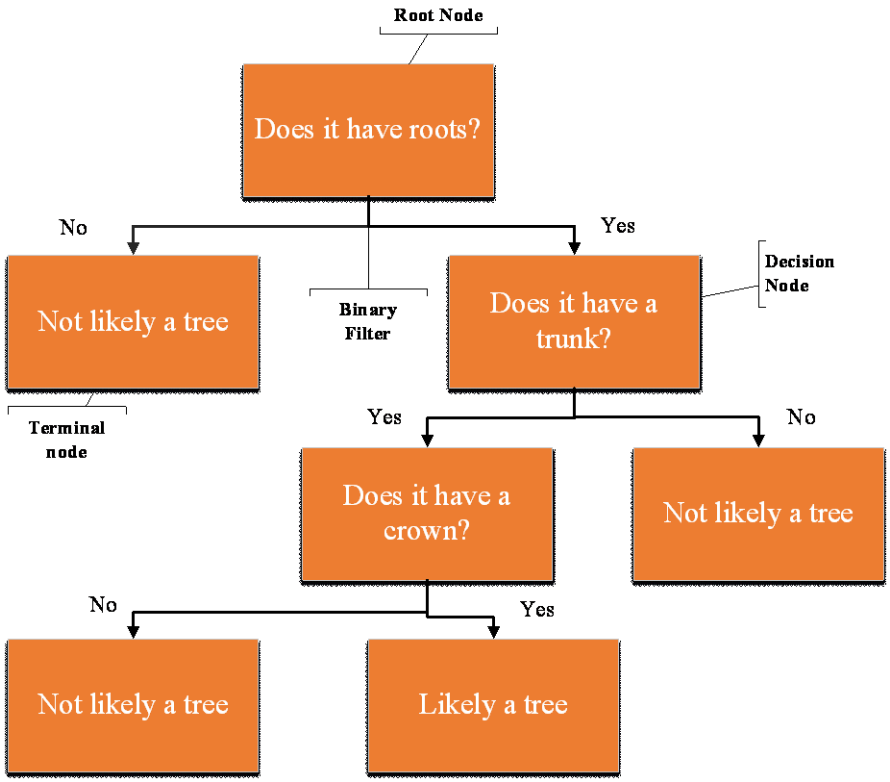

Figure 2.12 - Decision tree flowchart example

2.6.2.2 Random-Forest

Random-forest trees are an ensemble of decision trees trained using the bagging method. The bagging method creates a model that uses several independently built estimators to generate a prediction that is the average of the independently built estimators (Breiman 1996). Simply put, random-forest trees are several decision trees used to generate a prediction. The random-forest trees, in essence, are a collection of trees operating as a committee, often outperforming induvial decision trees. As an advantage of using the bagging method, random-forest trees are advantageous in preventing model overfitting (Breiman 2001). A visualization of a random-forest ML model is presented in Figure 2.13, where the output is the average of several independently built trees. The input data is fed into the random-forest model and is then fed into several decision trees simultaneously; each decision tree processes the input and provides a prediction. The predictions are then averaged, and the final prediction is generated.

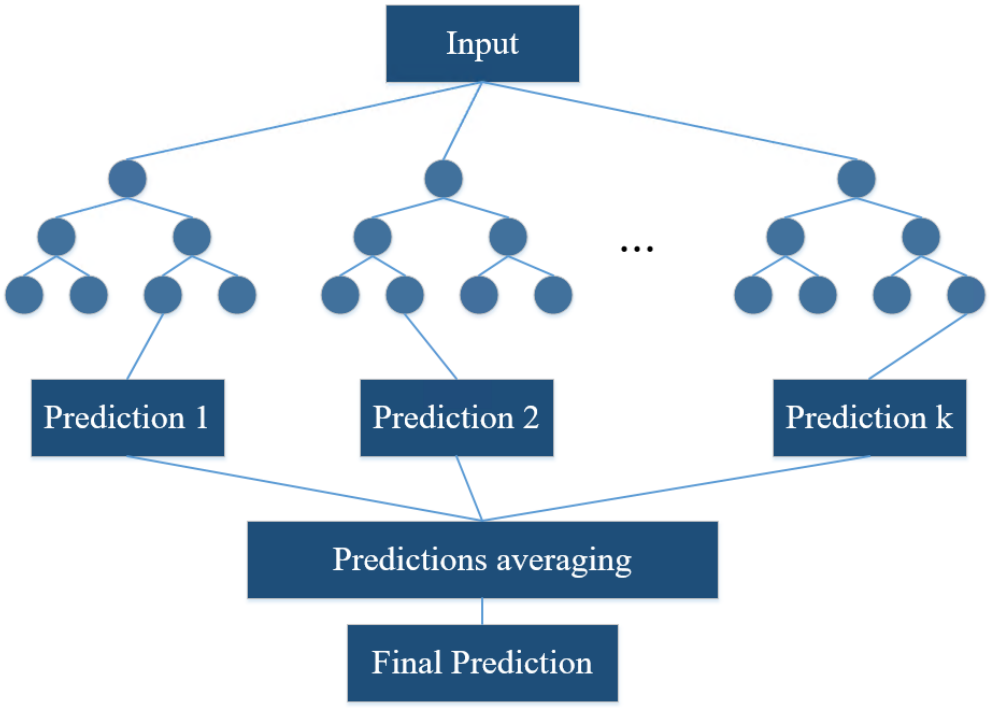

Figure 2.13 – Random-Forest ML Model representation of generating a prediction

2.6.2.3 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) have been commonly used in ML modelling for various applications. The ANN algorithm derives correlations between variables and is inspired by biological neurons. The general method of processing input in an ANN ML model starts with an input layer that feeds into layers of neurons (hidden layers) that each add a bias and applies a function. After the input has been changed in the hidden layers, an estimated output is presented in the output layer (Bishop 2006). There are several methods for developing ANN models. The most successful type of ANN modelling is the feed-forward neural network, also known as multilayer perception (Bishop 2006). A visual representation of an ML model built using the ANN algorithm is presented in Figure 2.14. The figure presents an input layer which contains neurons (highlighted in orange) that receive the input. The input is then fed into the hidden layers (highlighted in blue) to be transformed using biases and functions. The hidden layers then feed into the output layer (highlighted in yellow) where a prediction is presented.

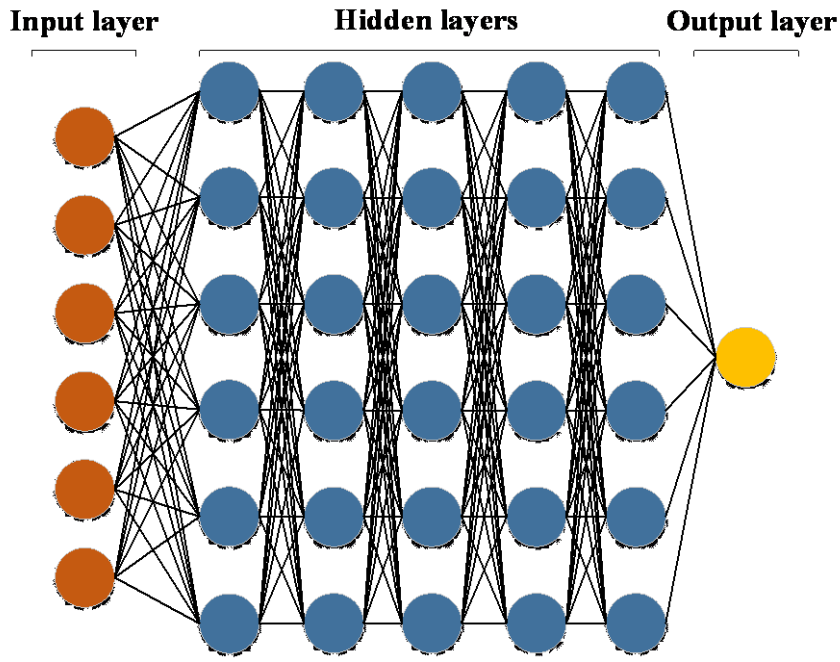

Figure 2.14 - Neural Networks ML Model layers representation

2.6.2.4 LightGBM Regressor

Light Gradient Boosting Machine (LightGBM) is another supervised ML algorithm that builds on the decision trees algorithm as ensembles. LightGBM uses two novel techniques in developing decision trees, the gradient-based one-side sampling and the exclusive feature building techniques, both part of the gradient boosting decision trees framework. The gradient-based one-side sampling technique is based on the idea that different data instances can have varied influences on a model's information gain. The technique keeps data instances that contribute more to the models' information gain while dropping data instances with small information gain. A larger gradient can represent a larger information gain by a data instance, hence the name gradient-based sampling. The exclusive feature bundling technique is used to create a model without nonzero values; in other words, it eliminates any features that have no significant contribution to the model's performance. A representation of model development using the LightGBM algorithm is presented in the figure below.

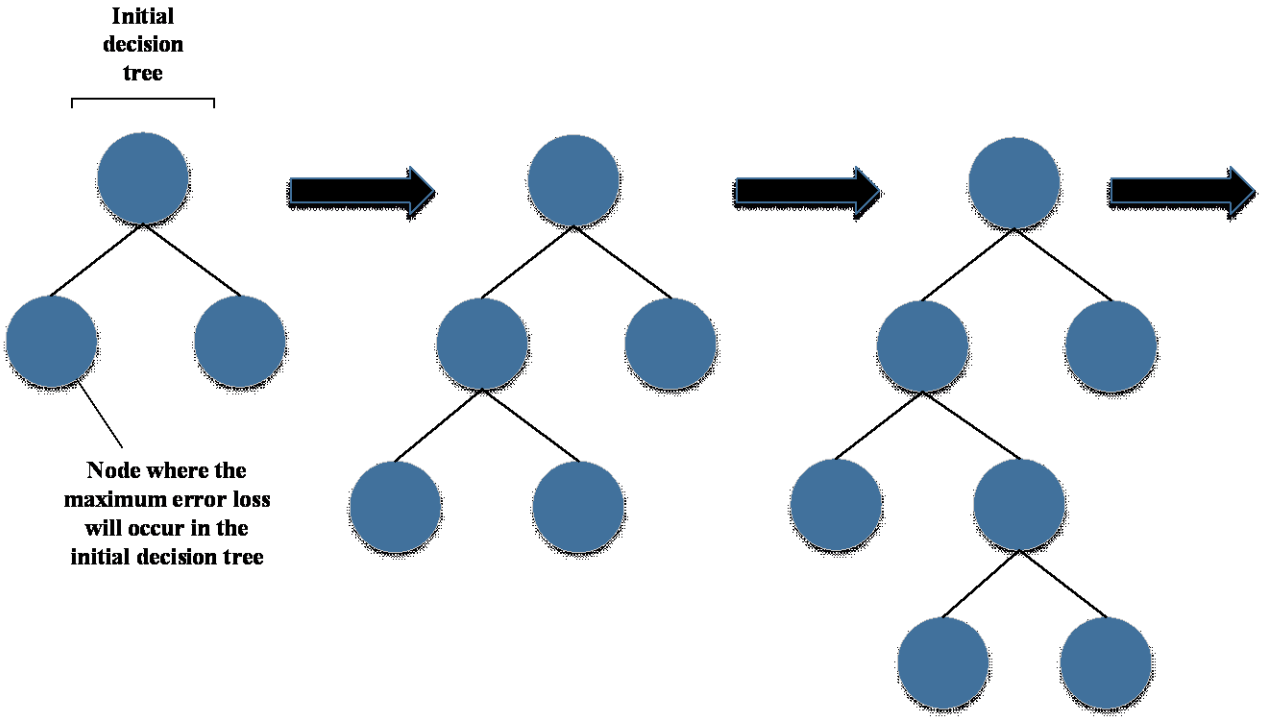

Figure 2.15 - LightGBM ML Model, Leaf-wise tree growth representation

The LightGBM algorithm combines gradient-based one-side sampling and feature bundling techniques to develop ensemble decision trees. The algorithm builds the trees vertically (leaf-wise) and keeps the leaves with the maximum error loss (Cover and Hart 1967, Ke et al. 2017, Microsoft 2021). The visual representation of LightGBM model generation is presented in Figure 2.15, highlighting the leaf-wise growth of the trees instructed by the LightGBM algorithm. The figure presents how the LightGBM algorithm develops decision trees. An initial decision tree is created, and the next leaf is added where the maximum error loss would occur. The process repeats itself until no more leaves can be added to improve the model's performance.

2.6.2.5 Support Vector Machine

The Support Vector Regression (SVR), also known as SVM, is an ML algorithm that utilizes Lagrange multipliers in model development (Bishop 2006). The SVM algorithm develops an ML model through the concept of a margin (the smallest distance between a decision boundary and any of the samples) and supports (Bishop 2006). The model's margin is identified using the Lagrange multiplier's

function and is used to create the decision boundary. The decision boundary is also referred to as hyperplanes; data points that fall within the hyperplanes are called support vectors. Through the use of hyperplanes and, inherently, the Lagrange multipliers, the SVM algorithm is useful in identifying outliers in a given dataset (Bishop 2006). The algorithm's use of margins and supports reduces the ability of data outliers or data noise to interfere with the model's performance, thus, reducing the model's generalization error (Bishop 2006). A representation of the hyperplanes and support vectors concept in SVM is presented in Figure 2.16. The red hyperplane in the figure represents the optimal hyperplane and the blue hyperplanes represents the decision boundary. Points marked in yellow represent the support vectors that fall within the decision boundary and are used in generating the model's prediction.

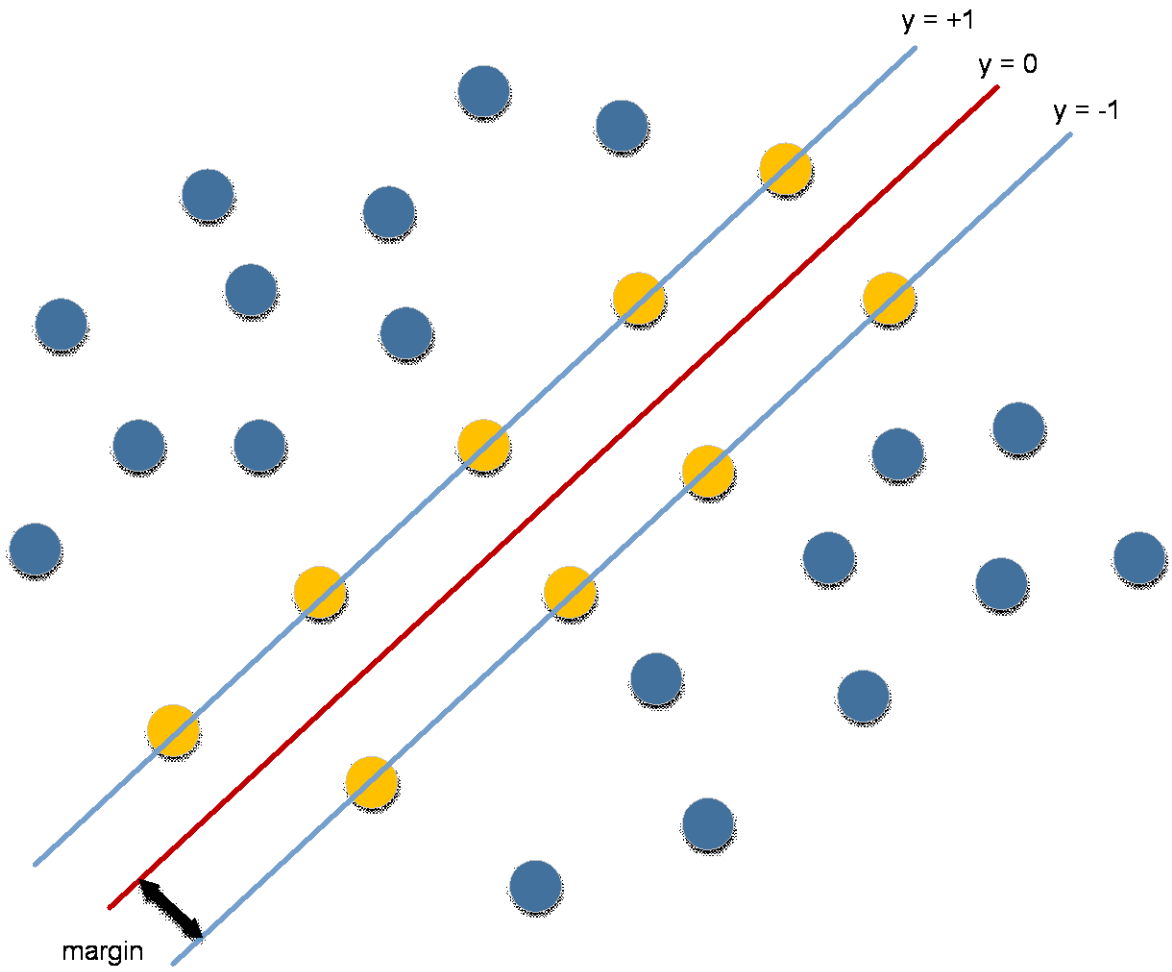

Figure 2.16 - SVM support vectors within a defined margin representation

2.6.2.6 KNN (K-Nearest-Neighbour)

The KNN utilizes the Euclidean distances to classify neighbouring data according to a predefined parameter “k” value, hence, the K in K-Nearest-Neighbour (Cover and Hart 1967). The Euclidean distances between data points are calculated, and the nearest “k” number of data points is used in providing the model’s prediction. The algorithm is generally a classification ML algorithm; however, the algorithm has regression applications that can be used similarly to the SVM algorithm. A representation of the “k” within KNN is presented in Figure 2.17. In the figure, the first circle uses a “k” value of four; thus, the closest four neighbours are used to make a prediction. In the first neighbourhood with a “k” value of four, three neighbours are highlighted in blue and one in orange; thus, the data point can be predicted to be similar to the blue neighbours. In the second circle, the “k” value is fifteen, consisting of seven neighbours highlighted in orange and eight highlighted in blue; thus, the data point can be predicted to be similar to the blue neighbours.

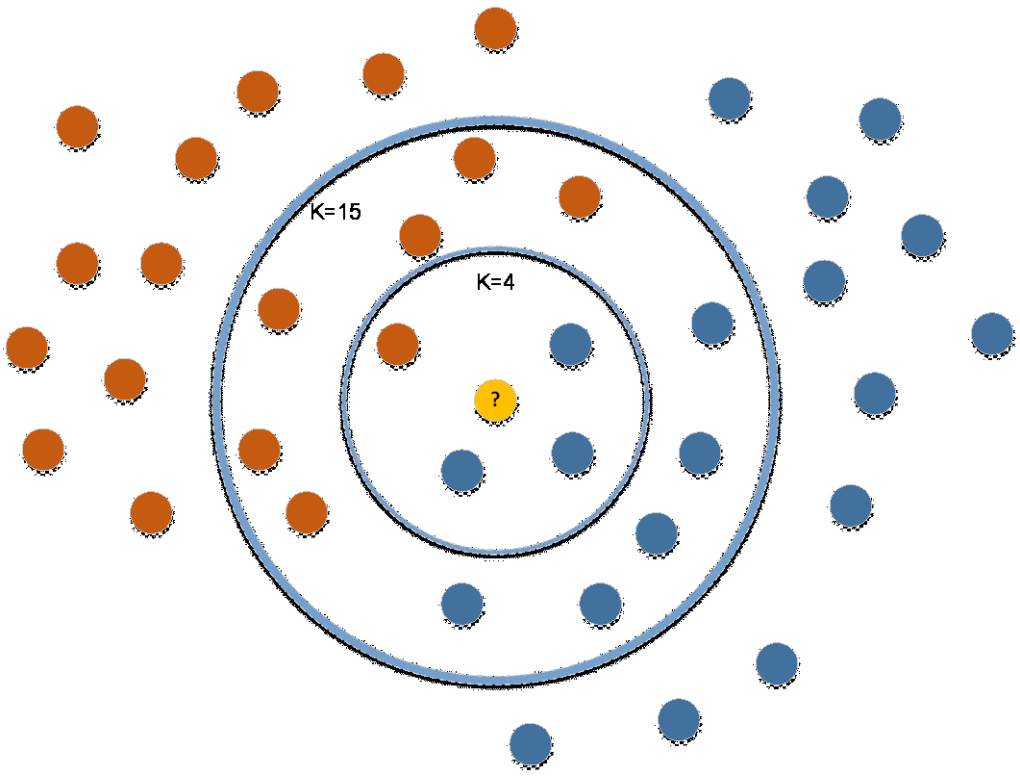

Figure 2.17 - K-Nearest-Neighbor algorithm neighbourhood representation

2.6.2.7 XGBoost Regressor

XGBoost is an ML algorithm based on gradient boosting machine decision trees framework (constructs ensemble decision trees). The gradient-boosting machine decision framework was previously discussed for the LightGBM algorithm. In the LightGBM algorithm, decision trees are built vertically or leaf-wise; in contrast, the XGBoost algorithm constructs trees horizontally or level-wise. The algorithm builds the model parallelly to increase the ML model's speed and performance. Additionally, XGBoost takes advantage of simultaneous processing, tree pruning, and regularizations to avoid overfitting the model while following the gradient-boosting machine decision framework (Chen and Guestrin 2016). A representation of the XGBoost algorithm building a decision tree is presented in Figure 2.18. The figure presents the level-wise decision tree construction where instead of each leaf being constructed where the maximum error loss would occur, a level is constructed at a time containing multiple leaves where the maximum error loss would occur.

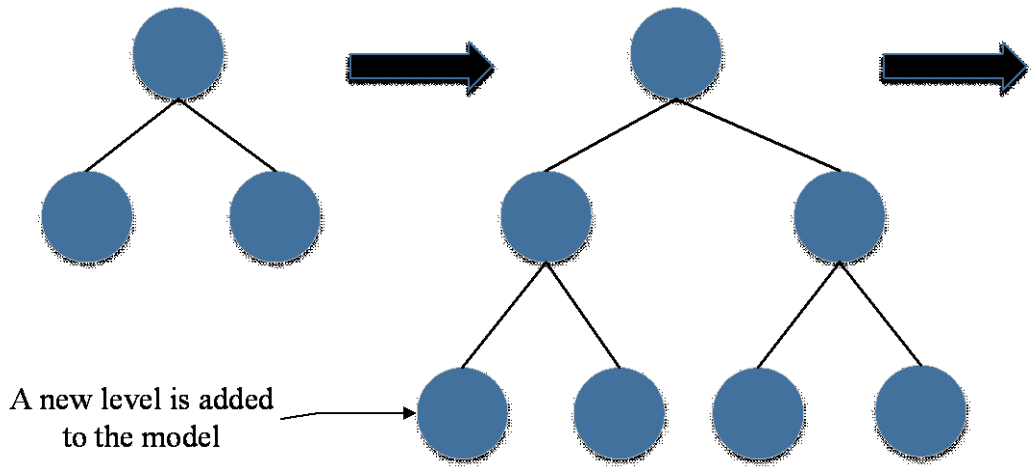

Figure 2.18 - XGBoost ML Model, Level-wise tree growth representation

2.6.2.8 CatBoost Regressor

CatBoost is an ML algorithm that utilizes the gradient boosting machine decision trees framework to develop ensemble decision trees. The algorithm creates decision trees sequentially, with each successive tree built with a smaller error than the prior tree (Dorogush et al. 2018). CatBoost

model's trees are constructed so that tree-level growth provides the lowest possible loss function. The process by which the CatBoost algorithm builds the decision trees is similar to that of the XGBoost. However, CatBoost imposes the rule that all nodes at the same leaf level would be tested with the same conditions, and the nodes are then given an index or a weight. In other words, the algorithm utilizes the weighted sampling version of the Stochastic Gradient Boosting to maximize the model's accuracy (Dorogush et al. 2018). A representation of CatBoost regression tree growth is presented in Figure 2.19. The figure shows an ensemble of decision trees developed using the CatBoost algorithm and presents balanced decision trees containing the same number of nodes built leaf-wise using the gradient boosting machine decision trees framework.

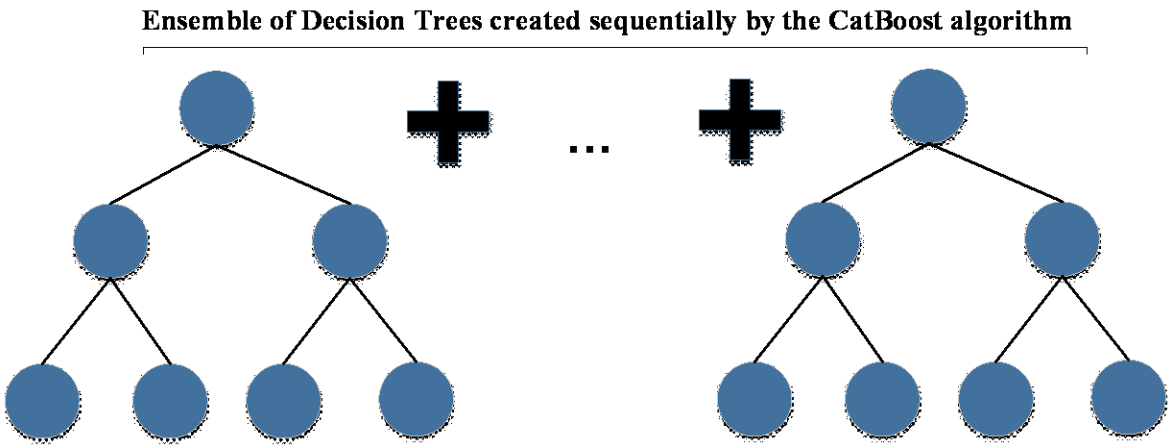

Figure 2.19 - CatBoost algorithm ensemble decision trees representation

2.6.2.9 Elastic Net, Ridge, and Lasso Regression

Elastic-Net, Ridge, and Lasso regression ML algorithms utilize "L1" and/or "L2" norm regularization in the model's development. The algorithms use the concept of shrinkage of data using a norm regularization method to assign coefficients to the variables in the model. Variables within the model with low contributions to the model are given a penalty, and their effects on the model's predictions are muted. The "L1" norm regularization method adds a penalty equal to the absolute value of the magnitude of the coefficient, resulting in some cases in a zero-coefficient variable and eliminating the

variable. The “L2” norm regularization adds a penalty to variables but does not eliminate variables. The Lasso regression ML algorithm utilizes “L1” norm regularization to shrink data points toward a mean point. The Lasso algorithm reduces data noise caused by adjusting for multicollinearity. Ridge regression algorithms utilize the “L2” norm regularization to shrink data closer to the population mean, improving the model’s least-square error. The Elastic-Net regression ML algorithm is a combination of both the Lasso and Ridge algorithms in that the Elastic-Net algorithm utilizes both “L1” and “L2” norm regularization to aid in fully utilizing the training data while minimizing nonzero data weights (Kim et al. 2007, Friedman et al. 2010). The Elastic-Net algorithm penalizes variables with little contribution to the model’s output but does not eliminate them all. The combination of “L1” and “L2” norm regularization helps keep relevant information in the model. A comparative representation of all three algorithms is presented in Figure 2.20. The figure shows the Lasso regression’s “L1” norm regularization data shrinkage range compared to the Ridge regression’s “L2” norm regularization. The shrinkage caused by the “L2” norm regularization is less severe than that of the “L1” norm regularization. The figure also shows Elastic-Net's use of both “L1” and “L2” to form a compromise in the data shrinkage.

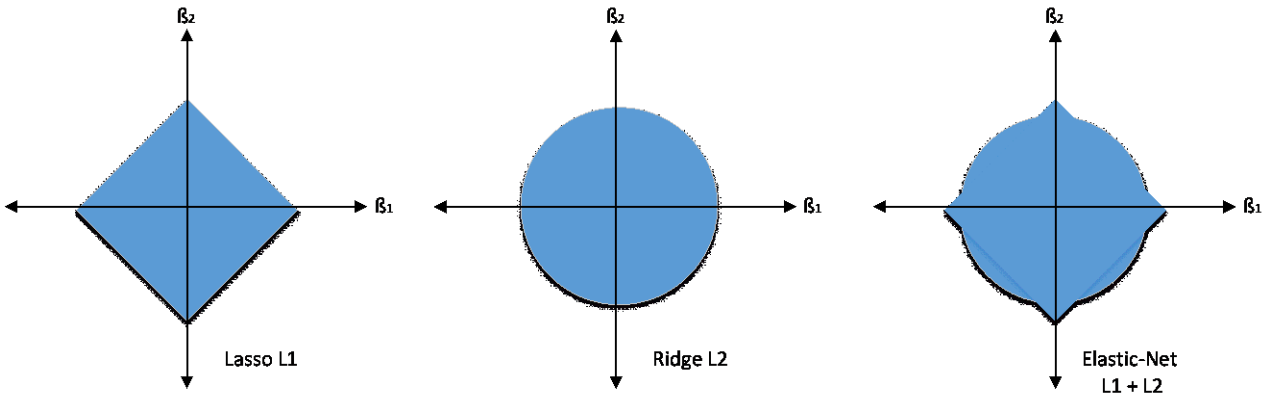

Figure 2.20 - Lasso, Ridge, Elastic-Net norm regularization representation

2.7 Model Evaluation

It is critical to evaluate the performance of the developed models and the variables used within the developed models. There are various statistical measures to evaluate the developed models; many of

them focus on the variance around a population's mean. Some of the most common measures that are used in model development include Mean Absolute Error *(MAE*), Mean Square Error (*MSE*), Root Mean Square Error (*RMSE*), Coefficient of determination (R^2), F-test, t-test, and p-values (Sarstedt and Mooi 2011, Dubitzky et al. 2013).

2.7.1 Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE)

The *MAE* or the Mean Absolute Deviation (*MAD*) is a statistical measure that measures how a predicted value differs from a measured value. The *MAE* is calculated by adding up the absolute differences between the predicted and observed values. The sum is then divided by the total number of observed data points to provide an indication of how close the average model's prediction is to the actual value (Willmott and Matsuura 2005). The *MAE* is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{2.3}$$

Where n is the total number of data points, Y_i is the observed value, and $\hat{Y}_i$ is the predicted value.

The *MSE* or the Mean Square Deviation (*MSD*) is another statistical measure that quantifies the deviation between the observed and predicted value (Hossein 2014, Olive 2014). The *MSE* is calculated in a similar way to that of the *MAE* as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{2.4}$$

The mathematical difference between *MAE* and *MSE* is that the former takes the absolute value of each difference between the observed and predicted values, while the latter takes the square value of the difference between the observed and predicted values. The *MSE* is more suitable than the *MAE* when needing to highlight the difference between the observed and predicted values greater than one while muting the effect of differences less than one. This is a result of the squaring, which effectively weights large errors more heavily than small ones.

The *RMSE*, or the Root Mean Square Deviation (*RMSD*), is another frequently used statistical measure that quantifies the difference between the observed and predicted values (Hossein 2014). Simply put, the *RMSE* is the square-rooted *MSE* and is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \quad (2.5)$$

The difference between the *RMSE* and the *MSE* is that *MSE* is measured in units that are squares of the model's independent variable. In contrast, the *RMSE* is measured in the same unit as the model's independent variable. Both the *RMSE* and *MSE* penalize larger differences between the observed and predicted values; however, the *RMSE* provides a measure of the model's error in the same units as the target variable that can provide a better insight into the model's error.

2.7.2 Coefficient of Determination (R^2)

The R^2 is a statistical measure that is widely used to describe the degree to which the model explains the observed variation relative to the mean. The coefficient is used as a benchmark to evaluate the performance of models and varies from 0 to 1. An R^2 of 0 indicates that the model completely fails to explain the observed variation, while an R^2 of 1 indicates that the model can perfectly explain the observed variation from the mean. The coefficient is calculated using two parameters, the explained variation of predicted values from the mean and the total observed variation (Draper 1998, Sarstedt and Mooi 2011). The explained variation of predicted values and total observed variation are calculated using the following equation:

$$Explained\ variation = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y}_i)^2 \quad (2.6)$$

$$Total\ variation = \sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2 \quad (2.7)$$

Using equations 2.6 and 2.7, the R^2 could be calculated by taking the explained variation and dividing it by the total observed variation as follows:

$$R^2 = \frac{Explained\ Variation}{Total\ Variation} \quad (2.8)$$

The coefficient R^2 can have limitations when assessing the performance of a regression model. Adding a slightly correlated independent variable to the regression model could cause the R^2 value to increase without significantly improving the model's predictive capability. The adjusted R^2 coefficient can be used to decide which combination of independent variables produces the best possible model (Draper 1998, Devore 2016). The adjusted R^2 represents the degree to which the model can explain the observed variation while accounting for the number of independent variables in the model. The adjusted R^2 is calculated as follows:

$$Adjusted\ R^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k-1} \quad (2.9)$$

Where n is the total number of observations and k is the number of independent variables in the model. The adjusted R^2 provides a quantitative method to choose between alternative regression models while considering the number of independent variables used (Sarstedt and Mooi 2011, Devore 2016).

2.7.3 Null Hypothesis, F-test, T-test, and P-values

Model development generally attempts to prove that the independent variables have an effect on a dependent variable. There are two general hypotheses formulated for the model development: a null hypothesis and an alternative hypothesis. A null hypothesis, denoted by H_0, states that there is no statistical significance or relationship between independent and dependent variables. The alternative hypothesis, denoted by H_1, states that there is a meaningful relationship between the dependent and independent variables, which can be explained by the developed model. Various statistical tests help to decide whether to reject the null hypothesis or the alternative hypothesis. For an adequate regression, the desired outcome is to reject the null hypothesis and prove the alternative hypothesis. The most commonly used statistical measures include the F-test, t-test and p-values (Sarstedt and Mooi 2011, Haldar 2013, Goodman 2017).

The F-test or F-statistic is a statistical measure that uses F-distributions to compare the model's variance and is computed as the ratio of the mean square between to the mean square within the distribution. The mean square between-group value is computed using the between-group variation (Schumacker and Tomek 2013). The between-group variation is calculated as follows:

$$SS_B = \sum_{j=1}^{k} n_j(\bar{X}_j - \bar{X})^2 \quad (2.10)$$

Where SS_B represents the between-group sum of squared deviations, k represents the overall number of groups, n_j represents the number of observations in that group, $\bar{X}$ represents the population's mean, and the $\bar{X}_j$ is the mean in the jth group (Sarstedt and Mooi 2011). Similarly, the mean square within is computed using the within-group variation and is calculated as follows:

$$SS_W = \sum_{j=1}^{k} \sum_{i=1}^{nj} (X_{ij} - \bar{X}_j)^2 \quad (2.11)$$

Where SS_W represents the within-group sum of squared deviations and X_{ij} represents the observation's value in the ith and jth groups. The between-group variation and the within-group variation have to be normalized to be used in the F-test. The normalization is done by dividing the variations by their degrees of freedom to obtain their mean squares (Sarstedt and Mooi 2011, Schumacker and Tomek 2013). The mean square between and mean square within are calculated as follows, respectively:

$$MS_B = \frac{SS_B}{k-1} \quad (2.12)$$

$$MS_W = \frac{SS_W}{n-k} \quad (2.13)$$

The normalized between and within group variations by their degrees of freedom, resulting in the mean squares above, is used to calculate the F-test value (Sarstedt and Mooi 2011, Schumacker and Tomek 2013). The F-statistic is calculated as follows:

$$F = \frac{MS_B}{MS_W} \quad (2.14)$$

The t-test is another commonly used statistical measure to aid with whether to accept or reject the null hypothesis. The calculation of the t-test value is easier to calculate than the F-test (Sarstedt and Mooi 2011, 2011). The t-test can be simply computed as follows:

$$t = \frac{\bar{x}-\mu}{s_{\bar{x}}} \tag{2.15}$$

Where $\bar{x}$ represents the sample mean, μ is the population mean, and s_x is the standard error. The standard error is the standard deviation, *s*, divided by the square root of the total number of observations (Sarstedt and Mooi 2011, Schumacker and Tomek 2013). The standard error can be calculated as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}}{\sqrt{n}} \tag{2.16}$$

Where n represents the total number of observations and x_i is the ith observation's value. It is important to recognize that the measures mentioned above are, in one form or another, related to the model's variance. The tests mentioned above aid with the decision on whether to reject the null or the alternative hypothesis. The decision to reject the null hypothesis is constricted by the model's p-value and α-value. The p-value is the probability of the null hypothesis being correct; in other words, the probability that the alternative hypothesis is true. The α-value is a threshold set to determine whether the p-value is small enough to reject the null hypothesis. The threshold is often set at 0.05 for most regression modelling applications. The model's p-value needs to be equal to or below the defined threshold to reject the null hypothesis. The 0.05 threshold means that there is less than a 5% probability that the null hypothesis is correct. Similarly, the confidence level in rejecting the null hypothesis is denoted by subtracting the α-value from 1; thus, the confidence level for an α-value of 5% is 95% (Sarstedt and Mooi 2011, Schumacker and Tomek 2013). The relationship between the t-value and p-value is presented in Figure 2.21. In the figure, the α-value is set at 5%, and the respective t-test value is denoted as t-critical. The t-critical value splits the curve into two parts; the area on the left of the t-critical value is called the acceptance region, where we accept the alternative hypothesis and reject the null hypothesis. The t-test value denotes the probability of error in rejecting a true

null hypothesis. In this case, the t-statistic is larger than the t-critical value, and thus the null hypothesis can be rejected with a confidence level of 95% (Sarstedt and Mooi 2011).

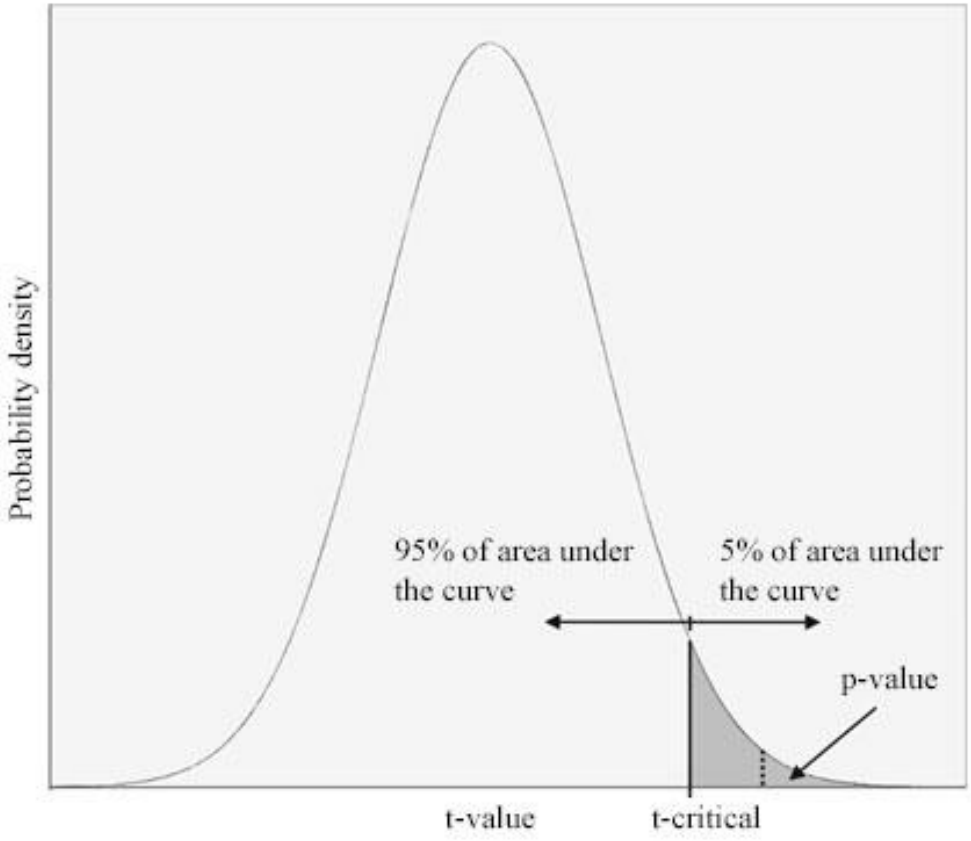

Figure 2.21 - The relationship between t-value and p-value (modified from Sarstedt and Mooi 2011)

Chapter 3 : Development of International Roughness Index (*IRI*) Models for cold regions using PMS data from the province of Alberta

Abstract

International Roughness Index (*IRI*) is an indicator of pavement performance measured by its longitudinal profile. Over the years, numerous empirical *IRI* prediction models have been developed based on measured data to estimate the effects of climate, traffic loading, pavement structure, surface distresses, and other factors contributing to road roughness. In practice, local calibration of *IRI* models remains a challenge for road agencies. Currently, the *IRI* model used by Alberta Transportation (AT) is only a calibration of a general model. The development and validation of more accurate and precise regression models are necessary. In this study, representative control sections are selected from a 7-year Pavement Management System (PMS) database provided by AT; the new models are developed to estimate *IRI*'s change as a function of multiple key variables. The site-specific *IRI* models developed in this chapter resulted in models superior to that of the Mechanistic-Empirical Design Guide (MEPDG). The results of this study indicate that the proposed models help analyze pavement deterioration trends in addition to planning maintenance and rehabilitation (M&R) activities at the network level. A Life-Cycle Analysis (LCA) and a Life-Cycle Cost Analysis (LCCA) comparing the MEPDG model with the one developed in this study confirm that the latter will result in relatively lower costs. Through this study, AT and other agencies can make better predictions about pavement deterioration.

3.1 Introduction

Throughout history, roads connected people to resources and one another. Pavement planning and maintenance have always been a challenge for various transportation agencies. Maintaining infrastructure continues to be a significant concern for governments around the globe. The larger the road network, the greater the challenge is. The province of Alberta in Canada has more than 31,400 kilometres of road, or about 64,000 lane kilometres (Alberta Transportation 2020). Alberta Transportation (AT) utilizes a

sophisticated Pavement Management System (PMS) for various pavement-related decisions. The system includes various tools to aid with decision-making for optimizing the serviceability conditions of the pavement network. One of the most critical aspects of the PMS is its pavement performance model, most commonly the International Roughness Index (*IRI*).

The *IRI* is an important indicator for assessing pavements' condition. Over time, pavement deterioration can be represented by a rise in *IRI* with respect to time. Changes in *IRI* values indicate changes in pavement roughness resulting from traffic, structural integrity, distress, age, and climate. A theoretical *IRI* of 0.0 denotes a perfectly smooth road surface and proportionally increases with road roughness (Sayers 1995b). Different regions exhibit unique traffic behaviours accompanied by their unique distress and climate patterns. Pavement predictive models are generally developed based on specific regions' datasets; therefore, developing a universal and accurate pavement predictive model is a challenge on its own. The *IRI* models require local calibration by transportation agencies throughout the world. Alberta's PMS database, maintained by AT, can be utilized to create an empirical model that predicts *IRI* values over a specified period of time.

This study aims to develop *IRI* models that are more representative of the province of Alberta than the currently utilized models. Through developing more accurate *IRI* models, road networks can be designed and planned more effectively, resulting in safer, more comfortable roads and, if used appropriately, lower expenditures. Furthermore, this study examines the impact of each independent variable on the developed *IRI* models' output. The most impactful variables on *IRI* are then used to develop reduced *IRI* models. In addition, this study uses the Mechanistic-Empirical Design Guide (MEPDG) *IRI* model as a benchmark to compare the developed model through a case study using the life cycle and cost analysis over a 50-year period, showing the environmental and economic impacts, respectively.

3.2 Background

Pavement roughness affects ride quality, travel times, and road safety (Robbins and Tran 2016). Roughness directly impacts fuel consumption, repair costs, greenhouse gas emissions, vehicle maintenance, and vehicle efficiency. In light of the growing attention to climate change, pavement roughness clearly plays a critical role. Nevertheless, recent efforts to measure pavement roughness, including the *IRI*, have been mainly conducted without much consideration of climate change.

Among the first experiments to measure pavement roughness were those conducted using high-speed profilers on Quarter-Car simulations in the late 1960s (Howe et al. 2003). The high-speed profilers could measure the actual profile over different wavelengths to reflect the vehicle's vibrations. Research on the Quarter-Car simulation produced a model that replicated the Bureau of Public Roads (BPR) Roughometer (Dillard and Allen 1959). Soon thereafter, a commercial version of the system was released, including the Quarter-Car for analyzing road roughness. The World Bank developed the currently used *IRI* system in 1986 (Sayers et al. 1986). The World Bank's *IRI* results from a correlation experiment conducted in Brazil called the International Road Roughness Experiment (IRRE), which indicated that a standardized index could and should be developed. This led to the *IRI* development being added as an objective of the research program, and the *IRI* was developed using the Quarter-Car operating at a standard speed of 80 km/hr.

A number of attempts have been made since the *IRI* system was standardized to improve the predictive capabilities of *IRI* models by making use of various datasets. From the literature, Table 3.1 shows some of the most frequently cited *IRI* predictive models developed using regression analysis based on the Long-Term Pavement Performance (LTPP) program (George 2000, Choi et al. 2004, Khattak et al. 2014, ARA Inc. 2020) or local agencies databases worldwide (Mactutis et al. 2000, Lin et al. 2003, Al-Suleiman (Obaidat) and Shiyab 2003, Albuquerque and Núñez 2011, Owolabi et al. 2012, Joni et al. 2020a). The model summaries provided in Table 3.1 also include the coefficient of determination (R^2) and the number of data points used in developing the model (N), where available. The most common variables used in the

IRI models above include age, traffic, and the at-construction *IRI*. Only a few models used soil parameters in predicting *IRI*. The soil parameters were the *PI* and P_{200} of the pavement's subgrade (Choi et al. 2004, Khattak et al. 2014). Some models cited in Table 3.1 had a structural number or factor as a predictive variable and might encapsulate other variables (George 2000, Albuquerque and Núñez 2011, Khattak et al. 2014, Choi and Do 2019, ARA Inc. 2020). Models are mainly applicable to the range of data used for their development and are often most accurate at a regional level. There is a wide variation in R^2 for the above-mentioned models, ranging from 0.35 (George 2000) to 0.94 (Lin et al. 2003, Albuquerque and Núñez 2011).

Table 3.1 – Summary of various *IRI* models reported in the literature

Model	**Model Variables**	***R^2***	***N***
(Albuquerque and Núñez 2011)	ESAL, mean precipitation, potential evapotranspiration, and SN	0.87-0.94	20
(Al-Suleiman (Obaidat) and Shiyab 2003)	Age	0.61-0.80	440
(Choi et al. 2004)	AC, ESAL, P_{200}, SN, and TO	0.71	117
(George 2000)	Age, ESAL, and MSN	0.35	690
(Joni et al. 2020b)	high and medium severity potholes, high severity ravelling and corrugation, medium severity alligator cracking, medium severity patching, and polished aggregate	0.78	395
(Khattak et al. 2014)	Age, CTI, ESAL, FN, IRI_0, PI, and TO	0.47	632
(Lin et al. 2003)	Alligator cracking, bleeding, corrugation, manholes, patches, potholes, rutting, and stripping	0.94	125
(Mactutis et al. 2000)	fatigue %area, IRI_0, and Rut Depth	0.71	317
(ARA Inc. 2020) / MEPDG	Average rut depth, fatigue %area, SF, and transverse cracking length	0.56	1926
(Owolabi et al. 2012)	patches, severity level of rut, and severity of longitudinal crack	0.78	-

AC: Asphalt Content, CTI: Cumulative Temperature Index, ESAL: Equivalent Single Axle Load, MSN: MSN (Modified SN), TO: Thickness Overlay, FN: Functional Classification, SF: Site Factor, SN: Structural Number, P_{200}: % passing 0.075 sieve.

The variables used in modelling *IRI* in Table 3.1 ranged significantly from one another. The *IRI* model developed by Al-Suleiman (Obaidat) and Shiyab (2003) only had *Age* as the predictive variable. In contrast, Lin et al. (2003) developed a model that used alligator cracking, bleeding, corrugation,

manholes, patches, potholes, rutting, and stripping as predictive variables. The most common variables in modelling *IRI* in the cited models are *Age*, *ESAL*, and rutting. The number of data points used in developing the *IRI* models also varied significantly. The model developed by Albuquerque and Núñez (2011) used 20 data points in model development, while the MEPDG *IRI* model was developed using 1926 data points. None of the most cited *IRI* models presented in Table 3.1 had any emphasis on cold region pavements. The majority of the *IRI* models focused on warm region pavements. The reviewed *IRI* models above show a clear gap in modelling *IRI* for cold region pavements.

3.3 Regression Analysis Framework

The regression analysis *IRI* models developed in this study are developed using a methodological framework. The regression analysis framework used in developing the *IRI* models is presented in Figure 3.1.

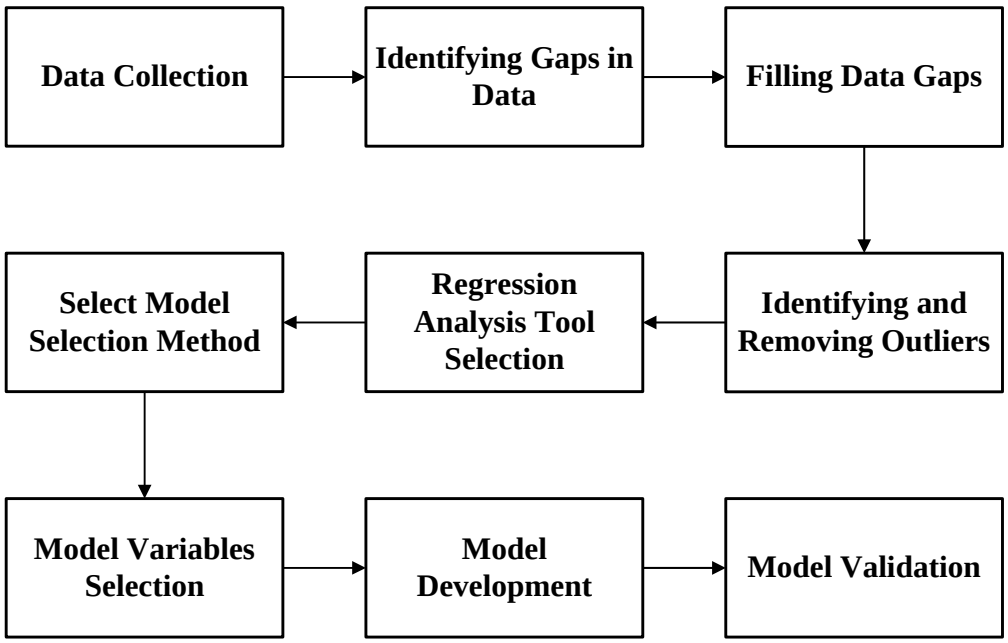

Figure 3.1 - Regression analysis framework used in developing the models in this chapter

The first step in creating the models is to collect the data to be used in the model development. This is followed by removing any meaningless outliers, selecting the appropriate tool for the

model development, selecting the variables selection method, creating the model, followed by testing and validating the model. The regression analysis framework depicted in Figure 3.1 starts with data collection, in this case, data extraction from the PMS database. Once the data has been collected, the limitations and inconsistencies in the extracted data are identified and resolved. One limitation could be that the extracted data does not include variables of interest. An example of inconsistencies would be the illogical progression of *IRI* values. Inconsistencies of the *IRI* values could be caused by equipment or record errors, among other reasons. This is followed by identifying and removing data outliers and then identifying the tools to be used in the regression analysis model development. The tools include software to identify variables' significance and train the model. The model variable selection method is then selected for training the model, and the variables with the highest statistical significance are then selected. The model is then trained using the selected variables, followed by model validation.

3.4 Alberta Transportation PMS Database

Transport agencies utilize PMS to perform pavement-related tasks, such as pavement design, maintenance, and rehabilitation. The accuracy of pavement performance predictive models is crucial for the PMS's efficiency (Hudson et al. 1979). Alberta maintains an extensive PMS database which is utilized in this study. The dataset provided by AT includes comprehensive pavement history, traffic, and distress variables for the period 2014 to 2020. The variables in the provided dataset included but were not limited to pavement section location, pavement type, length of the section, base soil type, base construction year, base thickness, surface thickness, last activity, last activity year, Annual Average Daily Traffic (*AADT*), Equivalent Single Axle Load (*ESAL*), *IRI*, transverse cracking area, longitudinal wheel path cracking length, other cracking areas, and rutting depth.

The most common type of pavement category in the road network is Granular Base Course (GBC), particularly Asphalt Concrete Pavement (ACP) and asphalt concrete Overlay pavement (OL). This study

focuses on both pavement types, the OL and ACP pavement types. The pavements classified as ACP in the network were identified as staged and non-staged pavements. For simplicity, no distinction was made between staged and non-staged pavements for the development of the *IRI* models.

3.5 Database Limitations

On a detailed examination of the AT dataset, some inconsistencies and limitations were observed. Inconsistencies in the dataset included recorded pavement thickness of zero and *IRI* values that were lower than the preceding year. All sections that exhibited such inconsistencies were removed from the dataset. Similarly, a handful of sections had "unknown" base soil types that required identification. The "unknown" base soil types were assigned using Alberta Agriculture and Forestry agency's tool for viewing soil types in Alberta (Alberta Agriculture and Forestry 2020). The soil types were listed according to the Unified Soil Classification System (USCS) except for "CI" which is described as "Clays of medium plasticity, gravelly clays, sandy clays, silty clays."

Limitations in the dataset included the omission of soil properties. Thus, the Plasticity Index (*PI*) and Percent passing 200 (P_{200}) were chosen as model variables to capture the soil properties pertaining to the freeze and thaw effect, as is the field's practice (ARA Inc. 2020). The values assigned to *PI* and P_{200} are the means of the ranges specified in the MEPDG of new and rehabilitated pavement structures by the National Cooperative Highway Research Program (NCHRP) (ARA Inc. 2001b).

Another limitation of the dataset was that the climate data was not included. Current and historical weather stations were used to compile climate data from the nearest weather station (Alberta Agriculture and Forestry 2021a). The compiled dataset included: precipitation, accumulated precipitation, mean temperature, minimum temperature, maximum temperature, relative humidity, and wind speed. For model development, several climates and temperature indices were derived from the variables above.

The presence of outliers is inevitable in any large dataset. Analysis of the data revealed several extreme values for some of the variables. Outliers were identified and removed in the dataset using the outer and inner fence approaches (NCSS 2021). The equations for both approaches are as follows:

$$Lower\ Inner\ Fence\ =\ Q_1\ -\ (1.5)\ IQR \tag{3.1}$$

$$Upper\ Inner\ Fence\ =\ Q_3\ +\ (1.5)\ IQR \tag{3.2}$$

$$Lower\ Outer\ fence\ =\ Q_1\ -\ (3)\ IQR \tag{3.3}$$

$$Upper\ Outer\ fence\ =\ Q_3\ +\ (3)\ IQR \tag{3.4}$$

Q_1 represents the 25th percentile, Q_3 represents the 75th percentile, and *IQR* represents the interquartile range (the difference between Q_3 and Q_1). The percentage of the data points in the inner and outer fence datasets was 37% and 85% for OL pavements, while for ACP pavements, 42% and 85%, respectively.

3.6 Development of Regression Model

Regression analysis is a mathematical approach to developing a relationship between a dependent and independent variable(s). This mathematical analysis relies heavily on available data and the number of variables utilized. Multiple linear regression is a technique that is used when a dependent variable is a function of various other variables. A multiple linear regression model can be expressed in the form as follows:

$$y =\ \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots +\ \beta_k x_k +\ \varepsilon \tag{3.5}$$

where $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients, $x_1, x_2, \ldots, x_k$ represent the independent variables, y represents the dependent variable, and ε represents the unknown error. The variables are selected based on statistical tests such as F-statistics, t-statistics, and p-values. Similarly, the ability of a regression model to describe the relationship between the dependent and independent variables can be assessed by estimating the coefficient of determination (R^2), the Mean Absolute Error (*MAE*), the Mean Square Error (*MSE*), and

the Root Mean Square Error (*RMSE*). The R^2 is the measure of how well a prediction can be made from the independent variable(s) and is a quantitative scale that ranges from 0 to 1, where 0 indicates that the model fails to accurately predict the data and 1 represents a perfect fit. It is important to note that the R^2 value can proportionally increase with the number of independent variables, often without improving the model's predictive capability; for that reason, the adjusted R^2 value is used as a more representative indicator.

Variable selection plays a significant role in the accuracy of a regression model. There are various statistical regression variable selection methods. Based on the most commonly used selection methods, three selection methods were considered in this study (Sweet and Grace-Martin 2010, Rahman et al. 2021):

- Forward selection: The forward selection method starts by building a predictive equation, one variable at a time. The variable with the highest correlation to the *IRI* is added based on the p-value and the t-statistic. The process then repeats itself until no more variables with significant correlations can be added to the model.
- Backward elimination (also known as the backward deletion method): The backward elimination method is the forward selection process in reverse. A model that has all the variables in an *IRI* predictive equation is first generated. The equation variables are then eliminated one at a time, with the least significant correlated variables eliminated first. This process repeats itself until all the remaining variables are highly correlated with the dependent variable.
- Stepwise selection method: The stepwise method is a combination of the previous two methods. The stepwise method is a modified version of the forward selection method where a variable is added, and with each step, a check is performed to see whether the other variables' significance has been reduced below a tolerance level. If any of the variables fall below the tolerance level, they are then removed.

As a result of applying each of the variable selection methods previously described, three models were created, and the best *IRI* model was selected after model validation. Statistical tools such as R^2, the adjusted R^2, the *RMSE*, the *MSE*, the *MAE*, F-statistics, and p-values aided in identifying and selecting the best model out of the three developed. The model development and validation were done in R through RStudio. R is a language developed for statistical computing and graphics that provides various statistical and graphical techniques. RStudio is an integrated development environment for R (R Core Team 2020).

3.7 Results

3.7.1 Regression Analysis IRI Models Results

In this study, a total of 27 variables were considered for the development of the *IRI* models. Fourteen of these variables were related to the climate. The variables considered for the *IRI* models' development were the *IRI* at construction, *PI*, P_{200}, base thickness, age, surface thickness, *ESAL*, transverse cracking, longitudinal cracking, other cracking area, rutting depth, annual minimum air temperature, annual maximum air temperature, average monthly minimum air temperature, average monthly maximum air temperature, average monthly average air temperature, humidity average, annual precipitation, cumulative temperature index, precipitation index, freezing index, thawing index, moisture index, site factor as defined by the 2008 MEPDG, site factor as defined by the 2020 MEPDG, heat index, and Thornthwaite moisture index. From the variables listed, the variables annual minimum air temperature, annual maximum air temperature, average monthly minimum air temperature, average monthly maximum air temperature, average monthly average air temperature, humidity average, annual precipitation, cumulative temperature index, precipitation index, freezing index, thawing index, moisture index, heat index, and Thornthwaite moisture index are directly related to climate. Additionally, the site factor defined by the 2008 MEPDG and the site factor defined by the 2020 MEPDG also includes climate parameters. The variables in the models were selected using the t-statistics and p-values. From the aforementioned variables considered in the models' development, the variables selected in the models are the *IRI* at construction, *PI*, P_{200}, *ESAL*, age, surface thickness, transverse cracking, rutting depth, annual

precipitation, and the other cracking area as defined by AT. Box and scatter plots for the models' selected variables are presented in Appendix A. The predictive models were developed using the entire, the outer fence, and the inner fence datasets. The statistics of the models are first presented for all three datasets, followed by the best-fit models. Table 3.2 presents the R^2, adjusted R^2, *RMSE, MSE, MAE*, F-statistics, and p-value of the models for both the ACP and OL pavement types. The adjusted R^2 differs from the R^2 as it considers the number of variables in the model (degrees of freedom). The adjusted R^2 value drops as the number of independent variables increases, provided that the model's fit increase does not make up for the loss of a degree of freedom (Sweet and Grace-Martin 2010).

Table 3.2 – Relevant models statistics ACP and OL pavements datasets (Analysis of variance table)

	R^2	**Adjusted R^2**	**RMSE**	**MSE**	**MAE**	**F-statistic**	**p-value**
OL - full dataset	0.58	0.57	0.346	0.120	0.241	301	< 2.2e-16
OL - Outer Fence	0.59	0.59	0.294	0.086	0.209	260.6	< 2.2e-16
OL - Inner Fence	0.58	0.58	0.190	0.036	0.145	116	< 2.2e-16
ACP - full dataset	0.31	0.30	0.457	0.209	0.327	51.61	< 2.2e-16
ACP – Outer Fence	0.33	0.32	0.384	0.147	0.301	44.93	< 2.2e-16
ACP - Inner Fence	0.20	0.19	0.321	0.103	0.263	12.81	< 2.2e-16

A review of Table 3.2 indicates the outer fence dataset provides a better overall fit than the full and inner fence datasets. The models developed using the outer fence datasets provided acceptable results for OL pavements with an R^2 of 0.59. The ACP model developed using the ACP outer fence dataset yielded an R^2 of 0.33. For comparison, the MEPDG *IRI* model using the same dataset yielded an R^2 of 0.23. The developed ACP model does show some improvement compared to the MEPDG model; however, the ACP model does not show adequate results. Apart from the model by Al-Suleiman (Obaidat) and Shiyab (2003), all the other *IRI* models reported in Table 3.1 have variables that are not included in AT's dataset and hence cannot be used for a similar comparison. The models developed by Al-Suleiman (Obaidat) and Shiyab (2003) have reported values of R^2 ranging from 0.61 to 0.80. However, *Age* is the only independent variable in their model. Given that these models are only correlated to *Age* and do not consider pavements of different thicknesses and other variables contributing to pavement roughness, such

as traffic and pavement distress information, the models have very limited applicability and are not appropriate for model comparison.

The developed OL and ACP *IRI* models using the outer fence datasets are presented as equations 3.6 and 3.7, respectively:

$$IRI = -0.369 + IRI_0 + 0.0128 * PI + 0.0172 * P_{200} + 0.00371 * Age - 0.000802 * surfthickness - 0.000104 * ESAL - 0.00120 * TrcAr + 0.0257 * OtherCAr + 0.0548 * RUT + 0.00000738 * PPT \quad (3.6)$$

$$IRI = -0.280 + IRI_0 + 0.00816 * PI + 0.00230 * P_{200} + 0.00126 * Age + 0.000109 * surfthickness + 0.0000225 * ESAL + 0.00450 * TrcAr + 0.0130 * OtherCAr + 0.0901 * RUT + 0.000152 * PPT \quad (3.7)$$

Where *IRI* is given in m/km, *Age* is the age of the pavement since construction or since rehabilitation in years (whichever is smaller), *surfthickness* is the total surface thickness in millimetres, *TrcAr* is the transverse cracking as a percentage of area, *OtherCAr* represents other cracking percentage as a percentage of area, *RUT* is the 80th percentile rut depth for the section in millimetre (i.e., 80% are less than & 20% are greater than the value), and *PPT* is the annual precipitation in millimetre. The remaining variables were previously defined in the text.

The developed *IRI* models' relevant statistics were provided in Table *3.2* presented earlier in the text. Visual plots such as the actual vs predicted values plots provide a representation of the models' fit and predictive capabilities. Figure 3.2 presents the actual vs predicted values plots for the developed models.

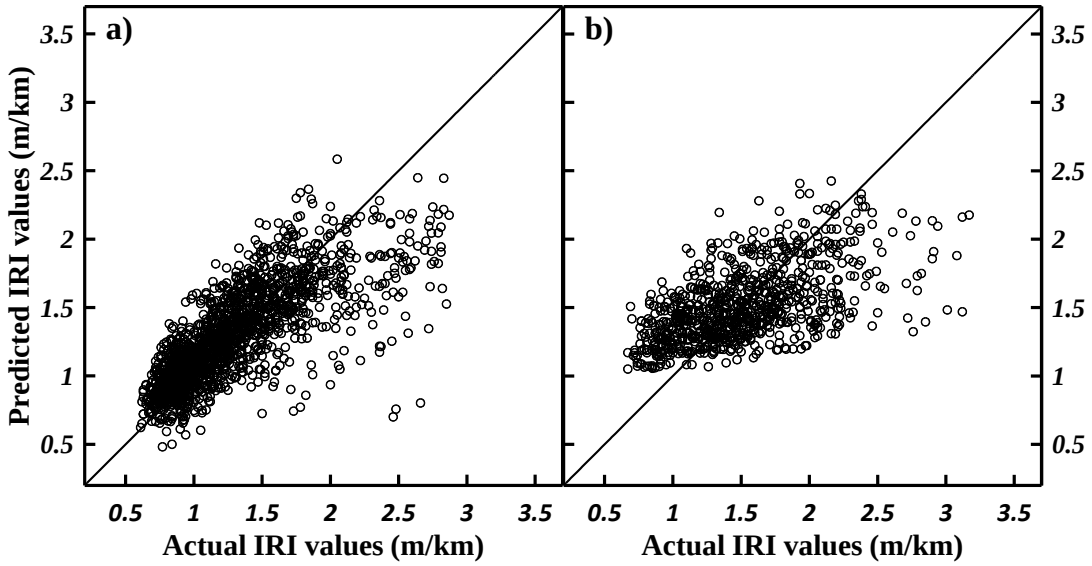

Figure 3.2 - Actual vs Predicted *IRI* values (m/km) for (a) OL *IRI* model (b) ACP *IRI* Model

From the figure, the actual vs predicted values plot for OL pavements appears to fit better as a result of a more accurate model than the ACP model. The actual vs predicted *IRI* plots in Figure 3.2 could be examined further through the use of residual plots as it plots the difference between the actual and predicted *IRI* values. Residual plots provide a graphical representation of the difference between the target variable's observed and the predicted value. Residual plots can be classified into two categories: random patterns and non-random patterns. Points that are randomly dispersed around the horizontal axis suggest that the model's assumptions are appropriate, whereas non-random patterns indicate otherwise (Sweet and Grace-Martin 2010). The residual plots for both the OL and ACP *IRI* models are provided in Figure 3.3. The figure shows the residuals randomly distributed around the horizontal axis, suggesting that the linear model assumption is valid.

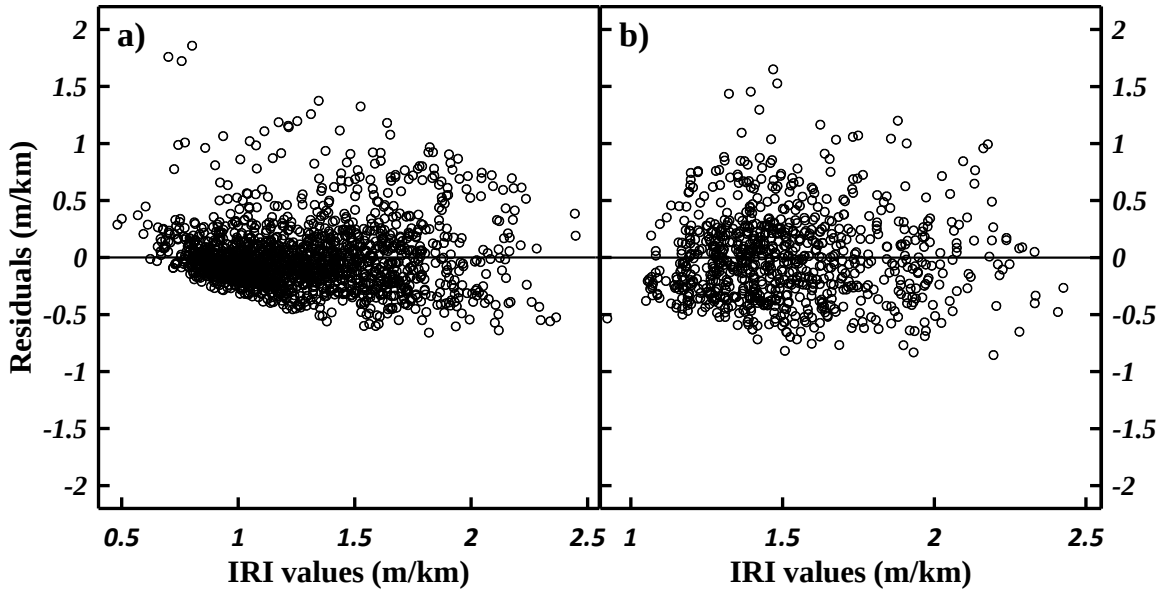

Figure 3.3 - Residual plots for the (a) OL *IRI* model (b) ACP *IRI* Model

The impacts of each of the models' variables on *IRI* can be examined to provide further insight into the models' structure. A sensitivity analysis is carried out to provide an understanding of how each of the independent variables contributes to the models' output. The most common method for visually presenting the sensitivity analysis results is in the form of a tornado plot, where the independent variables with the greatest contribution to the target variable's estimate are at the top, and the least significant contributors are at the bottom. Using the two developed *IRI* models, sensitivity analyses are conducted and arranged in a tornado plot format in order to identify the variables with the most significant effect on the *IRI* models' outputs. The arithmetic average of the independent variables' values in the models was used as a base case, and an average *IRI* of 1.28 and 1.53 was computed for OL and ACP models, respectively. Afterward, each variable's value was varied between the maximum and minimum values from their mean values, while the other variables' values were the same as in the base case. The sensitivity analysis plot in the form of a tornado plot for the OL model is presented in Figure 3.4.

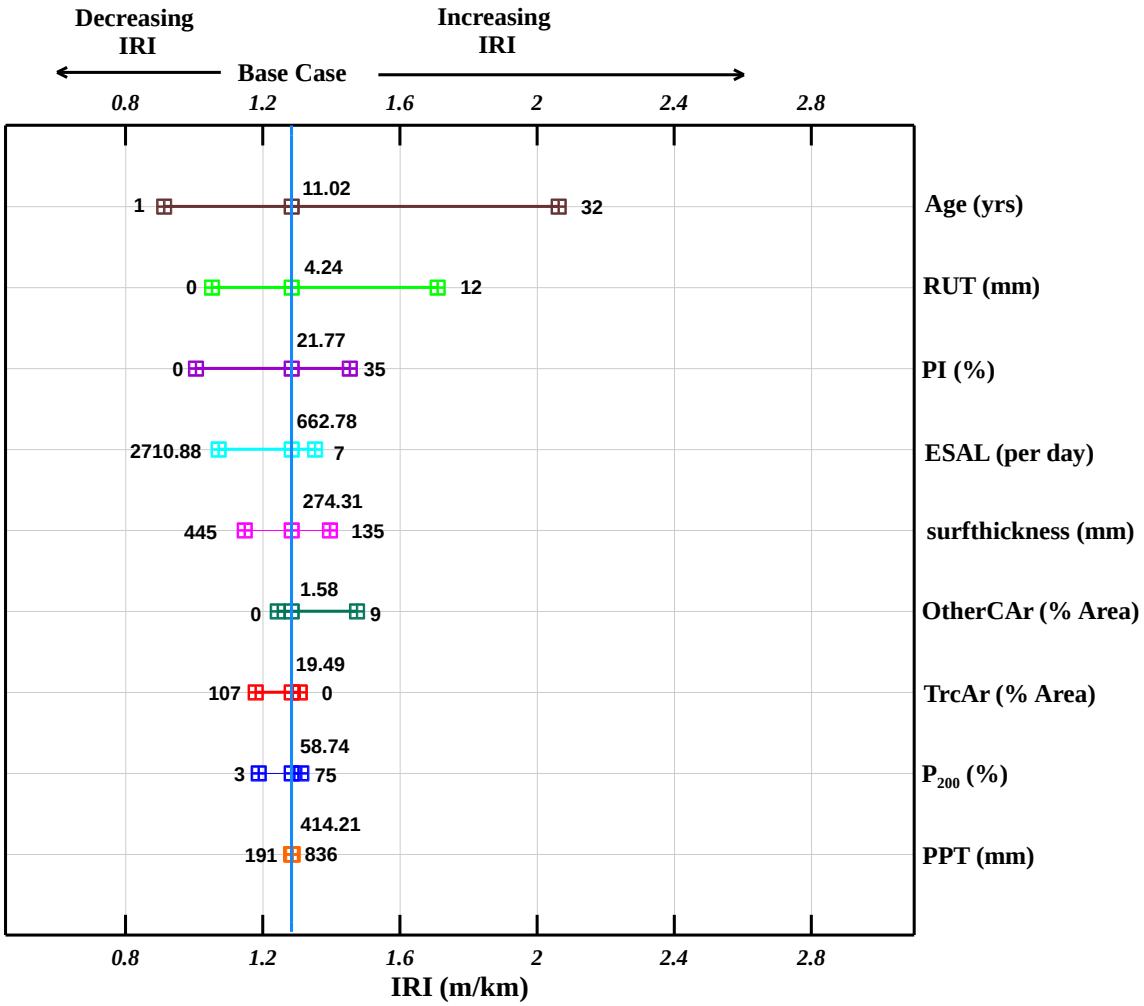

Figure 3.4 - Regression analysis developed OL *IRI* model sensitivity analysis Tornado plot

The sensitivity analysis presented in Figure 3.4 shows *Age* as the most impactful variable to the OL pavement *IRI* model's output. This indicates that the age since the last rehabilitation or construction of the pavement, whichever is smaller, plays a significant role in estimating the *IRI* value in the model. Following closely, *RUT* is the second most impactful variable as a result of rutting being the physical depressions in pavements. Moreover, *PI* appears to significantly impact the model's output closely after *RUT*. The *PI* variable's impact on the model shows that having a low or no *PI* is linked to a smoother pavement while, in contrast, a higher *PI* leads to a rougher pavement. The relationship between pavement roughness and *PI* results from the plasticity index's relation with frost-heave effects in soils. Soils with higher *PI* values can potentially have larger base volume changes due to frost-heave effects. The *PI* and

P_{200} for the base materials are often used to relate these frost-heave volume changes. In some instances, they are represented as a single term, namely, the weighted plasticity index, WPI, which is the product of the two variables divided by a hundred (ARA Inc. 2020). These two soil properties have been widely used in the MEPDG for predicting the freeze and thaw effect for different soil types (ARA Inc. 2020). Additionally, the *surfthickness* and *ESAL* variables are inversely related to *IRI*. For the case of *surfthickness*, one possible explanation could be that thicker pavements can be less prone to pavement distresses as opposed to thinner pavements. The inverse relationship of *ESAL* and *IRI* is possibly explained by the *ESAL* variable's correlation with the variables in the model, especially distress variables. For instance, a larger traffic load will likely produce more severe pavement distresses, leading to rougher pavements which translates to greater *IRI* values. The variable *ESAL* could be considered a categorical variable where the variable reflects the importance of the pavement based on its traffic load. For instance, larger *ESAL*-designed roads reflect a pavement designed to be more resilient to distresses caused by larger repetitive traffic loads and are given a more attentive Maintenance and Rehabilitation (M&R) schedule, resulting in smoother pavements. Moreover, other forms of cracking defined by AT as *OtherCAr* has a significant impact on the roughness model. Finally, the *TrcAr* impact on the *IRI* appears to be inversely related to *IRI*. This might be due to collinearity with other variables, such as *RUT* and *OtherCAr*, where a pavement with high transverse cracking is likely to show large rutting and other cracking values. This is not surprising as larger cracking in the pavements is expected to contribute to pavement roughness.

The sensitivity analysis results presented in the tornado plot indicate that the variable *PPT* has virtually no impact on *IRI*. The insignificant contribution of *PPT* to the predicted *IRI* is most likely due to the variable being correlated to other independent variables, such as the distress variables present in the model. The correlation of *PPT* with the other variables in the model results in the *PPT* variable posing little to no impact on the *IRI* model's output. The variable *PPT* on itself is independent and is not affected by the other variables in the model. The correlation between *PPT* and the other variables in the model results from other variables being dependent on variations in *PPT*, such as pavement distresses caused due to

precipitation. Additionally, the variable's insignificant contribution to the predicted *IRI* could be linked to the temporal resolution of *PPT*. A similar sensitivity analysis is conducted for the ACP model and is presented in Figure 3.5.

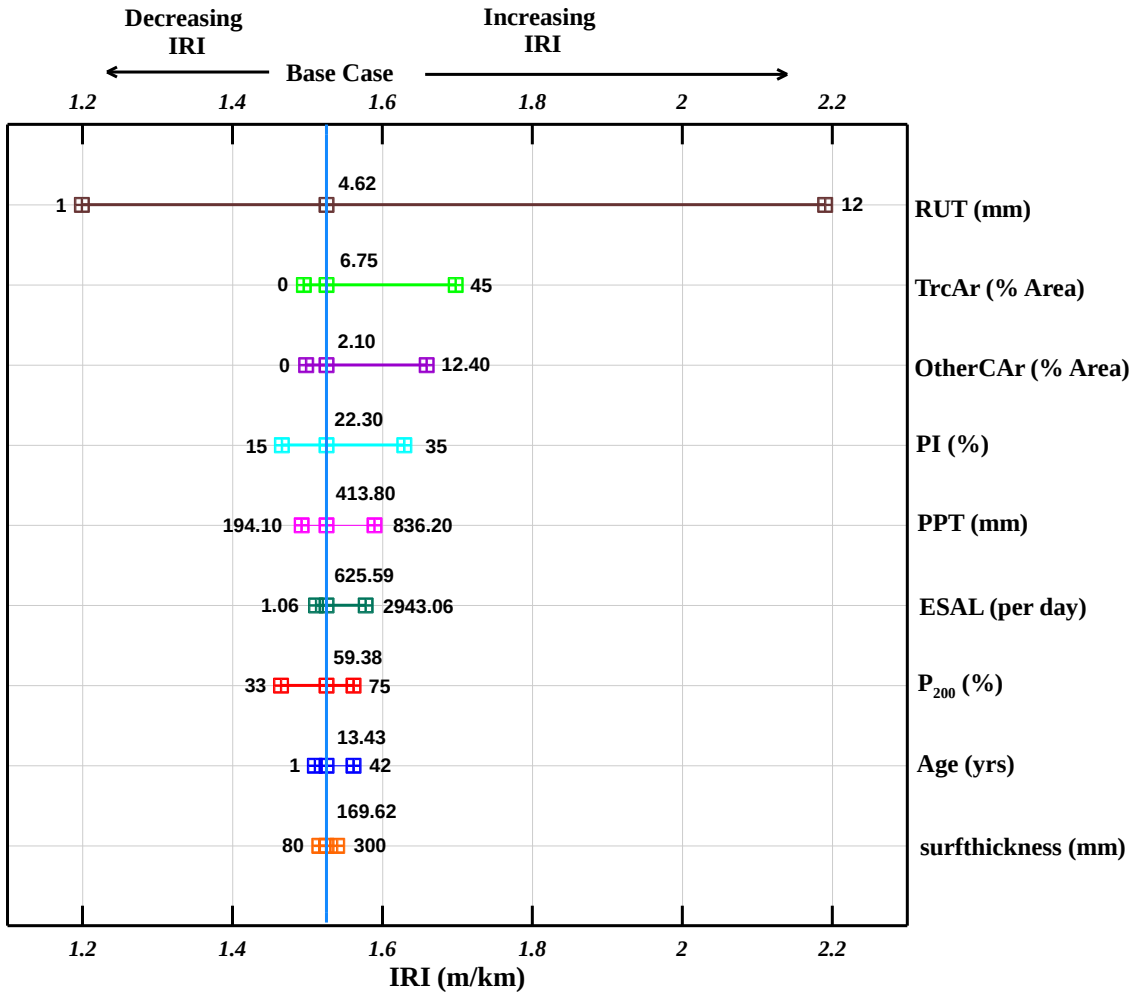

Figure 3.5 - Regression Analysis developed ACP *IRI* model sensitivity analysis Tornado plot

The sensitivity analysis conducted for the ACP *IRI* model in the form of a tornado plot shows that the ACP model is most sensitive to the variations in the *RUT* variable, much more than any other variable in the model. The variable *TrcAr* is the second most significant variable in the model. This could be the result of transverse cracking running perpendicular to the direction of the road, resulting in rougher pavements. Similar to the OL model, *OtherCAr* and *PI* have a high impact on the model's output. Furthermore, the variable *PPT* has a larger impact on the ACP model than on the OL model. This could

result from lower collinearity between *PPT* and the other variables in the ACP model than in the OL model. Additionally, ACP pavements could be more prone to damage caused by *PPT* than OL pavements. The *ESAL* effect on the ACP model's output is observed to be proportional rather than the inversely proportional relationship observed in the OL model. The P_{200} variable has a relatively smaller yet significant impact than the variables listed above. Lastly, both *Age* and *surfthickness* have a minor effect on the ACP model's output. The latter variable's effect on the model's output mirrors what was seen in the OL *IRI* model, while the former does not.

The conducted sensitivity analysis plots for both *IRI* models were utilized to identify the five most impactful independent variables for each of the models. The most impactful variables are then used to create reduced *IRI* models for both pavement types. For OL pavements, the sensitivity analysis defined *Age*, *RUT*, *PI*, *OtherCAr*, and P_{200} as the most impactful to the model's output. The variables were used to create the reduced model for OL pavements, and the model is presented in equation 3.8 as follows:

$$IRI = -0.670 + IRI_0 + 0.0442 * Age + 0.0333 * RUT + 0.0150 * PI + 0.0209\, OtherCAr + 0.00105 * P_{200} \quad (3.8)$$

All the variables used in the equation above were previously defined in the earlier sections. The model's statistics are presented in Table 3.3. The model's adjusted R^2 is around 0.55, a small reduction from the more complex developed model that showed an adjusted R^2 of 0.57. However, the more complex OL *IRI* model had larger values for the error statistics: *RMSE*, *MSE*, and *MAE*. For instance, the more complex model had an *RMSE* of 0.346 and an *MAE* of 0.241, compared to 0.307 and 0.219 for the reduced model. This comparison between the more complex *IRI* model for OL pavements and the reduced model emphasizes the importance of considering all the model's statistics rather than exclusively considering the R^2 and adjusted R^2 values as the sole indicators of the model's accuracy.

Table 3.3 – Reduced *IRI* models statistics

	R^2	**Adjusted R^2**	**RMSE**	**MSE**	**MAE**	**F-statistic**	**p-value**
OL	0.55	0.55	0.307	0.0942	0.219	404	< 2.2e-16
ACP	0.32	0.32	0.386	0.149	0.302	65.8	< 2.2e-16

Similar to the OL *IRI* model, the sensitivity analysis conducted for the ACP model was used to identify the five most impactful variables to develop a reduced model. The five most impactful variables in the ACP *IRI* model are *RUT*, *TrcAr*, *OtherCAr*, *PI*, and *PPT*. The variables are used to develop a reduced *IRI* model for ACP pavements, and the *IRI* model is presented as follows:

$$IRI = -0.0830 + IRI_0 + 0.0963 * RUT + 0.00413 * TrcAr + 0.0141 * OtherCAr + 0.00598 * PI + 0.000166 * PPT \quad (3.9)$$

All the variables used in equation 3.9 are as described earlier in the text. The reduced ACP *IRI* model's statistics are also presented in Table 3.3. The model's F-statistic and p-value are acceptable to reject the null hypothesis and thus validate the ACP *IRI* model. The model's R^2 of 0.32 is slightly smaller than the more complex model's R^2 of 0.33. However, as mentioned previously, R^2 does not account for the number of variables present in the model and their relevance to the model's performance, thus, the adjusted R^2 is used. The adjusted R^2 for both the complex and reduced ACP *IRI* models are the same at 0.32. The error statistics for both models are similar and only differ in the third decimal place. The model statistics clearly indicate that the more complex model does not provide any significant statistical improvement over the reduced model.

The actual vs predicted *IRI* values plots are plotted to visualize the model's accuracy in predicting the *IRI* value for both pavement types. The plots for the reduced models for both pavement types are presented in Figure 3.6. The plots are similar to what was previously observed for the actual vs predicted plots for the more complex models. The actual vs predicted *IRI* plots show that for the OL model, the model is the most accurate at *IRI* values of 0.75 to 1.2. The accuracy for the OL model deteriorates significantly at *IRI*

values larger than 1.2, especially at *IRI* values between 2 and 3. The ACP model itself is of low accuracy compared to the OL model. The ACP model is the least accurate between the *IRI* values of 2 and 3.5. The plots support the hypothesis that the *IRI* model developed for OL pavements provides a better fit than the model developed for ACP. Nonetheless, the reduced *IRI* models for both pavement types provide acceptable results for this research purpose.

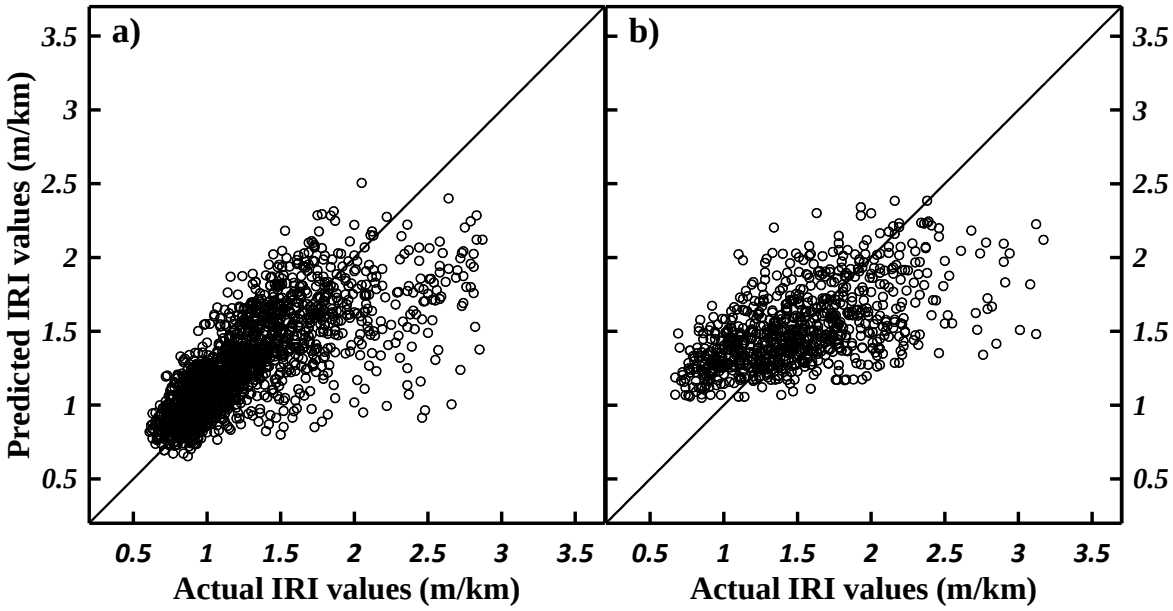

Figure 3.6 - Actual vs Predicted *IRI* values (m/km) for (a) OL *IRI* reduced model (b) ACP *IRI* reduced model

A useful representation of the model's accuracy is done with the use of residual plots. Residual plots, as previously defined, illustrate the difference between the observed and predicted values. The residual plots for both pavement types using the reduced *IRI* models are presented in Figure 3.7. The residual plots for both pavement types are distributed randomly around the horizontal axis; thus, the linear model assumption is acceptable for the developed models (Sweet and Grace-Martin 2010). The figure is similar to what was observed for the more complex models' residual plots presented earlier in the text. Figures 3.6 and 3.7 support using the reduced models as a viable alternative to the more complex *IRI* models developed in this research.

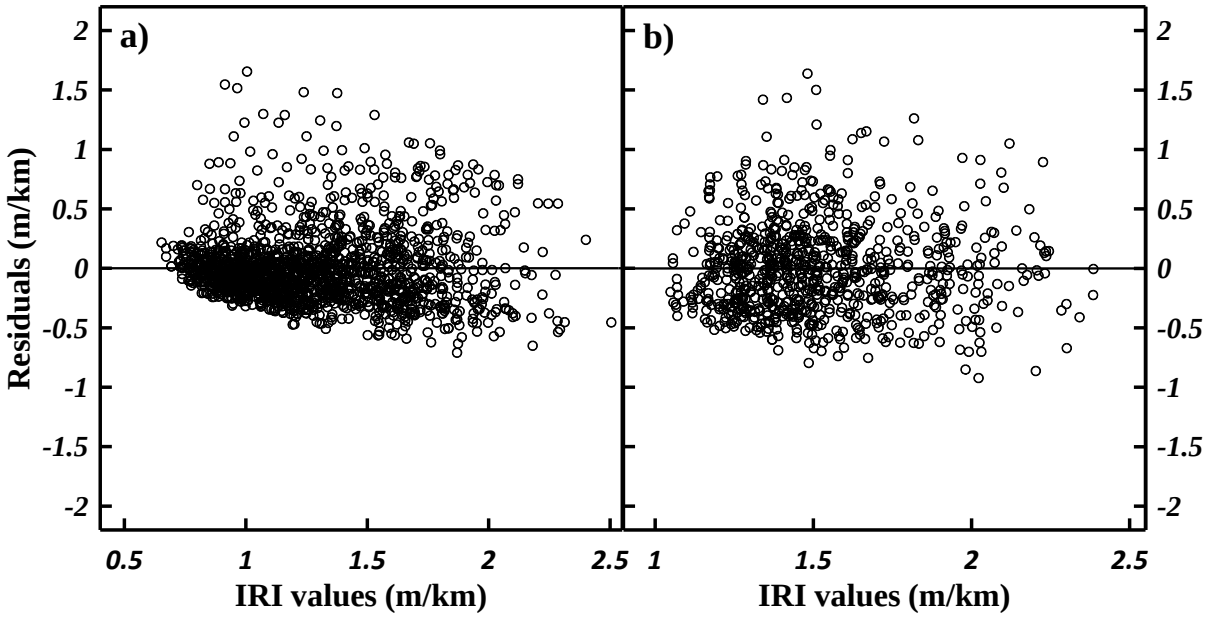

Figure 3.7 - Residual plots for the (a) OL reduced *IRI* model (b) ACP reduced *IRI* model

A sensitivity analysis was performed using the reduced models to provide insight into the impact of each of the independent variables on the dependent variable, *IRI*. This analysis is conducted solely to understand how each independent variable impacts the predicted *IRI* values. The sensitivity analysis conducted for the OL pavement types using the reduced *IRI* model is presented in Figure 3.8. The conducted sensitivity analysis is presented in a tornado plot format to show the relative impact of each of the model's variables on the model's output.

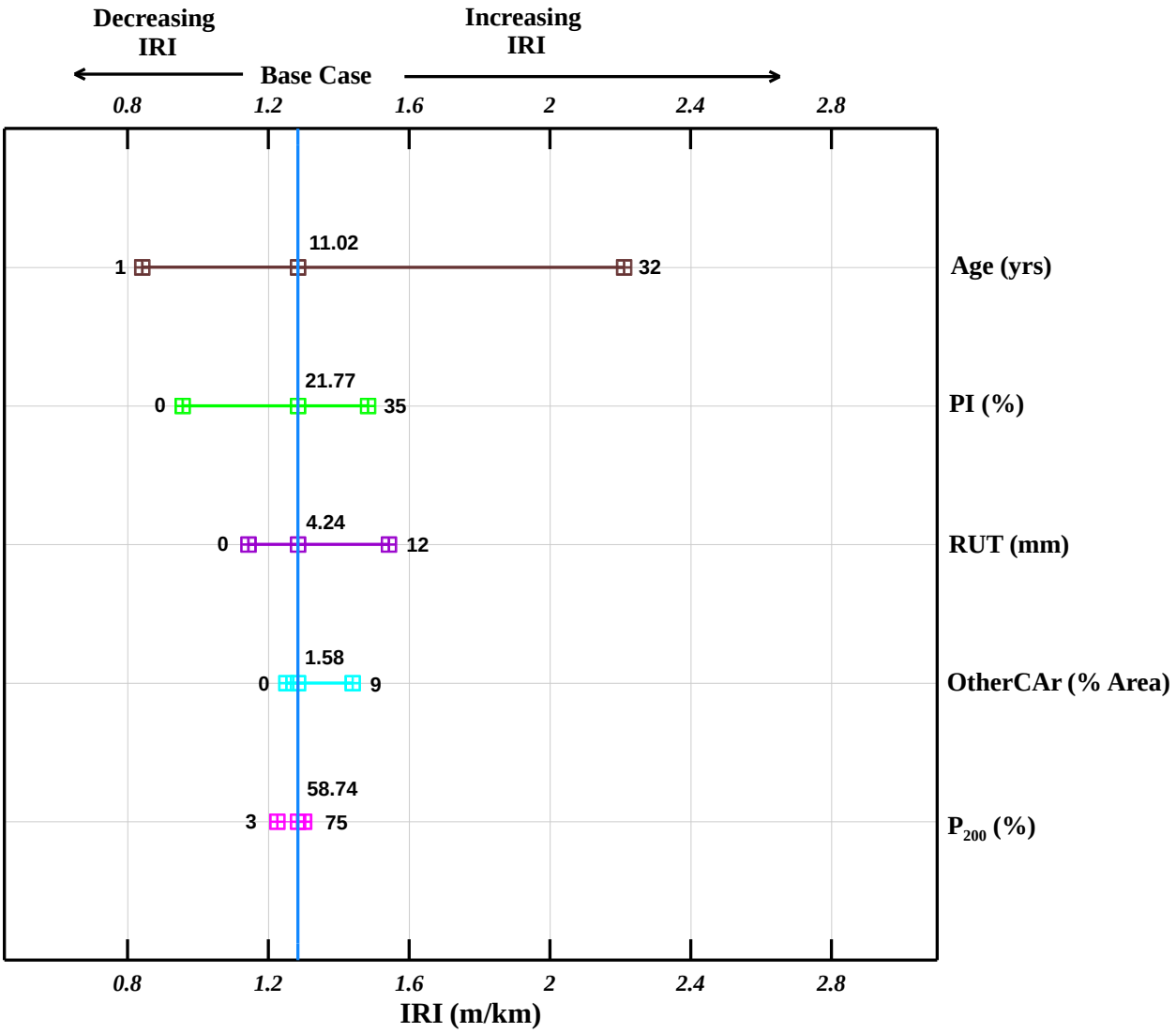

Figure 3.8 - Regression analysis developed OL *IRI* reduced model sensitivity analysis Tornado plot

The sensitivity analysis conducted for the reduced OL *IRI* model presented in a tornado plot format shows that the relative impacts of the five independent variables mostly agree with the sensitivity analysis conducted for the more complex model. Figure 3.8 identifies *Age* as the most impactful variable, followed by *PI*, *RUT*, *OtherCAr*, and P_{200} in that order. The figure showcases that the variables in the reduced model for OL pavements have similar impacts on the model's output to that of the more complex model. Additionally, the ranking of importance for the variables in the reduced model is identical to the more complex model.

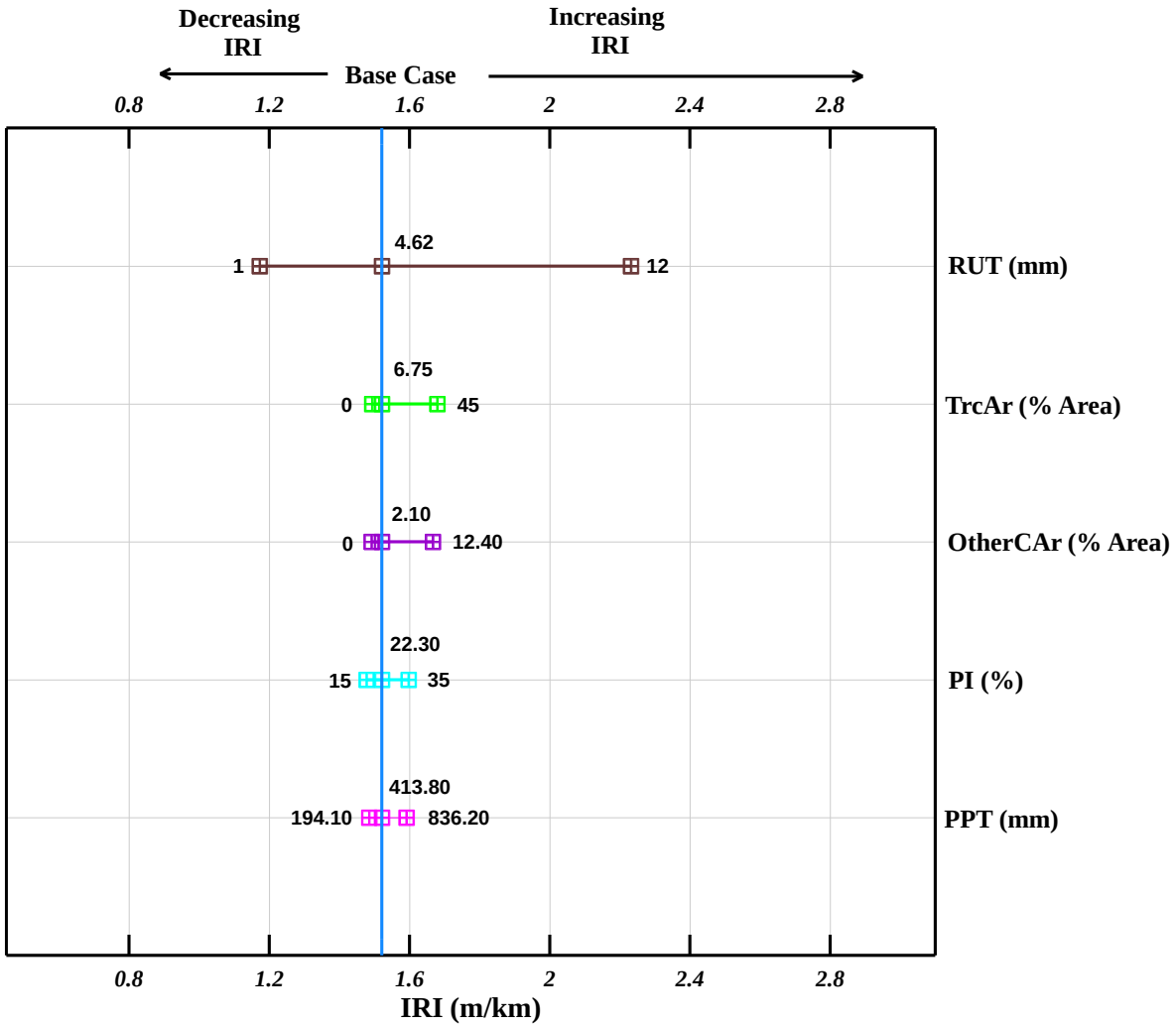

Figure 3.9 - Developed ACP *IRI* regression analysis reduced model sensitivity analysis Tornado plot

A similar sensitivity analysis is conducted for the reduced ACP *IRI* model and is presented in Figure 3.9. This figure provides an understanding of the impact of the independent variables selected for the reduced ACP *IRI* model on predicting *IRI*. Comparing the more complex ACP *IRI* model to the reduced model shows that the variables' impact ranking for the reduced model mirrors that of the more complex model. The figure shows that *RUT* is the most impactful variable, followed by *TrcAr*, *OtherCAr*, *PI*, and *PPT*. The analysis confirms that rutting is by far the most important variable and has the most impact on the *IRI* model's output, similar to what was observed in the more complex model.

3.7.2 LCA, LCCA, and MEPDG Model's Comparison

A case study is conducted to compare the regression analysis developed *IRI* model with the MEPDG model using an LCA and LCCA. Since the MEPDG model is used as a benchmark in this case study, the MEPDG model was tested using the same data used to develop and test the new ACP *IRI* model. The MEPDG model resulted in an R^2 value of 0.23, a mere 43.5% drop from the R^2 obtained using the developed ACP *IRI* model.

The LCA and LCCA for comparing the models were conducted using a representative section in Alberta's PMS database over a 50-year period. Details on how items in the LCA and LCCA are calculated are provided in Appendix B. The selected section has a length of 619 metres and is a part of the Trans-Canada highway around 140 kilometres east of Calgary and around 4 kilometres north of Bassano, with a posted speed of 110km/h. In AT's PMS database, the section is identified as Hwy 11A:06 (West). The selected pavement section is a two-lane roadway with a width of 7.5 metres, a GBC thickness of 500 millimetres, and a surface thickness of 200 millimetres.

An M&R schedule is required as input for the LCA and LCCA. The *IRI* models are used to create M&R schedules for the aforementioned pavement section. The selected 50-year period is expected to have multiple M&R activities. The activities are performed based on AT pavement thresholds. According to AT guidelines, *IRI* values (m/km) lower than 1.5, greater than or equal to 1.9 and between 1.5 and 1.9 correspond to good, poor and fair ratings, respectively (Jurgens and Chan 2005). This rating is for highways with a posted maximum speed of 110 km/h. The threshold for poor pavement is raised to 2.1 m/km for highways with a maximum speed below 110 km/h (Jurgens and Chan 2005). The first M&R activity is a rout and seal crack repair performed at an *IRI* of 1.5 m/km. The second M&R activity is a cold mill and inlay and is performed at an *IRI* of 1.9m/km. In the M&R schedules, the rout and seal crack repairs lead to an *IRI* drop of 0.2 m/km (Soleymani et al. 2008). The cold mill and inlay M&R activity is reflected in the M&R schedule by setting age and distresses to 0 (Soleymani et al. 2008). The M&R schedule for the models is presented in Figure *3.10*.

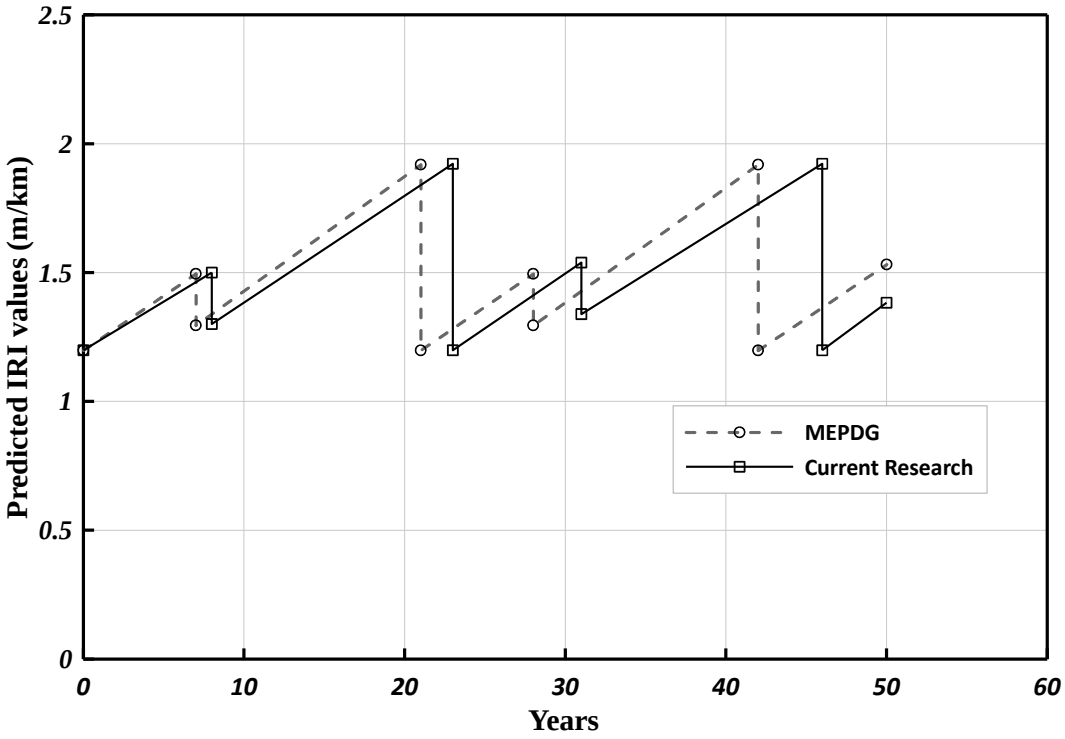

Figure 3.10 – M&R schedule using the developed and the MEPDG model for the case study section

Figure 3.10 illustrates how pavement deteriorates over the 50-year case study period for each of the models under consideration. The aforementioned *IRI* trigger thresholds for M&R of 1.5 and 1.9 m/km are chosen for maintenance and major rehabilitation for both models, respectively. The figure shows that the MEPDG model over-predicts the roughness of ACP pavements. This implies that if the MEPDG model is used, the M&R activities will be carried out earlier, not utilizing most of the pavement's lifetime. Therefore, using an *IRI* model with better predictive capabilities will help agencies and municipalities accurately forecast pavement maintenance costs. The difference between the forecasted M&R cost forecasts is due to the inflation effects presented through the net present value. Producing accurate M&R activity schedules is very useful for agencies to secure the public funding required to maintain and rehabilitate pavements on time while maintaining appropriate road roughness levels for safety and comfort.

Using the produced M&R schedule for both *IRI* models, the LCA and LCCA are used to further compare the developed and MEPDG *IRI* models. The LCA examines the environmental impact of a project, while

the LCCA examines its economic impact. The use of LCA and LCCA is conducted with the aid of Athena Pavement LCA software by the Athena Sustainable Materials Institute (Athena Sustainable Materials Institute 2018). This LCA is performed in accordance with North American standard practices and ISO 21930 and 21931 (Athena Sustainable Materials Institute 2018). An LCA is conducted to compare the developed and MEPDG model. The maintenance schedule and pavement roughness presented in Figure 3.10 is used to conduct the LCA. The result of this analysis provides the global warming potential and smog potential, as presented in Figure 3.11.

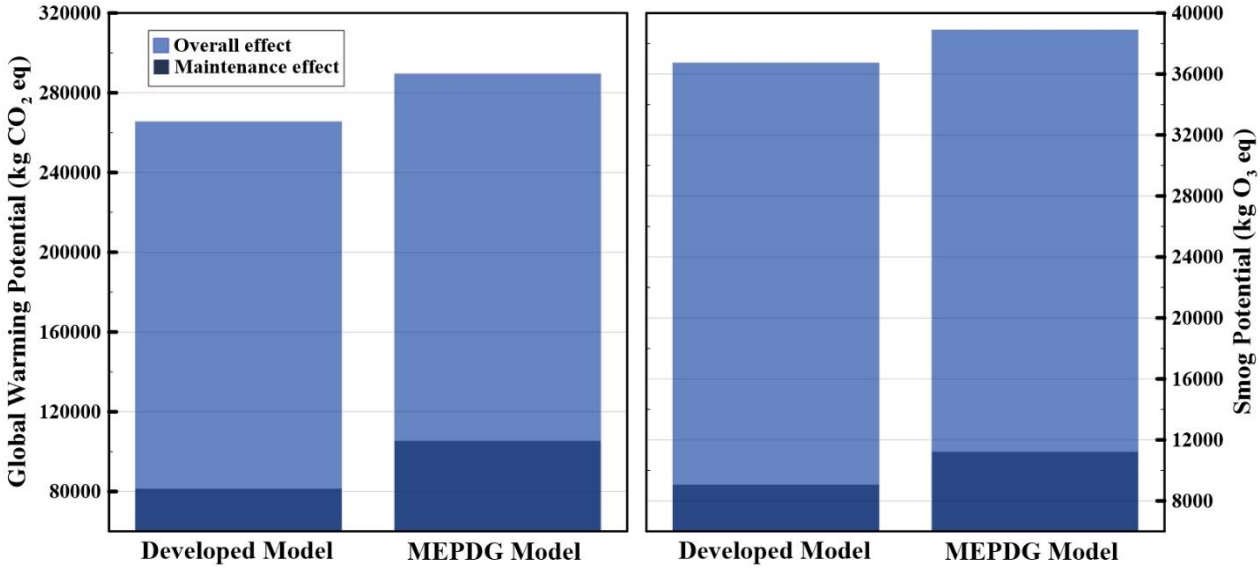

Figure 3.11 - Global warming and smog potential for the developed and the MEPDG model from LCA

Figure 3.11 shows both the total global warming and smog potentials for the developed and the MEPDG model. The effect of maintenance, inclusive of the effect of roughness, is highlighted within the overall impact in the figure. The overall impact includes all pavement activities' impacts, including but not limited to: site preparation, construction, maintenance, used materials, and pavement roughness impact. The figure illustrates the impact of pavement roughness and alternative M&R schedules in terms of environmental factors such as global warming and smog potential. The developed ACP model presents lower global warming potential and smog potential than the roughness obtained by the MEPDG model.

The pavement roughness predicted by the MEPDG model led to higher estimates of CO_2 (global warming potential) and O_3 (smog potential) equivalents.

As part of the LCA, Figure 3.12 compares the total energy consumption between the two models.

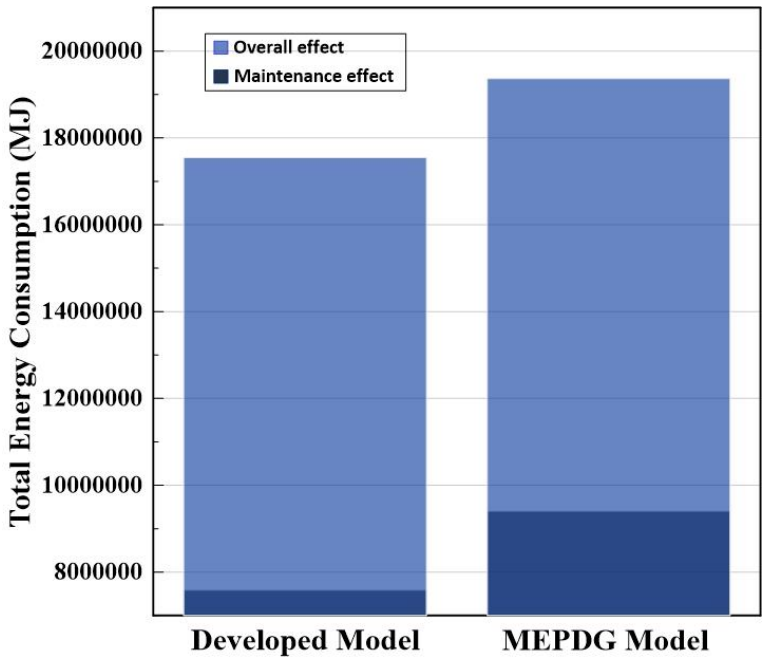

Figure 3.12 – Total Energy Consumption

Figure 3.12 illustrates the total energy consumption due to the roughness predicted by the two models. The total energy consumption prediction is higher when using the MEPDG model due to a higher pavement roughness and maintenance schedule. Pavement roughness, as presented in Figure *3.10*, shows that the MEPDG model predicted higher *IRI* values than the ACP models. For instance, between the years 10 and 20 of the life cycle, the MEPDG *IRI* model predicted a higher *IRI* value at any given year than the ACP model. The prediction of higher *IRI* values is considered in the LCA through the reduction of fuel efficiency of vehicles. The use of the MEPDG model predicts an increase in the total energy consumption, which leads to higher global warming and smog potential. A more representative *IRI* model will provide a more realistic estimate of the environmental impact.

