Deterioration Models for Flexible Pavements in the Pacific Northwest Using the Long-Term Pavement Performance (LTPP) Database

Tariq Lamei, S.M.ASCE
Department of Civil and Environmental Engineering, University of Idaho
875 Perimeter Drive MS | Moscow ID 83844-0901
Email: lame7244@vandals.uidaho.edu

Emad Kassem, Ph.D., P.E.
Department of Civil and Environmental Engineering, University of Idaho
875 Perimeter Drive MS | Moscow ID 83844-0901
Email: ekassem@uidaho.edu

Word Count: 6,973 words + 0 table (250 words per table) = 6,973 words

Submitted [July 31st, 2024]

ABSTRACT

Prediction of pavement performance deterioration is crucial for effective transportation infrastructure management and maintenance planning. This study focuses on developing deterioration models for flexible pavements in the Pacific Northwest, leveraging the extensive data available in the InfoPave Long-Term Pavement Performance (LTPP) tool. Specifically, historical records of International Roughness Index (IRI) measurements are analyzed to understand the performance degradation patterns of flexible pavements over time. Utilizing Kaplan-Meier (K-M) survival time analysis (reliability analysis), we investigate the survival probabilities of flexible pavements based on IRI measures identified by the federal measures over varying time intervals. By analyzing these survival curves, we aim to derive deterioration models that accurately estimate the percentage of flexible pavements surviving within specified IRI thresholds over time and identify trends in deterioration patterns affected by parameters such as maintenance treatments, average annual daily truck traffic, number of freeze and thaw cycles, average temperature, and the presence of cracking. The developed models provide valuable understanding into the deterioration trends of the Pacific Northwest's flexible pavements and offer predictive tools for infrastructure asset managers to optimize maintenance strategies and allocate resources efficiently on the network level. This research contributes to the advancement of pavement management practices by utilizing robust statistical methods to analyze large-scale longitudinal flexible pavement performance data to enhance the resilience and durability of transportation infrastructure.

Keywords: International Roughness Index, Deterioration Models, Flexible Pavement, Infrastructure Resilience, LTPP.

INTRODUCTION

Approximately 94% of the paved roads in the United States are surfaced with asphalt, underscoring the critical role these structures play in national infrastructure and daily life. The integrity and functionality of pavements are paramount not only for efficient transportation systems but also for road safety. Each year, the American Society of Civil Engineers (ASCE) evaluates the state of U.S. infrastructure, revealing significant concerns related to road conditions. For instance, the 2009 report highlighted that out of 41,059 motor vehicle fatalities, approximately 14,000 were attributable to poor road and highway conditions. Furthermore, these adverse road conditions also contributed to over 2.49 million injuries [1]. Studies further indicate that 15% to 18% of traffic accidents occur on wet pavements, which are often exacerbated by poor pavement conditions.

The importance of robust pavement management systems is increasingly recognized at the national level. The Moving Ahead for Progress in the 21st Century (MAP-21) federal transportation legislation mandates U.S. state highway agencies to develop performance-based approaches in their pavement management decision-making processes [2]. This legislative framework supports the need for continuous improvements in pavement evaluation and maintenance strategies, emphasizing the significance of long-term data collection and analysis for informed decision-making. In the broader economic context, road transportation is integral to the functioning of societies, influencing between 10% and 20% of the Gross Domestic Product (GDP). Roads facilitate personal mobility, including access to services, goods, and leisure, making the reliability of these transport pathways crucial for the economic stability and growth of nations [3]. Given these stakes, it becomes imperative to invest in research that seeks to extend the life of pavements and enhance their safety features through improved performance deterioration models and maintenance practices.

The Long-Term Pavement Performance (LTPP) program [4], initiated by the Federal Highway Administration (FHWA), has provided an invaluable database for pavement research, enabling the development and validation of numerous predictive models across various studies. This comprehensive collection of data spans hundreds of pavement sections and includes detailed information on pavement structure, performance, and environmental conditions, making it a cornerstone for advanced pavement deterioration and performance modeling. Recent studies illustrate the diversity and depth of analysis possible with LTPP data. For example, Jia et al. [5] utilized the LTPP data to develop a novel framework for characterizing the transverse profile shapes of rutting on asphalt pavements. Their approach, which integrates data acquisition, profile adjustment, categorization, and indicator analysis, demonstrates the utility of LTPP data in enhancing our understanding of specific distress mechanisms such as rutting. Similarly, Younos et al. [6] employed data from 89 LTPP sections to develop multi-input performance prediction models based on the Pavement Condition Index (PCI), highlighting the dataset's role in validating comprehensive pavement condition assessments. Further extending the application of LTPP data, Karam and Noorvand [7] focused on modeling the rutting potential in asphalt concrete subjected to repeated traffic loads, particularly in high-temperature environments. Their predictive model incorporates materials properties, traffic, and climatic data from the LTPP InfoPave database, showcasing the versatility of LTPP data in addressing region-specific pavement issues.

In addition to predictive modeling, Luo et al. [8] explored the effectiveness of preventive maintenance (PM) treatments using Specific Pavement Studies (SPS-3) data from the LTPP program. Their study evaluates the short-term impacts of various PM treatments on pavement performance, thereby contributing to more effective maintenance strategies. Moreover, the integration of advanced analytical techniques with LTPP data continues to expand. Li et al. [9] demonstrated the use of machine learning (ML) methods to predict the modulus of asphalt concrete layers, employing LTPP data to train and test their models. This approach not only enhances the accuracy of mechanical empirical rehabilitation analysis but also provides tools for managing pavement networks without extensive direct testing data. The research by Kaloop et al. [10] underscores the predictive capabilities enabled by LTPP data through the development of an IRI prediction model using Gaussian Process Regression and Locally Weighted Polynomials. Their work emphasizes the role of LTPP data in refining performance indicators critical to pavement management. Zhang and Wang [11] and Khan et al. [12] have further broadened the scope of LTPP data utilization by incorporating geospatial and decision tree analyses to explore the impacts of environmental factors and traffic loads on pavement performance, and by employing experimental design techniques to analyze data-driven pavement design strategies. Through these diverse applications, the LTPP data continues to be a pivotal resource in advancing the field of pavement engineering, providing a robust foundation for developing reliable, region-specific deterioration models that address the unique challenges faced by pavement infrastructures across different geographic and climatic conditions. This research builds on previous research to improve the capabilities of pavement deterioration modeling specifically in the Pacific Northwest by leveraging the LTPP dataset in conjunction with innovative analytical methodologies.

Literature Review

The development of pavement deterioration prediction models is a crucial aspect of pavement management, enabling engineers and policymakers to make informed decisions regarding maintenance, rehabilitation, and overall infrastructure planning. A rich history of research utilizing various methodologies showcases the advancements and challenges in this field. One of the key performance indicators in pavement management is the IRI, which serves as a primary metric for evaluating pavement condition and ride quality. Kaloop et al. [10] developed a sophisticated IRI prediction model using Gaussian Process Regression and Locally Weighted Polynomials. This model utilized multiple variables from the LTPP data, including pavement age, initial IRI, and various distress indicators, demonstrating the complexity of factors influencing IRI. Beyond IRI, the Pavement Condition Index (PCI) has also been a focal point of study. Younos et al. [6] proposed multi-input performance prediction models based on PCI, using observations from specific pavement studies within the LTPP database. This approach highlights the utility of comprehensive datasets in developing reliable predictive models without maintenance influences. Justo-Silva et al. [3] provided a broader perspective by reviewing various modeling techniques used in pavement performance prediction. They emphasized the stochastic nature of pavement deterioration and the need for complex deterministic or probabilistic models, offering guidelines for future model development.