Aside from environmental advantages, a more representative predictive model for *IRI* can also provide financial advantages. The LCCA is an engineering technique that is based on well-established principles

of economic analysis to analyze the differences in long-term efficiency between competing options in an analytical and fact-based manner (Sweet and Grace-Martin 2010, Moges et al. 2017). Based on the aforementioned *IRI* trigger thresholds, an LCCA was conducted to compare the MEPDG model and the ACP model developed in this study in terms of Net Present Value (NPV). The LCCA is conducted using identical input parameters to identify the cost differences as a result of using the two *IRI* models. LCCA's total expected cost is calculated as follows (Moges et al. 2017):

$$TEC = RMC + CC - SV \tag{3.12}$$

Where *TEC* is the total expected cost, *RMC* represents the routine maintenance cost, *CC* represents the construction cost, and *SV* is the salvage value. One of the most significant impacts on the LCCA is the terminal *IRI* value at the end of the life cycle, as this impacts the salvage value. The salvage value is calculated as follows (Moges et al. 2017):

$$SV = \frac{ULLOL\ x\ CCOLL}{TELOL} \tag{3.13}$$

Where *ULLOL* is the usable life in the last overlay in years, *TELOL* is the total expected life of the last overlay in years, and *CCOLL* is the construction cost of the last overlay in years. An LCCA of 50 years is taken to compare the two models using the discount rate of 4% suggested by AT (Moges et al. 2017). The NPV is used in the LCCA to determine the net benefit in today's dollar value. The selected road segment's pavement properties, soil properties, and distresses were inputted into both models. A summary of the results can be found in Table 3.4. The unit rates used for various construction, maintenance, and rehabilitation activities were estimated and included associated costs such as labour. The utilized unit rates are presented in Appendix B. In the LCCA, the site preparation and construction activity includes excavation of existing roadway, base, subbase, and asphalt works for the case study pavement section. The unit rates are from AT's reported unit price average based on the three lowest bids on recent tenders (Alberta Transportation 2021). Individual rates can fluctuate depending on the task and agency. In this analysis, unit rates were merely used to illustrate the relative advantage of one model over another.

Table 3.4 – LCCA Comparison of the developed and MEPDG model

Developed Model	***Year***	***Site Preparation and Construction***	***Maintenance and Rehabilitation***	***Total Agency Cost***
	0	$657,998.49	$0.00	$657,998.49
	9	-	$18,529.07	$18,529.07
	23	-	$47,027.41	$47,027.41
	32	-	$765.38	$765.38
	46	-	$39,654.07	$39,654.07
	50	-	-$18,831.37	-$18,831.37
	Sum	$657,998.49	$87,144.56	$745,143.05
MEPDG Model	***Year***	***Construction***	***Maintenance and Rehabilitation***	***Total Agency Cost***
	0	$657,998.49	$0.00	$657,998.49
	7	-	$20,041.04	$20,041.04
	21	-	$50,864.85	$50,864.85
	28	-	$895.39	$895.39
	42	-	$46,389.65	$46,389.65
	50	-	$0.00	$0.00
	Sum	$657,998.49	$118,190.93	$776,189.42

The M&R activities performed using both models are assumed to be identical and only differ in schedule, as presented in Figure 3.10. The first activity is mainly a rout and seal of cracks as well as spot repairs. The second activity is the milling of the surface and the replacement with a new asphalt surface. Shortly following that, a rout and seal of cracks is performed. Finally, the pavement undergoes surface milling, a full-depth base repair, and is then resurfaced with a new asphalt surface. The maintenance cost difference, including the salvage cost between the two models, is 26% on a percentage basis. In comparison with the MEPDG model, the application of the developed model will provide a financial benefit as a result of a differing M&R schedule and the aforementioned 4% discount rate used by Alberta.

3.8 Conclusions

Regression analysis models were developed in this chapter to predict the International Roughness Index (*IRI*) for Asphalt Concrete Pavements (ACP) and Asphalt Concrete Overlay pavements (OL) roads

in the province of Alberta by using the PMS dataset provided by Alberta Transportation. The developed *IRI* models were found to be superior in performance to the general models, such as the MEPDG model. In other words, *IRI* models using locally specific data perform better than general models. Additionally, the regression analysis models for *IRI* provide acceptable results in some instances.

In the variables considered for the *IRI* models' development, 14 of the variables were climate-related. The considered climate variables appear to have no or weak correlation with *IRI*. This does not necessarily mean that climate does not affect pavement roughness. One can conclude that the effect of climate is indirectly incorporated in the pavement distress measurements, and no meaningful relationship between the pavement roughness and climate variables is apparent at the temporal resolution of climates variables and *IRI* measurements.

Sensitivity analyses were conducted on the developed *IRI* models to identify the most impactful variables on the models' output. The sensitivity analyses helped conclude that for OL pavements, the age of the pavement, the 80th percentile rut depth, plasticity index, other cracking as defined by Alberta transportation, and the percent passing 200 sieve were the most impactful in the model. In addition, the 80th percentile rut depth, transverse cracking, other cracking as defined by Alberta transportation, plasticity index, and the annual precipitation were the most impactful for ACP pavements. One can conclude that the most impactful variables are dependent on the pavement type and could be used to understand the deterioration of pavements of different types over time.

Reduced *IRI* models were developed using regression analysis and the most impactful variables for both pavement types. The reduced models appear to be as efficient as their more complex counterparts and perhaps can help agencies in identifying which parameters to measure.

Furthermore, soil types and information related to soil plasticity appear to be important from the perspective of *IRI* model development. Considering that some of this information might be available at

the time of construction or can be collected at that time quite easily, road agencies should make it a priority to include it in the PMS database.

A case study was conducted to examine the models developed in this chapter's performance in comparison to the MEPDG model using a selected road section part of the Trans-Canada highway over a 50-year period. The life cycle cost analysis, with a discount rate of 4%, showed a 26% reduction in maintenance and rehabilitation costs from the alternative schedule presented by the developed model in this chapter. The results of this study clearly indicate that the development of accurate models is useful in terms of maintenance and rehabilitation schedules and results in significant savings.

Further research examining other datasets should be conducted using the developed model to further confirm this study's findings. A comparison of the developed model to other models published in the literature should be made using a life-cycle analysis and a life-cycle cost analysis to solidify this study's findings. Additional research needs to be carried out to assess whether the use of machine learning could further improve the *IRI* model.

Chapter 4 : Development of International Roughness Index (*IRI*) Models using Machine Learning

Abstract

The Pavement Management System (PMS) is a heavily utilized tool by transport agencies for pavement-related decisions such as pavement design, maintenance, and rehabilitation. The PMS can be developed to aid decision-makers in creating optimum strategies for delivering, evaluating, and maintaining pavements in serviceable conditions over time. An integral part of the PMS is the pavement roughness indicator, such as the commonly used International Roughness Index (*IRI*). The index attempts to represent the effect of climate, traffic, pavement structure, distresses, loading, and other factors that contribute to road roughness. The index itself aids in establishing a standardized method to assess pavement roughness for evaluating and maintaining pavements over time. Achieving accurate results from the *IRI* models often require transportation agencies to calibrate *IRI* models or develop their own predictive *IRI* models. Commonly used *IRI* models, such as the *IRI* model in the Mechanistic-Empirical Pavement Design Guide (MEPDG), heavily uses regression analysis and lack the use of emerging tools such as Machine Learning (ML) to develop the models. This study focuses on examining the effectiveness of some of the most popular and newly emerging ML algorithms for pavement-related applications, such as in creating predictive *IRI* models. Representative control sections from the province of Alberta were used for most of the last decade from Alberta Transportation's PMS database for this study. The extracted dataset is used to identify key pavement features/variables to create a set of predictive *IRI* models with the use of supervised ML algorithms, namely, Artificial Neural Networks, CatBoost, Decision Trees, Elastic-Net, K-Nearest-Neighbours, Lasso, LightGBM, Random-Forest, Ridge, Support Vector Machines, and XGBoost. The best-performing algorithms from this study were the gradient-boosted ensemble ML algorithms, specifically XGBoost and CatBoost. Additionally, other ensemble ML algorithms, such as the Random-Forest and LightGBM, ranked among the best-

performing algorithms in this study. The study results demonstrate significant success in performance with the coefficient of determination (R^2) values of up to 0.99 for training datasets and 0.87 for testing datasets. The models developed in this study provide at least a three folds increase in predictive capabilities compared to the MEPDG *IRI* model.

4.1 Introduction

Maintaining safe and serviceable pavements is the highest priority for transport agencies. Transport agencies employ a sophisticated management process, most commonly the Pavement Management System (PMS). The PMS is defined by the American Association of State Highway and Transportation Officials (AASHTO) as *"the effective and efficient directing of the various activities involved in providing and sustaining pavements in a condition acceptable to the travelling public at the least life cycle cost"* (AASHTO 1985). Pavement performance indicators are crucial to the PMS's performance. One of the most popular pavement performance indicators is the International Roughness Index (*IRI*). The *IRI* is a key indicator used to assess the roughness of pavements. The monitoring of pavement roughness is important as rougher pavements increase fuel consumption and operating costs, and reduce safety and comfort for the drivers (AASHTO 2009). Theoretically, a perfectly smooth pavement is denoted with an *IRI* value of 0.0 and increases proportionally with roughness (Sayers 1995a). Changes in *IRI* are caused by traffic, age, distresses, and environmental factors that affect the pavement's structural integrity. As every road network is characterized by its own traffic patterns and environmental conditions, *IRI* predictive models that are typically created or calibrated by local transport agencies are desirable.

Regression analysis has historically been the foundation of *IRI* predictive models. For instance, the *IRI* models in the Mechanistic-empirical Pavement Design Guide (MEPDG) also use regression analysis instead of emerging technologies such as Machine Learning (ML). Recently, the use of ML regression has been gaining traction for construction and building materials applications (Gong et al. 2019, Marani

and Nehdi 2020, Aravind et al. 2021, Rahman et al. 2021, Sadat Hosseini et al. 2021, Pereira Dias et al. 2021, Song et al. 2021, Sun et al. 2021, Guo and Hao 2021, Kim et al. 2022, Peng and Unluer 2022, Shah et al. 2022). Similarly, *IRI* modelling and other pavement-related applications using ML have also been gaining popularity in recent years (Kargah-Ostadi et al. 2010, Gong et al. 2018b, Fakhri and Shahni Dezfoulian 2019, Choi and Do 2019, Kaya et al. 2020, Luo et al. 2021, Piryonesi and El-Diraby 2021). As a subset of AI, ML is an evolving branch of computational algorithms designed to emulate human intelligence by learning from observed data (EL Naqa and Murphy 2015). Generally, Artificial Intelligence (AI) and, inherently, ML has two main applications: regression and classification (James et al. 2013, Marsland 2014, dhage and Raina 2016, Charte et al. 2019). The ML application of classification is the prediction of a class or a category, while regression is the prediction of a quantity.

There are different categories of ML techniques, and they can be described as unsupervised learning, supervised learning, and reinforcement learning (Dietterich 1997). Each of the categories of ML techniques is unique in how they develop models. Unsupervised learning attempts to identify patterns without the inputs being labelled; meanwhile, supervised learning generates a model or a function that maps labelled inputs to desired outputs (Ayodele 2010). In contrast, reinforcement learning attempts to iteratively learn from mistakes to generate a path of maximum "reward" in its modelling environment (also known as the model's interactive medium)(Ayodele 2010).

Supervised learning is often used in *IRI* modelling as it uses already labelled data to generate the desired output. There have been numerous attempts to use ML modelling under supervised learning to develop pavement roughness predictive models, such as the *IRI*. For these, the *IRI* models attempt to predict *IRI* values (the label) using pavement-related variables (a.k.a the features). The most cited studies using ML techniques to develop *IRI* models are presented in Table 4.1. The table includes both the ML technique used, model developers, and the *IRI* model variables. The number of variables used in the cited *IRI* models varied between 4 and 16. The models in the table also show various variables used for modelling *IRI*. The most common variables used in the models are *Age*, traffic (*ESAL*/*AADT*), rutting,

Freezing Index *(FI)*, and precipitation. Some of the least common variables in the cited models include polish, ravel, shove, and manholes.

Table 4.1 – Most cited studies using ML techniques to develop *IRI* models

Technique	**Model author**	**Variables**
ANN	(Abdelaziz et al. 2020)	Age, fatigue %area, IRI_0, rut depth, and transverse cracking length
ANN	(Choi and Do 2019)	AADT, avg. max. temp., avg min. temp., avg. temp, deciding agent, ESAL, and total rainfall
ANN	(Chou and Pellinen 2005b)	Age, ESAL, FI, IRI_0, NoFT, and precipitation
ANN	(Hossain et al. 2019)	AADT, AADTT, AAMaH, AAMiH, AAP, annual avg. freezing index, and annual avg. temperature
ANN	(Kargah-Ostadi et al. 2010)	Age, FI, milling depth, P_{200}, previous IRI, surface layer thickness, TO, and the time since previous IRI
ANN	(Lin et al. 2003)	Alligator cracking, bleeding, corrugation, manholes, patches, potholes, rutting, and stripping
ANN	(Ziari et al. 2016b)	AADT, AADTT, average precipitation, FI, pavement thickness, and surface thickness
ANN, gene expression programming	(Mazari and Rodriguez 2016)	Age, ESAL, IRI_0, and SN
ANN	(Kaya et al. 2020)	Age, longitudinal cracking, previous IRI, rut, traffic, and transverse cracking
ANN, SVM	(Georgiou et al. 2018)	IRI_{t-6}, IRI_{t-5}, IRI_{t-4}, IRI_{t-3}, IRI_{t-2}, and IRI_{t-1}
Random-Forest	(Gong et al. 2018b)	Age, block, edge, fatigue, FI, pavement thickness, IRI_0, ESAL, longitudinal cracking, patch, polish, potholes, precipitation, ravel, rut, and shove
SVM	(Ziari et al. 2016a)	AADT, AADTT, age, avg. temperature, ESAL, freezing index, pavement thickness, precipitation, and surface thickness

AADT: Annual Average Daily Traffic; AADTT: Annual Average Daily Truck Traffic; AAMaH: Annual Average Maximum Humidity; AAMiH: Annual Average Minimum Humidity; AAP: Annual Average Precipitation; ESAL: Equivalent Single Axle Load; FI: Freezing Index; NoFT: Number of Freeze and Thaw cycles; P_{200}: % passing 0.075 sieve; SN: Structural Number; TO: Thickness Overlay

The ML *IRI* models reported in the literature include a variety of variables, with the R^2 ranging from 0.68 (Mazari and Rodriguez 2016) to 0.99 (Kaya et al. 2020). In Table 4.1, the Artificial Neural Networks (ANN) algorithm is the most commonly used ML algorithm, and its popularity extends beyond *IRI* modelling to a wide range of construction and building material applications. (Terzi 2007, Gopalakrishnan et al. 2017, Gong et al. 2018a, Fakhri and Shahni Dezfoulian 2019, Zhao et al. 2019, Mei and Gül 2020, Zhang et al. 2021, Lu et al. 2021, Mabrouk et al. 2022). The reliance on ANN in modelling *IRI* in the literature supports the need to investigate the effectiveness of other ML algorithms in modelling *IRI*.

The aforementioned literature review indicated the lack of a systematic comparison between the different ML algorithms. The cited *IRI* models earlier in the text also indicate higher reliance on ANN in modelling *IRI*. Several other techniques and ML algorithms which have shown promise in model development should be examined. The models developed by Gong et al. (2019) using the XGBoost algorithm to predict pavement cracking showed promising results. Additionally, models developed by Guo and Hao (2021) using the Random-Forest algorithm to predict the location of potential damage on asphalt pavement also showed promising results. Thus, the need to investigate ML algorithms besides ANN and examine their effectiveness in predicting *IRI* values is justified. Moreover, through the examination of the peer-reviewed literature, no reported studies provided a systematic comparison between various ML algorithms, nor were there studies found that provided a comparison between regression analysis models with ML models using the same data.

The objective of this study is to conduct a systematic comparison of various ML algorithms in predicting *IRI* using various statistical measures. The ML algorithms utilized include algorithms that others have used, such as ANN and emerging algorithms. This study is one of the first studies that uses eleven different supervised machine learning algorithms while providing a meaningful comparison between them in terms of performance. Additionally, the models developed using the ML algorithms are compared with the MEPDG model and the regression analysis developed model presented in Chapter 3.

The impact of each pavement feature/variable used in modelling *IRI* on the models' output is also examined to highlight the most impactful variables on *IRI*. Additionally, several researchers have found that reduced models can be far more effective than their more complex counterparts as they have less chance of producing overfitted models (Hawkins 2004, Zhang 2014, Lever et al. 2016). Therefore, the performance of reduced models is also compared to their more complex counterparts.

4.2 Machine Learning and its Application

Several algorithms fall under the category of supervised learning, including but not limited to Support Vector Machine (SVM), K-Nearest-Neighbor (KNN), decision trees, ensemble methods, regularized linear regression, and ANN. This breakdown structure of ML algorithms related to supervised learning is presented in Figure 4.1.

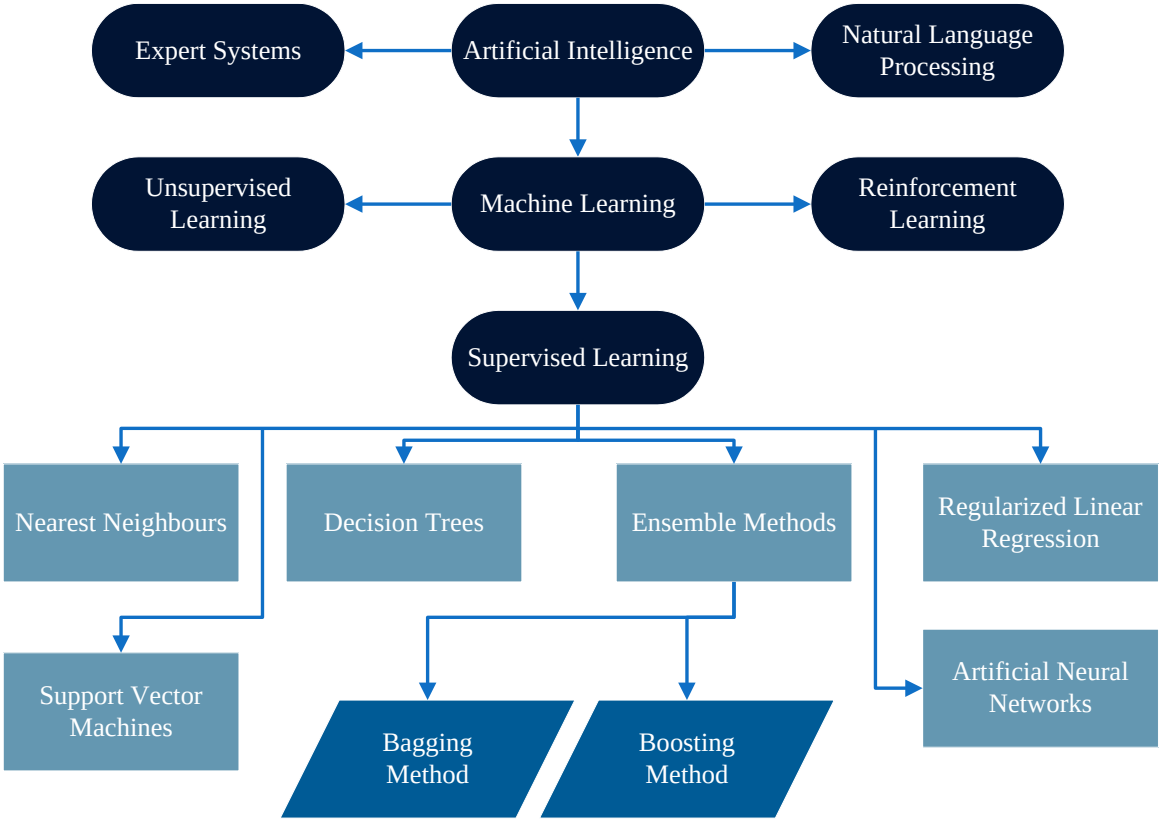

Figure 4.1 - Breakdown structure of ML to the learning categories and algorithms

The SVM algorithm develops a model by constructing a hyperplane in infinite-dimensional spaces (Bishop 2006). The SVM algorithm uses a defined margin (the distance between the decision boundary

and the closest data points) for output selection (Smola and Schölkopf 2004). The margin itself is identified using the Lagrange multiplier's function and is used to create the decision boundary. The algorithm uses the decision boundary to mute the effect of noise and outliers in the model.

Nearest neighbours (KNN) work similarly to SVM. The algorithm uses a predefined number of samples to select the closest neighbours for output prediction (Goldberger et al. 2005). The closest neighbours are identified using the Euclidean distances between the data points rather than the Lagrange multiplier's function as used in SVM. The predefined constant for the number of nearest neighbours to be used is denoted as "k", hence, K-nearest neighbour KNN)(Goldberger et al. 2005).

Decision trees differ significantly from KNN and SVM in the sense that the developed model structure does not depend on the closest data points but rather follows a top-down flow chart to generate an estimate. Decision trees are non-parametric models that predict output using simple decision rules inferred from the features of the training data (Breiman et al. 2017). Decision trees can be represented as flow charts with binary filters at the intermediate or decision nodes. The binary filters could be as simple as filtering if a value is greater than or less than a specified value.

The ANN algorithm differs from SVM, KNN, and decision trees. Unlike the aforementioned categories of supervised learning algorithms, ANN develops a mathematical function between variables to generate a prediction. The algorithm is based on the concept of biological neurons and consists of an input layer, a hidden layer(s), and an output layer (Bishop 2006). The hidden layers transform the values from the previous layers through a weighted linear summation and a non-linear activation function. After the values have been transformed in the hidden layers, the final transformed values are outputted in the output layer (Bishop 2006).

Another category of supervised learning algorithms is the regularized linear regression category. The regularized linear regression category includes Lasso, Ridge, and Elastic-Net algorithms. The algorithms within this category are based on linear regression and the algorithms "shrink" data closer to the

population while improving the least-square error (Friedman et al. 2010). The data "shrinkage" or "regularization" reduces the noises by considering and adjusting for multicollinearity (Kim et al. 2007). The Lasso algorithm uses the "L1" norm regularization, while Ridge utilizes the "L2" norm regularization in model development to reduce inefficiencies resulting from multicollinearity (Friedman et al. 2010). The Elastic-Net algorithms employ both "L1" and "L2" norm regularization in model training (Kim et al. 2007). The combination of the norm regularization methods aids in fully utilizing the training data while minimizing nonzero data weights (Friedman et al. 2010).

Ensemble methods are an ensemble (collection) of decision trees used to generate a prediction. Ensemble methods combine the predictions of several base estimators to improve the overall prediction. Ensemble methods can be divided into two families: the average or bagging method and the boosting method (Louppe and Geurts 2012). In the bagging method, several estimators are built independently, and the final estimation is the average of several estimates (Breiman 1996). In the boosting method, estimates are built sequentially and optimized so as to reduce the overall error (Drucker 1997). The ensemble methods algorithms used in this study are the XGBoost, CatBoost, Random-Forest, and LightGBM algorithms. The XGBoost, CatBoost, and LightGBM algorithms utilize the aforementioned boosting method, while the Random-Forest algorithm uses the bagging method. The XGBoost algorithm constructs an ensemble of decision trees horizontally or level-wise (Chen and Guestrin 2016). The CatBoost algorithm follows a similar process to the XGBoost in building the ensemble of decision trees. However, the CatBoost algorithm imposes the rule that all the nodes at the same horizontal level would test the same predictor with the same condition to weigh the nodes for model prediction (Dorogush et al. 2018). In contrast to the XGBoost and CatBoost algorithms, the LightGBM algorithm builds the ensemble decision trees vertically or leaf-wise (Cover and Hart 1967, Ke et al. 2017, Microsoft 2021). The Random-Forest algorithm differs significantly from the other three algorithms. The Random-Forest algorithm builds a collection of trees to operate as a committee to produce an averaged prediction.

4.3 Alberta Transportation PMS Database

Alberta Transportation (AT) maintains an extensive PMS database that is utilized for pavement-related decisions, such as pavement design, maintenance, and rehabilitation. The dataset used in this study is a part of the PMS database and includes records from 2014 to 2020. Within the dataset, asphalt concrete Overlay pavement (OL) and Asphalt Concrete Pavement (ACP) are the most prevalent pavement categories, both of which are classified under the Granular Base Course (GBC) pavement type. The extracted pavement features/variables for this study included measurements' geographical location, pavement type, section length, soil base type, base's construction year, pavement thickness, last activity year, *AADT*, *ESAL*, transverse cracking, longitudinal wheel path cracking, rutting depth, other cracking, and *IRI*. The statistical tools t-test and p-values were used to analyze the pavement features/variables within the dataset using a 95% confidence interval to identify and select those with the greatest statistical correlation to *IRI*. The data for variables that passed the 95% confidence interval are presented in Appendix A.

The dataset was analyzed for limitations and errors. The errors and limitations were identified by using pavement design expertise. Some erroneous values were observed in the database, such as a recorded pavement thickness of 0. A critical limitation of the database was missing base soil types, which were identified with the assistance of the Alberta soil information viewer (Alberta Agriculture and Forestry 2020). For the purpose of this study, the Plasticity Index (*PI*) and the Percentage passing no. 200 sieve (P_{200}); were assigned to all the base soils using the Mechanistic-Empirical Design Guide (MEPDG) for new and rehabilitated pavement structures by the National Cooperative Highway Research Program (NCHRP) (ARA Inc. 2001b). Additionally, Alberta's current and historical weather stations database was utilized to compile climate data from the nearest weather stations (Alberta Agriculture and Forestry 2021b). Climate data from the weather stations include but are not limited to accumulated precipitation since the first *IRI* reading, annual precipitation, mean temperature, minimum temperature, maximum temperature, wind speed, and relative humidity. Correlations between the climate variables and *IRI* are

examined and assessed using t-tests and p-values with a 95% confidence interval. The t-tests and p-values were used to identify the variables with the greatest statistical correlation to *IRI*. As a result of examining the correlations between climate variables and *IRI*, the average precipitation variable was found to be the most suitable for model development.

Analysis of the dataset revealed several unusual values for the variables mentioned above. Outliers in the dataset were identified and removed using the outer-fence method as follows:

$$Lower\ Outer\ fence\ =\ Q_1\ -\ (3)\ IQR \tag{4.1}$$

$$Upper\ Outer\ fence\ =\ Q_3\ +\ (3)\ IQR \tag{4.2}$$

where Q_1 represents the 25th percentile, Q_3 represents the 75th percentile, and *IQR* represents the interquartile range (the difference between Q_3 and Q_1). After removing the outliers, the dataset for OL and ACP pavement types comprised of 1633 and 839 data points, respectively.

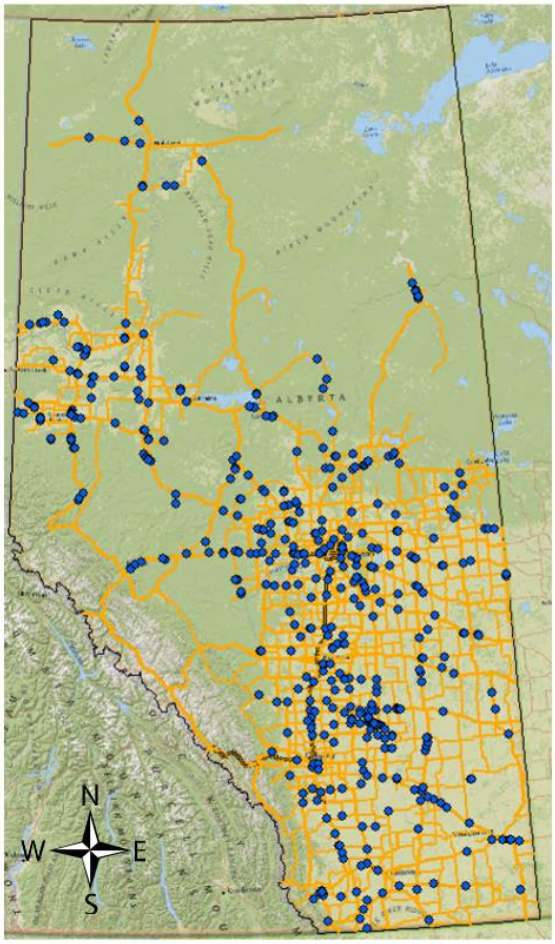

Figure 4.2 – Alberta's pavement measurement records geolocation in the dataset

The location of the pavement sections in the dataset is shown in Figure 4.2. It should be noted that blue dots do not depict the entire length of the road section but rather the start of the road section where measurements were taken. The review of this figure indicates that the dataset provided a good representation of the road sections for the province of Alberta in a spatial sense.

4.4 Methodology

The pavement features *PI*, P_{200}, *Age*, *ESAL*, pavement surface thickness, transverse cracking, rutting depth, and annual precipitation are used to develop the *IRI* models using various ML algorithms. The ML modelling was performed using Python through Jupyter Notebook while using the relevant Python libraries. The ML algorithms were imported into Python using the stand-alone ML algorithms and scikit-learn libraries. The utilized ML algorithms (including ANN) require hyperparameter tuning to produce the best possible model. The hyperparameter tuning is achieved through two strategies in this study: the grid-search method and the random-search method. The grid-search method is a technique for discovering an algorithm's optimal parameters by trial and error. The grid-search technique runs different combinations of parameters in a grid fashion while storing each iteration's score. Once all possible combinations are tested, grid-search outputs the best parameter combination along with its score. Random search techniques also work in a similar fashion to the grid-search method since various combinations are tested through trial and error to identify the best combination, but this search is run randomly rather than in the form of a grid. The random search technique provides a computationally less expensive solution to finding the optimal parameters.

The overall framework of the ML modelling used in this study is presented in Figure 4.3. The process begins with the collection of measurements that are then stored in the PMS database. After the dataset is extracted from the PMS database, it is subjected to pre-processing, including variable selection, removing outliers, and splitting the data randomly into training and test datasets (80:20 ratio). Following the pre-processing, the ML modelling process is initiated. The ML framework includes algorithm selection,

model training, model validation, and model tuning to come up with the final version of the ML regression model. The models are trained and hyperparameter-tuned using the training dataset. The hyperparameter tuning is optimized using the k-fold cross-validation technique to avoid overfitting.

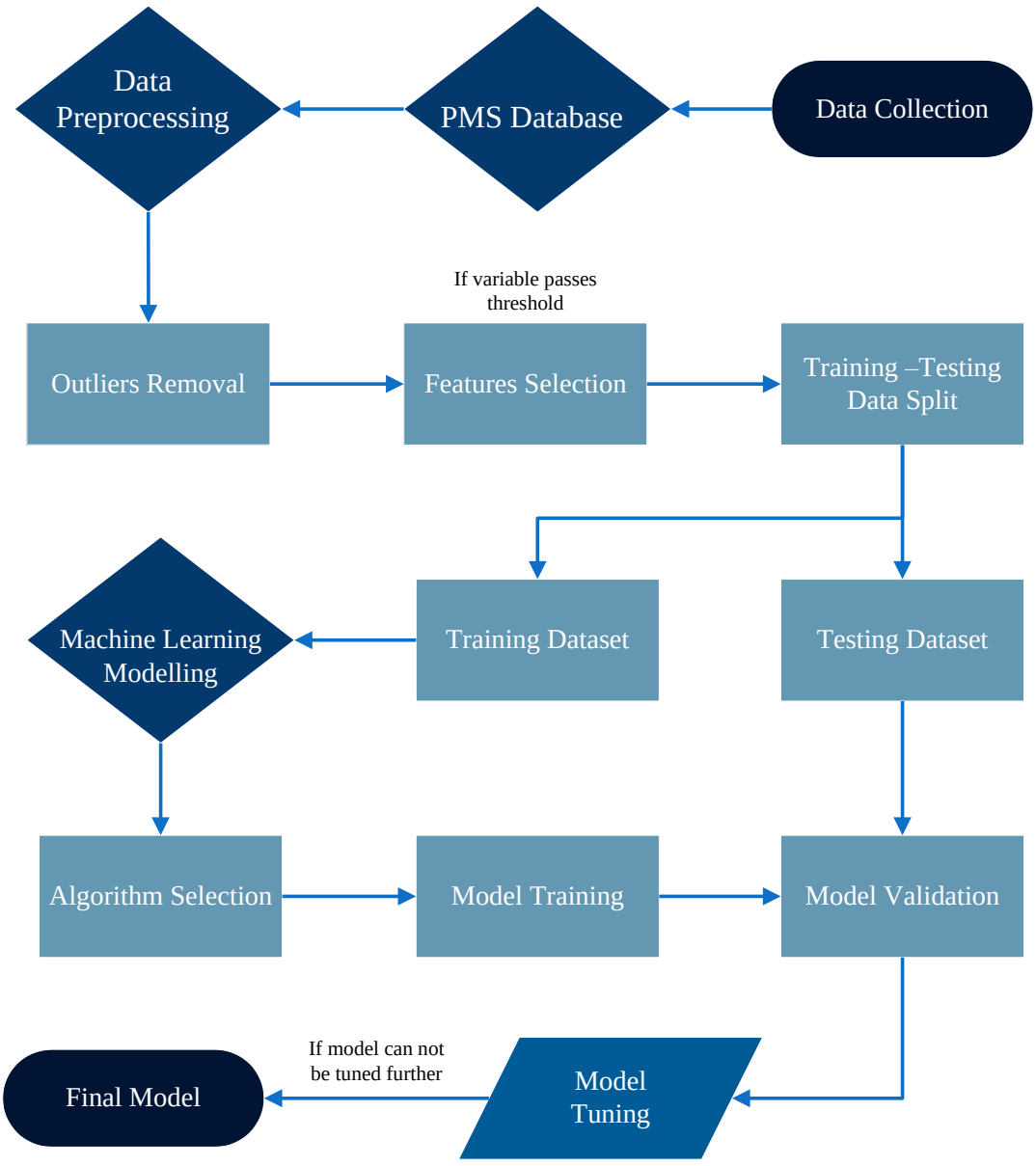

Figure 4.3 – Machine Learning modelling framework

Following the models' hyperparameter tuning, the developed models are compared to each other using statistical parameters, namely the coefficient of determination (R^2), Mean Absolute Error (*MAE*), Mean Square Error (*MSE*), and Root Mean Square Error (*RMSE*) based on the models' output. All the statistics mentioned above provide a measure of accuracy in one form or another for the models' *IRI* output and the actual *IRI* values in the dataset.

4.5 Results

The models are developed and examined as described in Section 4.4. The full/complex models are used to identify the ML algorithms that produce the best-fitting *IRI* models. The best-developed models were assessed using sensitivity analyses to examine the impact of each of the pavement variables on *IRI*. By understanding the impact of each of the pavement features in the developed models, lesser complex models can be developed using the most impactful pavement features. Less complex or reduced models can, in some cases, provide better fits for testing datasets. This is due to the fact that the more complex models are more prone to overfitting relative to reduced models (Hawkins 2004, Zhang 2014, Lever et al. 2016). Several studies suggest the need to examine the performance of reduced ML models (Hawkins 2004, Zhang 2014, Lever et al. 2016). The results of the full/complex ML *IRI* models are presented in Table 4.2. For the sake of brevity, the results of the full/complex ML *IRI* models are presented in Appendix C, including sensitivity analysis performed using the developed models.

The reduced ML models were developed using the five most impactful variables through the sensitivity analyses conducted for the more complex *IRI* models. The most impactful pavement features used in the development of the reduced models were the variables *RUT*, *ESAL*, *Age*, *PI*, and *surfthickness*. 4The reduced *IRI* models were developed using the same technique described for developing the more complex models. The reduced models were developed using XGBoost, CatBoost, Random-Forest, and LightGBM algorithms, which were identified as the best-performing algorithms for modelling *IRI*.

Table 4.2 – Full/Complex ML *IRI* models results

OL Models									**Variable impact ranking (1 = most impactful)**				
Algorithm	Test statistics				Train statistics								
	MAE	*RMSE*	*MSE*	R^2	*MAE*	*RMSE*	*MSE*	R^2	*Age*	*ESAL*	*PI*	*RUT*	*surfthickness*
XGBoost	0.127	0.176	0.031	0.873	0.019	0.029	0.001	0.996	1	2	3	4	5
CatBoost	0.124	0.177	0.031	0.871	0.042	0.056	0.003	0.985	1	3	4	2	5
Random-Forest	0.142	0.201	0.040	0.828	0.059	0.089	0.008	0.961	1	4	3	2	5
LightGBM	0.148	0.208	0.043	0.823	0.072	0.104	0.011	0.947	1	4	3	2	5
ANN	0.176	0.248	0.061	0.696	0.143	0.205	0.042	0.725	Sensitivity Analyses were not performed for these models				
Decision Trees	0.184	0.274	0.075	0.681	0.001	0.000	0.012	0.999					
Ridge	0.210	0.291	0.085	0.579	0.209	0.295	0.087	0.593					
Elastic-Net	0.211	0.292	0.085	0.578	0.209	0.295	0.087	0.593					
Lasso	0.215	0.294	0.087	0.570	0.213	0.299	0.089	0.583					
SVM	0.226	0.324	0.105	0.567	0.109	0.025	0.157	0.879					
KNN	0.276	0.378	0.143	0.412	0.205	0.287	0.082	0.595					
ACP Models									**Variable impact ranking (1 = most impactful)**				
Algorithm	Test statistics				Train statistics								
	MAE	*RMSE*	*MSE*	R^2	*MAE*	*RMSE*	*MSE*	R^2	*Age*	*ESAL*	*PI*	*RUT*	*surfthickness*
CatBoost	0.197	0.263	0.069	0.661	0.051	0.070	0.005	0.978	3	2	4	1	5
LightGBM	0.209	0.279	0.078	0.618	0.087	0.118	0.014	0.938	2	3	4	1	5
XGBoost	0.216	0.285	0.081	0.600	0.003	0.022	0.001	0.998	2	3	4	1	5
Random-Forest	0.216	0.299	0.089	0.563	0.002	0.023	0.001	0.998	2	3	4	1	5
ANN	0.242	0.337	0.114	0.501	0.186	0.246	0.060	0.724	Sensitivity Analyses were not performed for these models				
SVM	0.260	0.333	0.111	0.455	0.111	0.142	0.020	0.910					
KNN	0.263	0.121	0.348	0.404	0.138	0.200	0.040	0.820					
Decision Trees	0.283	0.364	0.132	0.350	0.154	0.208	0.043	0.784					
Lasso	0.321	0.423	0.179	0.281	0.295	0.377	0.142	0.325					
Ridge	0.321	0.423	0.179	0.281	0.293	0.375	0.141	0.333					
Elastic-Net	0.321	0.426	0.182	0.270	0.297	0.377	0.142	0.324					

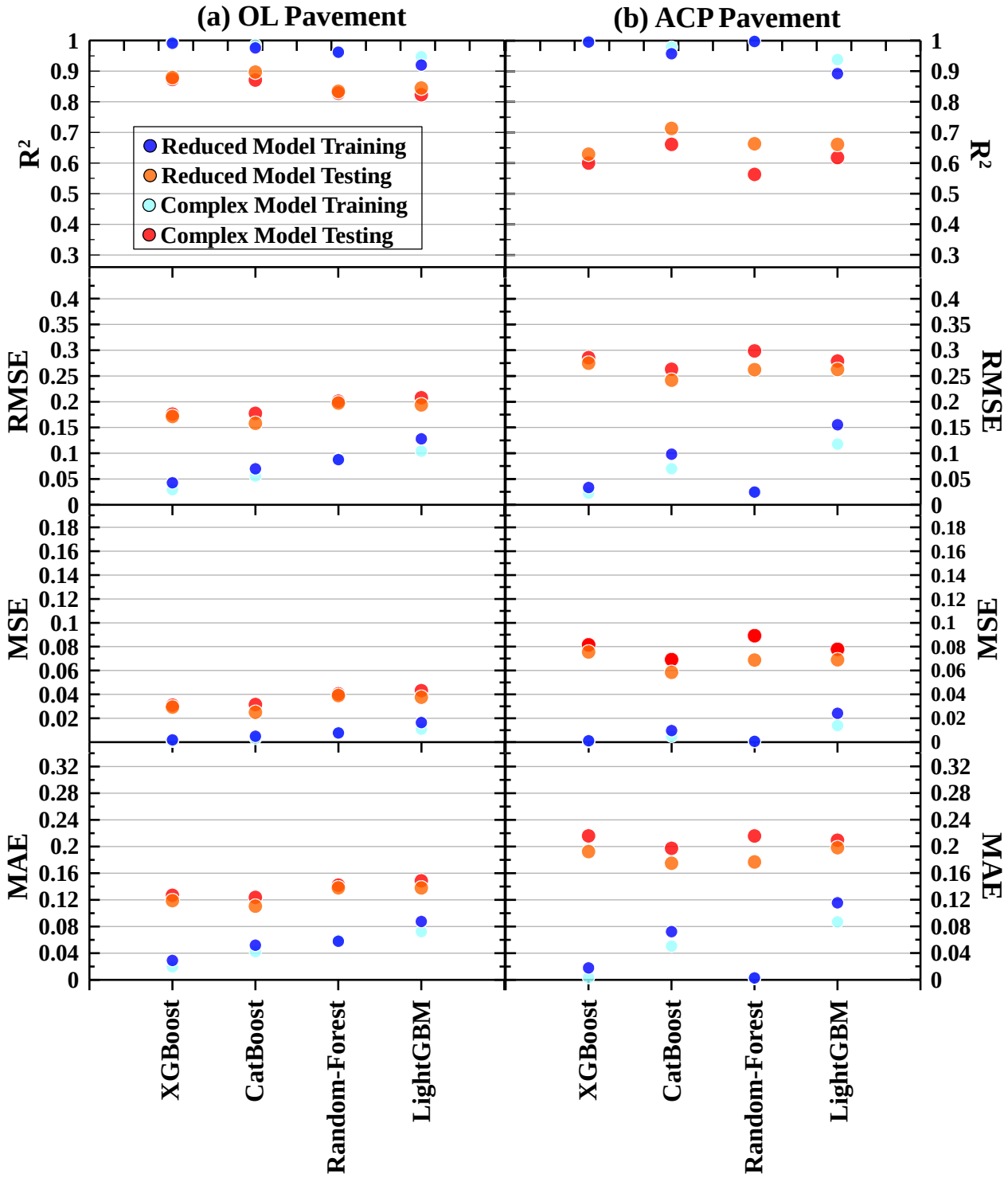

Figure 4.4 – Reduced and full/complex ML *IRI* models statistics for training and testing dataset

The statistical metrics for the reduced and full/complex models, including R^2, *RMSE*, *MSE*, and *MAE*, are presented in Figure 4.4 for both pavement types. The training statistics for the reduced *IRI* models provide a similar fit for both pavement types as compared to the more complex models. The reduced model using the XGBoost algorithm for OL pavements has an R^2 of 0.991 and an *RMSE* of 0.043, whereas the more complex ML *IRI* model using XGBoost for OL pavements has an R^2 of 0.996 and an *RMSE* of 0.029. Similarly, for ACP pavements, the model using XGBoost has an R^2 of 0.995 and an *RMSE* of 0.033, while the more complex model has an R^2 of 0.998 and an *RMSE* of 0.022. The *IRI* models' training dataset statistics for the reduced and complex models do not show a significant difference in the models' performance; however, the testing dataset statistics were more representative of the models' predictive capability than the training dataset statistics. Examining the testing dataset statistics, the reduced models appear to show better performance than the more complex models. This finding indicates that the more complex models were relatively overfitted compared to the reduced models. For instance, using the CatBoost algorithm, the reduced *IRI* model has an R^2 of 0.897 and an *RMSE* of 0.158 compared to the more complex model, which has an R^2 of 0.871 and an *RMSE* of 0.177 for OL pavements. Similar results were observed for ACP pavements for the CatBoost *IRI* models. The reduced CatBoost model has an R^2 of 0.713 and an *RMSE* of 0.242 compared to the more complex model, which has an R^2 of 0.661 and an *RMSE* of 0.263. The results of the reduced models indicate that the reduced ML *IRI* models have superior predicting capabilities than the more complex ML *IRI* models.

Additionally, the reduced models present acceptable results for both types of pavements for the purpose of this study. For a better representation of the models' fit, the actual vs predicted *IRI* and residual plots are created for both pavement types using the reduced and complex ML *IRI* models and are presented in Figure 4.5 and Figure 4.6, respectively. The actual vs predicted values and residual plots for the reduced *IRI* models are comparable in performance to the more complex *IRI* models.

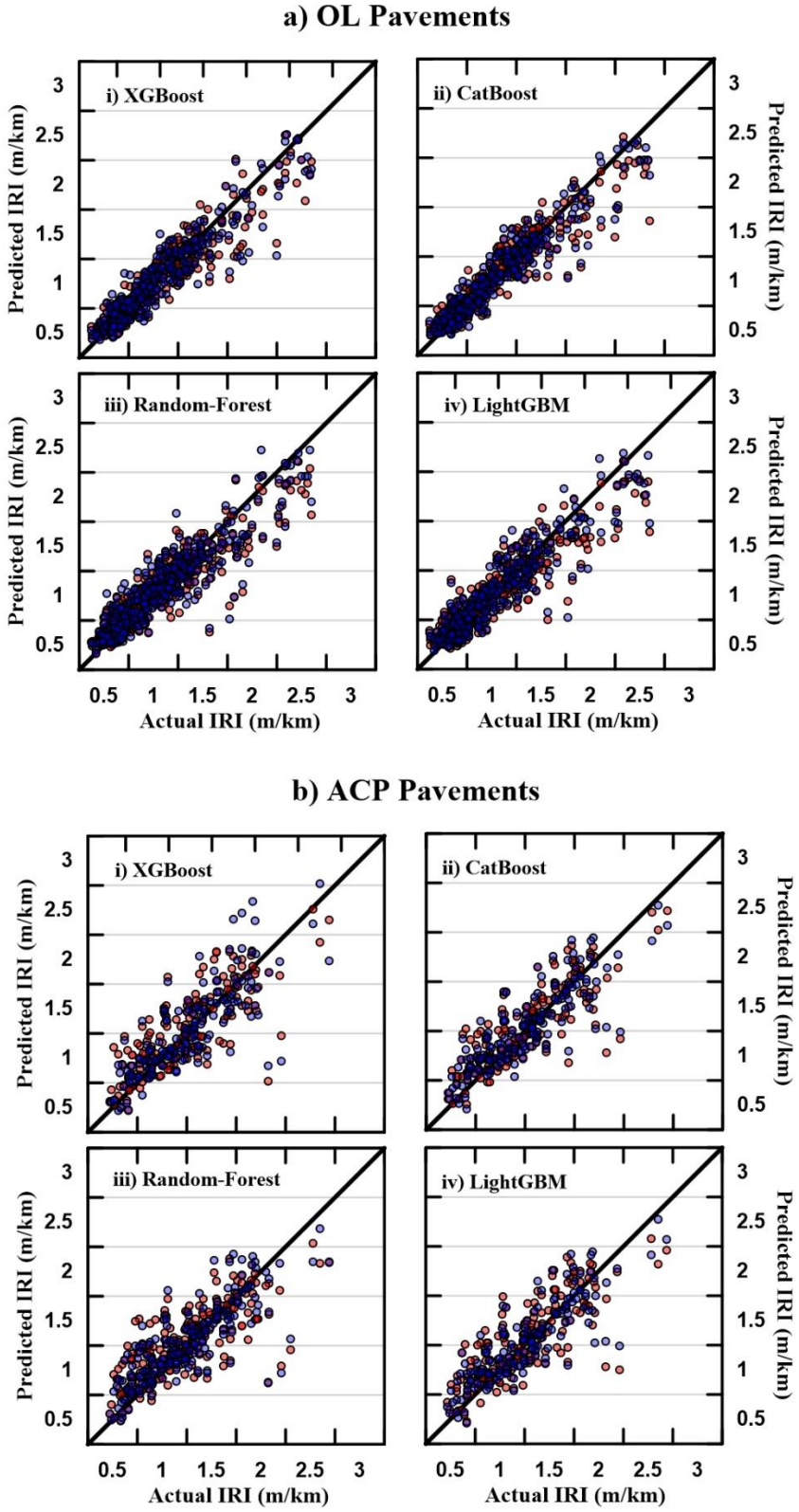

Figure 4.5 - Actual vs Predicted *IRI* values using the reduced (in blue) and full/complex (in red) ML *IRI* models for (a) OL pavement (b) ACP pavement

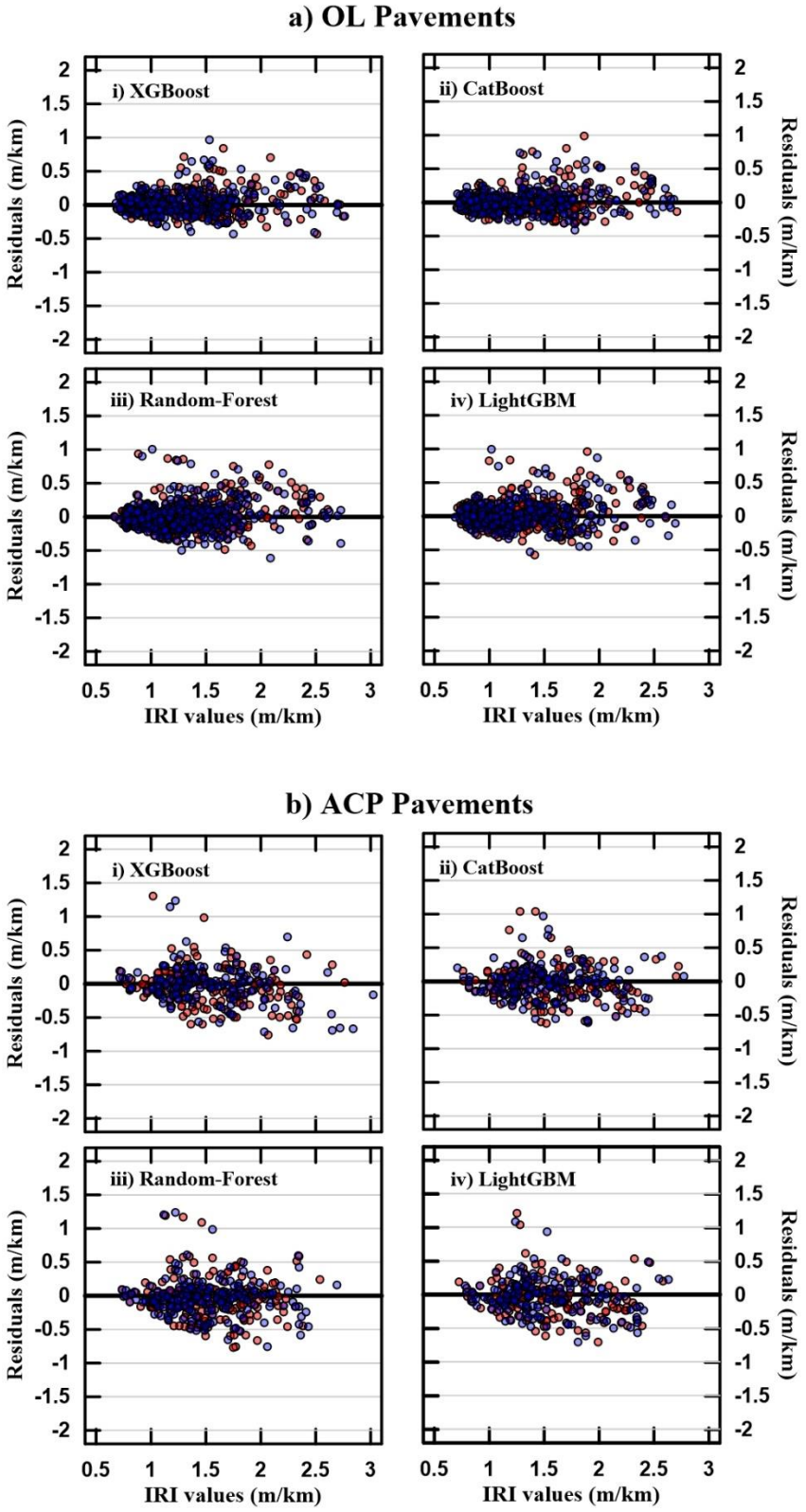

Figure 4.6 Residual plots using the reduced (in blue) and full/complex (in red) *IRI* models for (a) OL pavement (b) ACP pavement

From Figure 4.5, the actual vs predicted plots for both pavement types show adequate fits for the testing datasets. In Figure *4.6*, it can be observed that the residuals for both pavement types appear to be randomly distributed around the horizontal axis, thus suggesting the models are appropriate for the application of this study. The actual vs predicted and residual plots both provide additional evidence that the reduced models are a viable alternative to the full/complex *IRI* models. Additionally, Figure *4.5* and Figure *4.6* showcase that the ML models for OL are better fitting than the ACP models. For the residual plots, the ACP models appear to have a much larger residual spread than the OL models.

The ACP *IRI* model is compared to the MEPDG model. The outer fence ACP pavement testing dataset was used to test the MEPDG *IRI* model and yielded an R^2 of 0.23. In comparison, the ACP model yielded an R^2 of up to 0.71, resulting in around three folds increase in R^2 value. The ACP model developed in this study has far superior predicting capabilities than the MEPDG *IRI* model.

Comparing the regression analysis developed *IRI* models from Chapter 3 to the reduced ML models presented in this chapter, the use of ML in the development of the *IRI* models using the same training datasets resulted in far superior *IRI* models. The *IRI* model developed using regression analysis produced an R^2 of 0.59 and an *RMSE* of 0.294, while the ML-developed models produced an R^2 of up to 0.99 and an *RMSE* of up to 0.043 for OL pavements. Similarly, for ACP pavements, the regression analysis-developed model produced an R^2 of 0.33 and an *RMSE* of 0.384; meanwhile, the ML-developed model produced an R^2 of up to 0.99 and an *RMSE* of up to 0.033. The results show a significant increase in the model's predictive capabilities when using ML.

Sensitivity analyses were conducted for the reduced *IRI* models to examine each of the pavement feature's impact on the models' output. For the presented sensitivity analysis plots, the colour-coding legend is presented on the right side of the figure and presents the relative value of the independent variable, with blue being low values and magenta representing higher values. The y-axis has the independent variables presented in the order of impact on the models' output, with the most impactful variable being at the top. The x-axis has the quantified impact on the model's output due to the variation

in the corresponding independent variable from the y-axis while keeping all the other variables at their arithmetic mean. The analyses provide a visual representation of the features' impact around the mean line or the "0-impact" line.

The sensitivity analysis for the XGBoost, CatBoost, LightGBM, and Random-Forest *IRI* models are presented in the figures below. The pavement variables used in the reduced models are *RUT*, *ESAL*, *Age*, *PI*, and surfthickness, as mentioned earlier in the text. The first analysis is conducted for the XGBoost reduced models, as shown in Figure 4.7. The most impactful pavement features presented in Figure 4.7 for OL pavements significantly differ from the most impactful features for ACP pavements. The most impactful pavement features/variables of OL pavements are *Age*, *ESAL*, *PI*, *RUT*, and *surfthickness*, in that order. The most impactful pavement features/variables for ACP pavements are *RUT*, *ESAL*, *Age*, *surfthickness*, and *PI*.

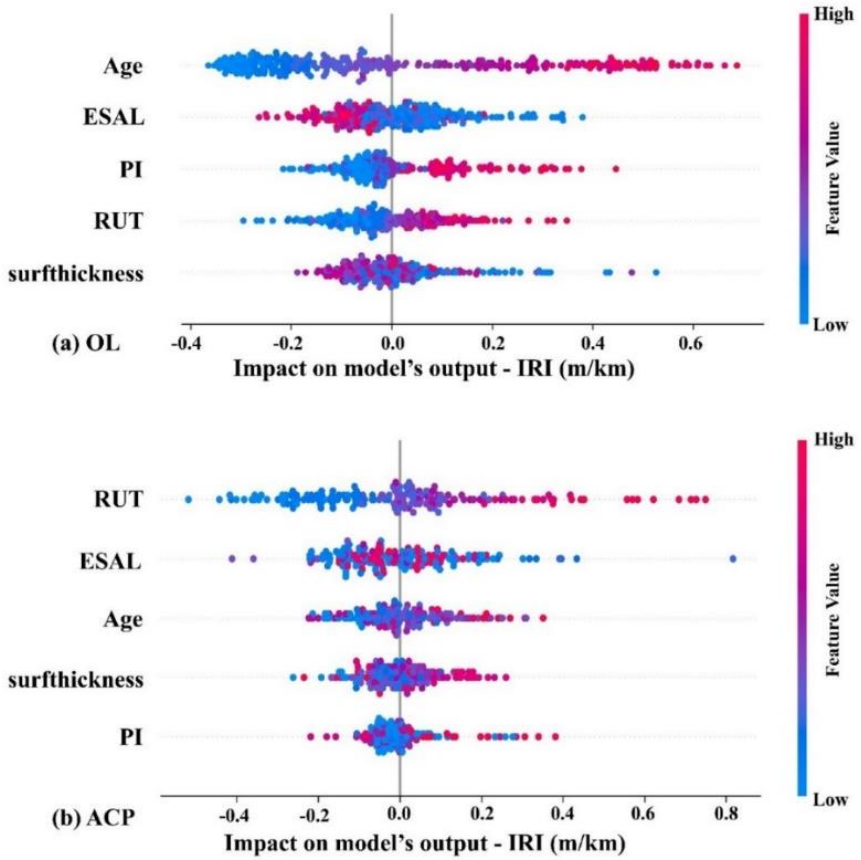

Figure 4.7 - XGBoost ML reduced *IRI* models' sensitivity analysis. a) OL pavements b) ACP pavement

The difference between the pavement features' impacts on the models' output indicates that the OL pavements are affected by different factors than the factors affecting ACP pavements. The *Age* of the pavement appears to be more impactful on OL pavements, possibly because OL pavements are rehabilitated ACP pavements. The traffic parameter *ESAL* appears to have a similar impact on both OL and ACP pavements. Interestingly, *ESAL* has an inverse impact on the OL models' output. This is indicative of the higher *ESAL* pavements being designed to deteriorate at a slower rate than lower *ESAL* pavements. Moreover, the variable *PI* appears to have a much more significant impact on OL pavements than on ACP pavements. This indicates that for OL pavements, the *PI* of the pavement's subgrade plays a more substantial role at later stages of the pavement's life cycle as OL pavements are essentially rehabilitated ACP pavements. The variable *RUT*, a type of pavement distress, impacts both OL and ACP pavements even though the variable has a more significant relative impact in the ACP model than in the OL model. The variable surfthickness has a low impact on both pavement types, as seen in the figure.

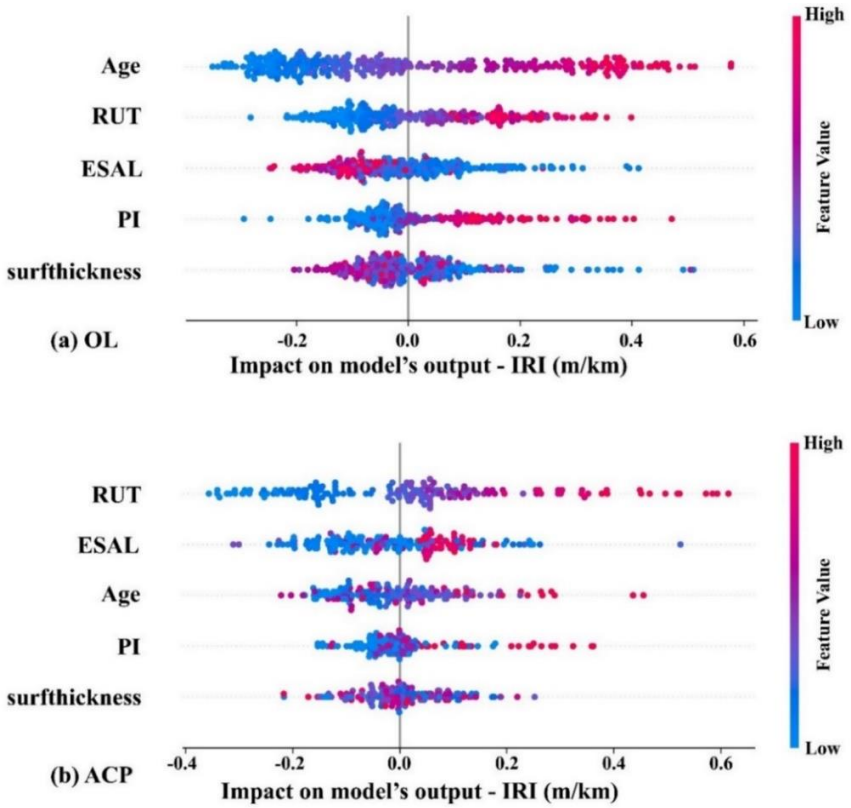

Figure 4.8 - CatBoost ML reduced *IRI* models sensitivity analysis. a) OL pavements b) ACP pavements

A sensitivity analysis similar to that conducted for the XGBoost *IRI* models is also conducted for CatBoost models and is presented in Figure 4.8. Comparing the CatBoost and XGBoost models' sensitivity analysis, the CatBoost models' variables impacts appear to differ more for OL pavements than ACP pavements. The variables *Age* and *surfthickness* have the same impact on OL pavements for both the CatBoost and XGBoost models. However, the variables *RUT*, *ESAL*, and *PI* relative impact appear to differ between the two models. This is due to the three variable's impact being relatively similar in magnitude on *IRI* and the correlation between the variables *RUT*, *ESAL*, and *PI*. In other words, larger traffic (*ESAL*) on higher *PI* pavements is likely to lead to more rutting (*RUT*), which would consequently lead to rougher pavements. In contrast, the CatBoost's sensitivity analysis for ACP pavements appears to be mostly in line with what was observed for the XGBoost model, except for the variables *surfthickness* and *PI*. The interchange in relative importance between *surfthickness* and *PI* suggests multicollinearity between the two variables.

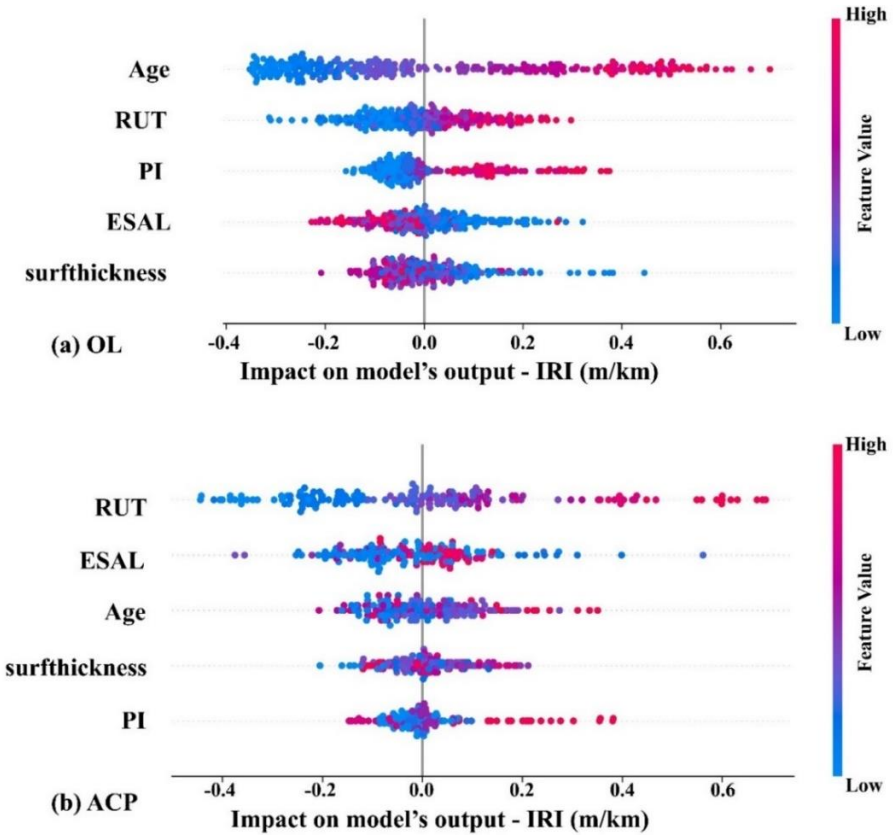

Figure 4.9 - LightGBM ML reduced *IRI* models sensitivity analysis. a) OL pavements b) ACP pavements

Figure 4.9 shows the sensitivity analysis conducted for the LightGBM models. The LightGBM model developed for OL pavements shows the variables *Age*, *RUT*, *PI*, *ESAL*, and *surfthickness* as the most impactful on the model's output, in that order. The sensitivity analysis results for the LightGBM model for OL pavements are close to that of CatBoost's OL model. The variables *PI* and *ESAL* are interchanged between the two models, suggesting a correlation between the variables. For ACP pavements, the most impactful variable on pavement roughness is *RUT*, followed by *ESAL*, *Age*, *surfthickness*, and *PI*. The sensitivity results for the LightGBM model are identical to that of the XGBoost model for ACP pavements.

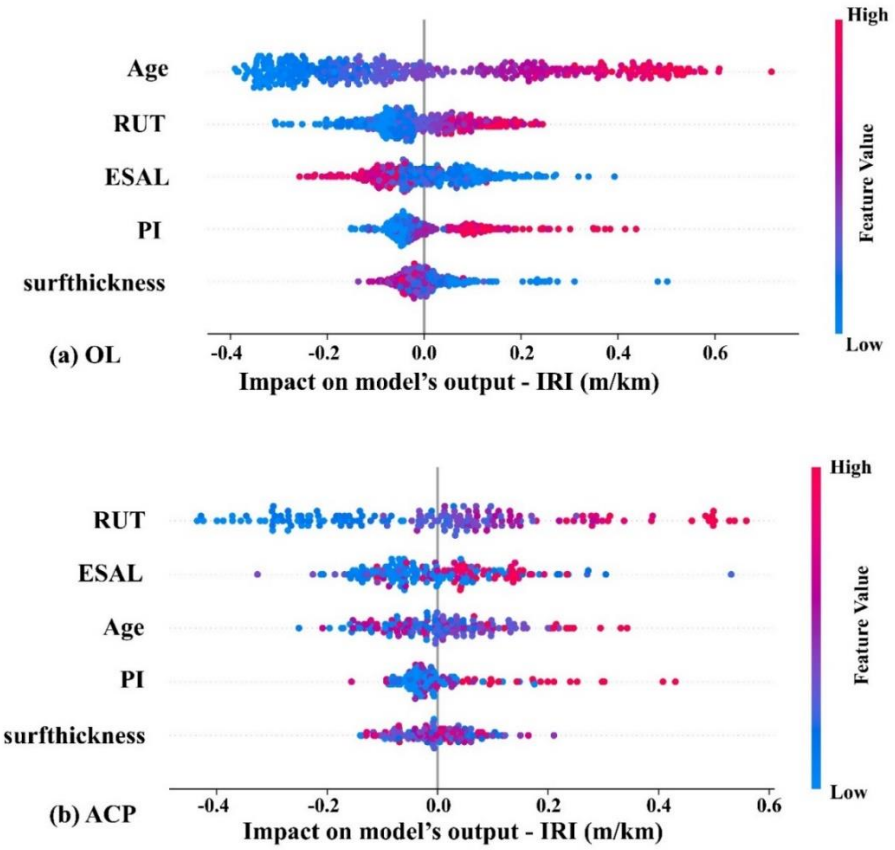

Figure 4.10 - Random-Forest ML reduced *IRI* models sensitivity analysis. a) OL pavements b) ACP pavement

A similar sensitivity analysis is conducted for the Random-Forest *IRI* models, presented in Figure 4.10. The analysis for OL pavements shows the variable *Age* as the most impactful on the model's output,

followed by *RUT*, *ESAL*, *PI* and *surfthickness*. For ACP pavements, the most impactful variables are *RUT*, *ESAL*, *Age*, *PI*, and *surfthickness*. The sensitivity analysis results for both pavement types using the Random-Forest *IRI* models are identical to that of the CatBoost *IRI* models' sensitivity analysis results. This finding supports the conclusions from the CatBoost *IRI* models' sensitivity analysis results.

The sensitivity analyses conducted for each of the reduced *IRI* models show that the models utilize the given variables effectively to predict the models' output. Generally, the sensitivity analyses for the reduced *IRI* models show that all the pavement features used to predict the models' output have a significant impact on the output. A handful of conclusions can be drawn by analyzing the models' sensitivity analyses for both pavement types. For instance, the pavement features *RUT*, *Age*, and *ESAL* significantly impact the models' output for ACP pavement type. Similarly, the pavement feature *Age* had a consistently significant impact on the models' output for OL pavements, while the pavement features *RUT* and *ESAL* had a significant impact for some but not for all the *IRI* models. Therefore, the pavement features in the *IRI* models for ACP pavement type have less correlation with other independent variables than in the reduced models for OL pavements. Nonetheless, the overall results of the reduced *IRI* models' sensitivity analyses provide acceptable results to use the *IRI* models as predictors of *IRI* for both pavement types.

4.6 Conclusions

This study developed Machine Learning (ML) models for predicting *IRI* using eleven different supervised machine learning algorithms while providing a meaningful comparison between them in terms of performance. The models were developed using the same datasets utilized in Chapter 3. Out of the examined algorithms, one can conclude that CatBoost, LightGBM, Random-Forest, and XGBoost are the best-performing ML algorithms in modelling *IRI*, all of which fall under ensemble methods. This indicates that ML algorithms that combine the prediction of several base estimators to improve the overall prediction fare well in performance. Historically, neural networks have been dominantly used for

developing *IRI* models; this study's findings conclude the need to examine other ML algorithms, such as ensemble methods.

Moreover, the models developed using ML provide *IRI* predictive capabilities far better than the use of existing models such as MEPDG. This is a significant improvement because of two reasons. First, the ML models use site-specific information for their development. Second, the ML algorithms do a significantly better job than regression analysis in the development of *IRI* models.

The performance of the ML models for both the asphalt concrete Overlay pavements (OL) and Asphalt Concrete Pavements (ACP) were superior to the regression analysis developed models. In addition, the OL models had better predictive capabilities than the ACP models, similar to what was observed for the regression analysis developed models. Although ML does a better job in the development of the *IRI* models, the importance of data should not be overlooked. The better performance of ML for OL pavements over ACP can be attributed to the difference in the size of the datasets.

Sensitivity analyses were conducted on the best-performing ML-developed models to identify the most impactful variables on the models' output. The sensitivity analyses concluded that the equivalent single axle load, the plasticity index, the age of the pavement since construction or since rehabilitation, the pavement's surface thickness, and the 80th percentile rut depth are the most impactful on the models' output for both pavement types.

Reduced models were developed using the best-performing ML algorithms and the variables with the most impact. The reduced models were found to provide an improvement in performance over their more complex counterparts. One can conclude that the reduced models are less prone to overfitting the training datasets and thus provide better-performing models than their more complex counterparts.

The reduced ML *IRI* models' predictive capabilities are further examined by comparing them to the MEPDG *IRI* model as a benchmark for ACP pavements. The reduced *IRI* model developed using ML

resulted in a three-fold increase in the coefficient of determination value. This study's findings conclude that the use of ML should be examined in the development of general models, such as the MEPDG model, as they are heavily reliant on regression analysis.

Further research examining other datasets should be conducted using the algorithms in this study to further confirm its findings. Additional research can be done to explore other ML algorithms and investigate the effectiveness of the best-performing machine algorithms highlighted in this study.

Chapter 5 Investigating the effect of climate on pavement roughness for different subgrade soil types using LTPP sites

Abstract

Pavements are an essential component of our modern infrastructure. It is in the transportation agencies' best interests to understand the impact of various environmental factors on pavements. Climate change is expected to cause around 1.5°C of warming globally in the next two decades, with regions of the contiguous United States expected to experience increases in precipitation and temperature. The existing literature on pavement roughness does not examine the effects of climate change on pavement roughness for different subgrade pavement types, especially in the absence of traffic. The purpose of this study is to identify the impact of climate on different subgrade types of pavements over time using regression analysis and machine learning. In this study, the pavement roughness is measured using the International Roughness Index (IRI) for both the center lane (*CLIRI*) and the mean of the wheel-path lanes (*MIRI*). The changes in *CLIRI* and *MIRI* are investigated over time, as changes in *CLIRI* are expected to be primarily influenced by climate, as the only traffic in the center lane is due to lane changes. The Long-Term Pavement Performance (LTPP) data, specifically the Specific Pavement Studies-1 (SPS-1) experiment data provided by the Federal Highway Administration (FHWA), is used in this study. The study investigated the effects of the climate parameters Freezing Index (*FI*) and precipitation (*PPT*) for different pavement subgrades soil types on pavement roughness over time. The subgrade soil types were captured by the parameters Plasticity Index (*PI*) and percent passing No. 200 sieve (P_{200}). This study found that the climate and subgrade soil type parameters can better describe the changes in *CLIRI* than in *MIRI*, with an R^2 value of up to 0.86 for the training dataset and up to 0.54 for the testing dataset. This study also found that soils with higher plasticity are more susceptible to freeze-thaw damage than soils with lower plasticity. Furthermore, finer-grained subgrade soils were found to have a greater change in pavement roughness due to climate factors than coarse-grained subgrade pavements. With climate change expected to increase precipitation in some contiguous United States

regions, fine-grained subgrade pavements in those areas are expected to deteriorate faster than coarse-grained pavements.

5.1 Introduction

Roads are an integral part of our transportation infrastructure and connect communities together. A road pavement deteriorates over time due to various factors causing pavement distress. The amount of distress and its effect is dependent on the properties of various pavement materials. A popular method of quantifying the state of deterioration of pavements is by measuring the pavement roughness. Pavement roughness affects fuel consumption, repair costs, vehicle maintenance costs, greenhouse gas emissions, and vehicle efficiency (Robbins and Tran 2016). Pavement roughness is often evaluated using a standardized index, namely, the International Roughness Index (IRI). The IRI system was developed by the World Bank in 1986 (Sayers et al. 1986). The changes in IRI values reflect changes in pavement roughness resulting from the effects of several factors, such as climate and pavement properties, over time. The IRI has a proportional relationship with pavement roughness and starts with the pavement roughness at construction (Sayers 1995b).

One of the most extensive pavement performance programs is the Long-Term Pavement Performance (LTPP) program, supported by the Federal Highway Administration (FHWA) to collect and analyze pavement data across the United States and Canada. The LTPP program was initiated in the 1980s by the National Research Council (NRC) of the United States. Various agencies were instrumental in helping the LTPP program achieve its potential. These agencies include the American Association of State Highway and Transportation Officials (AASHTO), the Canadian Strategic Highway Research Program (C-SHRP), highway agencies in the States (USA) and Provinces (Canada), the NRC through the Transportation Research Board (TRB) and Strategic Highway Research Program (SHRP), the FHWA and the International Highway Committee. Currently, the LTPP database (also known as the Pavement Performance Database (PPDB)) is the largest and most detailed pavement performance database, with more than 280 million records of pavement data and counting (FHWA 2015, 2022).

Several studies have attempted to model the IRI with various pavements parameters using the LTPP database, including but not limited to climate, age, subgrade properties, traffic, and pavement distresses (Mactutis et al. 2000, George 2000, Perera and Kohn 2001, Choi et al. 2004, Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2005, Chou and Pellinen 2005b, Kutay 2007, Puccinelli and Jackson 2007, Haider and Chatti 2009, Kargah-Ostadi et al. 2010, Nassiri et al. 2013, Khattak et al. 2014, Ziari et al. 2016a, 2016b, Mazari and Rodriguez 2016, Gong et al. 2018b, Hossain et al. 2019, ARA Inc. 2020, Abdelaziz et al. 2020). Several of these studies considered the effect of the subgrade soil properties, such as the Plasticity Index (*PI)* and the percent passing No. 200 sieve (P_{200}) on the IRI (Perera and Kohn 2001, Choi et al. 2004, Kutay 2007, Haider and Chatti 2009, Kargah-Ostadi et al. 2010, Nassiri et al. 2013, Khattak et al. 2014, ARA Inc. 2020). According to the Mechanistic-Empirical Pavement Design Guide (MEPDG), the subgrade materials' *PI* and P_{200} contribute to pavement roughness and have a significant impact on the rate of roughness progression (ARA Inc. 2020). Moreover, Kutay (2007) found that the *PI* of the subgrade soil plays a significant role, with higher *PI* value soils resulting in more pavement distress leading to a greater change in pavement roughness. Similarly, Haider and Chatti (2009) found that the pavements built on fine-grained soil subgrades generally showed a more significant increase in roughness than the pavements built on coarse-grained soil subgrades, especially in the wet-freeze regions. Additionally, Haider and Chatti (2009) also found that pavements located in wet climates had a more significant increase in pavement roughness than pavements located in dry climates. Furthermore, Perera and Kohn (2001) found that pavements' subgrades showed a strong correlation between the pavement roughness and the subgrade's material properties. Similarly, Nassiri et al. (2013) found that pavement roughness is linearly correlated with the subgrade's P_{200}.

The LTPP program consists of two main types of research: General Pavement Studies (GPS) and Specific Pavement Studies (SPS). The SPS experiments have various focuses; this study focuses on the SPS-1 experiment in particular. The SPS-1 experiment was created to investigate the climate effects on flexible pavement performance constructed on different subgrade types (Perera and Al-Rawashdeh 2017). This

research effort uses the SPS-1 sections data by the LTPP to identify climate and subgrade soil parameters contributing to pavement deterioration. A map of the States in which the SPS-1 sections are located is presented in Figure 5.1. The SPS-1 sections data provided by the FHWA included both Mean IRI (*MIRI*) data and Center-Lane IRI (*CLIRI*) data. The *MIRI* measurements are the mean of the IRI measurements from the left and right wheel paths, whereas the *CLIRI* measurements are the IRI measurements from the center of the lane.

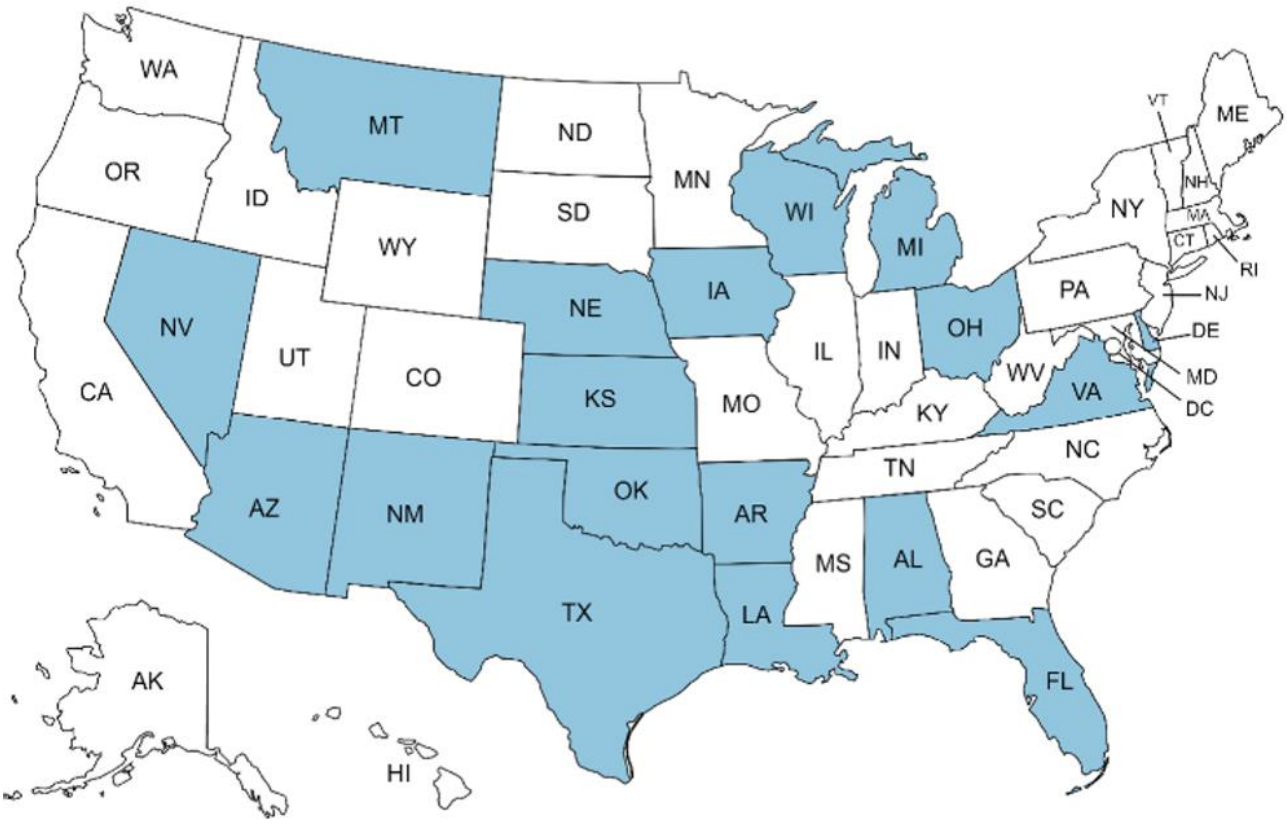

Figure 5.1 – The states in which the SPS-1 sections were used in this study's analysis

The purpose of this study is to evaluate the changes in both *MIRI* and *CLIRI* over time due to environmental and subgrade parameters using both regression analysis and Machine Learning (ML). The changes in *CLIRI* are expected to be mainly due to climate, as the only traffic the center lane receives is

when vehicles change lanes. Therefore, modelling the change in *CLIRI* and its comparison with *MIRI* provides a unique opportunity of investigating the effect of climate on pavement roughness in the absence of traffic.

The *IRI* modelling is also carried out using ML to identify whether ML can provide more informative models as compared to the traditional method of modelling pavement roughness using regression analysis. Additionally, it is also examined if ML can identify correlations between the variables where regression analysis has its limitations. The ML-developed models using the dataset are then used for a climate change impact study.

Moreover, this study also provides an insight into the effect of climate change by understanding the contribution of climate to pavement roughness for different pavement subgrade soils. Climate change can be observed on both a regional and a global scale. It is projected that Earth is on track to reach around 1.5°C of warming within the next two decades, dependent on global emissions (Masson-Delmotte, et al. 2021). On a regional scale, aside from an increase in temperature, the eastern United States region is expected to see an increase in precipitation intensity throughout the year. Additionally, the western, west-Midwest, and the west south-central United States regions are expected to face an increase in the frequency of extreme precipitation events, with the west-Midwest and the west south-central United States regions facing an increase in the average precipitation during the winter season (IPCC 2021). This study employs IRI models developed using ML to provide an insight into the effect of the regional climate change on pavement roughness over time for different subgrade soil-type pavements at two different locations within the contiguous United States.

5.2 LTPP data

The data used in this analysis was provided by FHWA and is a part of the LTPP program under the SPS-1 project. The SPS-1 project data used in this study is presented in Appendix D. The collected data is from various road sections across the United States of America. The dataset provided by the FHWA included *CLIRI*, *MIRI*, climate, and subgrade soil types for each of the road sections. The climate

information was computed from weather stations data and included the average annual precipitation (*PPT*) and the average annual freezing index (*FI*). The climate parameters *PPT* and *FI* are computed as the averages over the same period in which the changes in *CLIRI* and *MIRI* occur. Table 5.1 provides a summary of the climate parameters *PPT* and *FI* values, as well as the years over which the data was averaged (Perera and Al-Rawashdeh 2017).

Table 5.1 – Average annual precipitation and freezing index values

Project Location	**Years Over Which Data Were Averaged**	**Average *PPT* (Inches)**	**Average *FI* (°F Days/Year)**
Alabama	1993 – 2005	52	16
Arizona	1993 – 2006	7	0
Arkansas	1994 – 2007	47	115
Delaware	1996 – 2006	48	148
Florida	1995 – 2012	56	0
Iowa	1993 – 2001	41	713
Kansas	1993 – 2001	27	394
Louisiana	1997 – 2012	58	2
Michigan	1995 – 2012	32	823
Montana	1998 – 2010	14	940
Nebraska	1995 – 2000	27	671
Nevada	1995 – 2009	10	351
New Mexico	1995 – 2006	10	11
Ohio	1995 – 2012	42	578
Oklahoma	1997 – 2011	32	99
Texas	1997 – 2007	24	0
Virginia	1993 – 2010	45	92
Wisconsin	1997 – 2008	32	1613

States are classified in the provided LTPP dataset based on whether the majority of the sections in a state can be classified as fine-grained or coarse-grained subgrade soil types. The subgrade soil information in the dataset included the *PI* and P_{200} for each road section. Since the classification of a state as either the fine-grained or the coarse-grained subgrade soil type is based on the majority of the sections in that state, some pavement sections in a coarse-grained state will have *PI* values greater than 0. The *PI* and P_{200} data were not available for several road sections, regardless of whether they belonged in the fine-grained or

coarse-grained states. Thus, the median P_{200} values and the average *PI* values for each state are computed to assign the missing sections' data. The computed P_{200} median considers all the available values for each state, while the computed *PI* average only considers *PI* values greater than 0. The median P_{200} and the average *PI* values are used for road sections where the subgrade soil data was not available. The use of the median P_{200} and average *PI* values is based on how they are defined in the LTPP. Table 5.2 provides a summary of the median P_{200} values, the average *PI* values, and the subgrade soil type classification used for each state as defined by the FHWA. Note that no subgrade soil information was available for the LTPP sites in Delaware; however, the subbase soil's *PI* and P_{200} values were available. Thus, the subbase soil's *PI* and P_{200} values were used instead. Furthermore, Table 5.2 contains some sections where the average of *PI* values greater than 0 is denoted as "NA." The "NA" annotation is used in states where none of the subgrades' *PI* values are greater than 0.

Table 5.2 – Median P_{200} values, Average *PI* values, and subgrade soil type classification for each state

Project Location	**Layer Type**	**Median P_{200} (%) values**	**Average *PI* (%) values**	**Material Type**
Alabama	Subgrade	66	16.7	Fine-grained
Arizona	Subgrade	17	6.1	Coarse-grained
Arkansas	Subgrade	18	NA	Coarse-grained
Delaware	Subgrade	—	5.5	—
	Subbase	13	NA	Coarse-grained
Florida	Subgrade	14	NA	Coarse-grained
Iowa	Subgrade	93	27.7	Fine-grained
Kansas	Subgrade	36	6.8	Coarse-grained
Louisiana	Subgrade	94	20	Fine-grained
Michigan	Subgrade	67	10.3	Fine-grained
Montana	Subgrade	22	NA	Coarse-grained
Nebraska	Subgrade	96	20.3	Fine-grained
Nevada	Subgrade	45	11	Coarse-grained
New Mexico	Subgrade	68	30	Fine-grained
Ohio	Subgrade	71	15	Fine-grained
Oklahoma	Subgrade	44	18.4	Coarse-grained
Texas	Subgrade	8	NA	Coarse-grained
Virginia	Subgrade	42	8.3	Coarse-grained
Wisconsin	Subgrade	10	NA	Coarse-grained

The provided dataset included *CLIRI*, *MIRI*, and the *Age* of the pavement since the first recording of data. The Delta *CLIRI* and Delta *MIRI* values were computed by taking the final record minus the first record. Data exceeding 1.5 times the interquartile range (the 75th percentile minus the 25th percentile) were considered outliers but are still considered in the analysis as per the FHWA definition in the study examining the SPS-1 experiment (Perera and Al-Rawashdeh 2017).

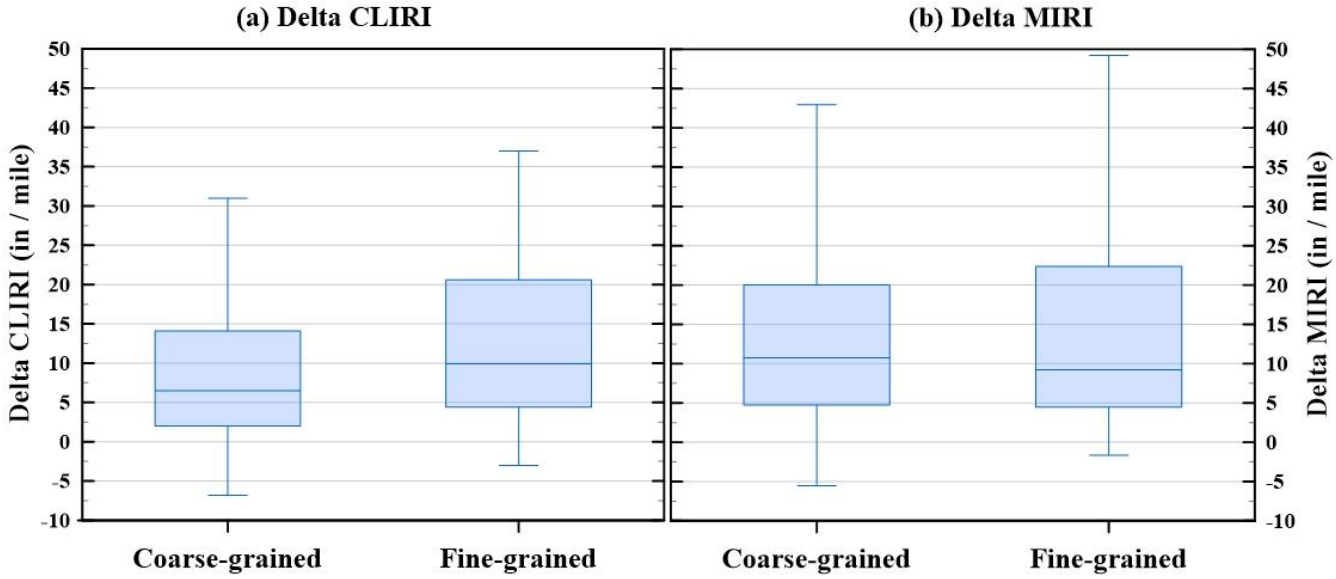

Figure 5.2 presents the values of Delta *CLIRI* and Delta *MIRI* in the form of box and whisker plots. This figure shows that coarse-grained subgrade pavements had higher recorded roughness change values for the wheel path compared to the center lane. For the fine-grained subgrade pavements, the upper range of the wheel path values was higher than that of the center lane, suggesting the change in pavement roughness is greater for the wheel paths. Additionally, the lower ranges of fine-grained subgrade pavements for both the wheel path and center lane were greater than that of the coarse-grained subgrade pavements. Moreover, the interquartile ranges of fine-grained subgrade pavements for both the center lane and wheel path were greater than that of the coarse-grained subgrade pavements.

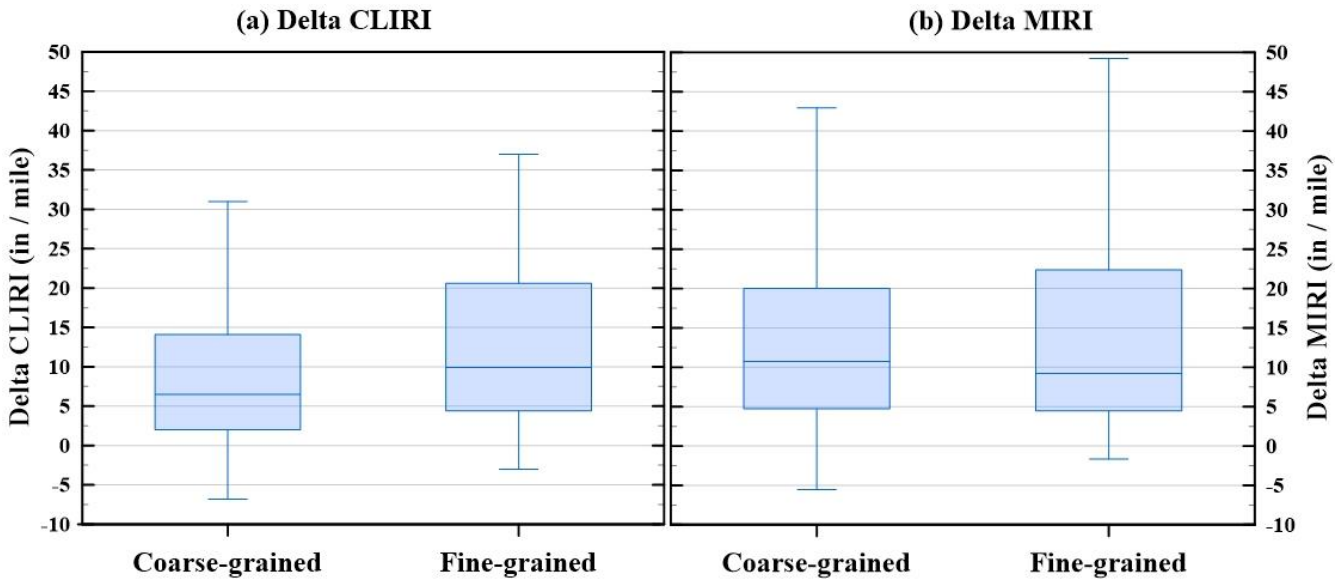

Figure 5.2 - (a) Delta *CLIRI* box and whisker plot and (b) Delta *MIRI* box and whisker plot for coarse-grained and fine-grained subgrade pavements

Similar to the Delta *CLIRI* and Delta *MIRI*, the *Age* is calculated by subtracting the time of the final record from the time of the first record. The range of *Age* values is presented in a box and whisker plot format in Figure 5.3. The figure shows that, on average, fine-grained subgrade pavements were younger than coarse-grained subgrade pavements. This could possibly be due to the fact that the fine-grained subgrade pavements need rehabilitation more frequently than their coarse-grained subgrade pavements counterparts.

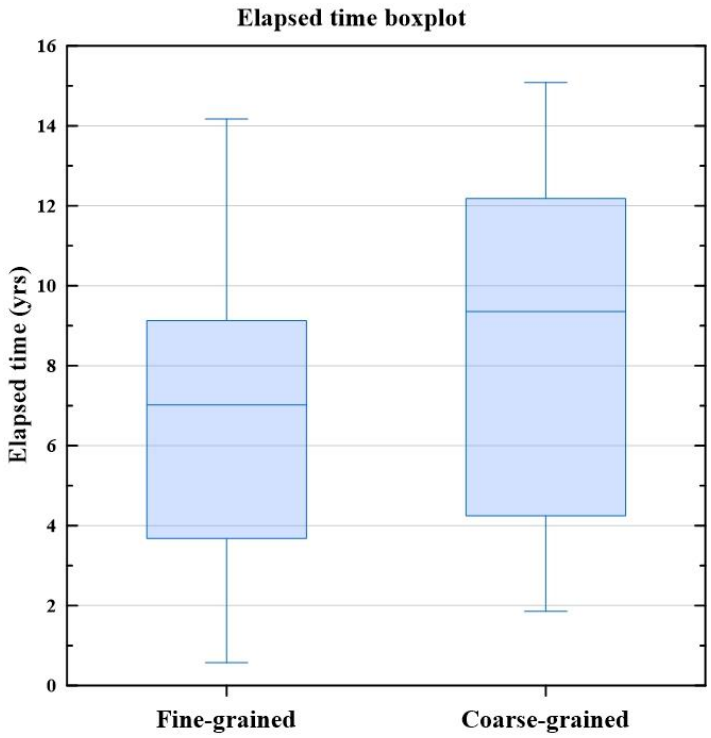

Figure 5.3 - *Age* box and whisker plot for fine-grained and coarse-grained subgrade pavements

5.3 Regression Analysis

The initial analyses of the data are performed using regression analysis. Regression analysis is one of the most commonly used mathematical methods for investigating the relationships between variables, in this case, climate, subgrade soil type, *Age* and pavement roughness (*CLIRI* and *MIRI*). Multiple linear regression models for Delta *CLIRI* and Delta *MIRI* are developed to investigate how well the variations in Delta *CLIRI* and Delta *MIRI* can be explained by climate, subgrade soil properties, and *Age*. The models correlate Delta *CLIRI* and *MIRI* to the independent variables *FI*, *PPT*, *PI*, and P_{200}. Three regression analysis models are developed for both Delta *CLIRI* and *MIRI*: the fine-grained, the coarse-grained, and considering both the fine-grained and coarse-grained soil types models. The model considering both the fine- and coarse-grained soil types is labelled as the "all-subgrades" model in this study. The ability to explain the variation in a model's output can be assessed by calculating the coefficient of determination (R^2). The coefficient R^2 ranges from 0 to 1, with 0 indicating that the model

fails to accurately predict an output given the input variables and 1 indicating a perfect fit to the data (Draper 1998, Mooi and Sarstedt 2011). Additionally, the Root Mean Square Error (*RMSE*) is used to quantify the difference between the predicted model output and the corresponding measured values. Regression analysis can also be used to determine the strength of the relationship between an independent variable like *PPT* and a dependent variable like Delta *CLIRI*. The strength of the correlation between two variables can be quantified by the variables' t-values and p-values. The regression model development mentioned above is done using R through RStudio. R is a software language developed for statistical computing and graphics that provides various statistical and graphical techniques. RStudio is an integrated development environment for R (R Core Team 2020).

5.4 Machine Learning Modelling

Another tool that is gaining popularity in examining and modelling data is ML. As a subset of AI, ML focuses on learning from a given dataset by attempting to mirror advanced human skills and decision-making. The ML process begins with training a model with a portion of the data (training dataset) using one or various ML algorithms. This is followed by testing the models with a testing dataset (the portion of the data that was not used in training the ML model). This study builds four different ML models for each subgrade soil type for Delta *CLIRI* and Delta *MIRI* using four supervised ML algorithms. These algorithms are XGBoost, CatBoost, Random-Forest, and LightGBM algorithms. The process of developing the ML models is described in Figure 5.4 in the form of a flow chart. The first step is splitting the dataset into training and testing datasets. The training dataset included 65% of the dataset selected randomly, while the testing dataset included the remaining 35%. The training dataset is then used to train the ML model using the ML algorithm of choice. The model is then tested using the testing dataset. The ML model is then hyperparameter-tuned iteratively until the best possible model is produced. The ML models are assessed and compared to one another using R^2 and *RMSE*. The developed ML models are also presented using plots such as actual vs predicted and residual values plots.

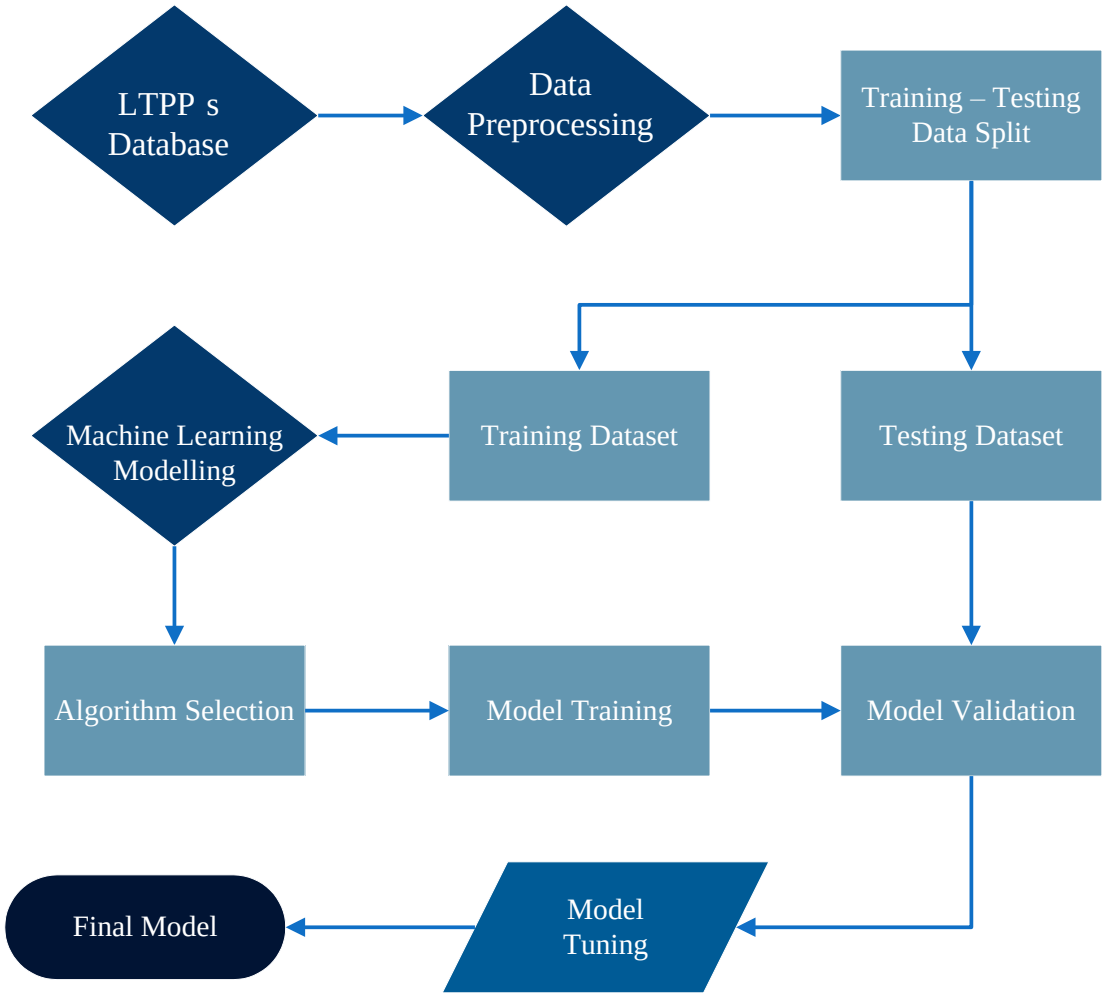

Figure 5.4 - Descriptive flowchart of the ML modelling performed in this study

5.5 Results

The regression analysis models are developed with a confidence interval of 95% or higher; thus, the independent variables' p-value must equate to or be less than 0.05 (Sarstedt and Mooi 2011, Schumacker and Tomek 2013). In contrast, the absolute t-value must be equal to or greater than 1.96 (Sarstedt and Mooi 2011, Schumacker and Tomek 2013). Negative t-values also indicate a negative correlation between the independent and the target variable. Table 5.3 displays the correlation between the dependent variable Delta *CLIRI* and each independent variable using their t-values and p-values.

Table 5.3 – Delta *CLIRI* regression analysis variables correlations

Subgrade type	**Fine**		**Coarse**		**All-subgrades**	
Correlating variable	**t-value**	**p-value**	**t-value**	**p-value**	**t-value**	**p-value**
Age	3.08	0.00290	4.56	1.14E-05	5.82	2.19E-08
FI	-3.07	0.00297	0.501	0.617	1.38	0.169
PPT	-4.42	3.33E-05	-2.39	0.0182	-2.97	0.00329
PI	-1.64	0.105	-3.58	0.000477	-0.144	0.886
P_{200}	2.39	0.0193	1.18	0.239	2.15	0.0328

Table 5.3 shows that based on the t-values and p-values for the fine subgrade model, the most correlated independent variable is *PPT*, followed by *Age*, *FI*, P_{200}, and *PI*. The significance of *PI* in the fine subgrade model is lower than the 95% confidence interval threshold using regression analysis. This could be due to the fact that the variable *PI* is correlated to other independent variables in the model, such as P_{200}, as suggested in the MEPDG, thus lowering the correlation metrics (ARA Inc. 2020). When the relationship between *PI* and Delta *CLIRI* is examined without the other variables, *PI* passes the confidence interval threshold with a p-value of 0.0243. For the coarse subgrade model, the most correlated independent variable is *Age*, followed by the variables *PI*, *PPT*, P_{200}, and *FI*. The variables P_{200} and *FI* in the model show a correlation lower than what is required by the 95% confidence interval. Coarse subgrade soils have no *PI* in definition; however, states with a majority of coarse-grained subgrade pavements are designated as "coarse-grained" subgrade states in the LTPP's database. Pavement sections within a "coarse-grained" subgrade state are all labelled as "coarse-grained" even if they were fine-grained subgrades. This practice of labelling states based on the majority of the pavement subgrade types causes variables such as *PI* to show a correlation in the coarse-grained model, even though coarse-grained soils have no plasticity. For the all-subgrades model, the dependent variable is most correlated with the independent variable *Age*, followed by *PPT*, P_{200}, *FI*, and *PI*. The variables *PI* and *FI* in the all-subgrades model show a correlation lower than required for the 95% confidence interval based on their t-values and p-values. The correlation between the dependent variable Delta *MIRI* and each of the independent variables is shown in Table 5.4 using both t-values and p-values.

Table 5.4 – Delta *MIRI* regression analysis variables correlations

Subgrade type	**Fine**		**Coarse**		**All-subgrades**	
Correlating variable	**t-value**	**p-value**	**t-value**	**p-value**	**t-value**	**p-value**
Age	4.58	1.87E-05	2.60	0.0105	2.99	0.00317
FI	1.05	0.297	0.0360	0.971	1.53	0.127
PPT	-1.79	0.0779	-2.37	0.0191	-2.39	0.0177
PI	-1.41	0.164	-1.45	0.149	0.195	0.845
P_{200}	-1.41	0.367	0.715	0.476	-0.187	0.852

Using the aforementioned t-value and p-value corresponding to the 95% confidence interval, each of the independent variable's correlation with the dependent variable Delta *MIRI* is examined. For the fine subgrade Delta *MIRI* model, the independent variable with the highest correlation to Delta *MIRI* is *Age*, followed by *PPT*, *PI*, P_{200}, and *FI*. However, with the exception of *Age*, all the independent variables in the model fail to meet the 95% confidence interval threshold. For the coarse subgrade Delta *MIRI* model, the most correlated independent variable is *Age*, followed by *PPT*, *PI*, P_{200}, and *FI*. Apart from the independent variables *Age* and *PPT*, all the other variables fail to meet the 95% confidence interval threshold for the coarse subgrade Delta *MIRI* model. For the all-subgrades Delta *MIRI* model, the most correlated independent is *Age*, followed by *PPT*, *FI*, *PI*, and P_{200}. Similar to what was observed for the coarse subgrade Delta *MIRI* model, the independent variables *FI*, *PI*, and P_{200} fail to meet the 95% confidence interval threshold.

The results from Table 5.3 and Table 5.4 indicate that the independent variables *FI*, *PPT*, *PI*, and P_{200} show better correlations for the Delta *CLIRI* variable than the Delta *MIRI* variable. The results indicate that the wheel path IRI is only correlated to *Age*. The Wheel path IRI data shows little correlation to climate and subgrade type, implying that the variables associated with these are not able to describe IRI. Meanwhile, the center lane IRI directly correlates to climate and subgrade variables. Additionally, the inclusion of Delta *MIRI* data might not be beneficial as it reduces the possibility of getting any information on the *CLIRI*, as Delta *CLIRI* is more correlated to other independent variables than Delta *MIRI*. The results indicate that Delta *CLIRI* and *MIRI* should be considered separately. Moreover, the results for the fine-grained, coarse-grained, and all-subgrades models for both Delta *CLIRI* and *MIRI* show variations in correlations with the predictive variables and should be examined separately.

Regression Analysis is used to develop Delta *CLIRI* and *MIRI* models for the fine-grained, coarse-grained, and all-subgrades models. The general equation for each of these models is presented below in equation 5.1.

$$Delta\,Y = C_1 * Age + C_2 * FI + C_3 * PPT + C_4 * PI + C_5 * P_{200} + Intercept \quad (5.1)$$

Where C_1, C_2, C_3, C_4, *and* C_5 are the coefficients for the predictive variables, *Y* is the target variable (*MIRI* and *CLIRI*), and *Intercept* is the intercept of the equation representing the target variable when all the other variables are equal to zero. The coefficients for the regression analysis developed models are presented in the table below.

Table 5.5 – Regression analysis models' coefficients

Variable / Coefficient	**Delta *CLIRI* coefficients**			**Delta *MIRI* coefficients**		
	Fine	**Coarse**	**All-subgrades**	**Fine**	**Coarse**	**All-subgrades**
Age	1.4000	0.8875	0.8188	3.1522	1.3961	0.9167
FI	-0.0099	0.0008	0.0019	0.0051	0.0002	0.0047
PPT	-0.3984	-0.1075	-0.1186	-0.2432	-0.2950	-0.2081
PI	-0.3436	-0.5517	-0.0161	-0.4438	-0.6177	0.0477
P_{200}	0.2375	0.0743	0.0785	0.1361	0.1244	-0.0149
Intercept	9.6743	3.4939	1.9748	-1.2966	13.0846	13.4942

The models developed for Delta *CLIRI* and Delta *MIRI* using regression analysis for all three subgrade types are assessed using R^2 and *RMSE*. Table 5.6 outlines the R^2 and *RMSE* values for all three subgrade types for the Delta *CLIRI* and Delta *MIRI* regression analysis models.