The integration of ML techniques has brought a new dimension to pavement deterioration modeling. Baykal et al. [13] utilized ensemble ML algorithms, including decision trees and boosting methods, to estimate IRI based on age, traffic loads, and structural parameters. Similarly, Liu et al. [14] applied ML to improve asphalt mix design, thereby enhancing pavement performance by predicting IRI values from the LTPP data. Innovative approaches have also been employed to overcome data variability challenges. Piryonesi and El-Diraby [15] segmented LTPP data into smaller, more cohesive groups based on location and functional class, which significantly improved model accuracy. This method underscores the importance of considering geographical and functional diversities in pavement analysis. Furthermore, the use of gene expression programming (GEP) by Imam et al. [16] to predict PCI from IRI exemplifies the novel application of genetic algorithms in pavement modeling, providing a fresh perspective on the relationship between different performance metrics. Also, the work by Tamagusko and Ferreira [17] systematically reviews ML models for predicting IRI, highlighting the transition towards more data-driven, accurate, and efficient modeling approaches in pavement management. Their synthesis of current methodologies provides valuable insights into the capabilities and future directions of pavement deterioration modeling.

Kaplan-Meier (K-M) survival analysis, traditionally used in medical research to estimate the survival time of patients, has found innovative applications in the field of pavement engineering. This nonparametric method provides a powerful tool for analyzing the survival probabilities of pavement systems over time, enabling researchers to better understand how various factors influence pavement longevity and performance [18]. In the realm of pavement research, K-M survival analysis is particularly useful for evaluating how long pavement remains in acceptable condition before reaching failure or requiring significant maintenance interventions. This approach allows for the estimation of median survival times and the development of survival probability curves for various pavement distresses. A notable application of this method was presented by Hatoum et al. [19], who utilized K-M survival analysis to study the performance of flexible pavements. Their research focused on assessing the impacts of using reclaimed asphalt pavement (RAP) versus virgin materials on the fatigue life of asphalt pavements. By examining test sections extracted from the LTPP program, the study developed separate fatigue survival curves for pavements constructed with RAP and those with virgin materials. This analysis not only tested the equality of survival probabilities between the two materials but also highlighted the significant effect of vehicle overloading on pavement deterioration. Moreover, the study explored the influence of combined factors such as environmental conditions, pavement materials, and traffic loads on pavement performance. Through nonparametric and parametric survival analyses, Hatoum et al. [19] identified the most significant subset of risk factors (covariates) affecting pavement longevity. This involved developing K-M survival probability curves for multiple pavement distresses, comparing their failure probabilities, and determining the median survival time for each type of distress.

The utilization of K-M survival analysis in this context demonstrates its versatility and value in pavement engineering. It provides a robust framework for analyzing time-to-event data, offering a clearer understanding of the durability and performance of pavement systems under various stressors. This approach not only aids in the predictive modeling of pavement deterioration but also supports the development of more sustainable, cost-effective, and environmentally friendly solutions in road construction. By integrating K-M survival analysis into pavement research, this study enhances the methodological toolkit available to engineers and researchers, enabling more precise predictions of pavement life and more informed decision-making in pavement management. This research utilizes these techniques to further investigate the survival probabilities of flexible pavements in the Pacific Northwest, aiming to develop detailed models that reflect the unique environmental and operational conditions of the region.

The study of pavement performance and deterioration modeling is enriched by a diverse array of research contributions that collectively advance our understanding and management of pavement systems. Among the significant works in this field, the research by Medina et al. [20] stands out for its innovative approach to correlating the IRI with the PCI. Their study established a relationship between these two critical indices using a concept analogous to the time-temperature superposition principle known in materials science, which they adapted as the time-deterioration superposition principle for pavements. Medina et al. [20] focused on the practical aspect of pavement management, particularly under the constraints of budget and resource availability that vary significantly across different jurisdictions. Recognizing that IRI data collection is more affordable and frequent compared to detailed pavement distress surveys, their study aimed to leverage the more readily available IRI data to infer PCI. This approach is particularly beneficial for agencies with limited resources, enabling them to maintain effective pavement management practices without the extensive data collection costs associated with comprehensive distress surveys. The analysis used data from the LTPP InfoPave database, focusing on flexible pavements within network samples from Arizona, California, and Wisconsin. Their findings revealed a robust correlation between the changes in IRI and PCI over time, with coefficients of determination ranging from 0.71 to 0.85. These results not only demonstrate the efficacy of the proposed modeling approach but also establish a set of general threshold limits for IRI that can serve as benchmarks for assessing pavement condition across various networks.

This research is relevant to the current study as it illustrates a successful application of deterioration modeling techniques that combine multiple performance indices, providing a comprehensive view of pavement condition. Furthermore, the use of a time-deterioration superposition principle suggests a novel methodology that can be adapted and expanded upon in other contexts, including the Pacific Northwest, where specific regional conditions such as climate and traffic patterns may influence pavement performance differently. This study aims to contribute to this vital area by developing reliable models for predicting the deterioration of flexible pavements in the Pacific Northwest. Utilizing the LTPP data, this research seeks to innovate in the modeling of pavement degradation, addressing unique regional challenges such as varying climate conditions and traffic patterns. By enhancing our understanding of pavement life cycles and deterioration processes, the findings from this study are expected to inform and refine pavement management strategies, leading to safer, more durable roads.

METHODS

The data used in this study were obtained from the InfoPave LTPP tool, a comprehensive database that offers a wealth of pavement performance data across various geographical regions in the US and Canada. Our focus was on flexible pavement sections in the Pacific Northwest, covering the states of Wyoming, Oregon, Idaho, Montana, and Washington. To narrow down our scope, we applied a filter for flexible pavement surface types within the Pacific Northwest states, resulting in the selection of 137 sections distributed as follows: 23 sections in Wyoming, 15 sections in Oregon, 24 sections in Idaho, 38 sections in Montana, and 37 sections in Washington. From the Pavement Structure and Construction tab, we obtained records of asphalt concrete (AC) surface treatments for the selected sections. The Performance tab provided us with data on AC distresses and Analysis Ready IRI datasets, which are key indicators of pavement condition and degradation over time. Traffic data, including truck volumes and Traffic Summary Statistics, were sourced from the Traffic tab. This data offers insights into the levels of truck traffic volumes affecting the pavements in the study area. Additionally, temperature data was collected from the Climate tab, which helps us understand the environmental factors influencing pavement performance with time.

Our study focused on data recorded over time, specifically examining IRI data for 129 out of the total 137 sections from the years 1989 to 2021. It is important to note that no IRI records were found for any section in the year 2006. This longitudinal data allowed us to track changes in pavement condition and develop models of deterioration trends. By utilizing this comprehensive dataset, we were able to perform a detailed analysis of flexible pavement performance in the Pacific Northwest region, contributing valuable insights into the long-term durability and resilience of transportation infrastructure in the region.

The data preparation process began with the identification of the number of record years for each pavement section. Sections with fewer than five years of recorded data were filtered out, resulting in a subset of 97 sections suitable for further analysis. This approach ensured the inclusion of sections with sufficient historical data to observe trends over time. Once the dataset was filtered, the remaining data were sorted chronologically from oldest to newest records, ensuring consistency in the timeline of each section. We aligned the corresponding independent variables, which included Annual Average Daily Truck Traffic (AADTT), the number of freeze and thaw cycles, average temperature, percentage of alligator cracking, length of transverse cracking, and length of longitudinal cracking. The maintenance history of each section was obtained from the LTPP data field reference guide and categorized according to predefined categories. This categorization facilitated a more systematic analysis of the impact of maintenance treatments on pavement performance. During data preparation, we encountered instances where discontinuities occurred in the dataset. These were cases where specific independent variables were not recorded in the same year as the corresponding IRI values. Such inconsistencies were addressed by cleaning the data and removing records with incomplete information, ensuring a high-quality and consistent dataset for subsequent analysis. By following this data preparation process, we achieved a comprehensive, cleaned, and filtered dataset ready for use in our analysis. This approach allowed for accurate modeling of flexible pavement deterioration patterns while taking into account the various factors influencing pavement performance over time.

To investigate the deterioration of flexible pavements over time, we adopted K-M survival time analysis, also known in engineering as reliability analysis or time to failure analysis. This statistical approach examines the incidence and timing of events, specifically the changes in IRI measures identified by federal standards over time. The K-M method is a nonparametric maximum likelihood estimator, commonly employed to estimate survival probabilities of pavement sections transitioning between different condition ratings. This method is particularly suitable for discontinuous data, as it can handle variability in the intervals between recorded observations and the occurrence of events. The K-M methodology calculates survival probabilities over time and provides insights into the expected lifespan of flexible pavements under different conditions. The analysis produces survival curves, which show the proportion of pavement sections that remain within specified IRI thresholds over time. These curves allow us to assess the rate of deterioration and identify key factors influencing pavement performance. Reliability data can be calculated using the K-M estimator by Equation 1.

$$\hat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right) for\, t_1 \leq\ t \leq\ t_k \tag{1}$$

where

$\hat{S}(t)$ is the K-M estimator,

d_j is the number of sections for which the event occurred (transitioned to the lower condition) at time t_j,

n_j is the number of sections at risk of event at time t_j, and

t_1 and t_k are the boundary for k distinct event times

One limitation of the K-M method is the reduction in the quantity of available data when time-blocking is used for analysis. Time-blocking, which involves dividing data into intervals based on specific factors (e.g., era of construction), can lead to a loss of information and reduced sample sizes. Despite this drawback, the K-M estimator remains a powerful tool for analyzing time-to-event data and uncovering trends in flexible pavement performance.

We developed two K-M models to investigate the survivability of flexible pavement sections. The first K-M model was designed to study the survival probability of flexible pavement sections classified as being in good condition according to the federal standards, which is defined as having an IRI value of less than 95 in/mi. This model provided insights into the expected lifespan of pavements that maintained a high level of quality. The second K-M model was performed to investigate the survival probability of flexible pavement sections classified as being in good or fair condition according to the federal standards, which is defined as having an IRI value of less than 170 in/mi. This broader range allowed us to explore a wider spectrum of pavement conditions and how they influence longevity. In both models, we considered the maintenance history of each section to determine the potential impact of surface treatments on the rate of IRI deterioration over time. By examining how maintenance treatments affected the slope of the K-M curves, we were able to identify potential points of change where specific treatments may have helped to slow down the deterioration process. Overall, K-M survival time analysis provides a robust framework for assessing the reliability and deterioration patterns of flexible pavements in the Pacific Northwest. By leveraging this approach, we can derive meaningful insights into pavement lifespan and develop predictive models to support infrastructure management and maintenance planning.

Multiple regression analysis is a statistical technique employed in this study to explore the relationship between the dependent variable and multiple independent variables. Specifically, it is used to analyze the relationship between the IRI, and various independent variables such as AADTT, the number of freeze-thaw cycles, average temperature, percentage of alligator cracking, length of transverse cracking, and length of longitudinal cracking. Forward stepwise regression, a subset of multiple regression, begins with an empty model and sequentially adds variables. At each step, the variable that most significantly enhances the model, typically measured by the increase in R-squared, is included. This process continues until no additional variables significantly contribute to the model's predictive power. The general equation for a multiple regression model is given in Equation 2, while Equation 3 was adopted in our study.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon \tag{2}$$

where

Y is the dependent variable,

$X_1, X_2, \dots, X_k$ are the independent variables,

β_0 is the intercept,
$\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients of the independent variables, and
ϵ is the error term, representing unexplained variation.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon \tag{3}$$

where
Y is the dependent variable (IRI value),
X_1: AADTT,
X_2: number of freeze and thaw cycles,
X_3: average temperature,
X_4: percent of alligator cracking,
X_5: length of transverse cracking,
X_6: length of longitudinal cracking,
β_0 is the intercept,
$\beta_1, \beta_2, \ldots, \beta_6$ are the coefficients of the independent variables, and
ϵ is the error term, representing unexplained variation.

Using forward stepwise regression facilitates the modeling process by incrementally selecting the most influential variables, providing a practical approach to identifying key factors affecting pavement roughness, especially when dealing with numerous potential predictors. This method is advantageous for ensuring that only statistically significant variables are retained in the final model, thereby enhancing model efficiency and interpretability. For the modeling and analysis, RStudio software was used. The models were run on a workstation equipped with an Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50 GHz processor and 16.0 GB of RAM, running Windows 10 Education. This setup provided the necessary computational power and environment to handle the data effectively and perform the analyses efficiently.

RESULTS

In this study, we developed two K-M models to investigate the survivability of flexible pavement sections in the Pacific Northwest, using different criteria for pavement condition based on IRI values.

Pavements in Good Condition (IRI < 95 in/mile)

The first model focused on the survival probability of flexible pavements considered to be in good condition (i.e., IRI value less than 95 in/mile). Figure 1 shows the K-M survival curve for this condition threshold demonstrates an initial high survival probability that gradually declines over time. The survival probability remains fairly stable in the early years, indicating good performance of the pavement sections initially. However, there is a noticeable change in the slope of the survival curve around the 5-year mark, where the survival probability begins to decrease at a faster rate, suggesting a significant deterioration in pavement quality over time. The smoothed survival curve, Figure 2, further highlights these changes, showing a more gradual but consistent decline in survival probability after approximately 5 years.