Table 5.6 – Regression analysis model statistics

Model output / Subgrade type	***R^2***			***RMSE***		
	Fine	**Coarse**	**All**	**Fine**	**Coarse**	**All**
Delta CLIRI	0.31	0.27	0.19	8.04	6.70	7.76
Delta MIRI	0.28	0.09	0.07	12.14	18.51	16.93

The Delta *CLIRI* regression analysis models show better performance than the Delta *MIRI* regression analysis models, as shown in Table 5.6. Overall, the models' statistics show that the models have no meaningful correlation metrics. The fine subgrade Delta *CLIRI* model appears to have a slightly higher R^2 of 0.31, while the Delta *MIRI* model has an R^2 of 0.28. The Delta *CLIRI* models appear to perform

significantly better than the Delta *MIRI* model developed using regression analysis for both the coarse subgrade and all-subgrades pavements. However, the regression analysis developed models for both Delta *CLIRI* and Delta *MIRI* do not present acceptable R^2 and *RMSE* values to estimate Delta *CLIRI* and Delta *MIRI* values accurately. Figure 5.5 visually represents the models' accuracy through actual vs predicted values plots. Figure 5.5 further proves that the independent variables can better describe Delta *CLIRI* and Delta *MIRI* for all subgrade types; however, the figure shows that both models do not provide sufficient model accuracy.

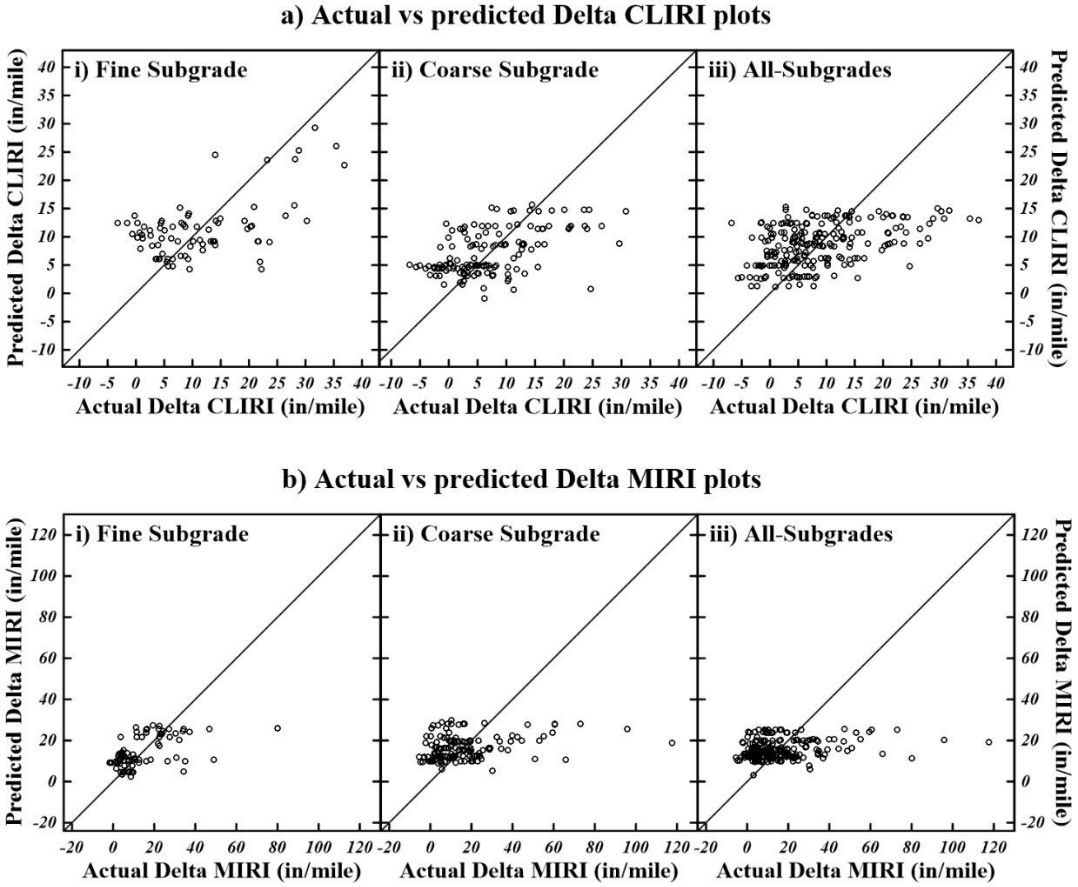

Figure 5.5 - Regression Analysis (a) actual vs predicted Delta *CLIRI* values plots and (b) actual vs predicted Delta *MIRI* values plots

The actual vs predicted values plots for both Delta *CLIRI* and Delta *MIRI* regression analysis models are used to create residual plots where the difference between the actual and predicted values is taken and plotted against the predicted values. The residual plots for both regression analysis models are presented in Figure 5.6.

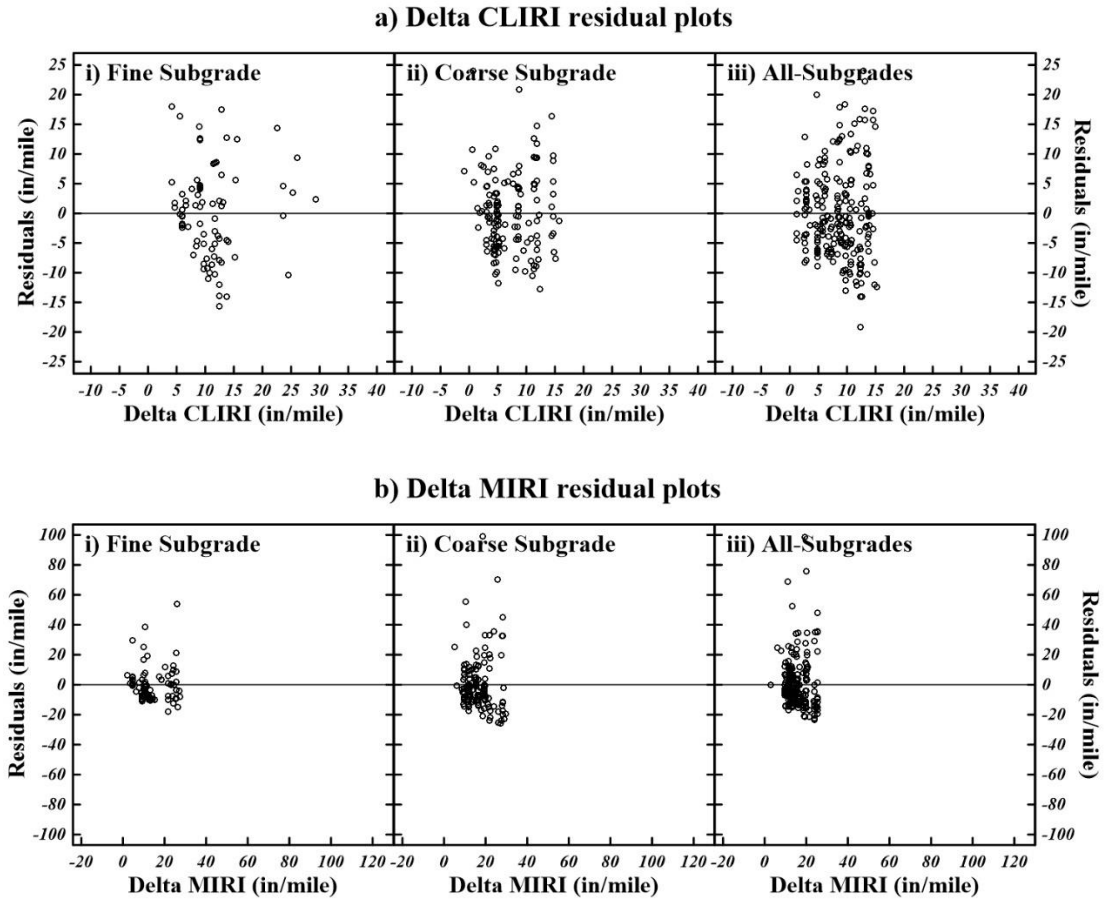

Figure 5.6 - Regression Analysis (a) Residual Delta *CLIRI* values plots and (b) Residual Delta *MIRI* values plots

The Delta *CLIRI* models' residual plots show an X-axis unbalanced pattern for the fine-subgrade model, a horn pattern for the coarse-subgrade model, and both a horn and an X-axis unbalanced pattern for the all-subgrades model. The horn-shaped residual pattern starts with residuals that are close together and spread more widely as the x-axis values increase (Sweet and Grace-Martin 2010). Similarly, the residual plots for Delta *MIRI* plots show a combination of an X-axis unbalanced pattern and horn pattern that appears to be worse than that observed in the Delta *CLIRI* residual plots. The horn-shaped and unbalanced residual patterns all signify that the Delta *MIRI* models' assumptions are inappropriate.

As presented above, regression analysis modelling does not provide adequate models; thus, ML is used to examine whether it can provide adequate models and insight into the correlations between the

independent and target variables. Using ML, models for both Delta *CLIRI* and *MIRI* are developed for each subgrade type. The developed ML models are evaluated and compared using the aforementioned R^2 and *RMSE*. Figure 5.7 presents the training and testing datasets statistics for the developed Delta *CLIRI* ML models. The training dataset statistics provide an overview of the models' output accuracy to the measured values used in training the models. The training dataset statistics do not represent how well the models perform to the data not used in training the model; thus, the testing dataset statistics are used to assess the models' accuracy.

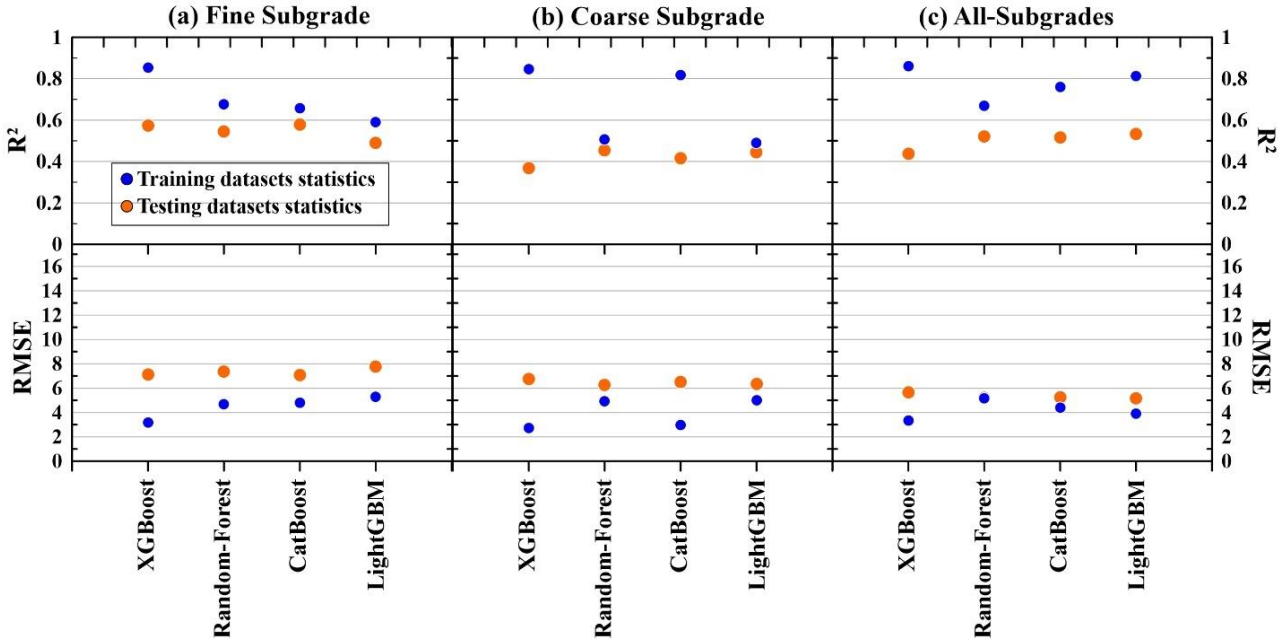

Figure 5.7 - Delta *CLIRI* ML models statistics summary for (a) Fine Subgrade (b) Coarse Subgrade (c) All-Subgrades

The Delta *CLIRI* ML models developed using XGBoost and CatBoost algorithms had the best model testing dataset statistics with an R^2 of 0.57 and 0.58, respectively, for the fine subgrade pavements. The models for coarse subgrade pavements had similar testing dataset statistics to one another. However, the training dataset statistics were significantly higher for the models developed using the XGBoost and CatBoost algorithms. Similarly, the models developed for all-subgrades had similar testing dataset results to one another apart from the XGBoost model. Additionally, the model developed using the Random-Forest algorithm showed lower model statistics for the training dataset compared to the other models.

Overall the best-performing ML algorithms for the Delta *CLIRI* models are the CatBoost and LightGBM models for all-subgrades. For the coarse subgrade type models, the overall best-performing algorithms are the XGBoost and CatBoost algorithms. Similarly, the fine subgrade models using the XGBoost and CatBoost algorithms showed the best results considering the training and testing datasets statistics.

Comparing the Delta *CLIRI* models developed using regression analysis and ML, it is clear that the ML models provide a large performance improvement. When comparing the training datasets statistics, the ML Delta *CLIRI* models provide an increase in the R^2 of up to 175%, 214%, and 353% for the fine, coarse, and all-subgrades, respectively. When comparing the *RMSE* metric, Delta *CLIRI* ML models reduce the error metric by up to 61%, 59%, and 57%, respectively.

Delta *CLIRI* models for all three subgrade types using the four different ML algorithms are used to create actual vs predicted values plots and are presented in Figure 5.8. This figure confirms that the Delta *CLIRI* models developed using XGBoost and CatBoost algorithms provide better performance models compared to the models developed using the Random-Forest and LightGBM algorithms. Figure 5.8 also shows that the model for all-subgrades developed using the LightGBM algorithm had a model fit similar to that of the XGBoost and CatBoost. Overall, the actual vs predicted value plots of the ML-developed models appear to be better than the models developed through regression analysis. This confirms the findings from the statistical metrics showing a large increase in the ML models' performance over the regression analysis models.

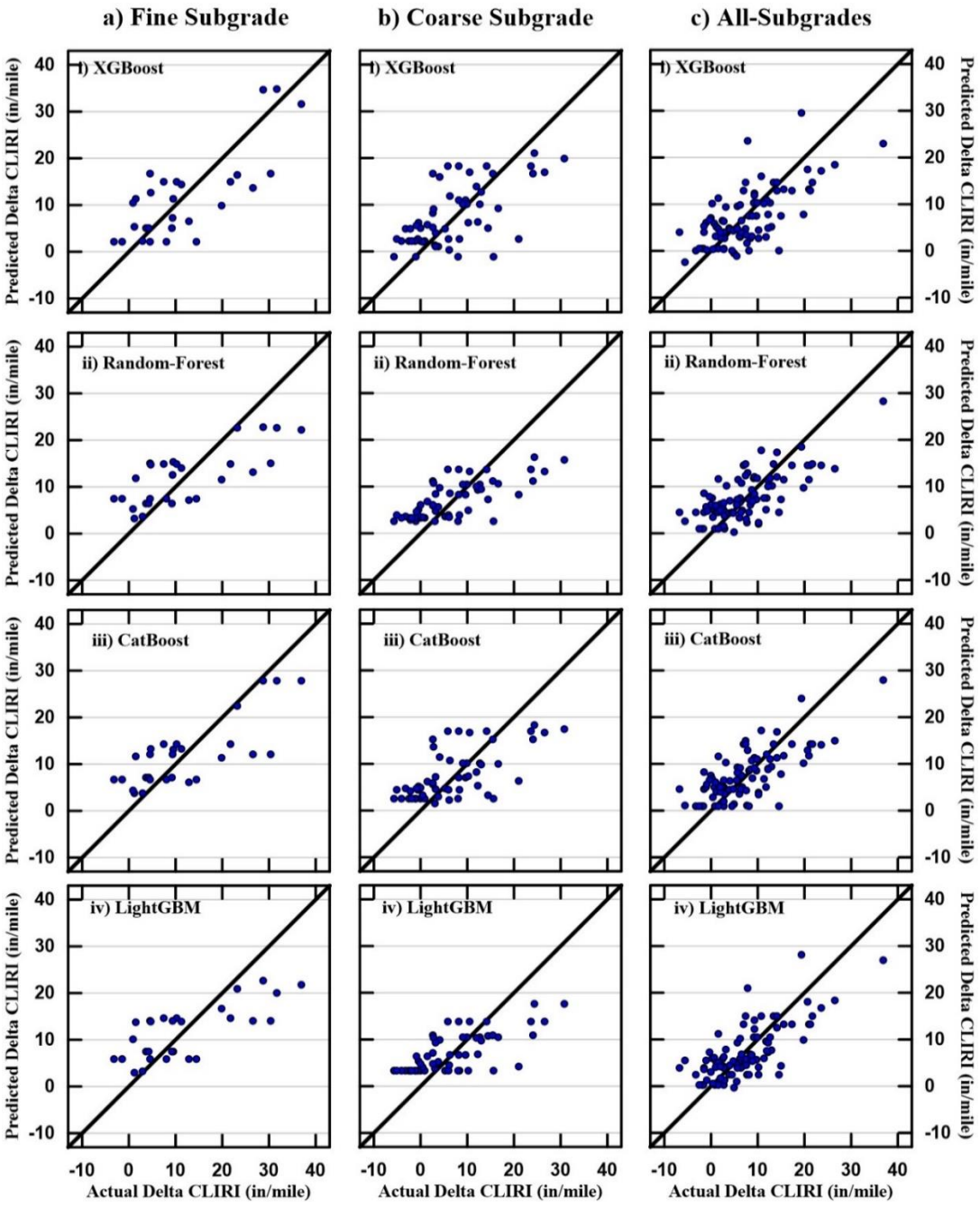

Figure 5.8 - Delta *CLIRI* ML models actual vs predicted Delta *CLIRI* values plots for (a) Fine Subgrade (b) Coarse Subgrade (c) All-Subgrades

The residual plots are plotted as the residual values against the predicted values and are presented in Figure 5.9 for the Delta *CLIRI* ML models.

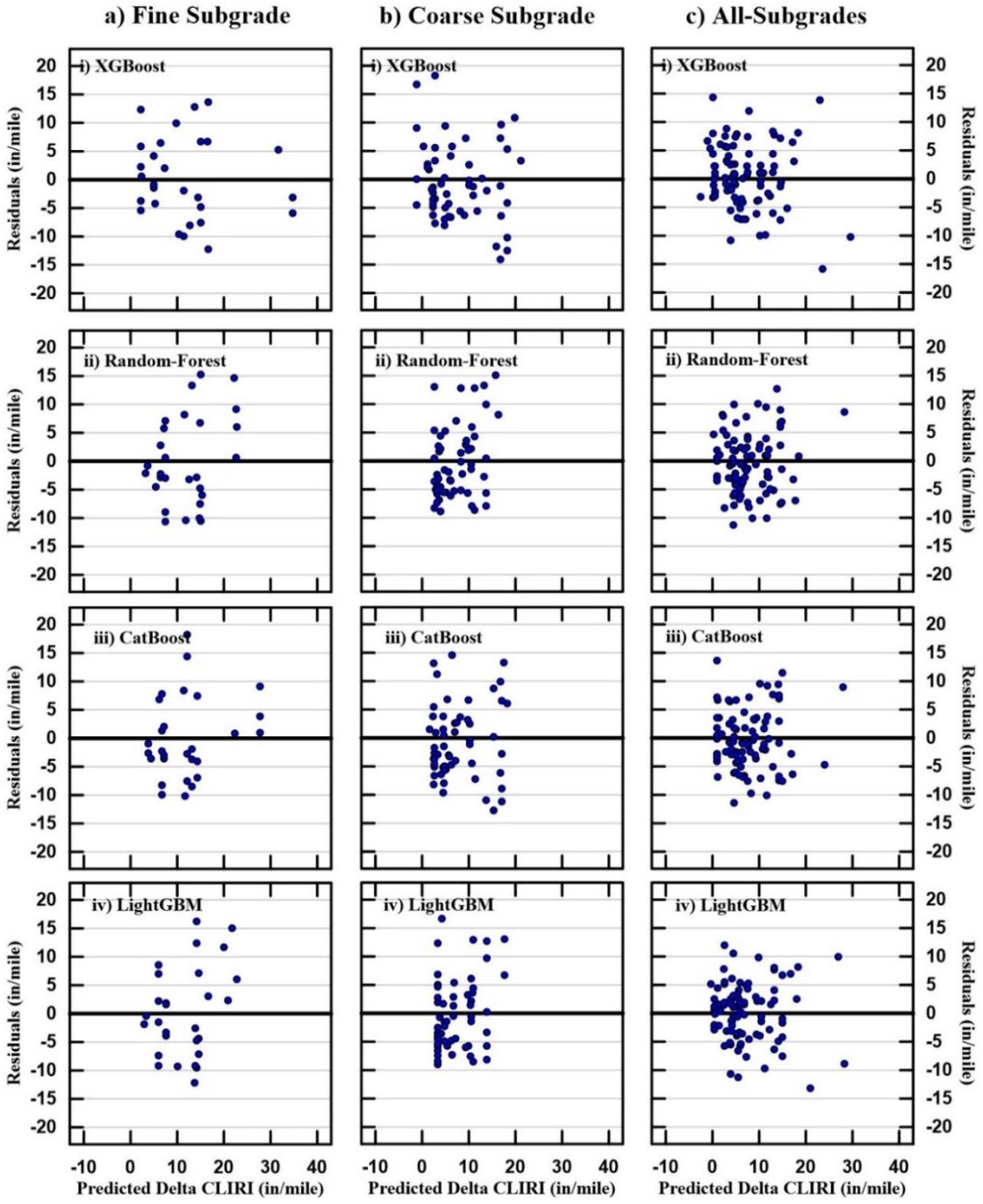

Figure 5.9 - Delta *CLIRI* ML models Residual Delta *CLIRI* values plots for (a) Fine Subgrade (b) Coarse Subgrade (c) All-Subgrades

Model residuals can be either positive or negative as they represent the difference between the observed and predicted values. From Figure 5.9, the fine subgrade models' residual plots show randomly distributed residuals for the models developed using the XGBoost and CatBoost algorithms. The residual

plots of the fine subgrade models developed using the Random-Forest and LightGBM algorithms revealed less random residual patterns compared to their counterparts. The coarse subgrade models' residual plots developed using the XGBoost algorithm presents a random residual pattern. In contrast, the models created using the CatBoost, Random-Forest, and LightGBM algorithms presented an X-axis unbalanced residual pattern. This finding indicates that the XGBoost model for the coarse-grained subgrade has appropriate model assumptions for the purpose of this study. Additionally, the all-subgrades residual plots support using the XGBoost and CatBoost models over the other two models. Overall, the models developed using ML through the XGBoost and CatBoost algorithms appear to be superior to the models developed using the two other algorithms when considering the models' R^2, *RMSE*, measured vs predicted values and residual plots; Thus, the models can be used in understanding the relationship between the independent variables *Age*, *FI*, *PPT*, *PI*, and P_{200} and the dependent variable Delta *CLIRI*.

The Delta *MIRI* models are also assessed using R^2 and *RMSE* statistics for training and testing datasets. Figure 5.10 presents the R^2 and *RMSE* statistics for the training and testing dataset for all the subgrade types models.

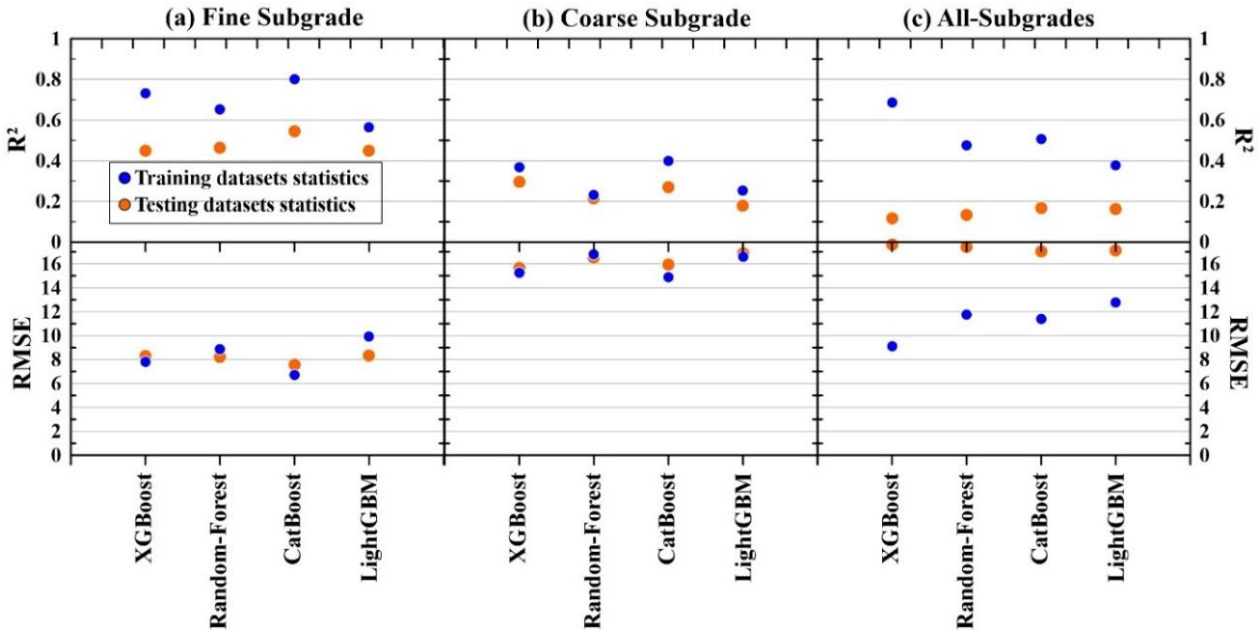

Figure 5.10 – Delta *MIRI* ML models statistics summary for (a) Fine Subgrade (b) Coarse Subgrade (c) All-Subgrades

From Figure 5.10, the performance of XGBoost and CatBoost algorithms is marginally better than that of the Random-Forest and LightGBM algorithms for Delta *MIRI* models. The XGBoost and CatBoost algorithms' performance improvement over the Random-Forest and LightGBM algorithms is especially obvious for the coarse subgrade models. The all-subgrades Delta *MIRI* models had varying results when examining the training and testing datasets statistics. The models developed for the all-subgrades type using the XGBoost and CatBoost algorithms also showed better results than when using the two other algorithms. However, when comparing the testing dataset statistics, the all-subgrades Delta *MIRI* model created using the CatBoost algorithm outperformed the models developed using the LightGBM, Random-Forest, and XGBoost algorithms. Overall, the models showed far lower R^2 values than the Delta *CLIRI* models.

Comparing the ML models results to the regression analysis models results in Table 5.6, the ML models appear to be significantly superior. The ML models provide an improvement in R^2 amounting up to 186%, 344%, and 879% for the fine, coarse, and all-subgrades, respectively. For *RMSE*, the error metric is reduced by up to 61%, 84%, and 77%, respectively. Comparing the ML models results for the Delta *MIRI* and *CLIRI* models, the Delta *MIRI* ML models provide a greater improvement than the regression analysis results. For instance, the Delta *CLIRI* ML model had an R^2 improvement of 353% for the all-subgrades model; in comparison, the Delta *MIRI* model had an R^2 improvement of 879%. Even as the model statistics indicate that the independent variables *Age*, *FI*, *PPT*, *PI*, and P_{200} are better at estimating Delta *CLIRI* than the Delta *MIRI* values, the Delta *MIRI* models developed using ML present a meaningful relationship between the variables and *MIRI*.

From the presented results above, ML does provide decent models for fine subgrades. The coarse and all-subgrades models do provide better results than the regression analysis results; However, the coarse and all-subgrade models do not provide suitable model statistics. The results confirm that fine-grained subgrades are more affected by climate, with a better variable correlation. The statistical metrics for the fine-grained models make it a good candidate for its use as a prediction tool. Additionally, the coarse-

grained ML models indicate some effect of climate, as opposed to none in regression; However, its relative contribution on the middle lane is not as significant as it gets muted due to repetitive loads due to traffic. Moreover, whenever possible, the development of IRI models should be aligned with the subgrade type, i.e., separate models for different subgrade types.

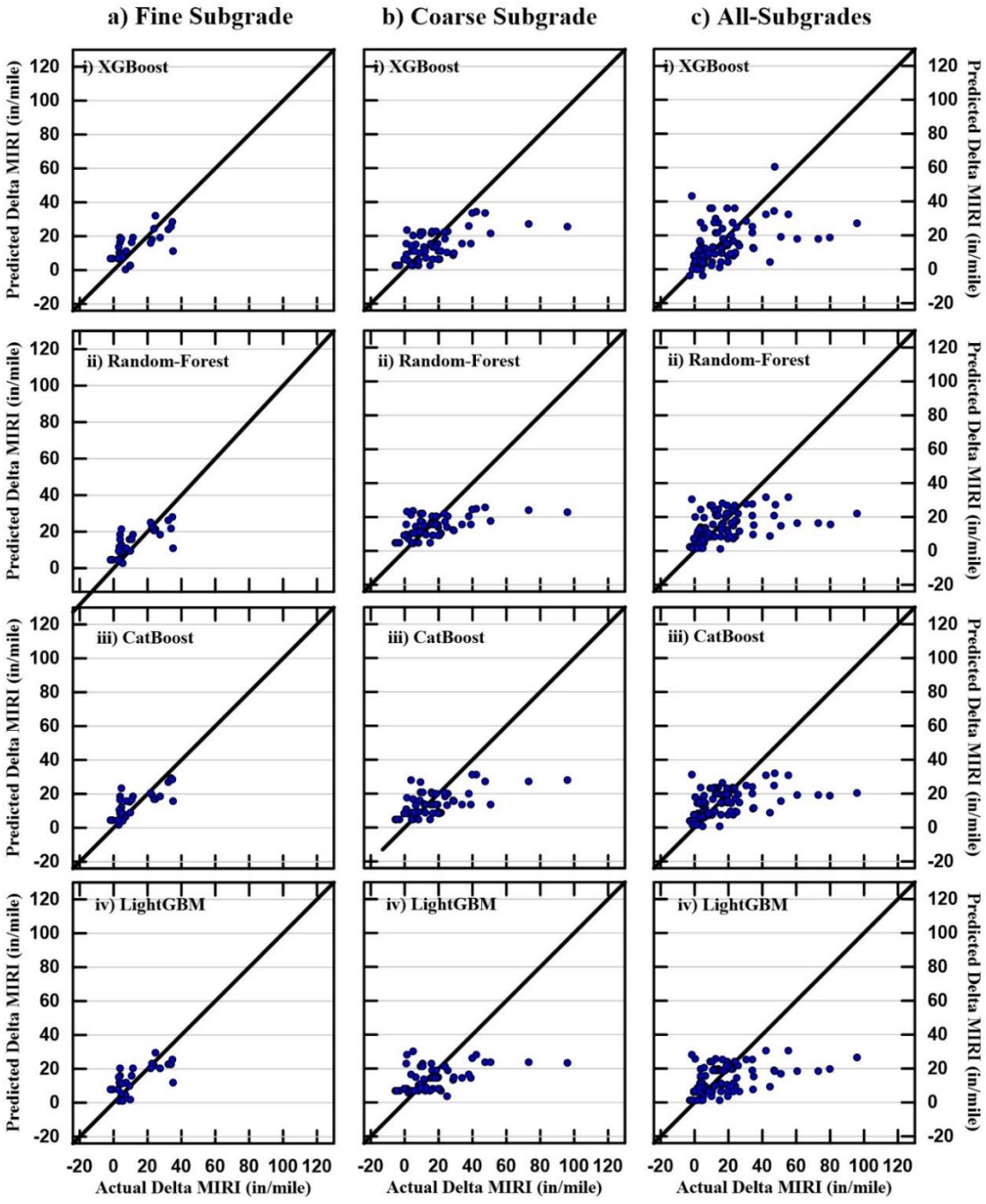

Figure 5.11 - Delta *MIRI* ML models actual vs predicted Delta *MIRI* values plots for (a) Fine Subgrade (b) Coarse Subgrade (c) All-Subgrades

The developed Delta *MIRI* models are used to create actual vs predicted values plots to visually assess the models' fit. The actual vs predicted values plots are presented in Figure 5.11 for all three subgrade types. The ML-developed Delta *MIRI* models' actual vs predicted plots for fine subgrades appear better fitting than the other subgrade types models in Figure 5.11. The fine subgrade models developed using the XGBoost and CatBoost algorithms showed predicted values that were randomly distributed around the regression line and provided a better data fit than the two other algorithms. The actual vs predicted values plot of the coarse subgrade models does not show a proper fit, confirming what was observed in the model statistics for both the training and testing datasets in Figure 5.10. The coarse subgrade models developed using the XGBoost and CatBoost algorithms show a slightly better fit than the models developed using the Random-Forest and LightGBM algorithms. The all-subgrades model's actual vs predicted plots appear to be visually close to one another, with the exception of the model developed using the XGBoost algorithm. However, the difference in the actual vs predicted plot for the model developed using the XGBoost algorithm does not significantly vary from the plots created using the three other models. Identifying the best-performing ML algorithms in modelling Delta *MIRI* would be helpful in understanding the pavement features' impact on *MIRI*.

The residual values plots for the Delta *MIRI* models are presented in Figure 5.12. The fine subgrade models' residual plots are almost identical, showcasing a random residual plot pattern. However, the all-subgrades and coarse subgrade type models show an X-axis unbalanced and horn-shaped residual pattern. Thus, based on the residual plots in Figure 5.12 and the summary of statistics for the Delta *MIRI* models in Figure 5.10, the models for coarse subgrade and all-subgrades pavements do not provide satisfactory results to be used outside the comparison context with their counterparts.

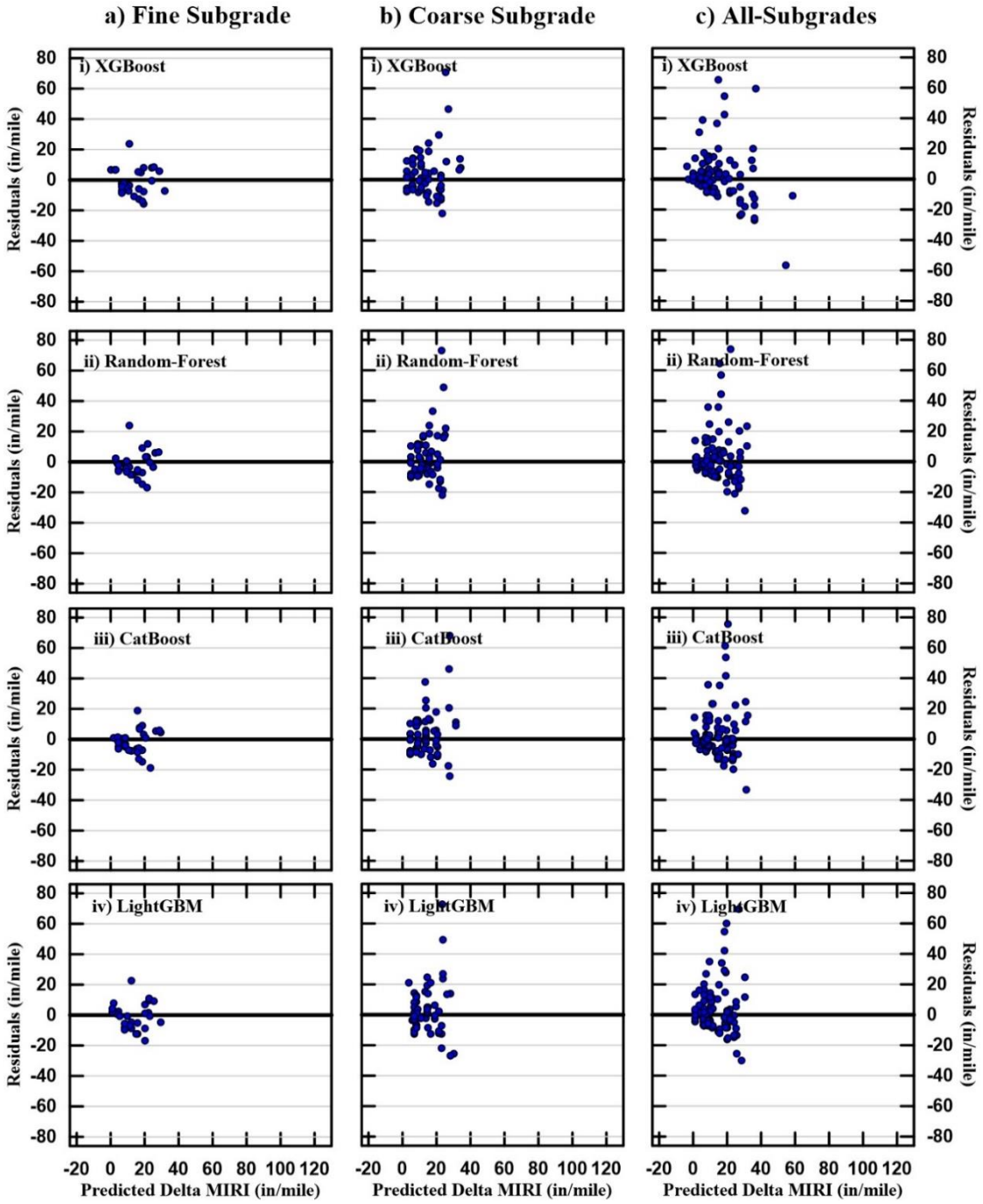

Figure 5.12 - Delta *MIRI* ML models residual Delta *MIRI* values plots for (a) Fine Subgrade (b) Coarse Subgrade (c) All-Subgrades

The ML-developed Delta *CLIRI* models show better accuracy results compared to the ML-developed Delta *MIRI* models. This finding suggests that the independent variables *Age*, *FI*, *PPT*, *PI*, and P_{200} can describe the dependent variable Delta *CLIRI* better than the dependent variable Delta *MIRI*. This can be

explained by the fact that the correlation between the independent variables and Delta *CLIRI* is stronger than the correlation between the independent variables and Delta *MIRI*. Additionally, there are other factors that contribute to the Delta *MIRI*, such as traffic, which would have a stronger correlation with Delta *MIRI* over Delta *CLIRI*, as the only traffic the center lane receives is when changing lanes. Moreover, the identified best ML algorithms for modelling Delta *MIRI* and *CLIRI* can be used to examine the individual impact of the pavement features *Age*, *FI*, *PPT*, *PI*, and P_{200} on Delta *MIRI* and *CLIRI*. The best-performing models can be used to create sensitivity analysis plots to examine the magnitude of impact the pavement features have on the models' output.

The results indicate that the effect of climate and subgrade type are the main contributors to the center lane IRI. This is indicated by the regression analysis and confirmed by the ML analysis. Additionally, the ML models indicate that the climate factors are more dominant for the middle lane IRI built on fine-grained subgrades. Moreover, the ML algorithms considered in this research show comparable results. The difference between the various algorithms appears marginal and supports that the ML algorithms examined in this chapter provide adequate results for IRI modelling. Finally, the works accomplished to this point in time from above indicate that ML has produced models that would provide decent predictive capabilities for Delta *CLIRI* and somewhat acceptable capabilities for Delta *MIRI* for fine-grained subgrades.

A sensitivity analysis was performed using the models developed with the XGBoost and CatBoost algorithms, as their performance was superior to the other algorithms. Sensitivity analysis helps in understanding the independent variables' impact on the models' output by varying the independent variables' values. The sensitivity analysis conducted is used to generate an average impact on model output for each of the variables. The average impact plot for the fine subgrade Delta *CLIRI* models is presented in Figure 5.13. The figure provides insight into how the independent variables affect the output of the two best-performing Delta *CLIRI* ML models.

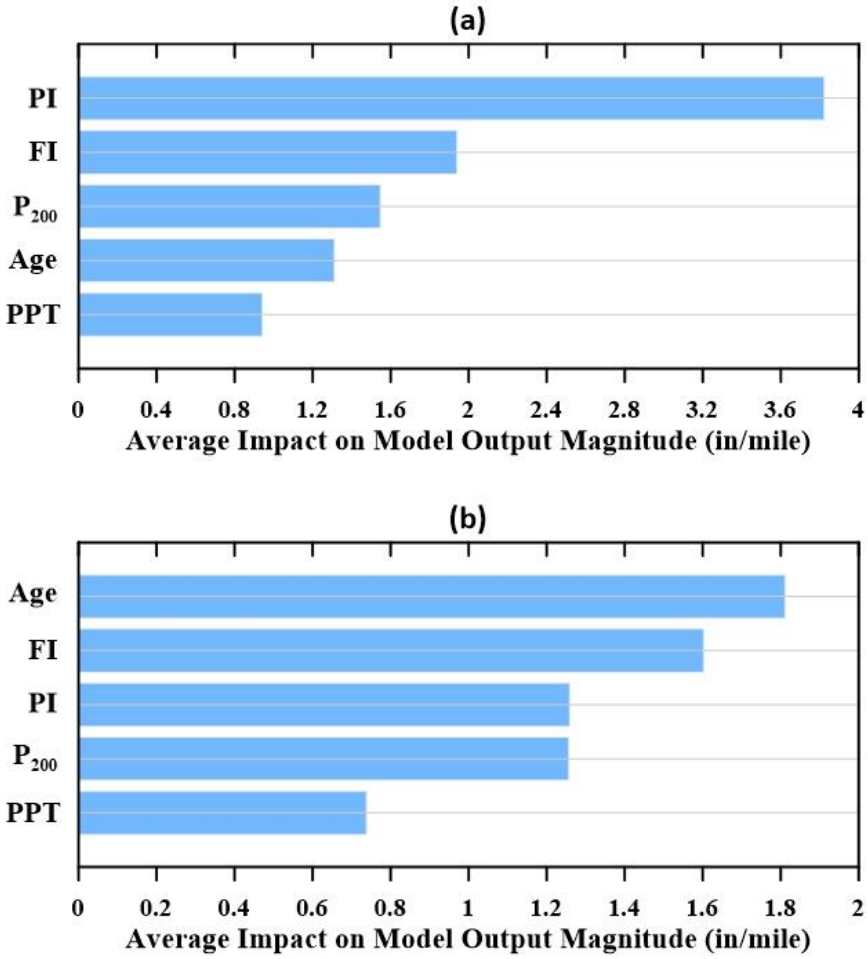

Figure 5.13 - Average impact on model output magnitude plot for Delta *CLIRI* ML models for fine subgrade type pavements using (a) XGBoost (b) CatBoost

From Figure 5.13, the variables in the Delta *CLIRI* ML models appear to be highly correlated with one another and thus could be interchangeable due to multicollinearity. For instance, the model built using the XGBoost algorithm shows *PI* as the most impactful variable, followed by *FI*, P_{200}, *Age*, and *PPT*. In contrast, the model built using the CatBoost algorithm shows *Age* as the most impactful on the model's output, followed by *FI*, *PI*, P_{200}, and *PPT*. In both models, *PPT* has the least impact on the model's output, while *FI* has the second most impact on both models' output. The variable *PI* appears to have a more significant impact on the XGBoost model than the CatBoost model; meanwhile, the variable *Age* appears to be more impactful for the CatBoost than the XGBoost model. Overall, the average impact plots for fine subgrades highlight the effect of climate, time, and subgrade soil properties on *CLIRI*. Several conclusions are drawn by making arguments from the average impact plot for the fine subgrade

pavements model. The larger the *PI* and *Age* values, the greater the impact on the Delta *CLIRI* models' output. Additionally, the effect of *PPT* on the models' output appears to be muted due to other variables' correlation with *PPT*. The variable *PPT* itself is not dependent on any other variable in the models. However, other variables in the models could be dependent on *PPT*, leading to muting the effect of *PPT* on the models' output. Moreover, climate parameters and subgrade soil properties appear to contribute significantly to the models' output.

Figure 5.14 presents the coarse subgrade Delta *CLIRI* models' average impact on the model output plot.

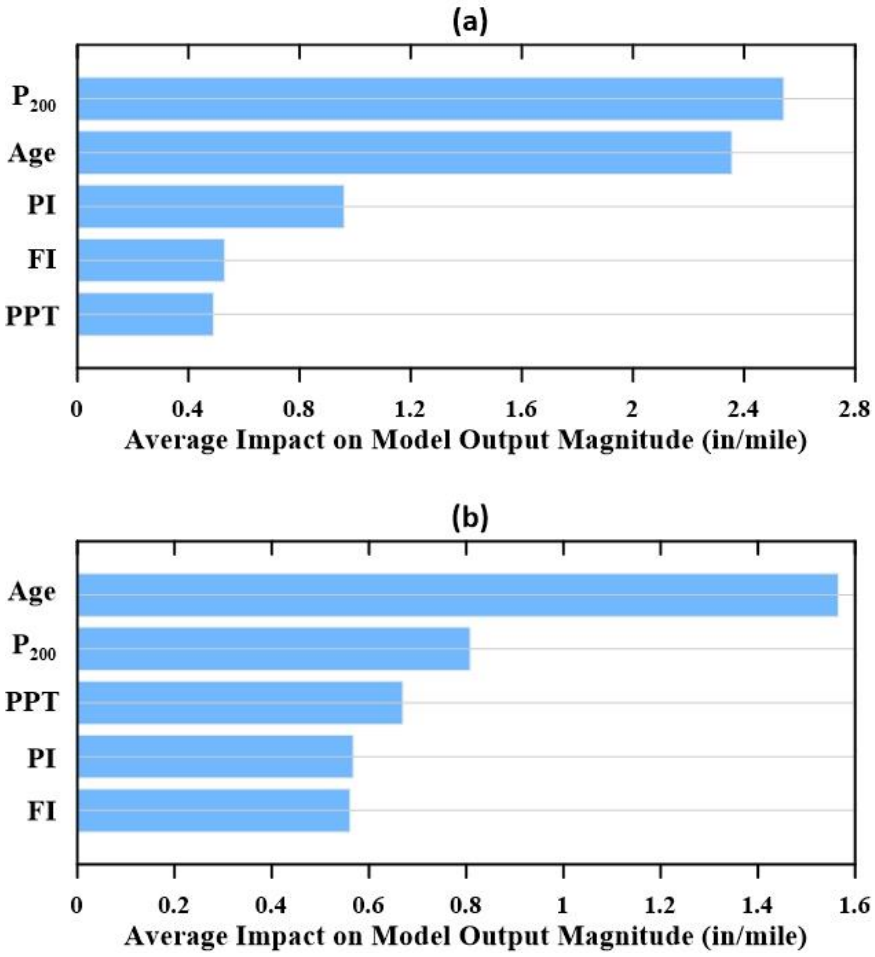

Figure 5.14 - Average impact on model output magnitude plot for Delta *CLIRI* ML models for coarse subgrade type pavements using (a) XGBoost (b) CatBoost

The average impact plot conducted for the XGBoost model and presented in Figure 5.14 shows that the variable P_{200} has the largest impact on the model's output, followed by *Age*, *PI*, *FI*, and *PPT*. The average impact plot for the CatBoost model shows the variable *Age* having the most impact on the model's output,

followed by P_{200}, *PPT*, *PI*, and *FI*. The variables *FI*, *PPT*, and *PI* in both models appear to pose the least impact on the models' output compared to the other independent variables in the models. This finding suggests that the Delta *CLIRI* values are not affected as much by climate parameters such as *FI* and *PPT* for coarse subgrade pavements, possibly due to the lack of plasticity in the subgrade's soil. Additionally, when comparing the two subgrade's soil properties, the P_{200} impact on the model is significantly greater than *PI*. Furthermore, the impact of climate parameters in the model could be related to geographically related climatic patterns where the coarse subgrade soils are used. For instance, southern states are more likely to have coarse-grained subgrades than fine-grained subgrades, based on Table 5.2.

Figure 5.15 presents the average impact plot for the Delta *CLIRI* models for the all-subgrades type pavements developed using the XGBoost and CatBoost algorithms.

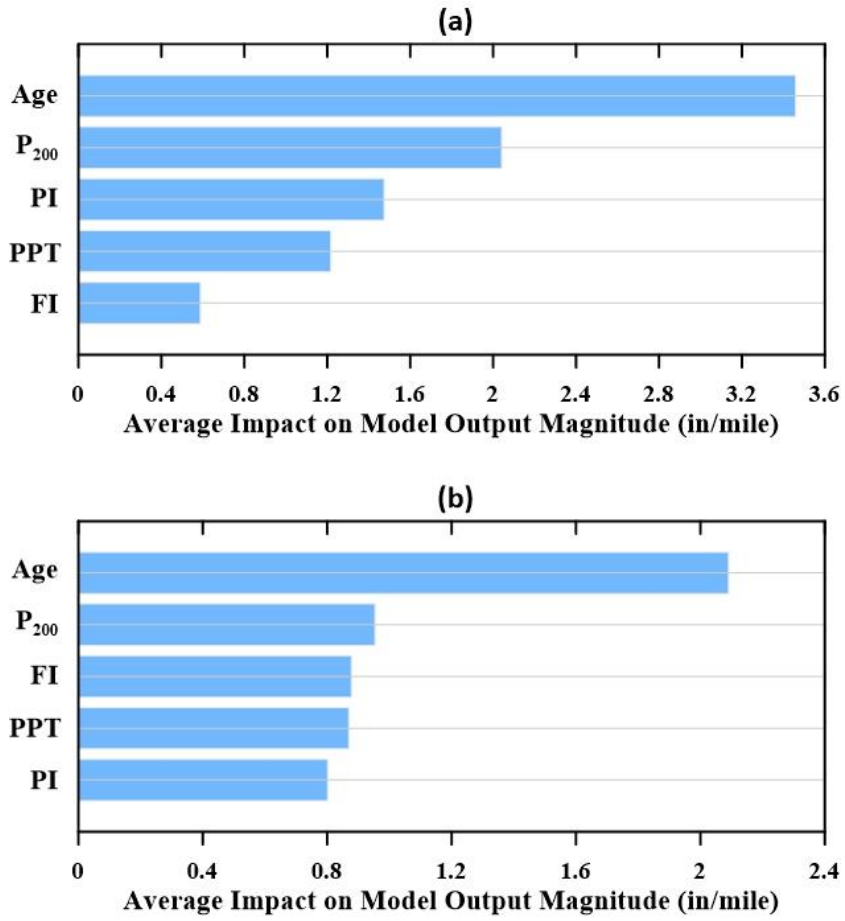

Figure 5.15 - Average impact on model output magnitude plot for Delta *CLIRI* ML models for all-subgrades type pavements using (a) XGBoost (b) CatBoost

From the figure above, both models have *Age* as the variable with the most impact on the model's output, followed closely by the variable P_{200}. The variable P_{200} is followed by the variables *PI*, *PPT*, and *FI* as the most impactful on the model's output for the XGBoost model. For the CatBoost model, the variable P_{200} is followed by the variables *FI*, *PPT*, and *PI* as the most impactful on the model's output. The interchangeability between *PI* and *FI* in the two models suggests the correlation between the two variables. As mentioned earlier in the text, pavements with higher soil plasticity are more prone to freeze-thaw damage than pavements with little to no soil plasticity.

5.6 Climate Change Impact Study

An impact study on the effects of climate change on pavements for two states in the contiguous United States is performed using sensitivity analysis. As stated earlier in the text, the contiguous United States is expected to experience a range of climate change impacts depending on the region. As recently reported by the Intergovernmental Panel on Climate Change (IPCC), the eastern United States region is expected to experience an increase in precipitation accompanied by a warming effect of at least 1.5°C (IPCC 2021). The average annual precipitation is directly incorporated into the models developed in this study, and the effect of temperature increase can be indirectly linked to the *FI*, which is also incorporated into the models developed in this study. The *FI* is calculated by taking the sum of the mean daily temperatures below the freezing temperature in a freezing season (ARA Inc. 2020). As a result, the temperature rise is expected to be inversely proportional to the *FI*. The sensitivity analysis conducted for the ML models has indicated that an increase in precipitation is expected to cause a greater change in pavement roughness for pavements on fine subgrades than on coarse subgrades. Similarly, as observed in the sensitivity analysis, a decrease in *FI* value is much more impactful on pavements for fine subgrades than on coarse subgrades. In other words, the sensitivity analysis studies show that *FI* is linked to IRI for fine subgrades pavements. The relationship implies that any decrease in *FI* in the future could potentially have a positive effect on IRI. This is due to lower *FI* values translating to a lower chance of freeze-thaw damage for fine-subgrades pavements.

The impact of *PPT* and *FI* have been examined separately and not jointly through the conducted sensitivity analysis in this chapter; Thus, the combined effect of an increase in *PPT* and a decrease in *FI* due to climate change should be examined and understood. A climate change impact study is conducted to examine the combined effect of an increase in *PPT* and a decrease in *FI* due to climate change. The case study is conducted to better understand the impact of climate change at the end of the next two decades. The states of Virginia and Ohio, located on the eastern side of the contiguous United States, are considered for this impact study. Pavements in Virginia are classified as coarse-grained subgrade pavements, whereas pavements in Ohio are classified as fine-grained subgrade pavements. Thus, the states of Virginia and Ohio can be used to understand the effects of increased *PPT* and decreased *FI* on coarse and fine subgrades, respectively.

According to the IPCC report, it is reasonable to assume a 10% increase in annual average precipitation and a 1.5°C increase in temperature for the selected region (IPCC 2021, Masson-Delmotte, et al. 2021). Based on the data provided by FHWA, for the considered range of years for the SPS-1 experiment, the states of Virginia and Ohio had, on average, 89 and 117 days below the freezing point, respectively. This adjustment in *FI* values is made by adding 1.5°C to the daily temperature for Virginia and Ohio and then recalculating the *FI* values for both states. This reduces the *FI* values for Virginia and Ohio from the values shown in Table 5.1 to 0 and 402, respectively. Additionally, the *PI* and P_{200} values presented in Table 5.2 are used in this impact study to represent the pavements' subgrade properties for each of the states. The first case for both states is set to the historical climate values shown in Table 5.1, while the second case is set to the future climate averages due to climate change as described above. The Delta *CLIRI* model developed for all subgrade types using the CatBoost algorithm is used for this case study as it has shown superior predictive capabilities relative to the other three models. The results of the conducted study are shown in Figure 5.16.

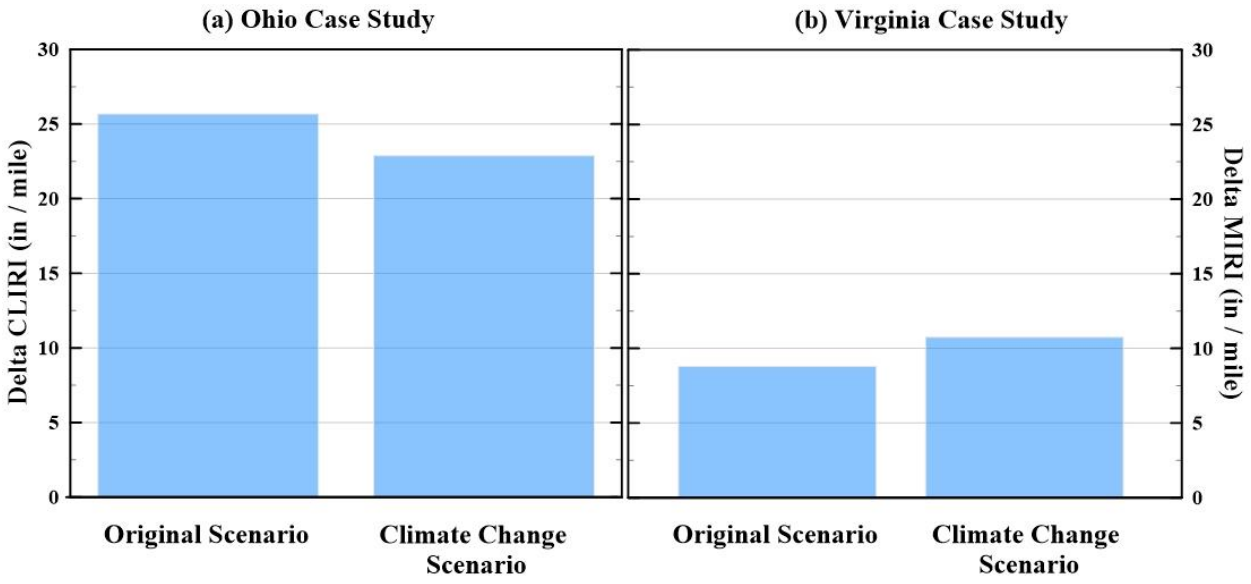

Figure 5.16 - Climate change case study for (a) the state of Ohio (b) the state of Virginia

Figure 5.16 shows that fine subgrade pavements in the state of Ohio had an 11 percent decrease in roughness projection due to the combined effect of a decrease in *FI* and an increase in *PPT*. In contrast, the state of Virginia shows a 22 percent increase in pavement roughness for coarse subgrade pavements as a result of the combined effect of a decrease in *FI* and an increase in *PPT*. The decrease in roughness is due to the increase in *PPT* being offset by the decrease in *FI* for fine subgrades. For coarse subgrades, *FI* is not as impactful as for the fine subgrades; therefore, the effect of the *PPT* increase is amplified. Additionally, the state of Virginia is geographically close to the state of Ohio; however, the state of Ohio had an average *FI* value that was more than six times that of the state of Virginia during the time periods listed in Table 5.1. Thus, the decrease in *FI* values combined with an increase in *PPT* appears to positively affect pavements in the state of Ohio, possibly due to the decrease in freeze-thaw damage. The findings of this study could also be potentially applicable to southern Canadian provinces as they are close to the northern states and have a similar climate.

5.7 Conclusions

Pavement roughness directly affects the safety of road users and global warming emissions. Pavement deterioration over time is caused by various factors that include the climate and soil subgrade

type. A common method for quantifying the deterioration of pavements is the measurement of pavement roughness. Pavement roughness is often quantified using the standardized index, the IRI.

This study examines the changes in both the *MIRI* and *CLIRI* (pavement roughness) over time due to environmental and subgrade variables using both regression analysis and Machine Learning (ML) methods. This study provides insight into the effect of climate change by examining the contribution of climate on different pavement subgrades using the LTPP database. The developed models are classified as fine-grained subgrade models, coarse-grained subgrade models, and all-subgrades models.

The regression analysis developed models are a function of the annual average freezing index (*FI*), the average annual precipitation (*PPT*), the percent passing No. 200 sieve (P_{200}), the Plasticity Index (*PI*), and the age of the pavement since the first recording of data (*Age*). The regression analysis results lead one to conclude that the independent variables *FI*, *PPT*, *PI*, and P_{200} are better at describing Delta *CLIRI* values than at describing Delta *MIRI* values. In general, the regression analysis developed models do not perform well, especially in describing Delta *MIRI*.

The ML-developed Delta *CLIRI* and Delta *MIRI* models using the best-performing ML algorithms identified in Chapter 4 show improvement in the predictive capabilities than the ones developed using regression analysis. From the statistical results of the models, one can conclude that the use of ML in model development results in better predictive capabilities. Additionally, the climate variables *FI* and *PPT* were found to be better at describing Delta *CLIRI* than Delta *MIRI*. This leads to the conclusion that the center lane's roughness is primarily impacted by climate, whereas the climate impact on the wheel-path lanes is muted.

Moreover, the models' performance varied depending on the pavement's subgrades. Fine subgrade pavement models had better predictive capabilities than coarse and all-subgrades models. This leads one to conclude that the pavements of different subgrades deteriorate differently. The results also confirm that climate affects fine-subgrade pavements more than coarse-subgrade and all-subgrades pavements. The

differing results based on subgrade types indicate that the development of the IRI models should be aligned with the subgrade type whenever possible.

Sensitivity analyses were conducted to better understand the impact of climate and subgrade properties on pavement roughness using the ML-developed models. From the results, some conclusions can be drawn, pavements with higher *PI* values result in a greater change in pavement roughness. Additionally, *FI* is highly correlated and impactful on pavement roughness for fine subgrade pavements. Moreover, the variable P_{200} significantly impacts the pavement roughness for coarse subgrade pavements, while *PI* had a lower effect on pavement roughness, this is due to the fact that coarse-grained soils have little to no plasticity. Furthermore, combining fine and coarse subgrade pavements into one model results in a significant reduction in the ability to develop a meaningful full relationship irrespective of the technique used for the model developed. In addition, the variables *PI* and *FI* were found to be highly correlated with one another. This concludes that higher plasticity soils are more prone to freeze-thaw damage than lower plasticity soils. Also, fine-grained subgrade pavements generally result in a greater change in pavement roughness than coarse-grained subgrade pavements.

The study results can also aid in drawing conclusions related to the effects of climate change on pavement roughness in the contiguous United States. The conclusion drawn for the climate change impact on pavements could potentially be valid for many fine-grained subgrade pavements in the contiguous United States. The conducted climate change impact study for pavements in the eastern United States region revealed that fine subgrade pavements models projected a decrease in pavement roughness, while coarse subgrade pavements models projected an increase in pavement roughness as a result of the combined impact of an increase in *PPT* and a decrease in *FI* caused by an increase in temperature due to climate change. The increase in pavement roughness is due to the offset created by the decrease in *FI* to the increase in *PPT*. This leads to the conclusion that *FI* is not as impactful for coarse subgrades as it is for fine subgrade pavements. This study's findings may also be applicable to southern Canadian provinces as they have similar climatic conditions to the locations considered for the United States.

The results of this study were used to understand the effect of climate on pavements with different subgrade properties. The study relies heavily on the SPS-1 data for its conclusions. One of the limitations is the limited number of data points, especially for fine-grained subgrade pavements. Additional research should be done to improve models as additional data becomes available. Models based on ML show promise and can be used to further examine the effect of climate change on pavement roughness and the rate of pavement deterioration based on different climate change scenarios.

Chapter 6 : Summary, Conclusions, and Recommendations for Future Research

6.1 Summary

The International Roughness Index (*IRI*) has been widely adopted worldwide; however, *IRI* modelling efforts for the index have largely relied on regression analysis. Machine Learning (ML) has been gaining popularity in recent years in modelling *IRI*. This research aimed to examine the effectiveness of ML in the development of *IRI* models as compared to the conventional regression analysis methods. This was accomplished for PMS data from the province of Alberta and LTPP data from the contiguous United States. This research also examines the effect of climate on pavement roughness for pavements with different soil subgrades over time to examine the effect of climate change on pavements.

This study is divided into three distinct sequential parts. The first part of this study examined Alberta Transportation's (AT) data from AT's Pavement Management System (PMS) database. The dataset provided by AT was first examined for any gaps or limitations in the extracted dataset and was mitigated for the purpose of this study. Regression analysis was then used to develop *IRI* models using AT's data. Two separate *IRI* models were developed using regression analysis, an Asphalt Concrete Pavement (ACP) and an asphalt concrete Overlay pavements (OL) model. A sensitivity analysis was then conducted using the developed models to examine the impact of independent variables on the pavement *IRI*. Reduced *IRI* models were then developed for both pavement types based on the results of the sensitivity analyses. The first section concludes with a comparison of the developed models with the Mechanistic-Empirical Pavement Design Guide's (MEPDG) *IRI* model. The comparison was carried out in terms of the prediction capability of the model as well as carrying out a Life Cycle Analysis (LCA) and a Life Cycle Cost Analysis (LCCA). The LCA and LCCA analyses provide insight to assess the environmental and economic impact using different *IRI* models.

The second part of this study utilizes the same AT dataset to develop *IRI* models using ML algorithms. This part of the study uses supervised learning under the ML subbranch of artificial intelligence. Generally, the supervised ML algorithms fall into the categories of decision trees, nearest neighbours, neural networks, support vector machines, ensemble methods, and regularized linear regression. The ensemble methods algorithms are broken down into bagging and boosting ensemble methods. The algorithms considered are decision trees, Random-Forest, artificial neural networks, LightGBM, support vector machines, K-Nearest-Neighbour, XGBoost, CatBoost, Elastic-Net, Ridge, and Lasso. This part of the study examines the effectiveness of using ML in modelling *IRI* and highlights the best-performing ML algorithms in modelling *IRI*. A sensitivity analysis was also performed using the ML-developed models, and reduced *IRI* models were created based on the sensitivity analysis results.

The final part of this study utilizes the Specific Pavement Studies-1 (SPS-1) data, which is a part of the Long-Term Pavement Performance (LTPP) program run by the Federal Highway Administration (FHWA) to examine the changes in pavement roughness due to climate for different subgrade soil types in the contiguous United States. In this part of the study, IRI data is available for the mean of the left and right wheel paths (*MIRI*) and the Center-Lane (*CLIRI*). The *CLIRI* was used to examine the effect of climate on different subgrade soil types in the absence of traffic, as the only traffic the center lane receives is from vehicles changing lanes. This part of the study examines the changes in *MIRI* and *CLIRI* in relation to climate and subgrade soil type using both regression analysis and ML. A sensitivity analysis was also conducted for the changes in *CLIRI* models to examine the impact of variations in climate and subgrade soil properties on pavement roughness in the absence of traffic in the contiguous United States. The study concludes with a case study that shows how the models in this section can be used to investigate the effect of climate change on pavements in the contiguous United States.

6.2 Conclusions

6.2.1 Significance of Climate on Pavement Roughness for Cold Region Pavements

Fourteen variables related to climate were considered in the *IRI* models' development using Alberta's PMS data. The climate variables considered in the *IRI* model development appeared to have no or weak correlation with *IRI*. This does not lead one to conclude that climate does not affect pavement roughness; However, at the temporal resolution of climate variables and *IRI* measurements, one may conclude that the effect of climate is indirectly integrated into the pavement distress measures and that no meaningful association between pavement roughness and climate variables is discernible.

6.2.2 Variables with Significant Impact on Pavement Roughness for Cold Region Pavements

The regression analysis developed *IRI* models using Alberta's PMS data were used to conduct sensitivity analyses to identify the most impactful variables on the models' output. From the sensitivity analysis results, one can conclude that the age of the pavement, the 80th percentile rut depth, plasticity index, other cracking as defined by Alberta transportation, and the percent passing 200 sieve were the most impactful for OL pavements. Additionally, for ACP pavements, the 80th percentile rut depth, transverse cracking, other cracking as defined by Alberta transportation, plasticity index, and the annual precipitation were the most impactful. One can conclude that the most impactful variables are based on the pavement type and could be utilized to explain the deterioration of different types of pavements over time.

Furthermore, the ML-developed *IRI* models helped in identifying the most impactful variables on the models' output while minimizing the effects of multicollinearity. The sensitivity analyses conducted for the ML-developed models helped conclude that the equivalent single axle load, the plasticity index, the age of the pavement since construction or since rehabilitation, the pavement's surface thickness, and the 80th percentile rut depth are the most impactful on the models' output for both pavement types.

Moreover, types of soil and soil plasticity information appear to be significant from the perspective of *IRI* model development. Given that some of this information may be available during construction or can be easily obtained at that time, road agencies should prioritize including it in the PMS database.

6.2.3 Performance of ML-Developed Models in Comparison to Regression Analysis Developed Models.

Regardless of the dataset used to develop the ML models, ML-developed models showed improvement in predictive capabilities over the regression analysis-developed models. The models' results help to generally conclude that the use of ML algorithms in model development results in better predictive capabilities. Additionally, one can also conclude that the use of ensemble methods ML algorithms result in the best-performing models in comparison to regression analysis and the other types of ML algorithms.

6.2.4 Performance of Reduced Models in Comparison to their More Complex Counterparts

Reduced *IRI* models were developed for both the regression analysis and the ML models using the most impactful variables. The reduced models' results conclude that they are as efficient as their more complex counterparts for the regression analysis models. Additionally, the reduced models can provide improvement in performance over their more complex counterparts for ML-developed models as they are less susceptible to overfitting the training dataset. Moreover, the reduced models can aid agencies in identifying which parameters to measure.

6.2.5 Performance of Data-Specific Models in Comparison to General Models

Comparisons were made for the developed *IRI* models using data from Alberta's PMS database with existing models such as the MEPDG *IRI* model. The ML-developed models were found to result in a three-fold increase in the coefficient of determination value. The results of the comparison concluded that site-specific information for the model's development results in a significant improvement in

performance. Additionally, models developed using ML perform significantly better than general models developed using regression analysis.

6.2.6 Economic Outcome of Adopting More Accurate Models

A case study was conducted to examine the developed models using Alberta's PMS data compared to the MEPDG *IRI* model using a selected road section part of the Trans-Canada highway over a 50-year period. The comparison showed a 26% reduction in maintenance and rehabilitation costs compared to using the MEPDG model to create a maintenance and rehabilitation schedule. The results of this comparison conclude that the use of more accurate models to create maintenance and rehabilitation schedules can result in significant savings.

6.2.7 Impact of Climate on Center Lane's Roughness

Separate pavement roughness models were created for the pavements' center lane and wheel paths using LTPP data to examine the effect of climate on center lanes in comparison to wheel paths, as the center lanes' only traffic is when vehicles change lanes. The results lead one to conclude that the freezing index, the average annual precipitation, the percent passing No. 200 sieve, the plasticity index, and the age of the pavement since the first recording of data can predict the center lane's roughness better than the wheel path's roughness. Moreover, using ML helps conclude that the freezing index and the average annual precipitation have a more pronounced effect on the center lane than the wheel paths. This leads to the conclusion that the roughness of the center lane is predominantly influenced by climate, whereas the climate impact on the wheel-path lanes is muted.

6.2.8 Importance of Subgrade's Properties on Pavement Deterioration

The LTPP database was used to develop pavement roughness models distinguished by the pavement's subgrades. Fine subgrade pavement models were found to outperform coarse and all-subgrades models in terms of predictive capabilities. The results lead to the conclusion that pavements of different subgrades degrade differently. The findings also help conclude that fine-subgrade pavements are

more affected by climate than coarse-subgrade and all-subgrades pavements. The different results based on subgrade types suggest that *IRI* model development should be linked with the subgrade type wherever possible.

6.2.9 Variables with Significant Impact on Pavement Roughness for the Contiguous United States

Sensitivity analyses were conducted using the ML models developed using data from the LTPP database to understand better the impact of climate and subgrade properties on pavement roughness. Some conclusions can be drawn from the results of the sensitivity analyses. Pavements with higher plasticity index values result in a greater change in pavement roughness. Moreover, the freezing index is highly correlated and impactful for fine subgrade pavements. Additionally, the percent passing No. 200 sieve has a significant impact on coarse subgrade pavements, while the plasticity index had no significant impact; this is due to the fact that coarse-grained soils have little to no plasticity. Furthermore, the combination of fine and coarse subgrade pavements into one model significantly reduces the ability to develop a meaningful full relationship regardless of the technique used in developing the models. In addition, higher plasticity soils are more prone to freeze-thaw damage than soils of lower plasticity. Finally, finer-grained subgrade pavements generally deteriorate at a greater rate than coarse-grained subgrade pavements.

6.2.10 Climate Change Impact Study for the Contiguous United States

A climate change impact study was conducted using data from the LTPP database for the contiguous United States. The conclusions drawn for the climate change impact on pavements may apply to many fine-grained subgrade pavements in the contiguous United States. This study's results concluded that climate change's impact on fine subgrade pavements varied from the impact on coarse subgrade pavements. The fine subgrade pavement models projected a decrease in pavement roughness. In contrast, coarse subgrade models projected an increase in pavement roughness due to the combined effect of an increase in the average annual precipitation and a decrease in the freezing index. The decrease in

pavement roughness projected for the fine subgrade pavements is due to the decrease in the freezing index being counteracted by the increase in the average annual precipitation. This leads one to conclude that for coarse subgrade pavements, the freezing index is not as impactful compared to fine subgrade pavements; thus, the effect of the increase in the average annual precipitation is amplified.

6.3 Contribution of this Research

This study developed *IRI* models with superior predictive capabilities for Alberta Transportation that can be used in their PMS. The more accurate *IRI* models will help Alberta Transportation maintain its pavements when it is most crucial for safety and rideability for the end-user. As a result of accurately predicting pavement conditions through the *IRI* models, Alberta Transportation would be able to secure the funding needed without causing service interruptions. The *IRI* models developed in this study can result in considerable financial savings for Alberta Transportation while ensuring safer roadways. This study also developed *IRI* models using ML using LTPP data. Moreover, the study highlights the best-performing ML algorithms for *IRI* modelling purposes. Additionally, this study also investigates the effect of climate on pavement roughness of different subgrade soil types using LTPP data. This study also provides insight into the effect of climate change on pavements in the contiguous United States.

6.4 Recommendations for Future Studies

There are some limitations tied to this research effort due to data and time constraints. The effectiveness of other ML algorithms not examined in this study effort in developing *IRI* models can and should be examined in the future. Moreover, the climate change case study only considers climate change in the contiguous United States; additional locations could be examined in and outside the contiguous United States. The following list is a recommendation for future research.

- Refinement of the *IRI* models as more data becomes available.
- Examining the effectiveness of other ML algorithms not used in this study in developing *IRI* models.

- Examining the effectiveness of other classes of ML, including unsupervised and reinforcement learning, in modelling *IRI*.
- Investigate the effectiveness of the best-performing ML algorithms in this study in modelling other pavement-related applications, such as modelling pavement distresses.
- Examine the effects of climate change on pavement roughness and the rate of pavement deterioration based on different climate change scenarios.
- Examining other climate change scenarios' impact on pavement roughness and incorporating these patterns into *IRI* modelling.

References

AASHTO. 1985. Guidelines on Pavement Management. American Association of State Highway and Transportation Officials.

AASHTO. 1990. AASHTO Guidelines for Pavement Management Systems. American Association of State Highway and Transportation Officials (AASHTO).

AASHTO. 2009. Rough roads ahead: fix them now or pay for them later. American Association of State Highway and Transportation Officials (AASHTO).

AASHTO. 2012. Pavement Management Guide, Second Edition.

Abdelaziz, N., Abd El-Hakim, R.T., El-Badawy, S.M., and Afify, H.A. 2020. International Roughness Index prediction model for flexible pavements. International Journal of Pavement Engineering, **21**(1): 88–99. doi:10.1080/10298436.2018.1441414.

Alberta Agriculture and Forestry. 2020. Alberta Soil Information Viewer. Available from https://soil.agric.gov.ab.ca/agrasidviewer/.

Alberta Agriculture and Forestry. 2021a. Current and Historical Alberta Weather Station Data Viewer. Available from https://acis.alberta.ca/acis/weather-data-viewer.jsp.

Alberta Agriculture and Forestry. 2021b. Alberta Climate Information Service (ACIS). Alberta Agriculture and Forestry.

Alberta Transportation. 2003. Surface Condition Rating (SCR) Manual. *In* Contract Administration Manual, 4.2. Government of Alberta Transportation.

Alberta Transportation. 2020. Transportation Annual Report 2020-2021. Government of Alberta.

Alberta Transportation. 2021. Unit prices and cost adjustments - Price averages and indexes for Alberta Transportation highway, bridge and water management construction tenders. Government of Alberta.

Albuquerque, F.S., and Núñez, W.P. 2011. Development of Roughness Prediction Models for Low-Volume Road Networks in Northeast Brazil. Transportation Research Record: Journal of the Transportation Research Board, **2205**(1): 198–205. doi:10.3141/2205-25.

Al-Suleiman (Obaidat), T.I., and Shiyab, A.M.S. 2003. Prediction of Pavement Remaining Service Life Using Roughness Data—Case Study in Dubai. International Journal of Pavement Engineering, **4**(2): 121–129. doi:10.1080/10298430310001634834.

ARA Inc. 2001a. Pavement management guide.

ARA Inc. 2001b. Appendix CC-1: Correlation of CBR values with soil index properties. *In* Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures.

ARA Inc. 2020. Mechanistic-Empirical Pavement Design Guide: A Manual of Practice. *In* 3rd edition. Washington, DC United States.

Aravind, N., Nagajothi, S., and Elavenil, S. 2021. Machine learning model for predicting the crack detection and pattern recognition of geopolymer concrete beams. Construction and Building Materials, **297**: 123785. doi:10.1016/j.conbuildmat.2021.123785.

Athena Sustainable Materials Institute. 2006. Life Cycle Perspective on Concrete and Asphalt Roadways: Embodied Primary Energy and Global Warming Potential. Transportation Association of Canada (TAC).

Athena Sustainable Materials Institute. 2018. User Manual and Transparency Document - Pavement LCA v3.1 – Web Application. Athena Sustainable Materials Institute.

Ayodele, T.O. 2010. Types of Machine Learning Algorithms. *In* New Advances in Machine Learning. *Edited by* Y. Zhang. InTech.

Bishop, C.M. 2006. Pattern recognition and machine learning. Springer, New York.

Breiman, L. 1996. Bagging predictors. Machine Learning, **24**(2): 123–140. doi:10.1007/BF00058655.

Breiman, L. 2001. Random Forests. Machine Learning, **45**(1). doi:https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. 2017. Classification And Regression Trees. *In* 1st Edition. Routledge, Boca Raton.

Carey Jr, W.N., and Irick, P.E. 1960. The Pavement Serviceability-Performance Concept. Highway Research Board, (250).

Charte, D., Charte, F., García, S., and Herrera, F. 2019. A snapshot on nonstandard supervised learning problems: taxonomy, relationships, problem transformations and algorithm adaptations. Progress in Artificial Intelligence, **8**(1): 1–14. doi:10.1007/s13748-018-00167-7.

Chen, T., and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,: 785–794. doi:10.1145/2939672.2939785.

Choi, J., Adams, T.M., and Bahia, H.U. 2004. Pavement Roughness Modeling Using Back-Propagation Neural Networks. Computer-Aided Civil and Infrastructure Engineering, **19**(4): 295–303. doi:10.1111/j.1467-8667.2004.00356.x.

Choi, S., and Do, M. 2019. Development of the Road Pavement Deterioration Model Based on the Deep Learning Method. Electronics, **9**(1): 3. doi:10.3390/electronics9010003.

Chou, S.-F., and Pellinen, T.K. 2005a. Assessment of Construction Smoothness Specification Pay Factor Limits Using Artificial Neural Network Modeling. Journal of Transportation Engineering, **131**(7): 563–570. doi:10.1061/(ASCE)0733-947X(2005)131:7(563).

Chou, S.-F., and Pellinen, T.K. 2005b. Assessment of Construction Smoothness Specification Pay Factor Limits Using Artificial Neural Network Modeling. Journal of Transportation Engineering, **131**(7): 563–570. doi:10.1061/(ASCE)0733-947X(2005)131:7(563).

City of Edmonton. 2021. Pavement and Road Management. Available from https://www.edmonton.ca/city_government/initiatives_innovation/pavement-and-road-management.

Cover, T., and Hart, P. 1967. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, **13**(1): 21–27. doi:10.1109/TIT.1967.1053964.

Darlington, J.R. 1970. . Evaluation and Application Study of the General Motors Corporation Rapid Travel Profilomefer. Michigan Department of State Highways,.

Devore, J.L. 2016. Probability and statistics for engineering and the sciences. Boston, MA : Cengage Learning.

dhage, S., and Raina, C. 2016. A review on Machine Learning Techniques. International Journal on Recent and Innovation Trends in Computing and Communication, **4**(3): 395–399.

Dietterich, T.G. 1997. Machine-Learning Research. AI Magazine, **18**(4): 40. doi:https://doi.org/10.1609/aimag.v18i4.1324.

Dillard, J.H., and Allen, T.M. 1959. Comparison of Several Methods of Measuring Road Surface Friction. Highway Research Board, (219).

Dorogush, A.V., Ershov, V., and Gulin, A. 2018. CatBoost: gradient boosting with categorical features support. arXiv:1810.11363 [cs, stat],.

Draper, N.Richard. 1998. Applied Regression Analysis. *In* 3rd edition. New York : Wiley.

Drucker, H. 1997. Improving Regressors using Boosting Techniques. : 9.

Dubitzky, W., Wolkenhauer, O., Cho, K.-H., and Yokota, H. 2013. Analysis of Variance (ANOVA) Tables. *In* Encyclopedia of Systems Biology. Springer, New York, NY.

EL Naqa, I., and Murphy, M. 2015. What Is Machine Learning? Springer.

Fakhri, M., and Shahni Dezfoulian, R. 2019. Pavement structural evaluation based on roughness and surface distress survey using neural network model. Construction and Building Materials, **204**: 768–780. doi:10.1016/j.conbuildmat.2019.01.142.

FHWA. 2015. The Long-Term Pavement Performance Program. Federal Highway Administration Washington.

FHWA. 2022. Long-Term Pavement Performance. Available from https://highways.dot.gov/research/long-term-infrastructure-performance/ltpp/long-term-pavement-performance.

Finn, F. 1998. Pavement Management Systems - Past, Present, and Future, Public Roads: 80 Years Old, But The Best Is Yet to Come. **62**(1).

Friedman, J., Hastie, T., and Tibshirani, R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, **33**(1). doi:10.18637/jss.v033.i01.

George, K.P. 2000. MDOT Pavement Management System: Prediction Models and Feedback System. Mississippi Department of Transportation.

Georgiou, P., Plati, C., and Loizos, A. 2018. Soft Computing Models to Predict Pavement Roughness: A Comparative Study. Advances in Civil Engineering, **2018**: 1–8. doi:10.1155/2018/5939806.

Gillespie, T.D. 2001. Everything You Always Wanted to Know about the IRI, But Were Afraid to Ask! : 14.

Gillespie, T.D., Sayers, M.W., and Segel, L. 1980. Calibration of response-type road roughness measuring systems. Transportation Research Board, National Research Council, Washington, D.C.

Goldberger, J., Roweis, S., Hinton, G., and Salakhutdinov, R. 2005. Neighbourhood Components Analysis. **15**: 8.

Gong, H., Sun, Y., and Huang, B. 2019. Gradient Boosted Models for Enhancing Fatigue Cracking Prediction in Mechanistic-Empirical Pavement Design Guide. Journal of Transportation Engineering, Part B: Pavements, **145**(2): 04019014. doi:10.1061/JPEODX.0000121.

Gong, H., Sun, Y., Mei, Z., and Huang, B. 2018a. Improving accuracy of rutting prediction for mechanistic-empirical pavement design guide with deep neural networks. Construction and Building Materials, **190**: 710–718. doi:10.1016/j.conbuildmat.2018.09.087.

Gong, H., Sun, Y., Shu, X., and Huang, B. 2018b. Use of random forests regression for predicting IRI of asphalt pavements. Construction and Building Materials, **189**: 890–897. doi:10.1016/j.conbuildmat.2018.09.017.

Goodman, M.S. 2017. Biostatistics for Clinical and Public Health Research. *In* 1st edition.

Gopalakrishnan, K., Khaitan, S.K., Choudhary, A., and Agrawal, A. 2017. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. Construction and Building Materials, **157**: 322–330. doi:10.1016/j.conbuildmat.2017.09.110.

Guo, X., and Hao, P. 2021. Using a Random Forest Model to Predict the Location of Potential Damage on Asphalt Pavement. Applied Sciences, **11**(21): 10396. doi:10.3390/app112110396.

Haas, R., Hudson, W.R., and Falls, L.C. 2015. Pavement Asset Management. Wiley.

Haider, S.W., and Chatti, K. 2009. Effects of Design and Site Factors on Roughness Development in Flexible Pavements. Journal of Transportation Engineering, **135**(3): 112–120. doi:10.1061/(ASCE)0733-947X(2009)135:3(112).

Haldar, S.K. 2013. Statistical and Geostatistical Applications in Geology. : 19.

Hawkins, D.M. 2004. The Problem of Overfitting. Journal of Chemical Information and Computer Sciences, **44**(1): 1–12. doi:10.1021/ci0342472.

Hossain, M.I., Gopisetti, L.S.P., and Miah, M.S. 2019. International Roughness Index Prediction of Flexible Pavements Using Neural Networks. Journal of Transportation Engineering, Part B: Pavements, **145**(1): 04018058. doi:10.1061/JPEODX.0000088.

Hossein, P.-N. 2014. Introduction to Probability, Statistics, and Random Processes. Blue Bell, Kappa Research, LLC.

Howe, J.G., Gorsich, D.J., Scott, A.J., Lee, D., Liang, C., Chrstos, J., Myers, T., and Allen, R.W. 2003. Development of a Road/Terrain Characterization Rating Tool. SAE International,: 12.

Huang, Y.H. 2003. Pavement analysis and design. Pearson.

Hudson, W.R., Haas, R.C.G., and Pedigo, R.D. 1979. Pavement management system development. Transportation Research Board, National Research Council, Washington, D.C.

IBM Cloud Education. 2020. Machine Learning. Available from https://www.ibm.com/cloud/learn/machine-learning.

IPCC. 2002. Climate Change 2001: The Scientific Basis. Foreign Affairs, **81**(1). doi:10.2307/20033020.

IPCC. 2021. IPCC: Regional fact sheet – North and Central America - sixth assessment report - Working Group I – The Physical Science Basis. IPCC- Intergovernmental panel on climate change:

James, G., Witten, D., Hastie, T., and Tibshirani, R. 2013. An Introduction to Statistical Learning. Springer New York, New York, NY.

Joni, H.H., Hilal, M.M., and Abed, M.S. 2020a. Developing International Roughness Index (IRI) Model from visible pavement distresses. IOP Conference Series: Materials Science and Engineering, **737**(1): 012119. doi:10.1088/1757-899X/737/1/012119.

Joni, H.H., Hilal, M.M., and Abed, M.S. 2020b. Developing International Roughness Index (IRI) Model from visible pavement distresses. IOP Conference Series: Materials Science and Engineering, **737**: 012119. doi:10.1088/1757-899X/737/1/012119.

Jurgens, R., and Chan, J. 2005. Highway Performance Measures for Business Plans in Alberta. Transportation Association of Canada (TAC),.

Kargah-Ostadi, N., Stoffels, S.M., and Tabatabaee, N. 2010. Network-Level Pavement Roughness Prediction Model for Rehabilitation Recommendations. Transportation Research Record: Journal of the Transportation Research Board, **2155**(1): 124–133. doi:10.3141/2155-14.

Kaya, O., Ceylan, H., Kim, S., Waid, D., and Moore, B.P. 2020. Statistics and Artificial Intelligence-Based Pavement Performance and Remaining Service Life Prediction Models for Flexible and Composite Pavement Systems. Transportation Research Record: Journal of the Transportation Research Board, **2674**(10): 448–460. doi:10.1177/0361198120915889.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Proceedings of the 31st International Conference on Neural Information Processing Systems,: 3149–3157. doi:10.5555/3294996.3295074.

Khattak, M.J., Nur, M.A., Bhuyan, M.R.-U.-K., and Gaspard, K. 2014. International roughness index models for HMA overlay treatment of flexible and composite pavements. International Journal of Pavement Engineering, **15**(4): 334–344. doi:10.1080/10298436.2013.842237.

Kim, H.-K., Lim, Y., Tafesse, M., Kim, G.M., and Yang, B. 2022. Micromechanics-integrated machine learning approaches to predict the mechanical behaviors of concrete containing crushed clay brick aggregates. Construction and Building Materials, **317**: 125840. doi:10.1016/j.conbuildmat.2021.125840.

Kim, S.-J., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D. 2007. An Interior-Point Method for Large-Scale -Regularized Least Squares. IEEE Journal of Selected Topics in Signal Processing, **1**(4): 606–617. doi:10.1109/JSTSP.2007.910971.

Kutay, M.E. 2007. Spectral Analysis of Factors Affecting Roughness in Flexible Pavements. Paris, France.

Lever, J., Krzywinski, M., and Altman, N. 2016. Model selection and overfitting. Nature Methods, **13**(9): 703–704. doi:10.1038/nmeth.3968.

Lin, J.-D., Yau, J.-T., and Hsiao, L.-H. 2003. Correlation Analysis Between International Roughness Index (IRI) and Pavement Distress by Neural Network. Transportation Research Board,: 22.

Louppe, G., and Geurts, P. (*Editors*). 2012. Ensembles on Random Patches. Springer Berlin Heidelberg, Berlin, Heidelberg.

Lu, Z., Feng, Z., Yao, D., Li, X., Jiao, X., and Zheng, K. 2021. Bonding performance between ultra-high performance concrete and asphalt pavement layer. Construction and Building Materials, **312**: 125375. doi:10.1016/j.conbuildmat.2021.125375.

Luo, X., Wang, F., Bhandari, S., Wang, N., and Qiu, X. 2021. Effectiveness evaluation and influencing factor analysis of pavement seal coat treatments using random forests. Construction and Building Materials, **282**: 122688. doi:10.1016/j.conbuildmat.2021.122688.

Mabrouk, G.M., Elbagalati, O.S., Dessouky, S., Fuentes, L., and Walubita, L.F. 2022. Using ANN modeling for pavement layer moduli backcalculation as a function of traffic speed deflections. Construction and Building Materials, **315**: 125736. doi:10.1016/j.conbuildmat.2021.125736.

Mactutis, J.A., Alavi, S.H., and Ott, W.C. 2000. Investigation of Relationship Between Roughness and Pavement Surface Distress Based on WesTrack Project. Transportation Research Record: Journal of the Transportation Research Board, **1699**(1): 107–113. doi:10.3141/1699-15.

Marani, A., and Nehdi, M.L. 2020. Machine learning prediction of compressive strength for phase change materials integrated cementitious composites. Construction and Building Materials, **265**: 120286. doi:10.1016/j.conbuildmat.2020.120286.

Marsland, S. 2014. Machine Learning An Algorithmic Perspective. CRC Press Taylor & Francis Group.

Mary, T. 2020. The Difference Between Artificial Intelligence, Machine Learning and Deep Learning. Available from https://community.intel.com/t5/Blogs/Tech-Innovation/Artificial-Intelligence-AI/The-Difference-Between-Artificial-Intelligence-Machine-Learning/post/1335666#:~:text=Machine%20learning%20is%20a%20class,outcomes%20on%20similar%20novel%20data.&text=Artificial%20neurons%20can%20be%20arranged,many%20layers%20of%20artificial%20neurons.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B. 2021. IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC.

Matin, A., Collas, P., Blain, D., Ha, C., Liang, C., MacDonald, L., McKibbon, S., Palmer, C., and Rhoades, K. 2004. Canada's Greenhouse Gas Inventory. Greenhouse Gas Division Environment Canada.

Mazari, M., and Rodriguez, D.D. 2016. Prediction of pavement roughness using a hybrid gene expression programming-neural network technique. Journal of Traffic and Transportation Engineering (English Edition), **3**(5): 448–455. doi:10.1016/j.jtte.2016.09.007.

Mei, Q., and Gül, M. 2020. A cost effective solution for pavement crack inspection using cameras and deep neural networks. Construction and Building Materials, **256**: 119397. doi:10.1016/j.conbuildmat.2020.119397.

Microsoft. 2021. Light Gradient Boosting Machine.

Miller, J.S., and Bellinger, W.Y. 2014. Distress Identification Manual for the Long-Term Pavement Performance Program. Federal Highway Administration.

Moges, M., Moges, M., Ayed, A., Viecili, G., and Eng, M. 2017. A Review and Recommendations for Canadian LCCA Guidelines. Transportation Association of Canada, St. John's, NL. p. 20.

Mooi, E., and Sarstedt, M. 2011. A Concise Guide to Market Research. *In* 2nd edition. Springer Berlin Heidelberg, Berlin, Heidelberg.

Nassiri, S., Shafiee, M.H., and Bayat, A. 2013. Development of Roughness Prediction Models Using Alberta Transportation's Pavement Management System. International Journal of Pavement Research and Technology, **6**(6): 714–720. doi:http://dx.doi.org/10.6135/ijprt.org.tw/2013.6(6).714.

National Cooperative Highway Research Program. 1997. Appendix B Distresses. Transportation Research Board.

NCSS. 2021. Stepwise Regression.

Olive, D.J. 2014. Statistical Theory and Inference. Springer International Publishing, Cham.

Owolabi, A.O., Sadiq, O.M., and Abiola, O.S. 2012. Development of Performance Models for a Typical Flexible Road Pavement in Nigeria. International Journal for Traffic and Transport Engineering, **2**(3): 178–184. doi:10.7708/ijtte.2012.2(3).02.

Paterson, W.D.O. 1986. International Roughness Index: Relationship to Other Measures of Roughness and Riding Quality. Transportation Research Board, (1084).

Peng, Y., and Unluer, C. 2022. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. Construction and Building Materials, **316**: 125785. doi:10.1016/j.conbuildmat.2021.125785.

Pereira Dias, P., Bhagya Jayasinghe, L., and Waldmann, D. 2021. Machine learning in mix design of Miscanthus lightweight concrete. Construction and Building Materials, **302**: 124191. doi:10.1016/j.conbuildmat.2021.124191.

Perera, R.W., and Al-Rawashdeh, A. 2017. Investigation of Increase in Roughness Due to Environmental Factors in Flexible Pavements Using Profile Data from Long-term Pavement Performance Specific Pavement Studies 1 Experiment. Federal Highway Administration (FHWA).

Perera, R.W., and Kohn, S.D. 2001. LTPP Data Analysis: Factors Affecting Pavement Smoothness. National Cooperative Highway Research Program Transportation Research Board National Research Council.

Peterson, D.E. 1987. Pavement management practices. Transportation Research Board, National Research Council, Washington, D.C.

Piryonesi, S.M., and El-Diraby, T. 2021. Climate change impact on infrastructure: A machine learning solution for predicting pavement condition index. Construction and Building Materials, **306**: 124905. doi:10.1016/j.conbuildmat.2021.124905.

Puccinelli, J., and Jackson, N. 2007. Development of Pavement Performance Models to Account for Frost Effects and Their Application to Mechanistic–Empirical Design Guide Calibration. Transportation Research Record: Journal of the Transportation Research Board, **1990**(1): 95–101. doi:10.3141/1990-11.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rahman, S., Bhasin, A., and Smit, A. 2021. Exploring the use of machine learning to predict metrics related to asphalt mixture performance. Construction and Building Materials, **295**: 123585. doi:10.1016/j.conbuildmat.2021.123585.

Robbins, D.M.M., and Tran, D.N.H. 2016. A synthesis report: value of pavement smoothness and ride quality to roadway users and the impact of pavement roughness on vehicle operating costs. : 20.

Sadat Hosseini, A., Hajikarimi, P., Gandomi, M., Moghadas Nejad, F., and Gandomi, A.H. 2021. Optimized machine learning approaches for the prediction of viscoelastic behavior of modified asphalt binders. Construction and Building Materials, **299**: 124264. doi:10.1016/j.conbuildmat.2021.124264.

Sarstedt, M., and Mooi, E. 2011. A Concise Guide to Market Research. *In* 2nd edition. Springer Berlin Heidelberg.

Sayers, M.W. 1995a. On the Calculation of International Roughness Index from Longitudinal Road Profile. Transportation Research Board, (1501): 12.

Sayers, M.W. 1995b. On the Calculation of International Roughness Index from Longitudinal Road Profile. Transportation Research Record, (1501): 12.

Sayers, M.W., Gillespie, T.D., and Paterson, W.D.O. 1986. Guidelines for conducting and calibrating road roughness measurements. World Bank, Washington, D.C., U.S.A.

Schumacker, R., and Tomek, S. 2013. Understanding Statistics Using R. Springer New York, New York, NY.

Shah, M.I., Javed, M.F., Aslam, F., and Alabduljabbar, H. 2022. Machine learning modeling integrating experimental analysis for predicting the properties of sugarcane bagasse ash concrete. Construction and Building Materials, **314**: 125634. doi:10.1016/j.conbuildmat.2021.125634.

Smola, A.J., and Schölkopf, B. 2004. A tutorial on support vector regression. Statistics and Computing, **14**(3): 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

Soleymani, H.R., Palsat, D., Mesher, D., and Campbell, P. 2008. The Effect of Pavement Crack Treatments on IRI and Surface Profile - A Case Study in Alberta.

Song, H., Ahmad, A., Farooq, F., Ostrowski, K.A., Maślak, M., Czarnecki, S., and Aslam, F. 2021. Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. Construction and Building Materials, **308**: 125021. doi:10.1016/j.conbuildmat.2021.125021.

Spangler, E.B., and Kelly, W.J. 1964. GMR Road Profilometer-A Method for Measuring Road Profile. Highway Research Record,: 28.

Sun, Z., Liu, H., Huyan, J., Li, W., Guo, M., Hao, X., and Pei, L. 2021. Assessment of importance-based machine learning feature selection methods for aggregate size distribution measurement in a 3D binocular vision system. Construction and Building Materials, **306**: 124894. doi:10.1016/j.conbuildmat.2021.124894.

Sweet, S.A., and Grace-Martin, K.A. 2010. Data Analysis with SPSS: A First Course in Applied Statistics. Pearson.

Terzi, S. 2007. Modeling the pavement serviceability ratio of flexible highway pavements by artificial neural networks. Construction and Building Materials, **21**(3): 590–593. doi:10.1016/j.conbuildmat.2005.11.001.

Transportation Research Board and National Academies of Sciences, Engineering, and Medicine. 2005. LTPP Data Analysis: Influence of Design and Construction Features on the Response and Performance of New Flexible and Rigid Pavements. Transportation Research Board, Washington, D.C.

de Ville, B. 2013. Decision trees: Decision trees. Wiley Interdisciplinary Reviews: Computational Statistics, **5**(6): 448–455. doi:10.1002/wics.1278.

Willmott, C.J., and Matsuura, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Inter-Research, **30**(1): 79–82. doi:10.3354/cr030079.

Zhang, X., Otto, F., and Oeser, M. 2021. Pavement moduli back-calculation using artificial neural network and genetic algorithms. Construction and Building Materials, **287**: 123026. doi:10.1016/j.conbuildmat.2021.123026.

Zhang, Z. 2014. Too much covariates in a multivariable model may cause the problem of overfitting. Journal of Thoracic Disease, **6**(9). doi:10.3978/j.issn.2072-1439.2014.08.33.

Zhao, H., Wu, D., Zeng, M., Tian, Y., and Ling, J. 2019. Assessment of concrete pavement support conditions using distributed optical vibration sensing fiber and a neural network. Construction and Building Materials, **216**: 214–226. doi:10.1016/j.conbuildmat.2019.04.195.

Ziari, H., Maghrebi, M., Ayoubinejad, J., and Waller, S.T. 2016a. Prediction of Pavement Performance: Application of Support Vector Regression with Different Kernels. Transportation Research Record: Journal of the Transportation Research Board, **2589**(1): 135–145. doi:10.3141/2589-15.

Ziari, H., Sobhani, J., Ayoubinejad, J., and Hartmann, T. 2016b. Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods. International Journal of Pavement Engineering, **17**(9): 776–788. doi:10.1080/10298436.2015.1019498.

Appendix A: Alberta Transportation PMS Dataset used in this Research

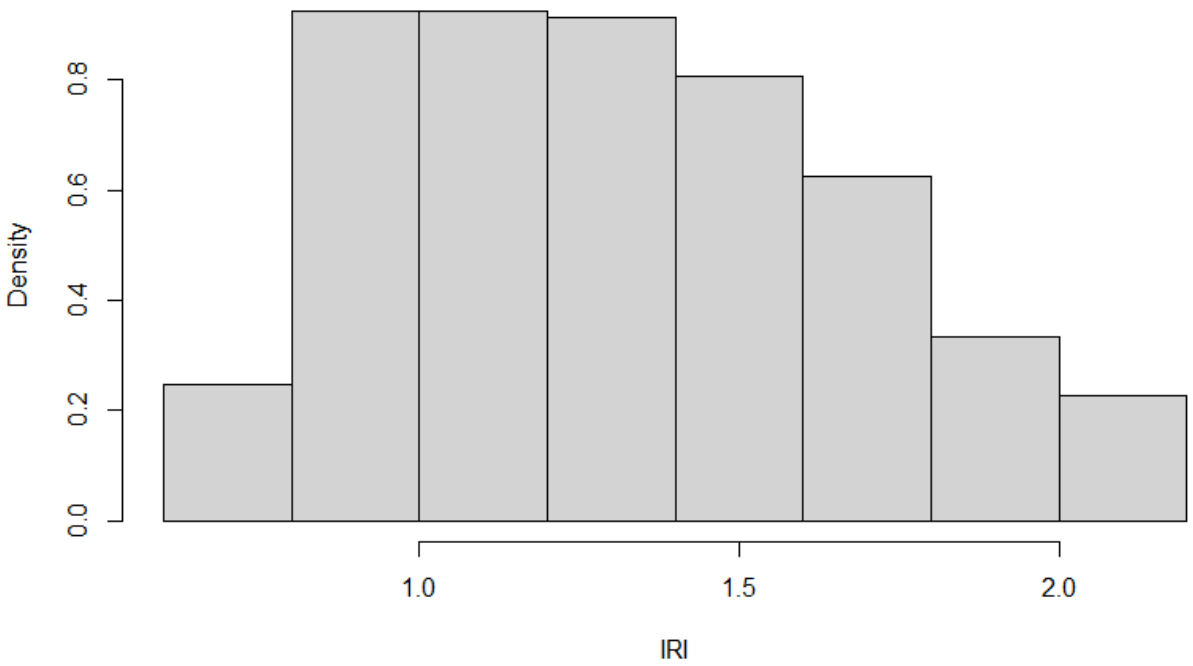

Figure A.1 – *IRI* values Boxplot for ACP pavements – inner-fence Dataset

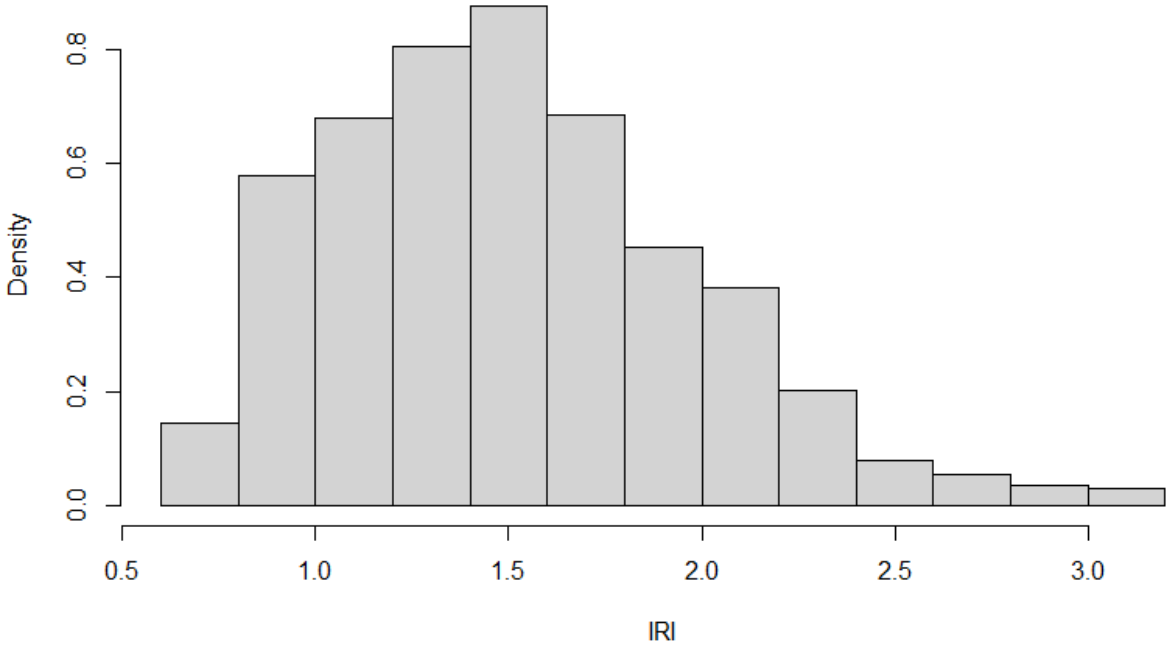

Figure A.2 – *IRI* values Boxplot for ACP pavements – Outer-fence Dataset

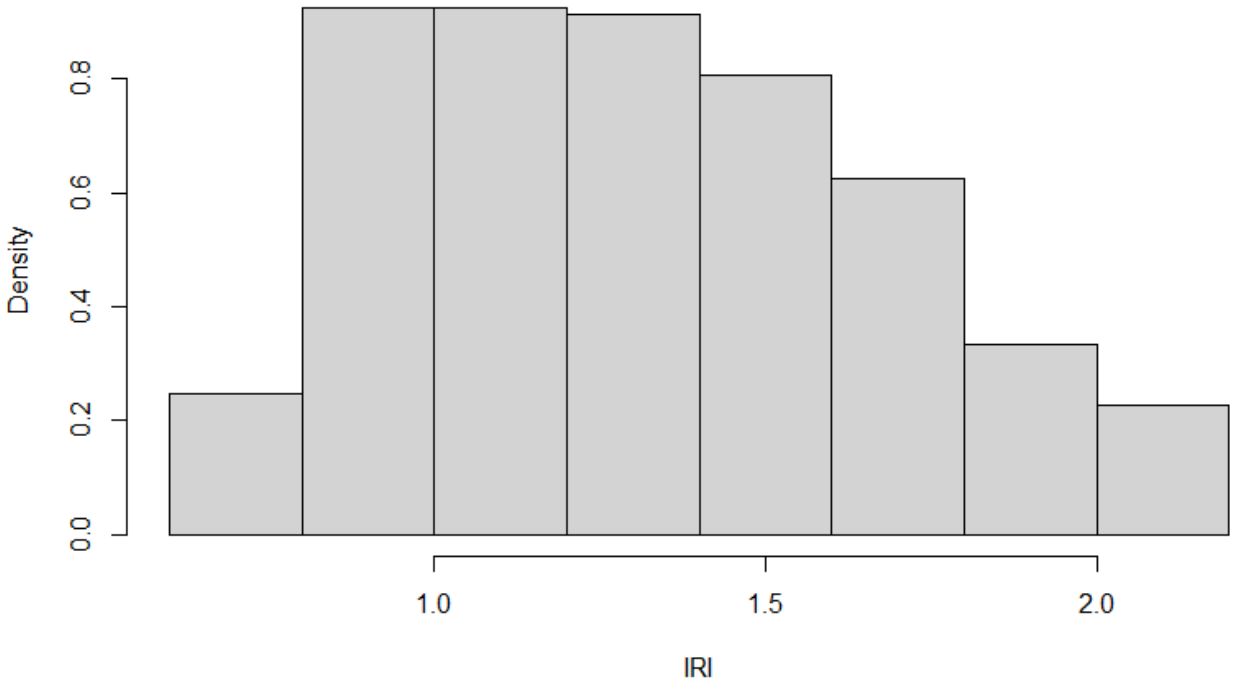

Figure A.3 – *IRI* values Boxplot for ACP pavements – Full Dataset

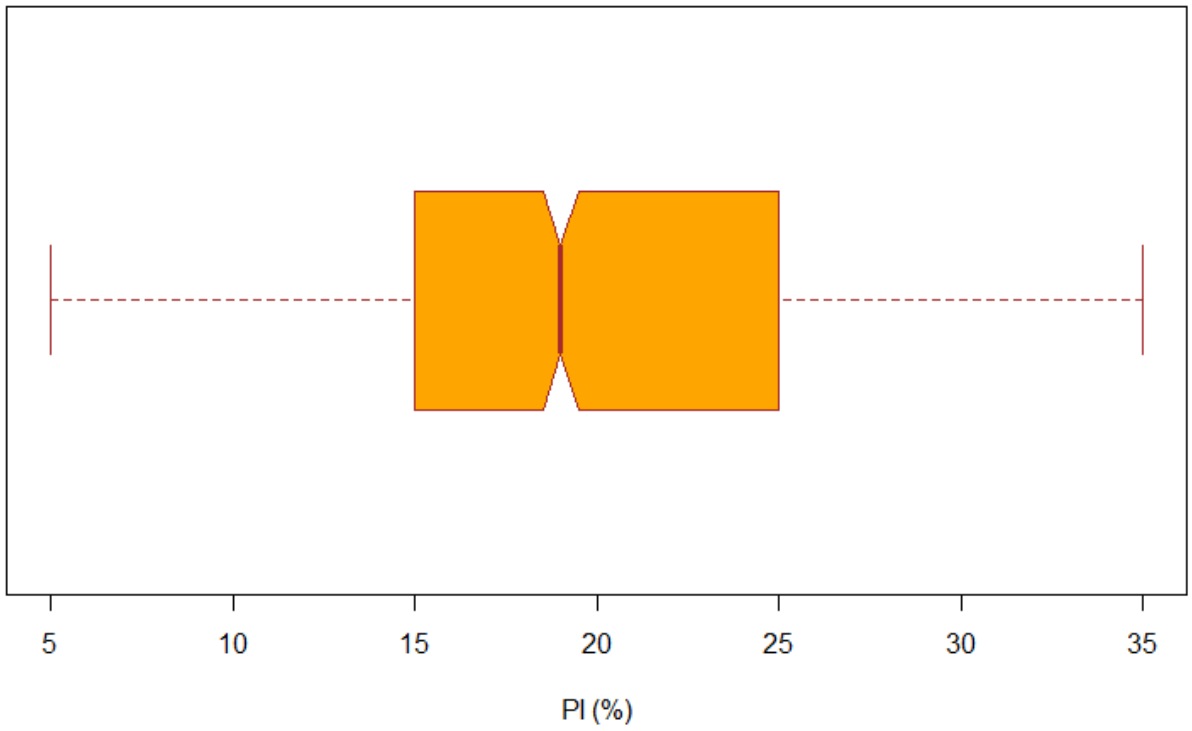

Figure A.4 – *PI* values Boxplot for ACP pavements – Full Dataset

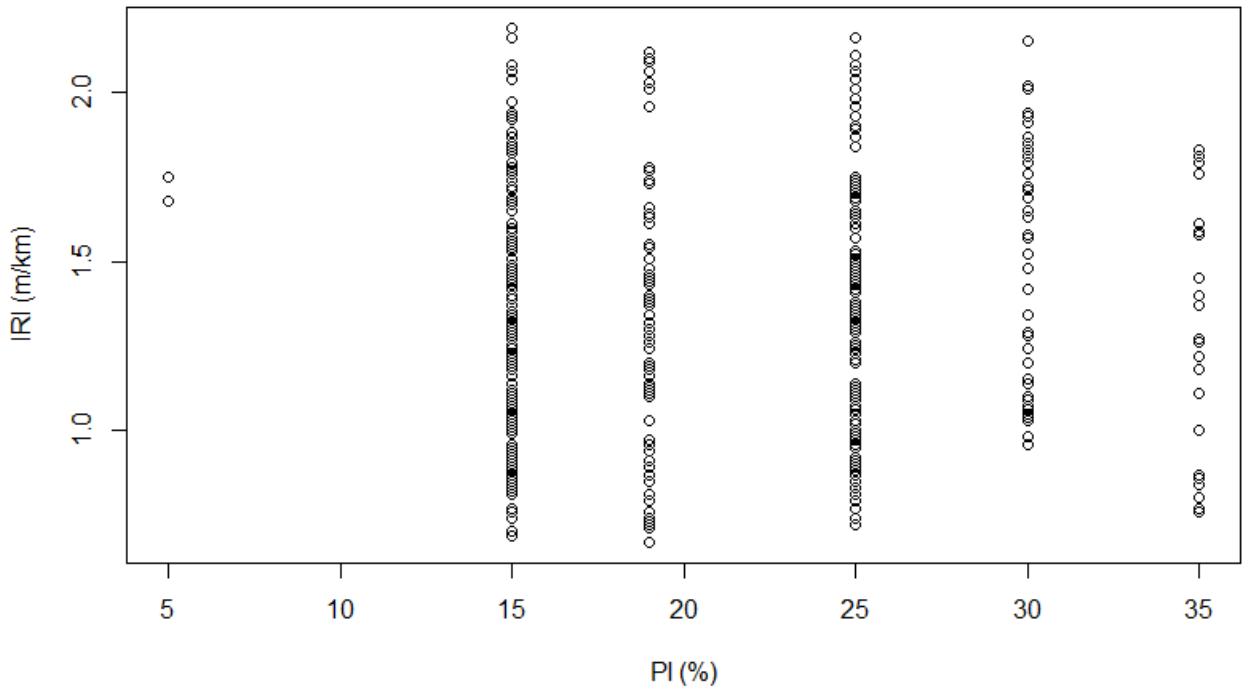

Figure A.5 – *PI* values scatter plot for ACP pavements – inner-fence Dataset

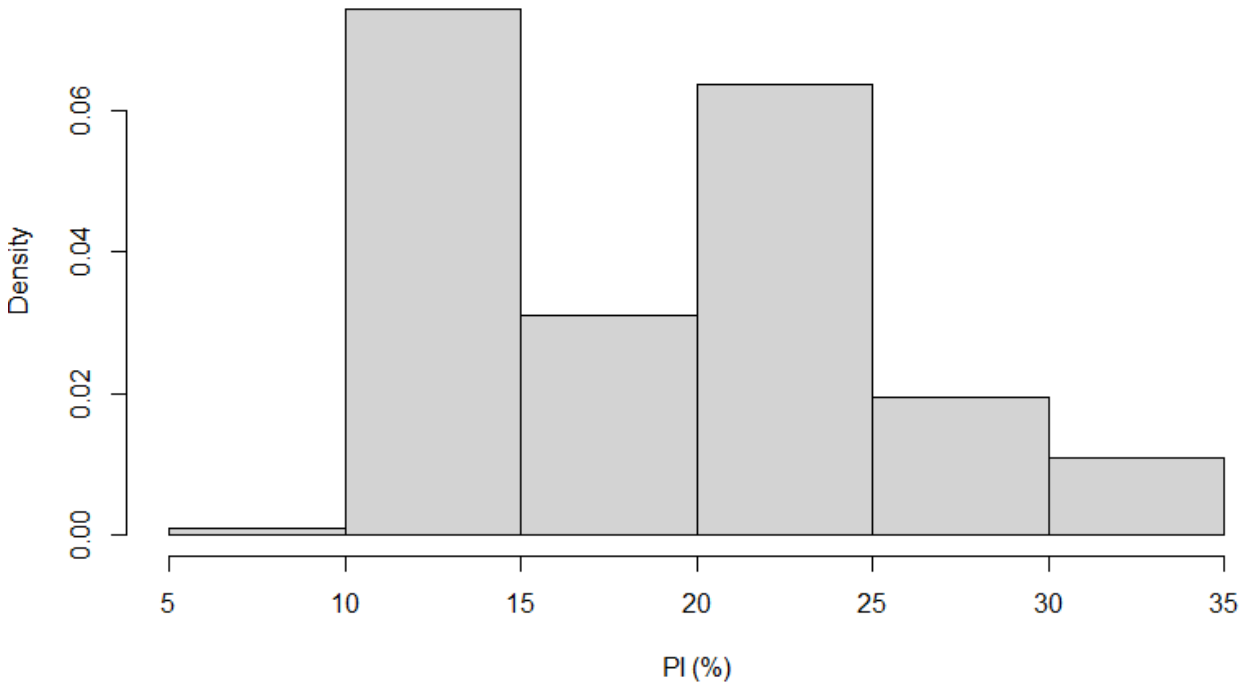

Figure A.6 – *PI* values histogram plot for ACP pavements – inner-fence Dataset

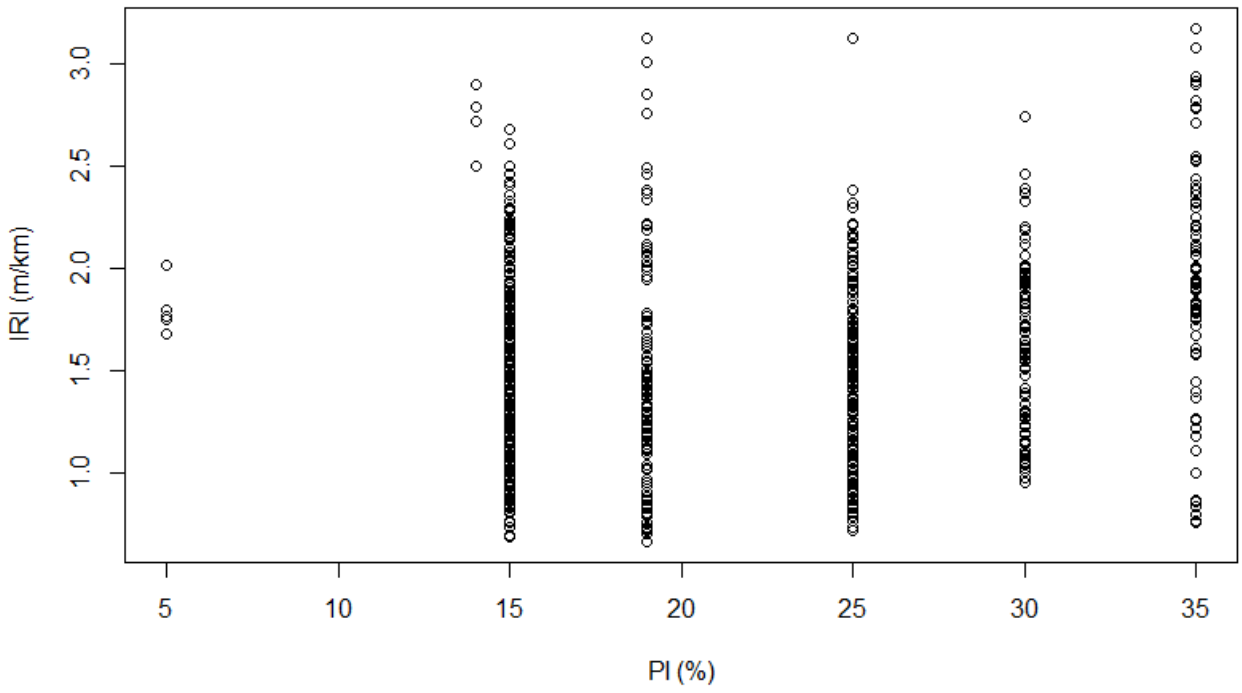

Figure A.7 – *PI* values scatter plot for ACP pavements – Outer-fence Dataset

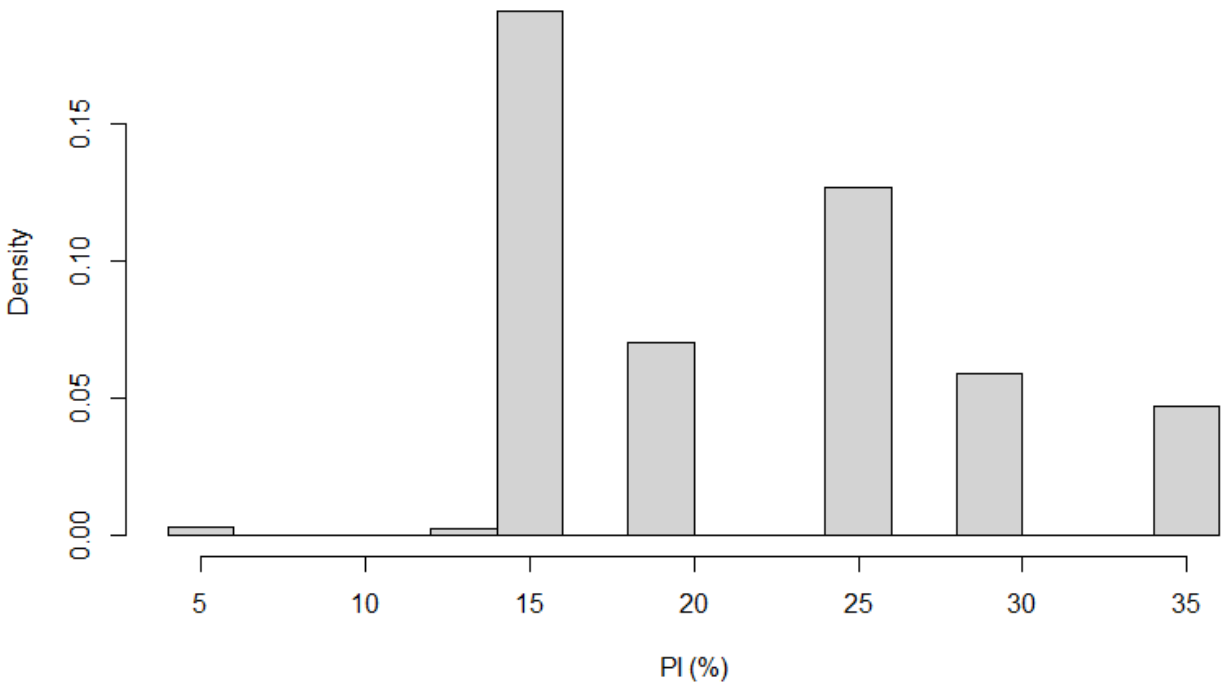

Figure A.8 – *PI* values histogram plot for ACP pavements – Outer-fence Dataset

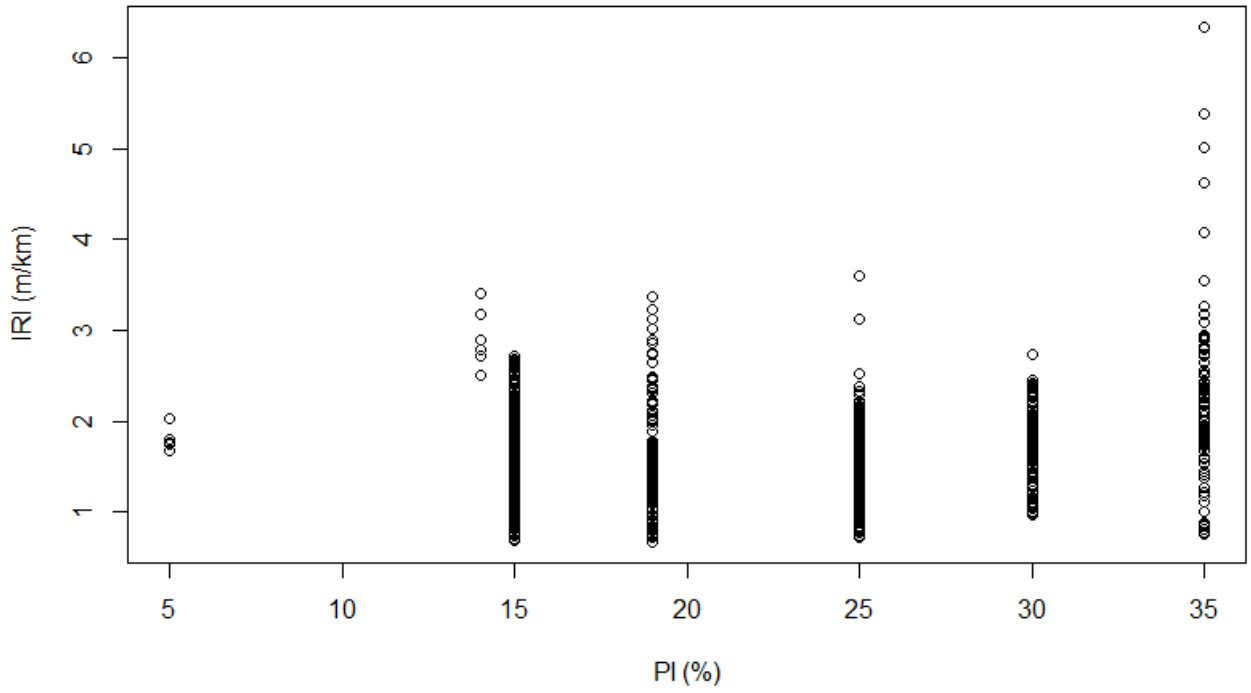

Figure A.9 – *PI* values scatter plot for ACP pavements – Full Dataset

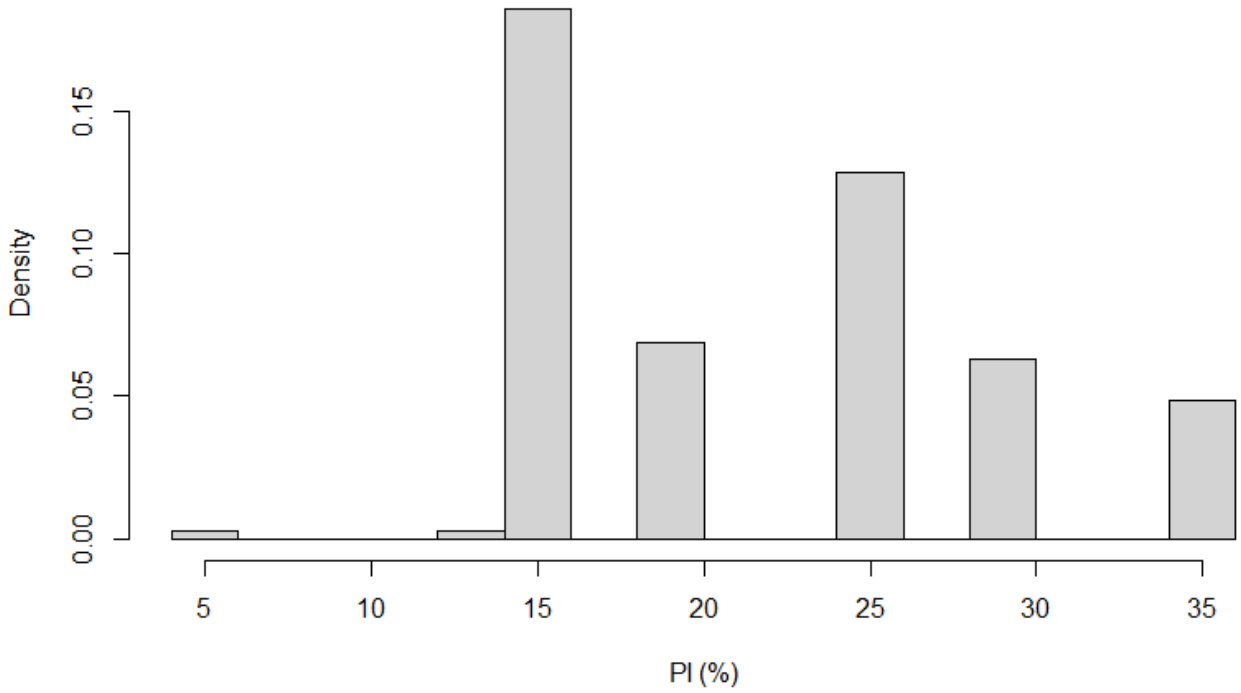

Figure A.10 – *PI* values histogram plot for ACP pavements – Full Dataset

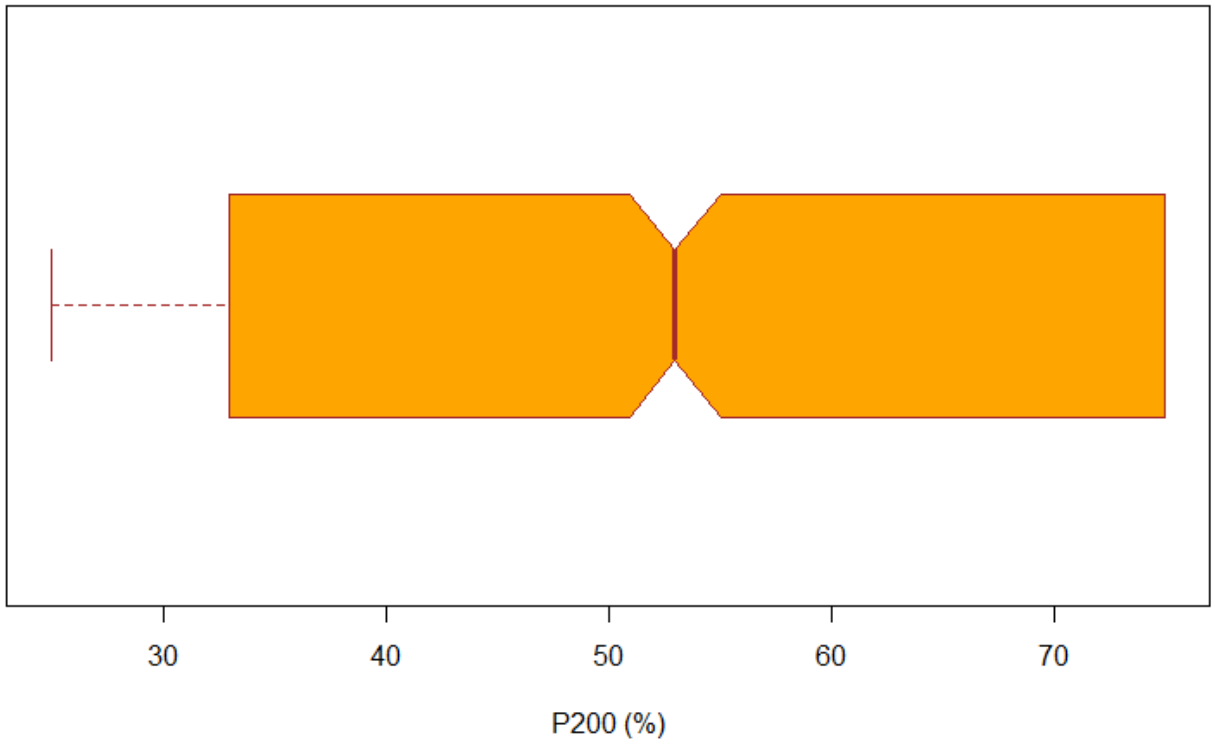

Figure A.11 – P_{200} values Boxplot for ACP pavements – Full Dataset

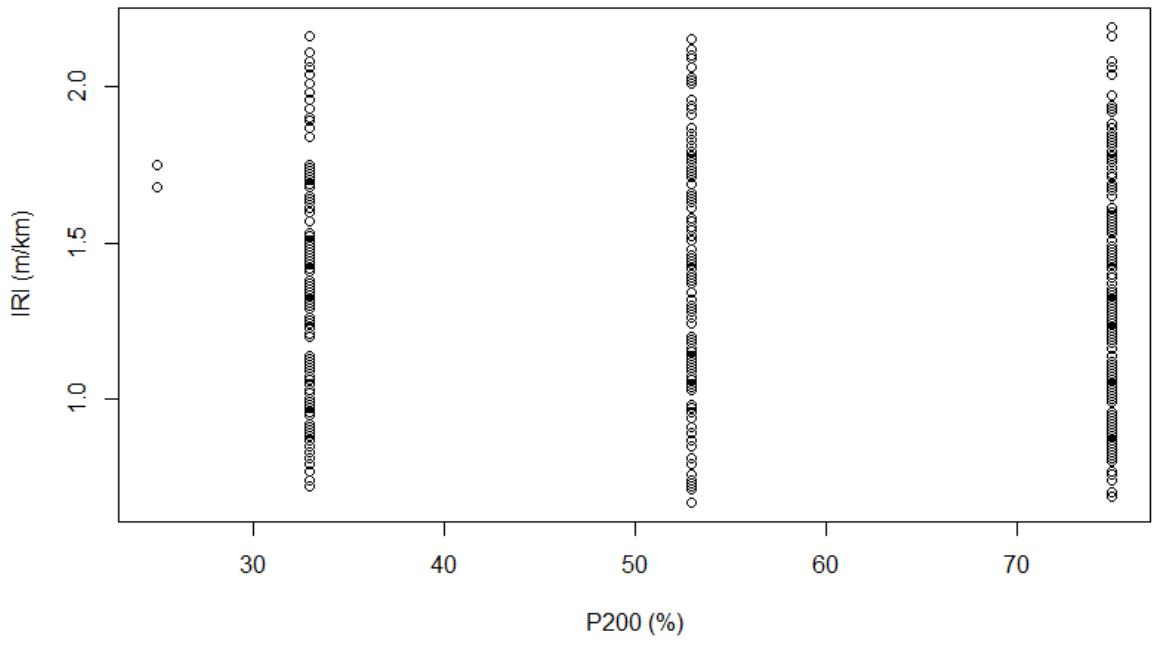

Figure A.12 – P_{200} values scatter plot for ACP pavements – inner-fence Dataset

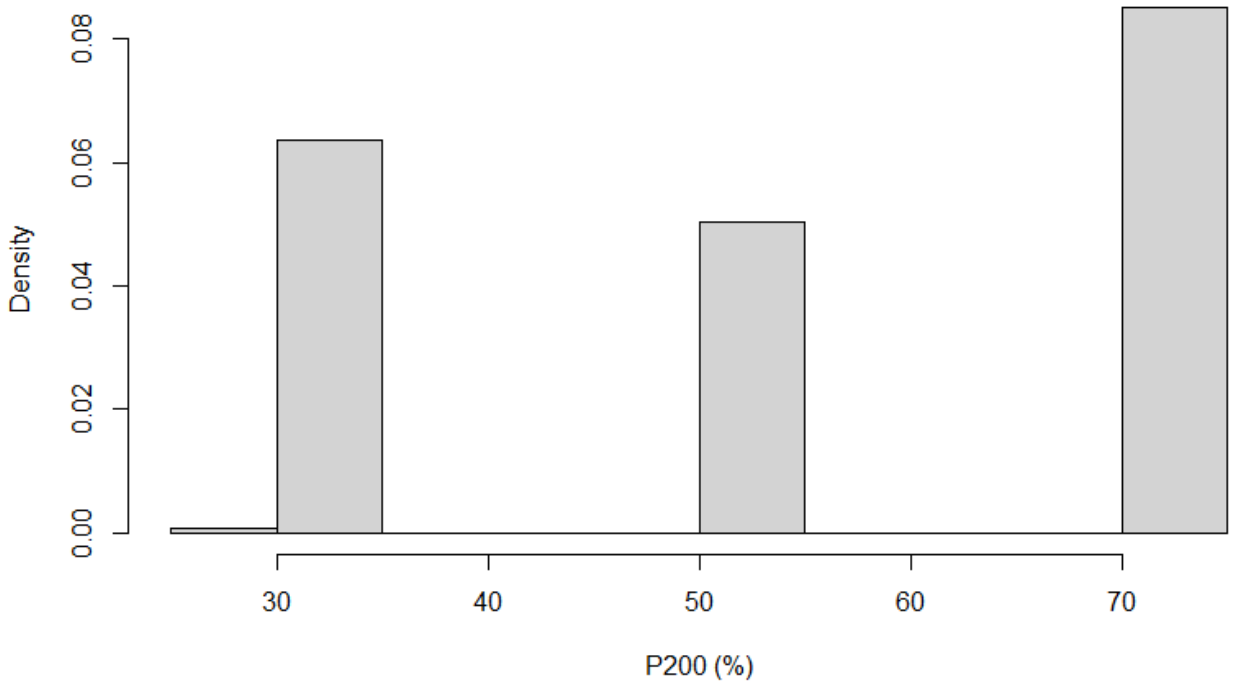

Figure A.13 – P_{200} values histogram plot for ACP pavements – inner-fence Dataset

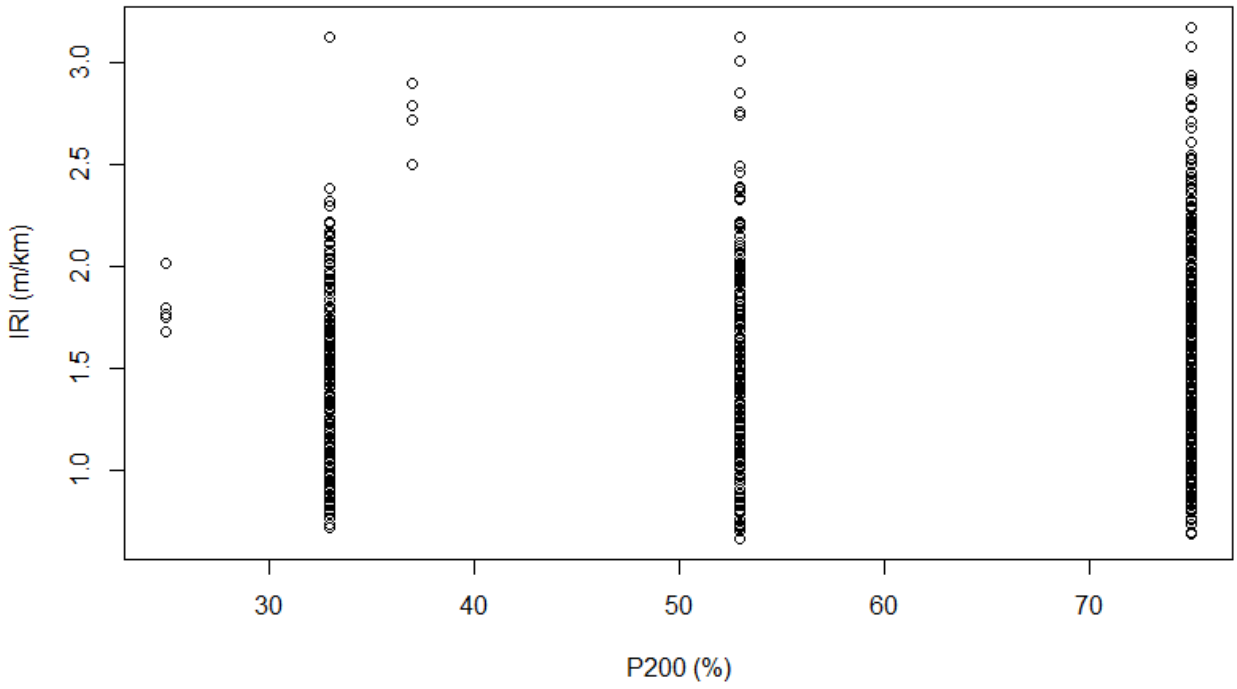

Figure A.14 – P_{200} values scatter plot for ACP pavements – Outer-fence Dataset

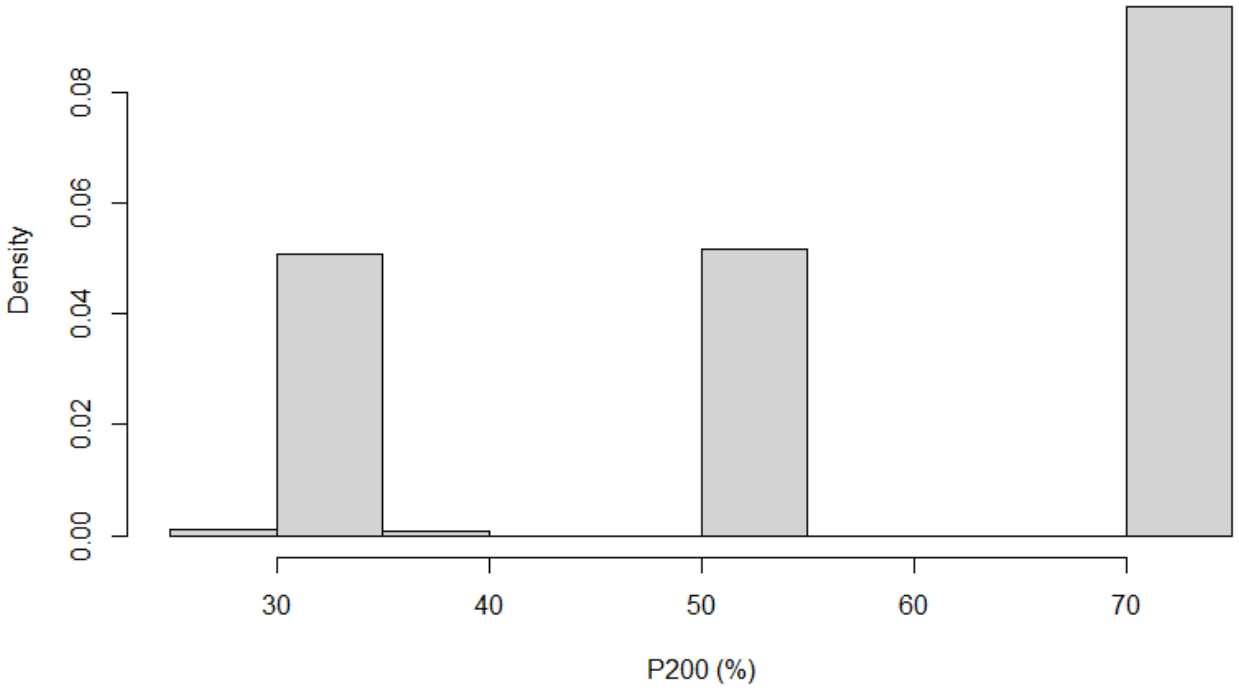

Figure A.15 – P_{200} values histogram plot for ACP pavements – Outer-fence Dataset

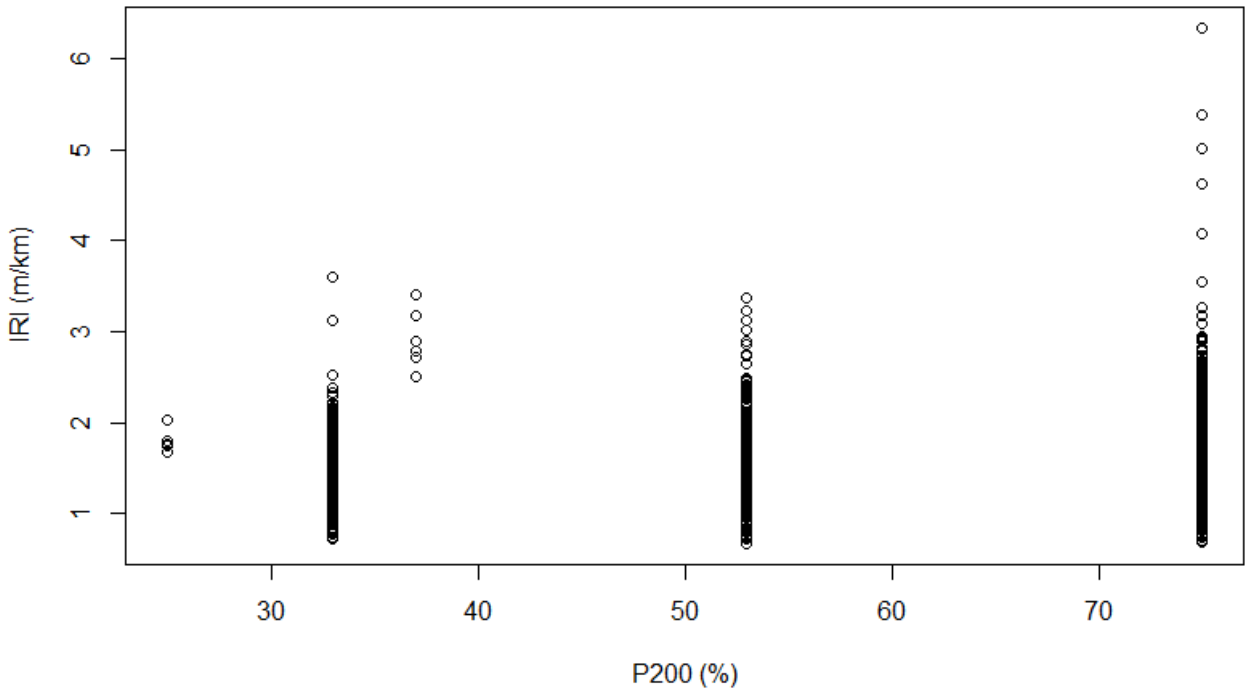

Figure A.16 – P_{200} values scatter plot for ACP pavements – Full Dataset

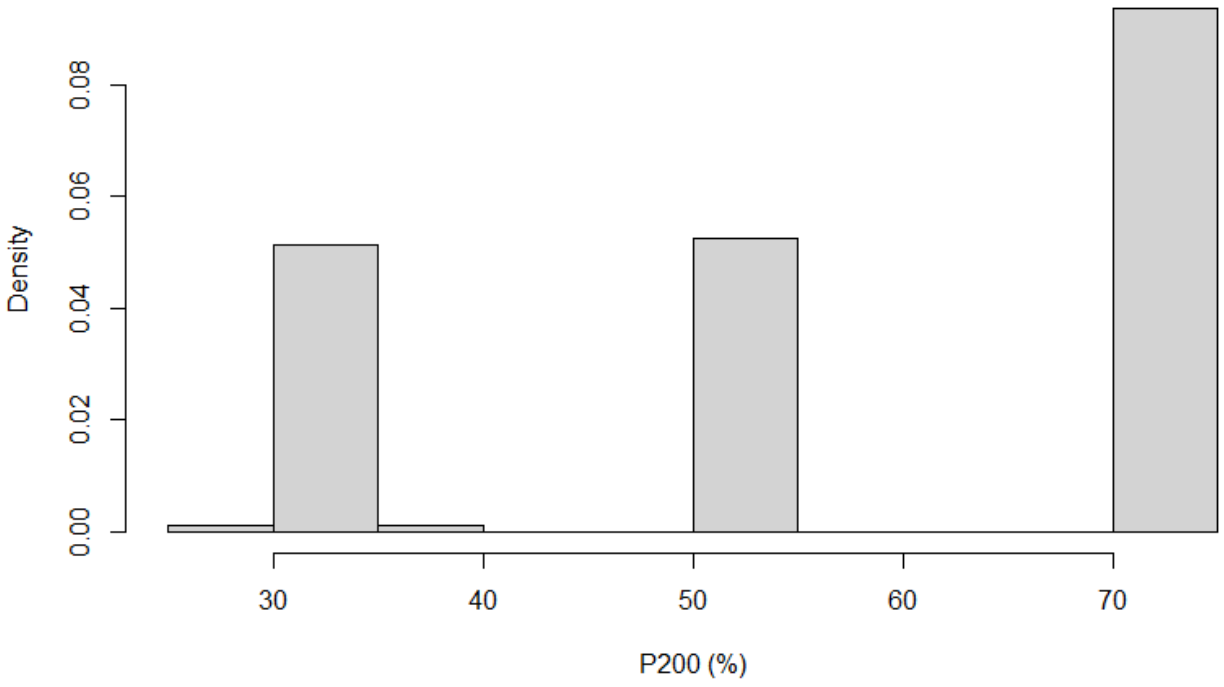

Figure A.17 – P_{200} values histogram plot for ACP pavements – Full Dataset

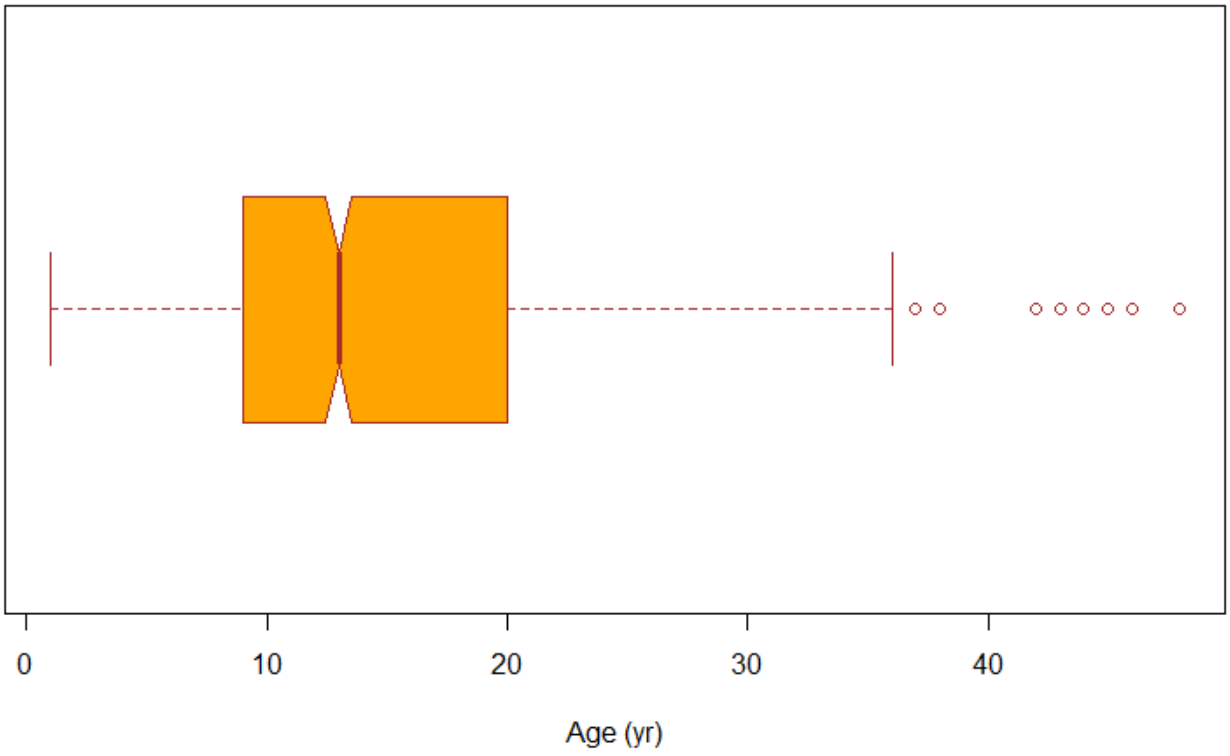

Figure A.18 – *Age* values Boxplot for ACP pavements – Full Dataset

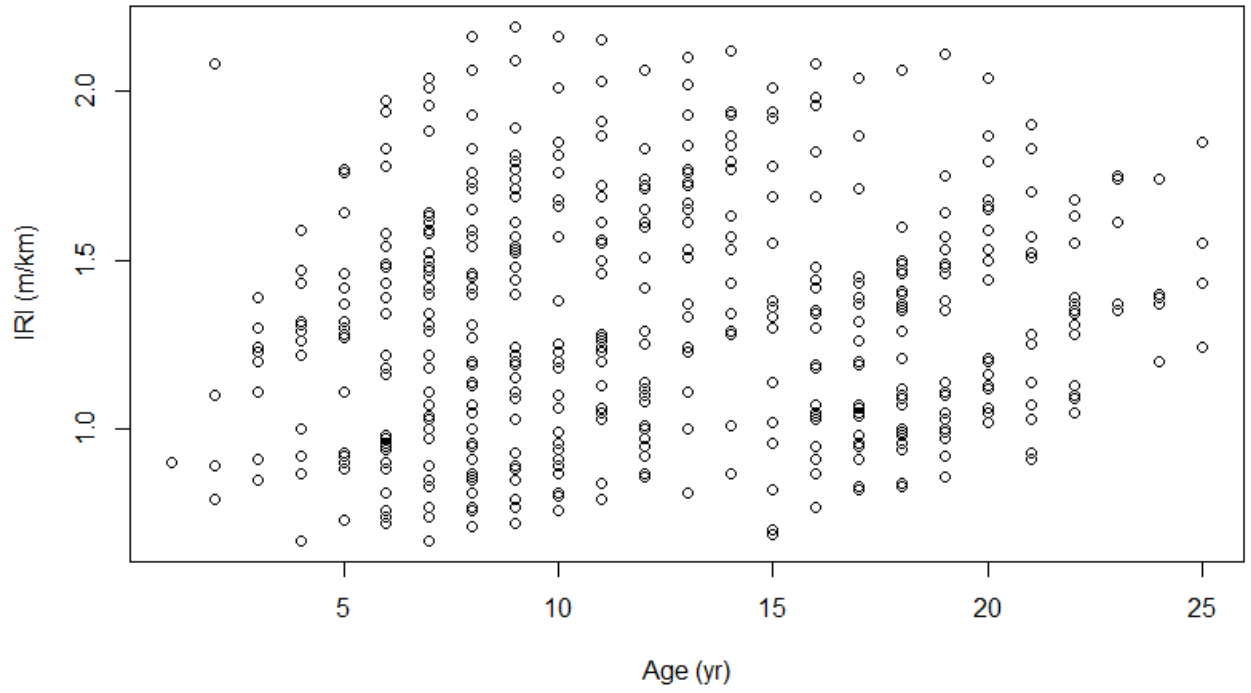

Figure A.19 – *Age* values scatter plot for ACP pavements – inner-fence Dataset

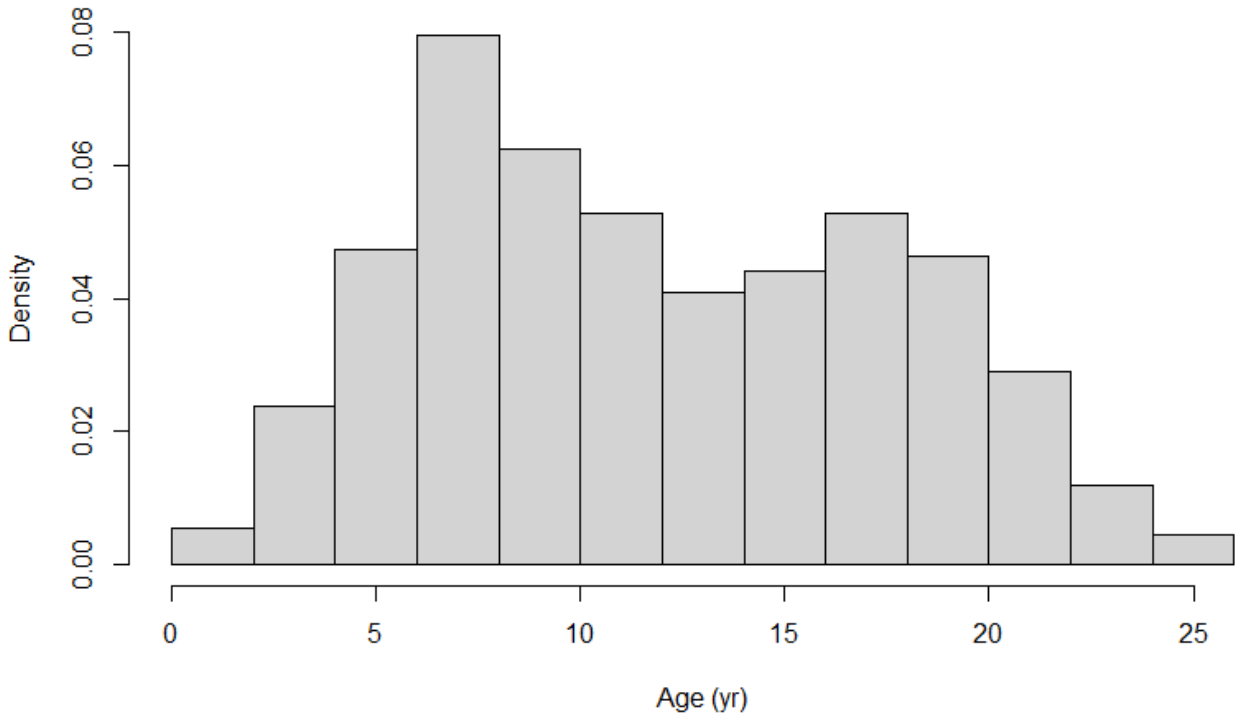

Figure A.20 – *Age* values histogram plot for ACP pavements – inner-fence Dataset

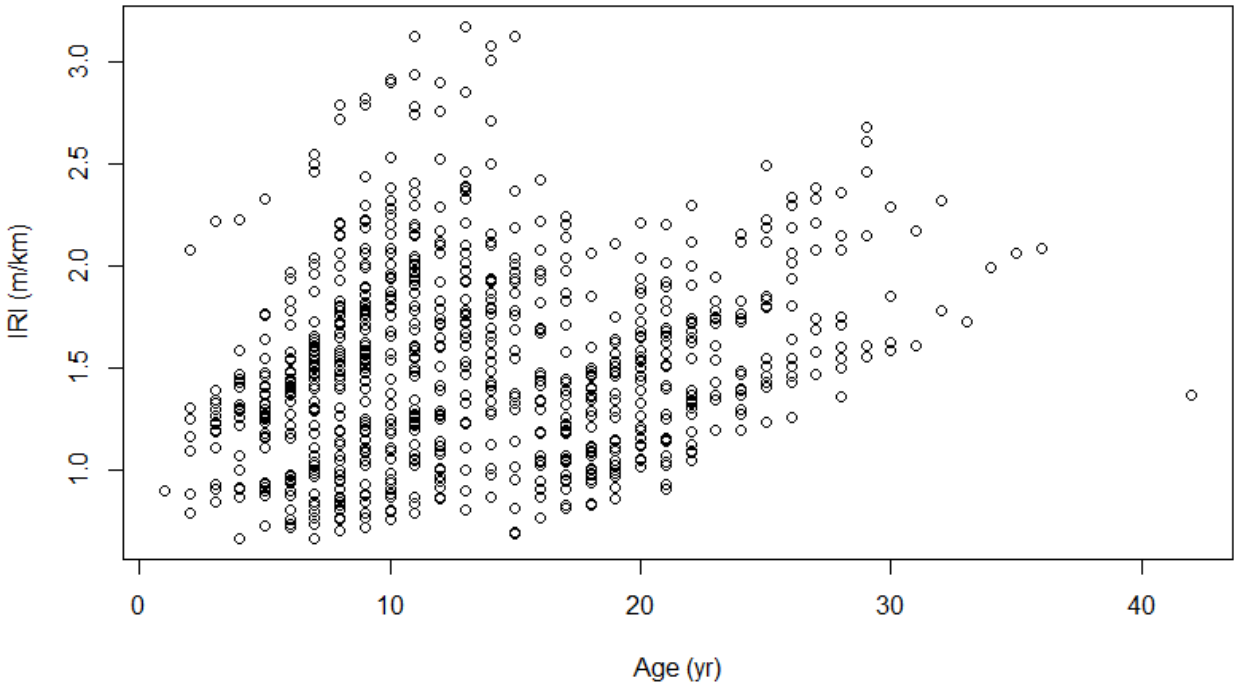

Figure A.21 – *Age* values scatter plot for ACP pavements – Outer-fence Dataset

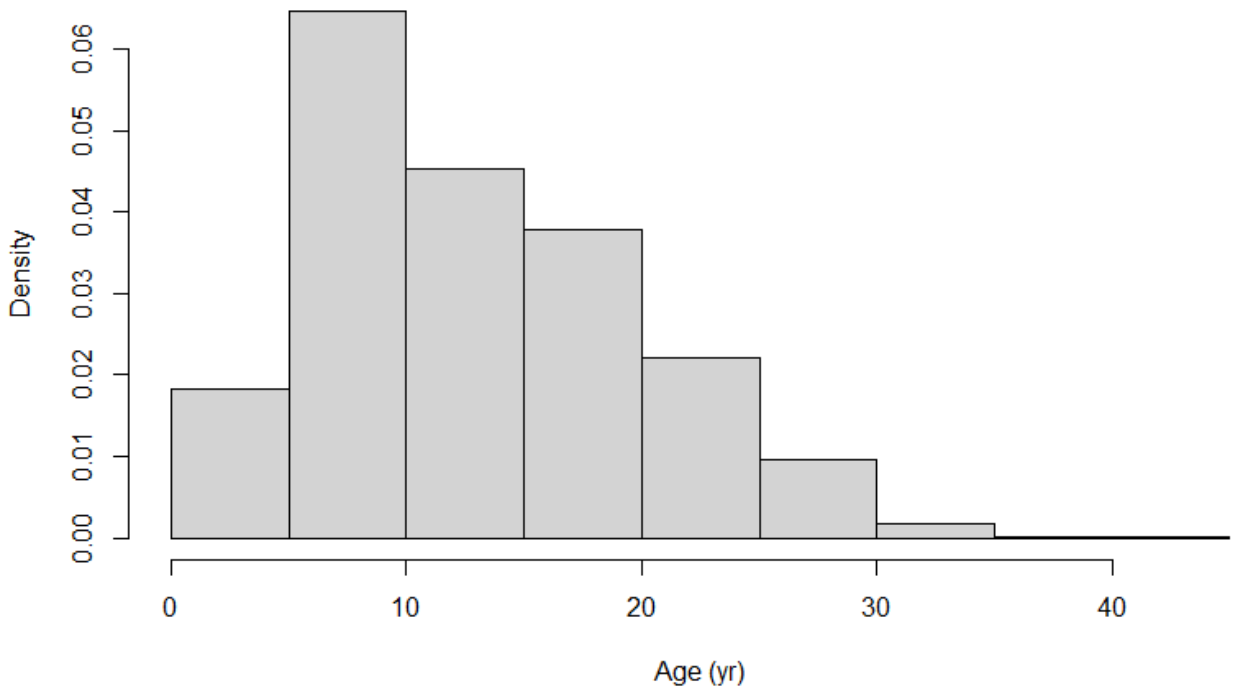

Figure A.22 – *Age* values histogram plot for ACP pavements – Outer-fence Dataset

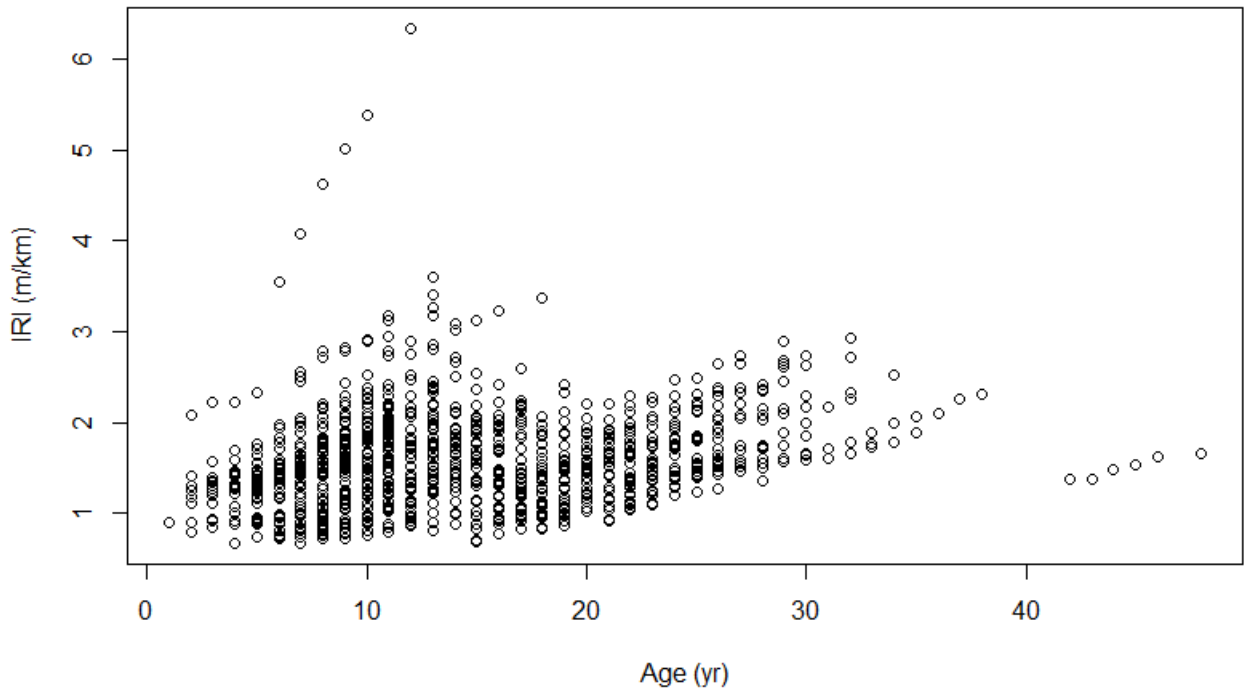

Figure A.23 – *Age* values scatter plot for ACP pavements – Full Dataset

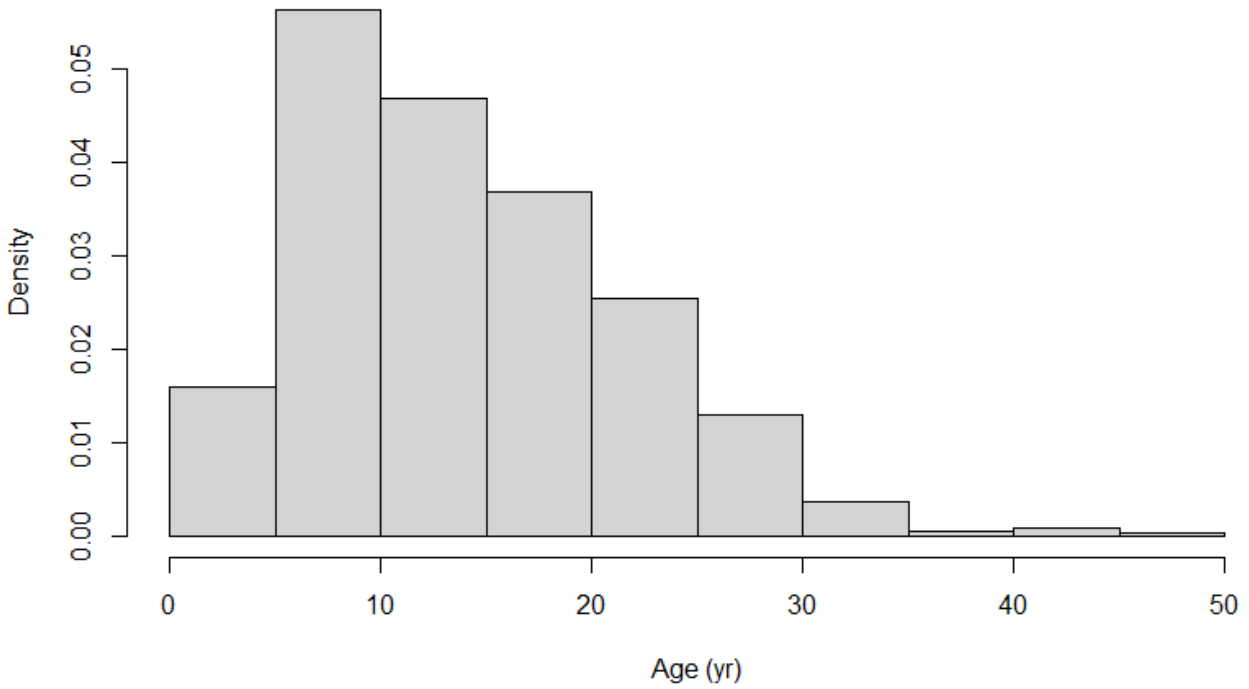

Figure A.24 – *Age* values histogram plot for ACP pavements – Full Dataset

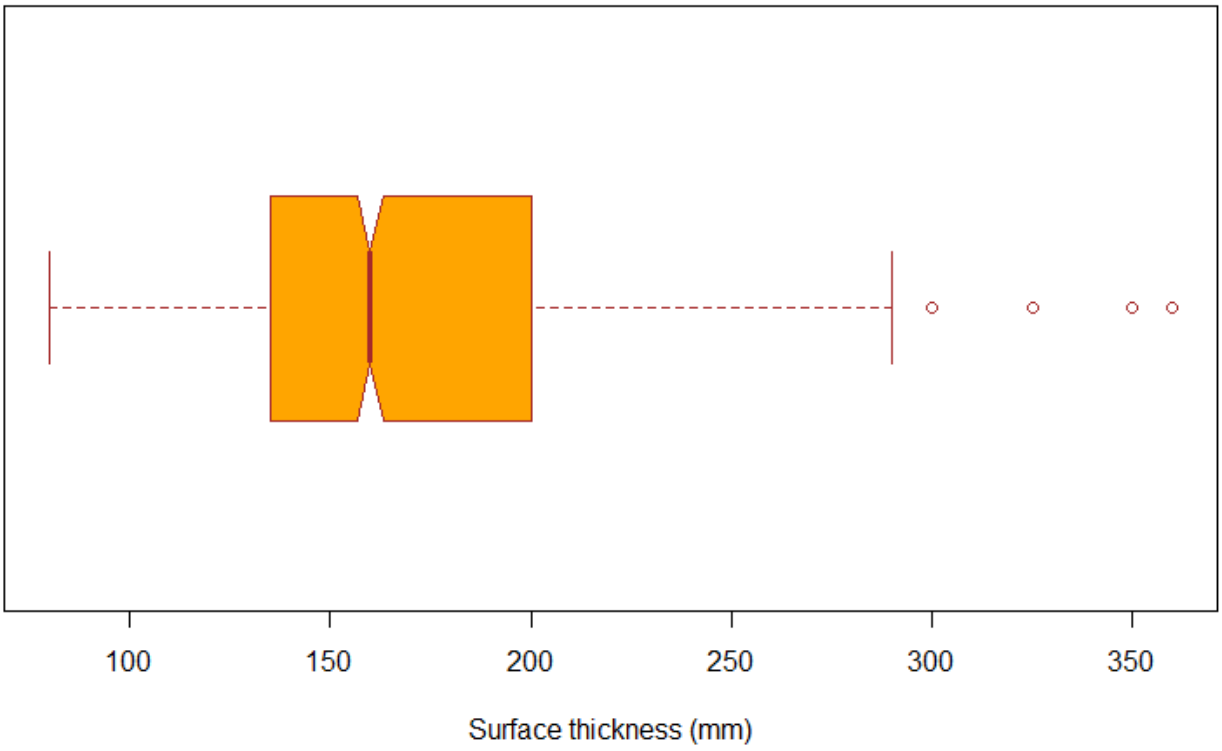

Figure A.25 – *Surfthickness* values Boxplot for ACP pavements – Full Dataset

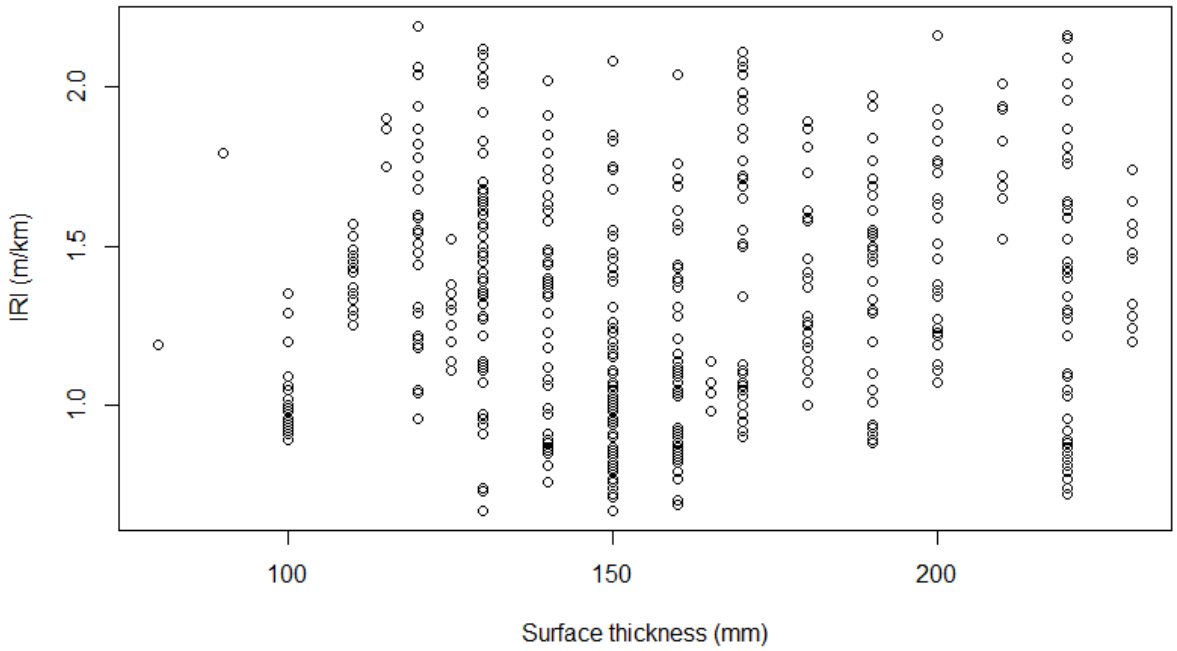

Figure A.26 – *Surfthickness* values scatter plot for ACP pavements – inner-fence Dataset

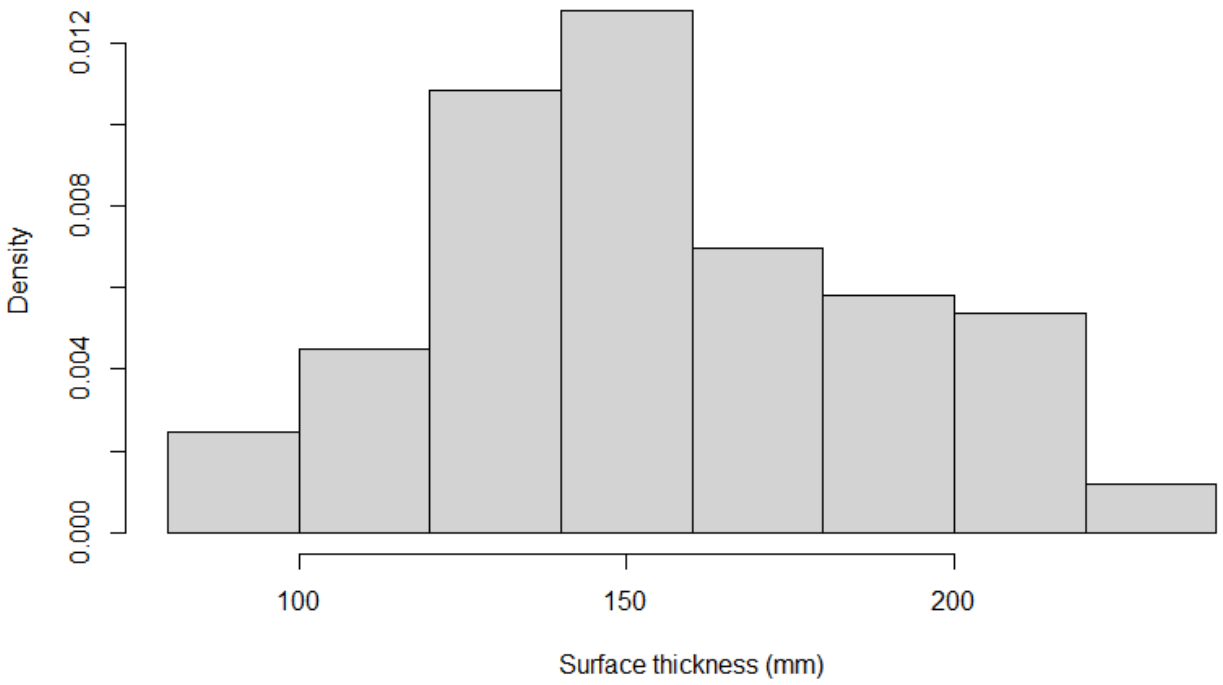

Figure A.27– *Surfthickness* values histogram plot for ACP pavements – inner-fence Dataset

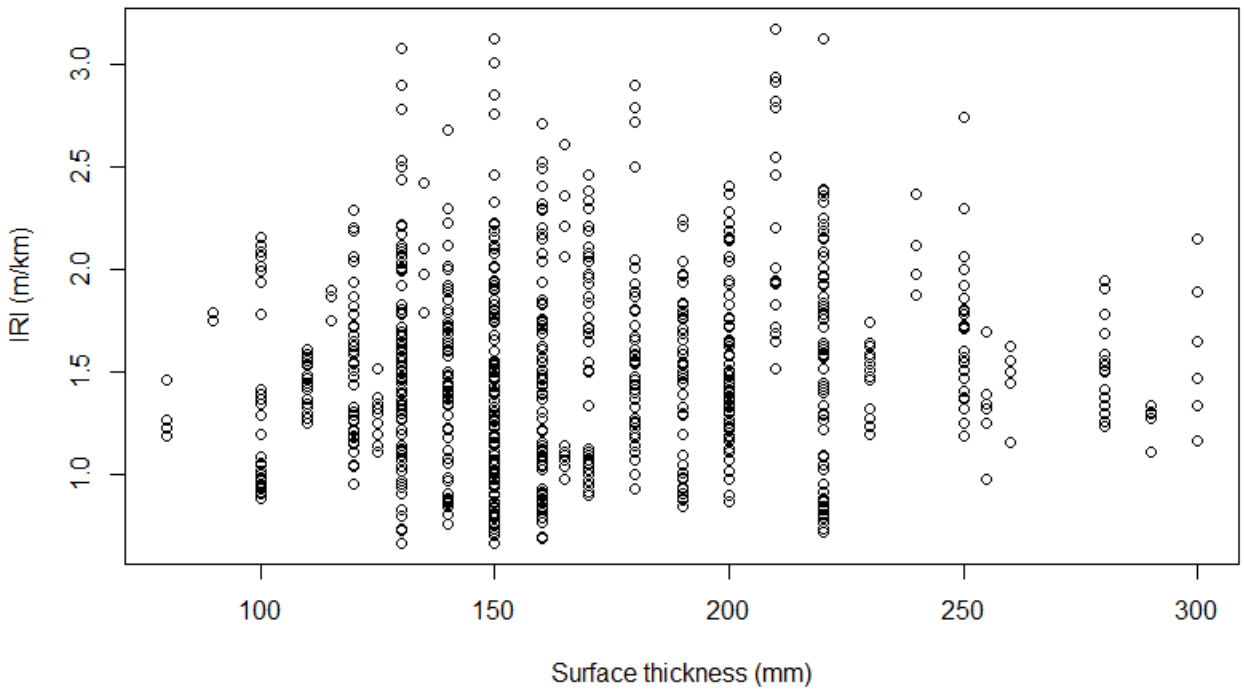

Figure A.28 – *Surfthickness* values scatter plot for ACP pavements – Outer-fence Dataset

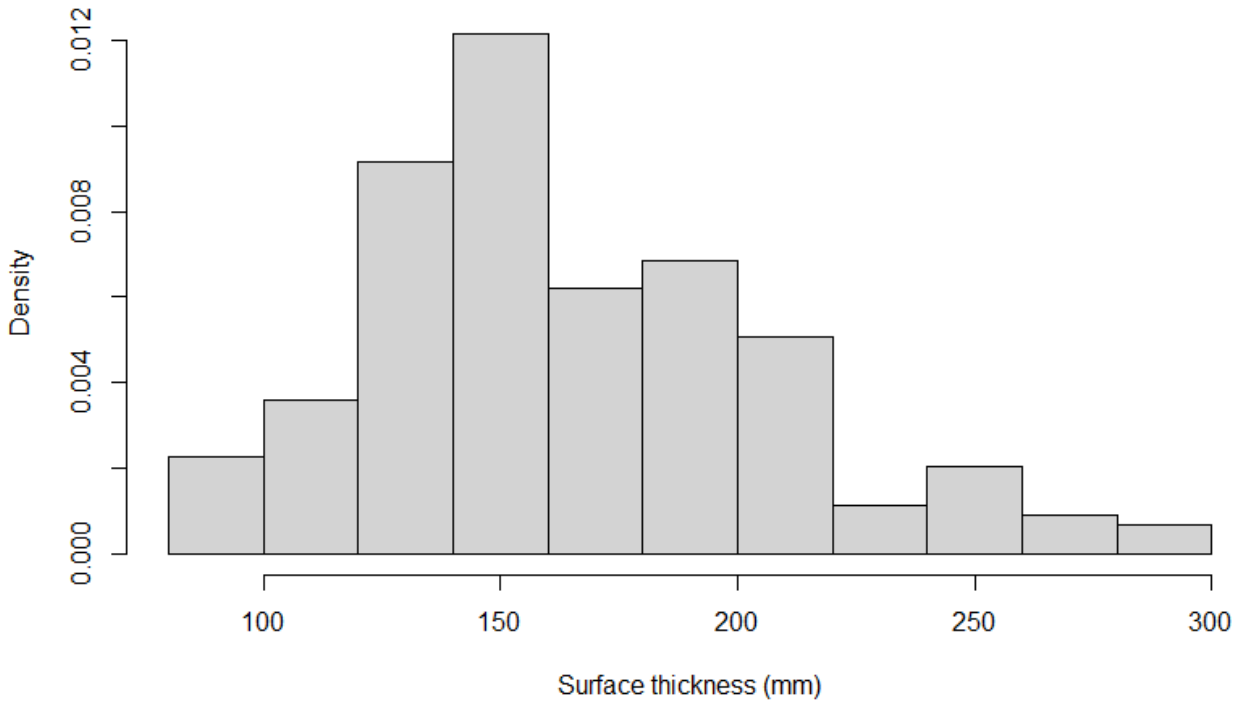

Figure A.29 – *Surfthickness* values histogram plot for ACP pavements – Outer-fence Dataset

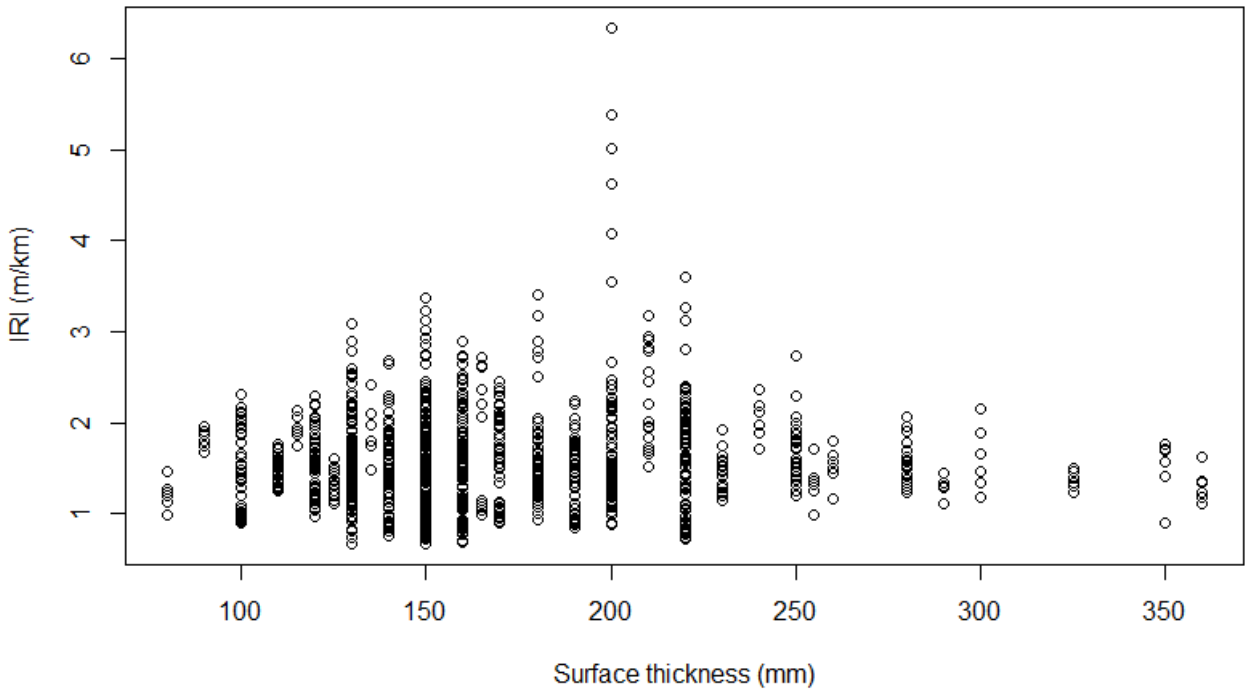

Figure A.30 – *Surfthickness* values scatter plot for ACP pavements – Full Dataset

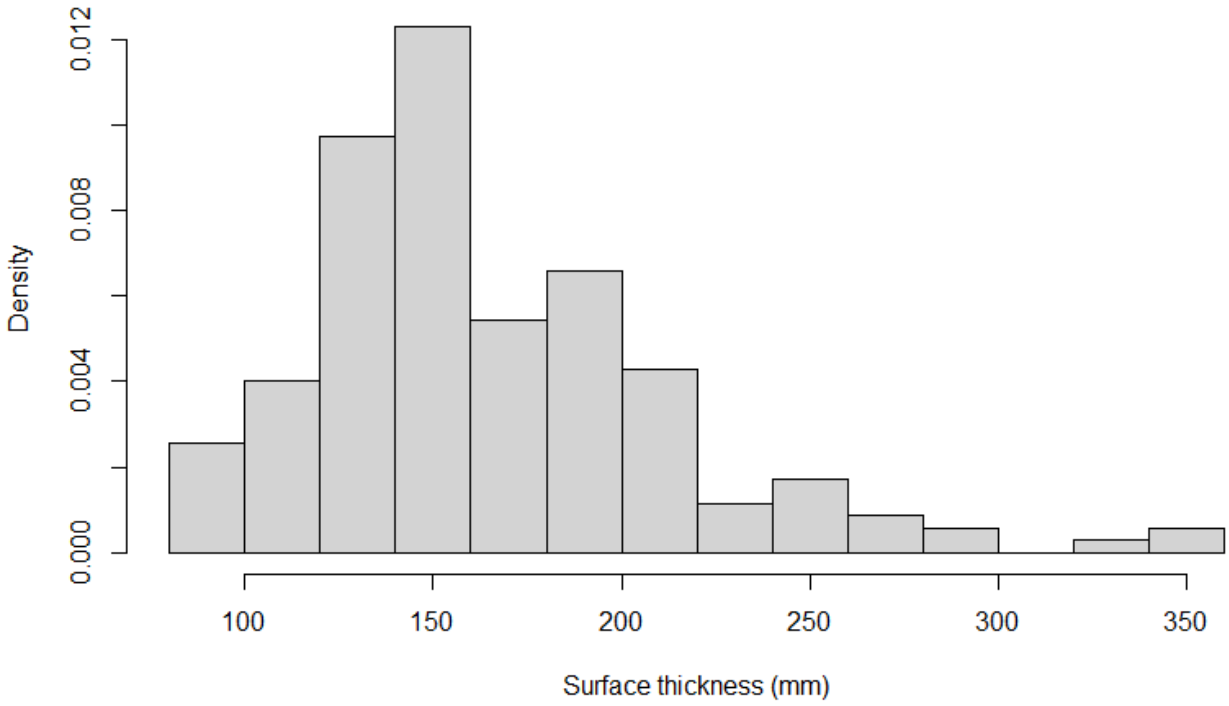

Figure A.31 – *Surfthickness* values histogram plot for ACP pavements – Full Dataset

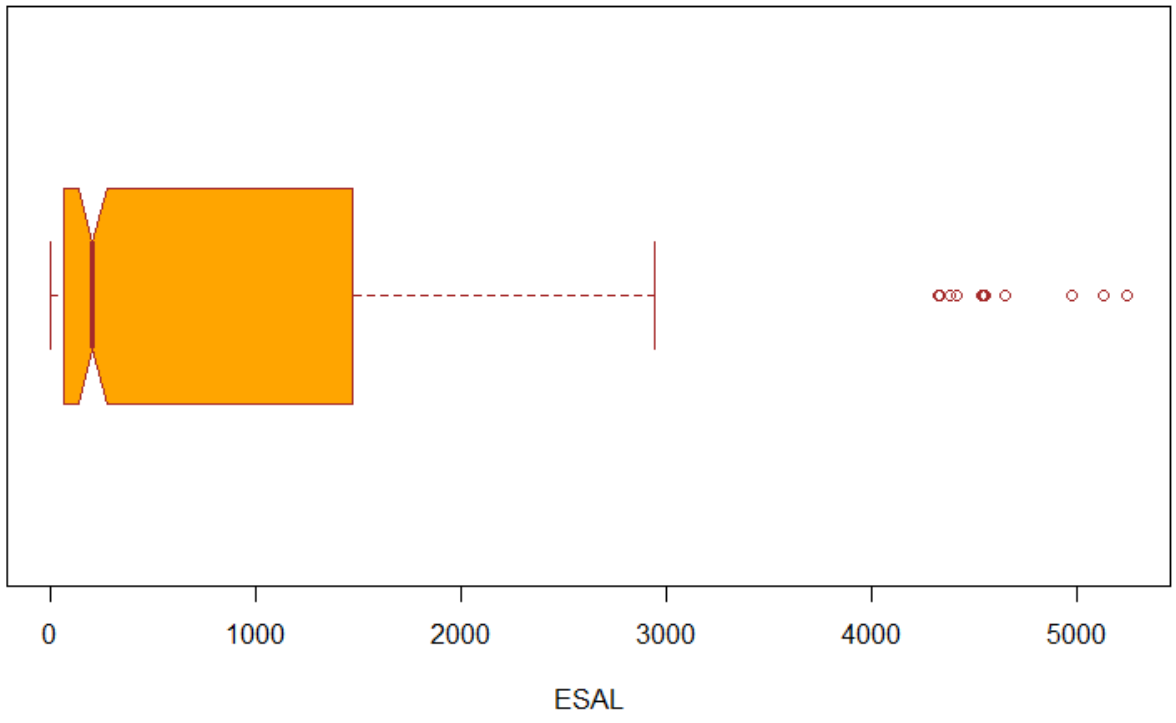

Figure A.32 – *ESAL* values Boxplot for ACP pavements – Full Dataset

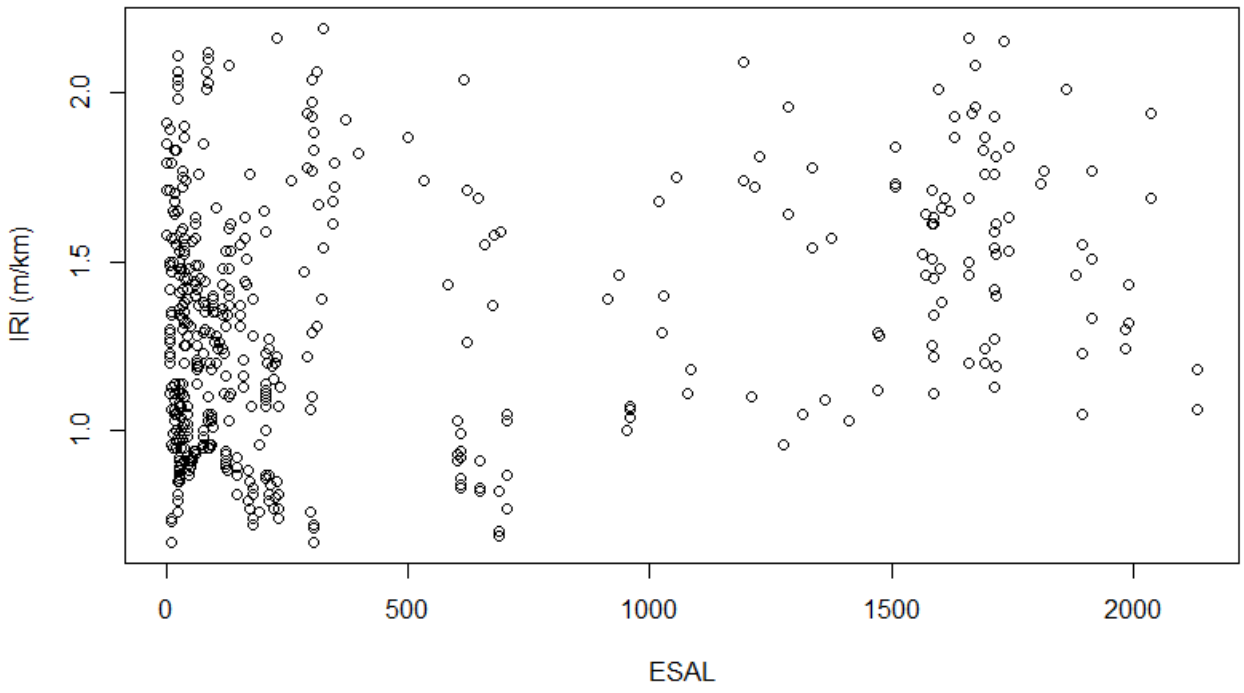

Figure A.33 – *ESAL* values scatter plot for ACP pavements – inner-fence Dataset

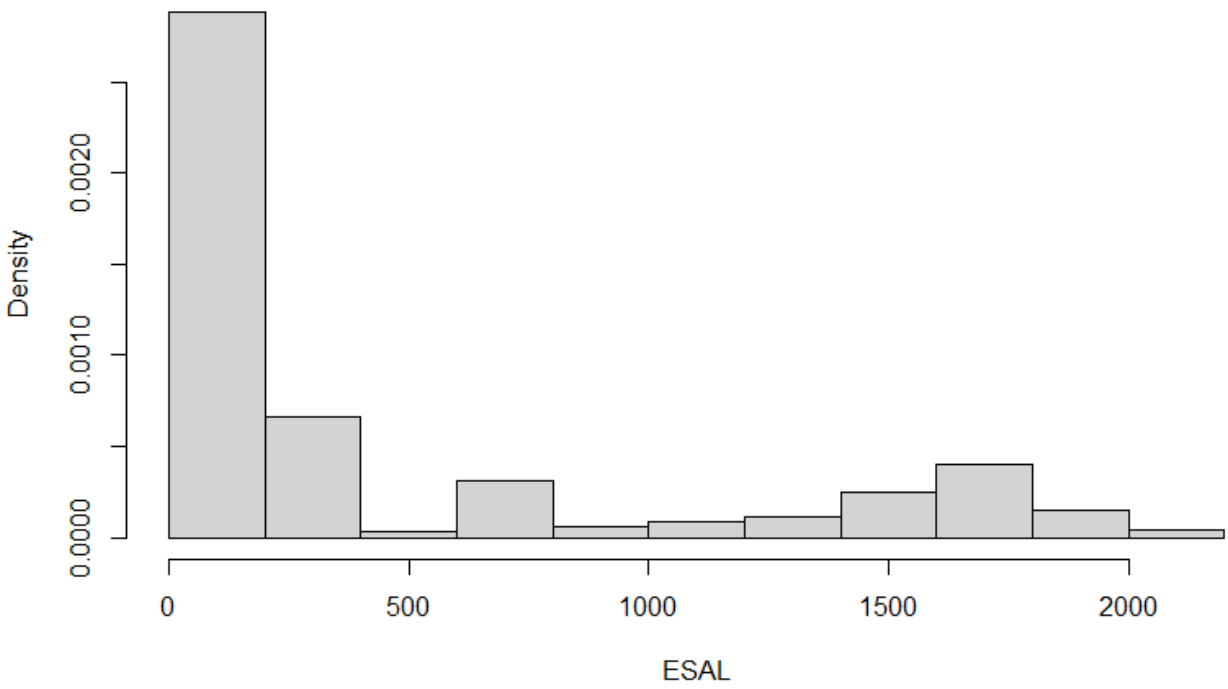

Figure A.34 – *ESAL* values histogram plot for ACP pavements – inner-fence Dataset

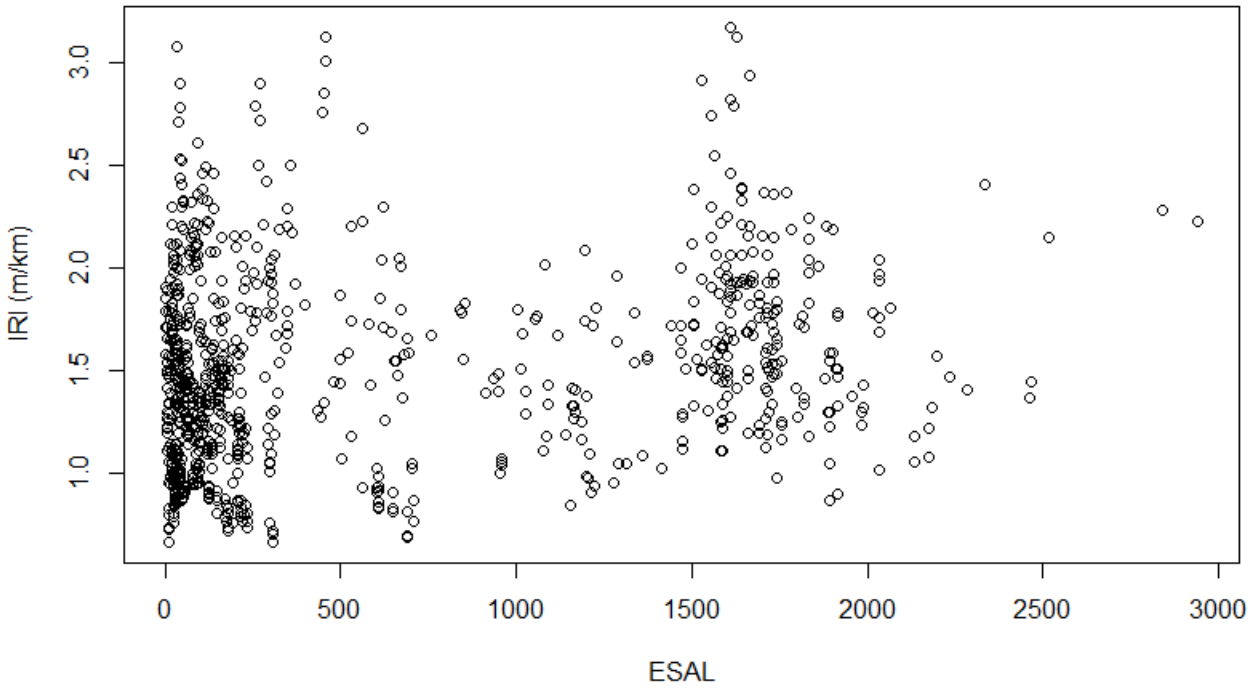

Figure A.35 – *ESAL* values scatter plot for ACP pavements – Outer-fence Dataset

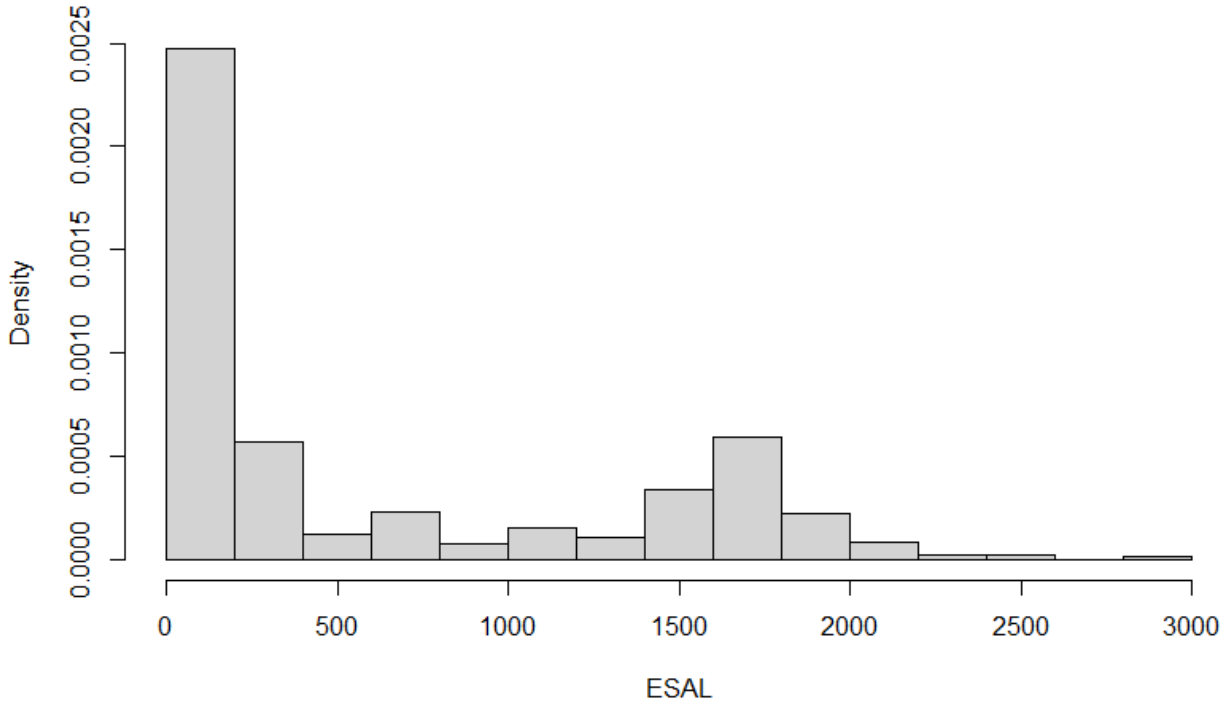

Figure A.36 – *ESAL* values histogram plot for ACP pavements – Outer-fence Dataset

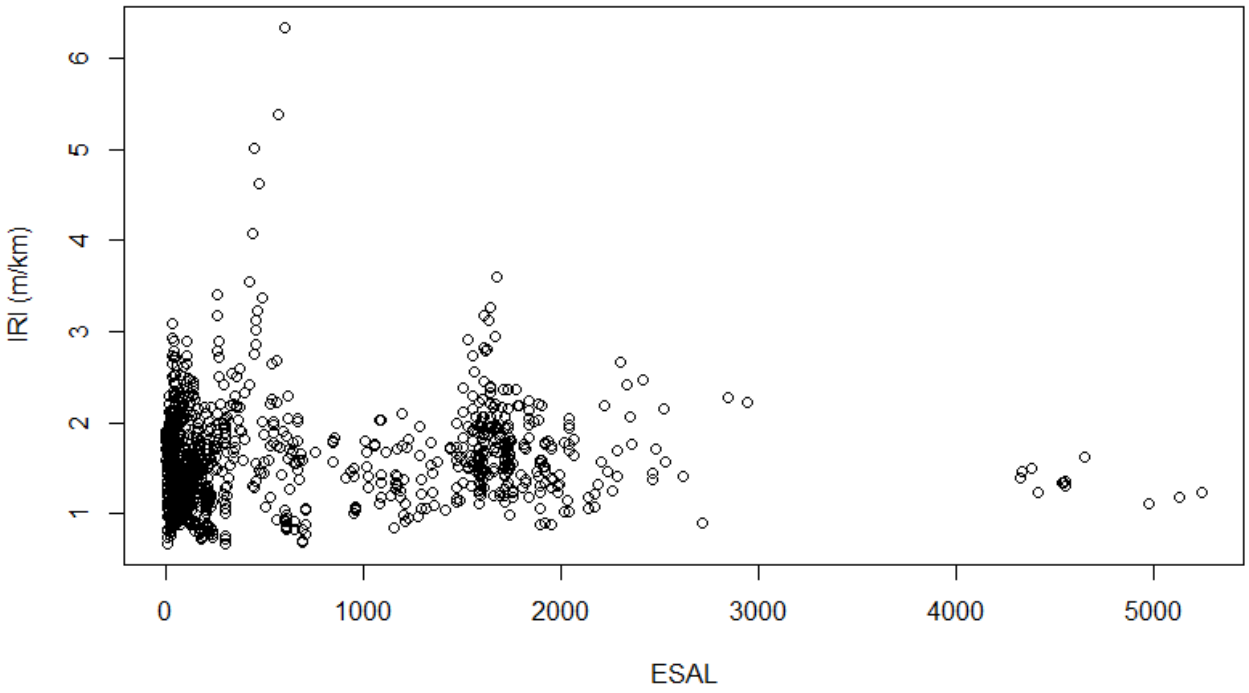

Figure A.37 – *ESAL* values scatter plot for ACP pavements – Full Dataset

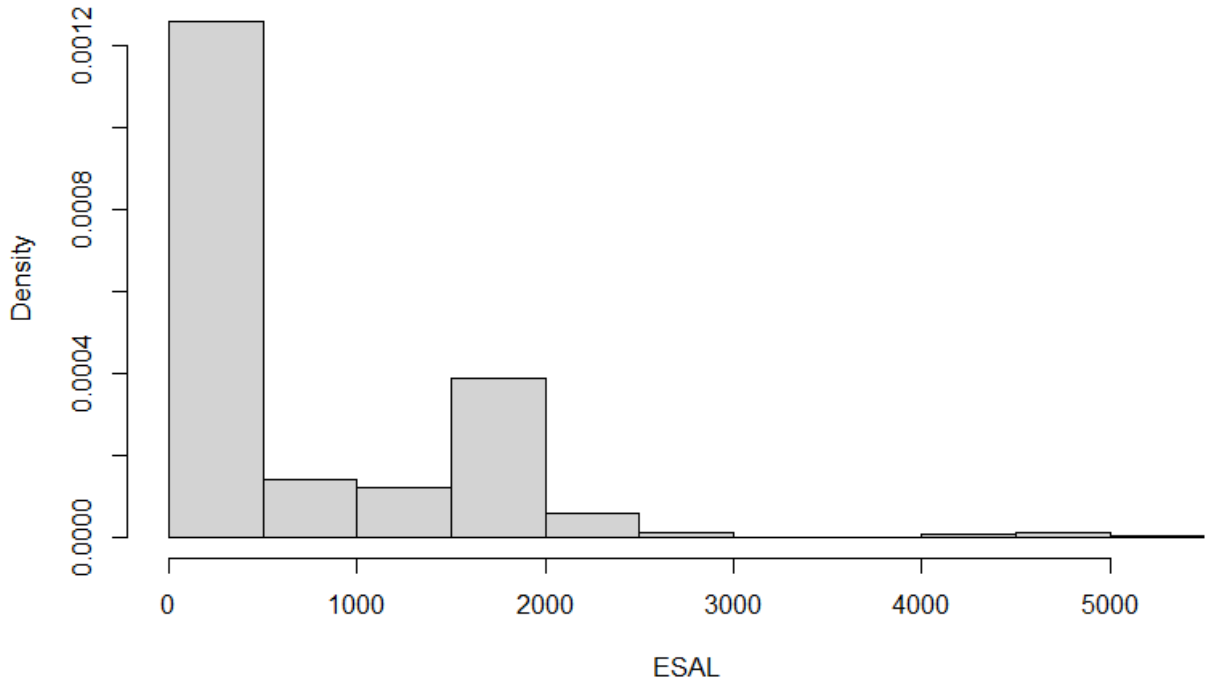

Figure A.38 – *ESAL* values histogram plot for ACP pavements – Full Dataset

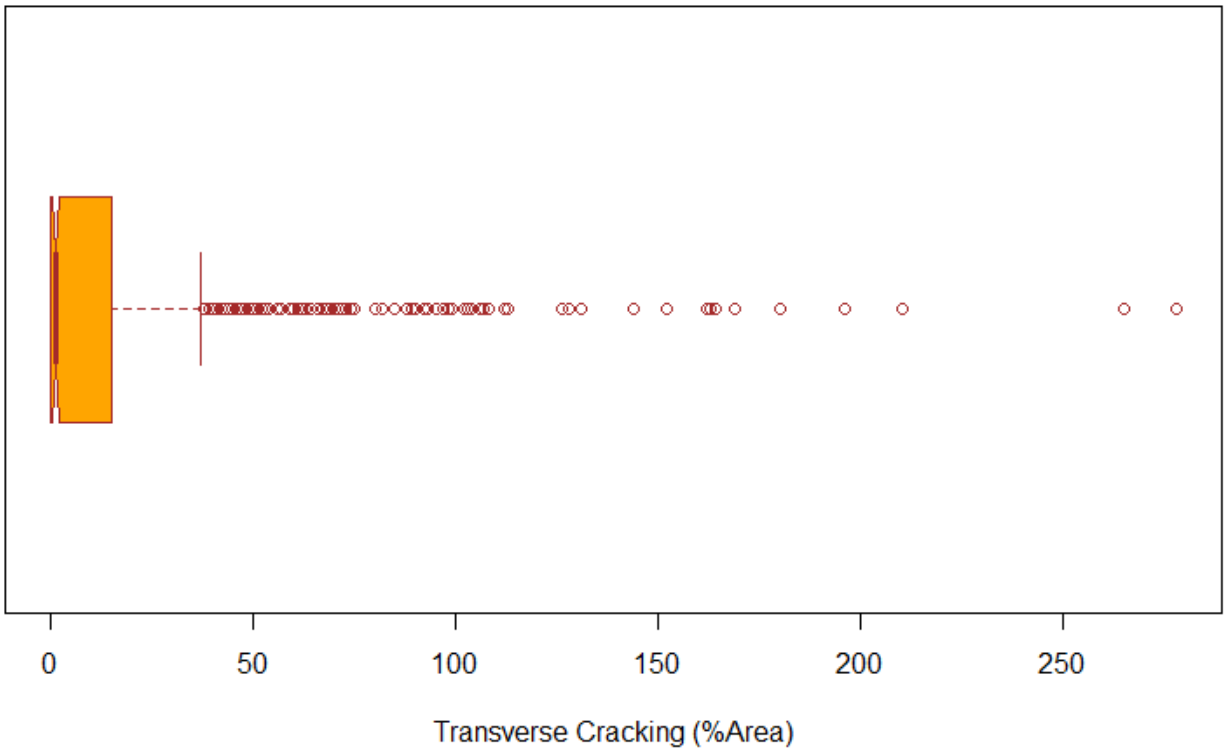

Figure A.39 – *TrcAr* values Boxplot for ACP pavements – Full Dataset

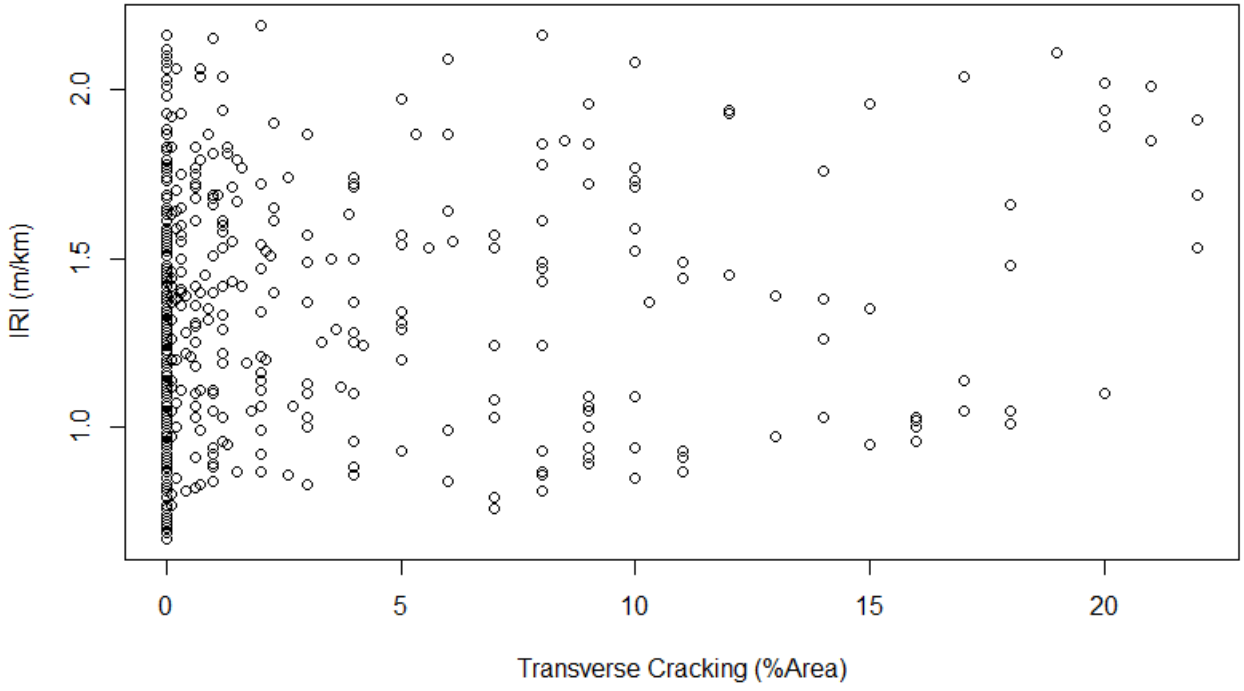

Figure A.40 – *TrcAr* values scatter plot for ACP pavements – inner-fence Dataset

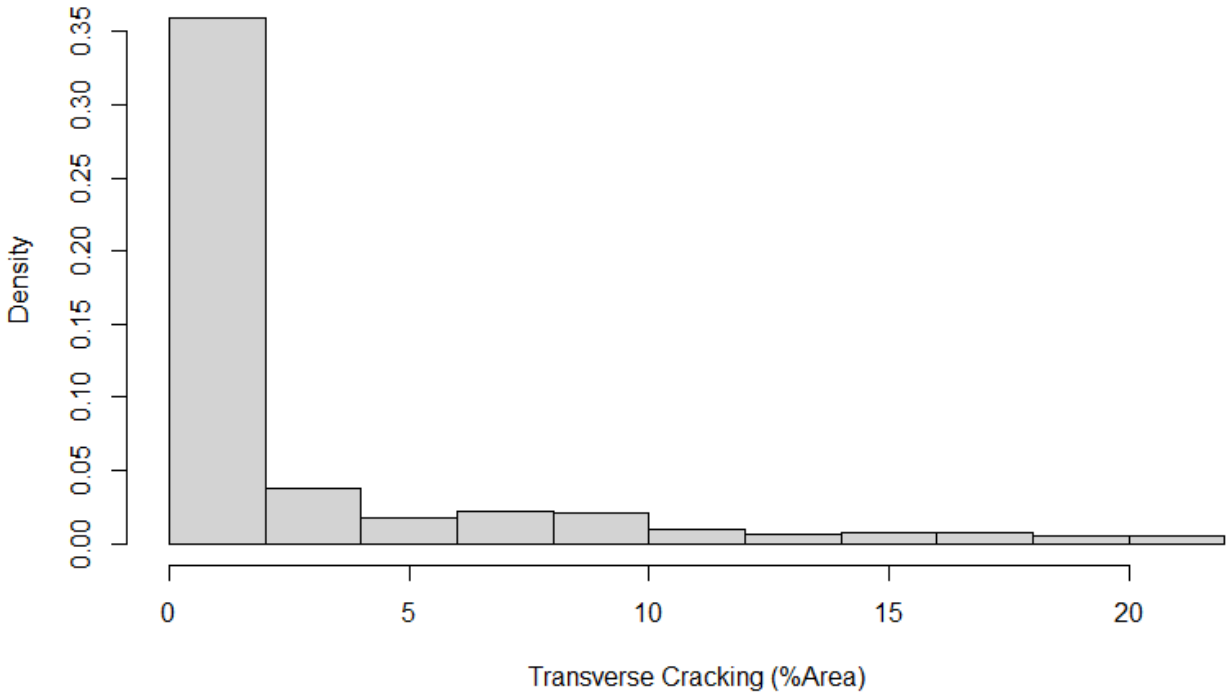

Figure A.41 – *TrcAr* values histogram plot for ACP pavements – inner-fence Dataset

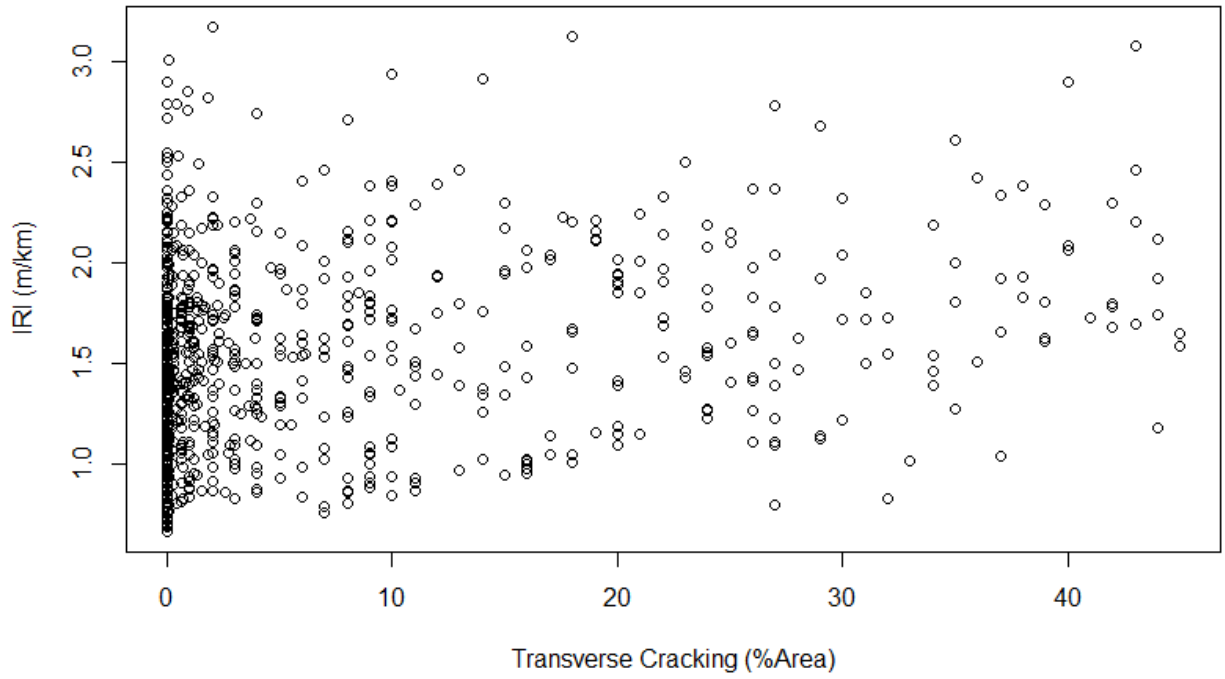

Figure A.42 – *TrcAr* values scatter plot for ACP pavements – Outer-fence Dataset

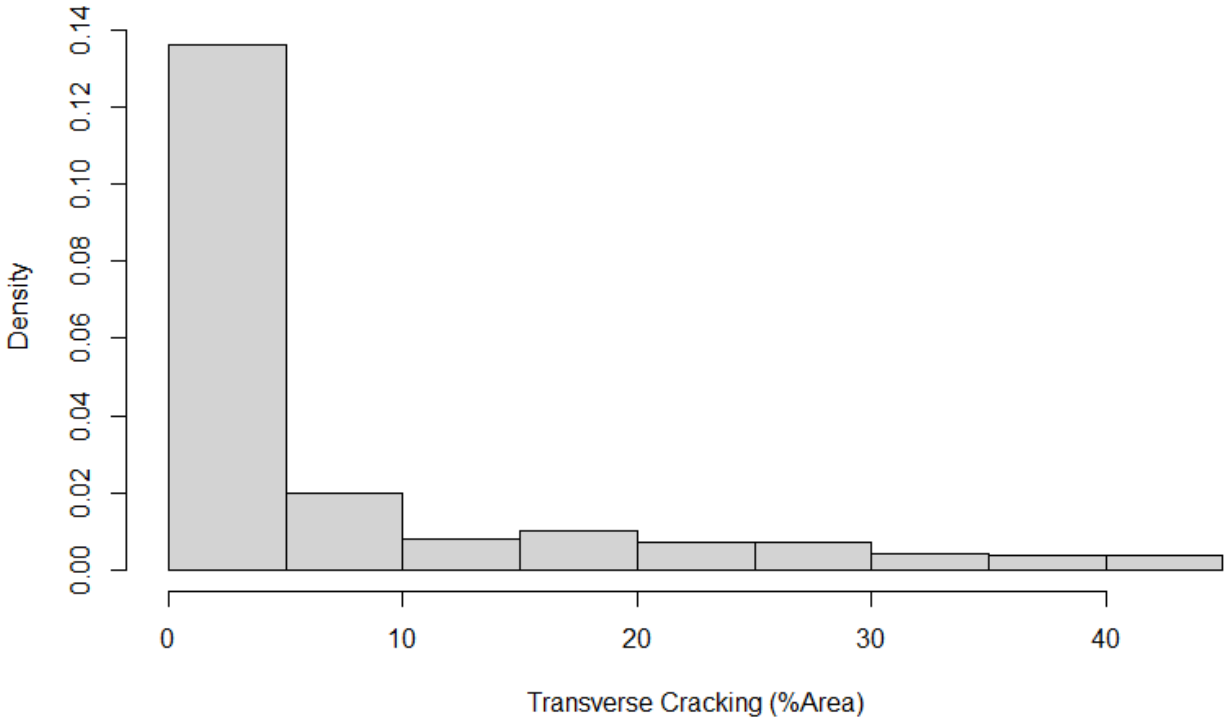

Figure A.43 – *TrcAr* values histogram plot for ACP pavements – Outer-fence Dataset