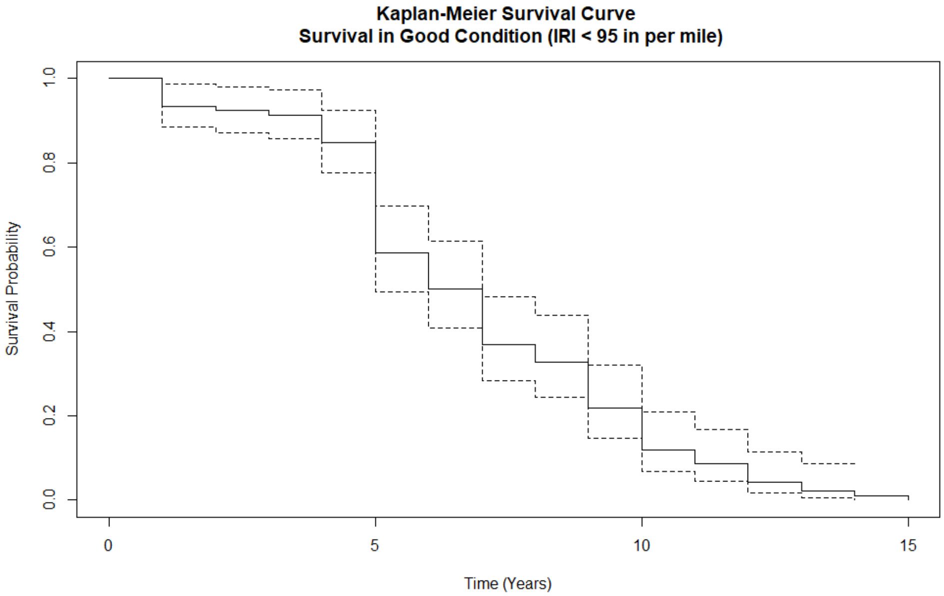

Figure 2: Survival probability in good condition (IRI < 95 in\mile)

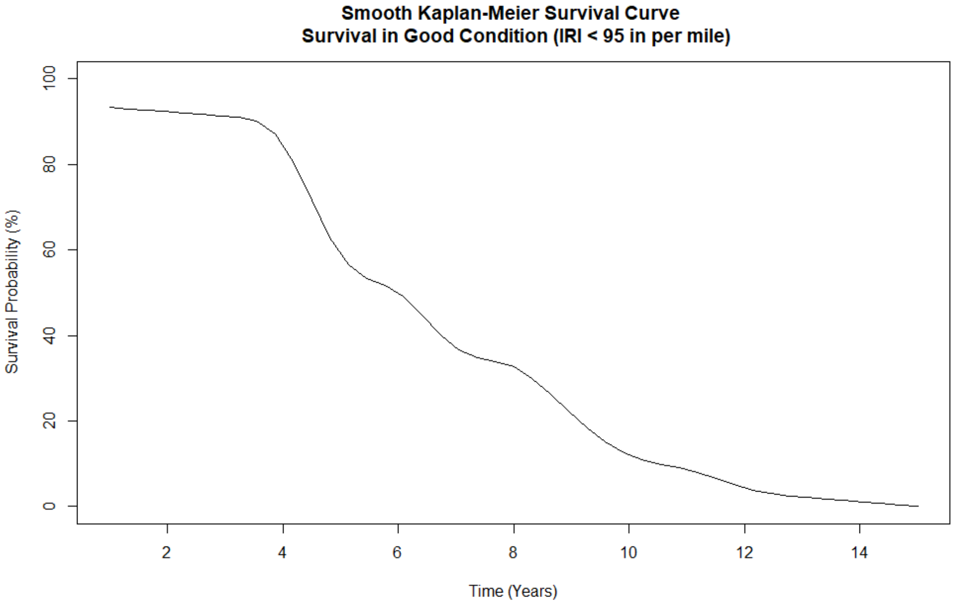

Figure 1: Smoothed survival probability in good condition (IRI < 95 in\mile)

Pavements in Good or Fair Condition (IRI < 170 in/mile)

The second model examined pavements categorized under good or fair condition (i.e., IRI value less than 170 in/mile). Figure 3 shows the survival curve, for this broader condition range, starts with a higher initial survival probability compared to the stricter good condition model. The curve exhibits a slower rate of decline in the initial years, but similar to the first model, a noticeable change in slope occurs around the 10-year mark. This indicates that while pavements may remain in fair condition for a longer period, they eventually begin to deteriorate more rapidly.

The smoothed survival curve, Figure 4, for this model emphasizes a more pronounced and steady decline after the initial stable period, which correlates with increased deterioration rates as the pavement ages.

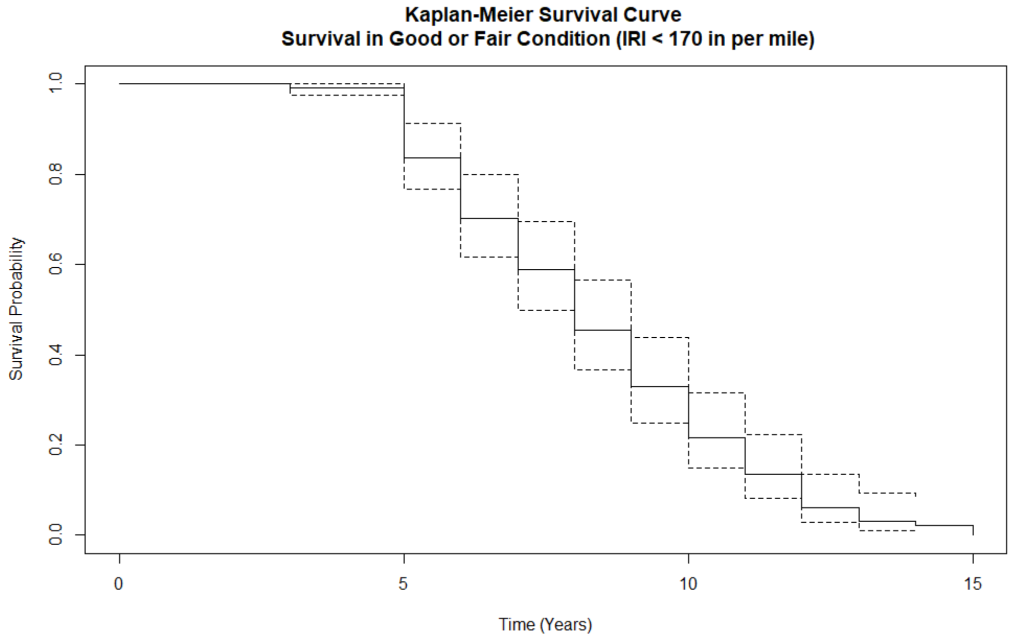

Figure 3: Survival probability in good or fair condition (IRI < 170 in\mile)