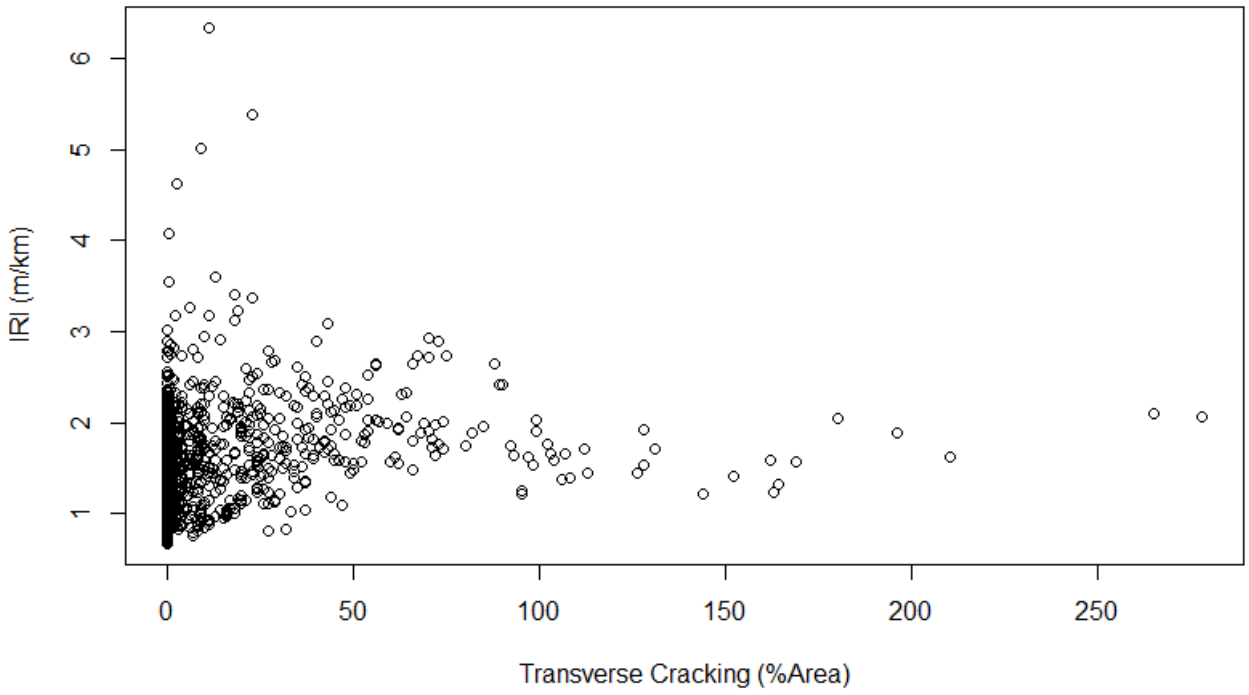

Figure A.44 – *TrcAr* values scatter plot for ACP pavements – Full Dataset

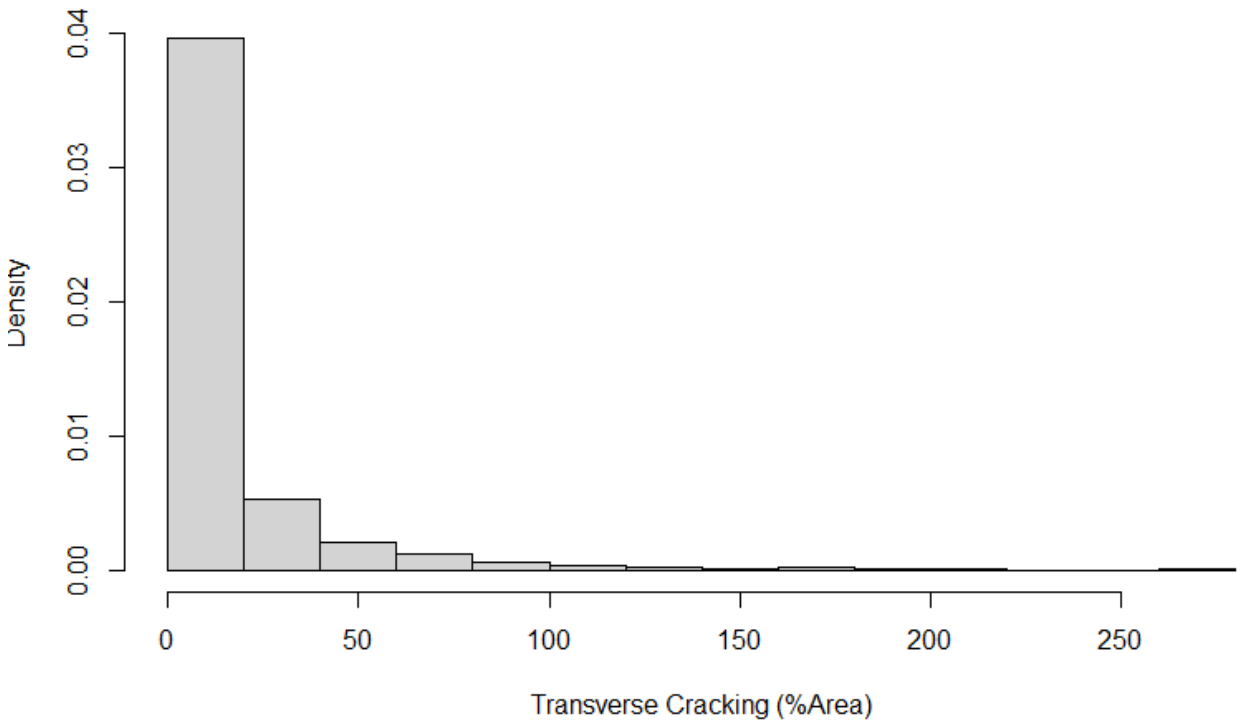

Figure A.45 – *TrcAr* values histogram plot for ACP pavements – Full Dataset

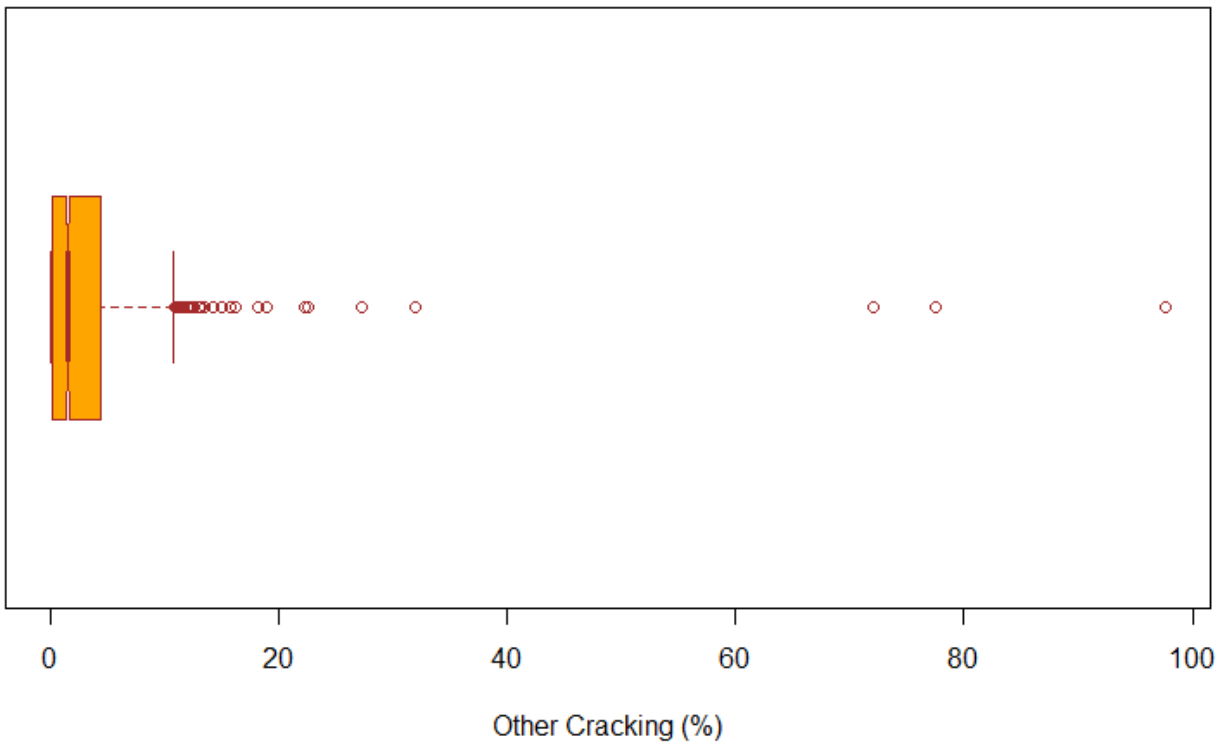

Figure A.46 – *OtherCAr* values Boxplot for ACP pavements – Full Dataset

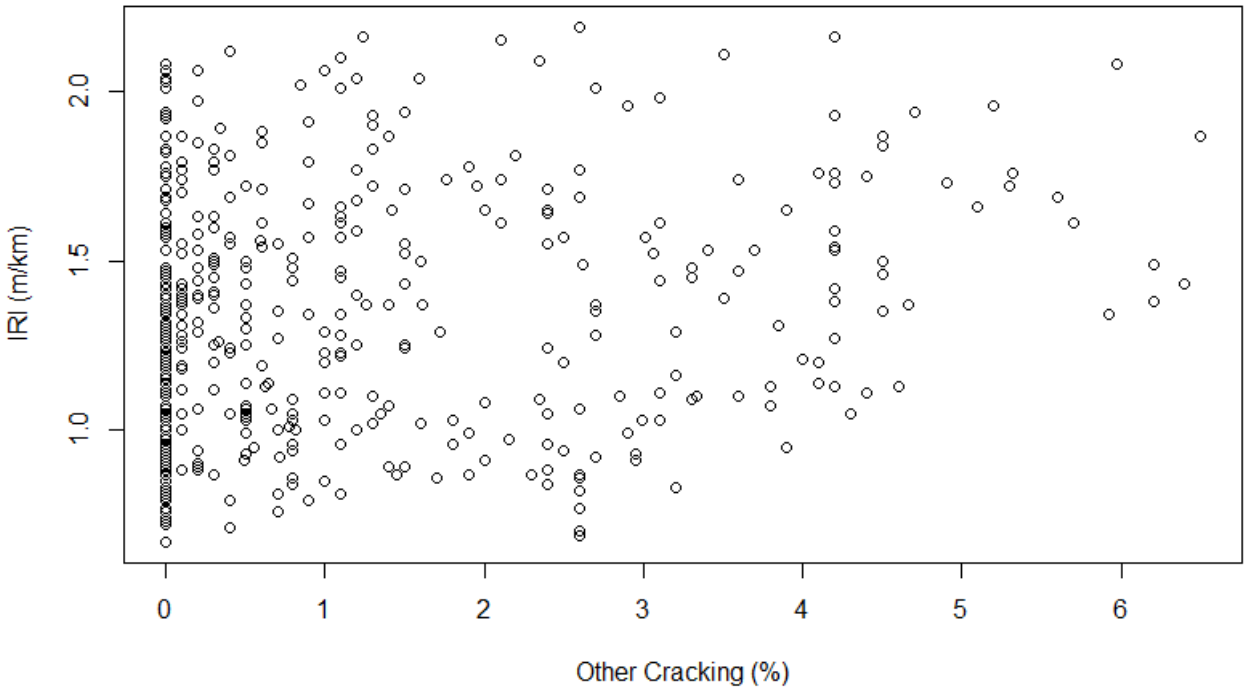

Figure A.47 – *OtherCAr* values scatter plot for ACP pavements – inner-fence Dataset

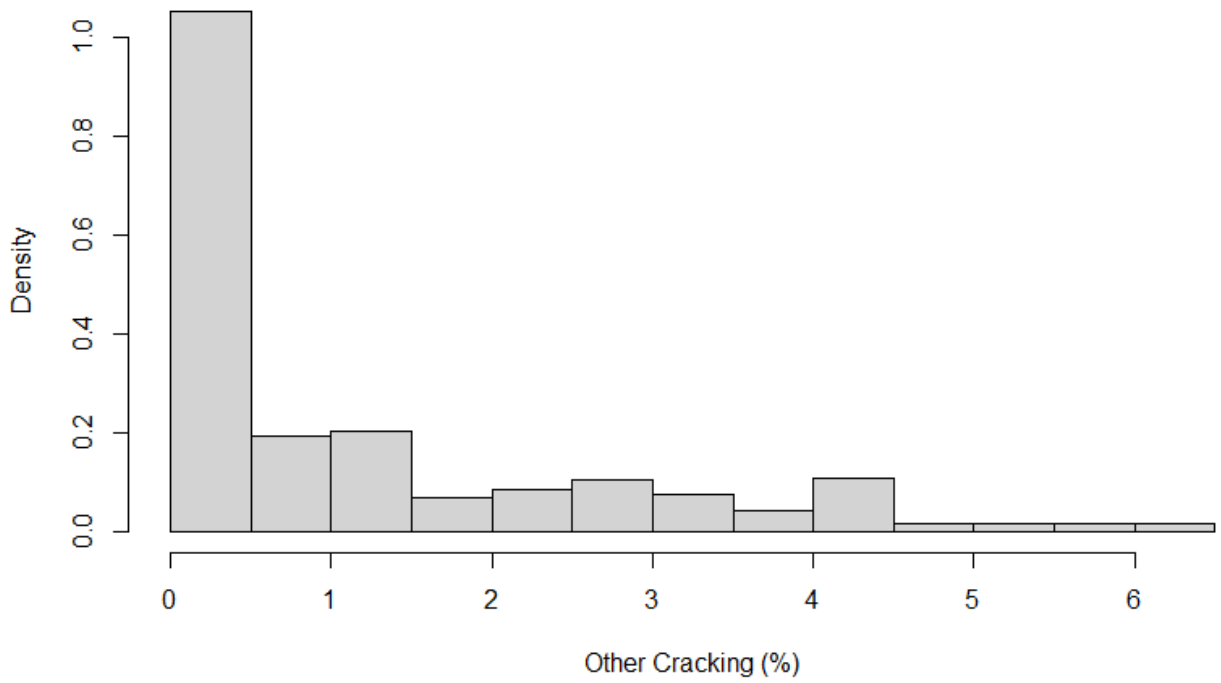

Figure A.48 – *OtherCAr* values histogram plot for ACP pavements – inner-fence Dataset

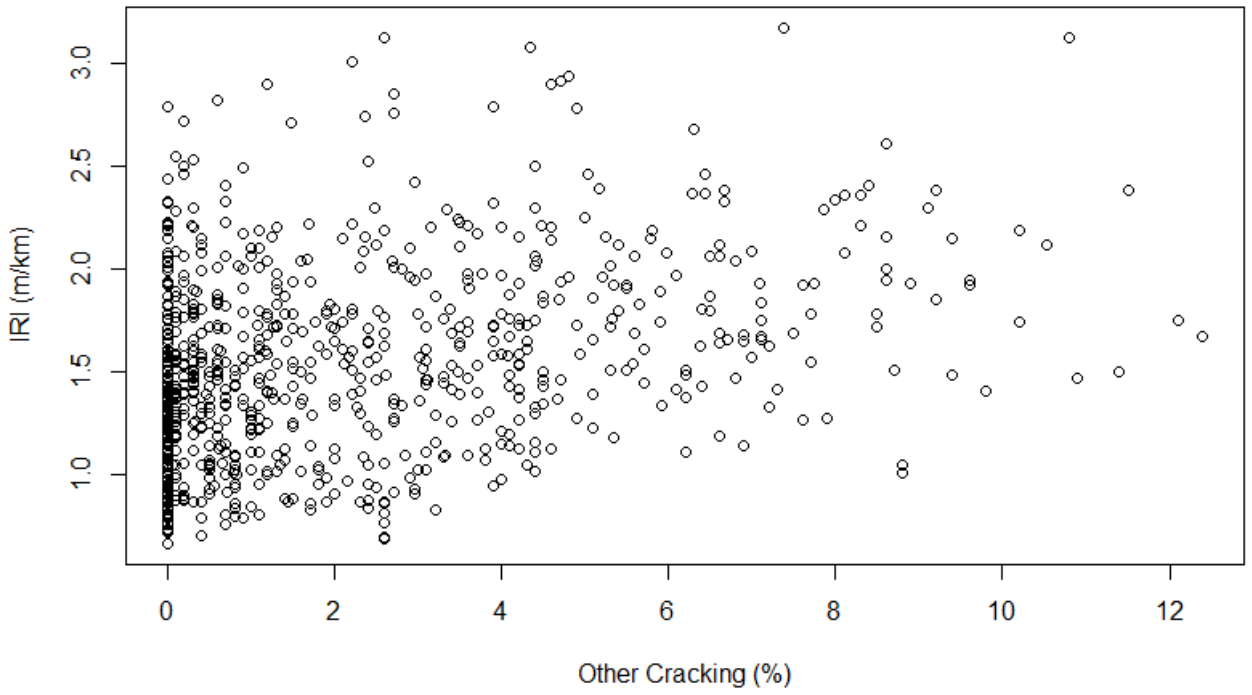

Figure A.49 – *OtherCAr* values scatter plot for ACP pavements – Outer-fence Dataset

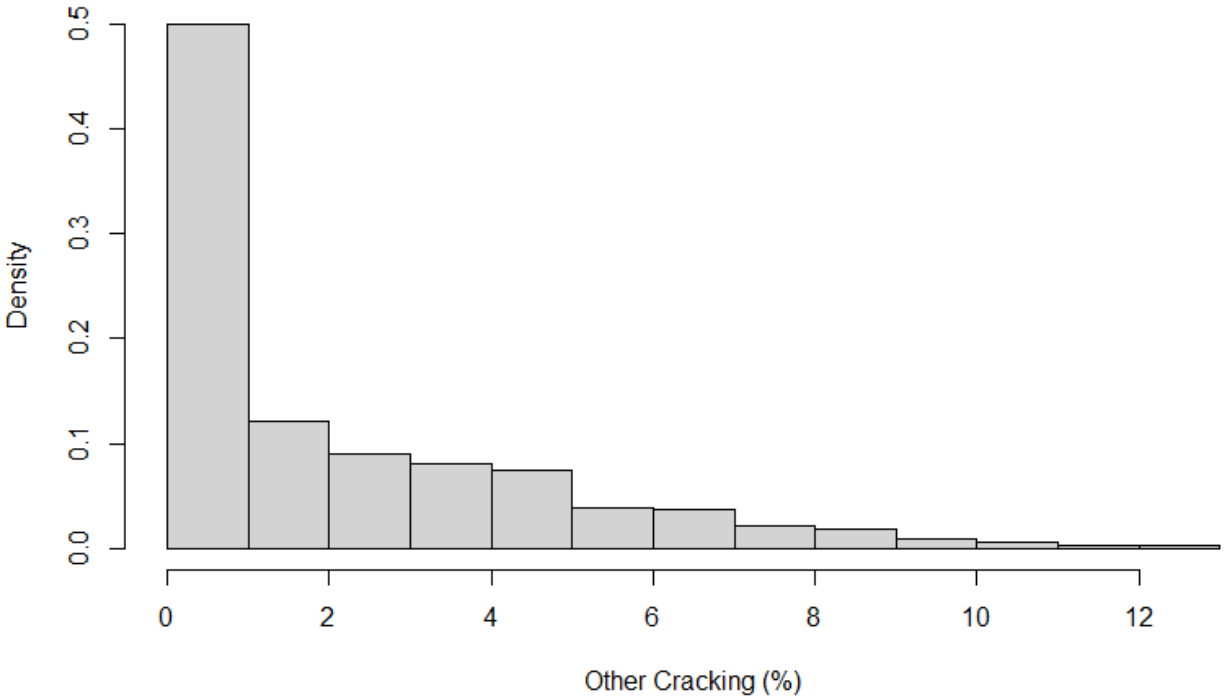

Figure A.50 – *OtherCAr* values histogram plot for ACP pavements – Outer-fence Dataset

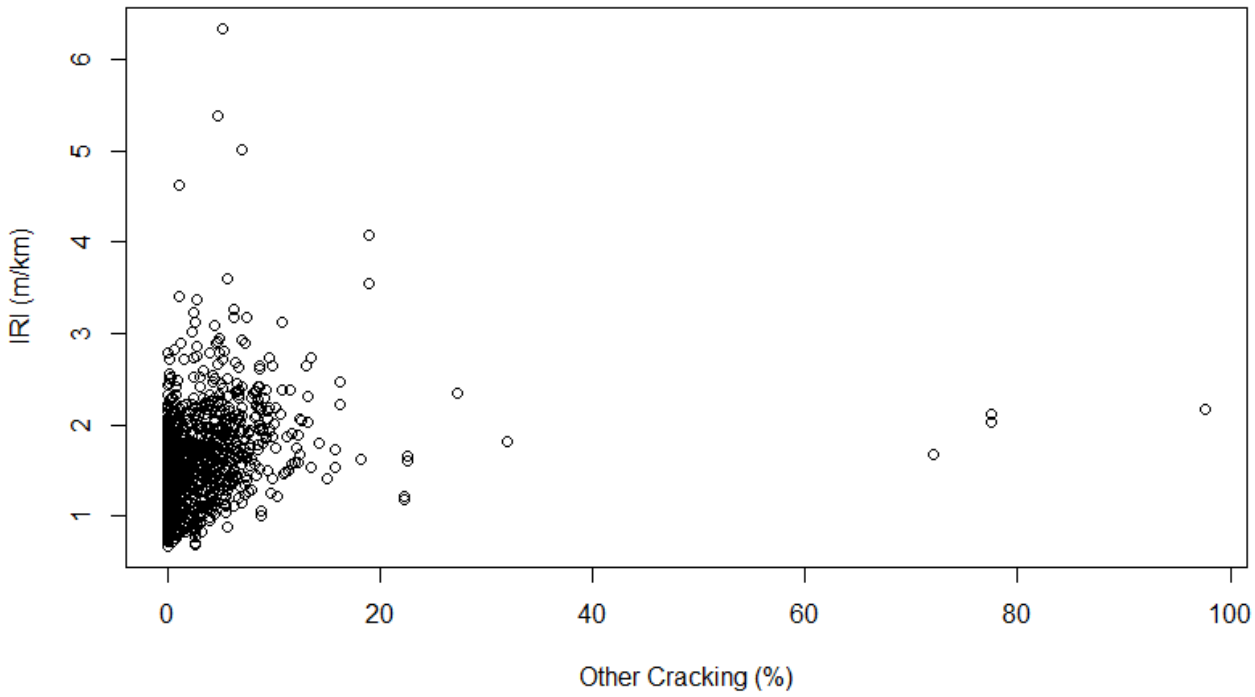

Figure A.51 – *OtherCAr* values scatter plot for ACP pavements – Full Dataset

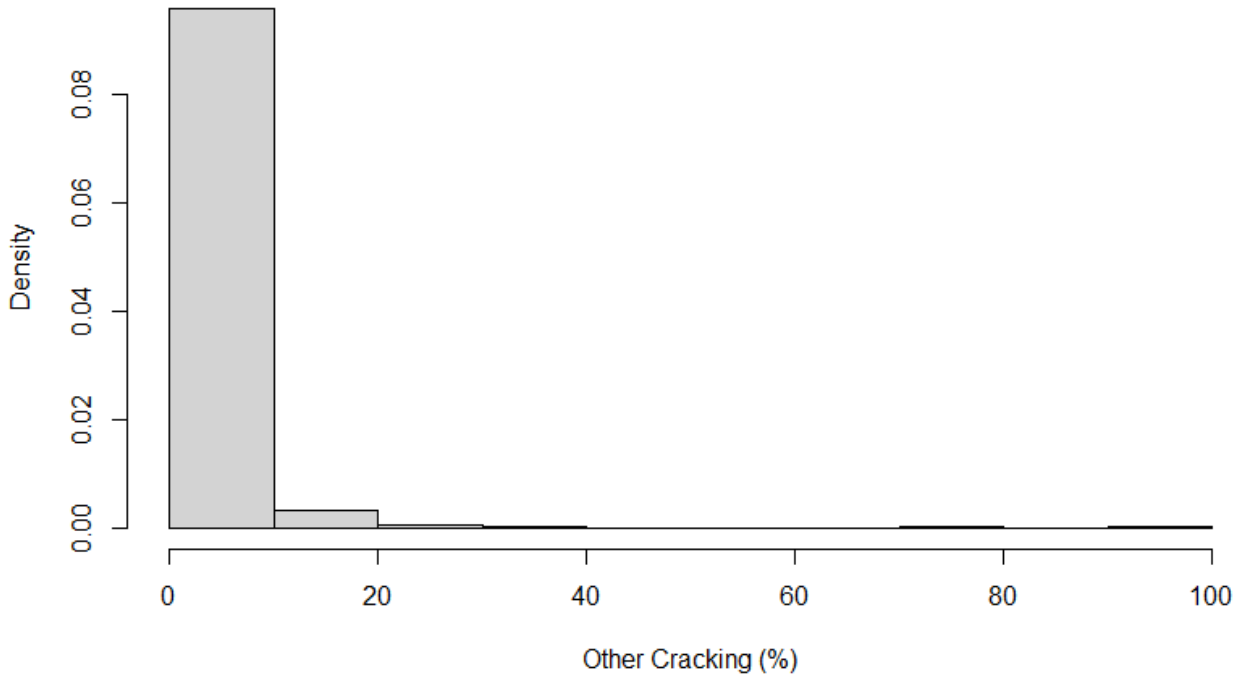

Figure A.52 – *OtherCAr* values histogram plot for ACP pavements – Full Dataset

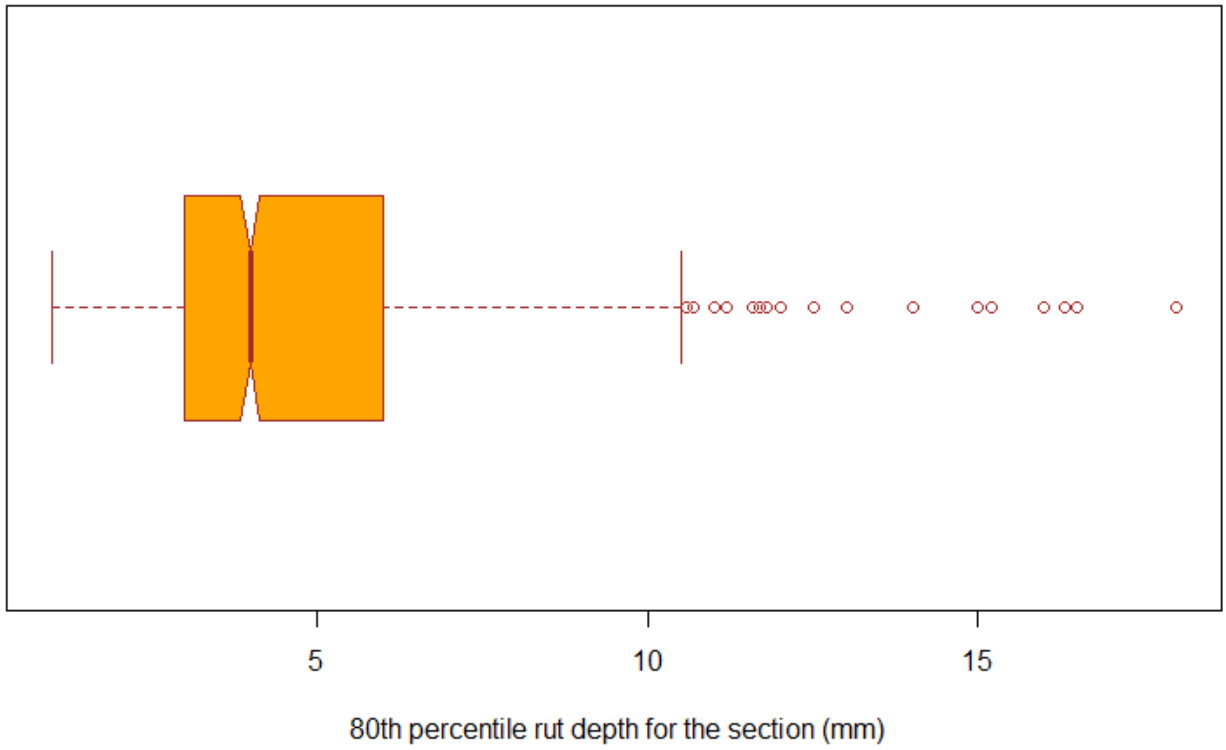

Figure A.53 – *RUT* values Boxplot for ACP pavements – Full Dataset

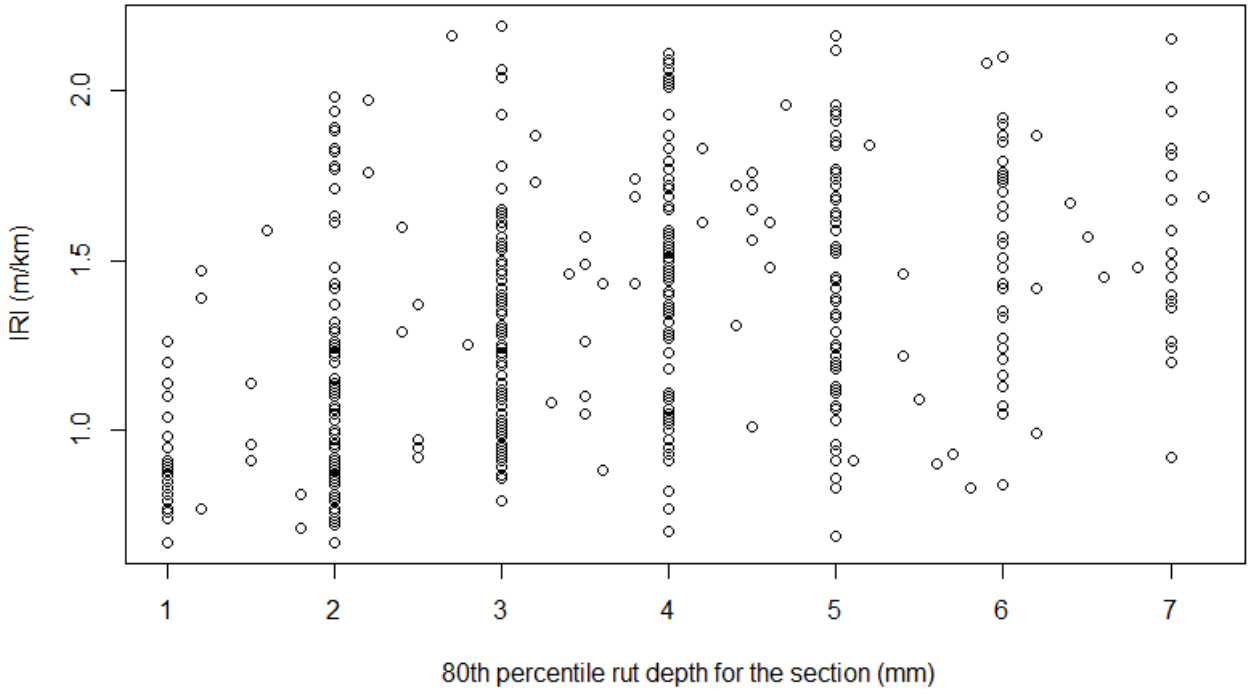

Figure A.54 – *RUT* values scatter plot for ACP pavements – inner-fence Dataset

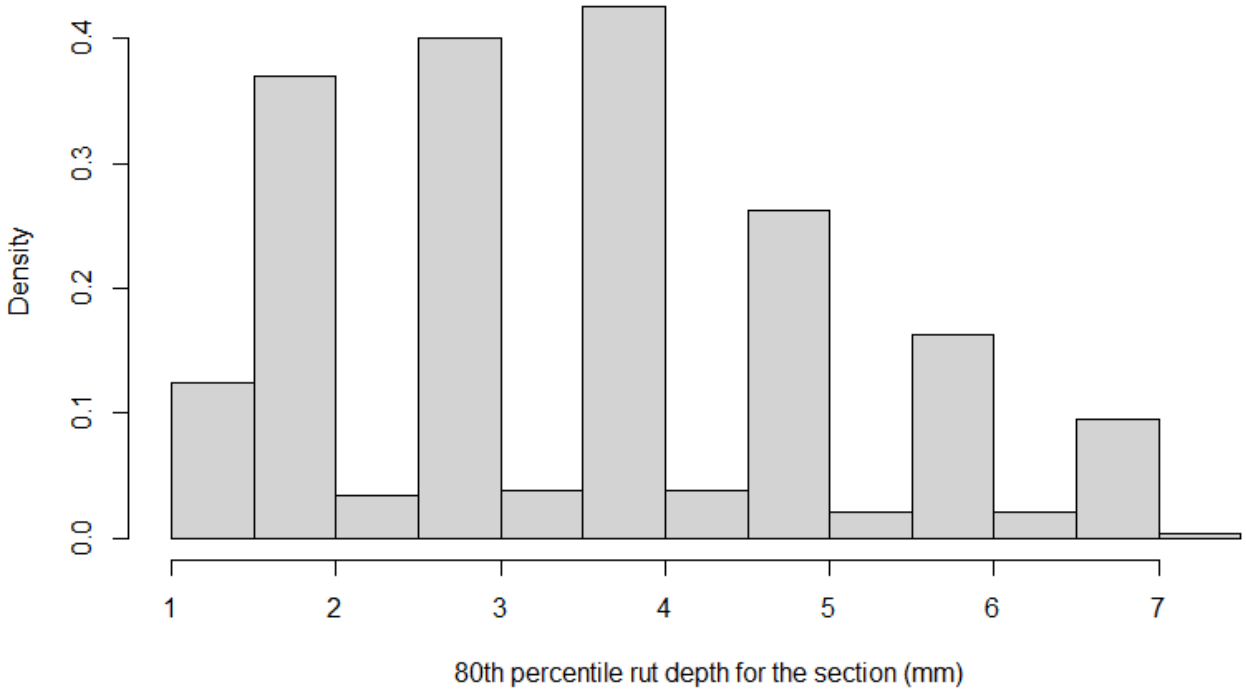

Figure A.55 – *RUT* values histogram plot for ACP pavements – inner-fence Dataset

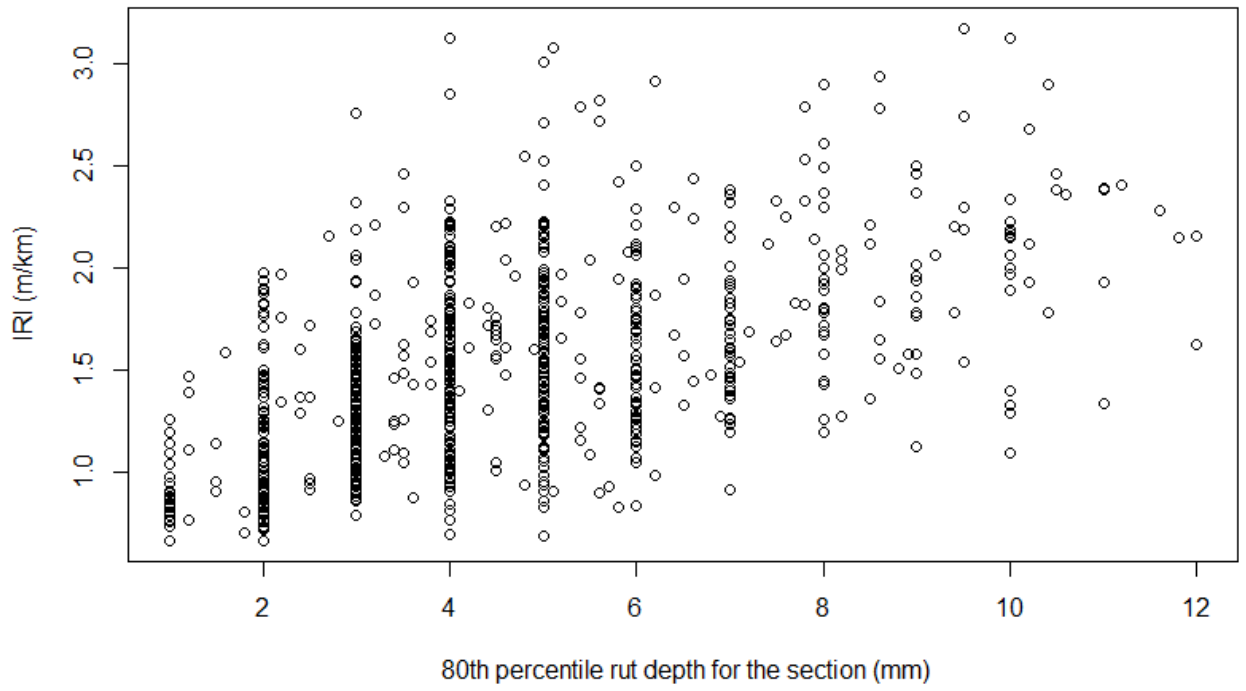

Figure A.56 – *RUT* values scatter plot for ACP pavements – Outer-fence Dataset

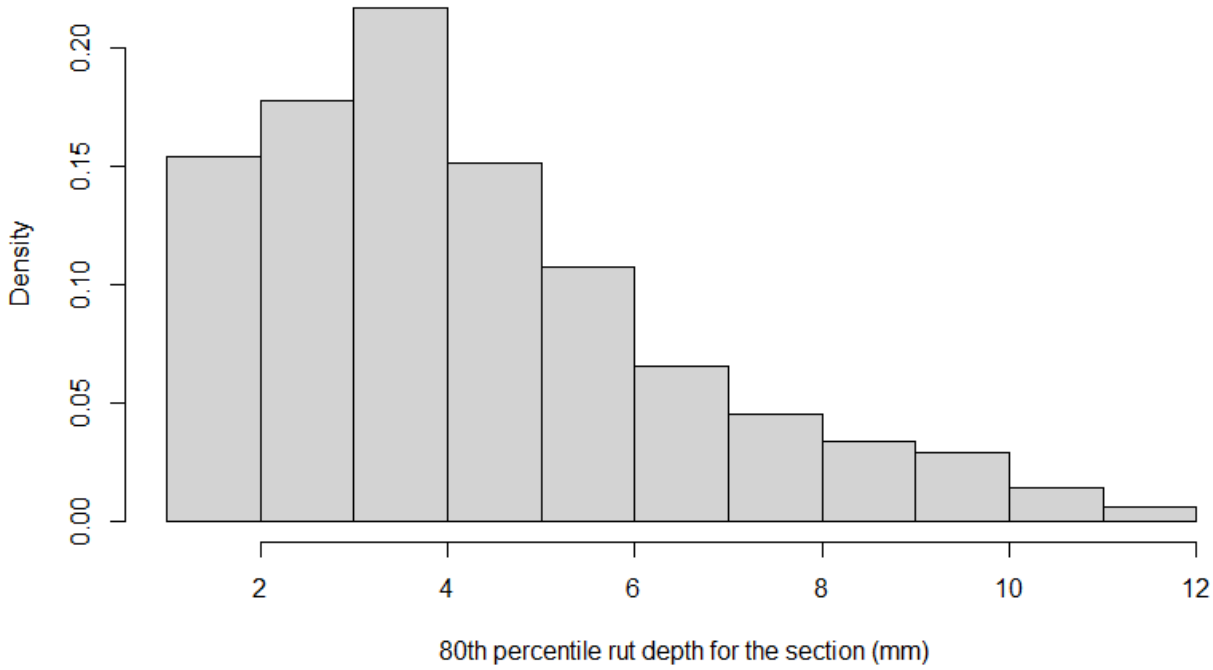

Figure A.57 – *RUT* values histogram plot for ACP pavements – Outer-fence Dataset

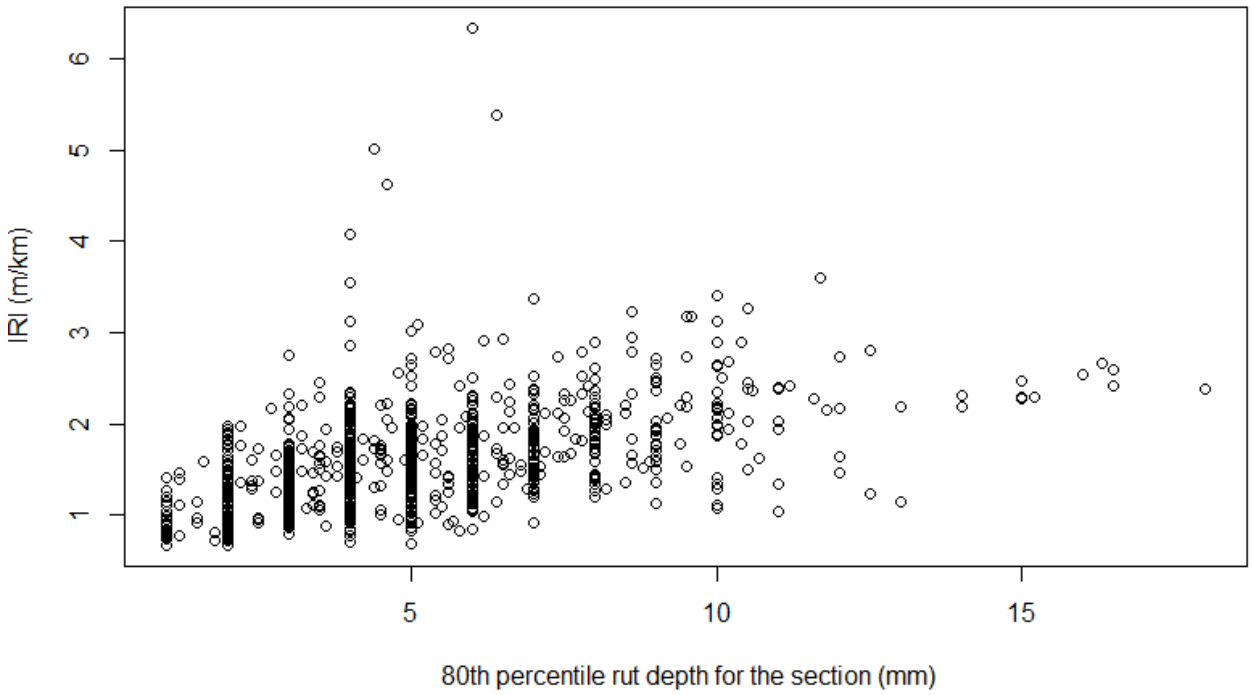

Figure A.58 – *RUT* values scatter plot for ACP pavements – Full Dataset

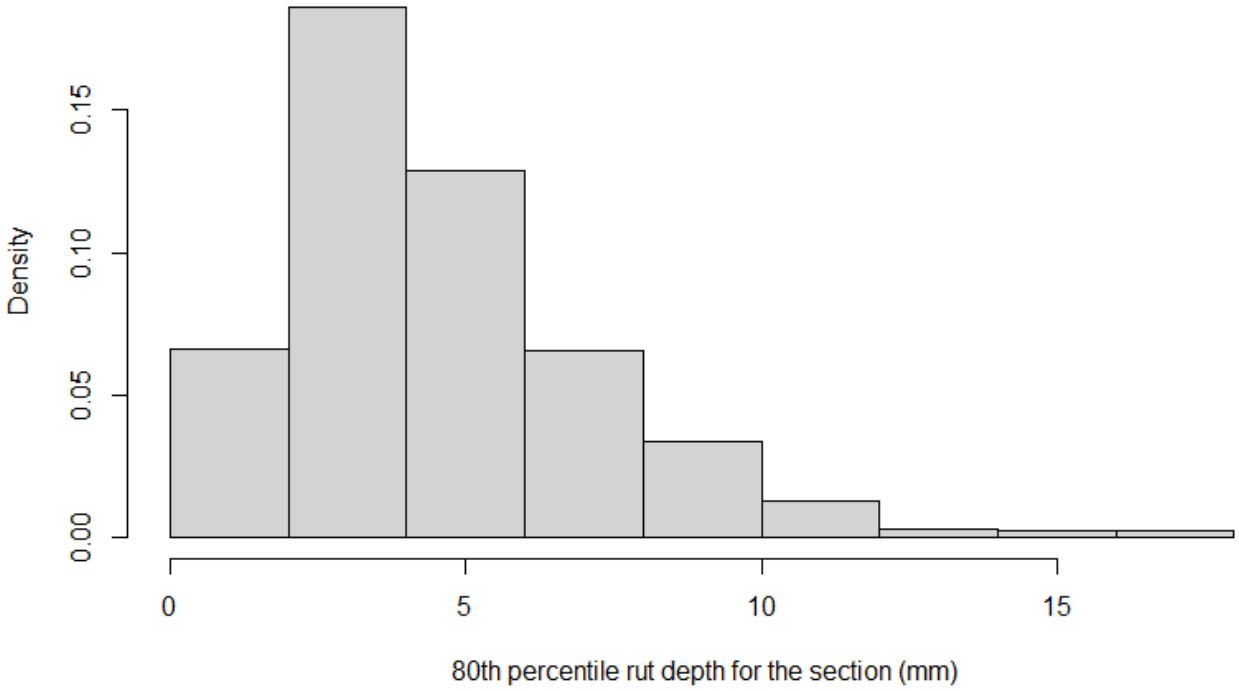

Figure A.59 – *RUT* values histogram plot for ACP pavements – Full Dataset

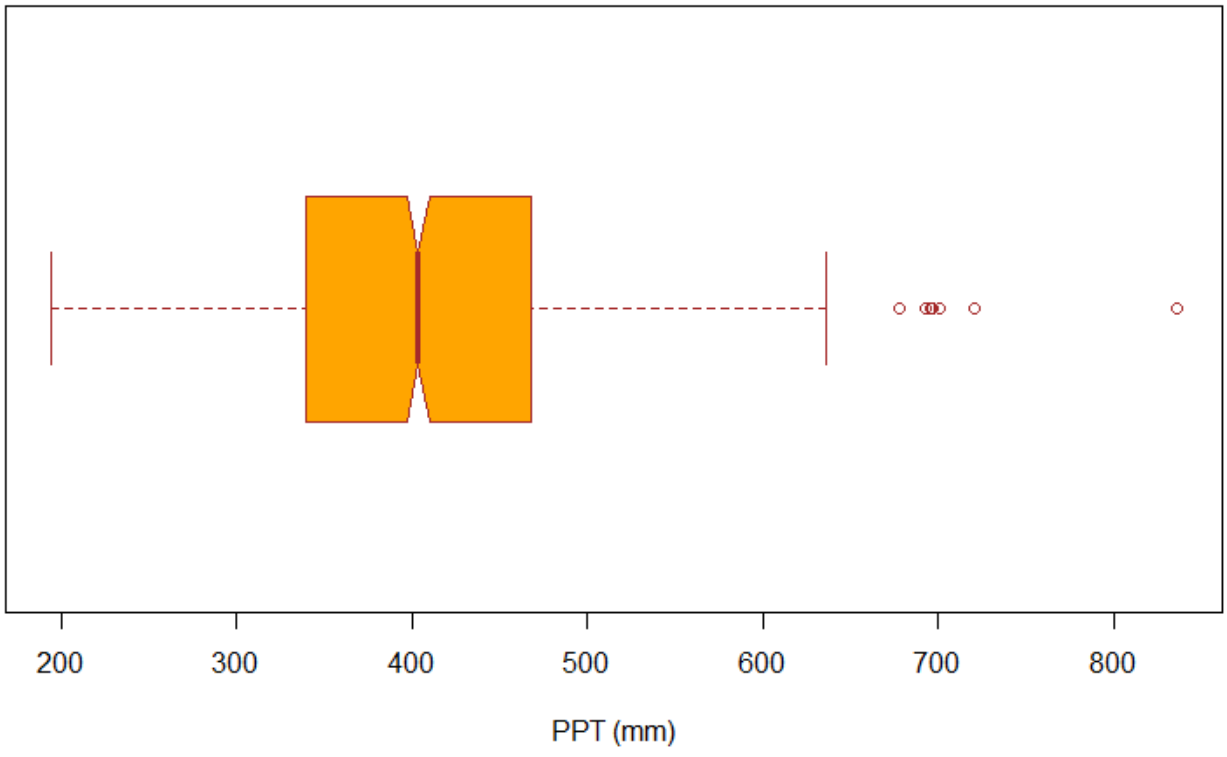

Figure A.60 – *PPT* values Boxplot for ACP pavements – Full Dataset

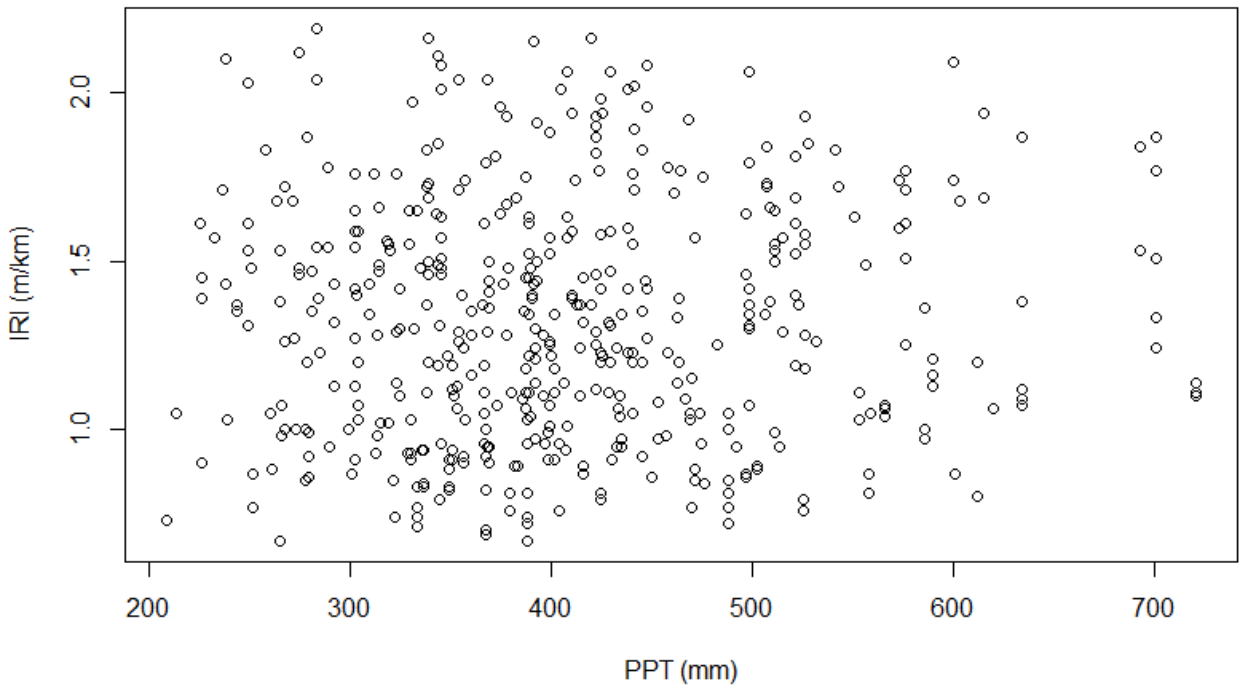

Figure A.61 – *PPT* values scatter plot for ACP pavements – inner-fence Dataset

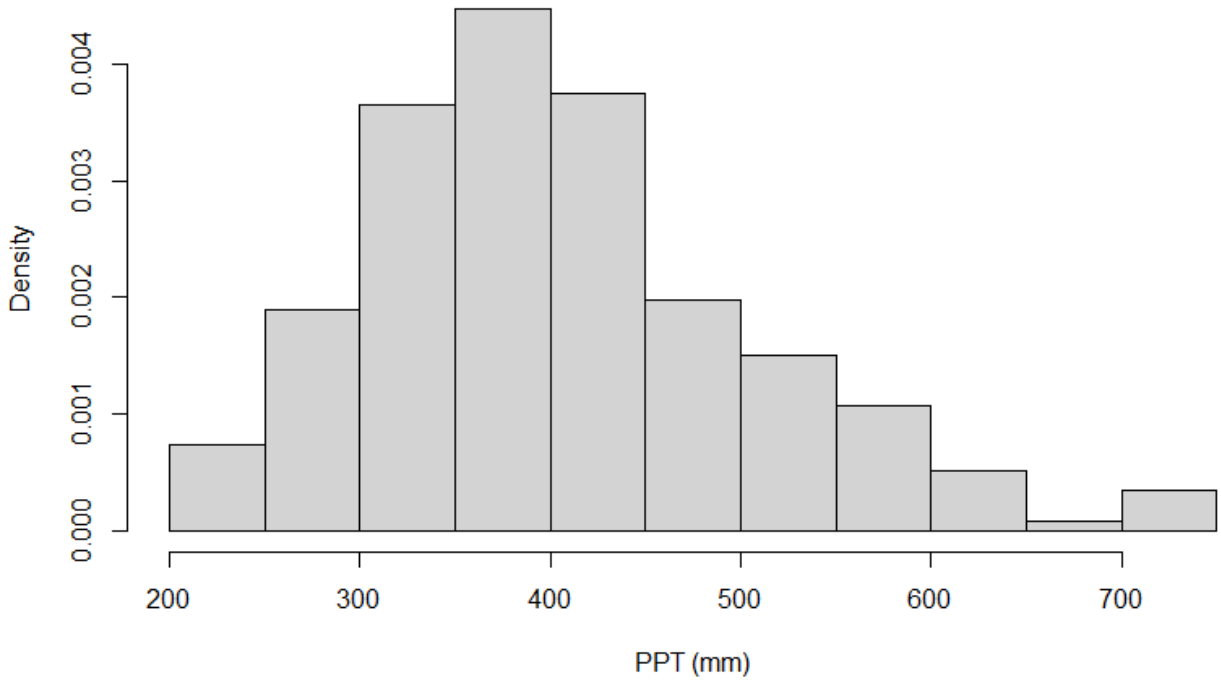

Figure A.62 – *PPT* values histogram plot for ACP pavements – inner-fence Dataset

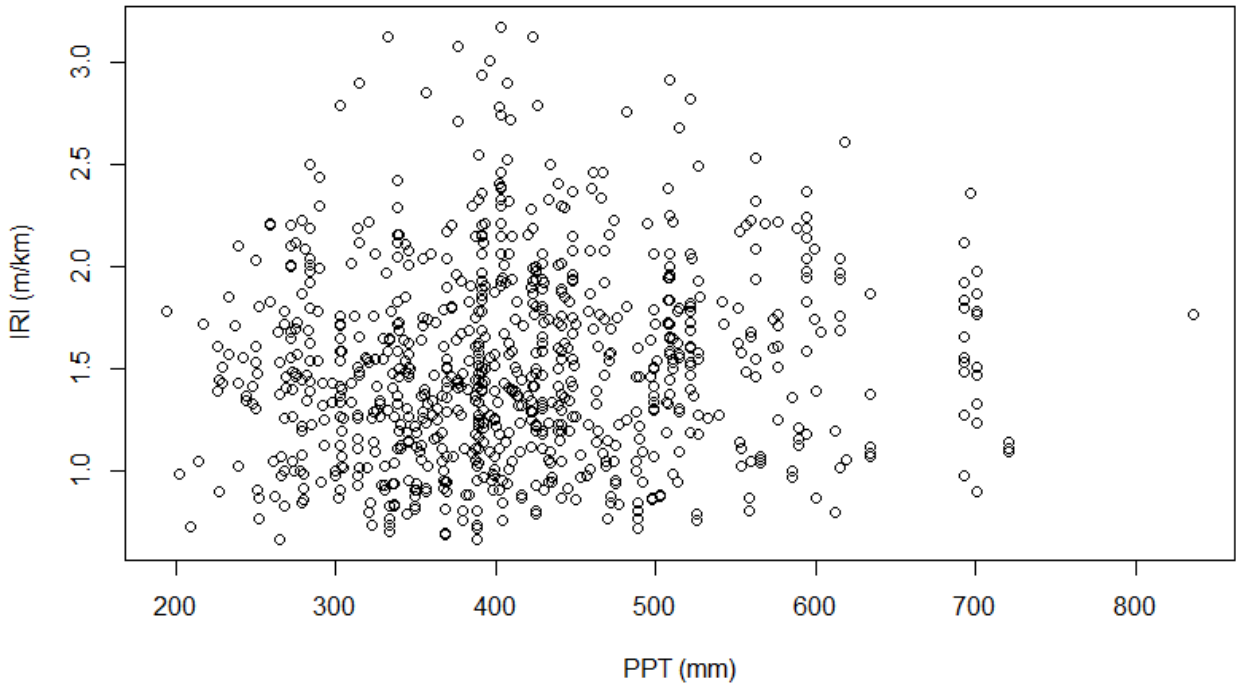

Figure A.63 – *PPT* values scatter plot for ACP pavements – Outer-fence Dataset

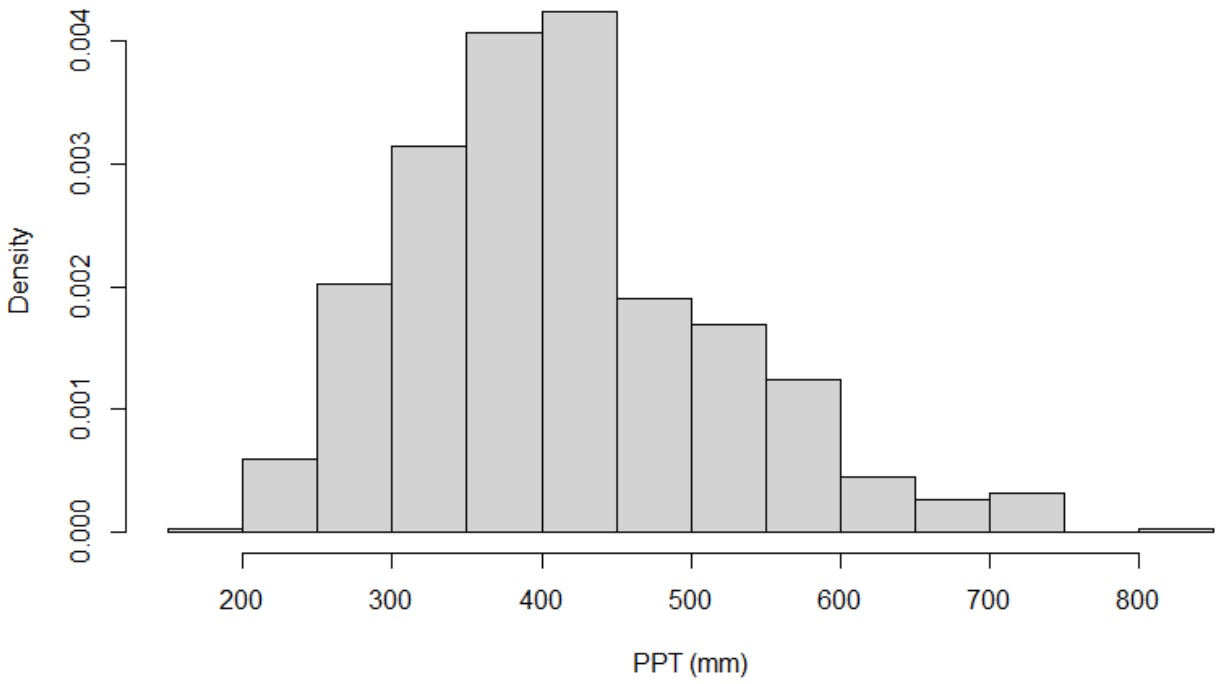

Figure A.64 – *PPT* values histogram plot for ACP pavements – Outer-fence Dataset

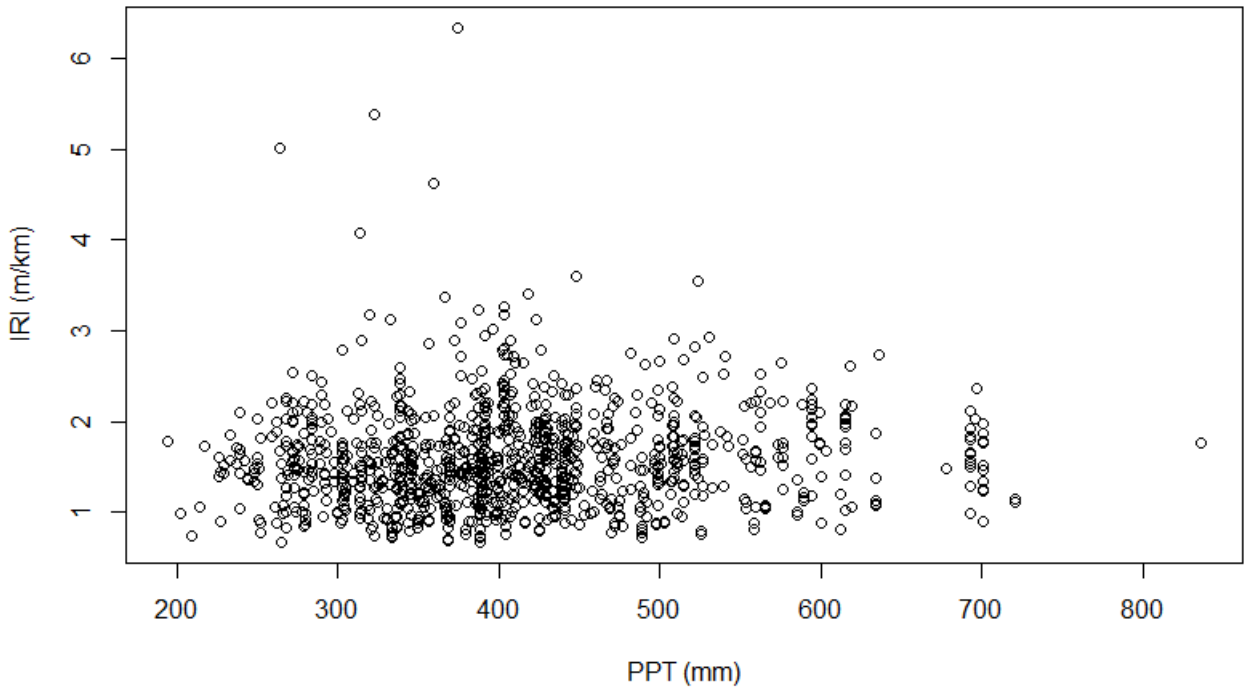

Figure A.65 – *PPT* values scatter plot for ACP pavements – Full Dataset

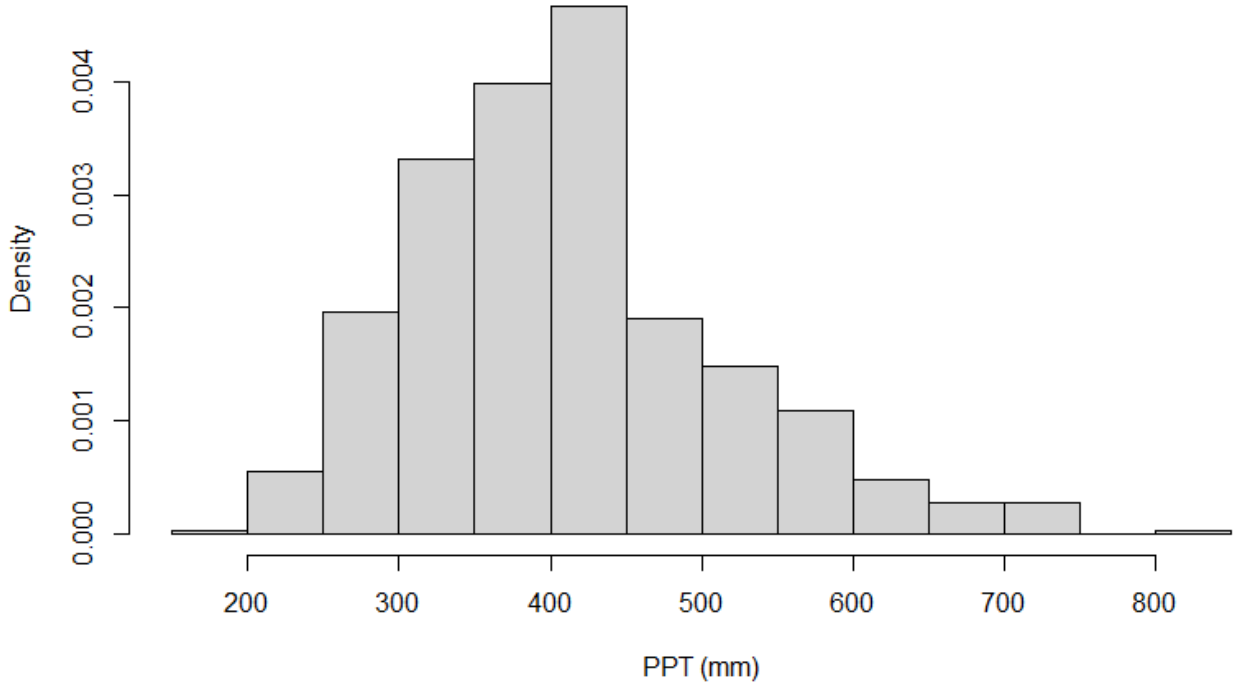

Figure A.66 – *PPT* values histogram plot for ACP pavements – Full Dataset

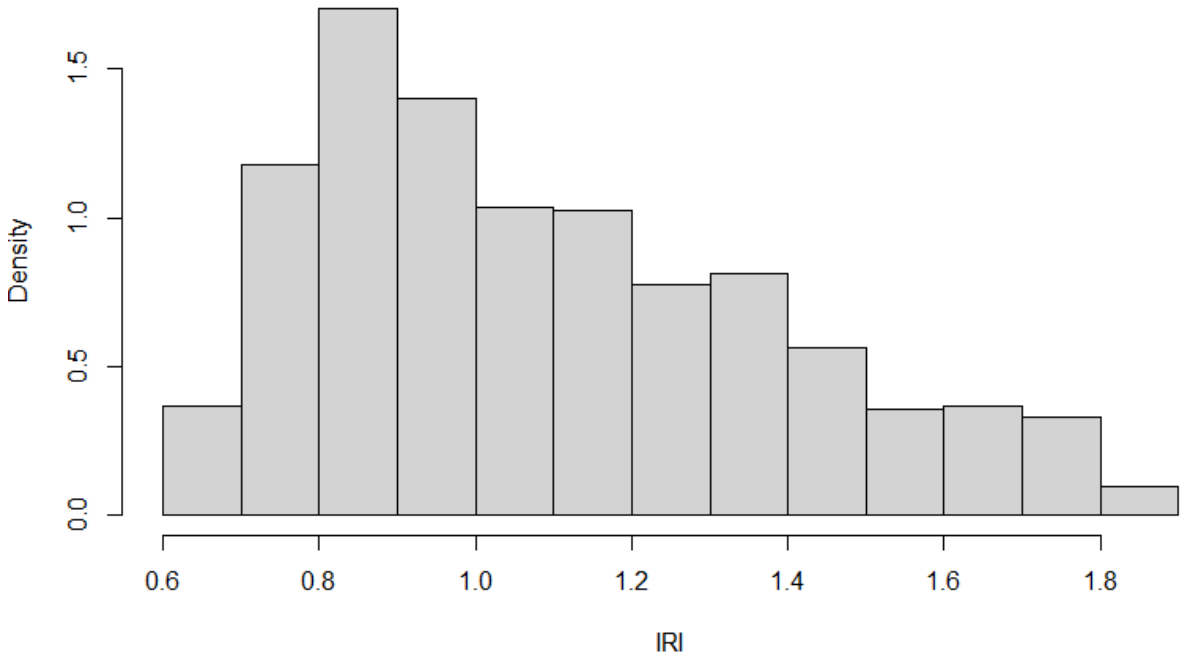

Figure A.67 – *IRI* values Boxplot for OL pavements – inner-fence Dataset

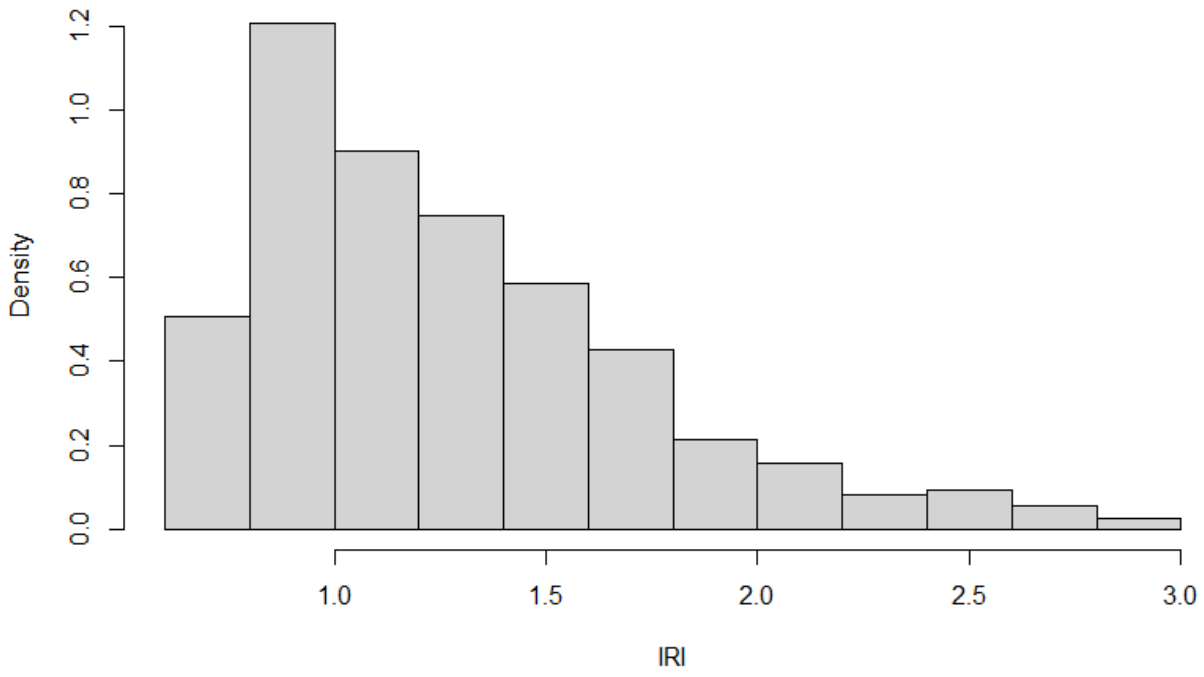

Figure A.68 – *IRI* values Boxplot for OL pavements – Outer-fence Dataset

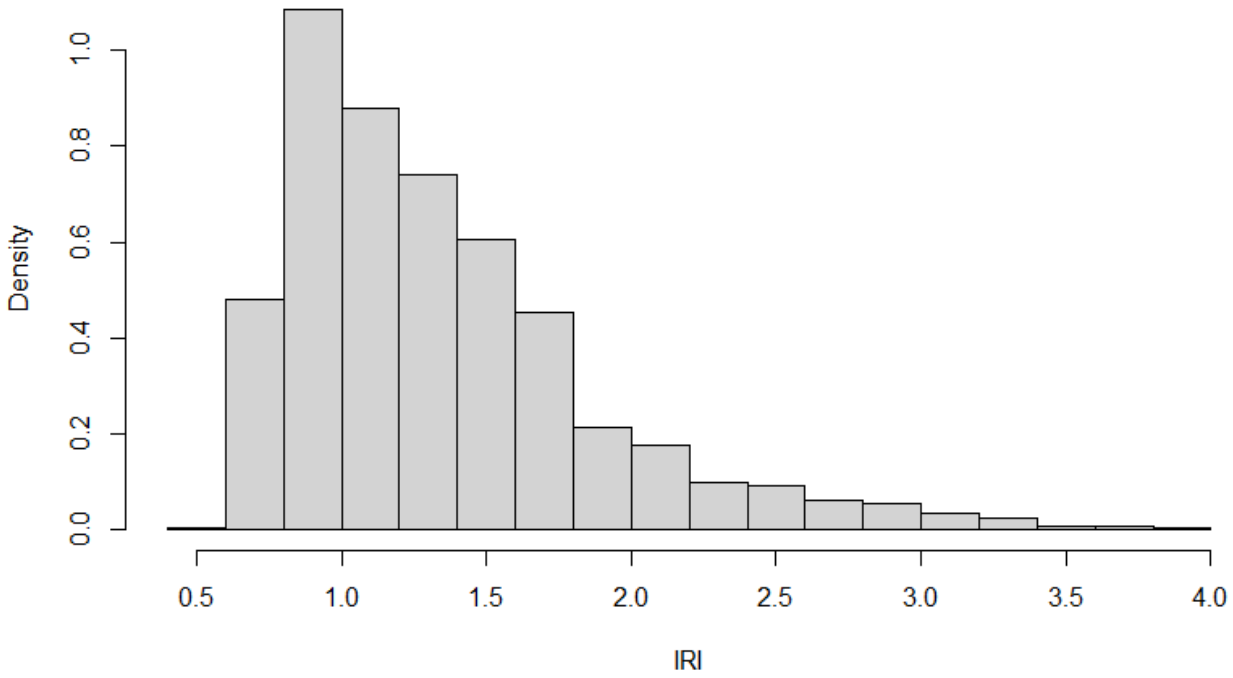

Figure A.69 – *IRI* values Boxplot for OL pavements – Full Dataset

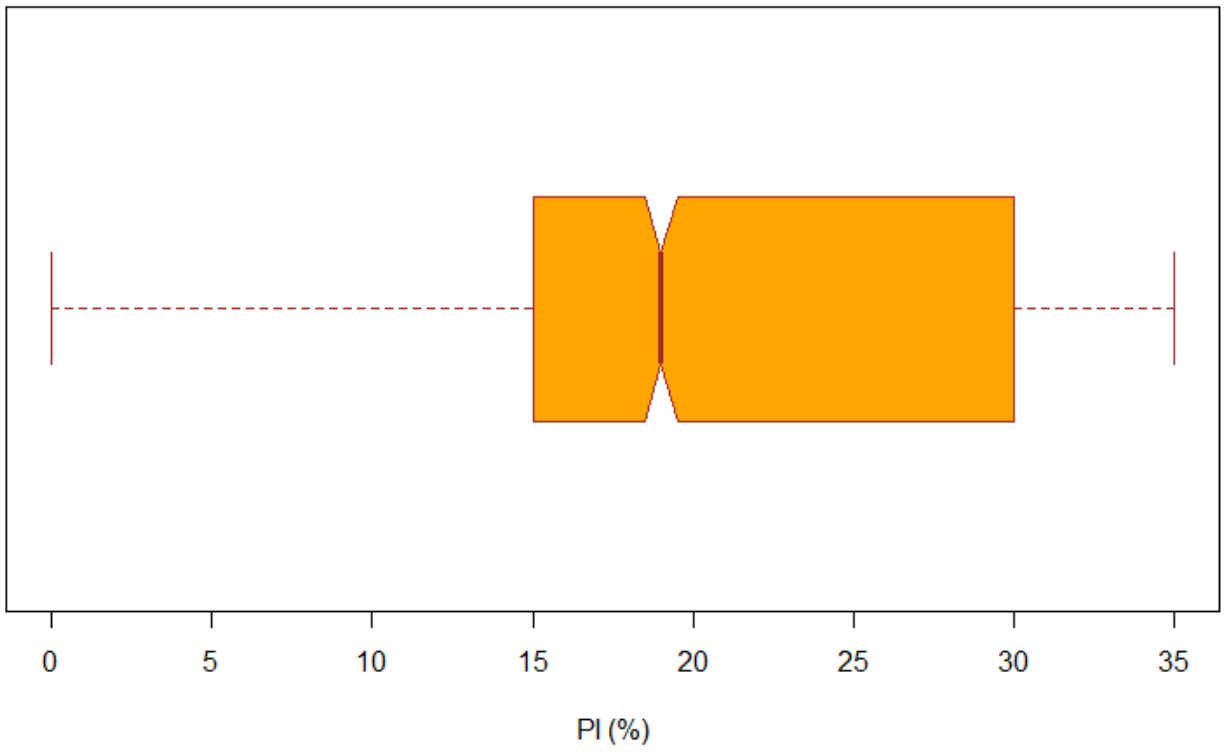

Figure A.70 – *PI* values Boxplot for OL pavements – Full Dataset

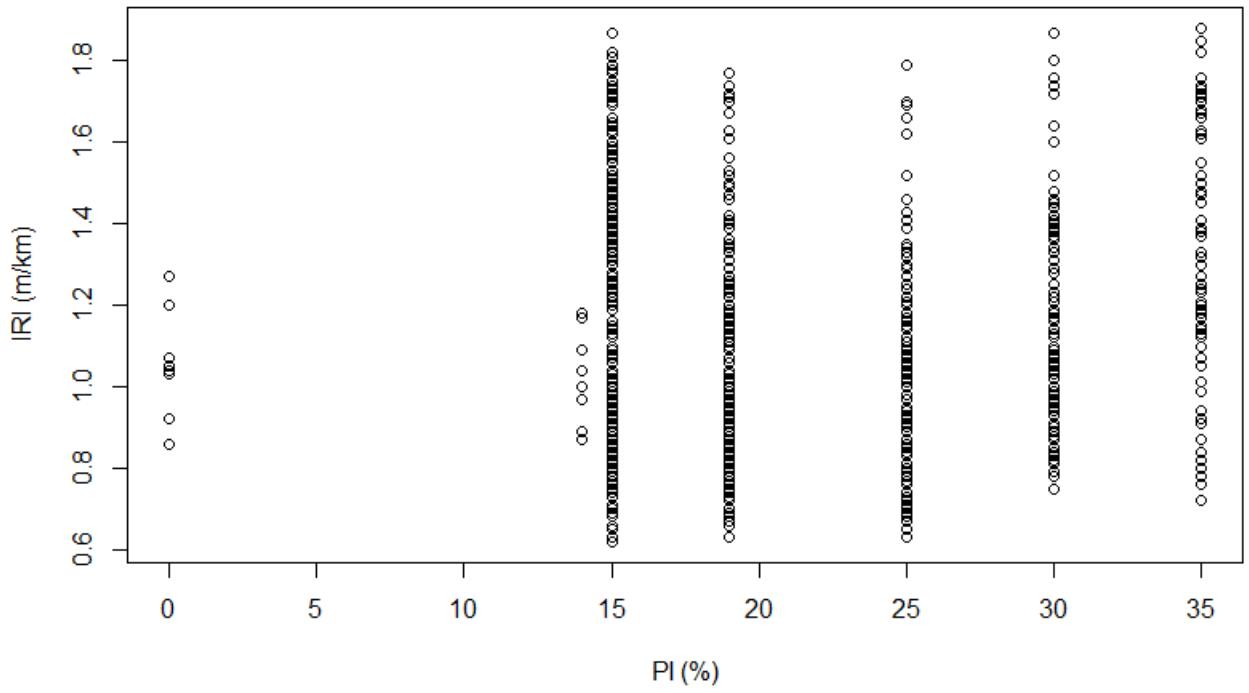

Figure A.71 – *PI* values scatter plot for OL pavements – inner-fence Dataset

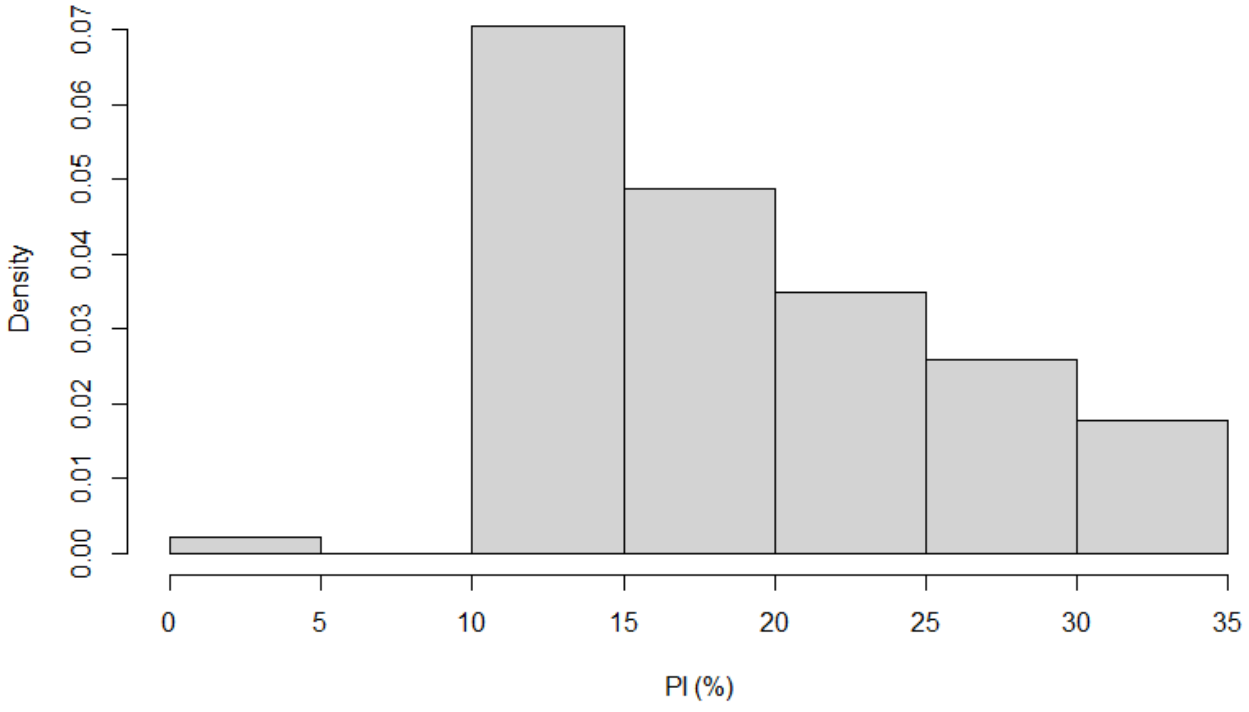

Figure A.72 – *PI* values histogram plot for OL pavements – inner-fence Dataset

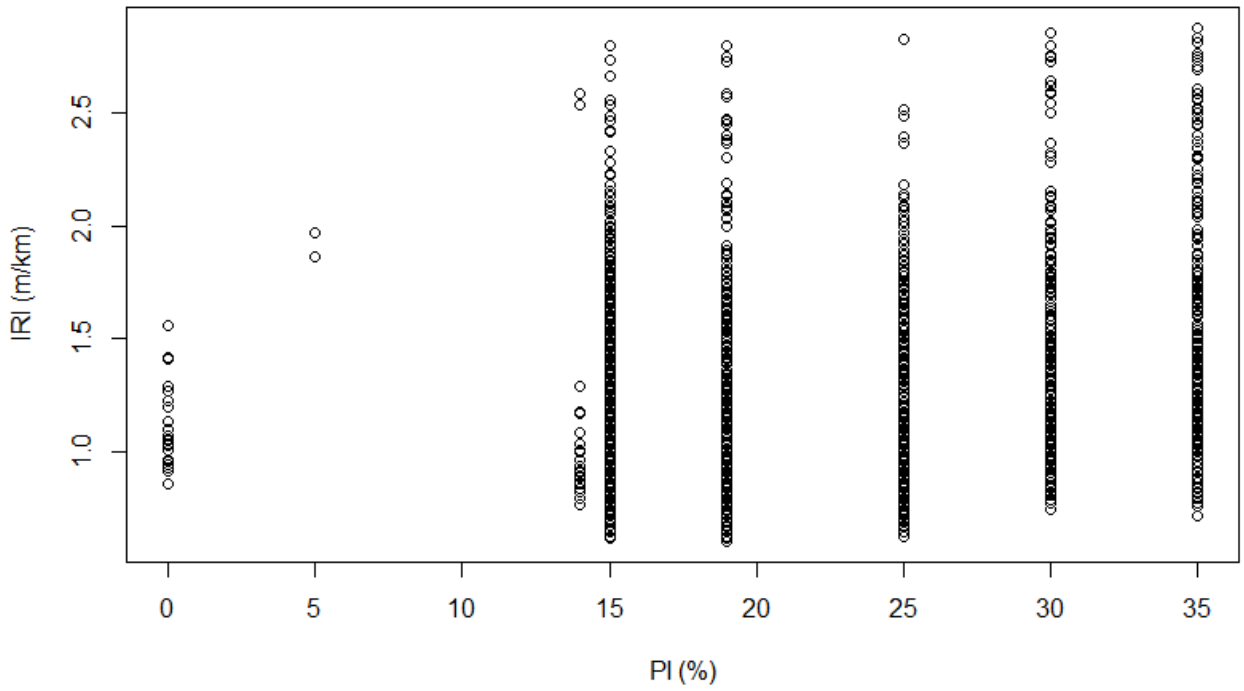

Figure A.73 – *PI* values scatter plot for OL pavements – Outer-fence Dataset

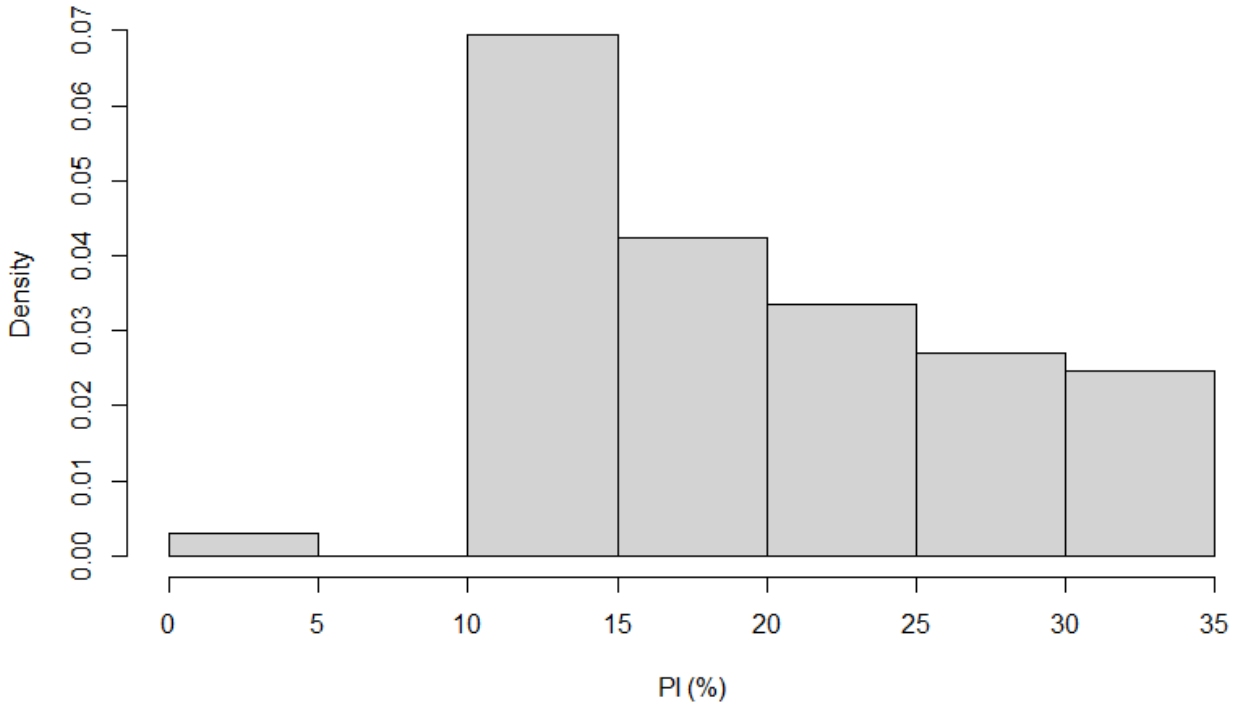

Figure A.74 – *PI* values histogram plot for OL pavements – Outer-fence Dataset

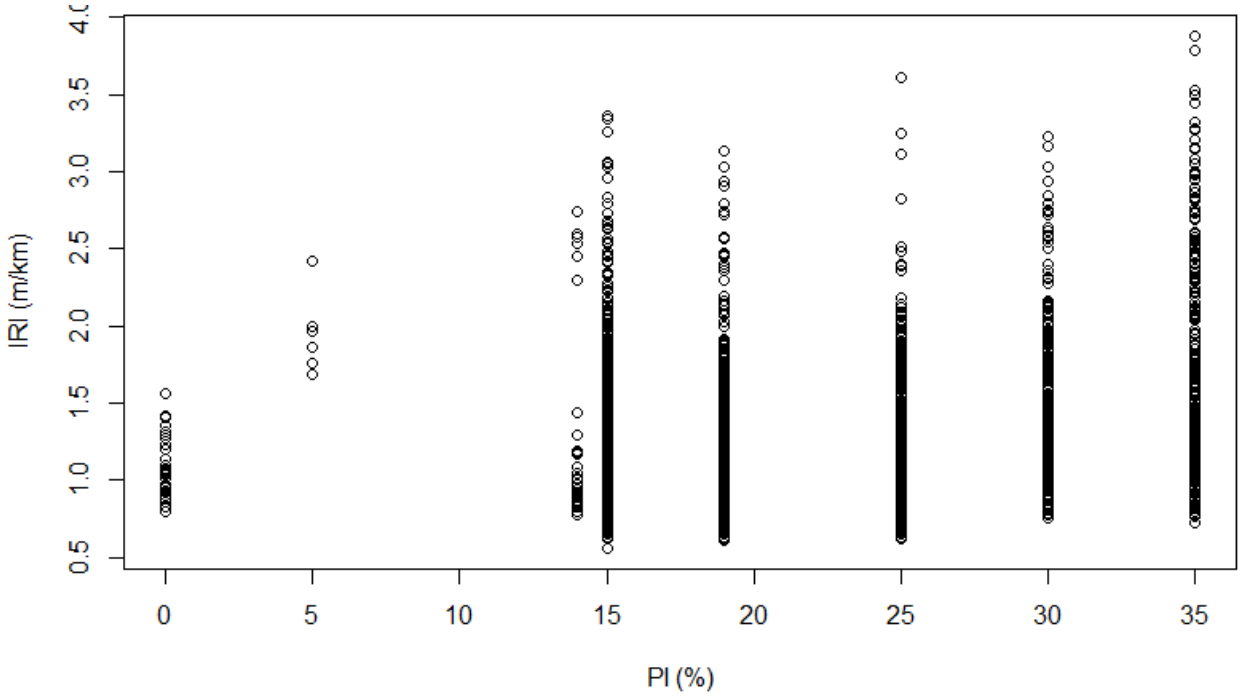

Figure A.75 – *PI* values scatter plot for OL pavements – Full Dataset

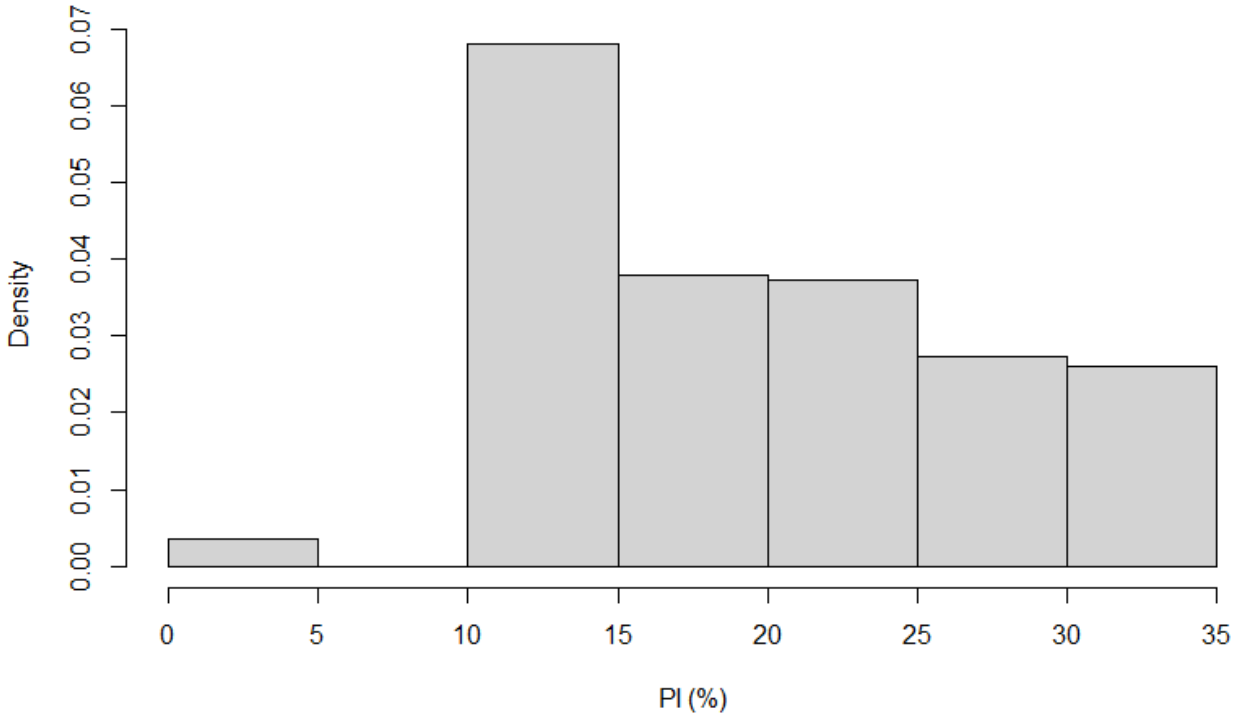

Figure A.76 – *PI* values histogram plot for OL pavements – Full Dataset

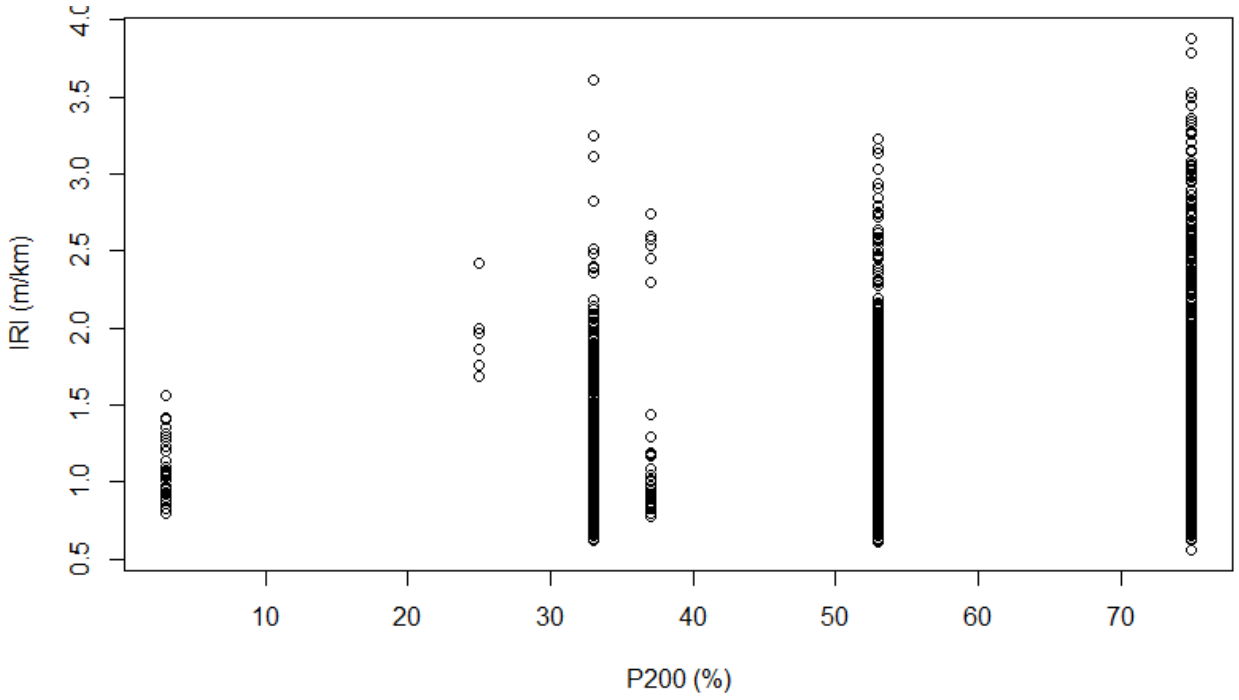

Figure A.77 – P_{200} values Boxplot for OL pavements – Full Dataset

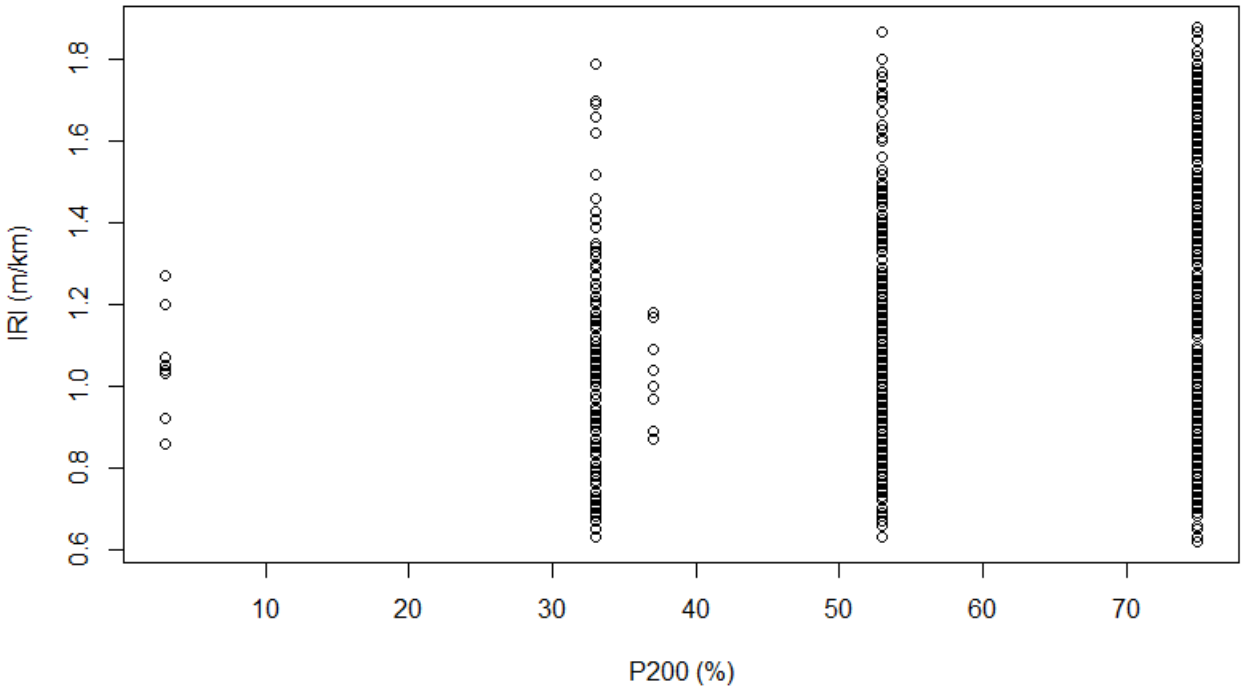

Figure A.78 – P_{200} values scatter plot for OL pavements – inner-fence Dataset

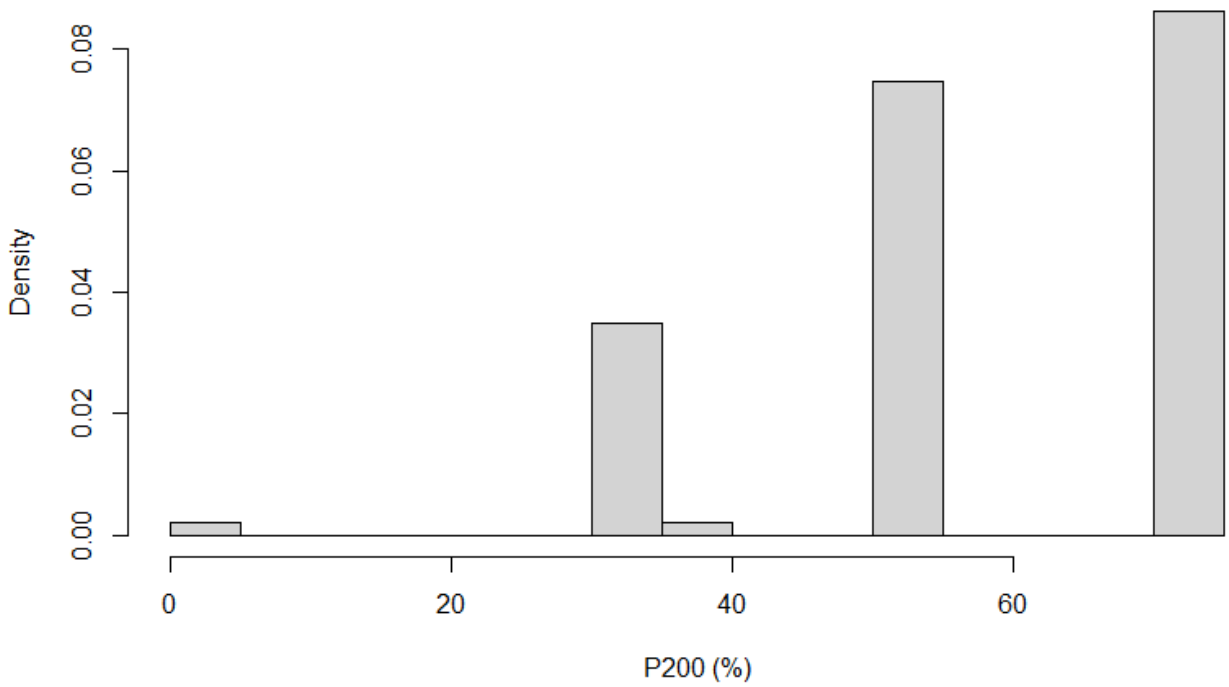

Figure A.79 – P_{200} values histogram plot for OL pavements – inner-fence Dataset

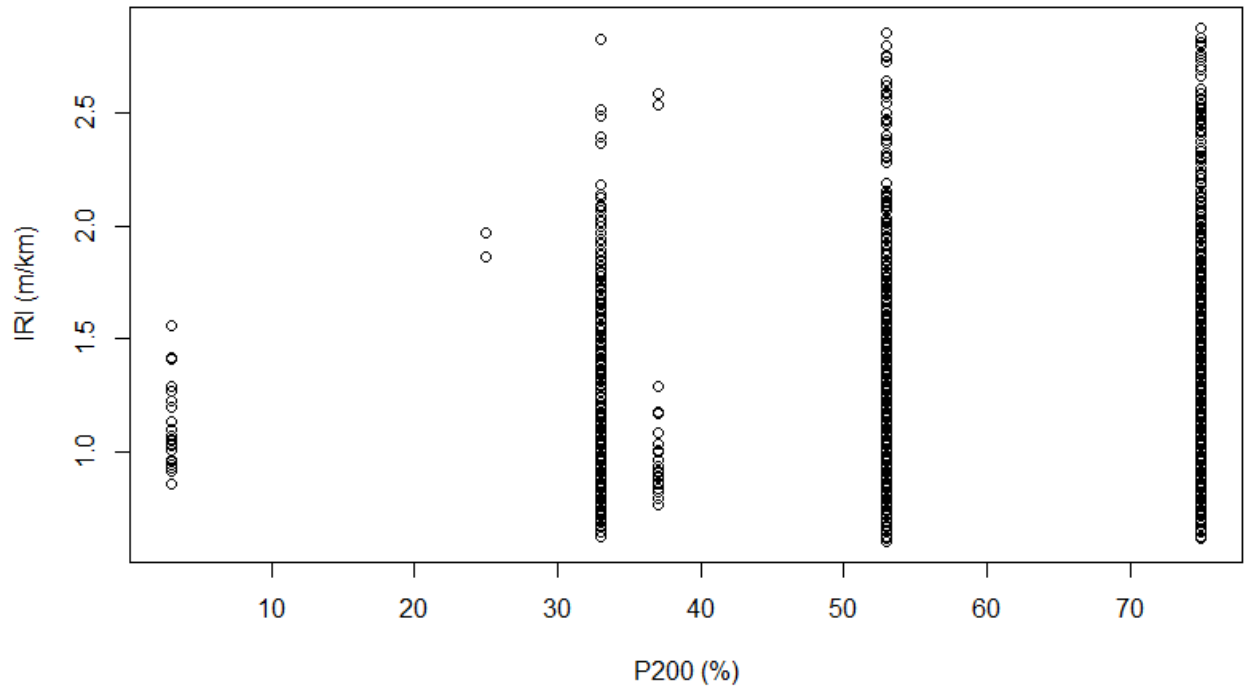

Figure A.80 – P_{200} values scatter plot for OL pavements – Outer-fence Dataset

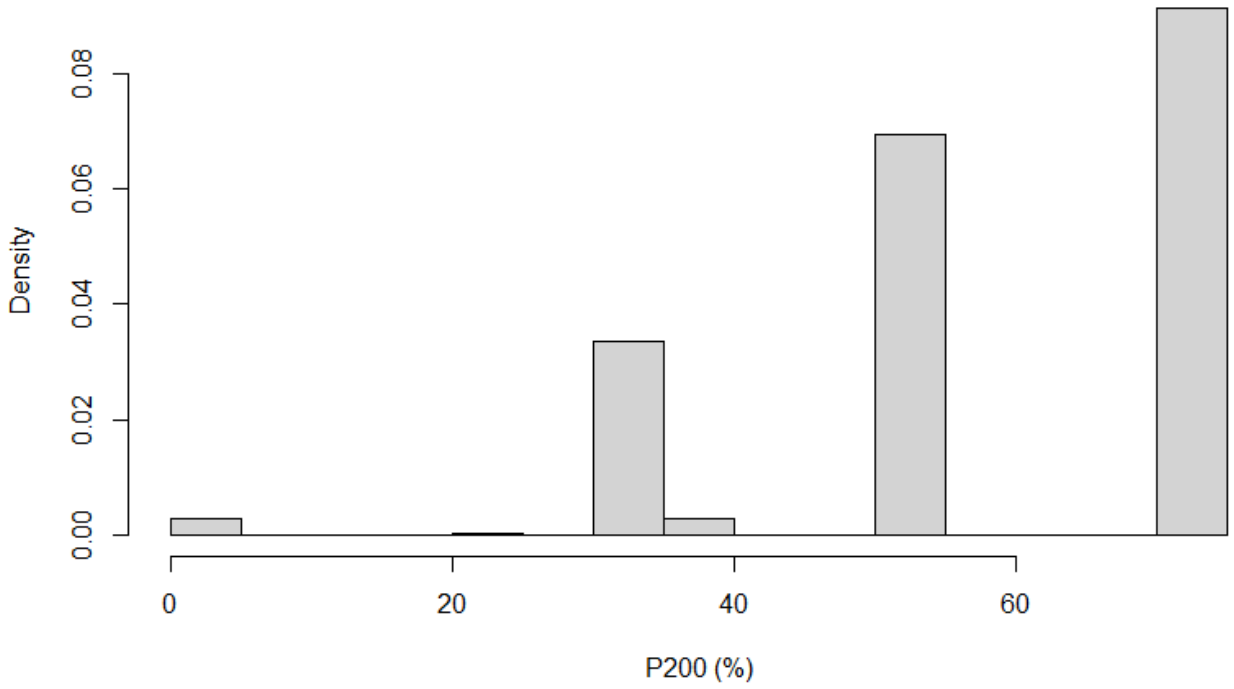

Figure A.81 – P_{200} values histogram plot for OL pavements – Outer-fence Dataset

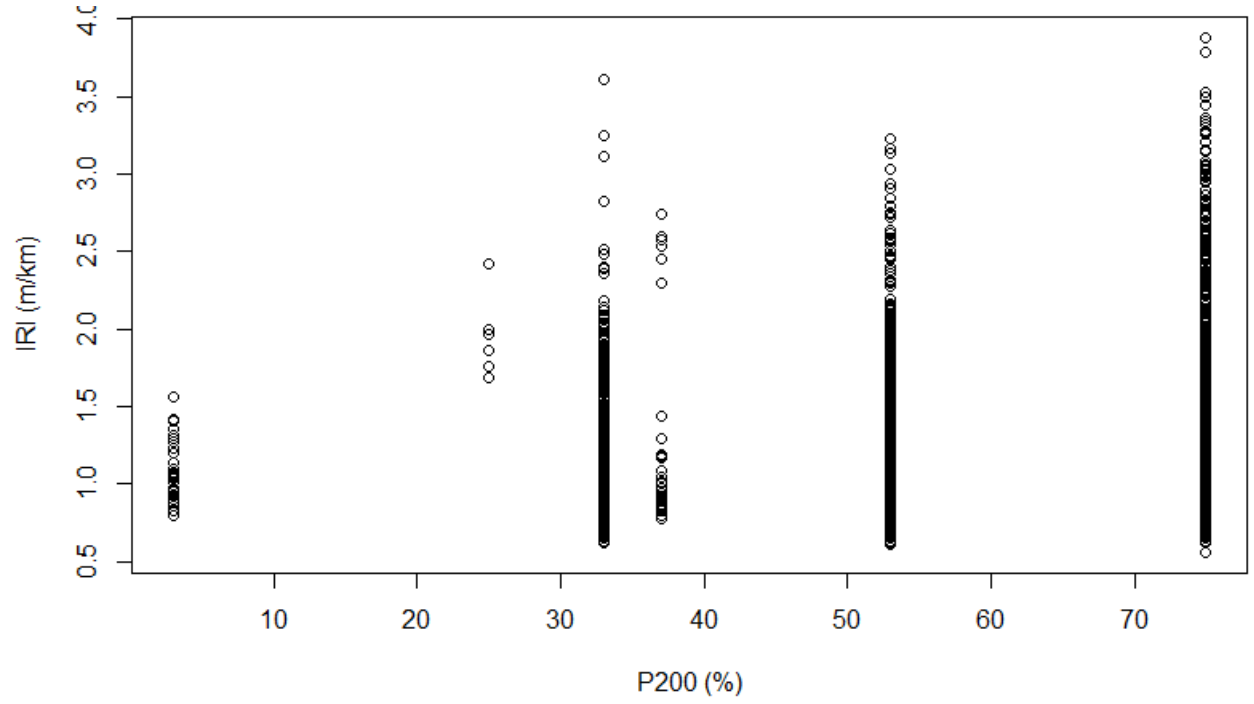

Figure A.82 – P_{200} values scatter plot for OL pavements – Full Dataset

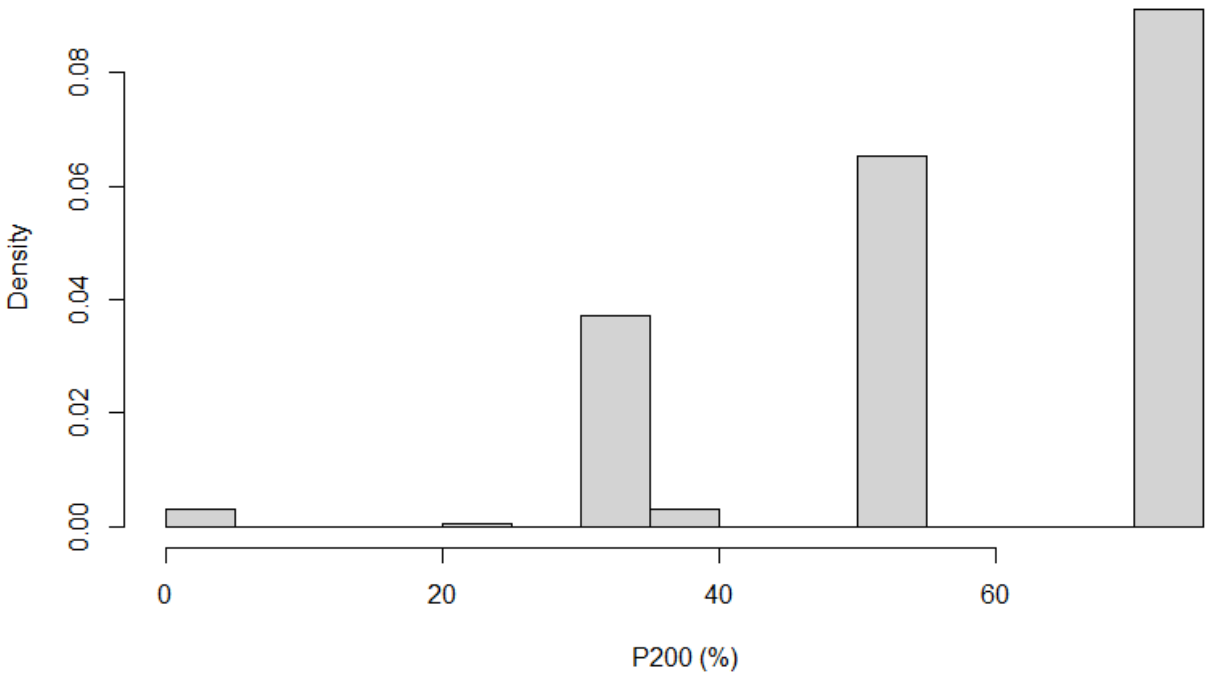

Figure A.83 – P_{200} values histogram plot for OL pavements – Full Dataset

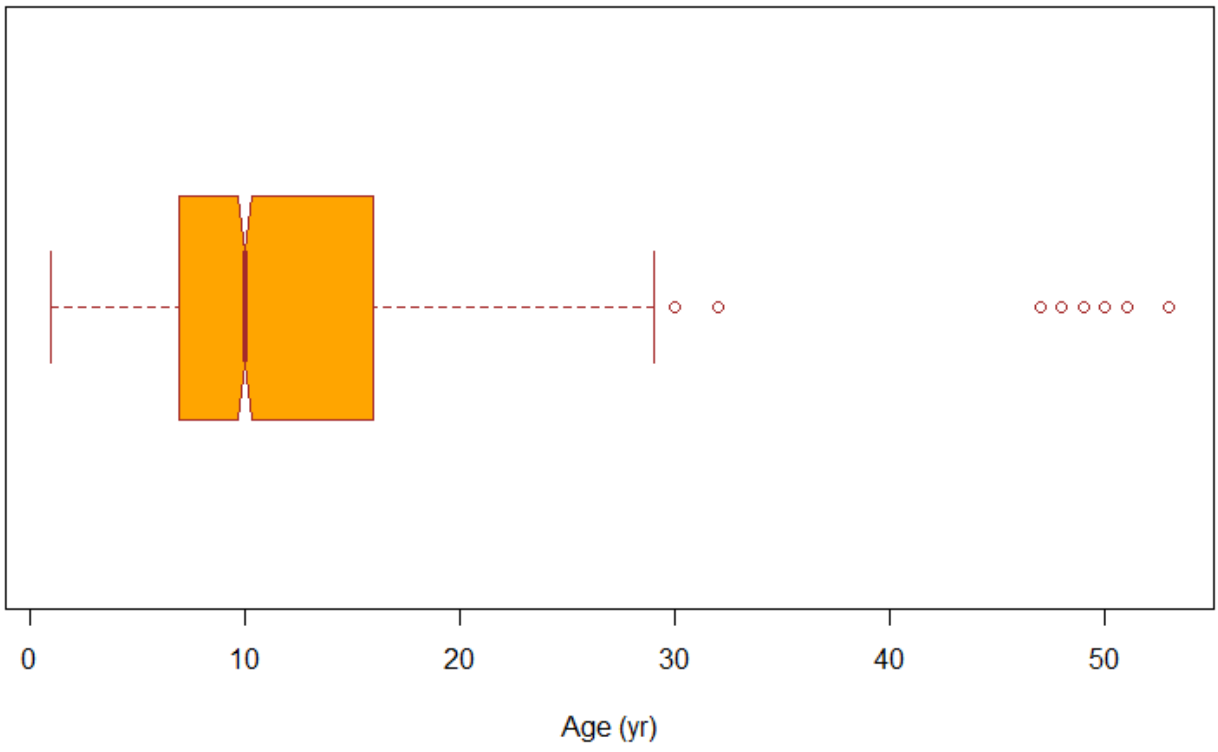

Figure A.84 – *Age* values Boxplot for OL pavements – Full Dataset

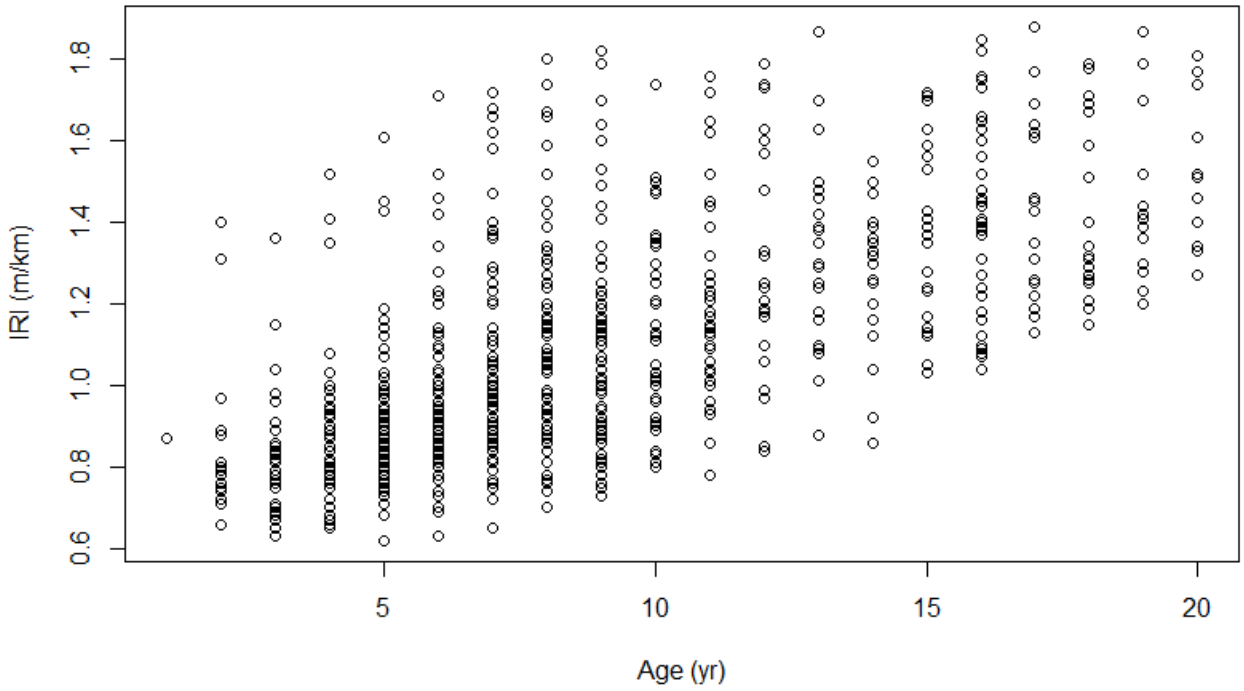

Figure A.85 – *Age* values scatter plot for OL pavements – inner-fence Dataset

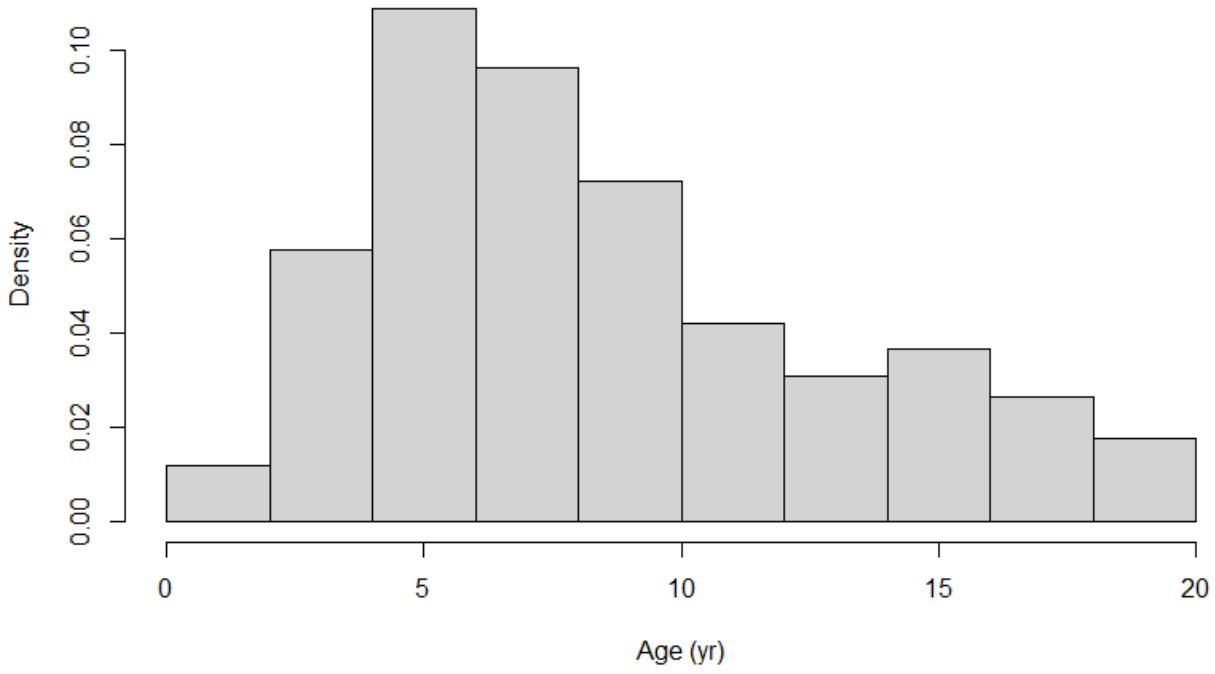

Figure A.86 – *Age* values histogram plot for OL pavements – inner-fence Dataset

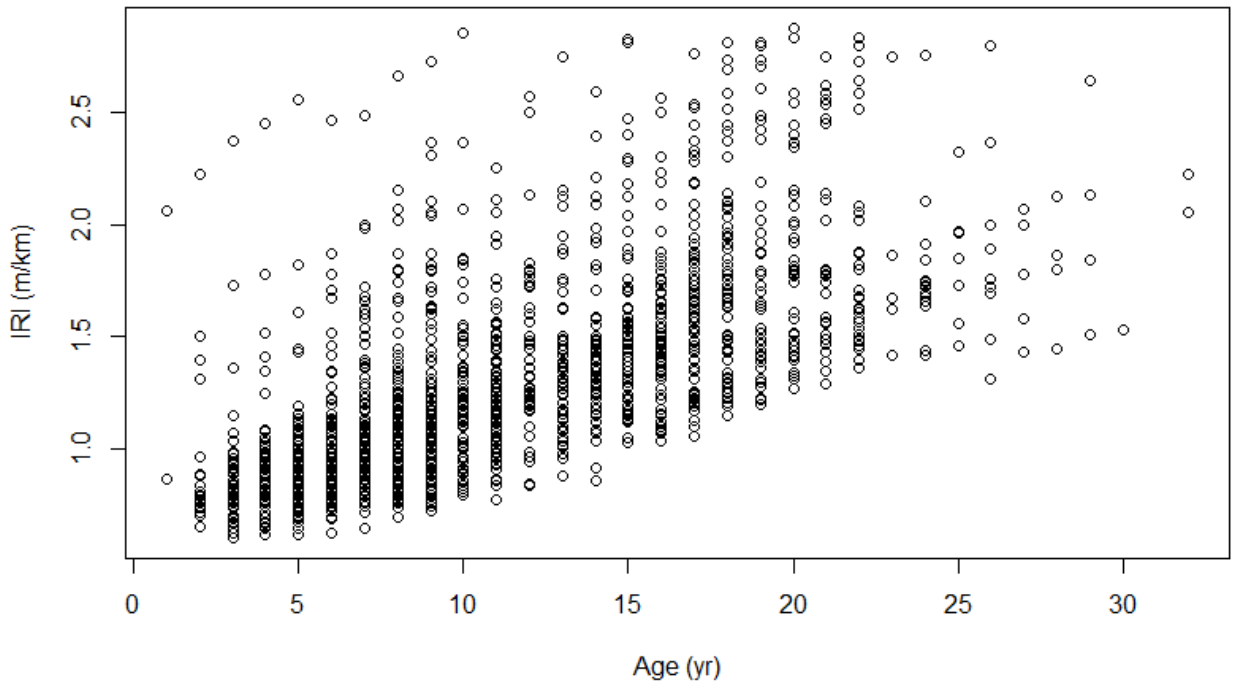

Figure A.87 – *Age* values scatter plot for OL pavements – Outer-fence Dataset

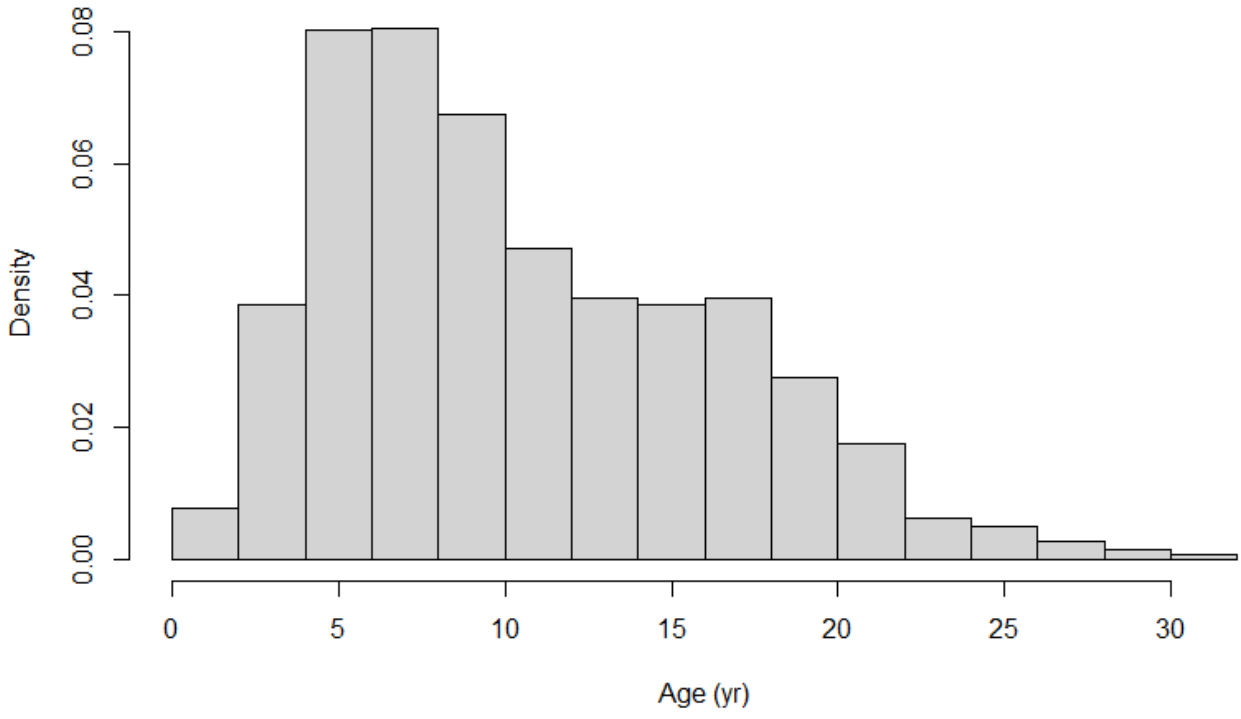

Figure A.88 – *Age* values histogram plot for OL pavements – Outer-fence Dataset

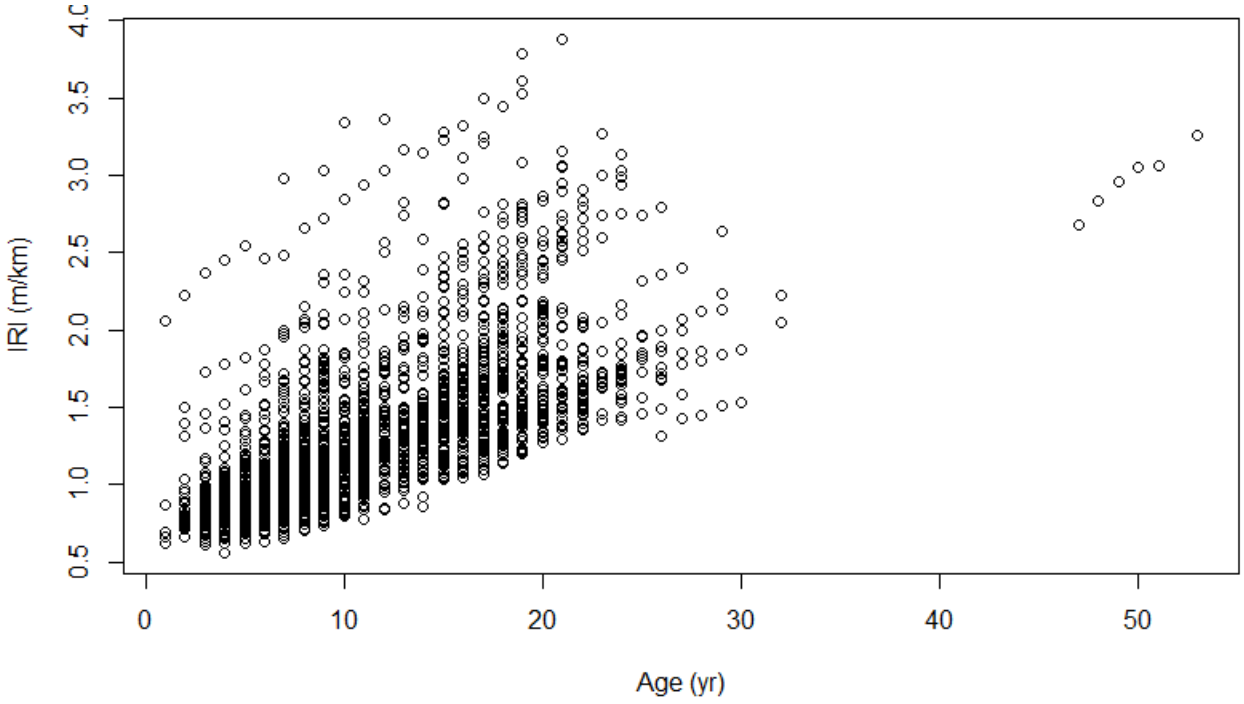

Figure A.89 – *Age* values scatter plot for OL pavements – Full Dataset

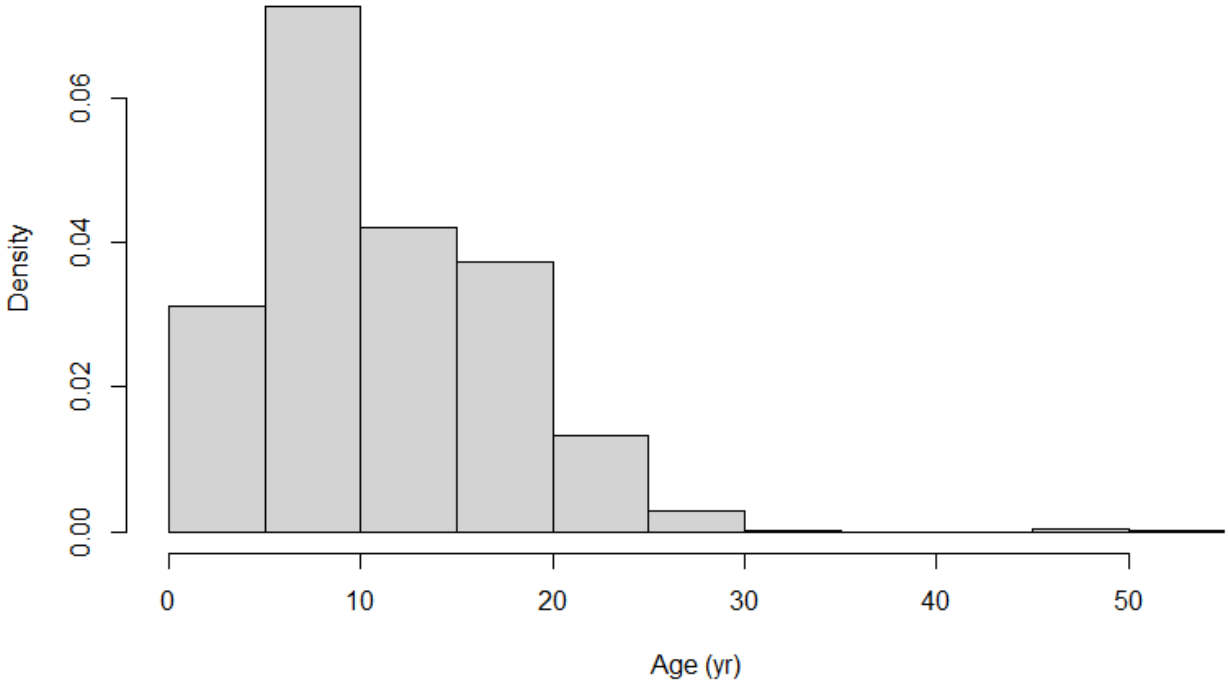

Figure A.90 – *Age* values histogram plot for OL pavements – Full Dataset

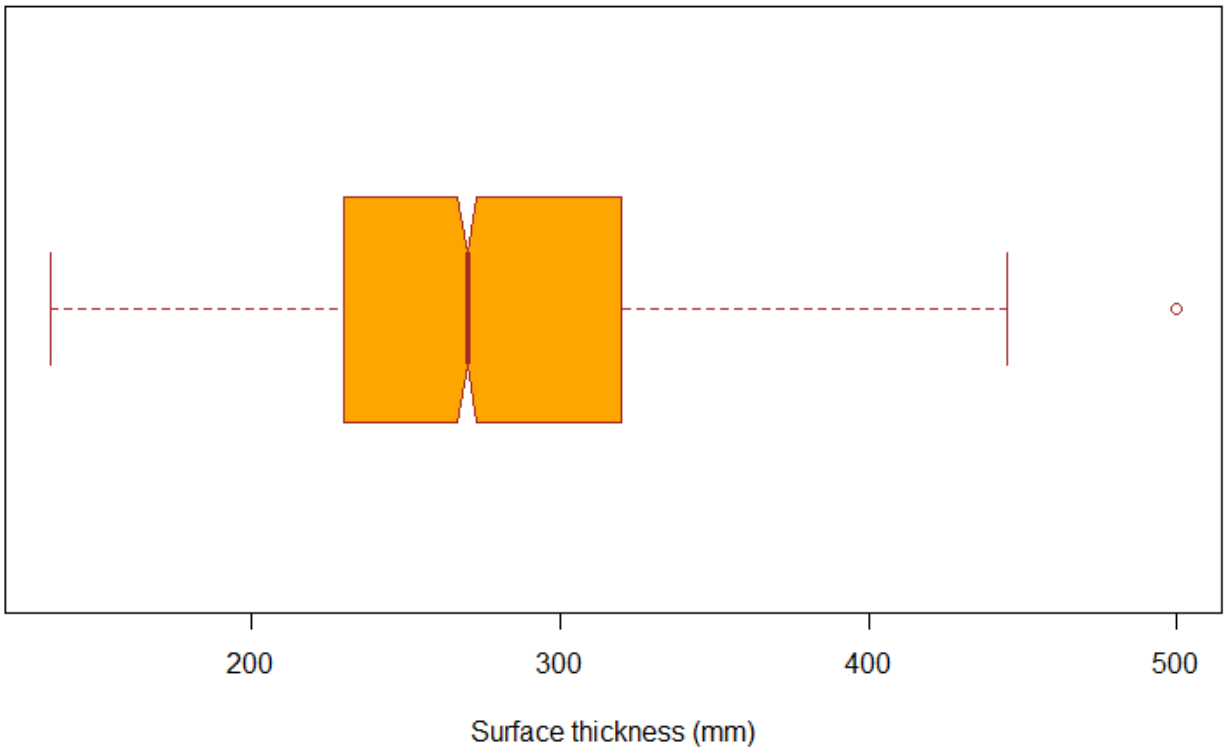

Figure A.91 – *Surfthickness* values Boxplot for OL pavements – Full Dataset

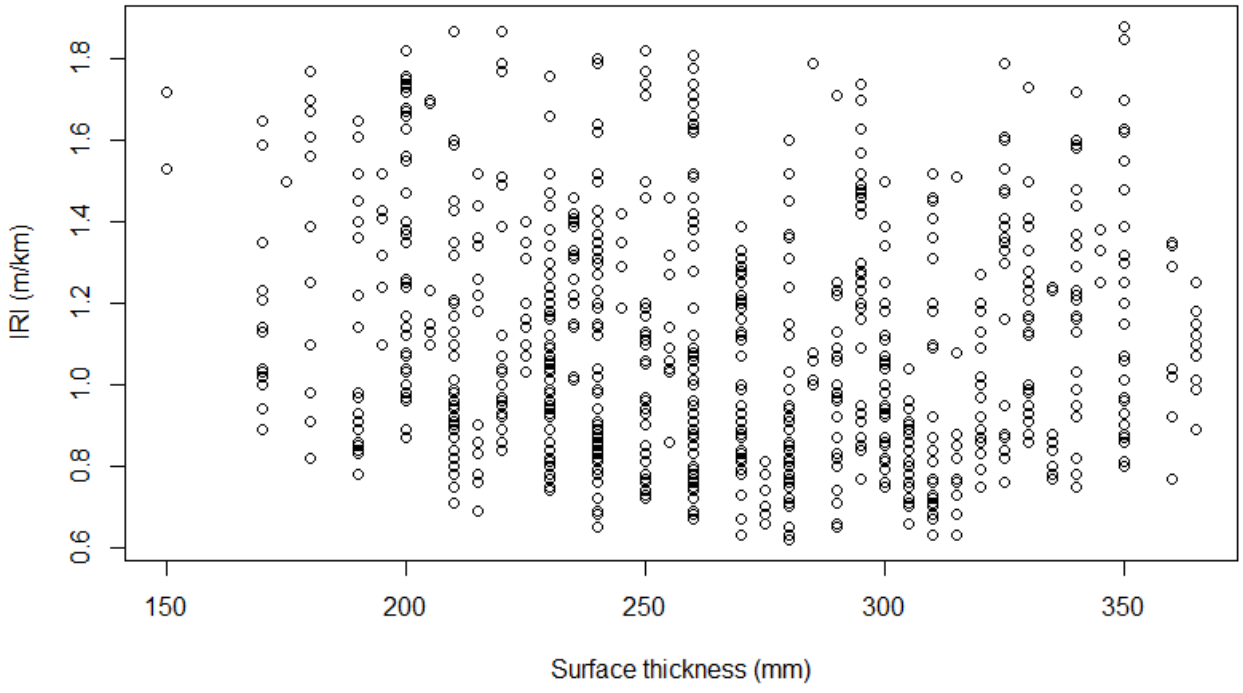

Figure A.92 – *Surfthickness* values scatter plot for OL pavements – inner-fence Dataset

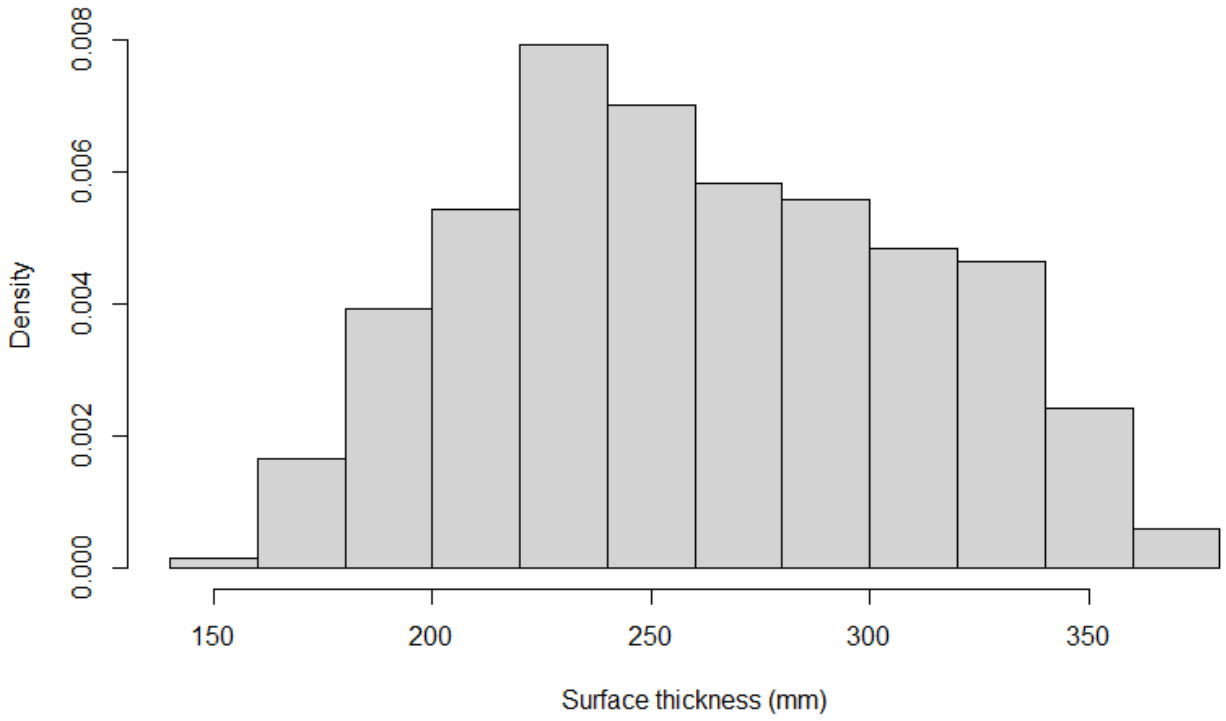

Figure A.93 – *Surfthickness* values histogram plot for OL pavements – inner-fence Dataset

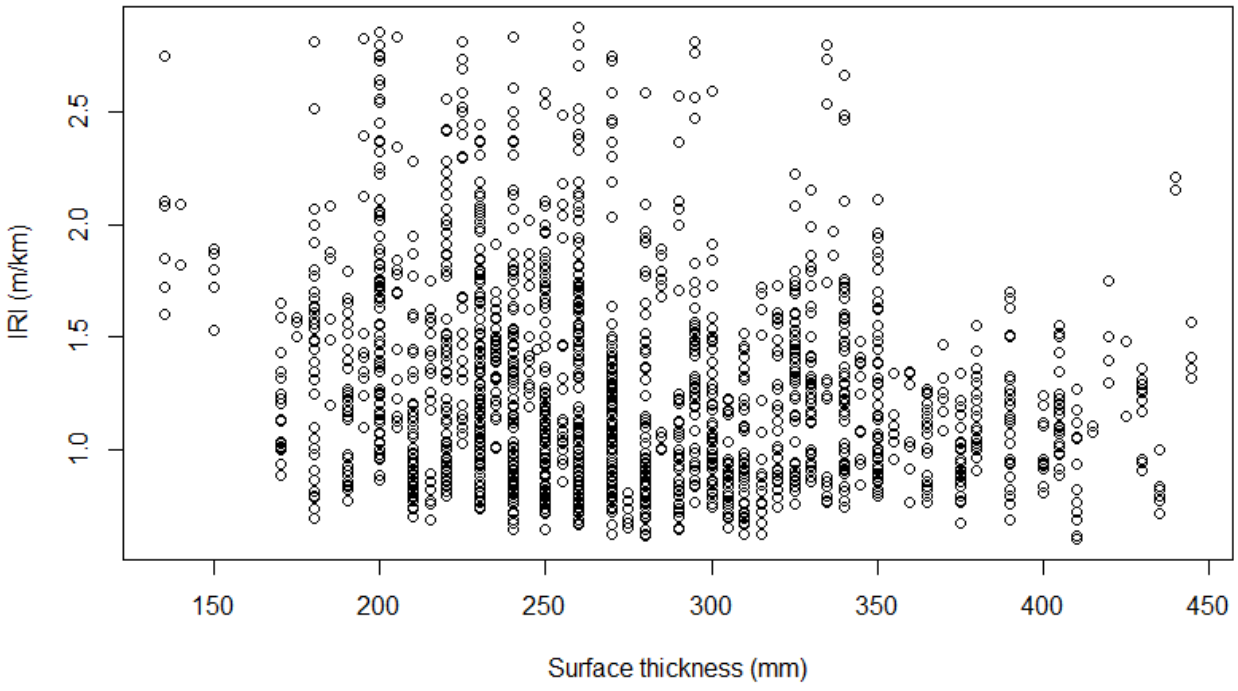

Figure A.94 – *Surfthickness* values scatter plot for OL pavements – Outer-fence Dataset

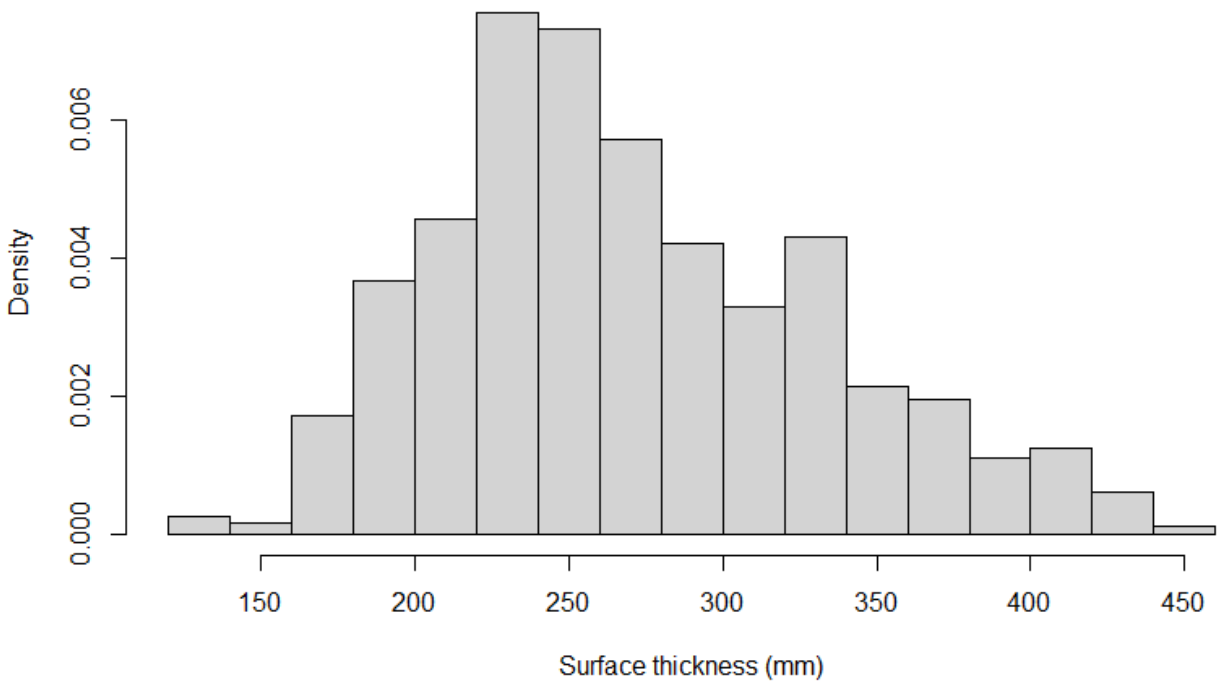

Figure A.95 – *Surfthickness* values histogram plot for OL pavements – Outer-fence Dataset

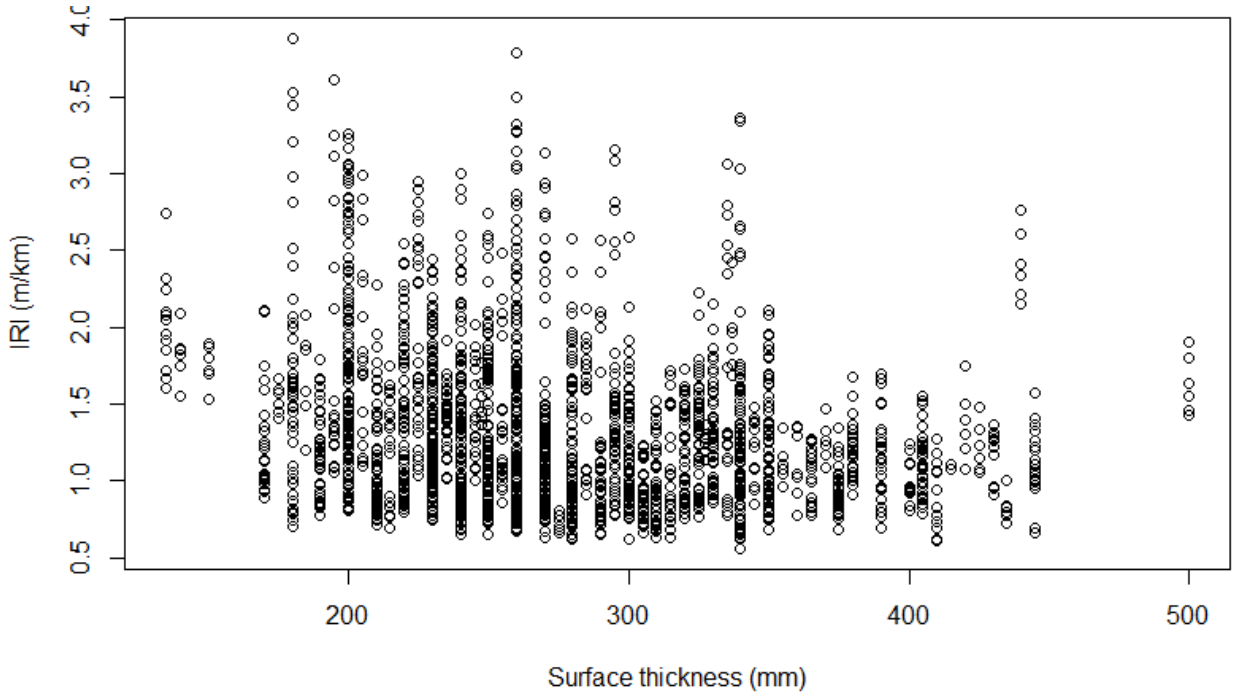

Figure A.96 – *Surfthickness* values scatter plot for OL pavements – Full Dataset

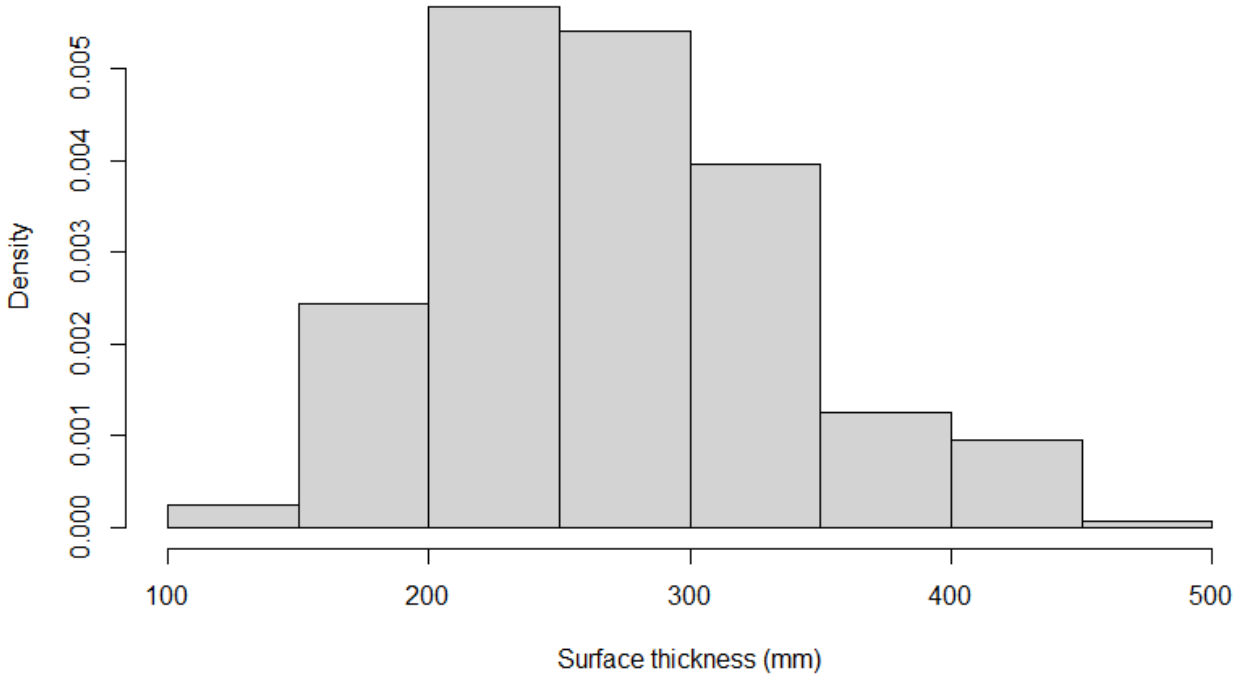

Figure A.97 – *Surfthickness* values histogram plot for OL pavements – Full Dataset

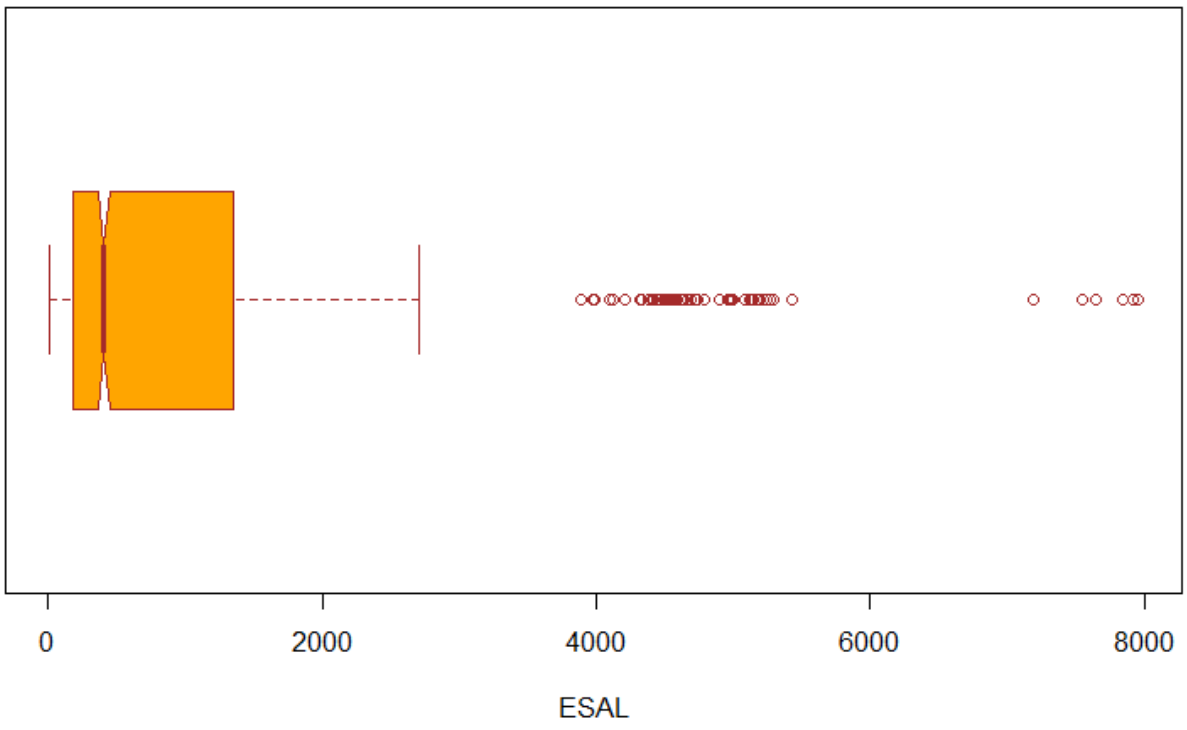

Figure A.98 – *ESAL* values Boxplot for OL pavements – Full Dataset

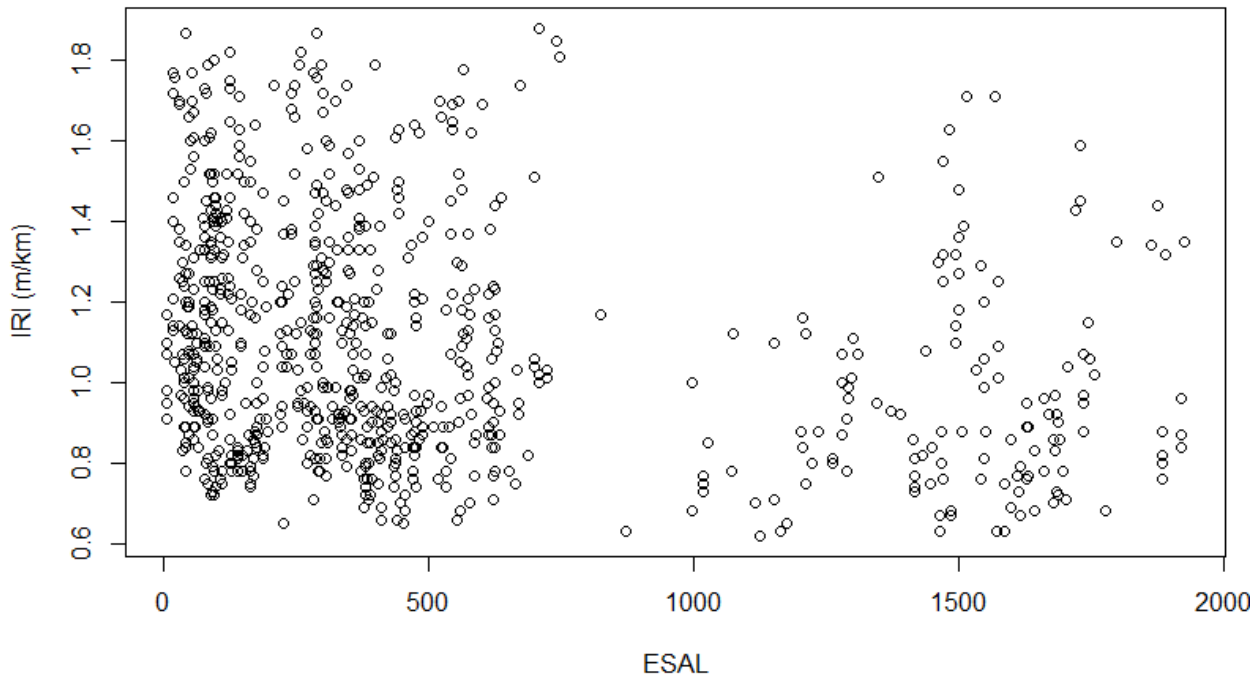

Figure A.99 – *ESAL* values scatter plot for OL pavements – inner-fence Dataset

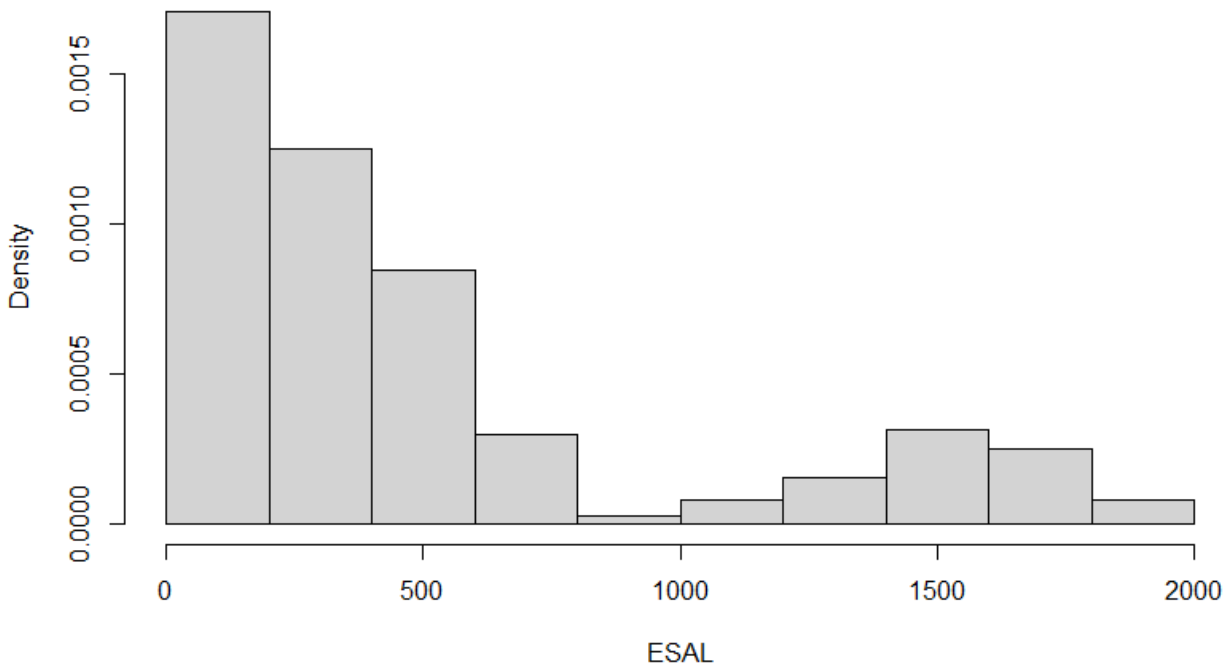

Figure A.100 – *ESAL* values histogram plot for OL pavements – inner-fence Dataset

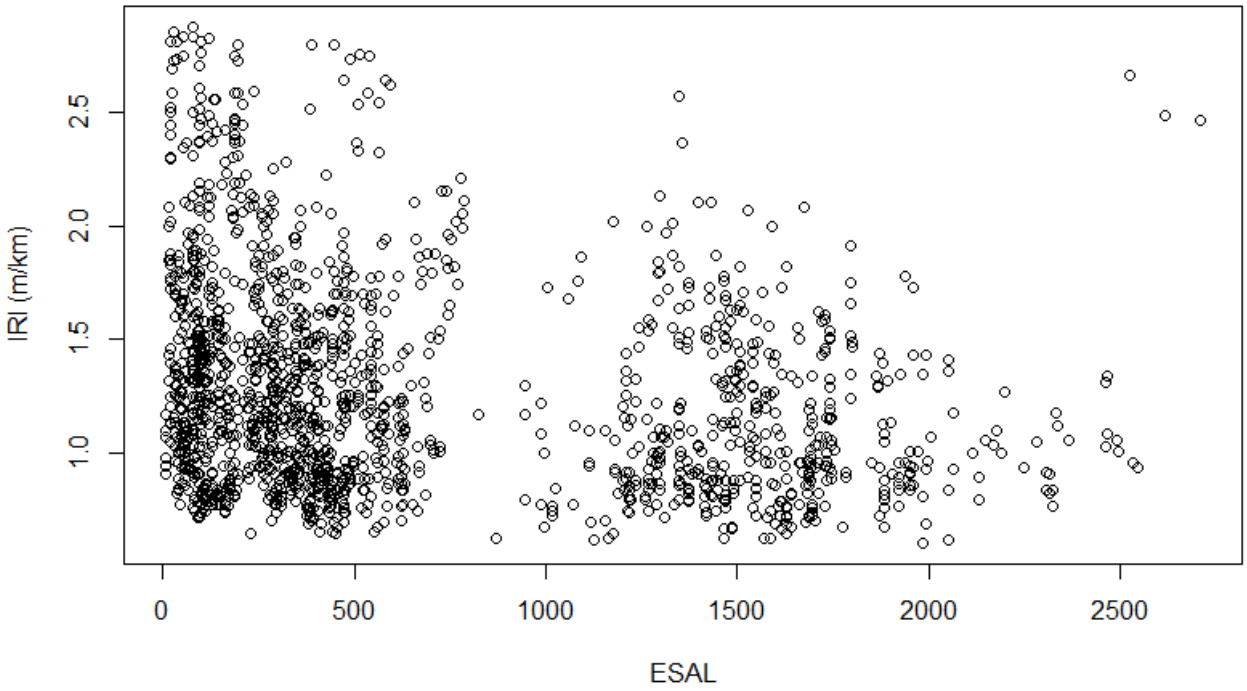

Figure A.101 – *ESAL* values scatter plot for OL pavements – Outer-fence Dataset

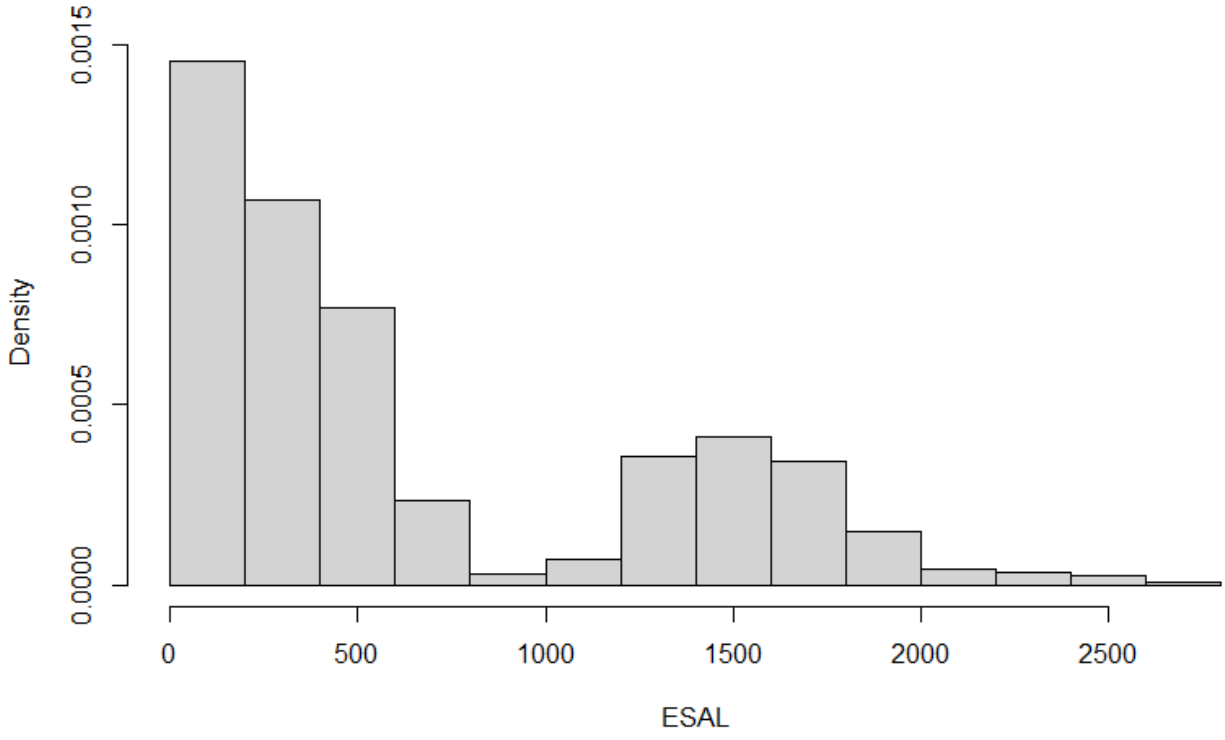

Figure A.102 – *ESAL* values histogram plot for OL pavements – Outer-fence Dataset

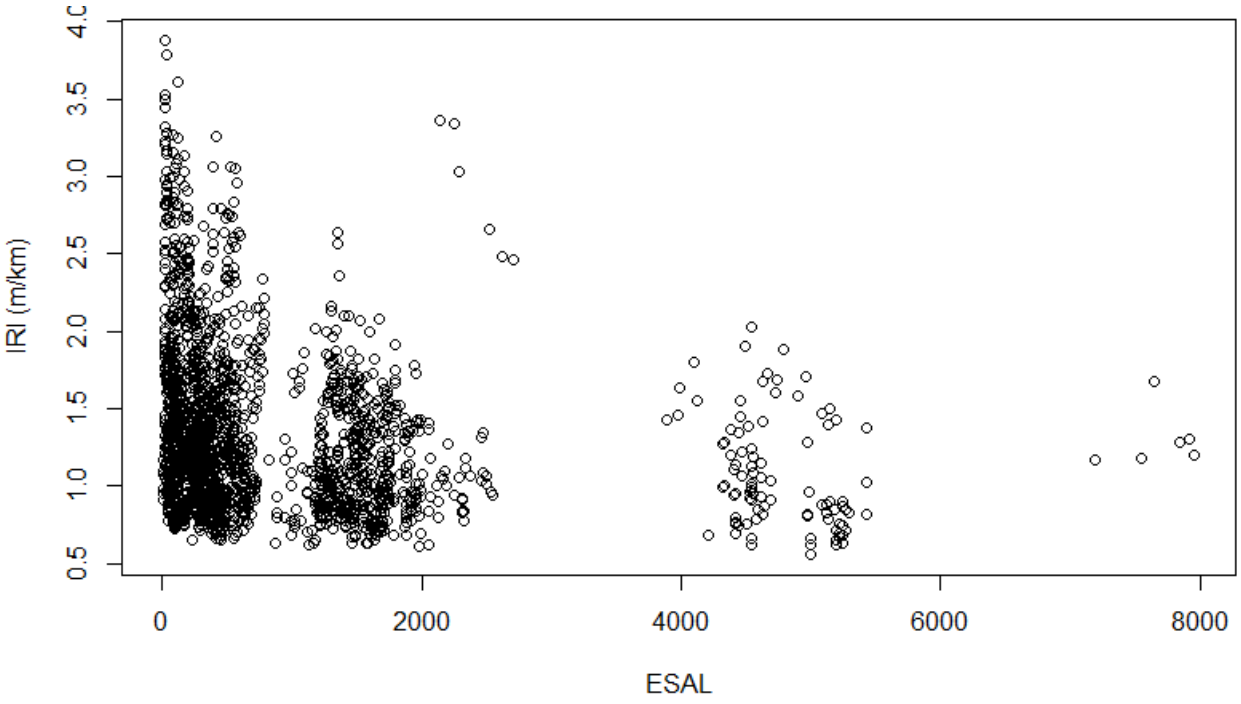

Figure A.103 – *ESAL* values scatter plot for OL pavements – Full Dataset

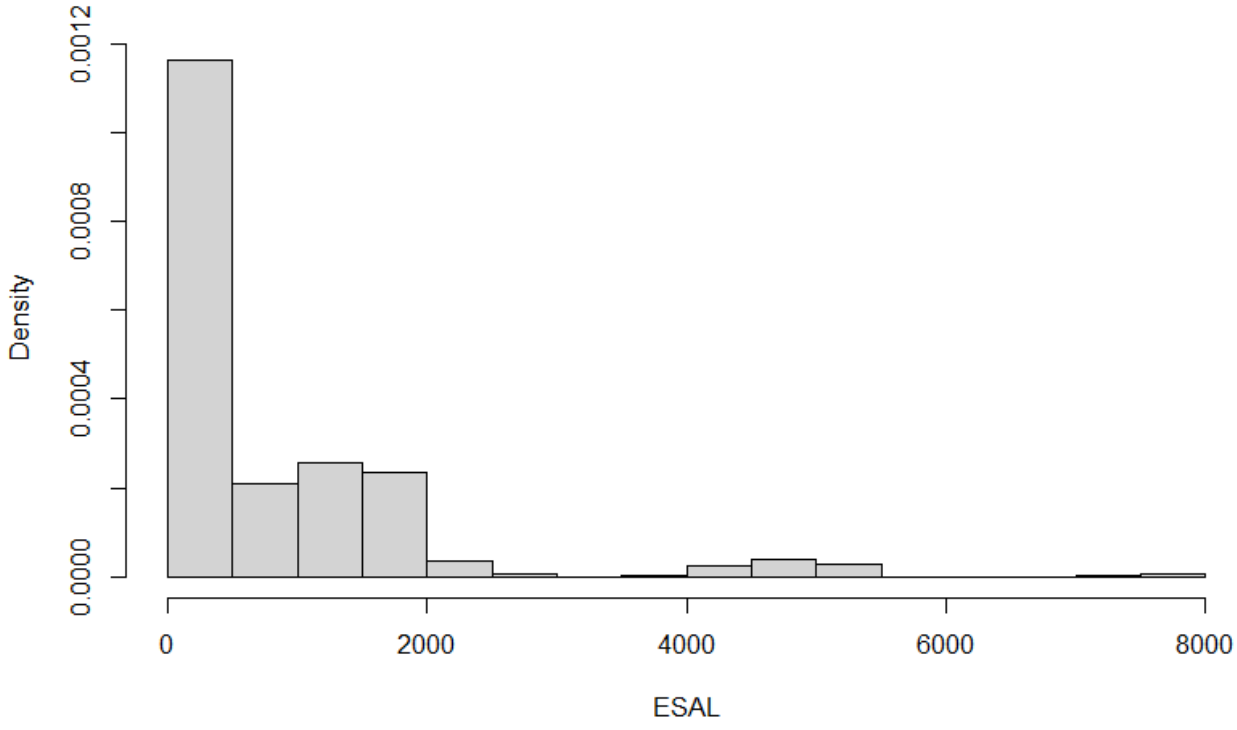

Figure A.104 – *ESAL* values histogram plot for OL pavements – Full Dataset

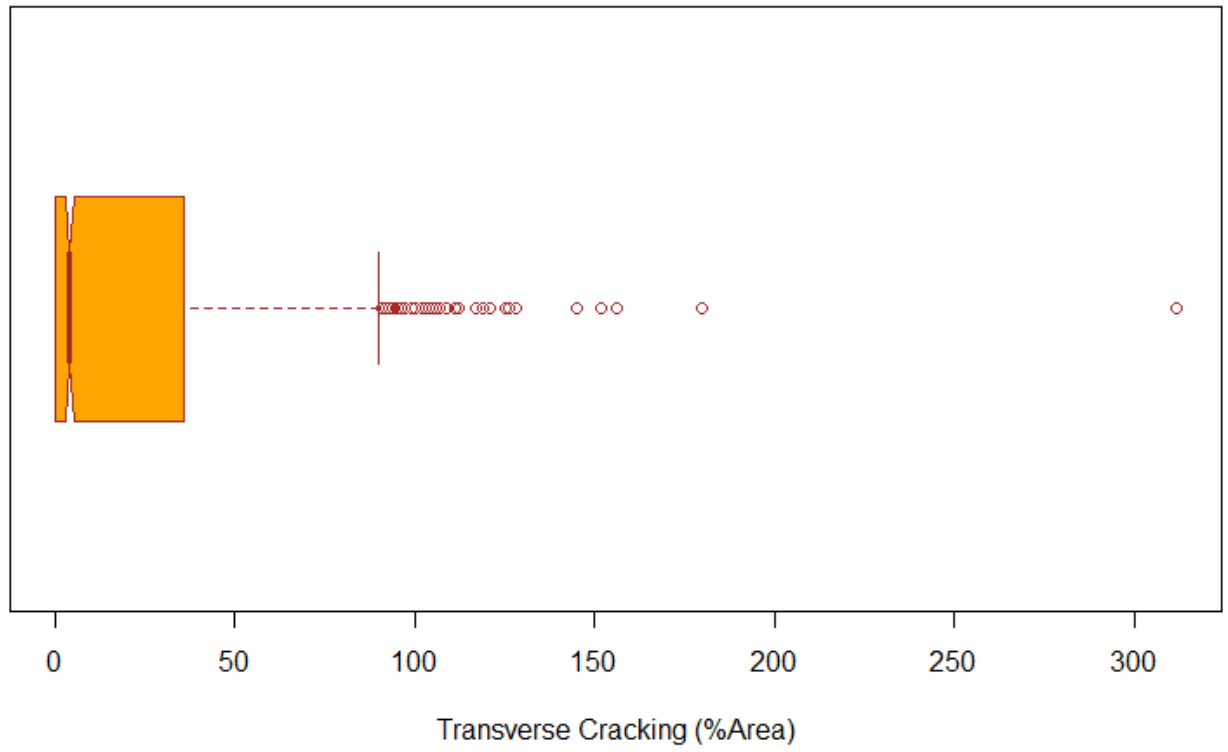

Figure A.105 – *TrcAr* values Boxplot for OL pavements – Full Dataset

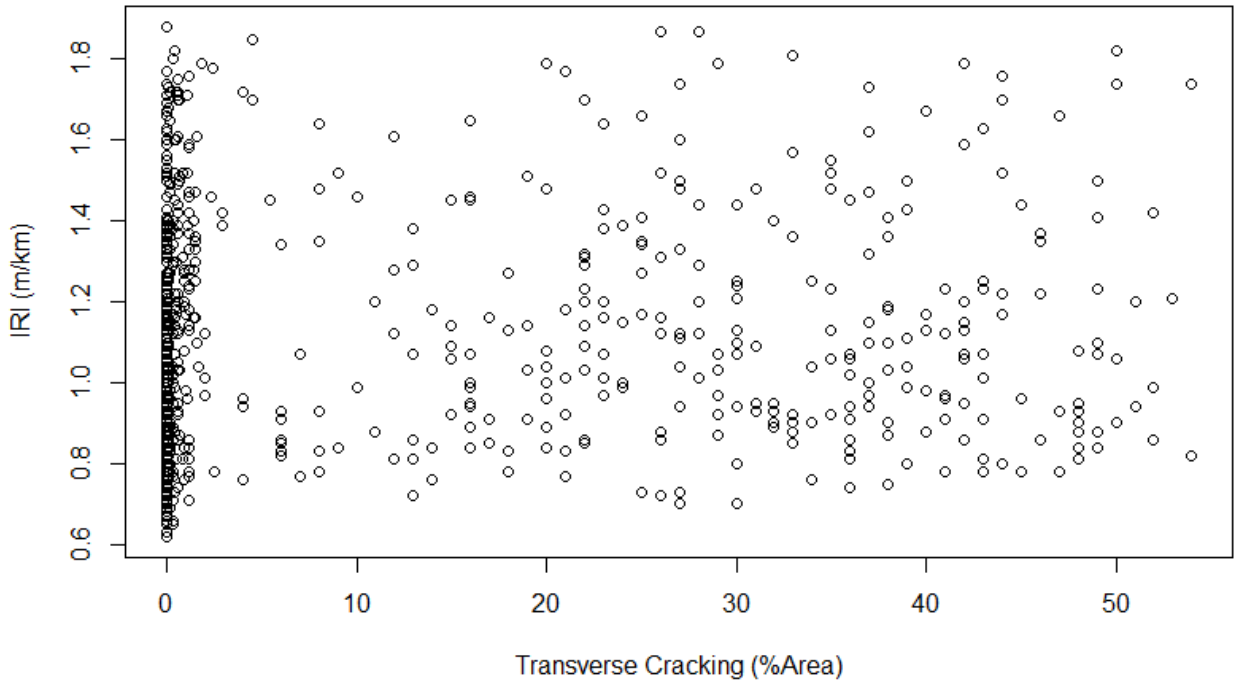

Figure A.106 – *TrcAr* values scatter plot for OL pavements – inner-fence Dataset

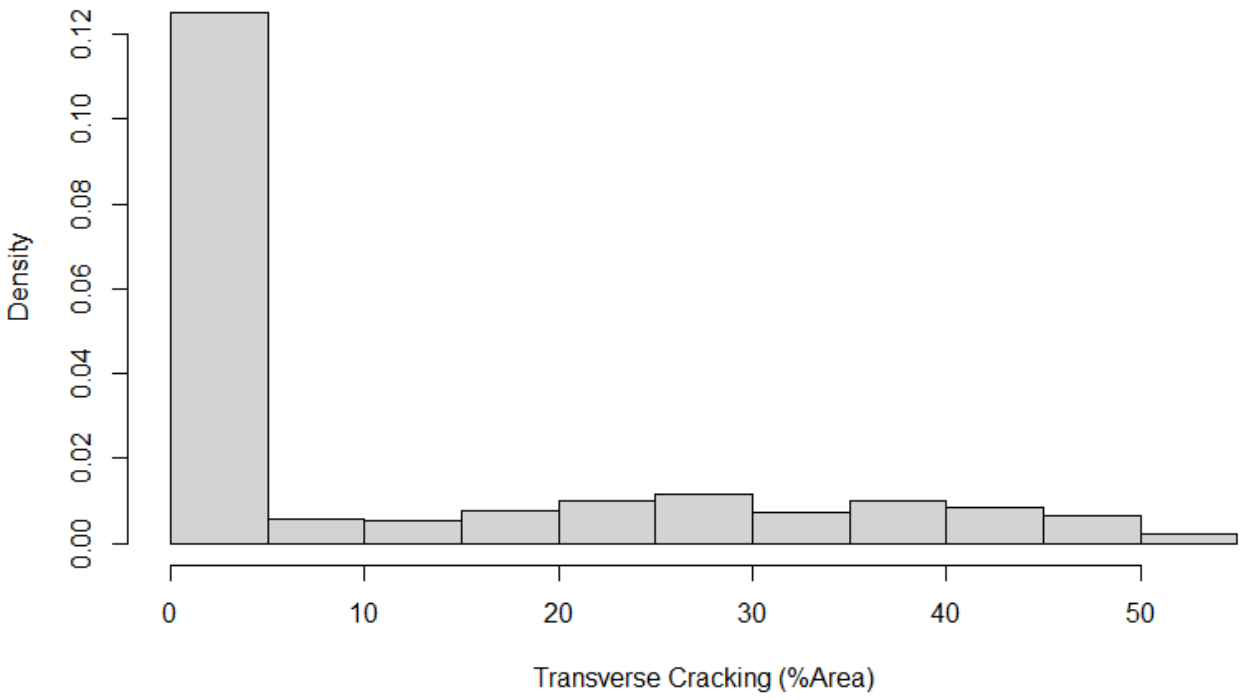

Figure A.107 – *TrcAr* values histogram plot for OL pavements – inner-fence Dataset

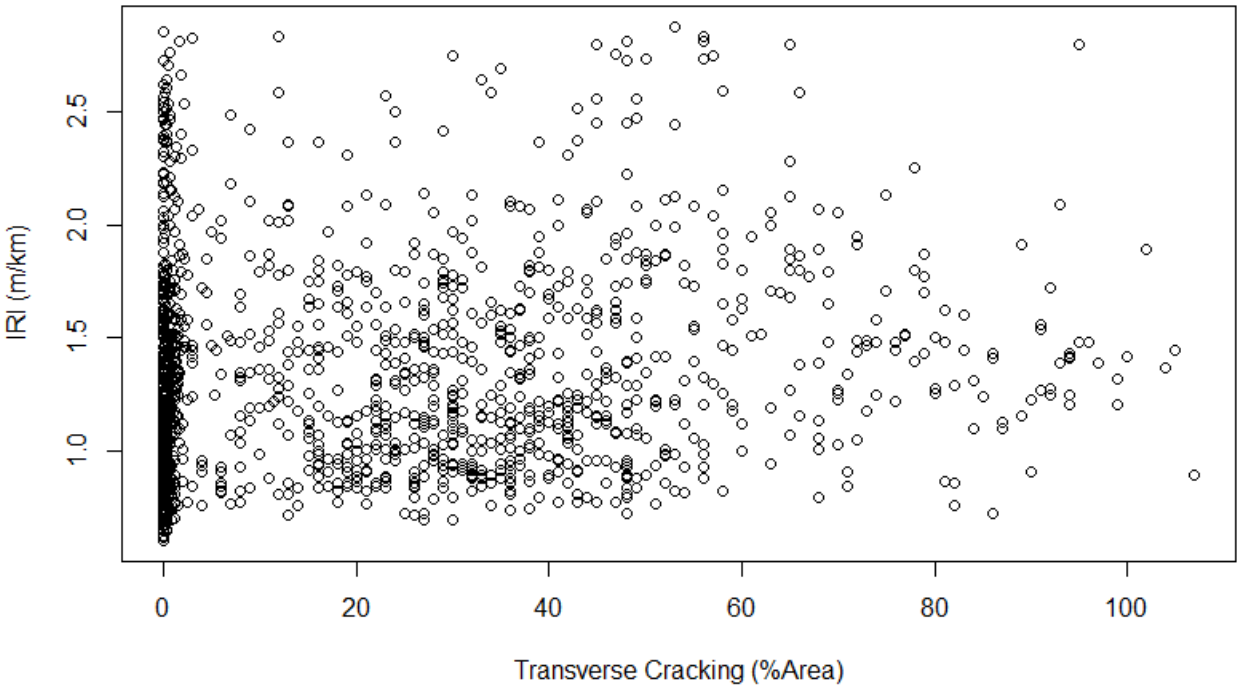

Figure A.108 – *TrcAr* values scatter plot for OL pavements – Outer-fence Dataset

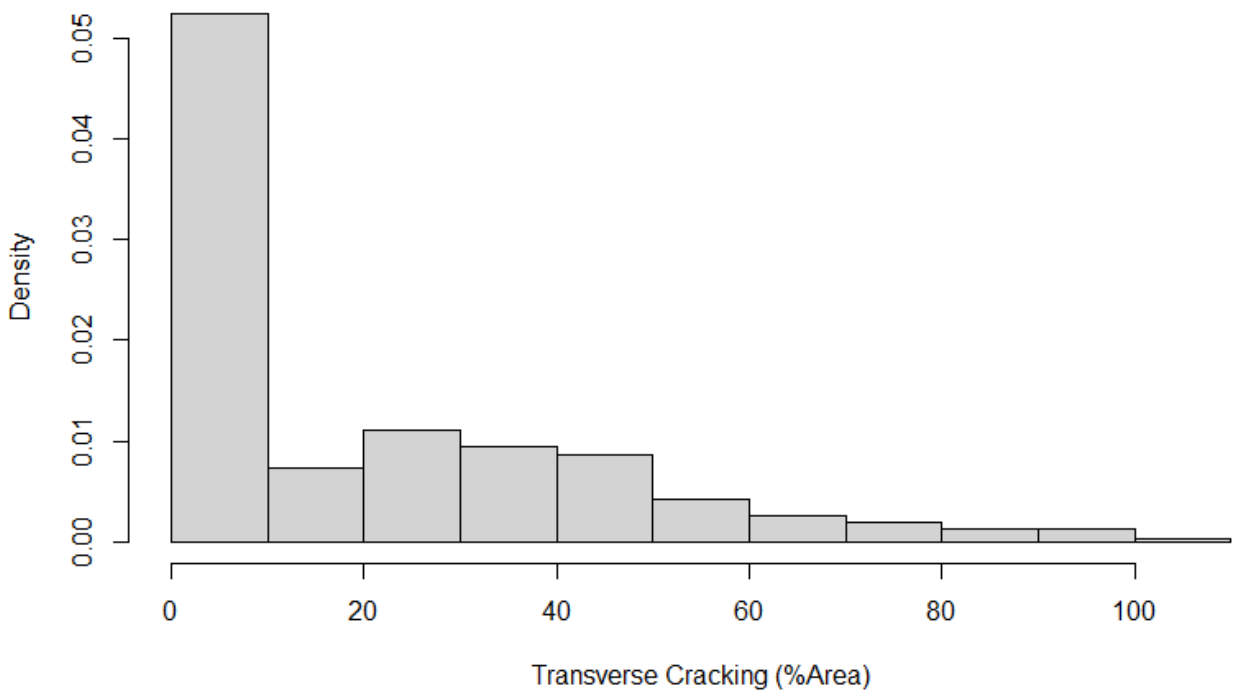

Figure A.109 – *TrcAr* values histogram plot for OL pavements – Outer-fence Dataset

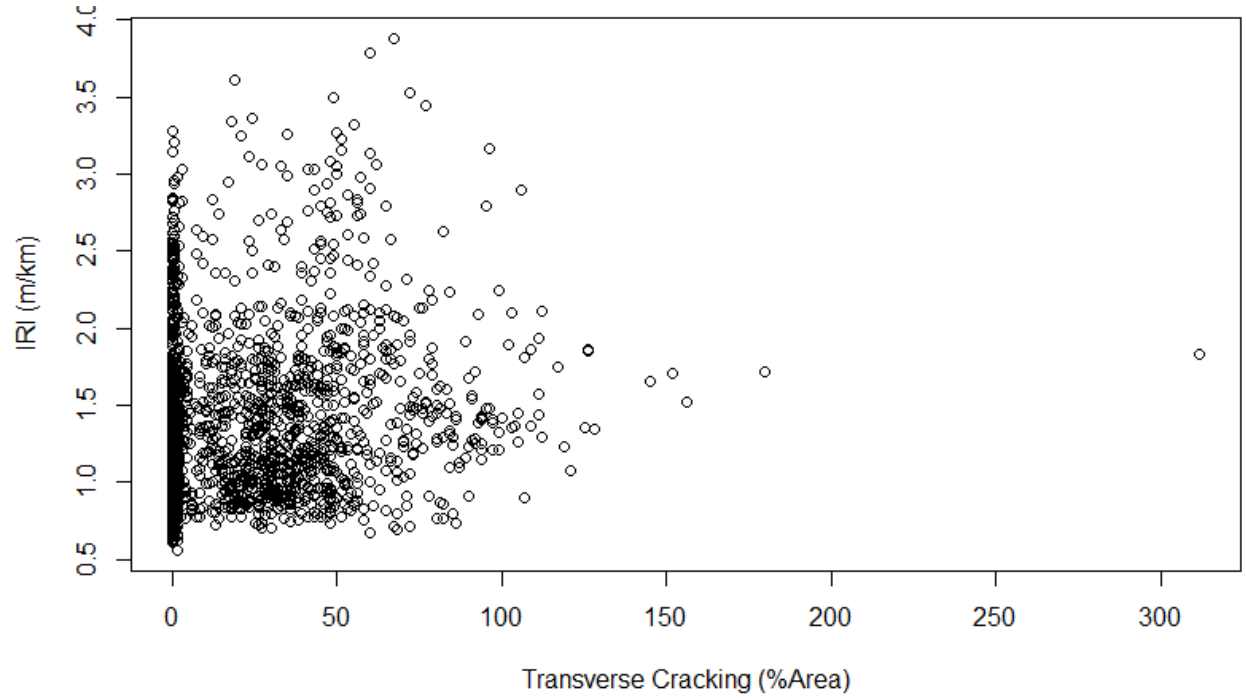

Figure A.110 – *TrcAr* values scatter plot for OL pavements – Full Dataset

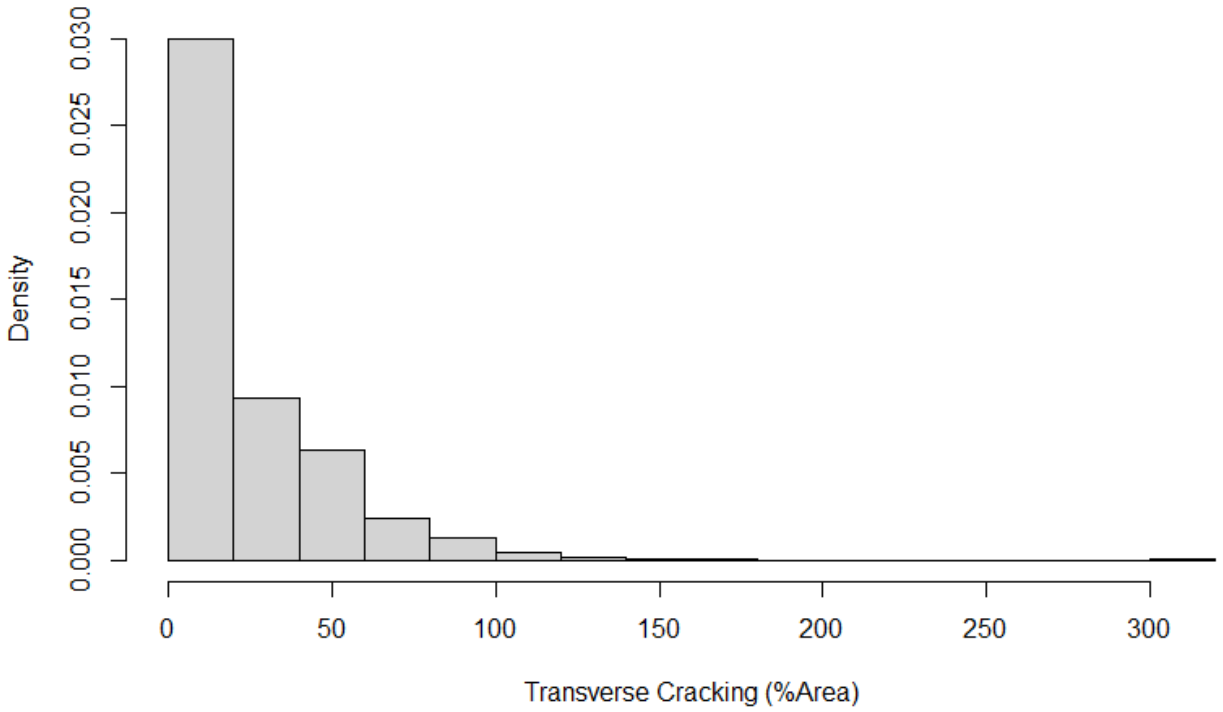

Figure A.111 – *TrcAr* values histogram plot for OL pavements – Full Dataset

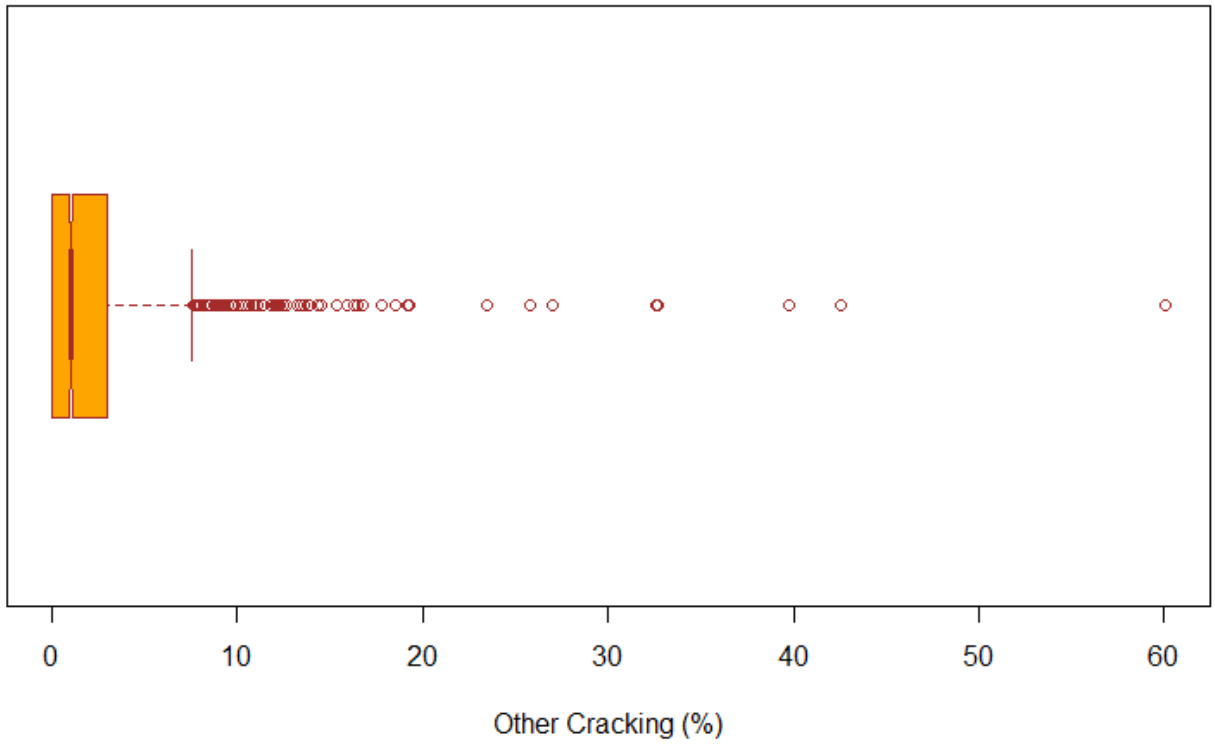

Figure A.112 – *OtherCAr* values Boxplot for OL pavements – Full Dataset

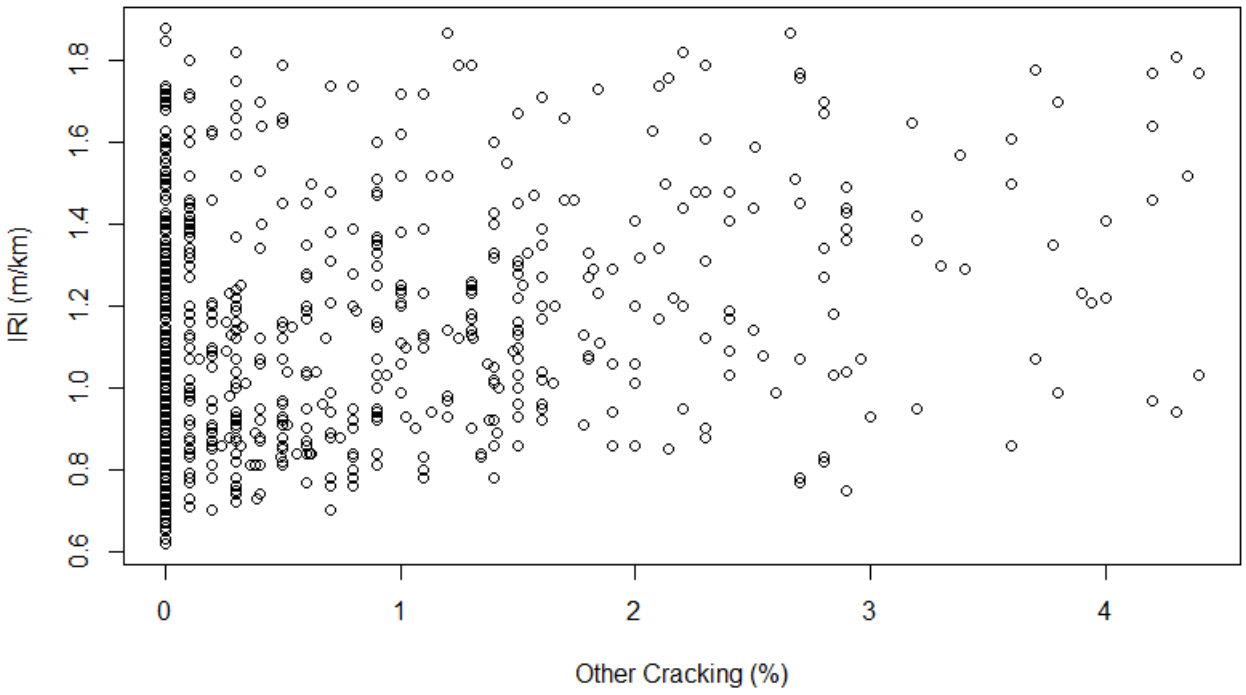

Figure A.113 – *OtherCAr* values scatter plot for OL pavements – inner-fence Dataset

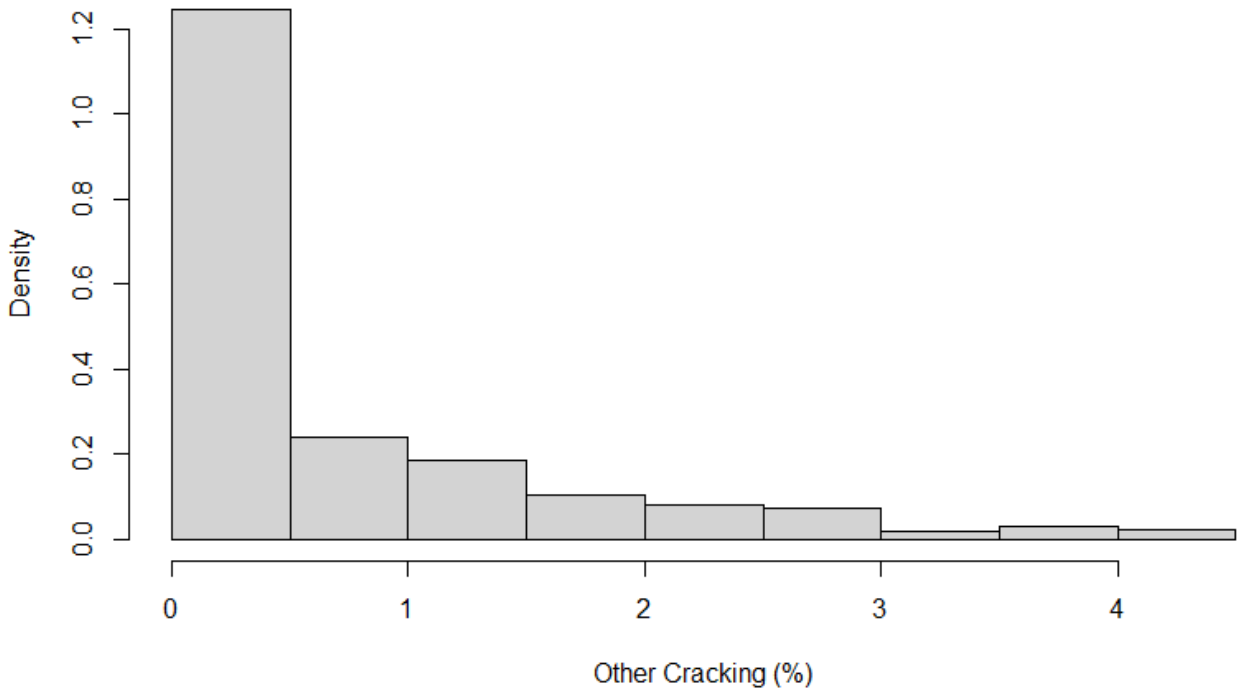

Figure A.114 – *OtherCAr* values histogram plot for OL pavements – inner-fence Dataset

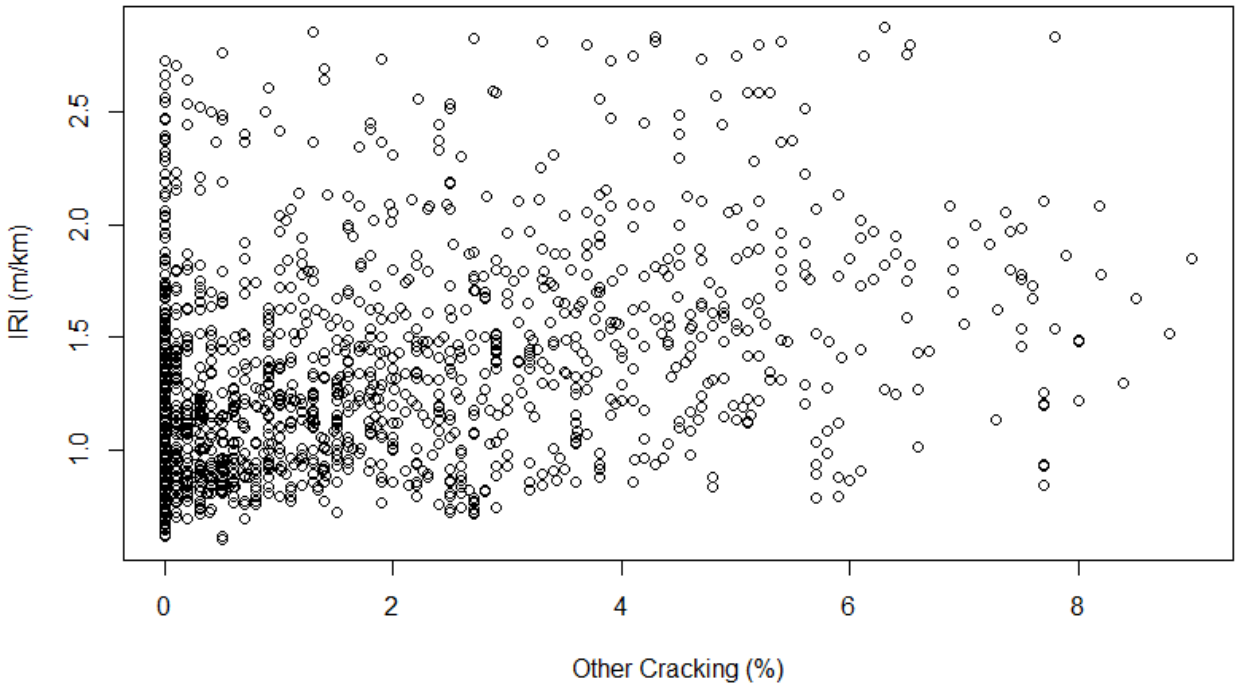

Figure A.115 – *OtherCAr* values scatter plot for OL pavements – Outer-fence Dataset

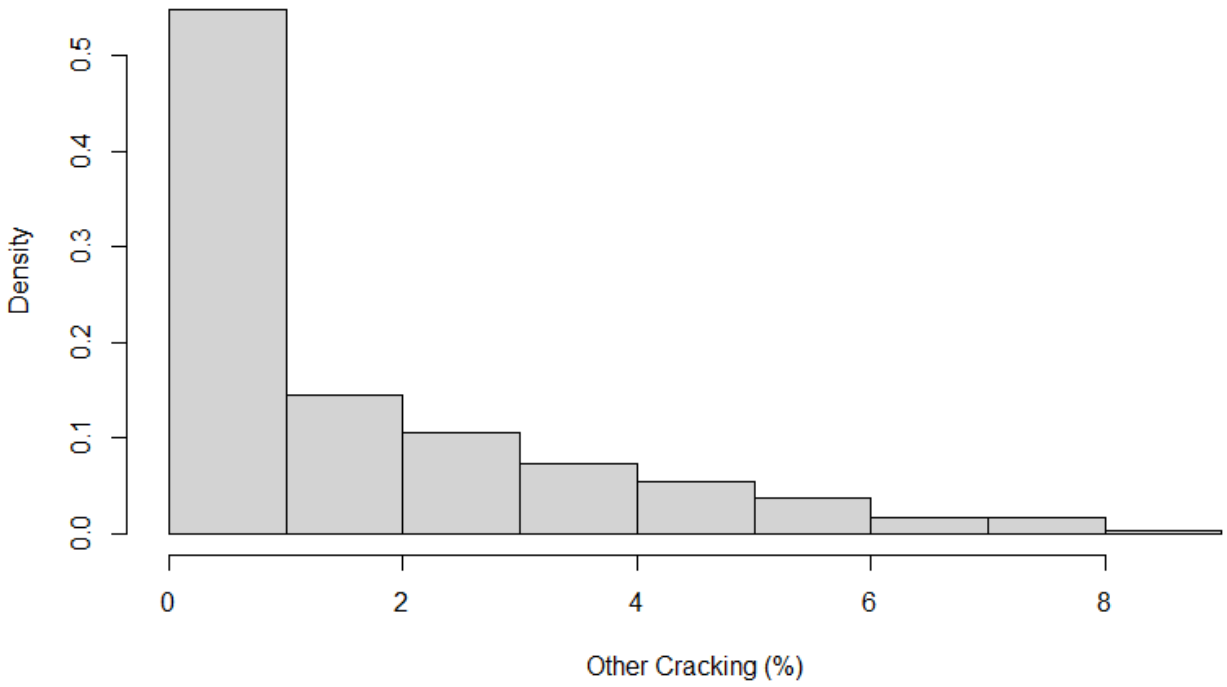

Figure A.116 – *OtherCAr* values histogram plot for OL pavements – Outer-fence Dataset

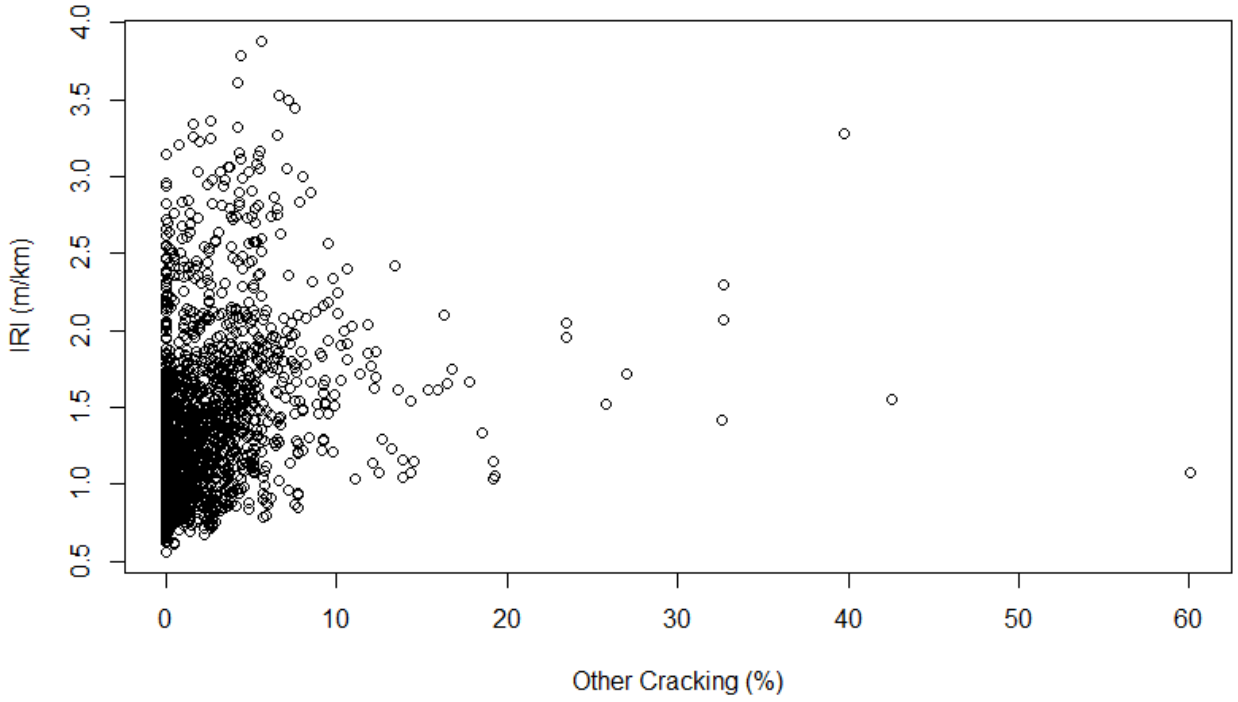

Figure A.117 – *OtherCAr* values scatter plot for OL pavements – Full Dataset

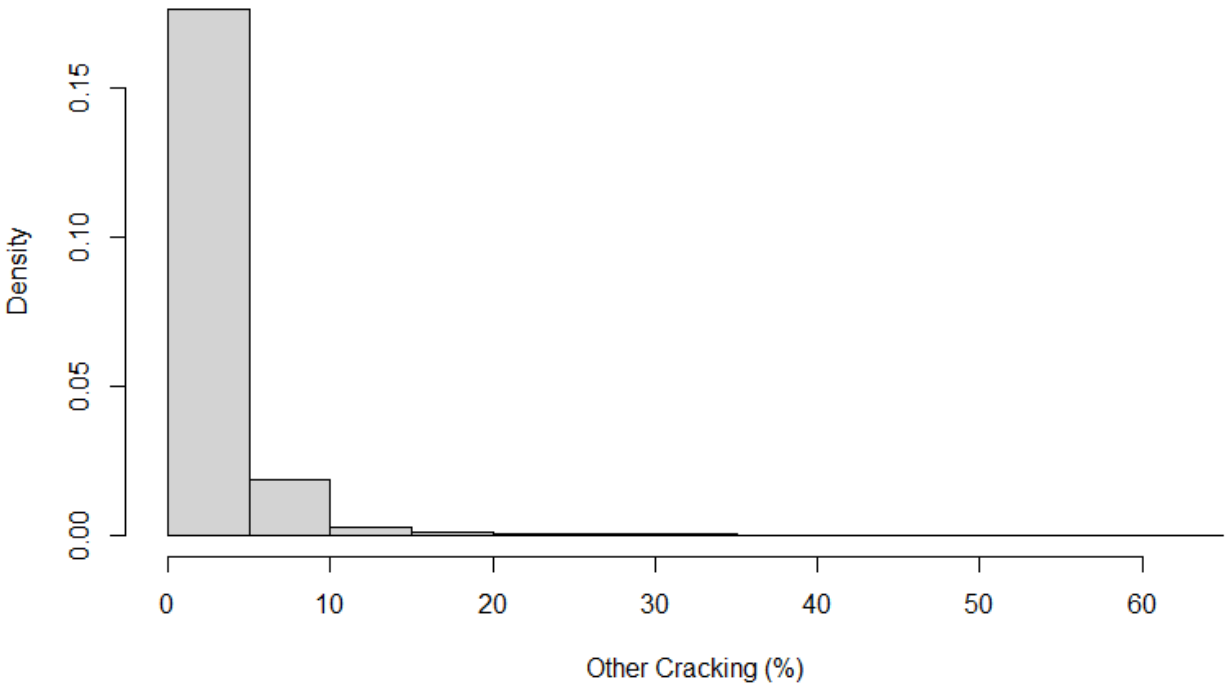

Figure A.118 – *OtherCAr* values histogram plot for OL pavements – Full Dataset

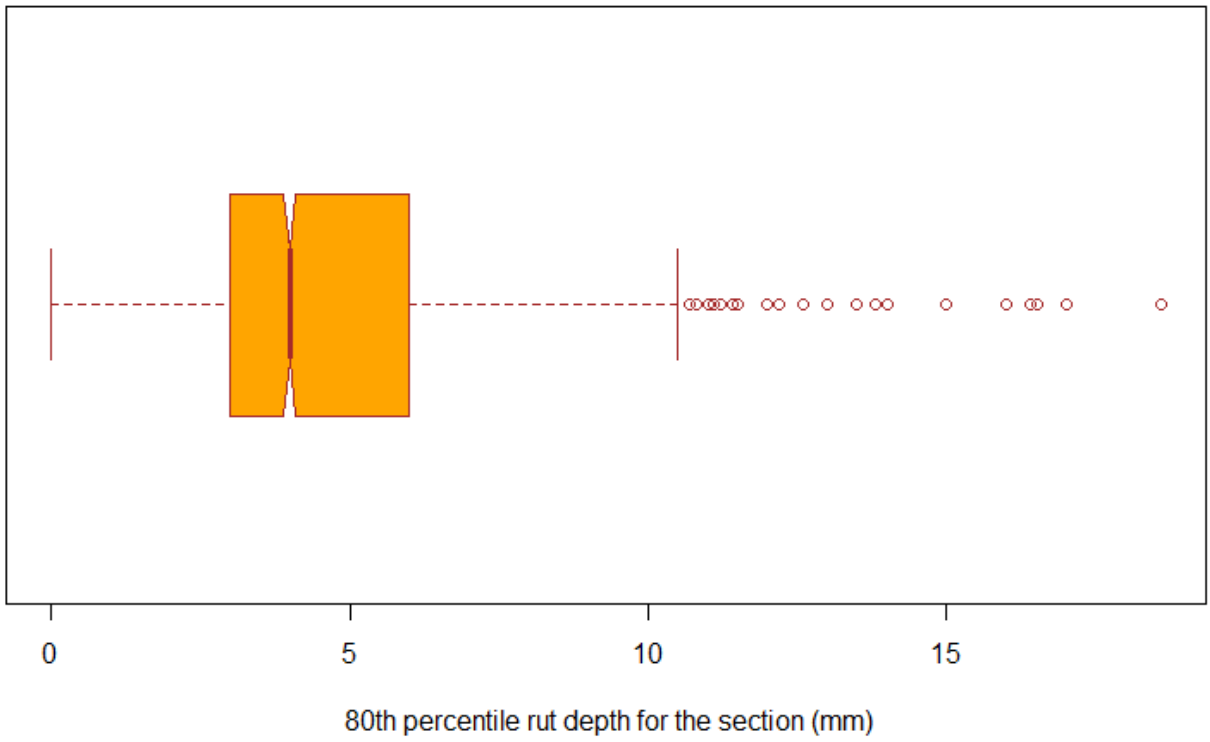

Figure A.119 – *RUT* values Boxplot for OL pavements – Full Dataset

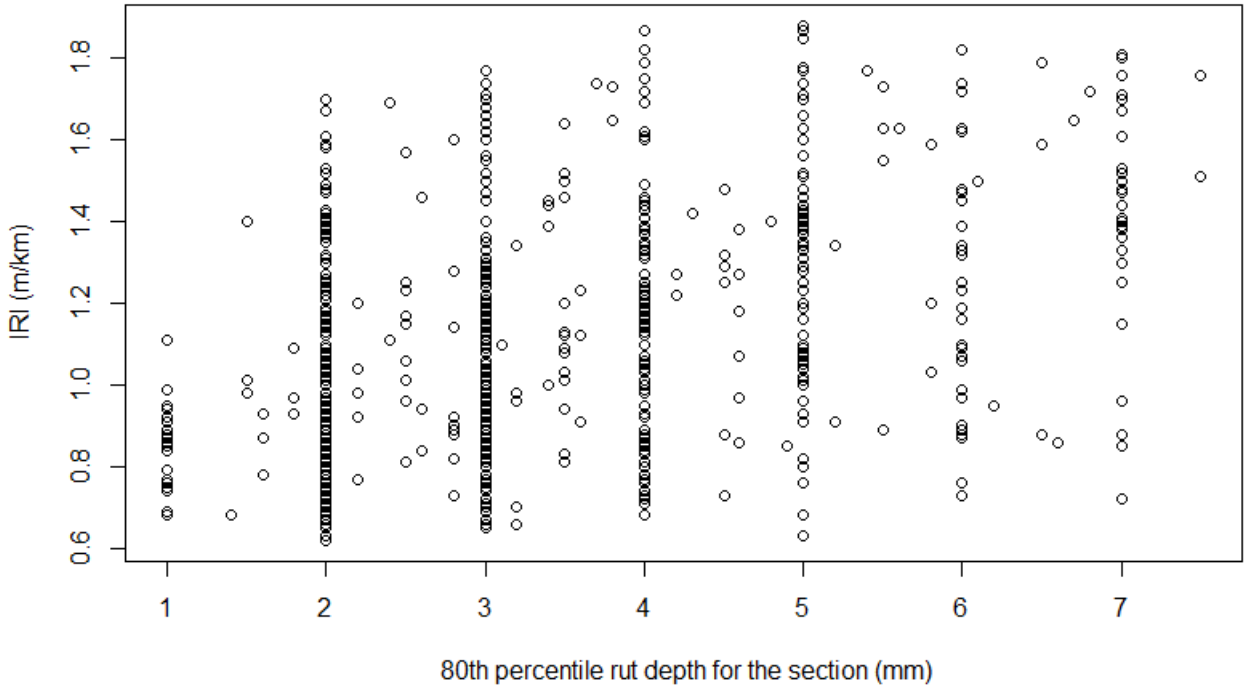

Figure A.120 – *RUT* values scatter plot for OL pavements – inner-fence Dataset

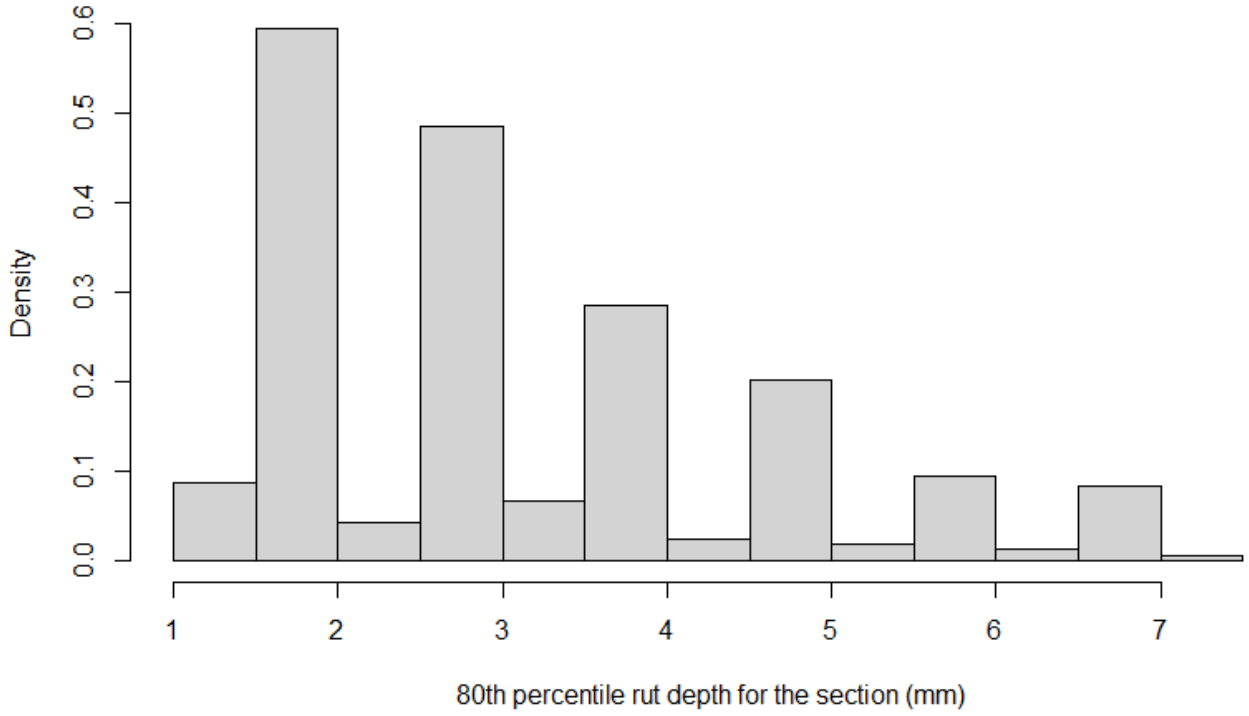

Figure A.121 – *RUT* values histogram plot for OL pavements – inner-fence Dataset