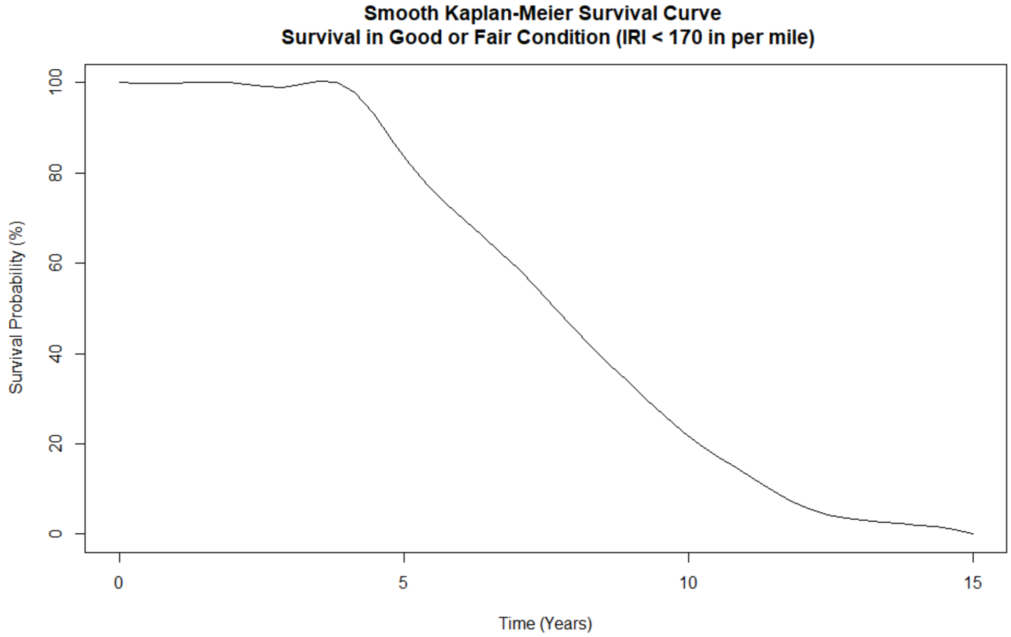

Figure 4: Smoothed survival probability in good or fair condition (IRI < 170 in\mile)

Analysis of Kaplan-Meier Survival Curves with Maintenance Interventions

Interestingly, at points where changes in the slope of the K-M curves are observed, corresponding maintenance history data indicates the application of surface treatments. These maintenance interventions appear to coincide with the changes in the rate of pavement deterioration, suggesting that timely maintenance could be effectively slowing down the deterioration process. Specifically, the points of slope change in both models seem to align with significant maintenance activities, which could be attributed to the retardation in the degradation of the IRI values over time. In the survival curve for pavement sections categorized as being in good condition, the treatments applied at the first 4 years of the pavement lifespan and at approximately 5, 7, 10, and 12 years appear to significantly alter the trajectory of the curve (Figure 5). Prior to the treatment at 5 years, the curve shows a severe decline, suggesting a natural progression of deterioration. After each treatment, however, the rate of decline in survival probability noticeably slows, indicating that the treatments are effective in extending the pavement's lifespan in good condition. This is particularly evident after the second and third treatments at around 7 and 10 years, where the slope of the curve flattens, suggesting a pronounced retardation in deterioration.

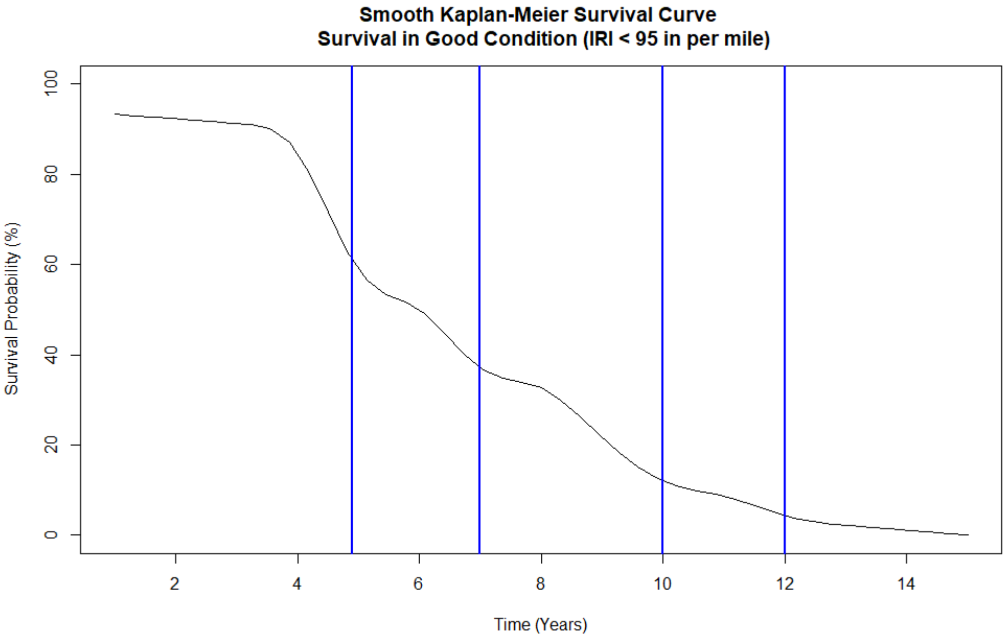

Figure 5: Maintenance interventions for pavements surviving in good conditions

For pavements maintained in good or fair condition, the treatments at the same points, the first 5 years of the pavement lifespan and at around 7, 10, and 12 years, also show a marked impact on the survival probabilities. The curve initially declines steadily, but each maintenance intervention correlates with a noticeable deceleration in the rate of deterioration (Figure 6).

These findings suggest that proactive and strategic maintenance plays a crucial role in extending the lifespan of pavement sections and maintaining them within acceptable IRI thresholds. The analysis demonstrates the effectiveness of surface treatments in impacting pavement survival rates, providing valuable insights for transportation infrastructure management and planning. These results are discussed further in the subsequent sections to explore their implications for pavement management practices and to compare them with existing literature on pavement deterioration modeling.

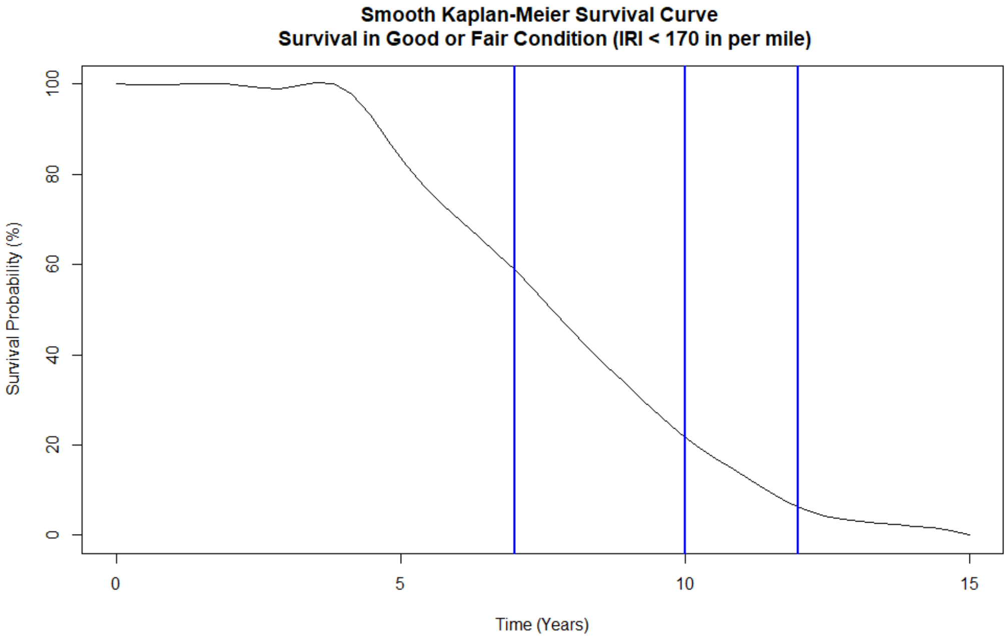

Figure 6: Maintenance interventions for pavements surviving in good or fair conditions

Regression Analysis

Our regression analysis aimed to explore the relationships between the IRI and a set of independent variables to identify key factors influencing pavement performance. We used a linear multiple regression model, as described in the methods section, and performed the analysis to select the most suitable model for the data. Initial exploration of the correlation matrix revealed varying levels of correlation between the independent variables and the response variable, IRI value, as shown in Figure 7. Specifically, average temperature (X_3), percentage of alligator cracking (X_4), and length of longitudinal cracking (X_5) demonstrated a moderate to strong correlation with IRI. In contrast, annual average daily truck traffic (X_1), the number of freeze and thaw cycles (X_2), and length of transverse cracking (X_6) showed lower correlation with IRI. This observation was consistent with the scatterplot matrix, which further corroborated the distinct relationships between IRI and specific independent variables.

Reducing the model by excluding variables with lower correlation could enhance its predictive capability, providing insights that can guide maintenance strategies and infrastructure planning. Performing the full model hypothesis test, we found that at least one factor significantly affected the IRI value of flexible pavement sections. Testing for collinearity indicated that collinearity was not a concern in our model. We continued with individual factor hypothesis testing, where the null hypothesis was that each factor is not significant. The tests revealed that annual average daily truck traffic (X_1) and length of transverse cracking (X_6) were not significant, as the null hypothesis was not rejected for these two factors. Consequently, they were considered for removal from the model. Upon inspection of the adjusted R-squared value, we found that 52.44% of the variation in IRI values could be explained by the relationships between all the variables in the model. However, further analysis was conducted to optimize the model. This included removing X_1 and X_6 and exploring potential model transformations to improve the correlation and explanatory power of the model. These adjustments were pursued to enhance the model's predictive capability and better capture the key factors influencing pavement deterioration. Assessment of model assumptions was conducted to ensure the mean of residuals approximated zero (indicating linearity), residuals followed a normal distribution, residuals were independent, and variance of residuals was homogenous. Further examination of diagnostic plots confirmed that all model assumptions were met. Several reduced models were tested to identify the most appropriate model, following a forward stepwise analysis approach.

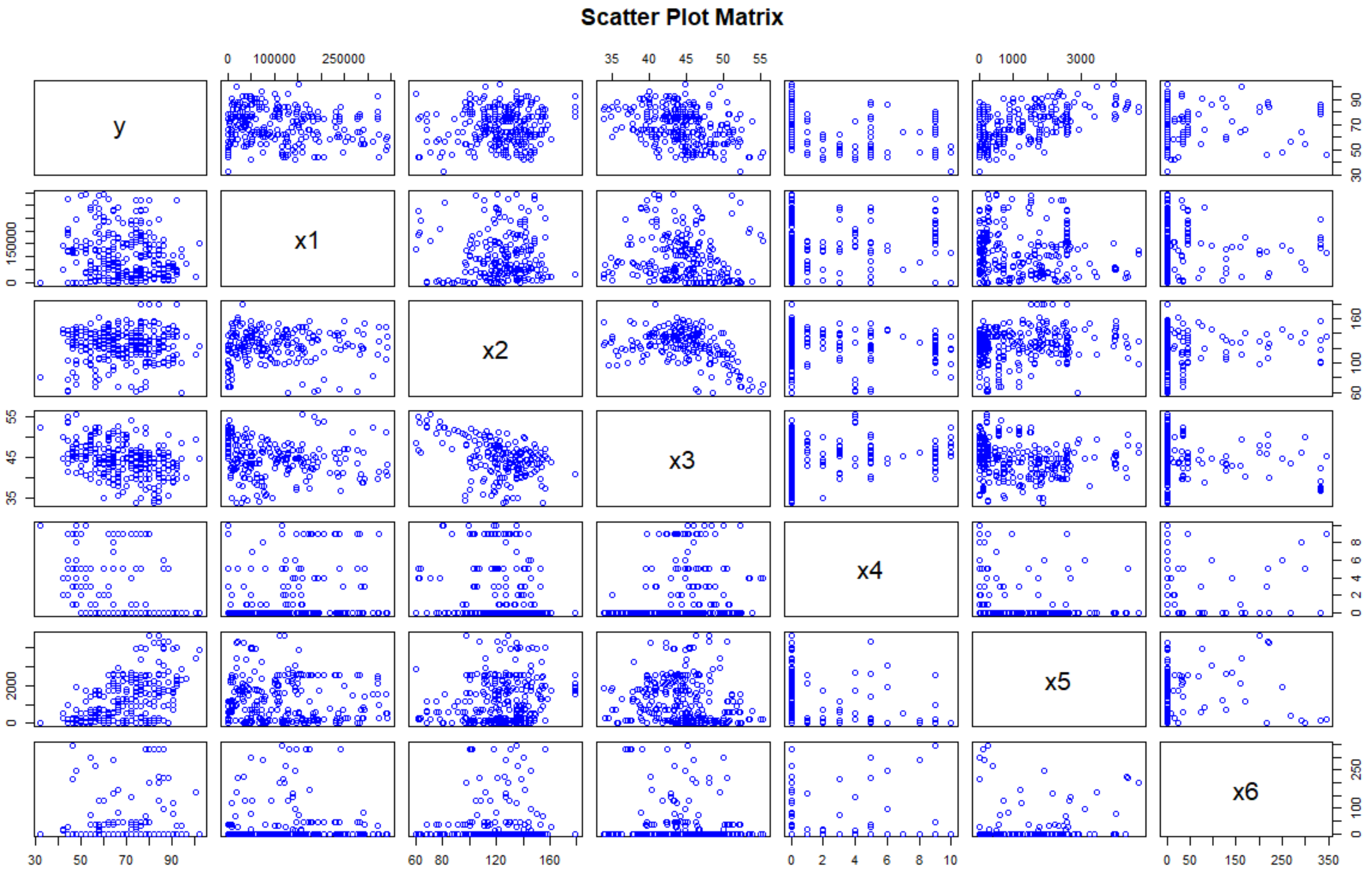

Figure 7: Full model scatter plot matrix

The selected model, excluding annual average daily truck traffic (X_1) and length of transverse cracking (X_6), showed a slightly higher R-squared value of 55.28% compared to the full model and all tested reduced models. Assessment of the selected model's assumptions was performed to ensure all assumptions were met. Diagnostic plots further confirmed that all assumptions were met. As a result, the chosen model is considered the most suitable for predicting IRI values based on the analyzed LTPP dataset. The selected model for this study is shown in equation (4). This model, which omits X_1 and X_6, optimally captures the key factors affecting pavement deterioration and serves as a valuable tool for understanding and forecasting the condition of flexible pavement sections.

$$\hat{Y} = 174 - 13.82\log(X_2) - 0.97\,X_3 - 6.38\log(X_4) + 0.006\,X_5 \tag{4}$$

The "Predicted vs. Actual" plot illustrates the comparison between the predicted and actual (IRI) values. The scatter plot, marked by a red line denoting perfect prediction alignment, reveals a general concordance between the predicted and actual IRI measurements, thereby underscoring some predictive accuracy of the model. However, the spread of data points around this line indicates notable prediction variability. Particularly, the predictions are more consistent around the median IRI range, whereas accuracy diminishes at the extremes, suggesting a potential underfitting issue where the model fails to capture more complex or extreme data patterns. The lack of data points clustering tightly around the red line further points to errors across the spectrum of IRI values, indicating room for model refinement or enhancement through additional features or improved data quality to boost prediction accuracy, especially for very low or high IRI values.