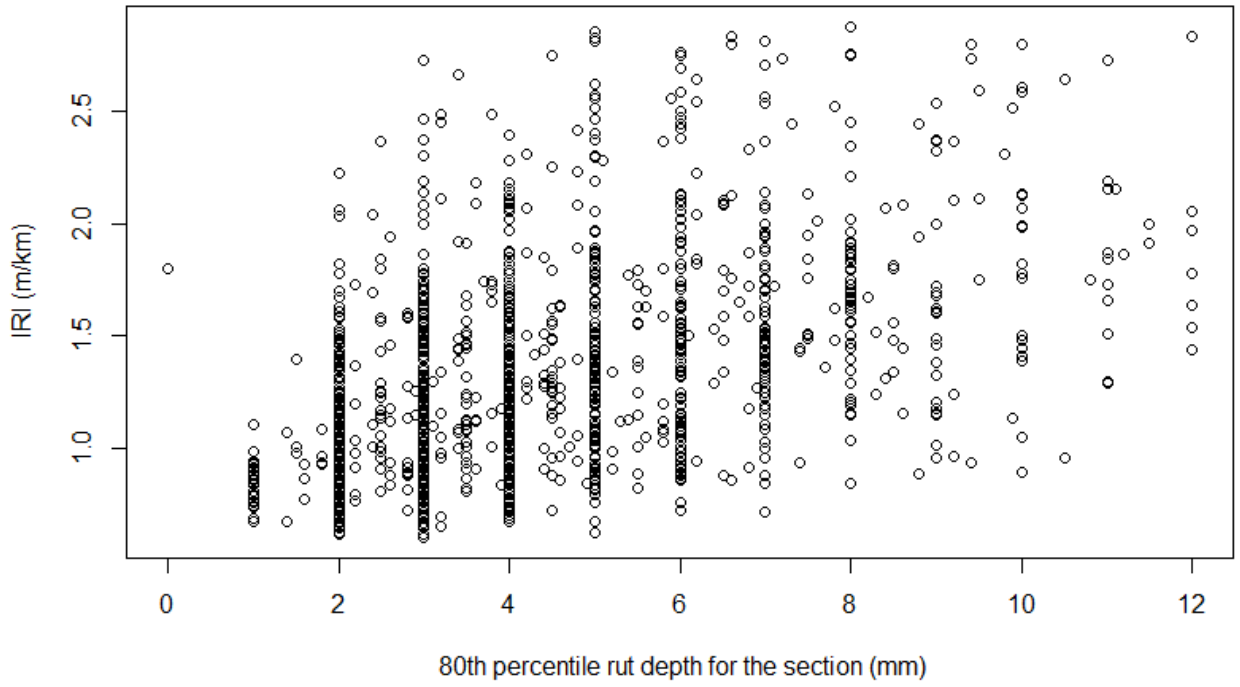

Figure A.122 – *RUT* values scatter plot for OL pavements – Outer-fence Dataset

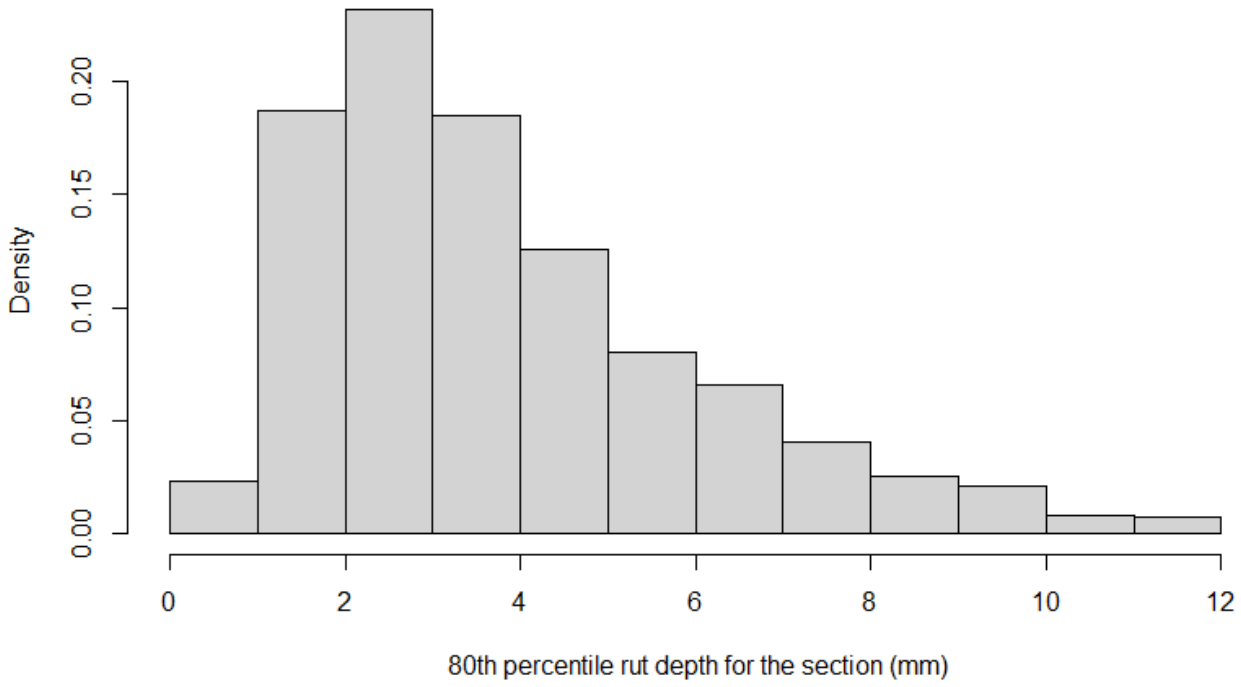

Figure A.123 – *RUT* values histogram plot for OL pavements – Outer-fence Dataset

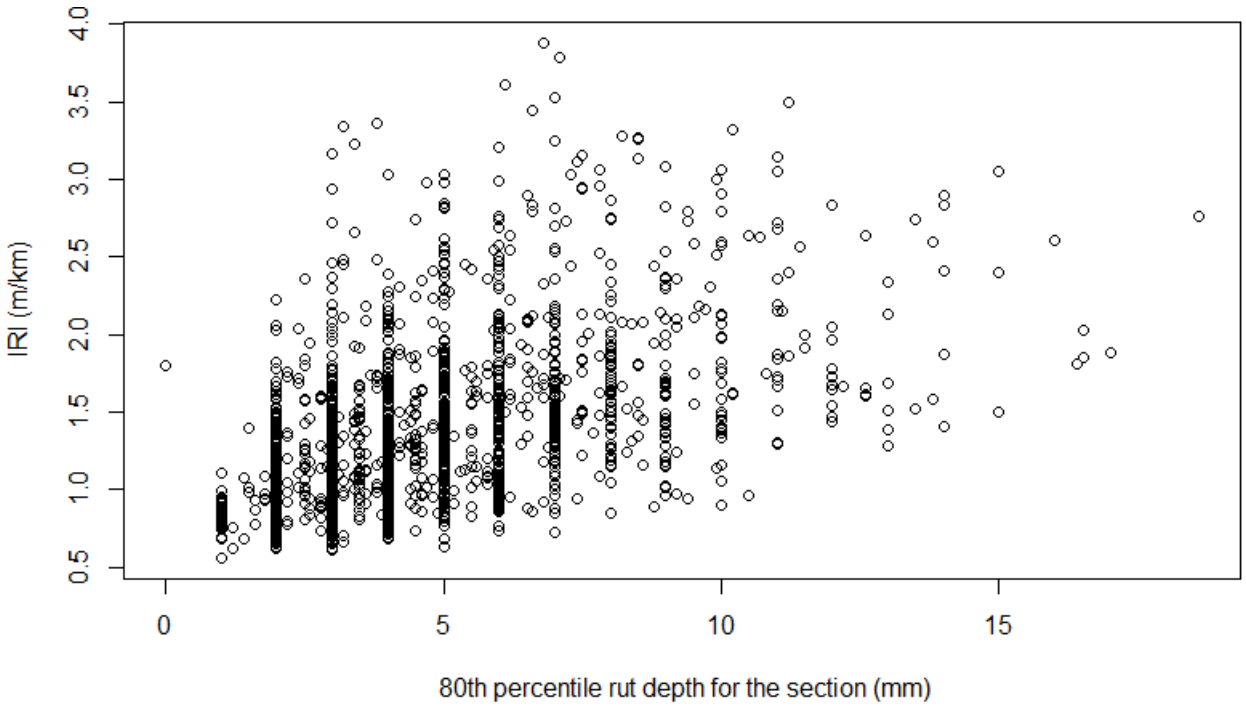

Figure A.124 – *RUT* values scatter plot for OL pavements – Full Dataset

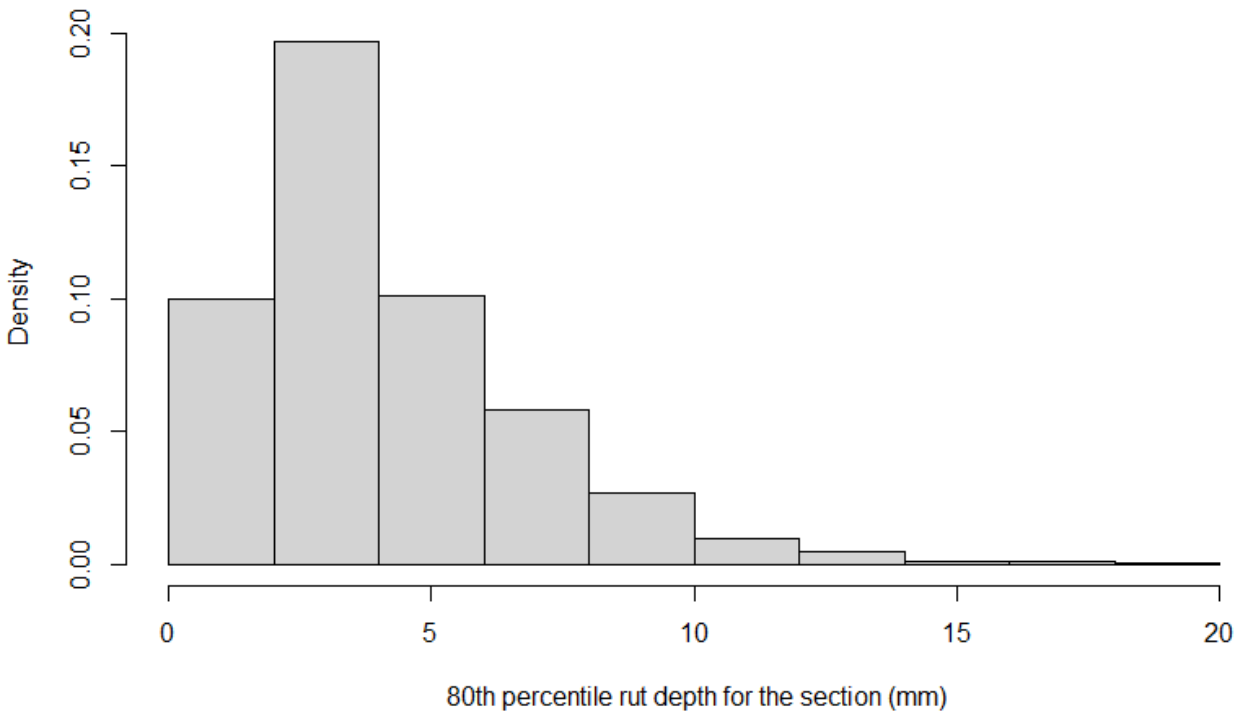

Figure A.125 – *RUT* values histogram plot for OL pavements – Full Dataset

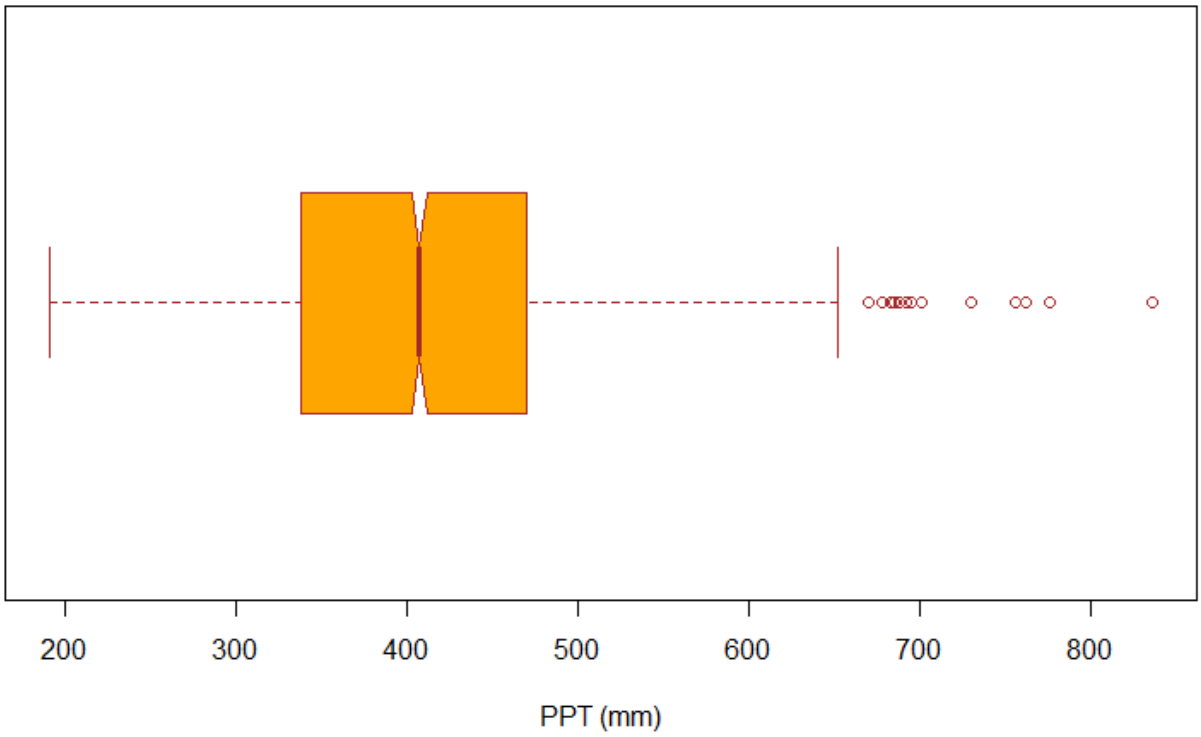

Figure A.126 – *PPT* values Boxplot for OL pavements – Full Dataset

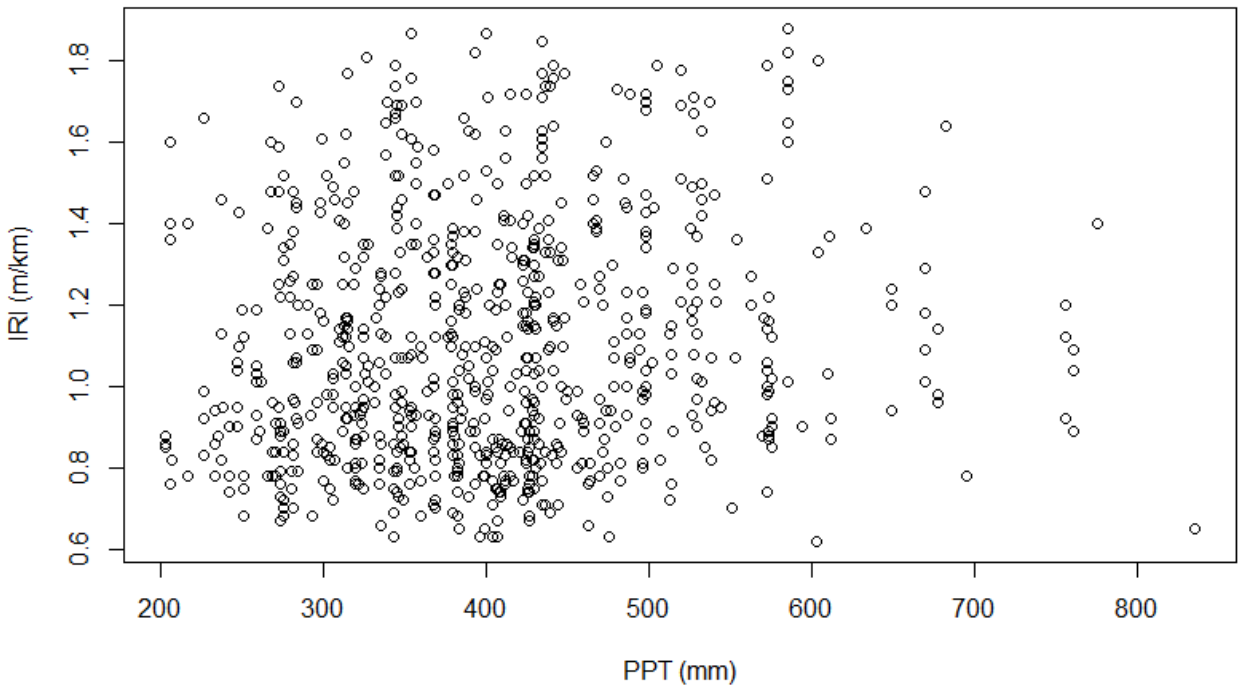

Figure A.127 – *PPT* values scatter plot for OL pavements – inner-fence Dataset

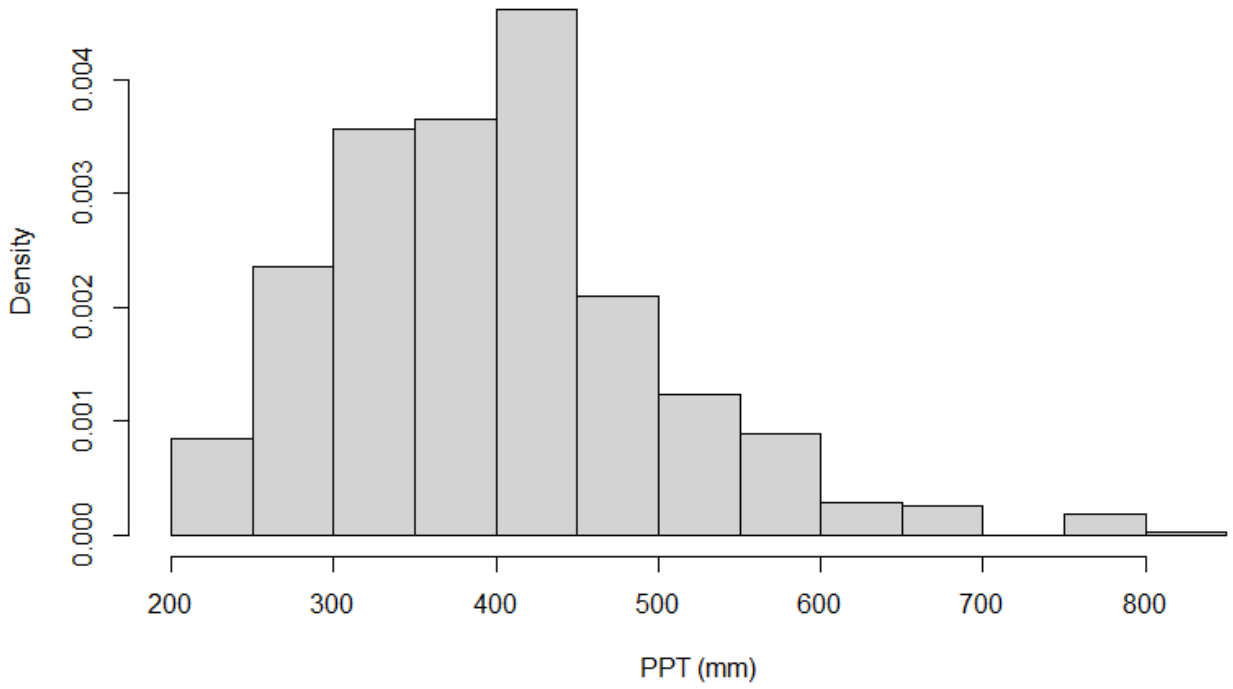

Figure A.128 – *PPT* values histogram plot for OL pavements – inner-fence Dataset

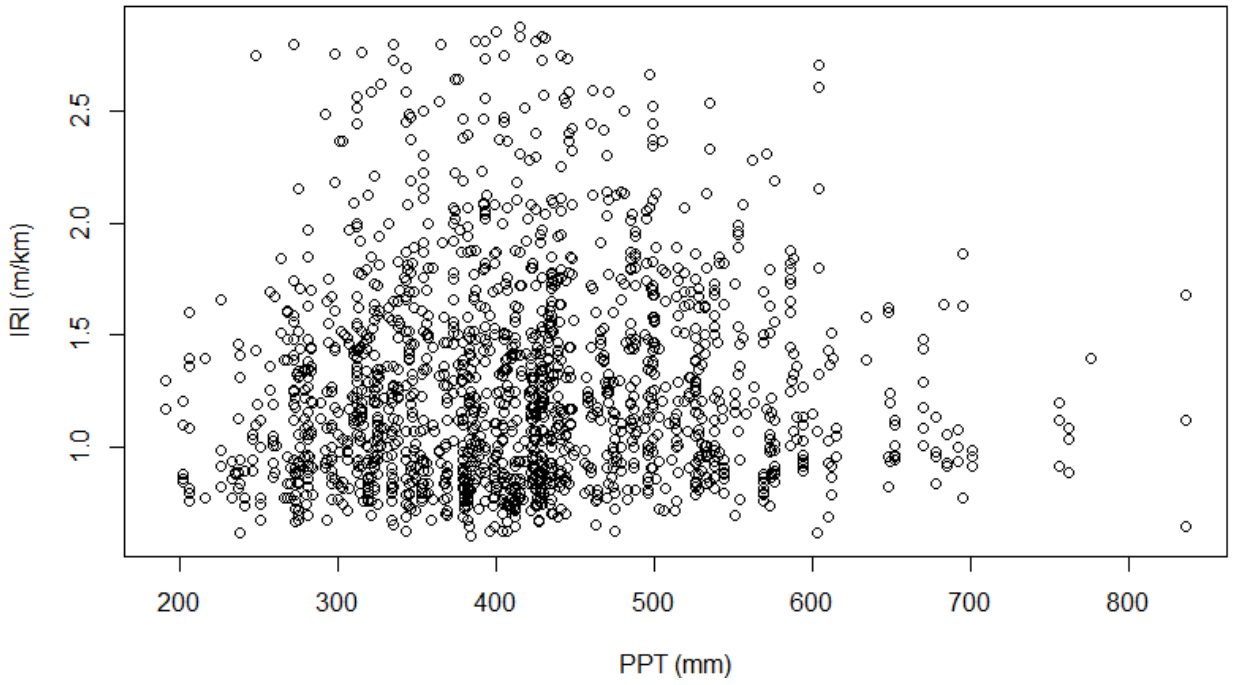

Figure A.129 – *PPT* values scatter plot for OL pavements – Outer-fence Dataset

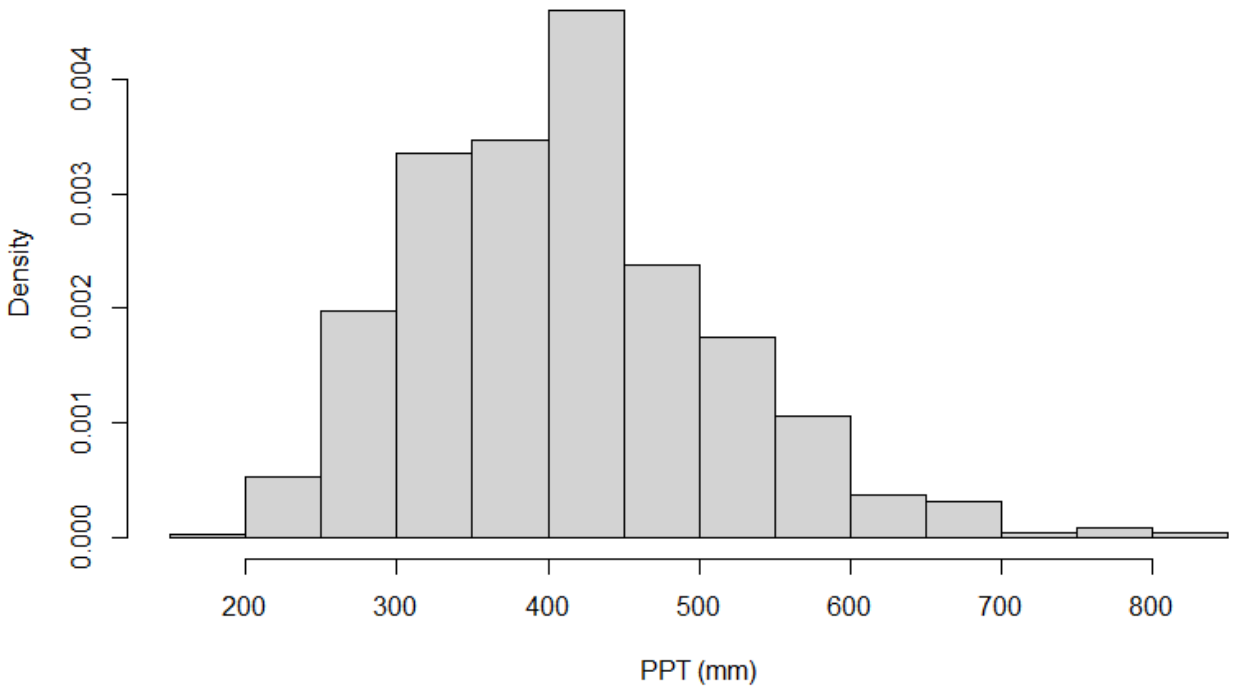

Figure A.130 – *PPT* values histogram plot for OL pavements – Outer-fence Dataset

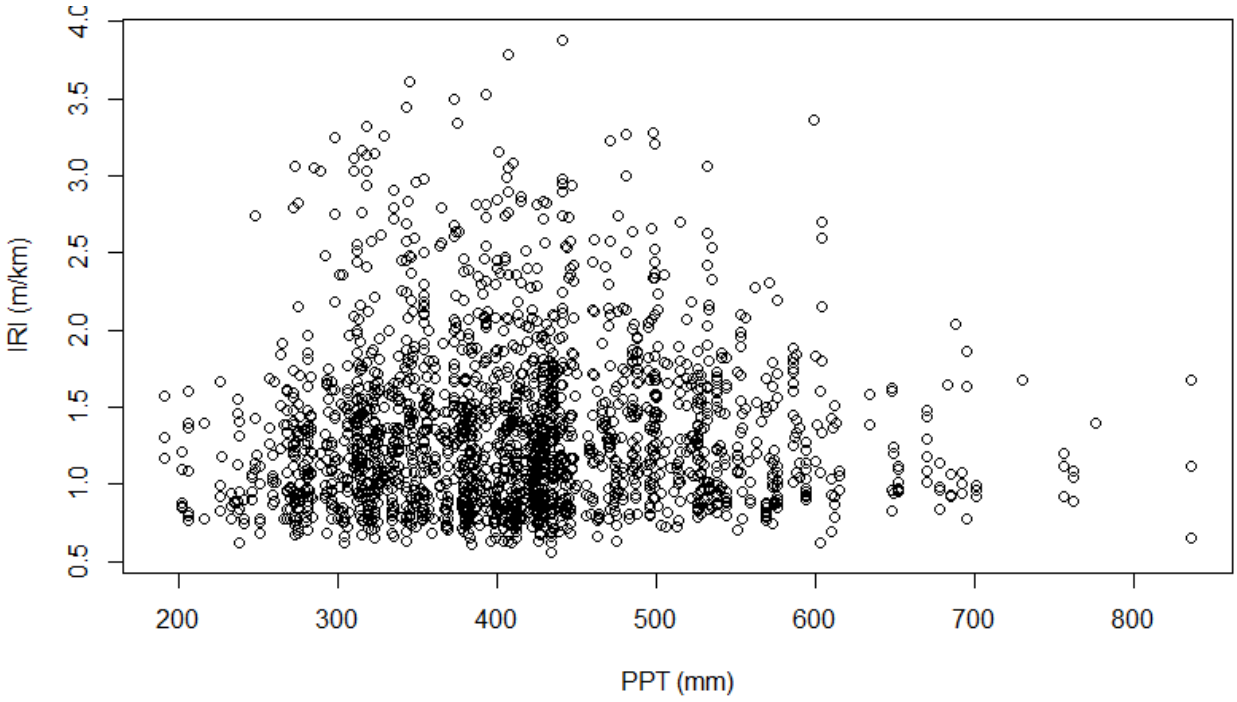

Figure A.131 – *PPT* values scatter plot for OL pavements – Full Dataset

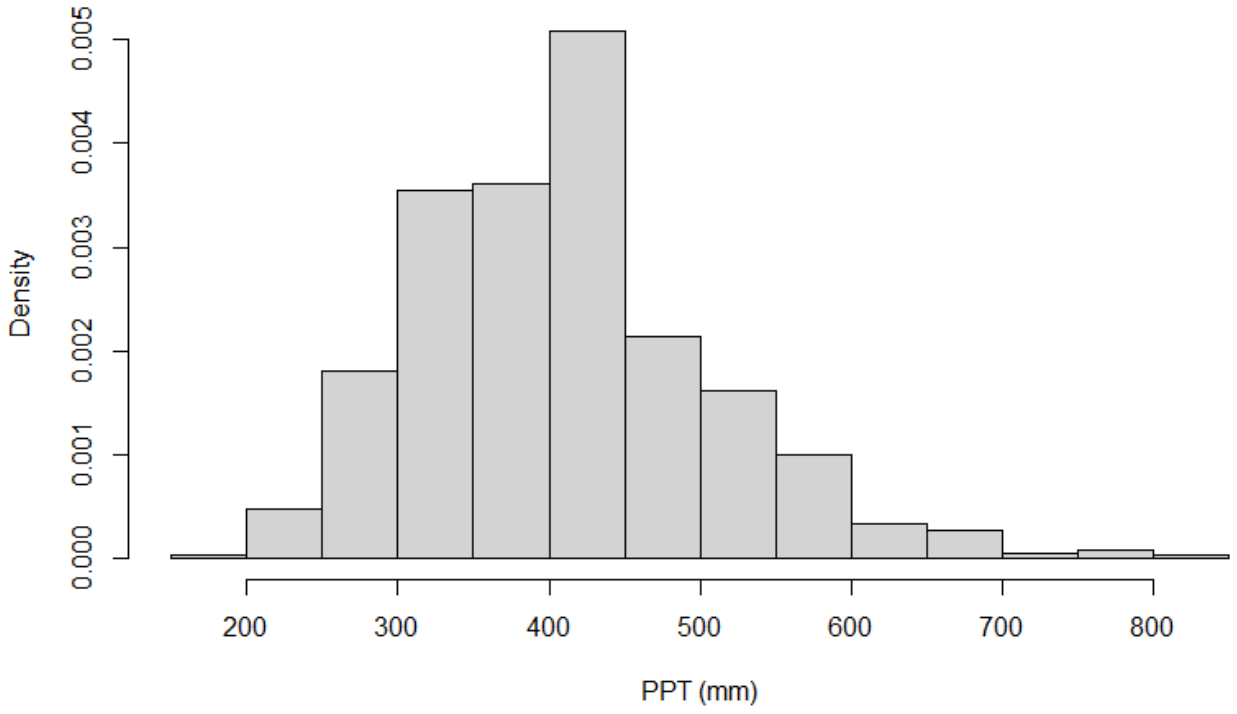

Figure A.132 – *PPT* values histogram plot for OL pavements – Full Dataset

Appendix B: Life Cycle Analysis (LCA) and Life Cycle Cost Analysis (LCCA)

The LCA and LCCA performed in this study were conducted through Athena Pavement LCA software by the Athena Sustainable Materials Institute (Athena Sustainable Materials Institute 2018). The analyses are performed in accordance with the North American standard practices as well as ISO 21930 and 21931 (Athena Sustainable Materials Institute 2018). The LCA mainly examines the site preparation, construction, and maintenance emissions as well as the emissions caused as an effect of pavement roughness. Among the analysis`s output are the global warming potential, smog potential, and total energy consumption estimates. The estimates consider all fossil fuel energy consumption in material manufacturing and related transportation for pavement construction, maintenance, and rehabilitation. The estimates also include the "inherited" energy consumption and emissions values of materials. The LCA also takes into consideration the gross combustion heat of any material which may be considered as energy but is not being used as a source of energy, for example, the use of bitumen asphalt. The total energy consumption estimate also considers fuel pre-combustion energy, the fuel conversion efficiency in generating electricity, and transmission line losses associated with the distribution of electricity (Athena Sustainable Materials Institute 2006).

The greenhouse gas emissions considered in the LCA were primarily carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), as they comprise 98.9% of the greenhouse gas inventory recognized by Enlivenment Canada (Matin et al. 2004). Out of the three gasses, carbon dioxide is commonly the reference standard for global warming and greenhouse gas effects. Thus, the global warming potential is quantified within the LCA using CO_2. The International Panel on Climate Change (IPCC) provides a "CO_2 equivalence effect" for other greenhouse gases and can be summed up under the Global Warming Potential Index (GWPI) as presented below (IPCC 2002).

$$GWPI\ (kg) = CO_2 kg + (CH_4\ kg * 23) + (N_2O\ kg * 296) \quad \text{(B.1)}$$

The greenhouse gas emissions in the LCA are presented through the global warming potential index. The global warming emissions due to site preparation, construction, maintenance, rehabilitation, and pavement roughness are all considered through the LCA. The description of the pavements used in the case study is included in the chapters in the text. The greenhouse emissions values for specific activities were internally calculated by Athena Pavement LCA.

As previously mentioned, the LCCA is also conducted using the Athena Pavement LCA software. The LCCA is an engineering technique that is based on established principles of economics that aid with analyzing the difference in long-term efficiency between alternative options in an analytical and fact-based manner (Moges et al. 2017). A common tool used in the LCCA to relatively quantify the value of money across different stages of a project with respect to time is the NPV concept (Moges et al. 2017). The NPV attempts to provide a present value for an entire project based on a discount rate (Moges et al. 2017). The discount rate used in the LCCA varies; in this study, a 4% discount rate is used throughout the LCCA (Moges et al. 2017). The LCCA in this study provides the total expected cost through the following equation:

$$TEC = RMC + CC - SV \tag{B.2}$$

Where *TEC* is the total expected cost, *RMC* represents the routine maintenance cost, *CC* is the construction cost and, *SV* is the salvage value. The salvage value quantifies the remaining life of the project in terms of NPV. The salvage value can be calculated as follows:

$$SV = \frac{ULLOL\ x\ CCOLL}{TELOL} \tag{B.3}$$

Where *ULLOL* is the usable life in the last overlay in years, *TELOL* is the total expected life of the last overlay in years, and *CCOLL* is the construction cost of the last overlay in years. The conducted LCCA in this study utilizes unit rates from Alberta Transportation's unit price average report (Alberta Transportation 2021). The induvial rates fluctuate depending on the task and agency. The rates used in this study are presented in the table below.

Table B.01 Average unit rates used in this study's LCCA

Description	**Unit**	**Average unit**
Supply of Aggregate - With Option	tonne	$1.27
Supply of Aggregate - No Option	tonne	$0.82
Subgrade Excavation	cubic metre	$16.66
Granular Fill	tonne	$53.75
Preparing Subgrade Surface (First Layer)	square metre	$1.70
Preparing Subgrade Surface (Second Layer)	square metre	$1.51
Portland Cement	tonne	$269.33
Site Clearing and Grubbing	square metre	$8.00
Geotextile	square metre	$5.84
Subsurface Drainage System	metre	$773.91
Filter Material	cubic metre	$107.50
Excavation - Channel	cubic metre	$9.66
Backfill - Non-Granular	cubic metre	$30.94
Concrete Slope Protection	square metre	$392.00
Surface Removal	square metre	$18.21
Partial Depth Repair	square metre	$1,034.08
Concrete - Class C	cubic metre	$1,514.84
Concrete - Class HPC	cubic metre	$2,380.50
Asphalt Concrete Pavement	tonne	$129.48
Borrow Excavation	cubic metre	$8.95
Borrow Excavation - Contractor Supplied	cubic metre	$21.41
Common and/or Borrow Excavation Loaded to Trucks	cubic metre	$12.98
Truck Haul of Common and/or Borrow Excavation	cubic metre	$2.17
Crack Repair - Spray Patch	metre	$5.37
Crack Repair - Shallow Mill and Fill	metre	$46.48
Gravel Surfacing	cubic metre	$50.74
Gravel Surfacing - Des. 4 Cl. 25	tonne	$202.52
Gravel Surfacing	tonne	$20.77
Asphalt Mix for Others	tonne	$64.20
Cold Milling Asphalt Pavement	tonne	$16.11
Cold Milling Asphalt Pavement	square metre	$2.64
Supply and Application of Fog Coat	square metre	$0.96
Chip Seal Coat (CRS-2P)	square metre	$10.24

Chip Seal Coat - Bridge Decks	square metre	$16.60
Asphalt Concrete Pavement - EPS Mix Type H1	tonne	$98.92
Asphalt Concrete Pavement - EPS Mix Type H2	tonne	$53.31
Asphalt Concrete Pavement - EPS Mix Type M1	tonne	$65.19
Asphalt Concrete Pavement - EPS Mix Type L1	tonne	$3,511.87
Asphalt Concrete Pavement - EPS Mix Type S1	tonne	$69.23
Asphalt Concrete Pavement - EPS Mix Type S3	tonne	$61.21
Asphalt Concrete Pavement - Superpave	tonne	$90.34
Concrete Base - Supply and Install	bases	$936.63
Concrete Base - Remove and Reinstall	bases	$255.18
Concrete Base - Remove and Dispose	bases	$287.72
Roadway Lines - Supplying Paint and Painting (Directional	kilometre	$876.97
Roadway Lines - Supplying Paint and Painting (Lane Dividing	kilometre	$856.80
Roadway Lines - Supplying Paint and Painting (Lane Dividing	kilometre	$252.96
Roadway Lines - Supplying Paint and Painting	kilometre	$18,271.99
Intersection Lines - Supplying Paint and Painting	intersection	$2,118.25
Interchange Lines - Supplying Paint and Painting	interchange	$1,198.53
Removal of Existing Painted Lines	metre	$2.51
Milled Rumble Strips for Stop Conditions	set	$2,464.33
Milled Rumble Strip - Shoulder	kilometre	$1,122.55
Milled Rumble Strip - Centreline	kilometre	$1,390.73
Thrie Beam Guardrail - Supply and Install	metre	$227.22
Impact Attenuator (NCHRP 350/MASH 2009 TL-3) - Supply and	unit	$3,039.53
Remove, Salvage and Reinstall Existing Guardrail	metre	$92.40
Supply of Guardrail Posts	post	$273.33
Remove and Dispose of Existing Guardrail	metre	$14.13
Flexible Guide Post/Delineator - Round - Supply and Install	post	$56.76
Trenching and Backfilling	metre	$17.45
Asphalt Surfacing - Remove and Dispose	cubic metre	$602.40
Concrete Curb and Gutter - Remove and Dispose	metre	$29.28
Concrete Surface - Remove and Dispose	square metre	$24.63
Solid Concrete Islands	square metre	$260.00
Solid Concrete Medians	square metre	$191.09
Granular Fill for Medians	tonne	$32.92
Median Asphalt Concrete Surfacing	tonne	$86.67

Median Concrete Surfacing	square metre	$76.68
Cutting of Pavement	metre	$11.74

Appendix C: Complex Machine Learning *IRI* models using Alberta PMS data

More complex ML *IRI* models than the ones presented in Chapter 4 are presented in this appendix. The *IRI* models were developed to predict *IRI* values from the pavement features: *PI*, P_{200}, *Age*, *ESAL*, pavement surface thickness, transverse cracking, rutting depth, and annual precipitation. The relevant statistics for the *IRI* models from each of the ML techniques based on the models' outputs are presented in Figure C.1. It is important to point out that some algorithms are more prone to overfitting than others (Bishop 2006). The *IRI* models developed using the decision-tree algorithm appear to provide the best fit for OL pavements, with an R^2 of 1.00 and an *RMSE* of 0.00 for the training dataset. The decision tree algorithm results indicate that the model is overfitted to the training dataset; thus, the testing dataset's results should be considered to indicate the models' predictive capabilities. The models are validated using the testing dataset to assess the model's accuracy for input values not used in the model development.

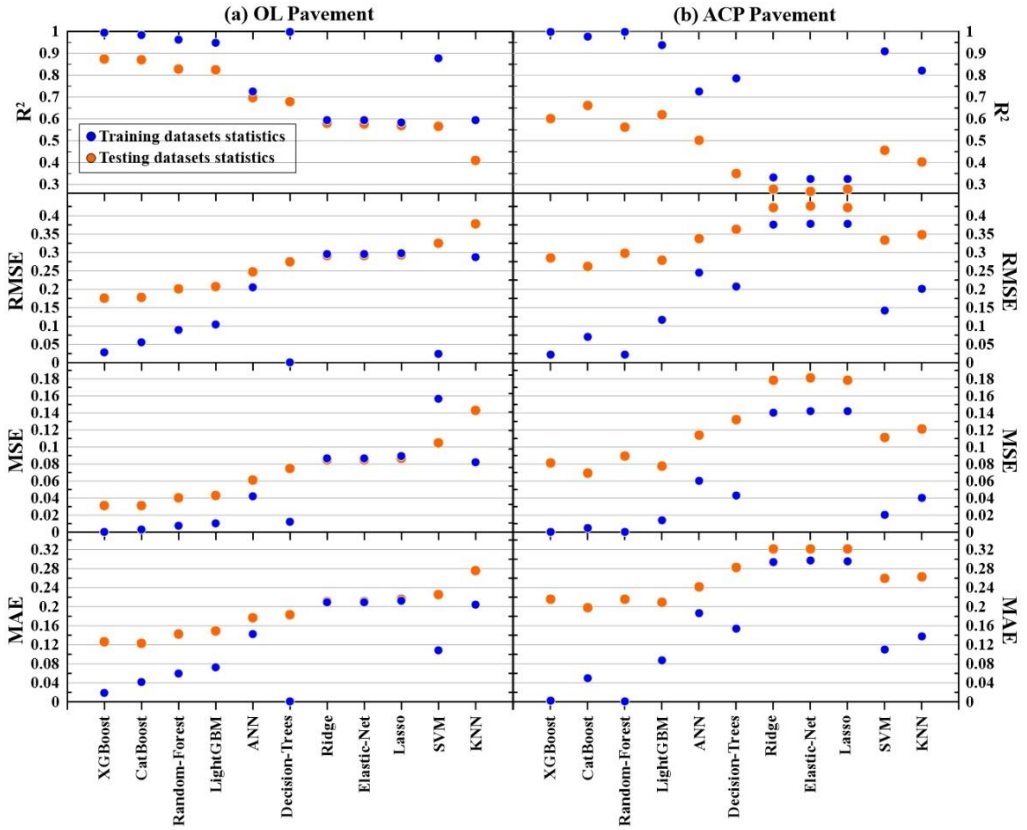

Figure C.1 – Relevant statistics for the *IRI* models from various machine learning algorithms. a) Overlay pavements b) ACP pavements.

When ranking each algorithm's performance, it is critical to consider the testing dataset's R^2 over the training dataset's R^2. Based on the testing dataset, the best-performing algorithms for both types of pavements are XGBoost, CatBoost, LightGBM, and Random-Forest. Based on these results, it is evident that ensemble ML algorithms perform the best for this study's datasets. The predicted and measured *IRI* values for the testing datasets for all the models are shown in Figures C.2 and C.3 for OL and ACP, respectively.

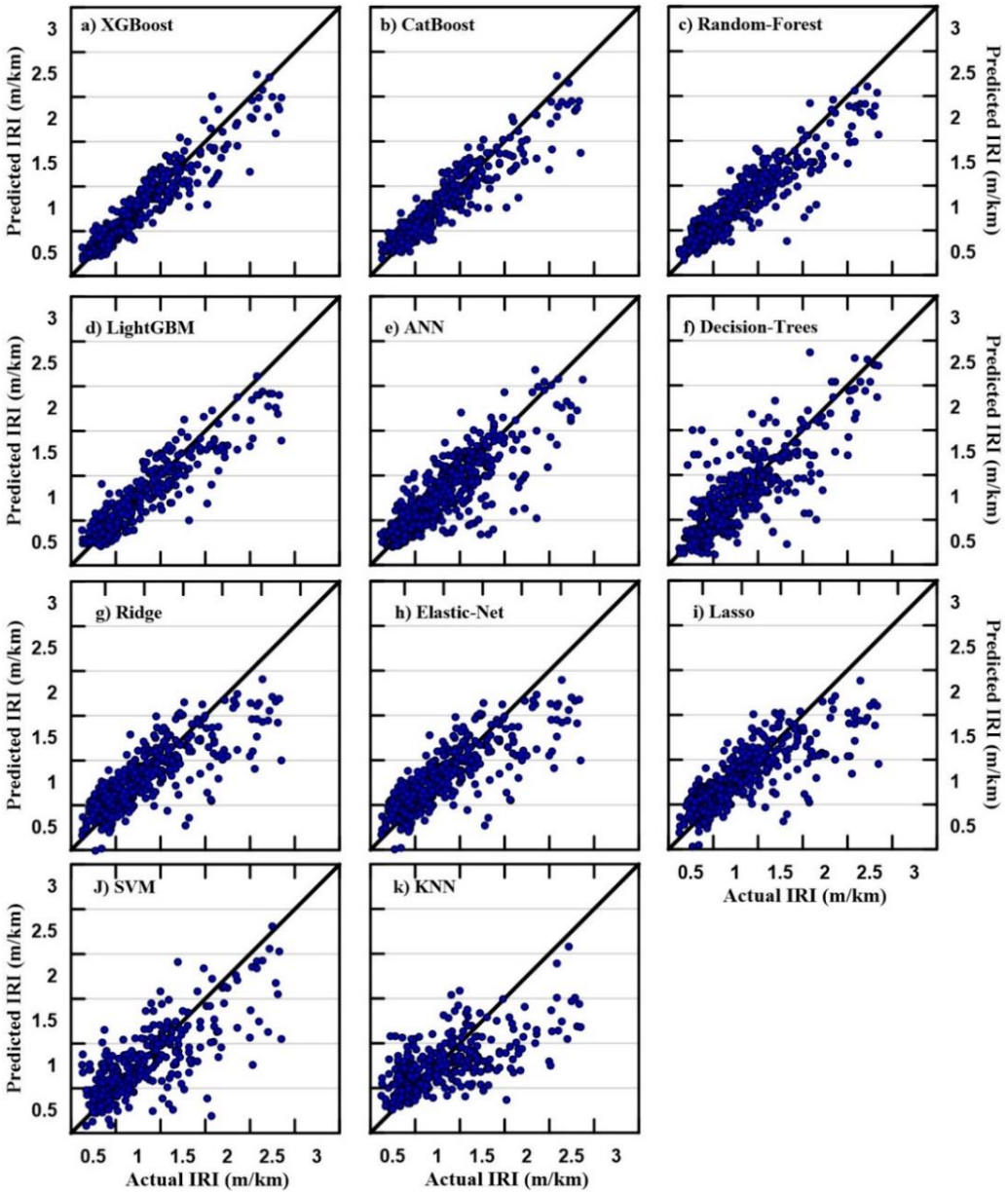

Figure C.2 - Actual vs Predicted *IRI* values for OL Models

From Figure C.2, XGBoost, Catboost, LightGBM, and Random-Forest are once again observed to be the best-performing models. Figure C.3 further confirms that the best-performing algorithms for the application of this study are Catboost, LightGBM, XGBoost, and Random-Forest.

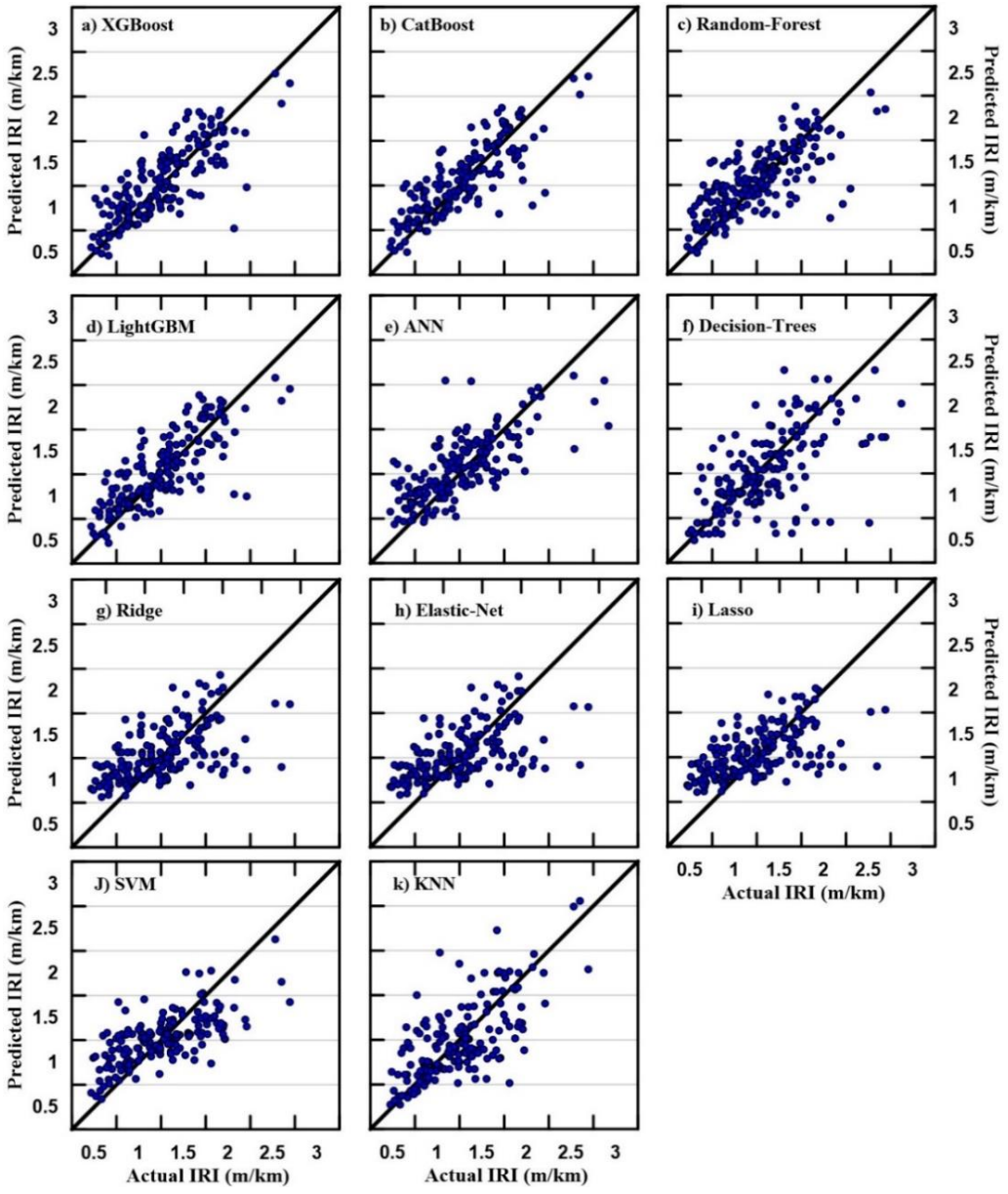

Figure C.3 - Actual vs Predicted *IRI* values for ACP models

The computational time for training each of the ML models is presented in Figure C.4. The figure illustrates the time taken to train each of the models and provides an insight into the complexity of the

model training for both the OL and ACP models. Aside from the statistical results, computational time helps identify the most effective algorithm with time in mind. For instance, the XGBoost and CatBoost algorithms produce similar model results for OL pavements; however, the computation time for training the CatBoost model is much greater than the computation time taken to train the XGBoost model. This finding illustrates that XGBoost is superior to CatBoost when considering resource use. Generally, Figure C.4 further supports that Ensemble ML algorithms are the best-performing ML algorithms, even while considering computational time.

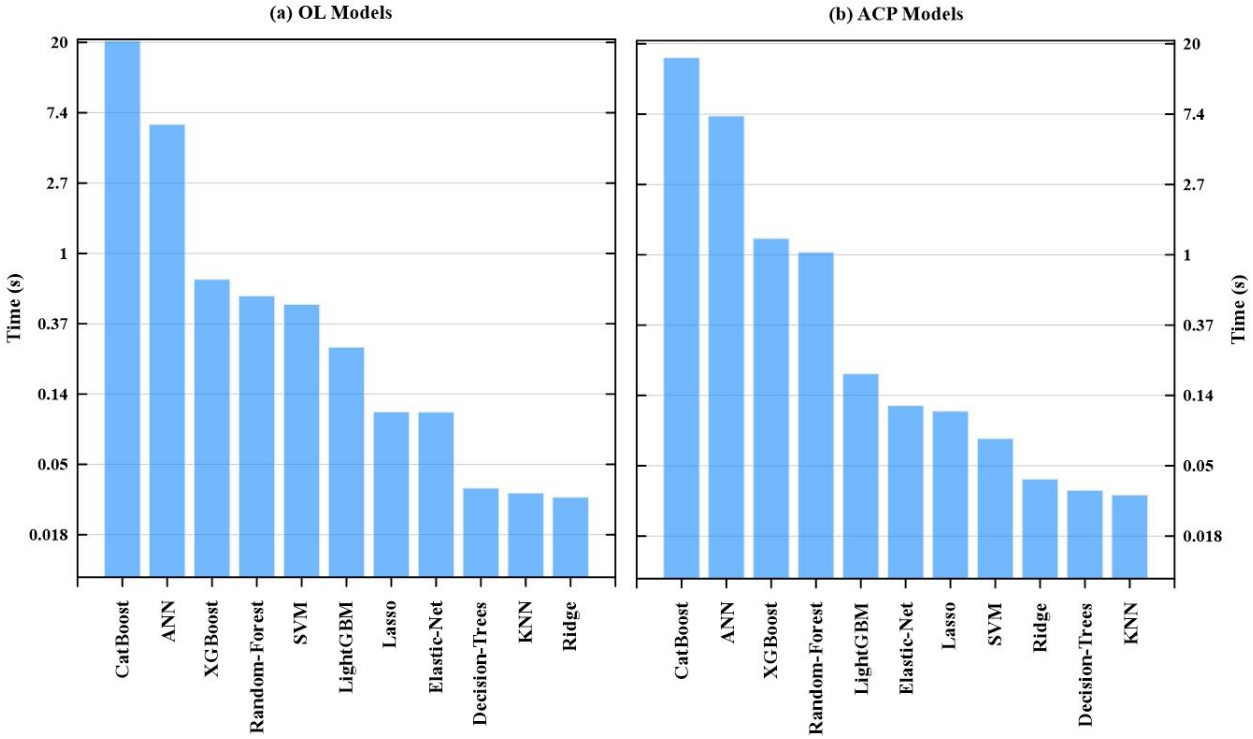

Figure C.4 – Model training computational time for ML *IRI* models (a) OL pavements (b) ACP pavements

It is important to understand the impact of each independent variable on the *IRI* values within the context of each model. For a visual representation of pavement features' impact, sensitivity analysis plots are created. Sensitivity plots help visualize the impact of each variable around the mean line (the 0-impact line). This analysis is carried out for the four previously mentioned best-performing algorithms. The first

analysis is presented in Figure C.5 using the developed XGBoost's *IRI* models for both OL and ACP pavements.

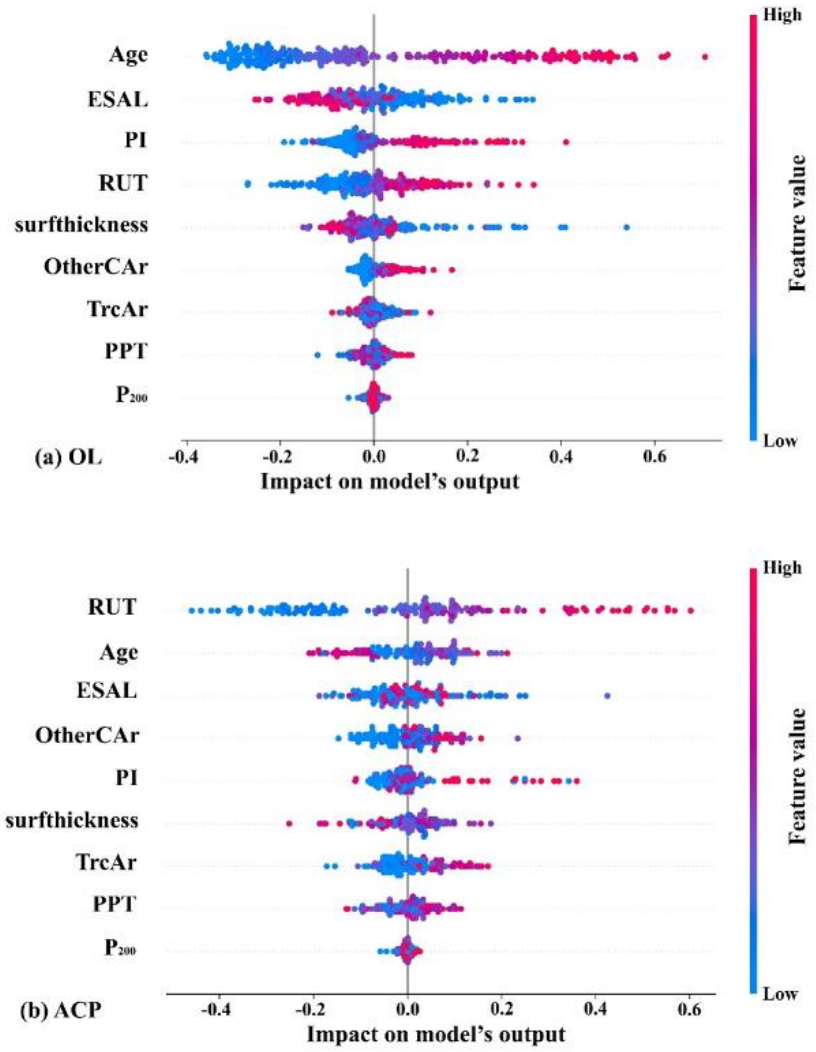

Figure C.5 - XGBoost ML *IRI* models sensitivity analysis. a) OL pavements b) ACP pavement

From Figure C.5, the most impactful pavement features for OL pavements differ from ACP pavements. The feature's variables were previously defined in the text, except for *OtherCA*r being the other cracking area and *surfthickness* being the surface thickness of the pavement. Generally, for the XGBoost models, pavement features: *Age*, *RUT*, and *ESAL* appear to have a higher effect on the output, regardless of the pavement type. Pavement features such as *PI* appear to have a lower impact on the ACP pavement's *IRI* model output compared to the OL pavement's *IRI* model output. Features with the lowest impact are

common between the two pavements and pose a smaller impact on the models' output as they are correlated with other pavement features in the model. A similar analysis is conducted for CatBoost and is presented in Figure C.6.

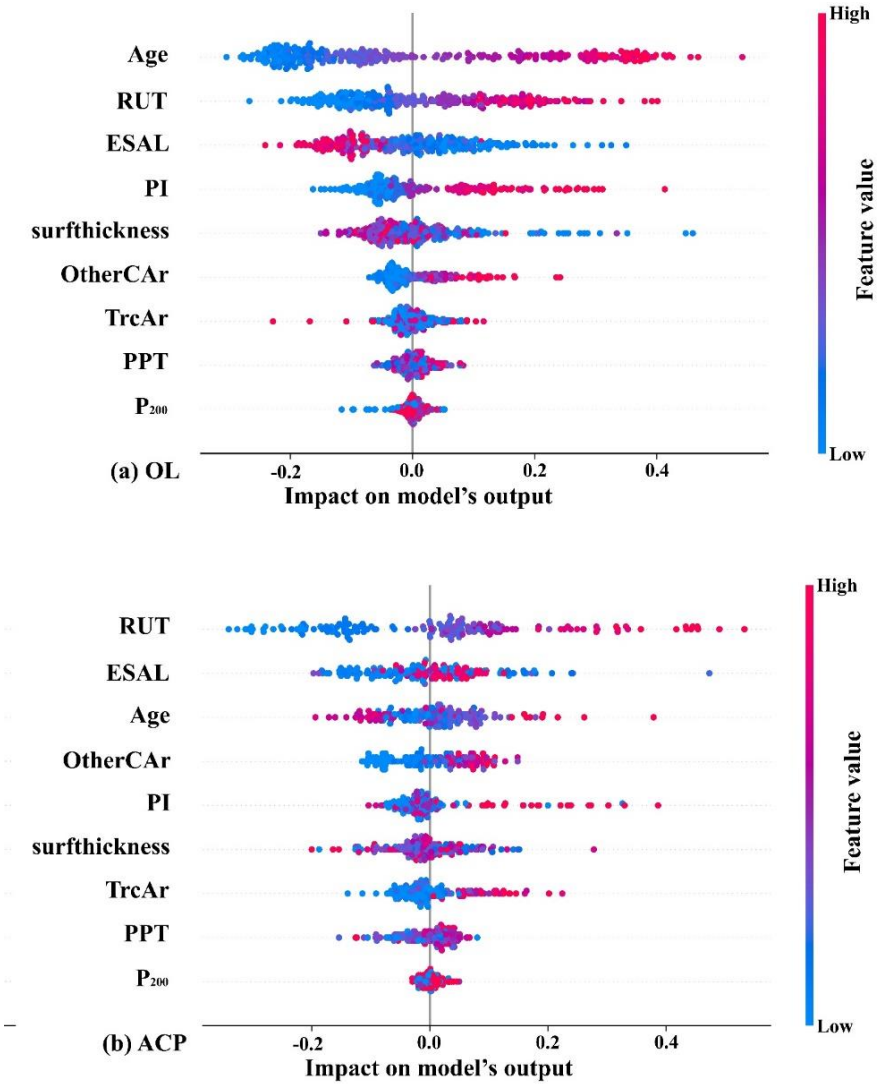

Figure C.6 - CatBoost ML *IRI* models sensitivity analysis. a) OL pavements b) ACP pavements

The sensitivity analysis conducted for CatBoost *IRI* models provides a different picture for some of the pavement features for both pavement types compared to XGBoost's sensitivity analysis. CatBoost models showcase a stronger pavement *RUT* impact for OL pavements than what was seen using XGBoost. The rutting impact appeared to increase from the 4th most important to the 2nd most important feature.

Conversely, the ACP pavement's *IRI* model developed using CatBoost demonstrated a stronger impact for *ESAL* than *Age*; The opposite was true for the XGBoost *IRI* model. The interchangeability in pavement features' importance suggests collinearity between the variables in the models. The remaining pavement features' impacts on the models' output for the CatBoost model are similar to the XGBoost model. A similar sensitivity analysis is conducted for the *IRI* models developed using the LightGBM algorithm and is presented in Figure C.7.

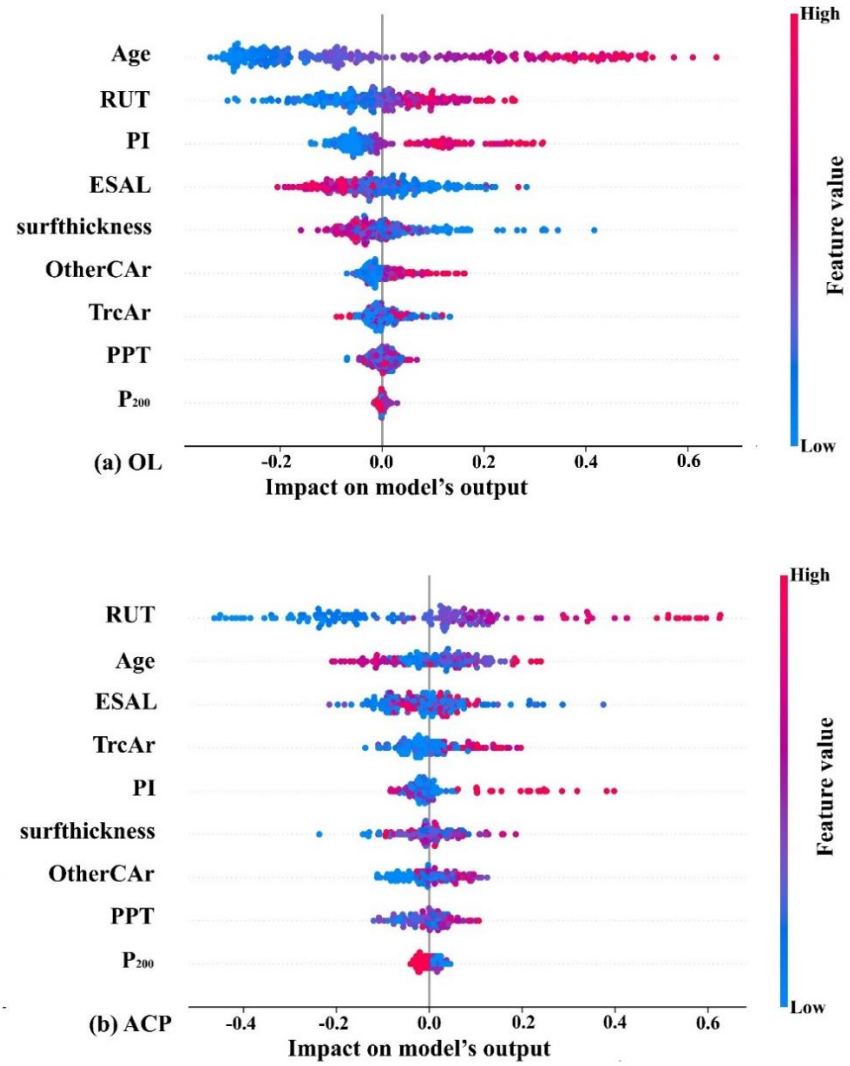

Figure C.7 - LightGBM ML *IRI* models sensitivity analysis. a) Overlay pavements b) ACP pavements

LightGBM *IRI* models' sensitivity analysis showcases the variables' impact strength for both pavements than the models developed using XGBoost and CatBoost algorithms. LightGBM *IRI* models' sensitivity

analysis shows a weaker *ESAL* impact for OL pavements than XGBoost and CatBoost models, a stronger *RUT* impact than the XGBoost model, and a stronger *PI* impact than the CatBoost model. For ACP pavements, *TrcAr* poses a much stronger impact for the LightGBM *IRI* model, while *OtherCAr* has a much weaker impact than in both XGBoost and CatBoost models. The impact change for ACP pavement features suggests a correlation between *TrcAr* and *OtherCAr*. The LightGBM *IRI* models' sensitivity analysis further supports the presence of collinearity between the models' variables. A similar sensitivity analysis is conducted for the Random-Forest *IRI* models, presented in Figure C.8.

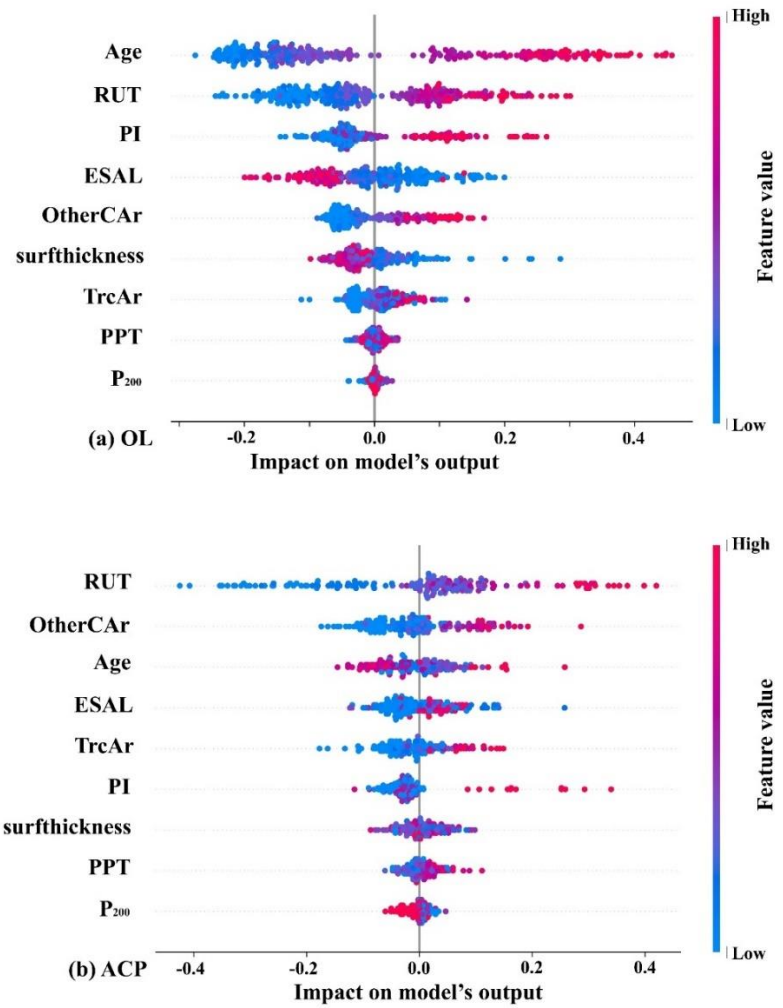

Figure C.8 – Random-Forest ML *IRI* models sensitivity analysis. a) Overlay pavements b) ACP pavements

The Random-Forest *IRI* model for OL pavements showcases very similar features' impacts to that of the LightGBM *IRI* model. For OL pavements, the impact of *OtherCAr* is stronger than *surfthickness* when compared to the LightGBM model's sensitivity analysis. This similarity in sensitivity analysis between Random-Forest and LightGBM models is not present for ACP pavements. Compared with XGBoost, CatBoost, and LightGBM *IRI* models, the sensitivity analysis shows a much stronger impact from *OtherCAr* and a lesser impact from *ESAL* and *surfthickness* on ACP pavements. Moreover, the analysis indicates a stronger impact from *TrcAr* than what is observed with XGBoost and CatBoost *IRI* models.

The sensitivity analysis conducted for XGBoost, CatBoost, LightGBM, and Random-Forest *IRI* models helps demonstrate how pavement features affect the *IRI* models' output. Some of the least impactful pavement features on all the developed *IRI* models' output were *PPT* and P_{200}. The analyses also demonstrated different impact strengths for some of the pavement features used in the *IRI* models due to the variables' collinearity. The most impactful pavement features for OL pavements *IRI* models were more consistent and included *Age*, *RUT*, *PI*, and *ESAL*. The pavement features that had the most impact on the *IRI* model's output for ACP pavements varied in strength; however, *RUT*, *Age*, and *ESAL* were some of the most impactful features.

The conducted sensitivity analyses are used to compute the average impact of the pavement features on the models' *IRI* output values to further understand the features' relationship with the models' output values. The average impact of each pavement feature highlights each feature's importance and considers the range of values for the feature. The pavement features' average impact values are plotted in Figure C.9 for the developed ML *IRI* models. As the pavement features' average impact values are built using the sensitivity analyses conducted earlier, similar results are found in identifying the most impactful pavement features. The pavement features *Age*, *RUT*, *PI*, and *ESAL* appeared to be the most impactful features for OL pavements, while *RUT*, *Age*, and *ESAL* were the most impactful features for ACP pavements. Nonetheless, when considering the five most impactful pavement features for both pavement types across the four highlighted ML algorithms, the variables *RUT*, *ESAL*, *Age*, *PI*, and *Surfthickness*

appeared to be the most impactful to the models' output. The figure shows varying pavement features' impact, possibly due to collinearity between the variables in the models.