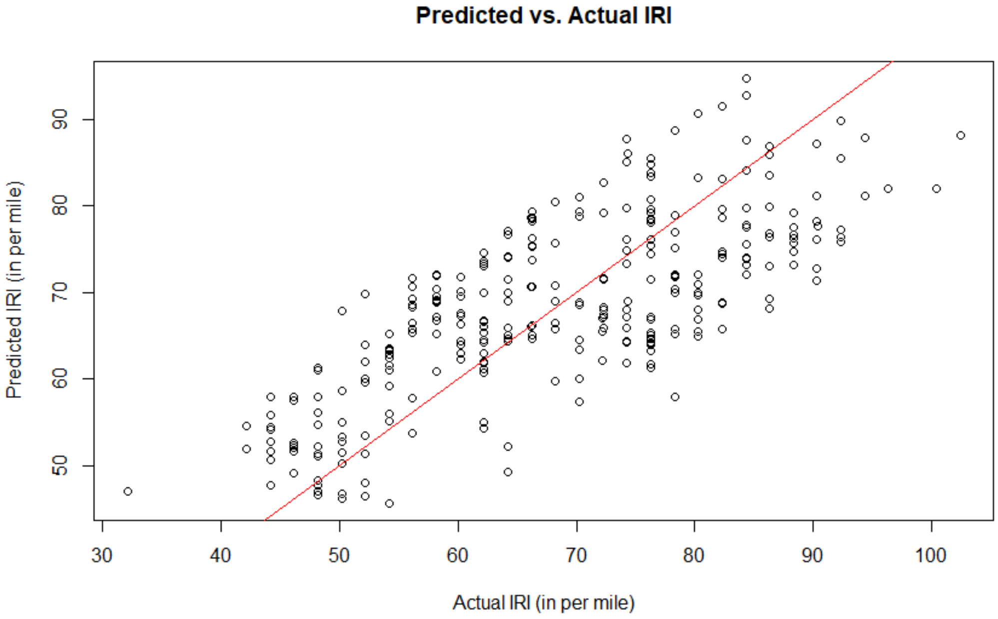

Figure 8: Predicted vs. Actual IRI (in\mile)

DISCUSSION

The findings of our study provide valuable insights into the deterioration patterns and factors affecting the IRI of flexible pavements in the Pacific Northwest. By leveraging the LTPP data and employing advanced statistical methods such as K-M survival analysis and multiple regression, we have been able to assess the impact of different variables on pavement performance over time. Our survival analysis revealed the significance of IRI values in understanding pavement conditions and lifespan. Specifically, we assessed the survival probabilities of flexible pavement sections in both good and fair conditions according to federal measures, identifying trends in how various factors influence pavement longevity. This information is crucial for transportation infrastructure management and maintenance planning, particularly in achieving federal targets for interstate and non-interstate National Highway System (NHS) pavements. Our regression analysis further highlighted the key predictors of IRI values, such as average temperature, percentage of alligator cracking, and length of longitudinal cracking, while identifying less significant factors like annual average daily truck traffic (AADTT) and length of transverse cracking. This nuanced understanding of the most impactful variables aids in prioritizing maintenance strategies and interventions.

The “Predicted vs. Actual” plot demonstrated a general alignment between predicted and actual IRI values, indicating some predictive accuracy of the model. However, the spread of data points around the perfect prediction line suggests notable prediction variability, particularly at the extremes of the IRI range. This indicates potential underfitting issues and suggests a need for model refinement to capture more complex data patterns and improve prediction accuracy, especially for very low or high IRI values. The practical implications of our findings are substantial. For transportation agencies, understanding the key factors affecting pavement performance allows for targeted maintenance strategies, enabling the efficient allocation of resources to achieve federal and state targets. Specifically, the results can guide decisions to maintain pavement in good condition, thereby supporting the Federal Highway Administration’s (FHWA) targets of limiting poor conditions to less than 4% for Interstate pavements and less than 8% for non-Interstate NHS pavements.

However, our study has limitations that may impact the results. The analysis may be subject to data quality issues due to the diverse sources and time periods covered in the LTPP data. Furthermore, our models were based on available data and may not fully capture the complex interplay of variables affecting pavement deterioration. Future research could explore alternative modeling approaches or additional data sources to enhance the predictive power of models and address underfitting issues. The development of flexible pavement deterioration models in this study, using historical IRI measurements and the K-M survival analysis, contributes significantly to the existing body of knowledge. This approach aligns with the findings of Medina et al. [20], who successfully related PCI and IRI through advanced deterioration modeling techniques. However, unlike Medina et al., who primarily focused on empirical relationships between PCI and IRI, this study extends the methodology by incorporating survival analysis to provide a probabilistic estimation of pavement lifespans under varying environmental and traffic conditions. This probabilistic approach offers a more nuanced understanding of pavement behavior over time, which is crucial for regions like the Pacific Northwest where environmental conditions can significantly influence pavement performance. The use of survival analysis highlights how different maintenance strategies can extend pavement life, a critical consideration for optimizing resource allocation in pavement management.

Furthermore, the integration of ML techniques, as seen in the previous studies [10], [13], and [17], suggests a growing trend towards employing sophisticated analytical tools in pavement management systems. The current study's application of these techniques underlines their potential to enhance predictive accuracy and operational efficiency. However, it also brings to light the challenges associated with integrating such complex models into practical applications, such as the need for extensive training data and computational resources. Future research should focus on refining these models to improve their usability in real-world settings, potentially by simplifying the models or enhancing their computational efficiency. Additionally, the development of hybrid models that combine traditional statistical methods with ML could provide a balanced approach, maintaining predictive power while reducing complexity and resource demands. In conclusion, the insights gained from this study contribute to the body of knowledge on pavement performance and can inform more effective maintenance strategies. With ongoing research and development, transportation agencies can leverage these findings to enhance the resilience and durability of flexible pavements in the region.

CONCLUSION

Pavement deterioration prediction is essential for effective transportation infrastructure management and maintenance planning. This study successfully developed deterioration models for flexible pavements in the Pacific Northwest by utilizing extensive data from the InfoPave LTPP tool. Historical records of IRI measurements were analyzed to understand performance degradation patterns of flexible pavements over time. Through K-M survival analysis, we assessed the survival probabilities of flexible pavements based on IRI measures as defined by federal standards over varying time intervals. This approach allowed us to derive models that estimate the percentage of flexible pavements surviving within specified IRI thresholds over time, considering factors such as maintenance treatments, average annual daily truck traffic, freeze and thaw cycles, average temperature, and the presence of various forms of cracking. These models offer valuable insights into the deterioration trends of flexible pavements in the Pacific Northwest and provide predictive tools for infrastructure asset managers. By understanding these trends, managers can optimize maintenance strategies and allocate resources efficiently at the network level.

Future research could focus on refining predictive models to address underfitting and improve accuracy, particularly for very low or high IRI values. Incorporating additional data sources or exploring alternative modeling techniques could enhance the models' predictive power. Investigating the impact of other variables not covered in this study may also provide further insights into pavement performance. In practice, the findings can guide transportation agencies in optimizing maintenance strategies to improve the resilience and longevity of flexible pavements. By focusing on the most impactful variables and targeting maintenance efforts accordingly, agencies can effectively allocate resources to achieve federal and state targets for pavement conditions. In conclusion, our study contributes to the body of knowledge on flexible pavement deterioration and provides practical guidance for transportation infrastructure management. Continued research and model development will further enhance our understanding of pavement performance and support the development of more effective maintenance strategies for the long-term durability of transportation infrastructure.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Lamei and Kassem; data collection: Lamei; analysis and interpretation of results: Lamei and Kassem; draft manuscript preparation: Lamei. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] E. Kassem, A. Awed, E. A. Masad, and D. N. Little, “Development of Predictive Model for Skid Loss of Asphalt Pavements,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2372, no. 1, pp. 83–96, Jan. 2013, doi: 10.3141/2372-10.

[2] O. Kaya, H. Ceylan, S. Kim, D. Waid, and B. P. Moore, “Statistics and Artificial Intelligence-Based Pavement Performance and Remaining Service Life Prediction Models for Flexible and Composite Pavement Systems,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2674, no. 10, pp. 448–460, Oct. 2020, doi: 10.1177/0361198120915889.

[3] R. Justo-Silva, A. Ferreira, and G. Flintsch, “Review on Machine Learning Techniques for Developing Pavement Performance Prediction Models,” *Sustainability*, vol. 13, no. 9, p. 5248, May 2021, doi: 10.3390/su13095248.

[4] “LTPP InfoPave - Home.” Accessed: May 02, 2024. [Online]. Available: https://infopave.fhwa.dot.gov/

[5] Y. Jia, S. Wang, J. Peng, Y. Gao, M. Liu, and W. Zhou, “Characterization of rutting on asphalt pavement in terms of transverse profile shapes based on LTPP data,” *Constr. Build. Mater.*, vol. 269, p. 121230, Feb. 2021, doi: 10.1016/j.conbuildmat.2020.121230.

[6] M. A. Younos, R. T. Abd El-Hakim, S. M. El-Badawy, and H. A. Afify, “Multi-input performance prediction models for flexible pavements using LTPP database,” *Innov. Infrastruct. Solut.*, vol. 5, no. 1, p. 27, Apr. 2020, doi: 10.1007/s41062-020-0275-3.

[7] J. Karam and H. Noorvand, “Developing a Rutting Prediction Model for HMA Pavements Using the LTPP Database,” *Int. J. Pavement Res. Technol.*, Jun. 2023, doi: 10.1007/s42947-023-00340-9.

[8] X. Luo, F. Wang, H. Gong, J. Tao, X. Qiu, and N. Wang, “Effectiveness Evaluation of Preventive Maintenance Treatments on Asphalt Pavement Performance Using LTPP Data,” *Int. J. Pavement Res. Technol.*, vol. 15, no. 5, pp. 1139–1154, Sep. 2022, doi: 10.1007/s42947-021-00078-2.

[9] M. Li, Q. Dai, P. Su, Z. You, and Y. Ma, “Surface layer modulus prediction of asphalt pavement based on LTPP database and machine learning for Mechanical-Empirical rehabilitation design applications,” *Constr. Build. Mater.*, vol. 344, p. 128303, Aug. 2022, doi: 10.1016/j.conbuildmat.2022.128303.

[10] M. R. Kaloop, S. M. El-Badawy, J. W. Hu, and R. T. Abd El-Hakim, “International Roughness Index prediction for flexible pavements using novel machine learning techniques,” *Eng. Appl. Artif. Intell.*, vol. 122, p. 106007, Jun. 2023, doi: 10.1016/j.engappai.2023.106007.

[11] K. Zhang and Z. Wang, “LTPP data-based investigation on asphalt pavement performance using geospatial hot spot analysis and decision tree models,” *Int. J. Transp. Sci. Technol.*, vol. 12, no. 2, pp. 606–627, Jun. 2023, doi: 10.1016/j.ijtst.2022.06.007.

[12] S. Khan, K. Hossain, C. Bazan, and S. T. Swarna, “Factors Influencing the Design of Semirigid Pavement in Different Climatic Regions Using LTPP Data, MEPDG Software, and the Design of the Experiment Method,” *J. Transp. Eng. Part B Pavements*, vol. 150, no. 1, p. 04023034, Mar. 2024, doi: 10.1061/JPEODX.PVENG-1200.

[13] T. Baykal, F. Ergezer, E. Eriskin, and S. Terzi, “Using Ensemble Machine Learning to Estimate International Roughness Index of Asphalt Pavements,” *Iran. J. Sci. Technol. Trans. Civ. Eng.*, Dec. 2023, doi: 10.1007/s40996-023-01320-6.

[14] J. Liu, F. Liu, C. Zheng, E. O. Fanijo, and L. Wang, “Improving asphalt mix design considering international roughness index of asphalt pavement predicted using autoencoders and machine learning,” *Constr. Build. Mater.*, vol. 360, p. 129439, Dec. 2022, doi: 10.1016/j.conbuildmat.2022.129439.

[15] S. M. Piryonesi and T. E. El-Diraby, “Examining the relationship between two road performance indicators: Pavement condition index and international roughness index,” *Transp. Geotech.*, vol. 26, p. 100441, Jan. 2021, doi: 10.1016/j.trgeo.2020.100441.

[16] R. Imam, Y. Murad, I. Asi, and A. Shatnawi, “Predicting Pavement Condition Index from International Roughness Index using Gene Expression Programming,” *Innov. Infrastruct. Solut.*, vol. 6, no. 3, p. 139, Sep. 2021, doi: 10.1007/s41062-021-00504-1.

[17] T. Tamagusko and A. Ferreira, “Machine Learning for Prediction of the International Roughness Index on Flexible Pavements: A Review, Challenges, and Future Directions,” *Infrastructures*, vol. 8, no. 12, p. 170, Nov. 2023, doi: 10.3390/infrastructures8120170.

[18] E. L. Kaplan and P. Meier, “Nonparametric Estimation from Incomplete Observations,” *J. Am. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, Jun. 1958, doi: 10.1080/01621459.1958.10501452.

[19] A. A. Hatoum, J. M. Khatib, F. Barraj, and A. Elkordi, “Survival Analysis for Asphalt Pavement Performance and Assessment of Various Factors Affecting Fatigue Cracking Based on LTPP Data,” *Sustainability*, vol. 14, no. 19, p. 12408, Sep. 2022, doi: 10.3390/su141912408.

[20] J. R. Medina, A. Zalghout, A. Gundla, S. Castro, and K. Kaloush, “Use of Time–Temperature Superposition Principle to Create Pavement Performance Master Curves and Relate Pavement Condition Index and International Roughness Index,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2675, no. 9, pp. 972–983, Sep. 2021, doi: 10.1177/03611981211004965.