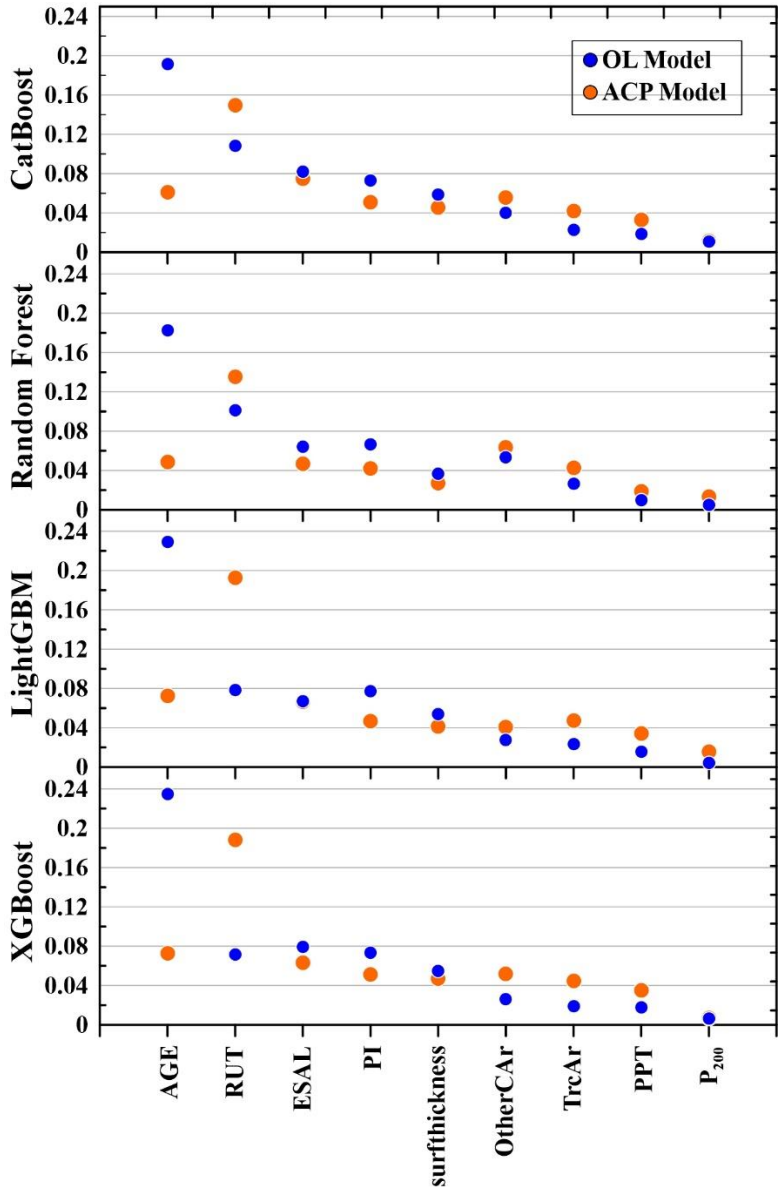

Figure C.9 - Average pavement features' impact on *IRI* models' output values

Appendix D: SPS-1 experiment Data from LTPP's database

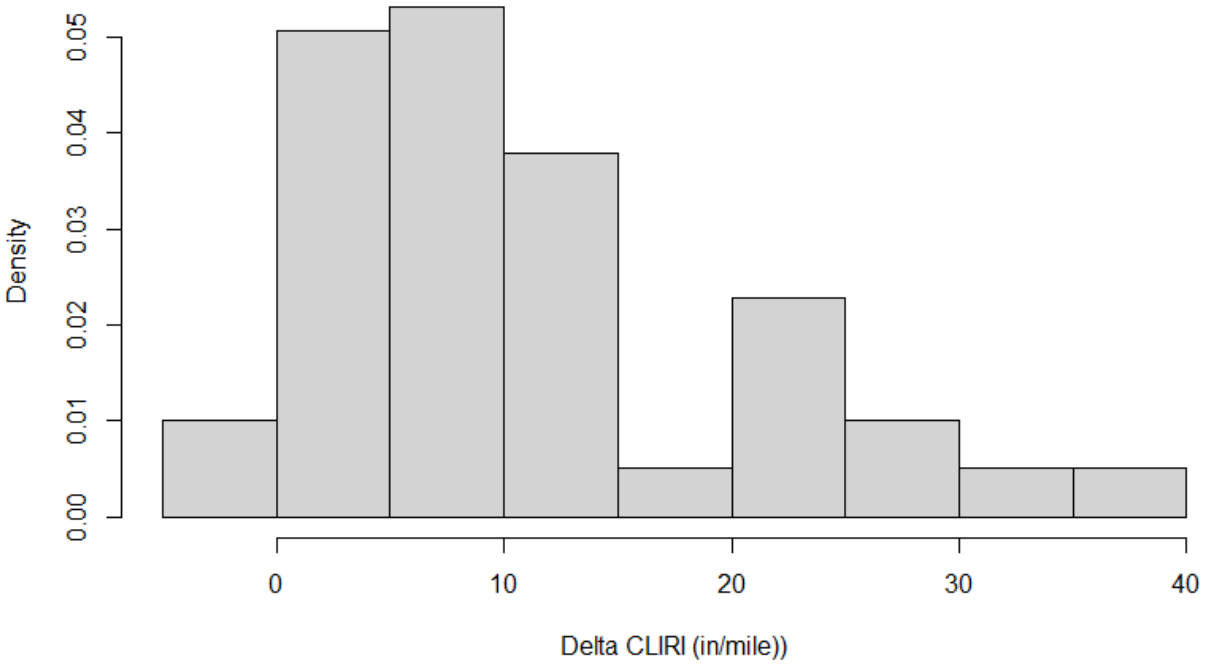

Figure D.1 – Delta *CLIRI* values histogram plot for fine subgrade pavements

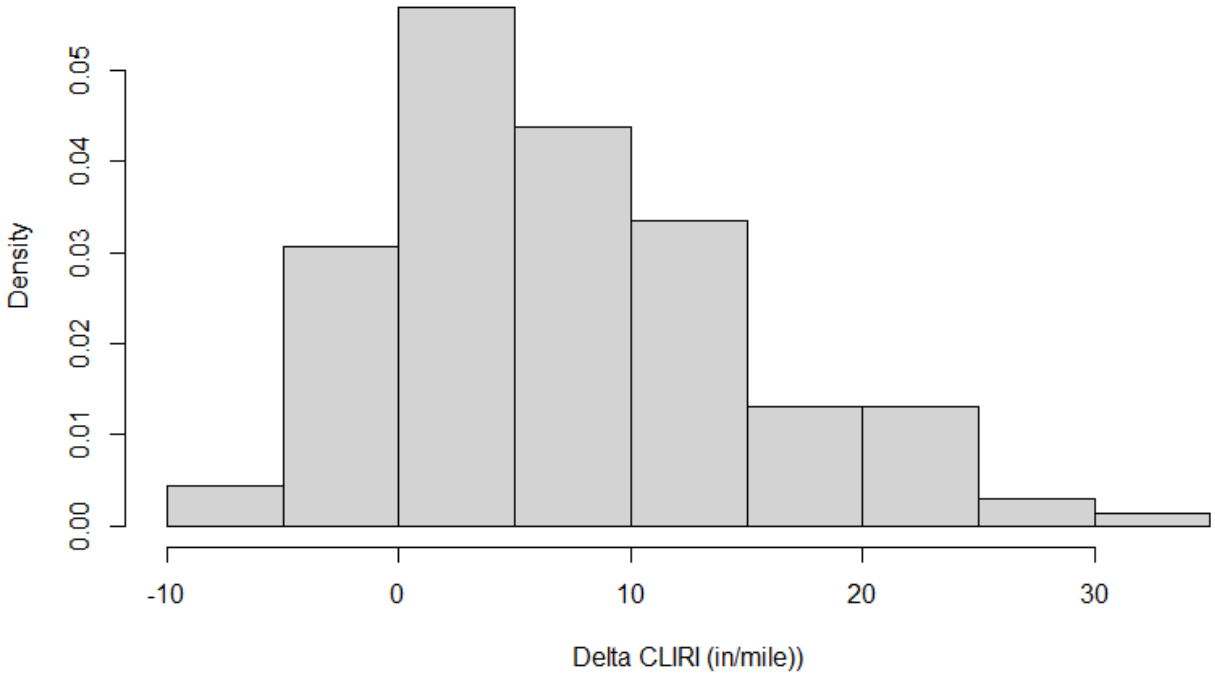

Figure D.2 – Delta *CLIRI* values histogram plot for coarse subgrade pavements

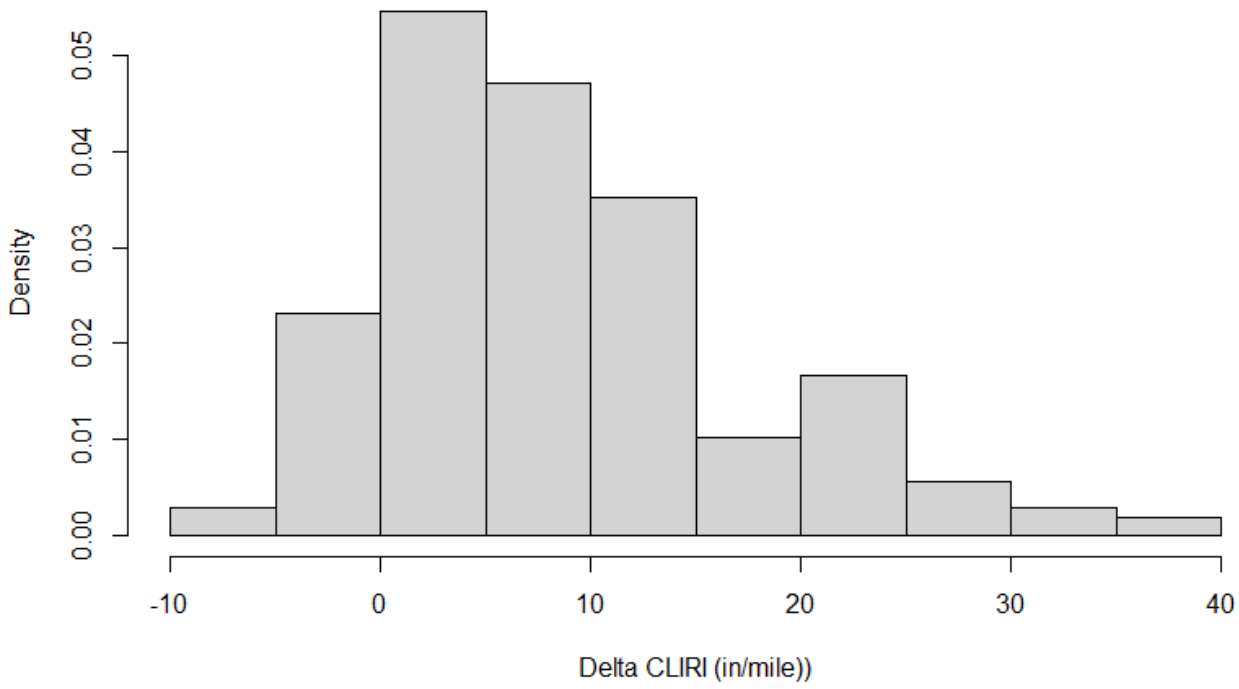

Figure D.3 – Delta *CLIRI* values histogram plot for all subgrade pavements

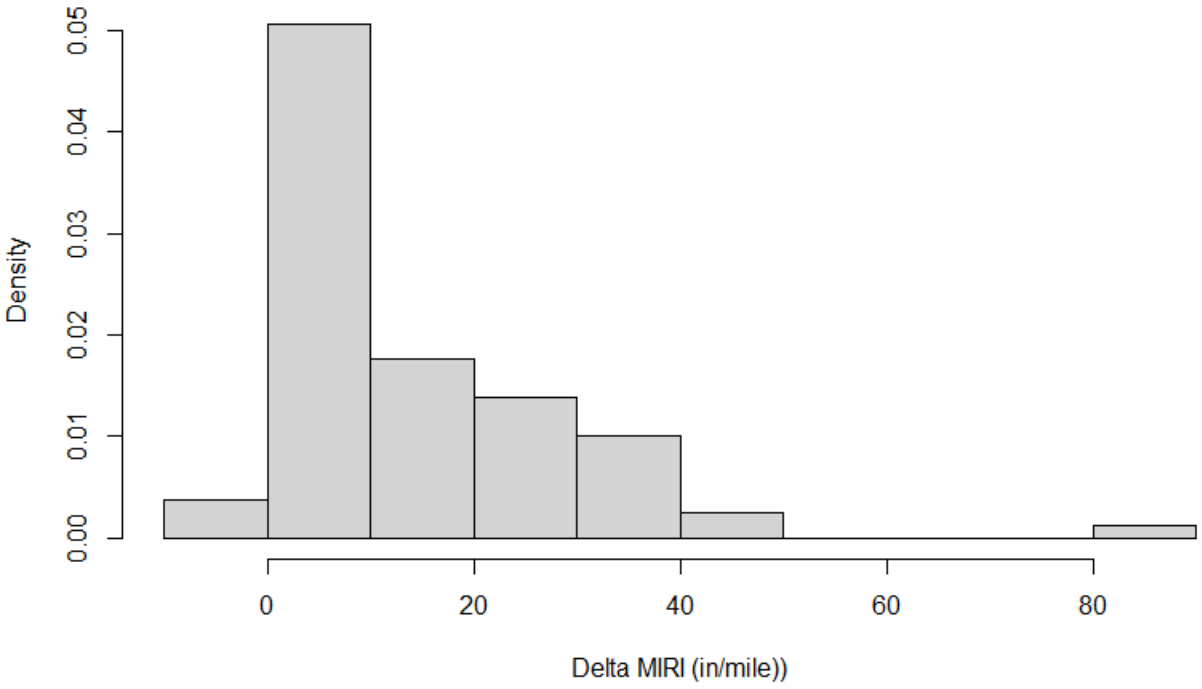

Figure D.4 – Delta *MIRI* values histogram plot for fine subgrade pavements

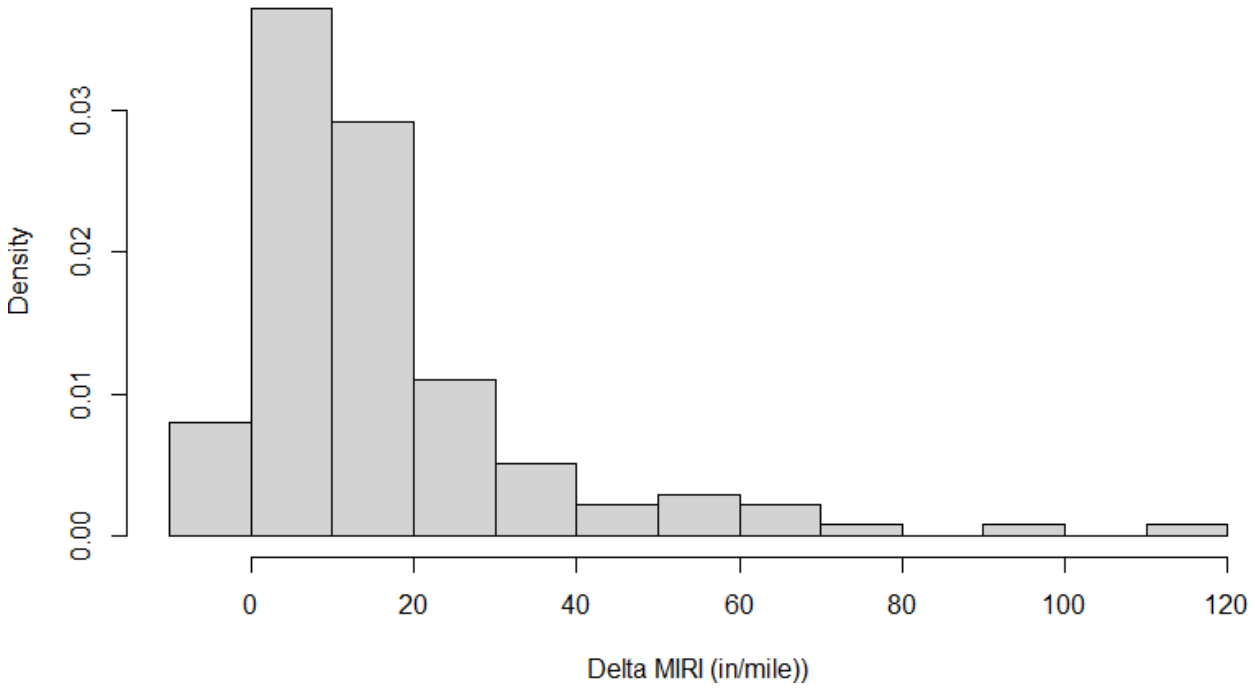

Figure D.5 – Delta *MIRI* values histogram plot for coarse subgrade pavements

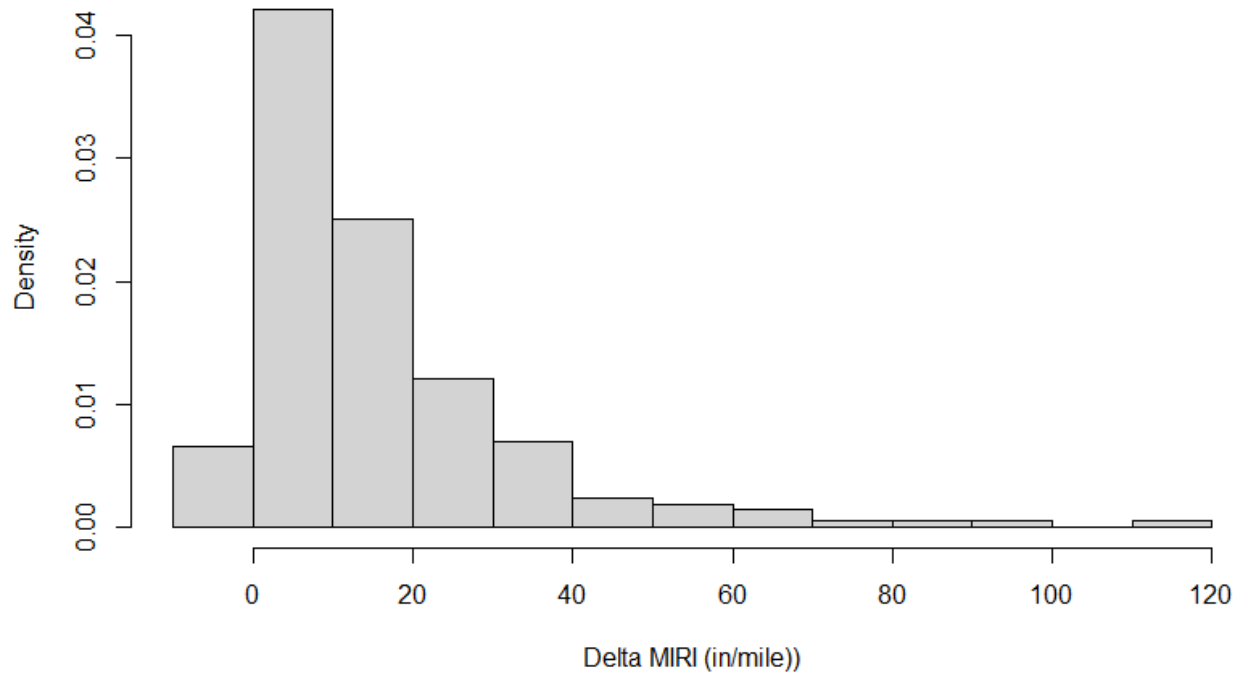

Figure D.6 – Delta *MIRI* values histogram plot for all subgrade pavements

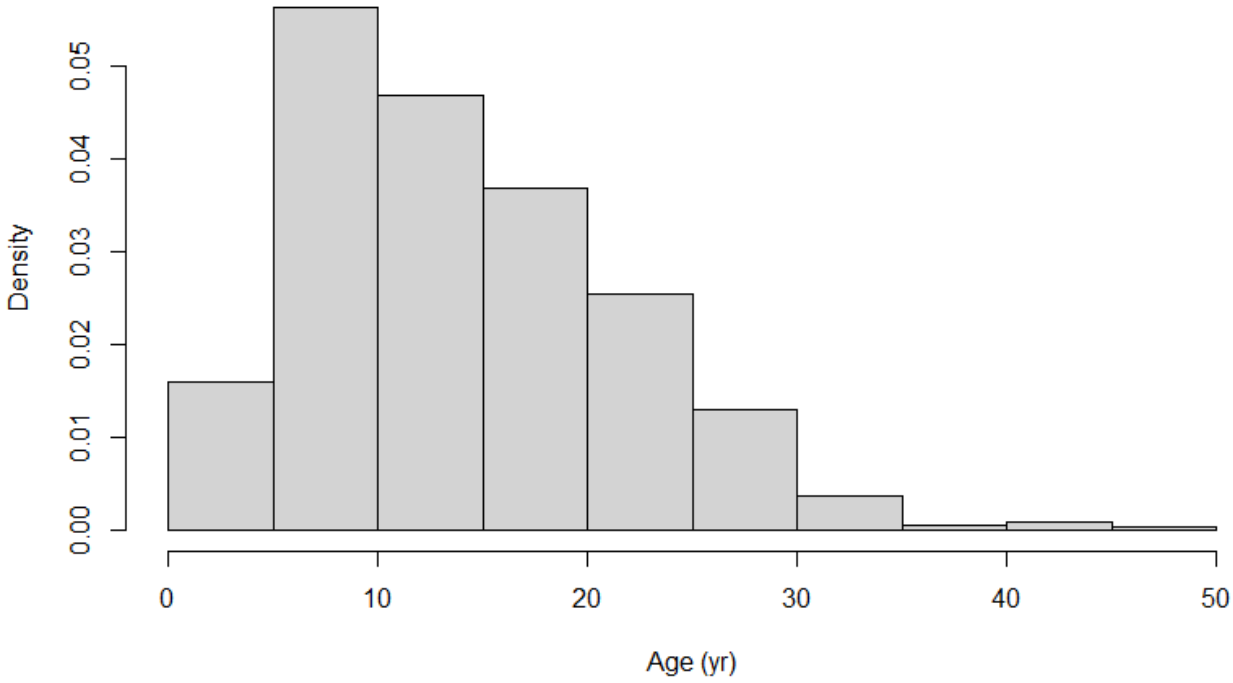

Figure D.7 – *Age* values histogram plot for fine subgrade pavements

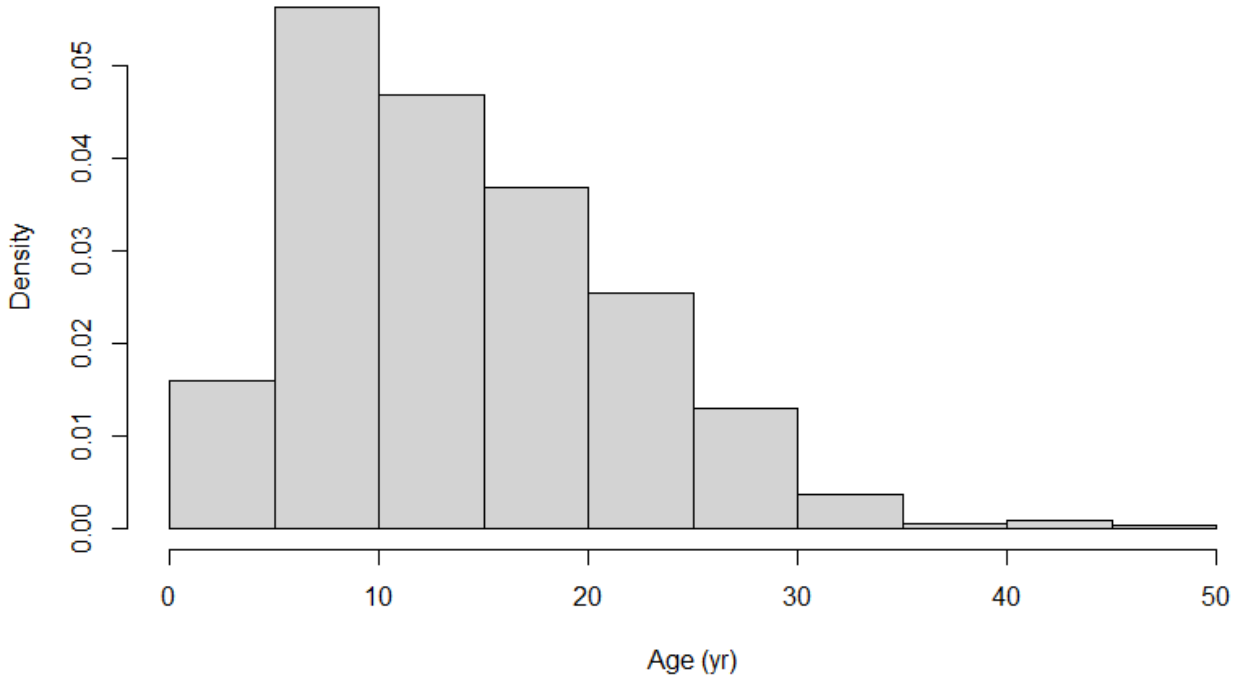

Figure D.8 – *Age* values histogram plot for coarse subgrade pavements

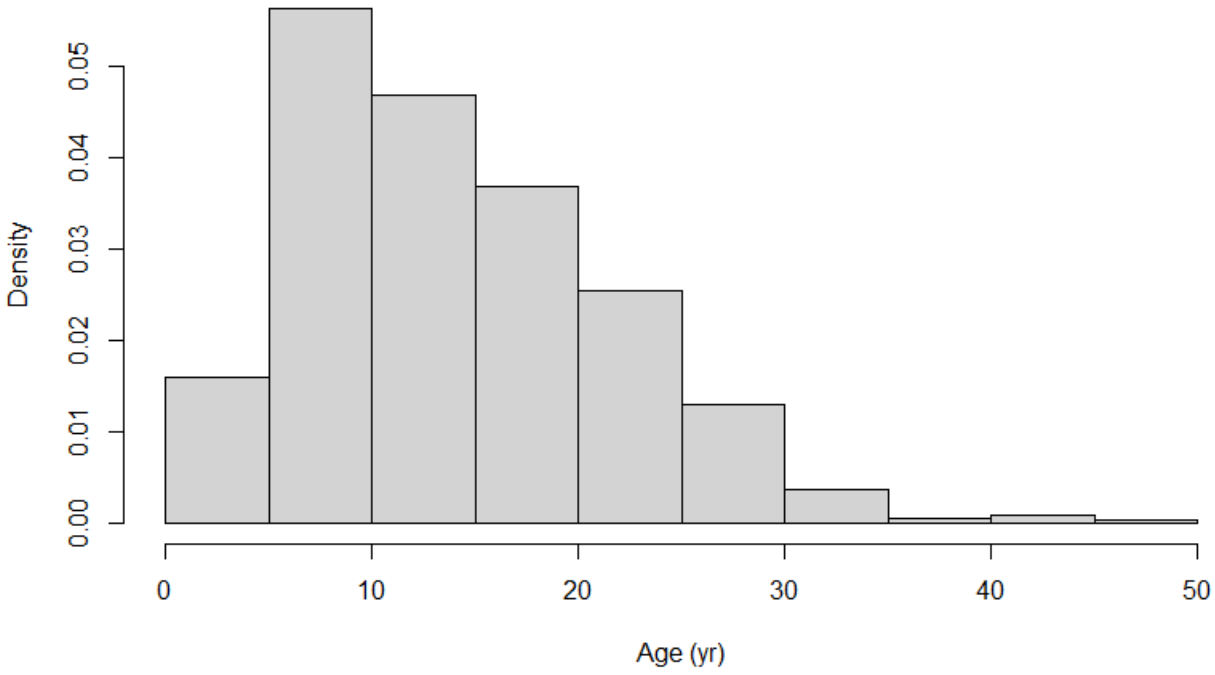

Figure D.9 – *Age* values histogram plot for all subgrade pavements

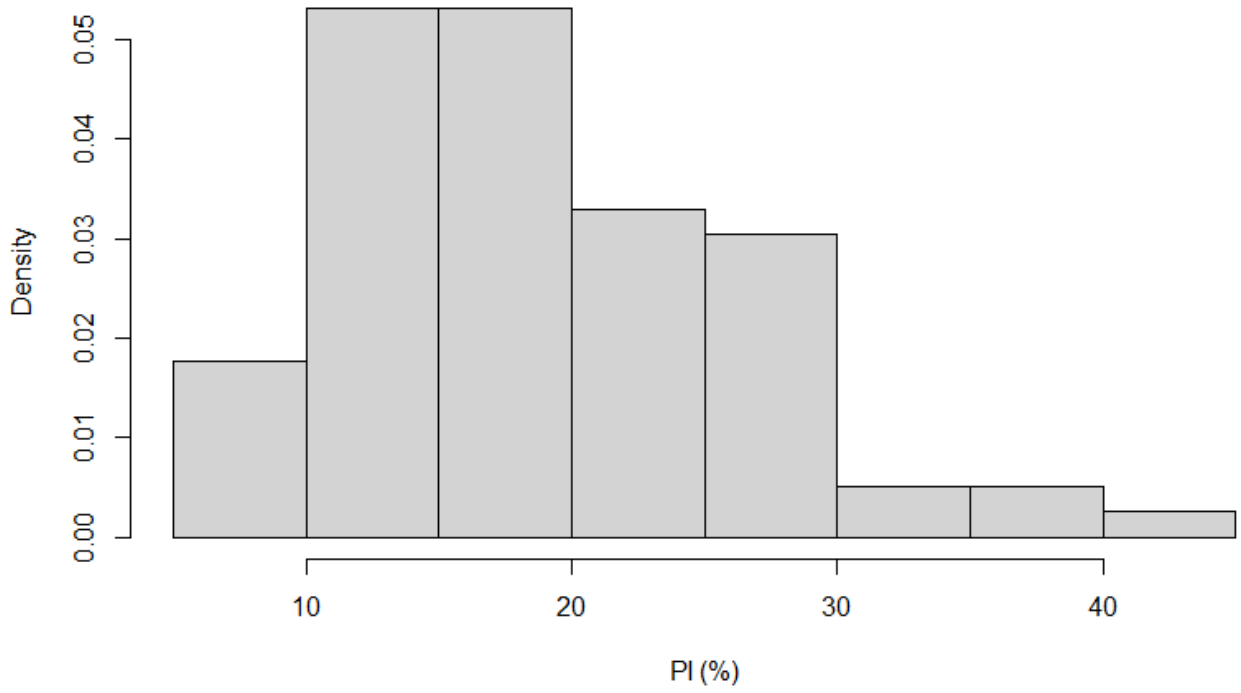

Figure D.10 – *PI* values histogram plot for fine subgrade pavements

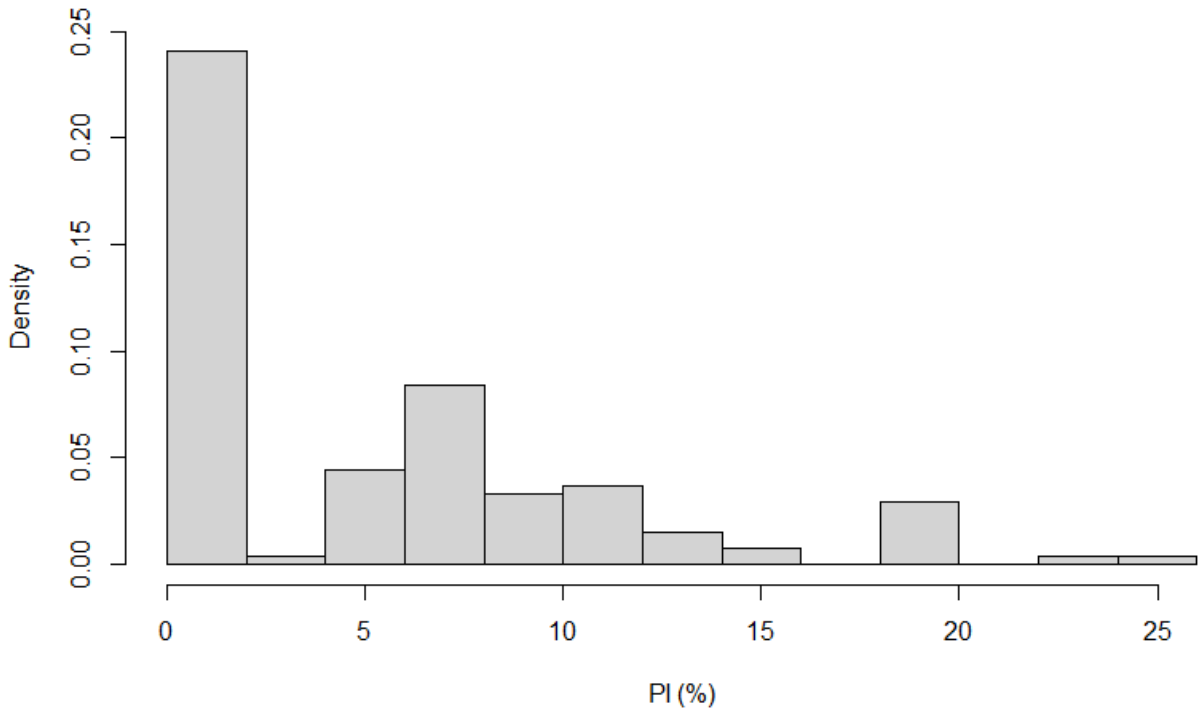

Figure D.11 – *PI* values histogram plot for coarse subgrade pavements

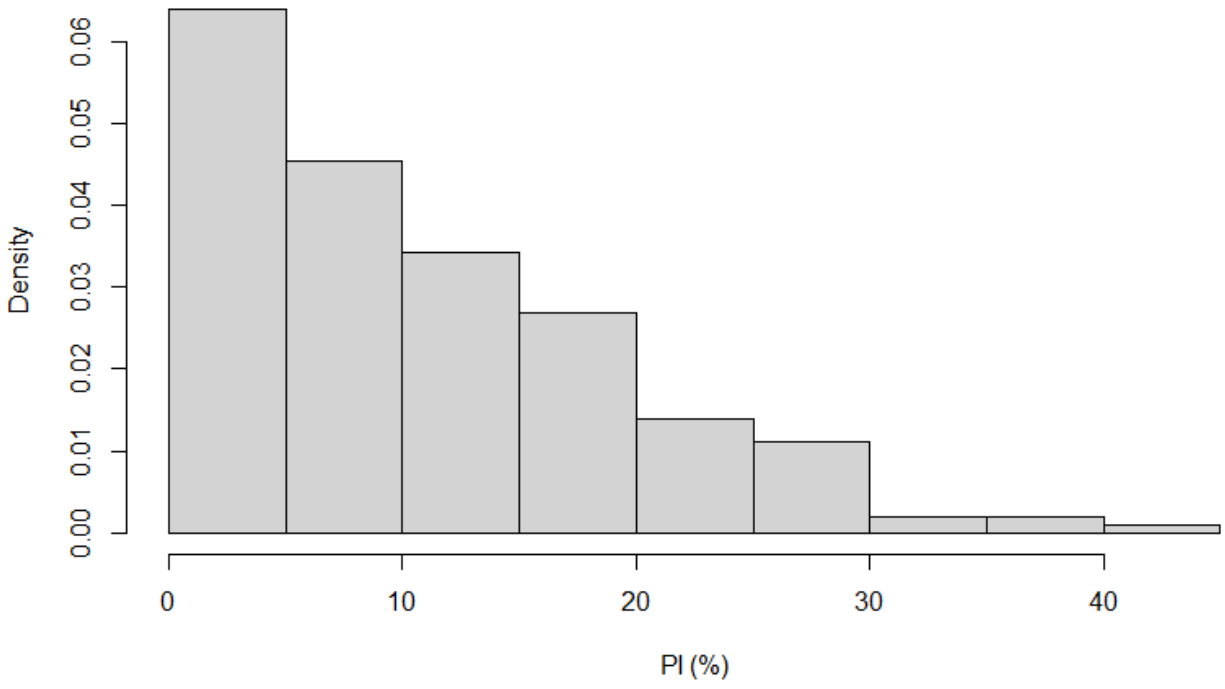

Figure D.12 – *PI* values histogram plot for all subgrade pavements

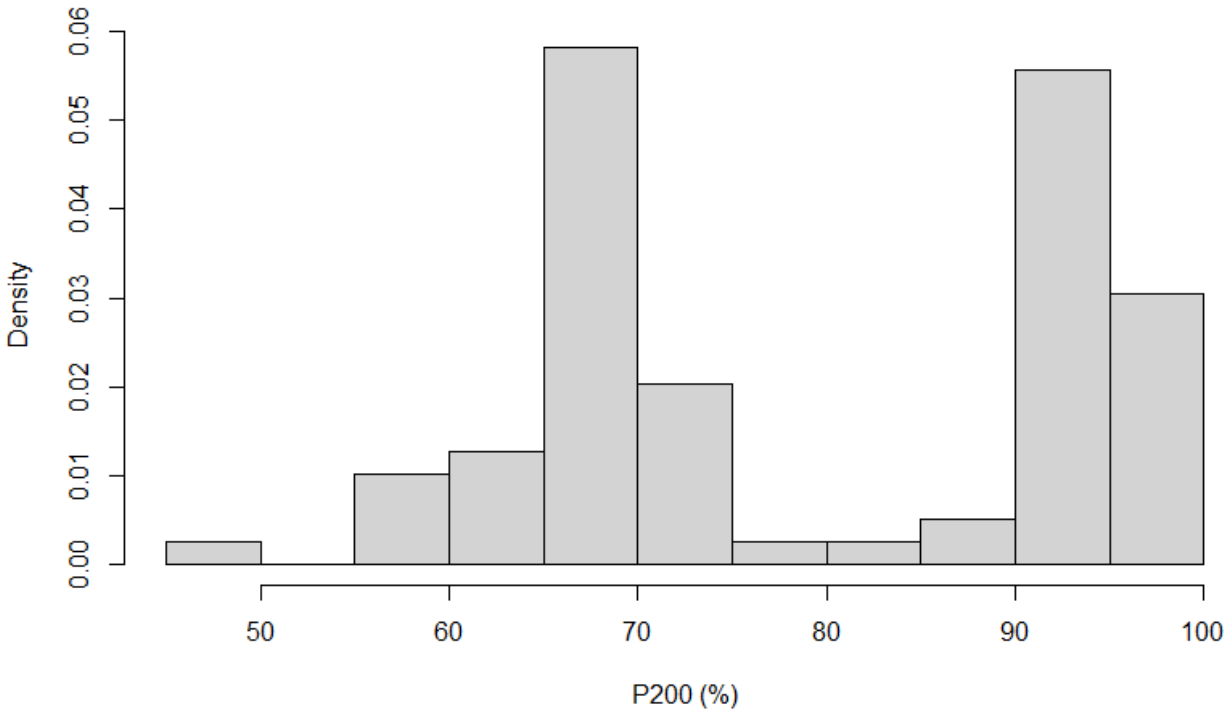

Figure D.13 – P_{200} values histogram plot for fine subgrade pavements

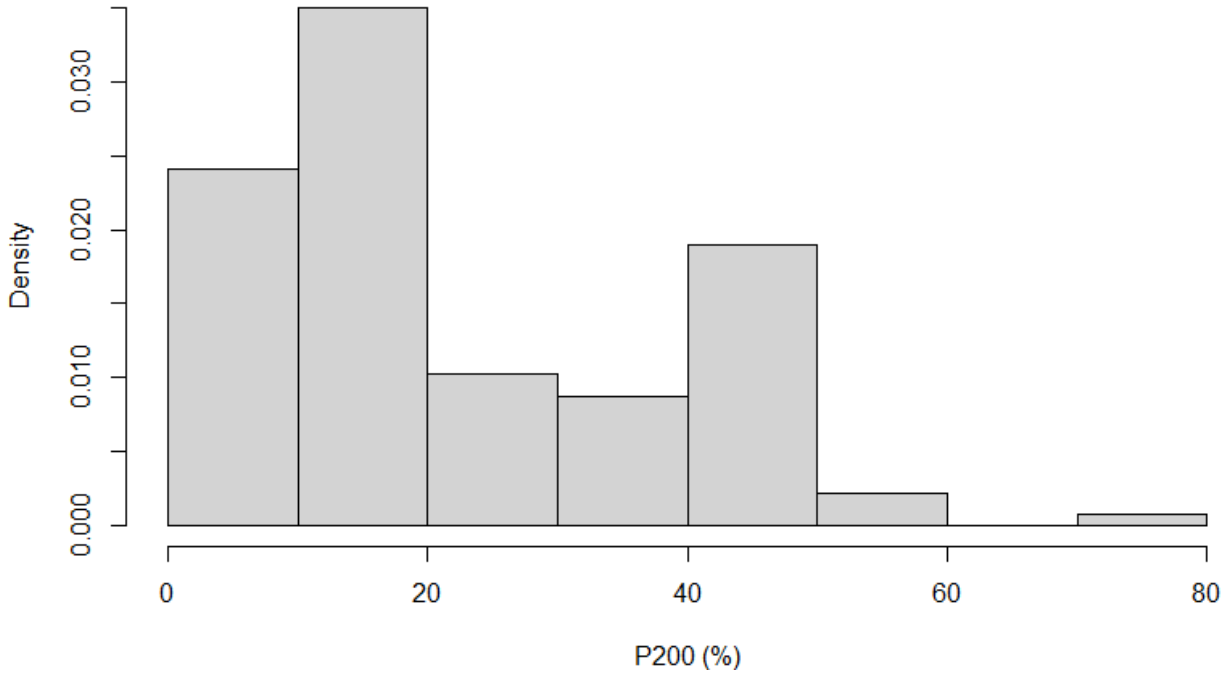

Figure D.14 – P_{200} values histogram plot for coarse subgrade pavements

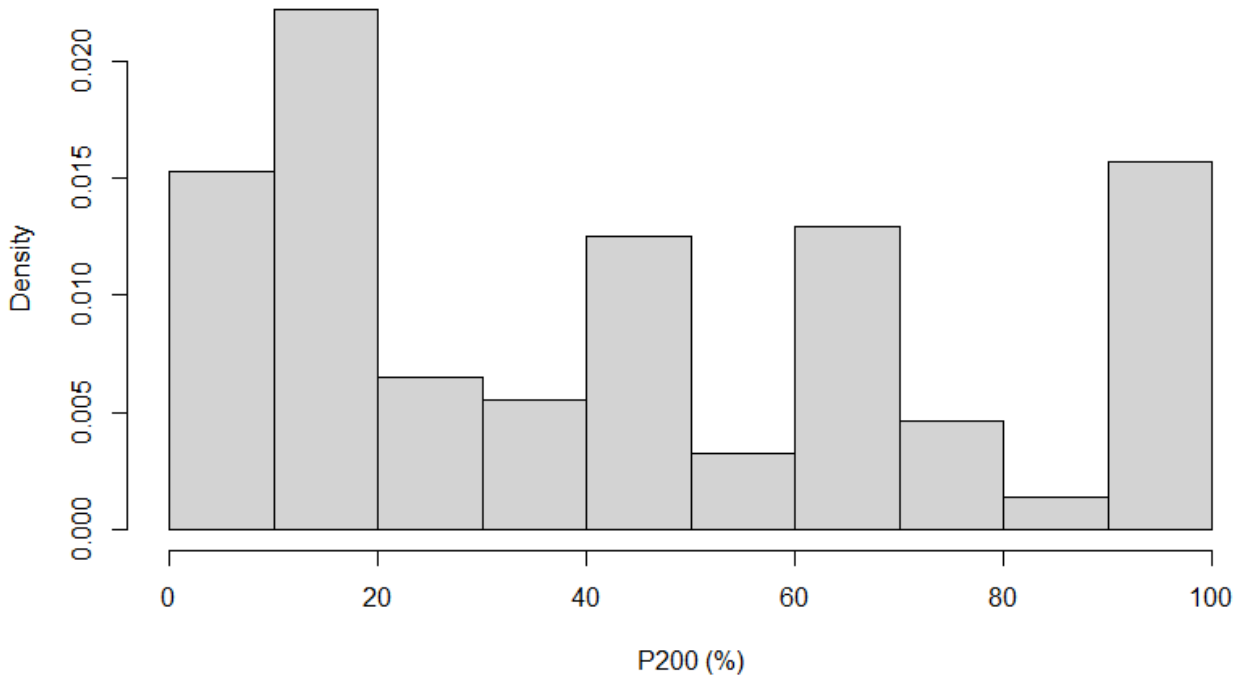

Figure D.15 – P_{200} values histogram plot for all subgrade pavements

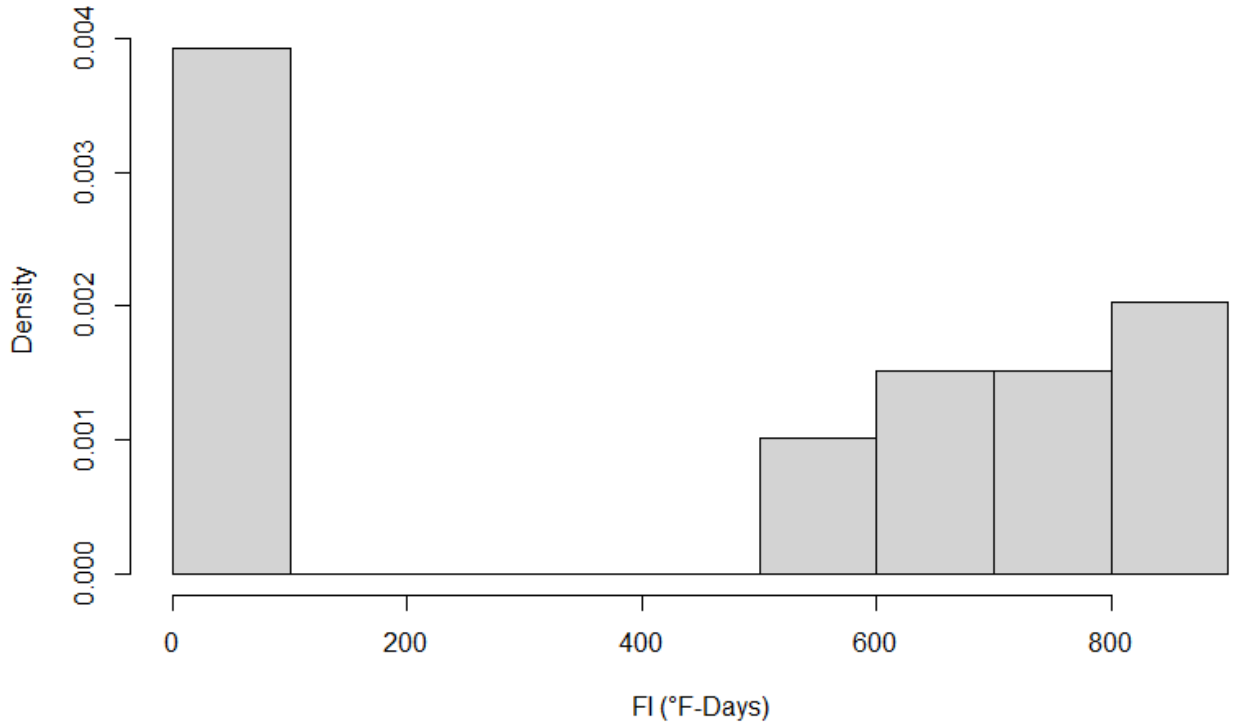

Figure D.16 – *FI* values histogram plot for fine subgrade pavements

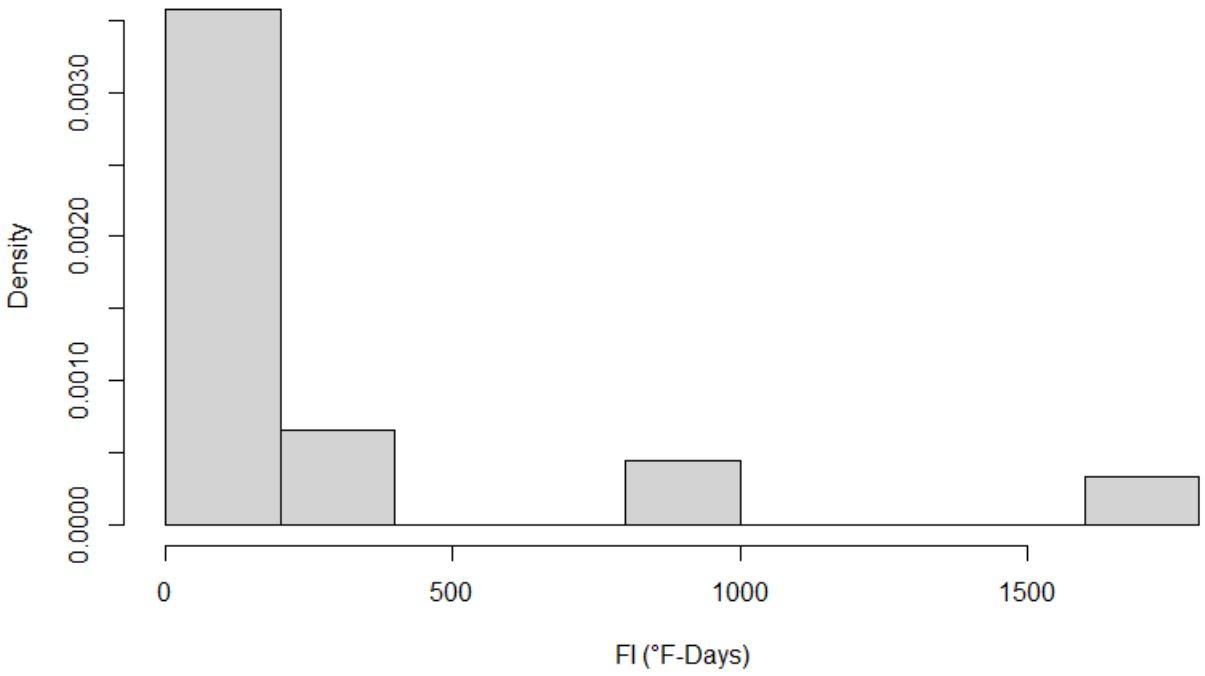

Figure D.17 – *FI* values histogram plot for coarse subgrade pavements

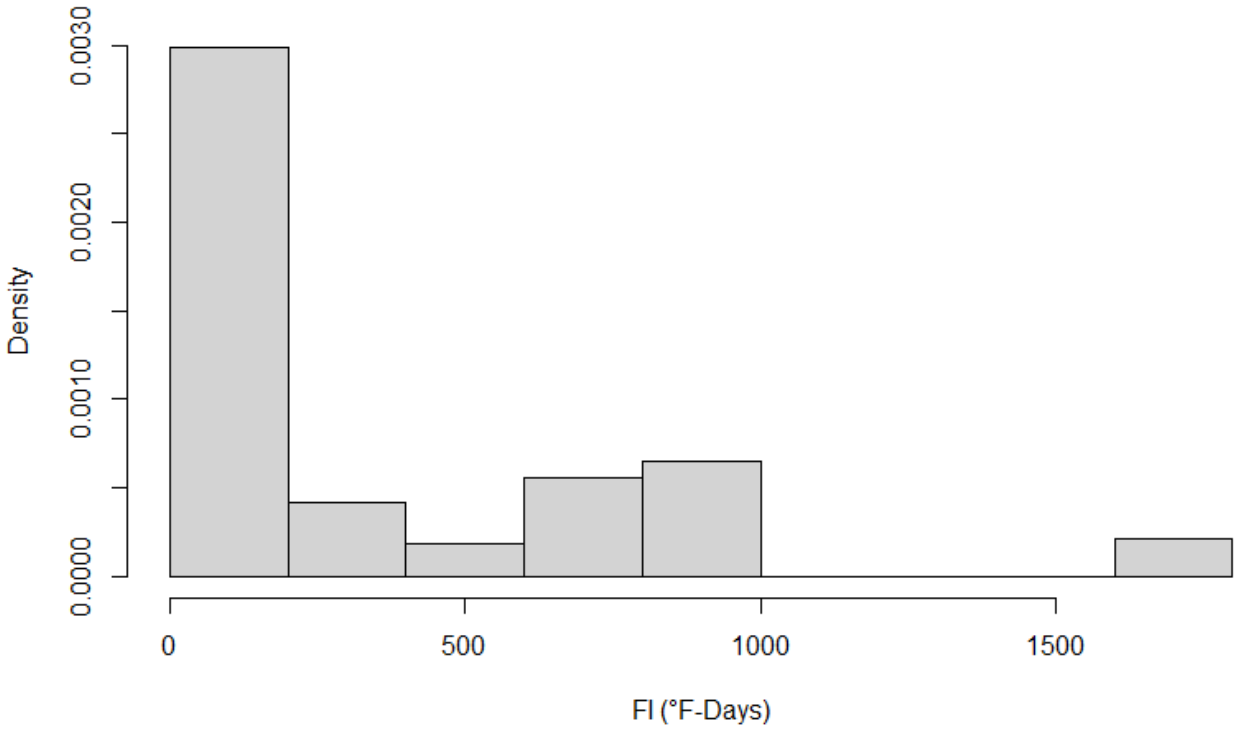

Figure D.18 – *FI* values histogram plot for all subgrade pavements

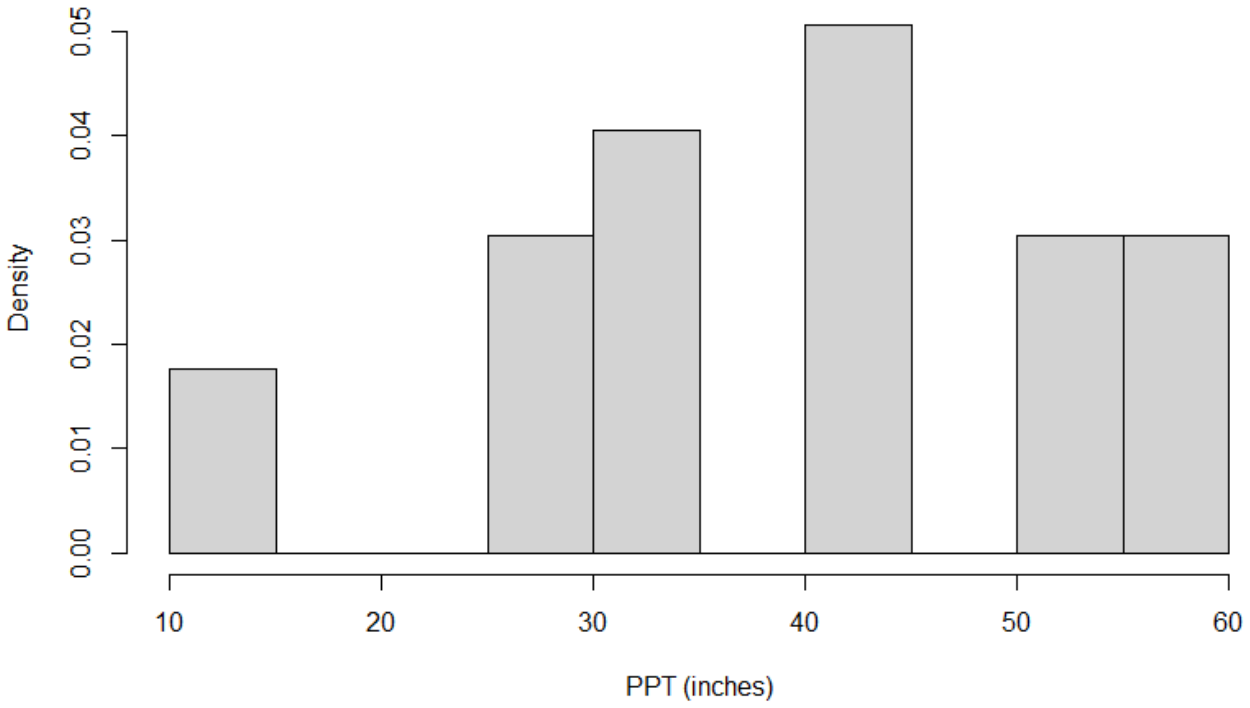

Figure D.19 – *PPT* values histogram plot for fine subgrade pavements

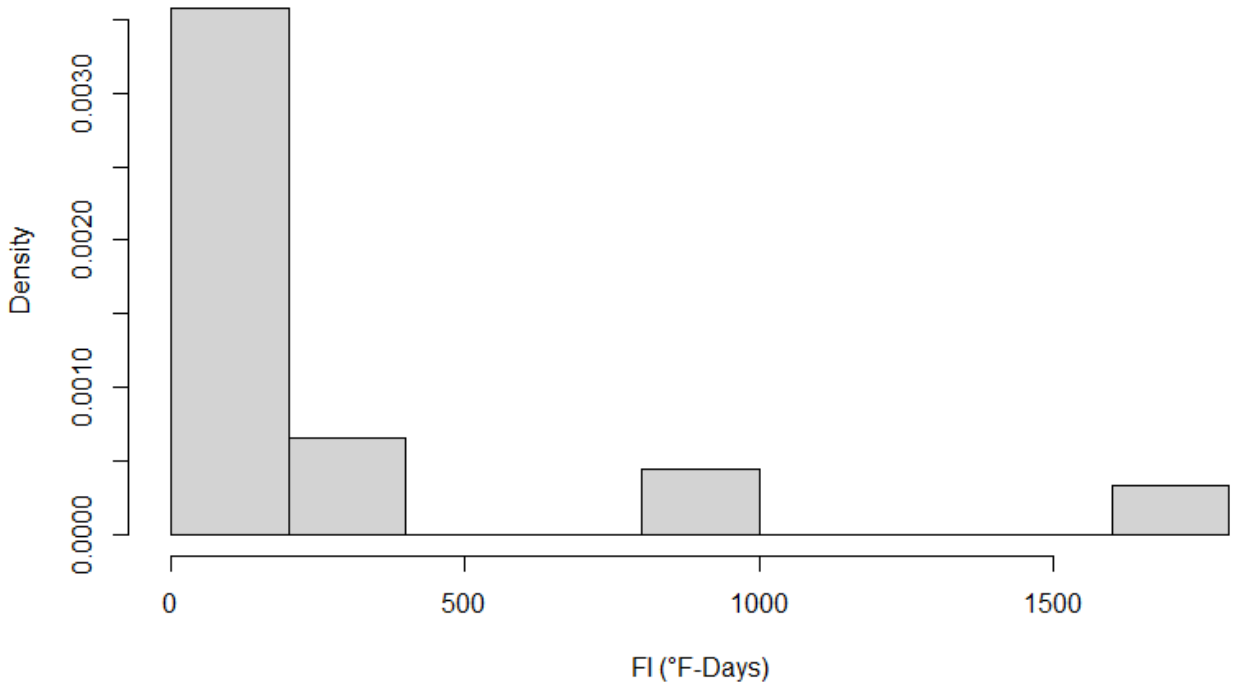

Figure D.20 – *PPT* values histogram plot for coarse subgrade pavements

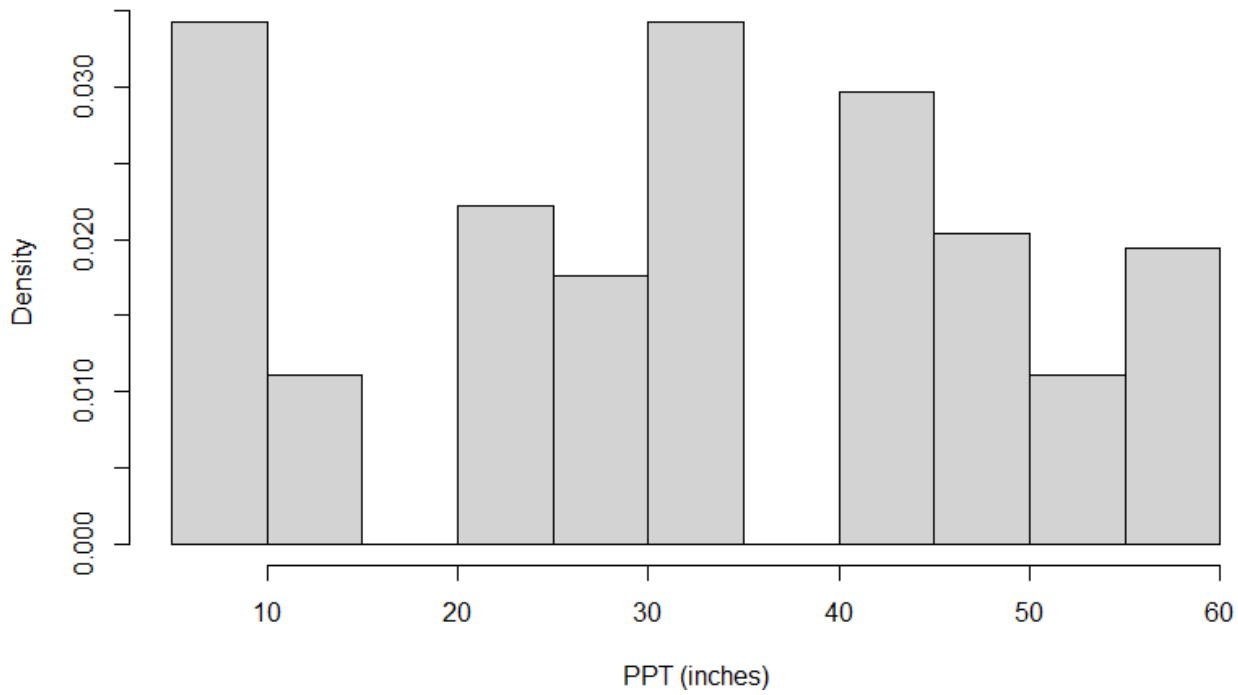

Figure D.21 – *PPT* values histogram plot for all subgrade pavements

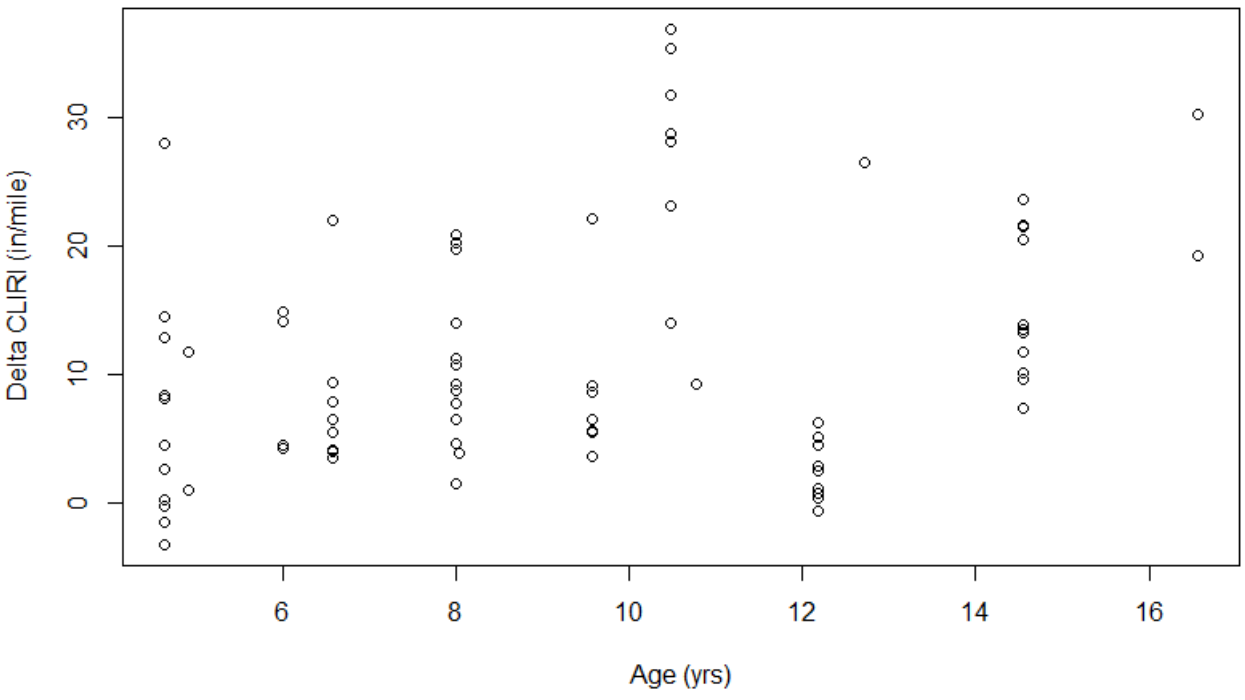

Figure D.22 – Delta *CLIRI* and *Age* values scatter plot for fine subgrade pavements

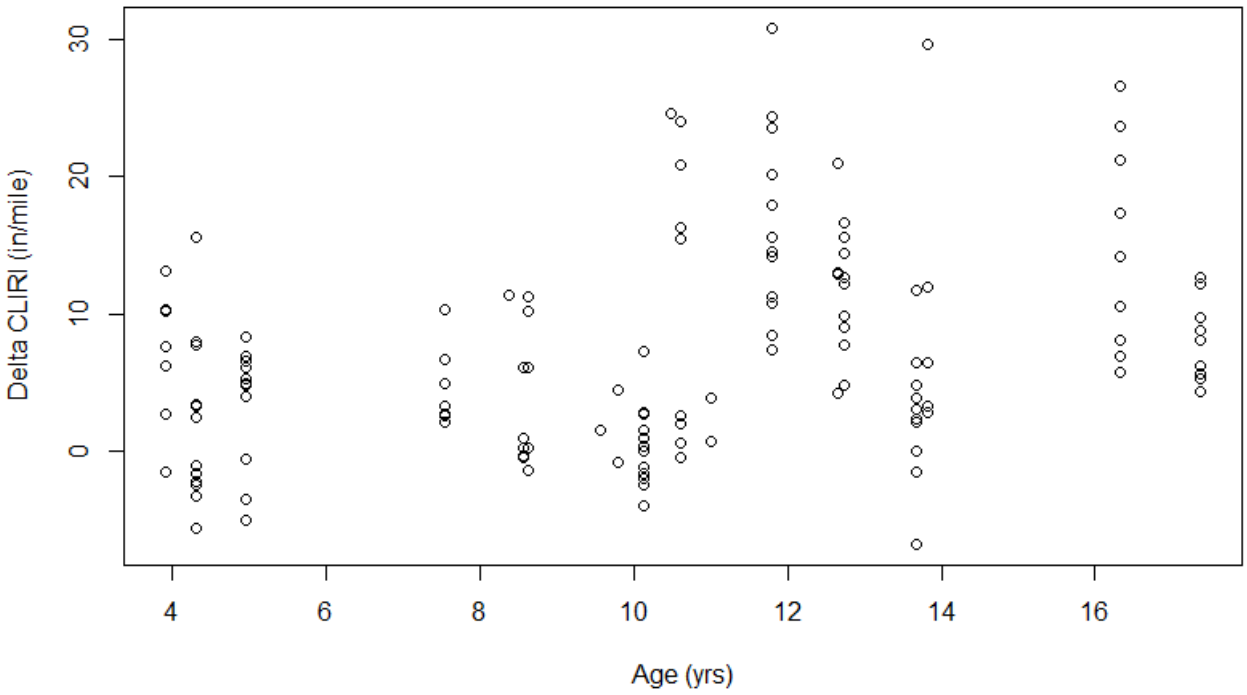

Figure D.23 – Delta *CLIRI* and *Age* values histogram plot for coarse subgrade pavements

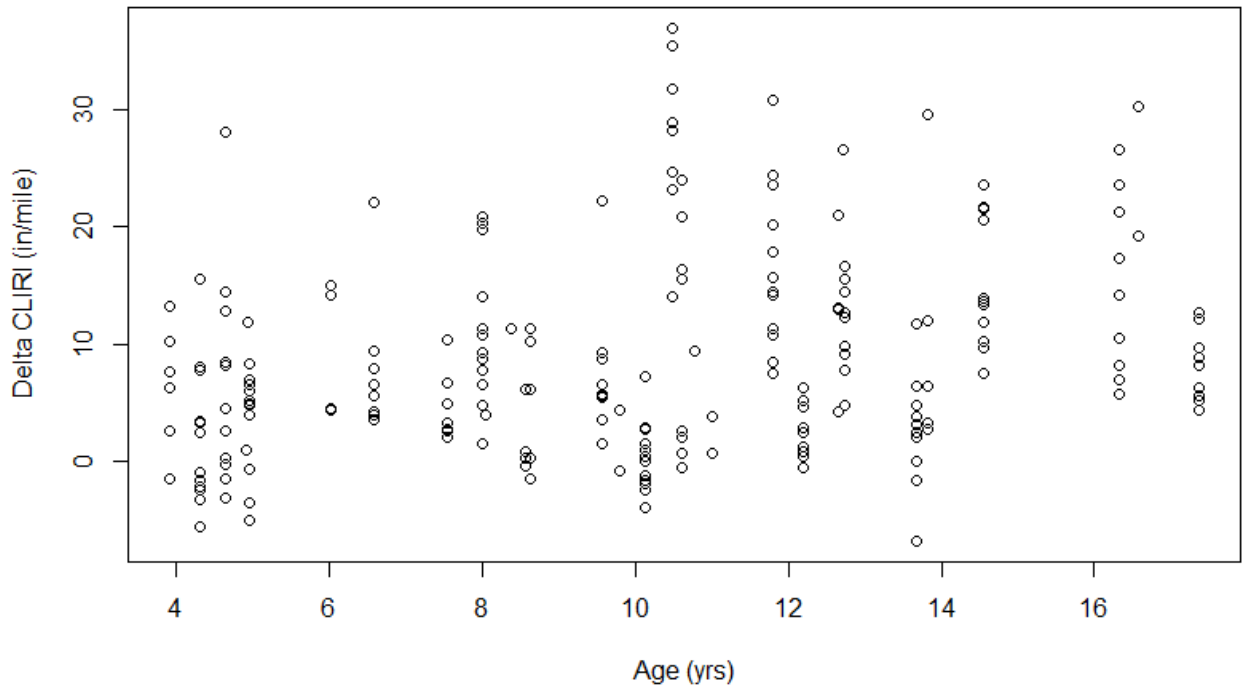

Figure D.24 – Delta *CLIRI* and *Age* values histogram plot for all subgrade pavements

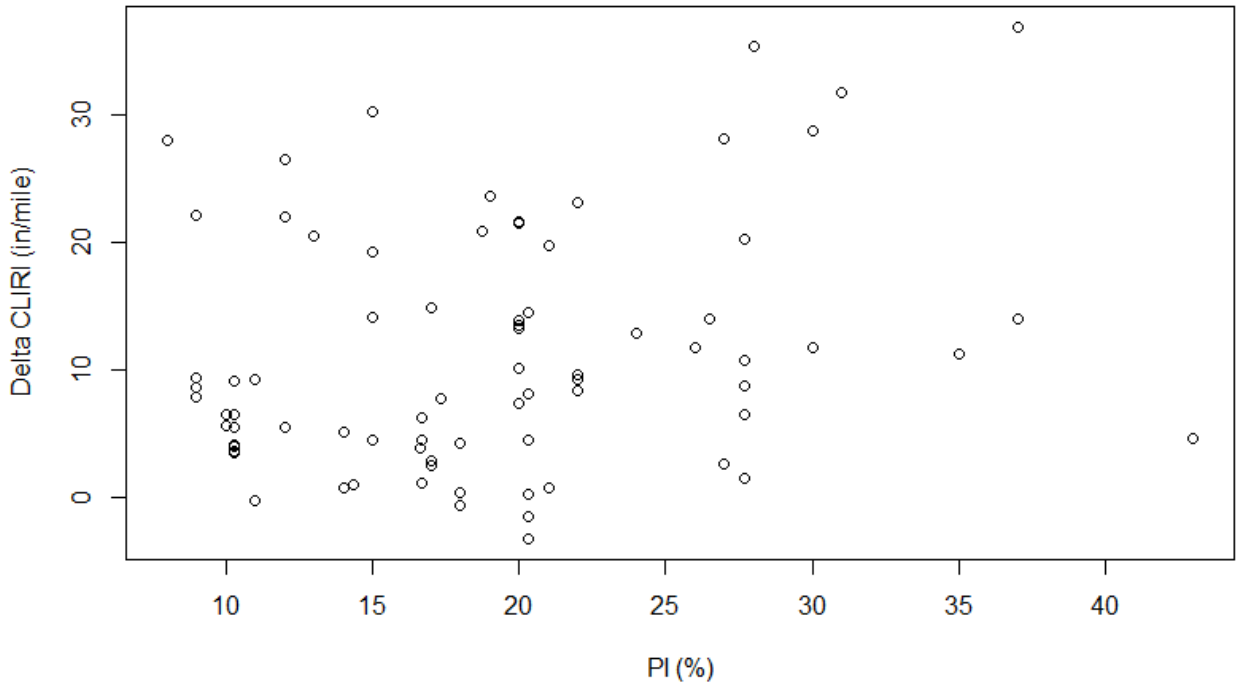

Figure D.25 – Delta *CLIRI* and *PI* values scatter plot for fine subgrade pavements

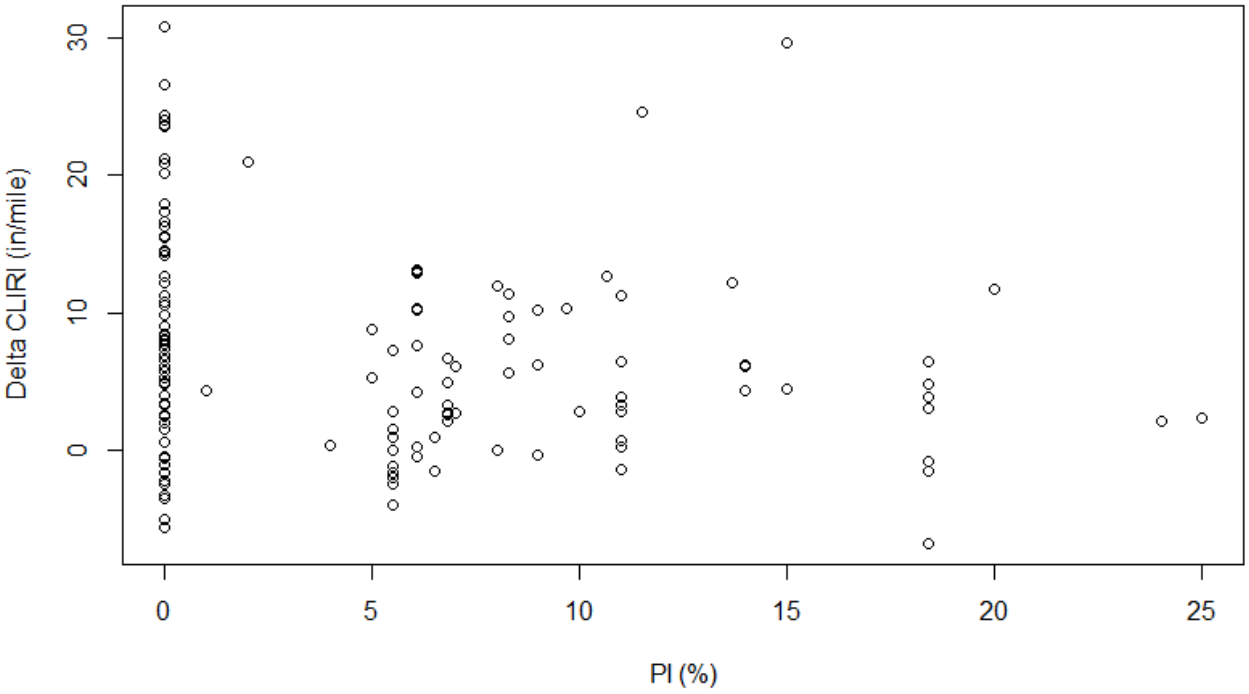

Figure D.26 – Delta *CLIRI* and *PI* values histogram plot for coarse subgrade pavements

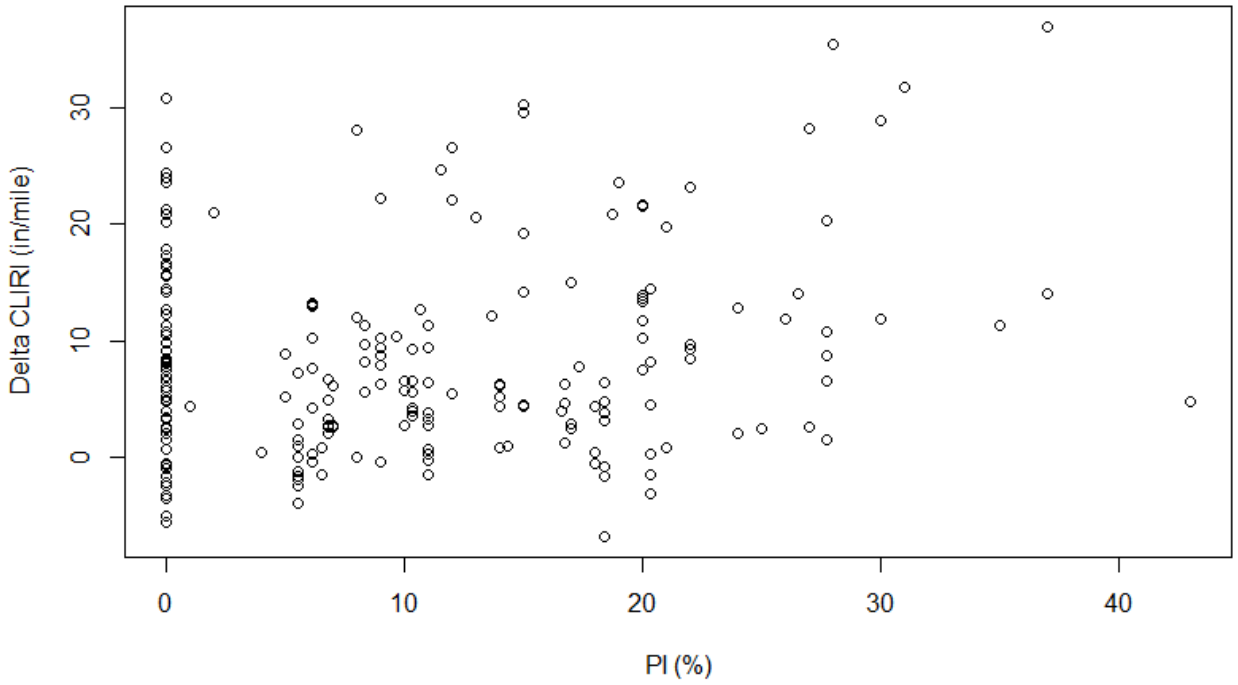

Figure D.27 – Delta *CLIRI* and *PI* values histogram plot for all subgrade pavements

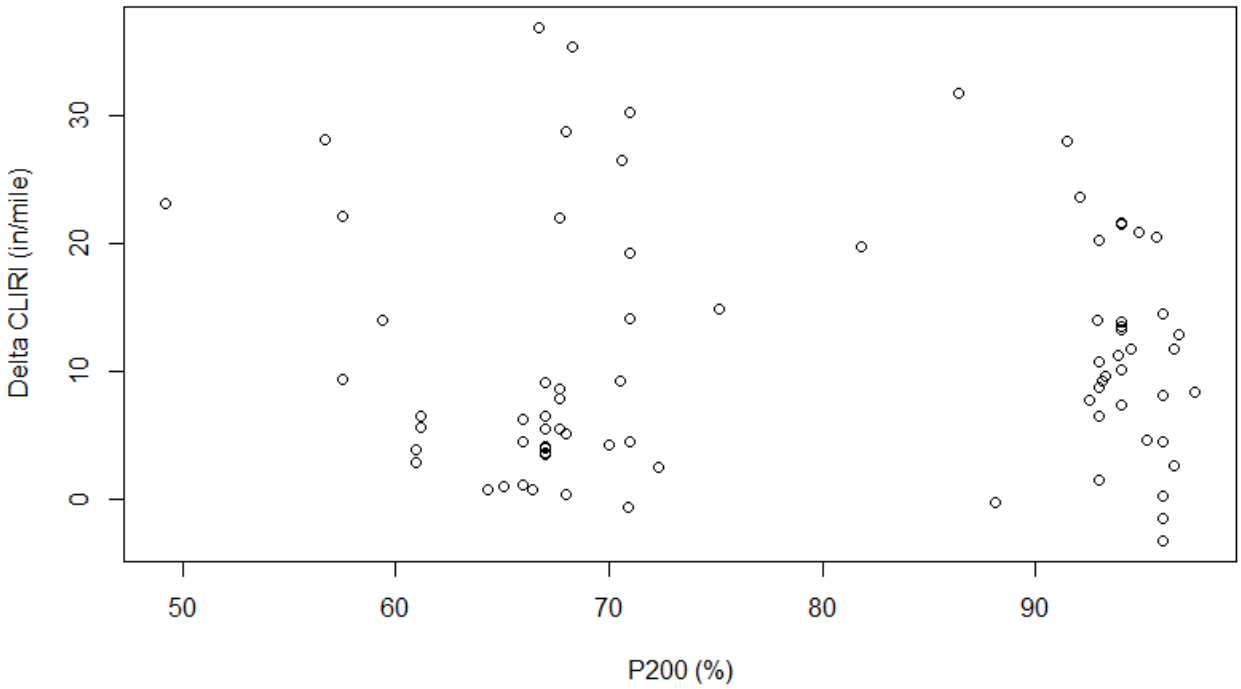

Figure D.28 – Delta *CLIRI* and P_{200} values scatter plot for fine subgrade pavements

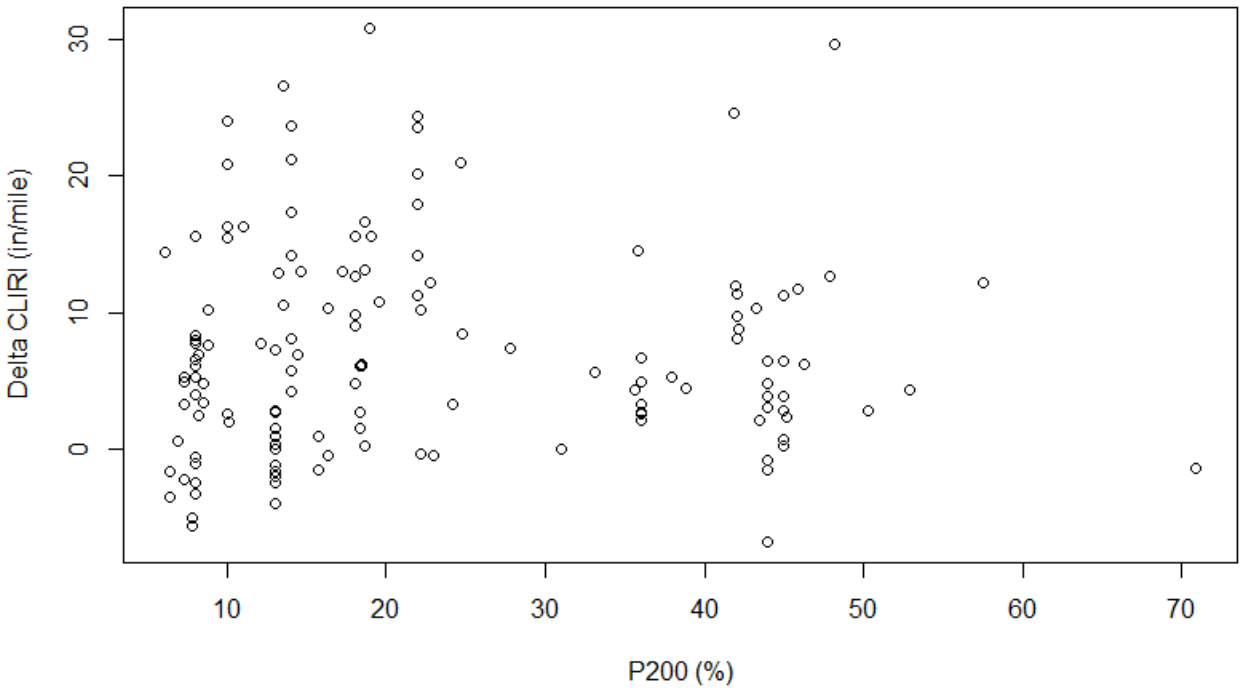

Figure D.29 – Delta *CLIRI* and P_{200} values histogram plot for coarse subgrade pavements

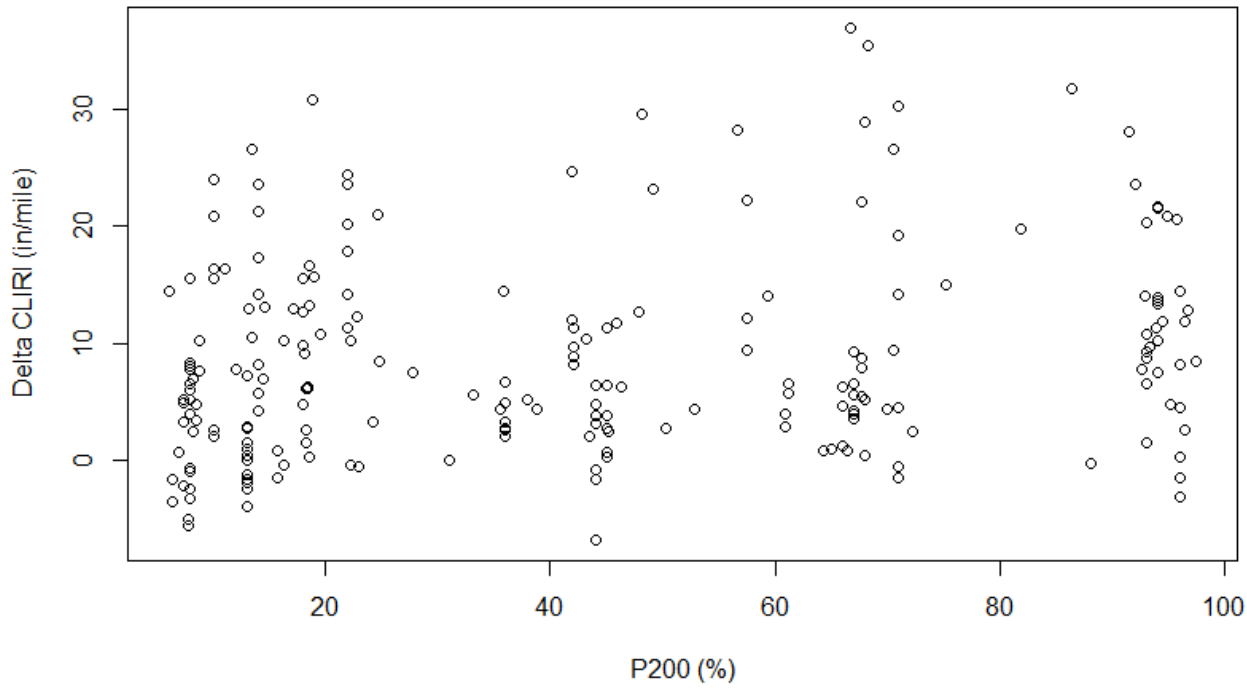

Figure D.30 – Delta *CLIRI* and P_{200} values histogram plot for all subgrade pavements

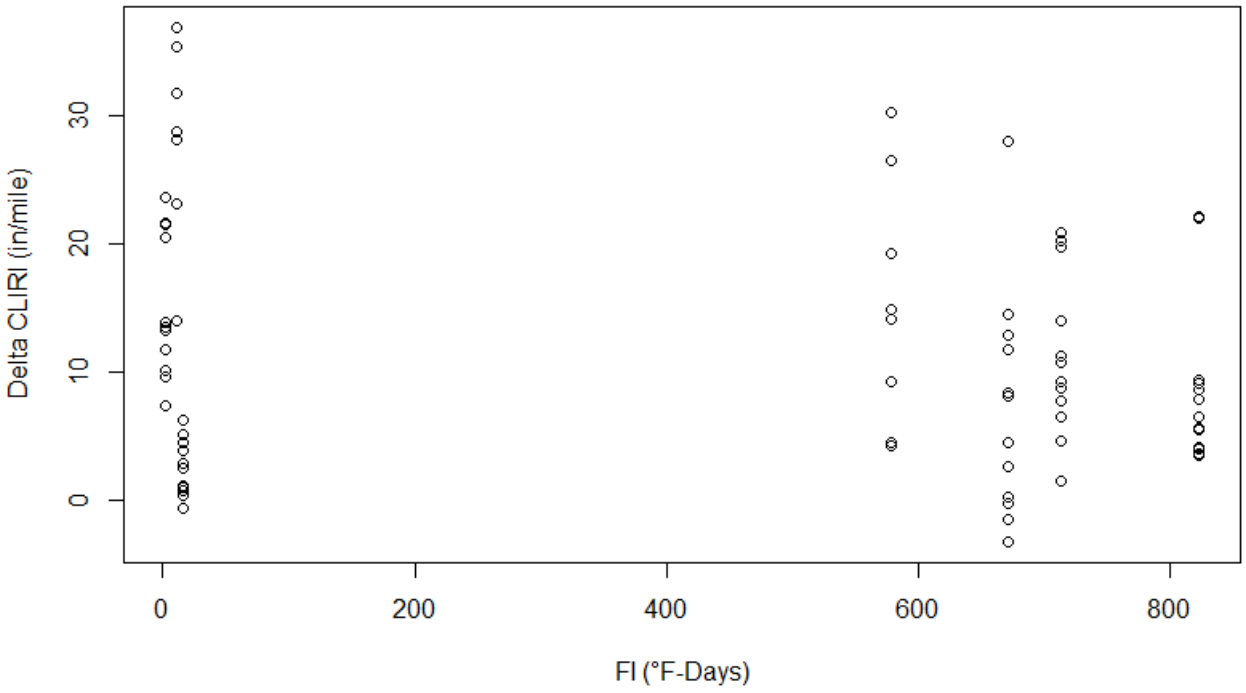

Figure D.31 – Delta *CLIRI* and *FI* values scatter plot for fine subgrade pavements

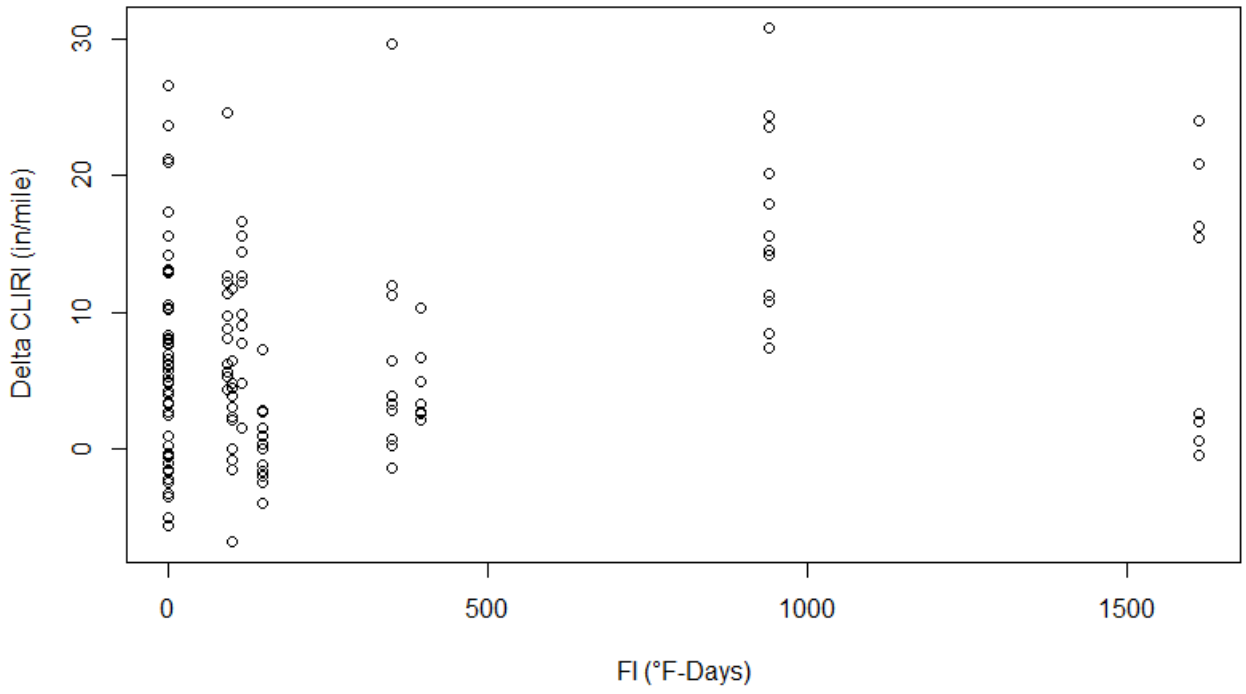

Figure D.32 – Delta *CLIRI* and *FI* values histogram plot for coarse subgrade pavements

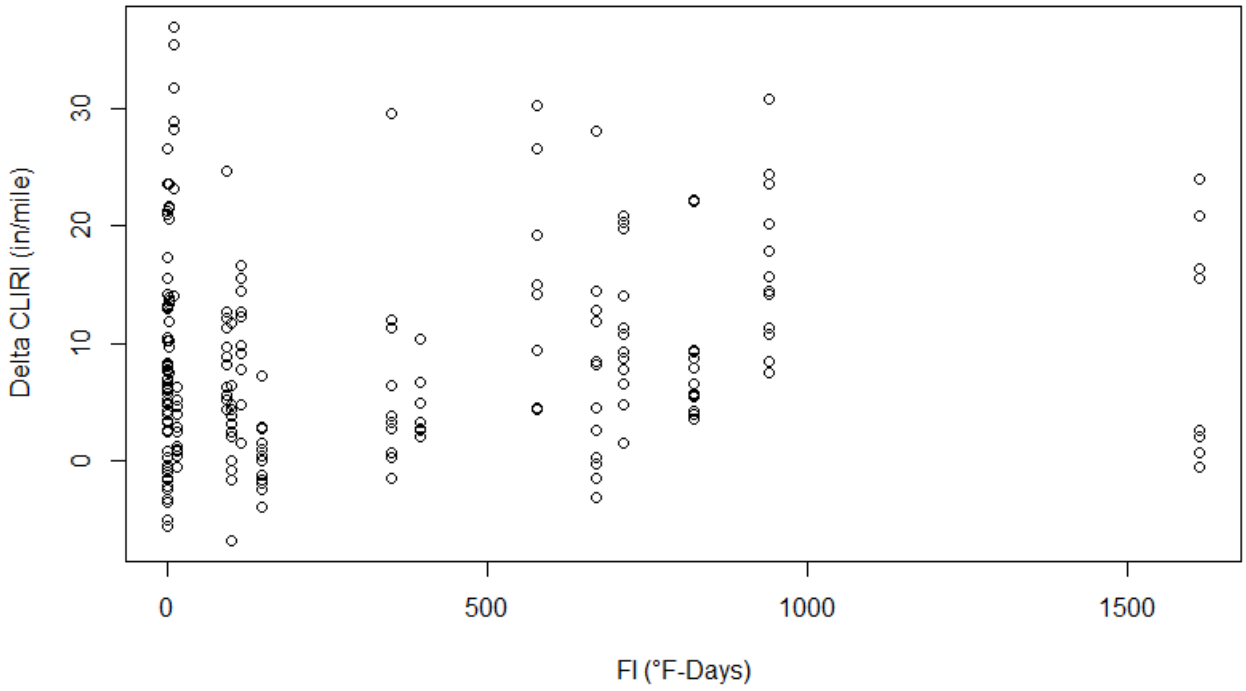

Figure D.33 – Delta *CLIRI* and *FI* values histogram plot for all subgrade pavements

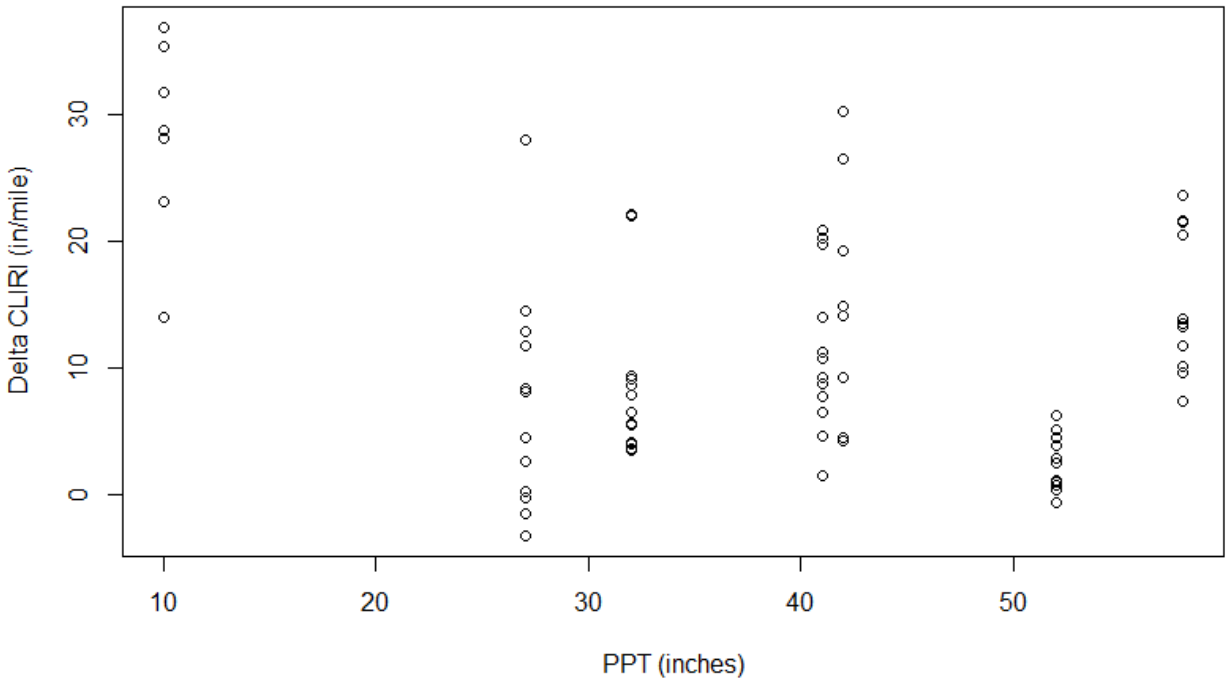

Figure D.34 – Delta *CLIRI* and *PPT* values scatter plot for fine subgrade pavements

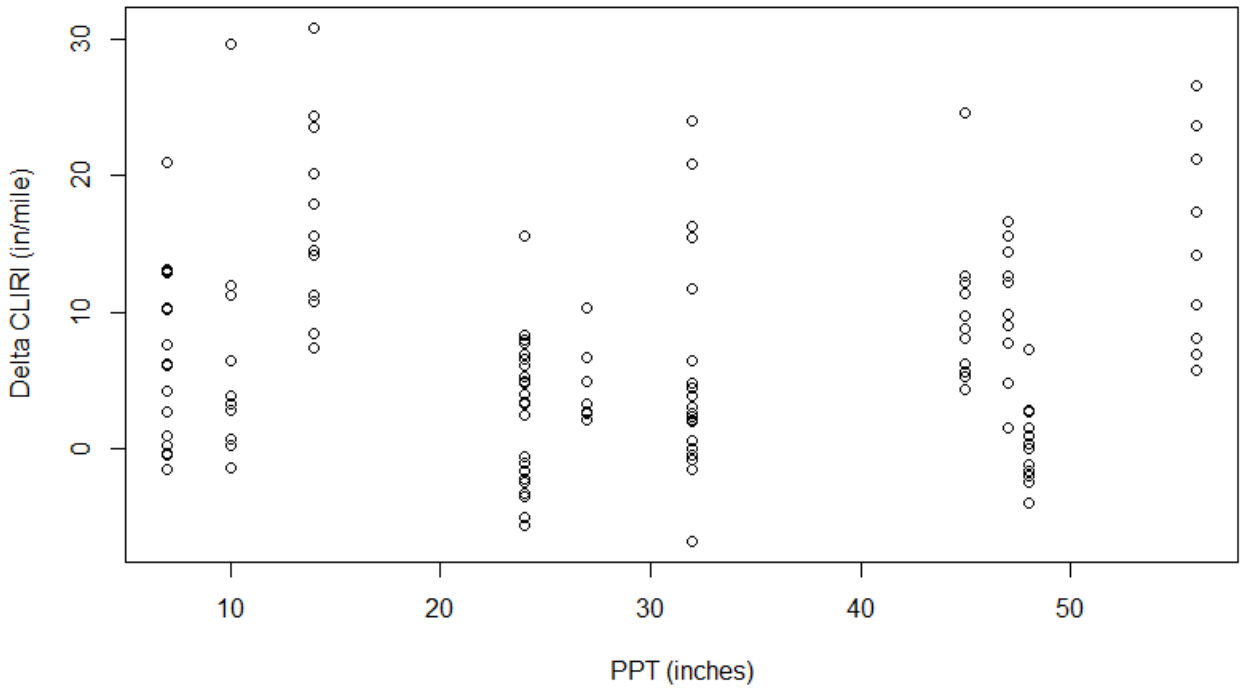

Figure D.35 – Delta *CLIRI* and *PPT* values histogram plot for coarse subgrade pavements

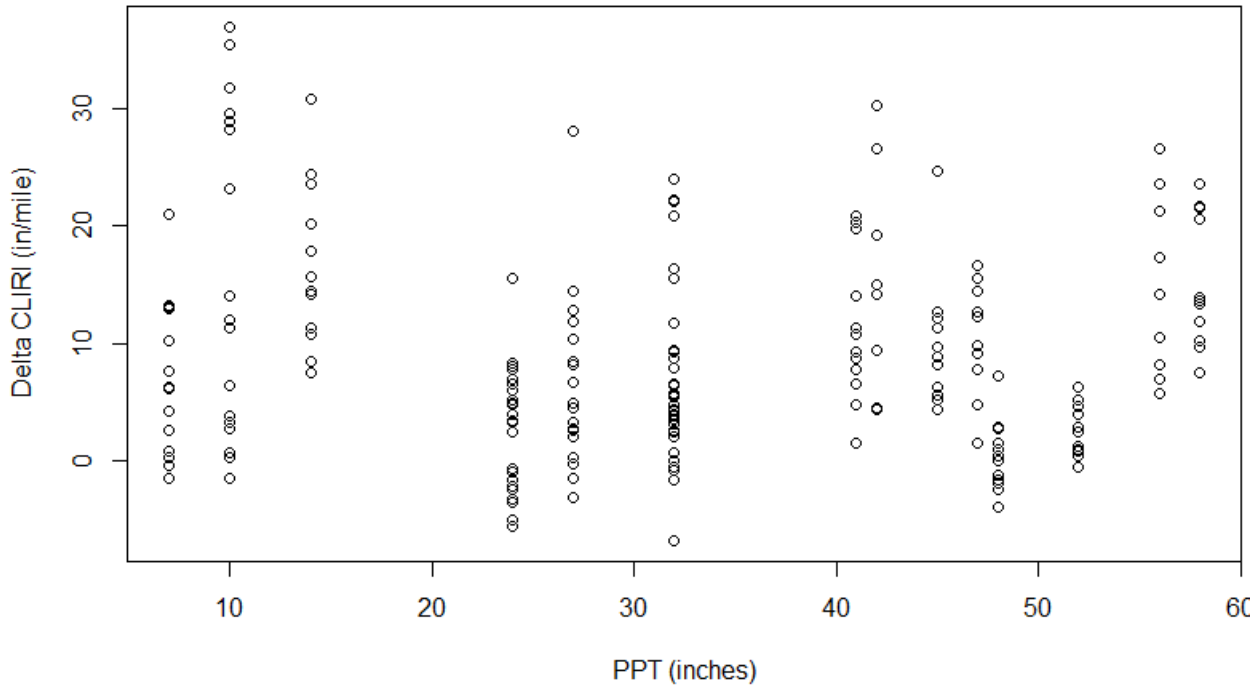

Figure D.36 – Delta *CLIRI* and *PPT* values histogram plot for all subgrade pavements

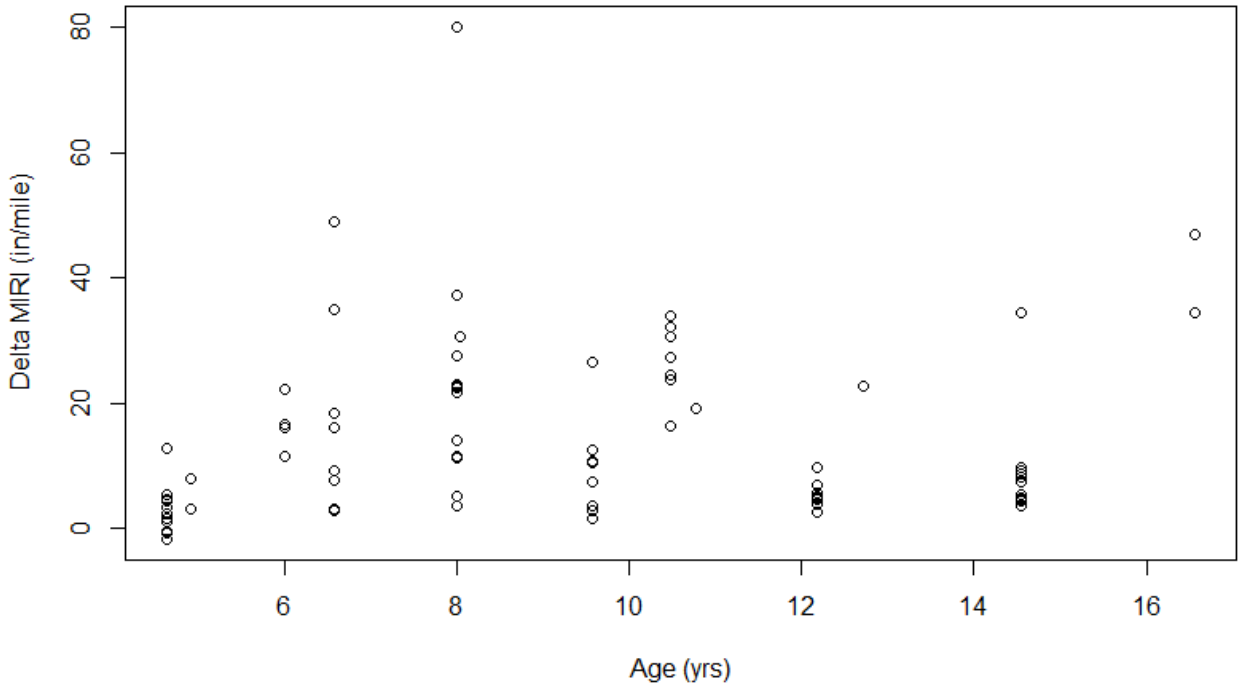

Figure D.37 – Delta *MIRI* and *Age* values scatter plot for fine subgrade pavements

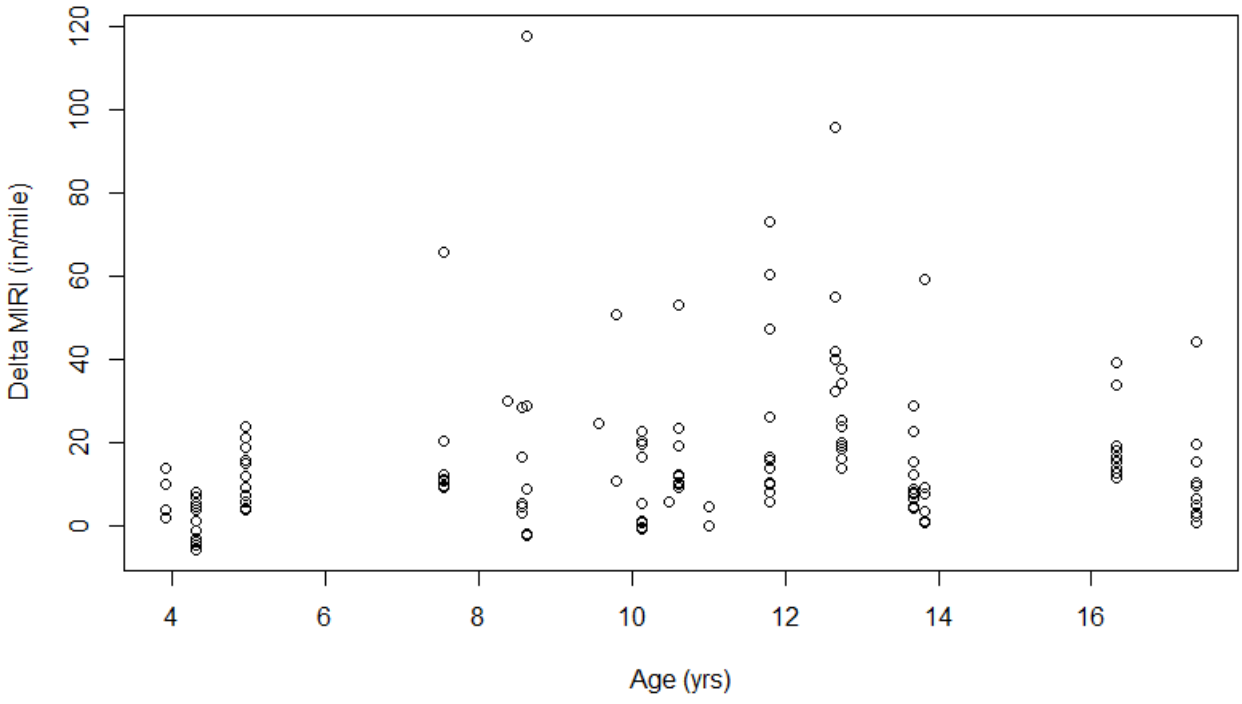

Figure D.38 – Delta *MIRI* and *Age* values histogram plot for coarse subgrade pavements

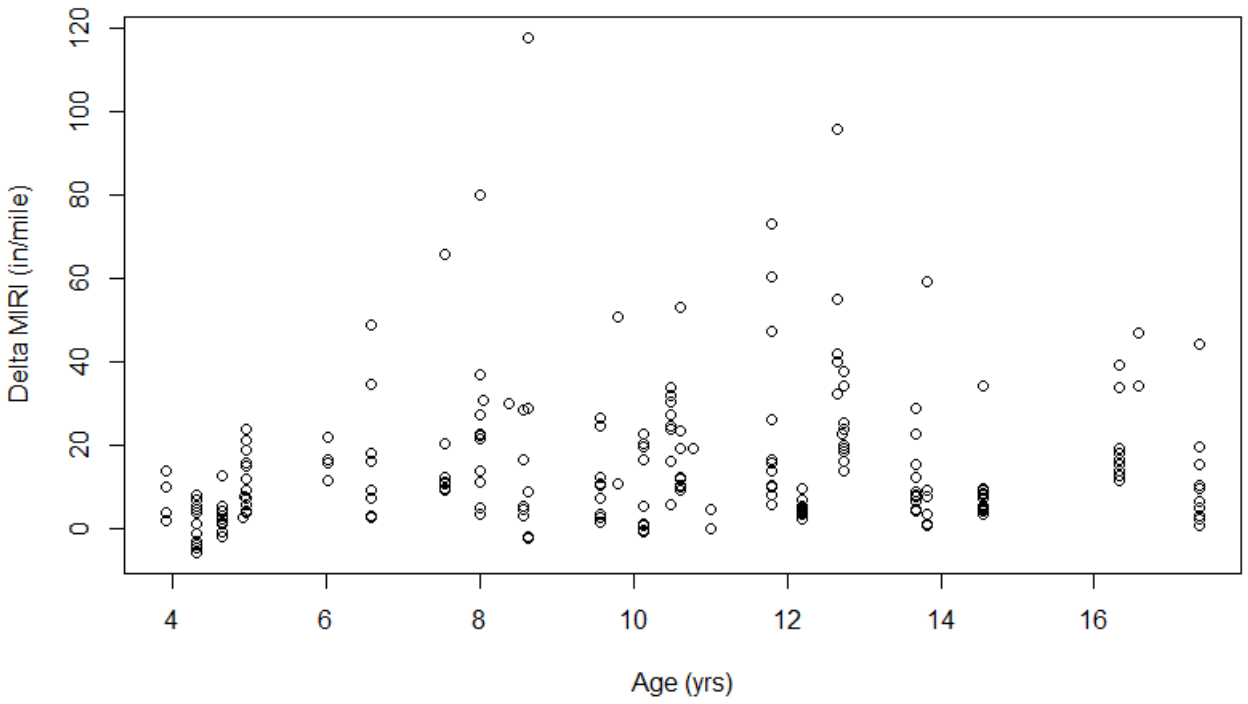

Figure D.39 – Delta *MIRI* and *Age* values histogram plot for all subgrade pavements

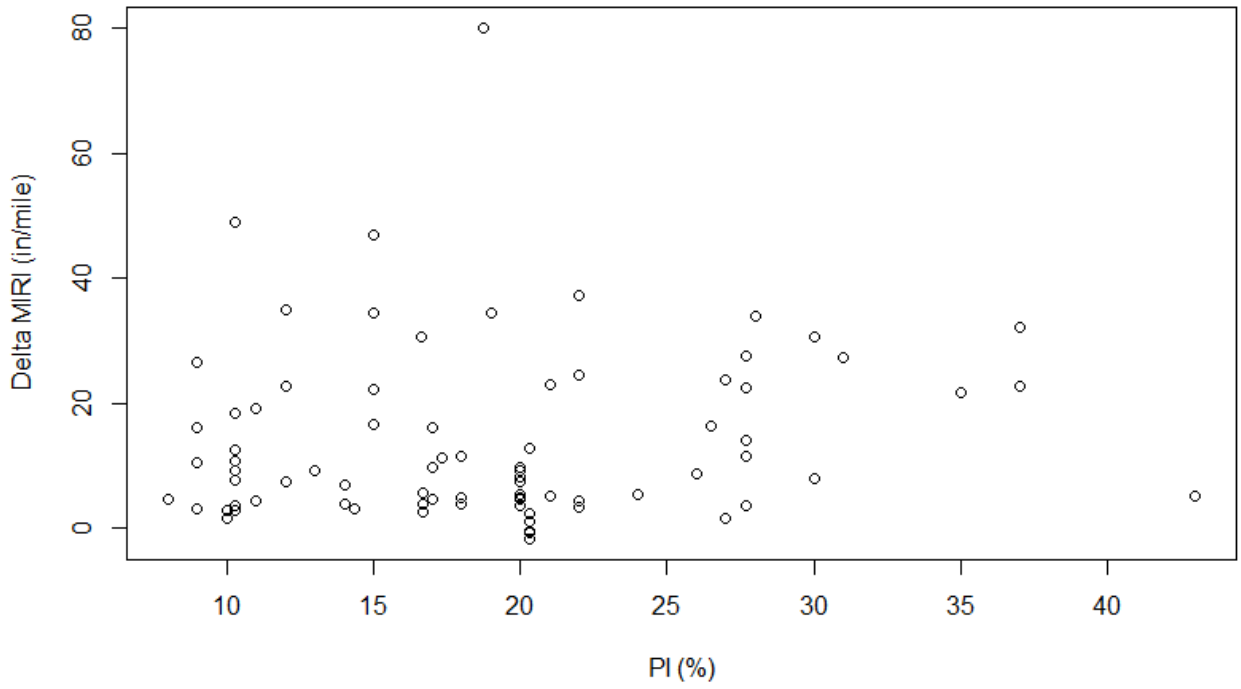

Figure D.40 – Delta *MIRI* and *PI* values scatter plot for fine subgrade pavements

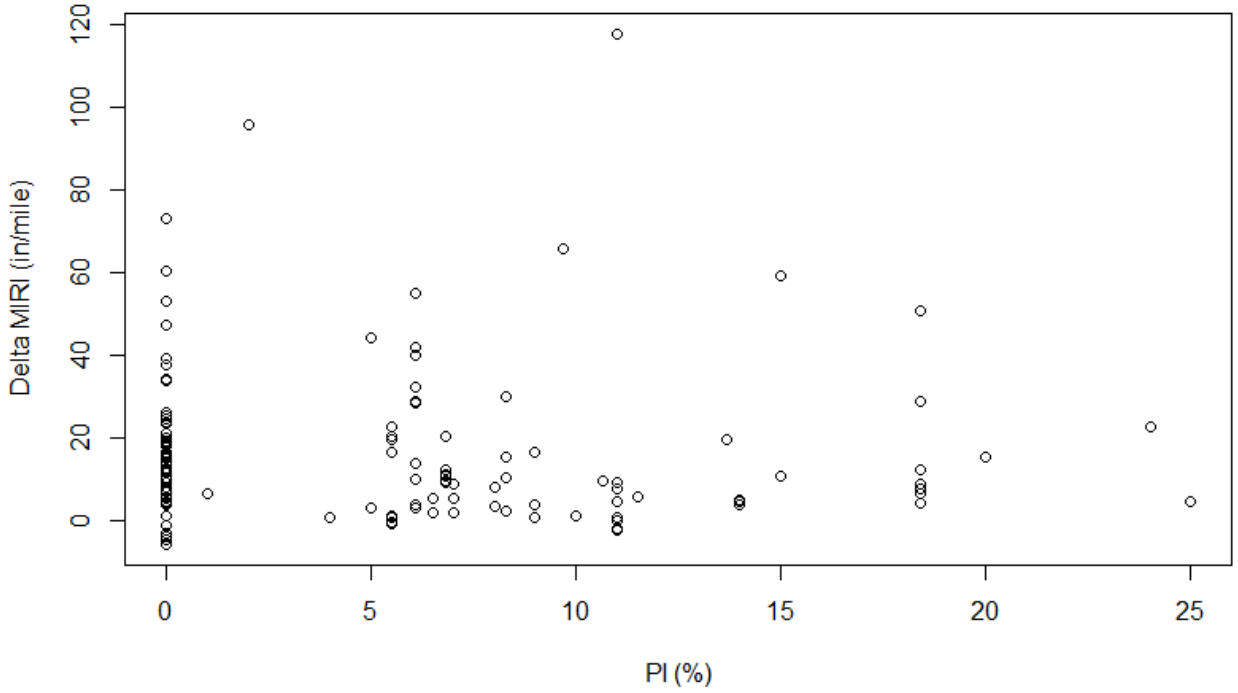

Figure D.41 – Delta *MIRI* and *PI* values histogram plot for coarse subgrade pavements

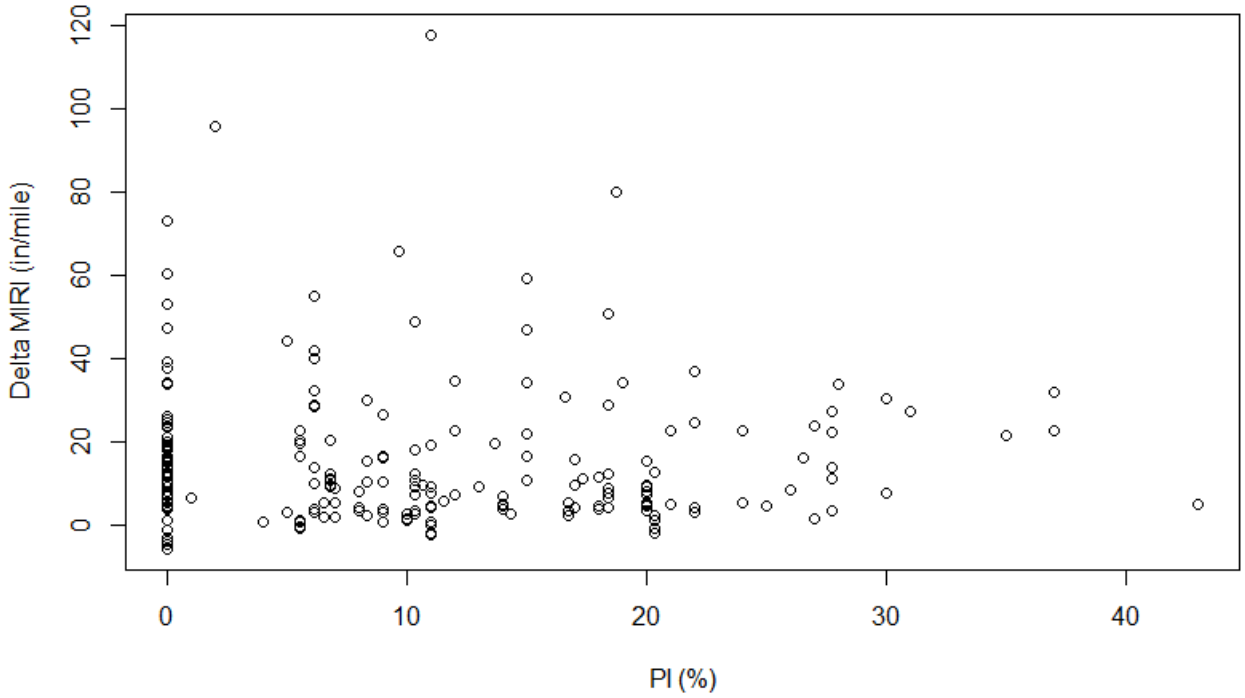

Figure D.42 – Delta *MIRI* and *PI* values histogram plot for all subgrade pavements

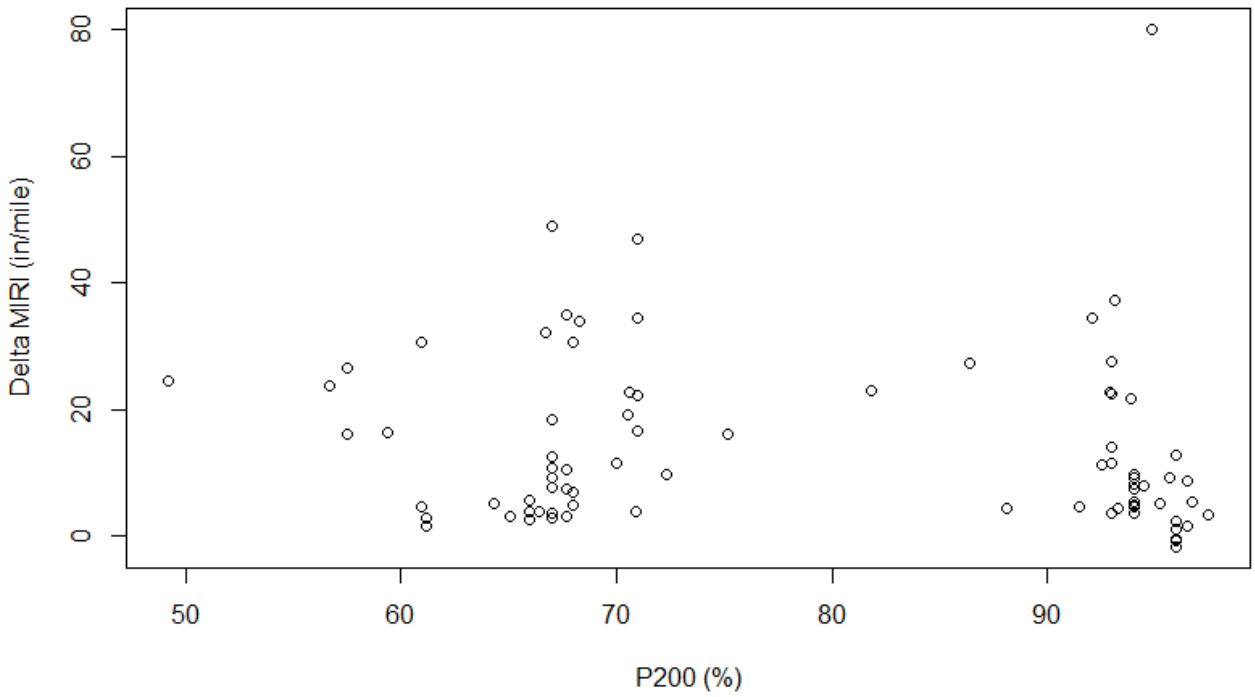

Figure D.43 – Delta *MIRI* and P_{200} values scatter plot for fine subgrade pavements

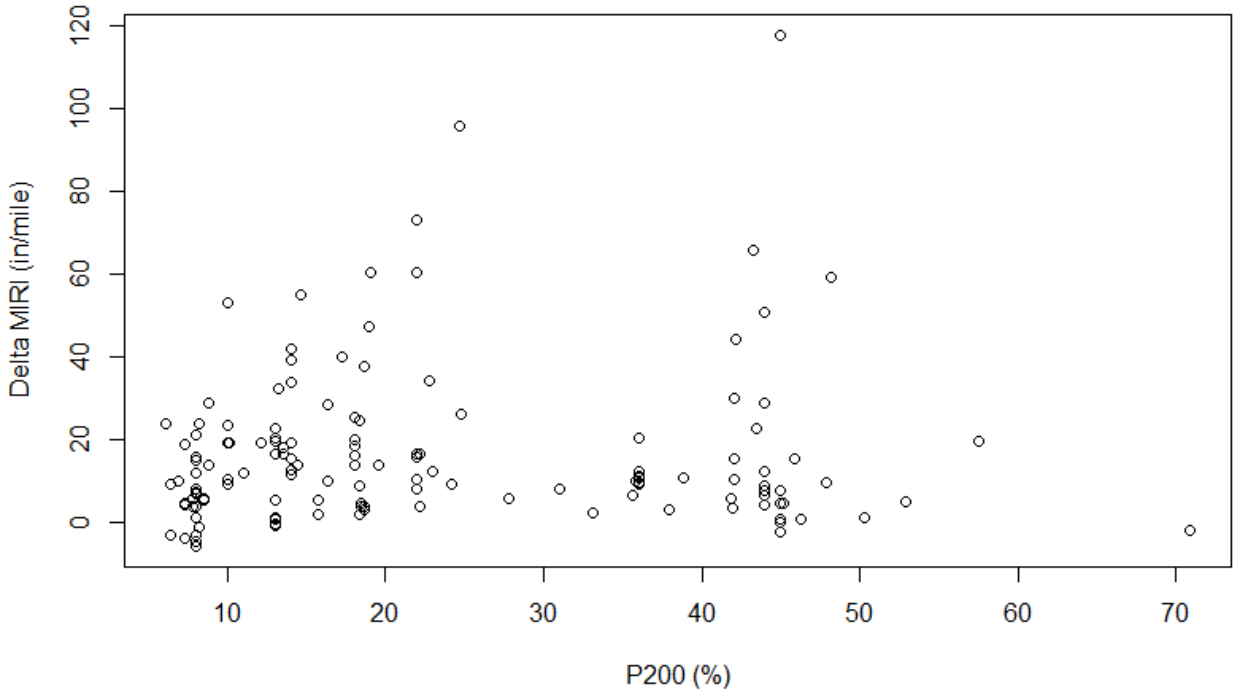

Figure D.44 – Delta *MIRI* and P_{200} values histogram plot for coarse subgrade pavements

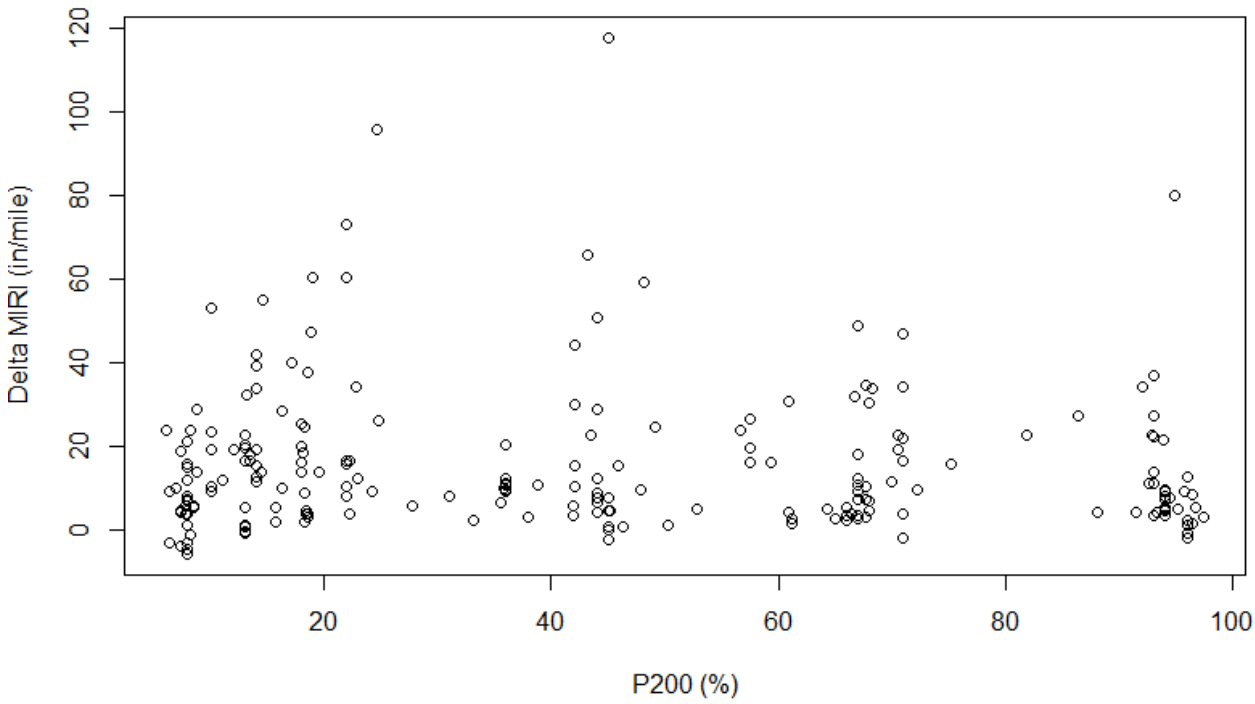

Figure D.45 – Delta *MIRI* and P_{200} values histogram plot for all subgrade pavements

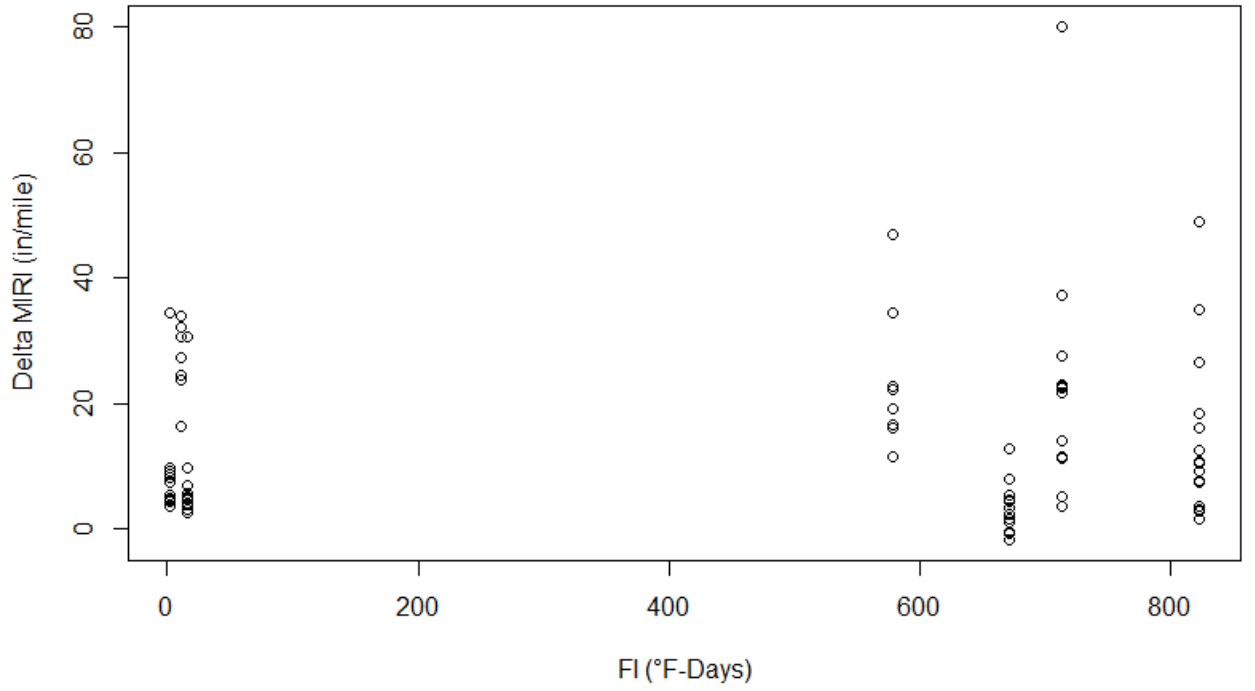

Figure D.46 – Delta *MIRI* and *FI* values scatter plot for fine subgrade pavements

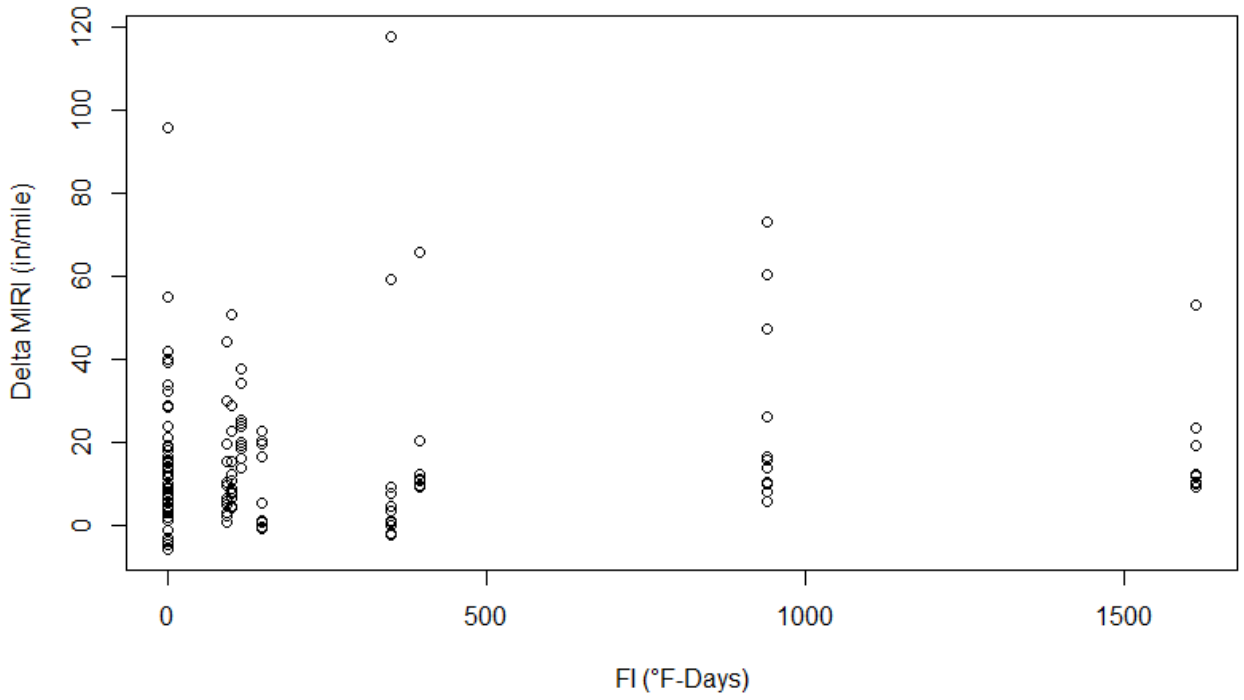

Figure D.47 – Delta *MIRI* and *FI* values histogram plot for coarse subgrade pavements

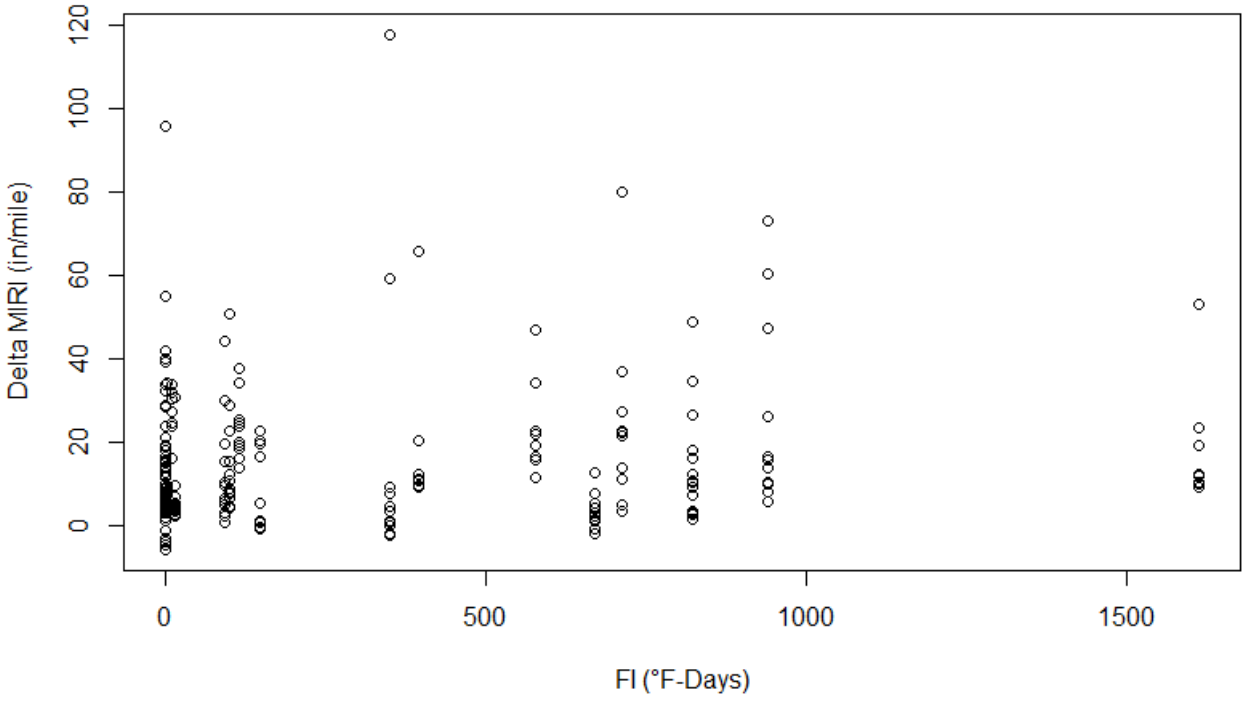

Figure D.48 – Delta *MIRI* and *FI* values histogram plot for all subgrade pavements

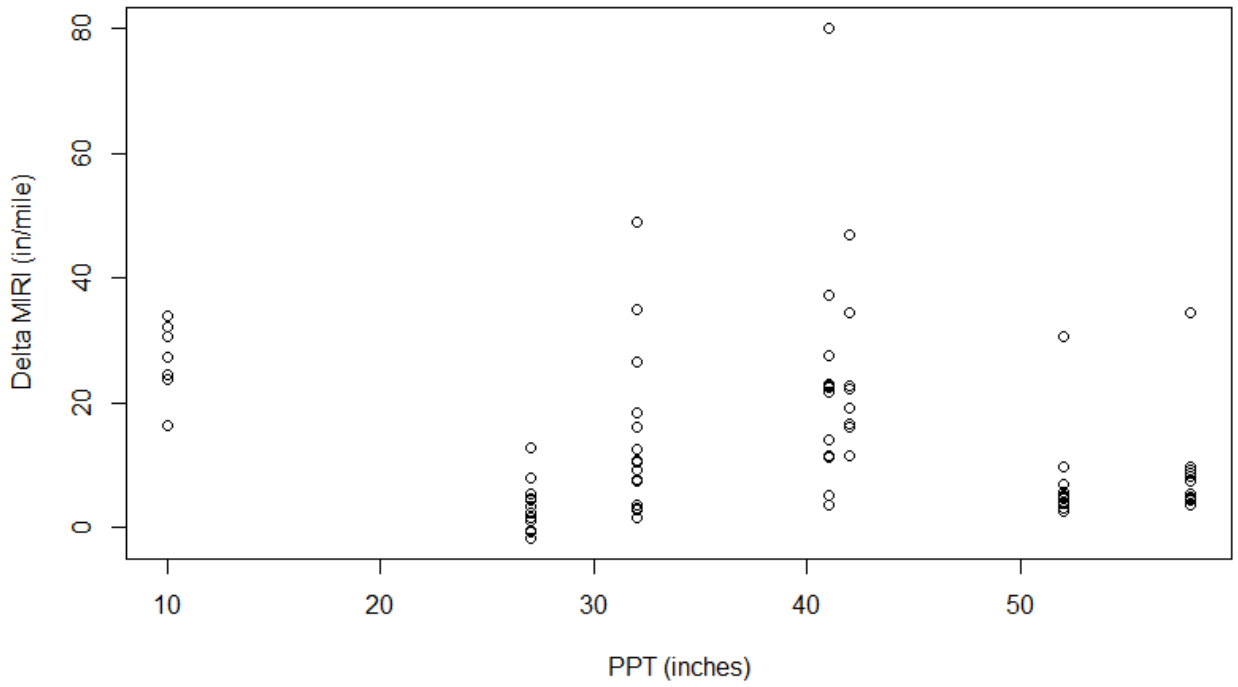

Figure D.49 – Delta *MIRI* and *PPT* values scatter plot for fine subgrade pavements

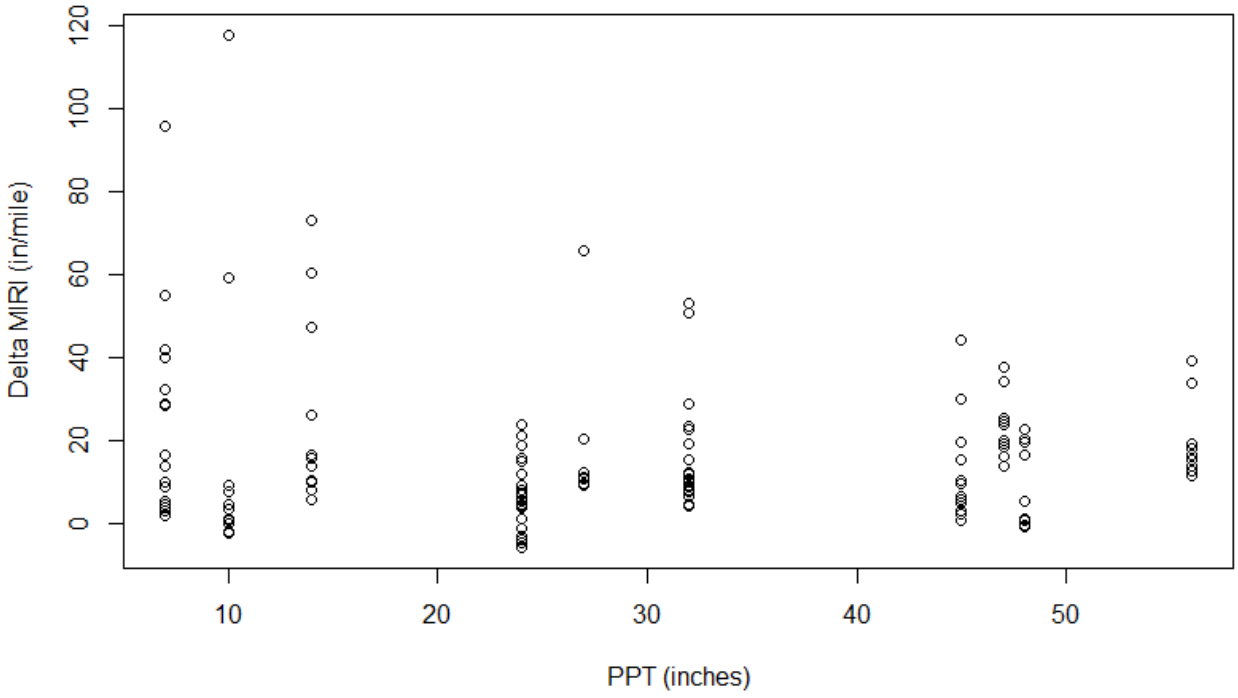

Figure D.50 – Delta *MIRI* and *PPT* values histogram plot for coarse subgrade pavements

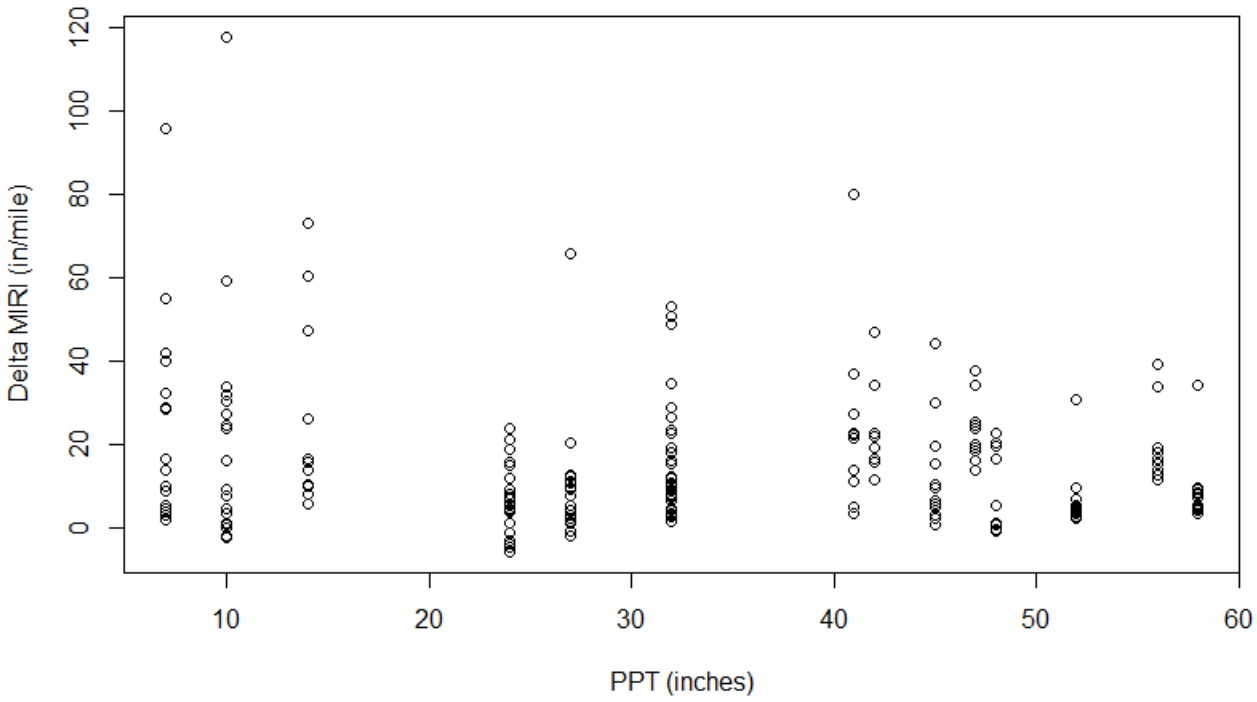

Figure D.51 – Delta *MIRI* and *PPT* values histogram plot for all subgrade pavements