ROADWAY SAFETY INSTITUTE

Human-centered solutions to advanced roadway safety

Accident Prediction Models using Macro and Micro Scale Analysis: Dynamic Tree and Zero Inflated Negative Binomial Models with Empirical Bayes Accident History Adjustment

Jacob Mathew

Rahim F. Benekohal

Juan C. Medina

Department of Civil and Environmental Engineering
University of Illinois Urbana-Champaign

Final Report

REGION 5

CTS 19-02

UNIVERSITY OF MINNESOTA
Driven to Discover℠

The University of Akron

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

SOUTHERN ILLINOIS UNIVERSITY EDWARDSVILLE

WESTERN MICHIGAN UNIVERSITY

Technical Report Documentation Page

1. Report No. CTS 19-02	2.	3. Recipients Accession No.	
4. Title and Subtitle: Accident Prediction Models using Macro and Micro Scale Analysis: Dynamic Tree and Zero Inflated Negative Binomial Models with Empirical Bayes Accident History Adjustment		5. Report Date February 2019 6.	
7. Author(s) Jacob Mathew, Rahim F. Benekohal, Juan C. Medina		8. Performing Organization Report No.	
9. Performing Organization Name and Address Department of Civil and Environmental Engineering University of Illinois Urbana-Champaign 205 N Mathews Ave, Urbana, IL 61801		10. Project/Task/Work Unit No. CTS#2015056 11. Contract (C) or Grant (G) No. DTRT13-G-UTC35	
12. Sponsoring Organization Name and Address Roadway Safety Institute Center for Transportation Studies University of Minnesota 200 Transportation and Safety Building 511 Washington Ave. SE Minneapolis, MN 55455		13. Type of Report and Period Covered Final Reports 14. Sponsoring Agency Code	
15. Supplementary Notes http://www.roadwaysafety.umn.edu/publications/			

16. Abstract (Limit: 250 words)

This report presents two ways to analyze accidents at highway rail grade crossings: a microscopic approach that looks at individual accidents at a crossing or a group of crossings, and a macroscopic approach to identify correlations between accident counts at crossings and crossing characteristics. The outcome of the microscopic approach is a data-driven dynamic tree that helps to visualize accident trends at a single crossing or a group of crossings. The dynamic tree is also used to identify new variables (crossing angle and distance to nearby highway intersection). The outcome of the macroscopic approach were new accident prediction models for crossings with gates, flashing lights, and crossbucks. Zero Inflated Negative Binomial models were used to predict the accident counts and the Empirical Bayes approach was used to adjust the predicted based on accident history at the crossing. Data from the state of Illinois was used to develop the model and data from four other states were used to validate the model. The newly developed models resulted in cumulative predicted accident distributions that closely represent the field data. The EB adjusted ZINB accident predictions value were significantly closer to the actual accident counts for the crossings than the USDOT models. More accurate predictions from the EB-adjusted ZINB model were obtained for the top 10, 20, 30, 40 and 50 locations with highest accident frequency for all three warning devices.

17. Document Analysis/Descriptors Railroad grade crossings, Crash analysis, Trees (Mathematics), Binomial distributions, Bayes' theorem		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 102	22. Price

Accident Prediction Models using Macro and Micro Scale Analysis: Dynamic Tree and Zero Inflated Negative Binomial Models with Empirical Bayes Accident History Adjustment

FINAL REPORT

Prepared by:

Jacob Mathew
Rahim F. Benekohal
Juan C. Medina
Department of Civil and Environmental Engineering
University of Illinois Urbana Champaign

February 2019

Published by:

Roadway Safety Institute
Center for Transportation Studies
University of Minnesota
200 Transportation and Safety Building
511 Washington Ave. SE
Minneapolis, MN 55455

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The contents do not necessarily represent the views or policies of the United States Department of Transportation (USDOT) or the University of Illinois. This document is disseminated under the sponsorship of the USDOT's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

The authors, the USDOT, and the University of Illinois do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to this report.

ACKNOWLEDGMENTS

The funding for this project was provided by the United States Department of Transportation's Office of the Assistant Secretary for Research and Technology for the Roadway Safety Institute, the University Transportation Center for USDOT Region 5 under the Moving Ahead for Progress in the 21st Century Act (MAP-21) federal transportation bill passed in 2012.

In addition, the authors would like to acknowledge assistance of Stephen Laffey of Illinois Commerce Commission for providing data for Illinois.

TABLE OF CONTENTS

CHAPTER 1: Introduction 1

CHAPTER 2: Dynamic Tree for Accident Visualization at Highway Rail Grade Crossings 2

2.1 Introduction to Chapter 2 2

2.2 Overview of static method: a tree to visualize accidents 2

2.3 Improved accident visualization: dynamic method 5

2.3.1 Method A (Absolute Sorting) 7

2.3.2 Method B (Nested Sorting) 9

2.3.3 Modified Method B (Modified Nested Sorting) 11

2.4 Static and dynamic methods using recent data from three states 13

2.4.1 Selected Crossings from Illinois (Recent Data) 14

2.4.2 Selected Crossings from California 16

2.4.3 Selected Crossings from Indiana 21

2.5 Handling special cases using the modified method b 26

2.6 Dynamic tree for corridors 28

2.7 Dynamic tree analysis for a group of single accident locations 33

2.8 Dynamic tree analysis using variables from inventory database 37

2.8.1 XAngle 37

2.8.2 HwyNDist 44

CHAPTER 3: Zero Inflated Negative Binomial Models with Empirical Bayes Accident History Adjustment 54

3.1 Introduction to Chapter 3 54

3.2 Data Source and Data Cleaning 54

3.2.1 Development Data 56

3.2.2 Validation Data 57

3.3 USDOT Accident Prediction Formula 58

3.4 Zero Inflated Negative Binomial (ZINB) Model 60

3.4.1 Previous ZINB for Crossings with Gates 61

3.4.2 Previous ZINB for Crossings with Flashing Lights 61

3.4.3 Previous ZINB for Crossings with Crossbucks 61

3.4.4 Crossings with Gates 68

3.5 Crossings with Flashing Lights 69

3.6 Crossings with Crossbucks 69

3.7 Empirical Bayes Accident History Adjustment 70

3.8 Comparison between the Models 70

3.8.1 Cumulative distribution of predicted accidents with respect to cumulative distribution of accidents 71

3.8.2 Predicted number of accidents at a location with respect to actual number of accidents at location 82

3.8.3 Relative ranking of crossings based on the number of accidents identified among top locations. 84

CHAPTER 4: Conclusions and Recommendations 88

4.1 Conclusions 88

4.2 Recommendations and Future Work 88

REFERENCES 90

LIST OF FIGURES

Figure 1: Crossing 173887G and Tree using Static Method 4

Figure 2: Tree using Dynamic Method A for hypothetical crossing 8

Figure 3: Tree using Dynamic Method A for crossing 173887G 9

Figure 4: Tree for Hypothetical Crossing using Dynamic Tree Method B. 10

Figure 5: Tree for crossing 173887G using Method B. 10

Figure 6: Modified Method B for Crossing 173887G. 13

Figure 7: Map and Sketch of 608311K 14

Figure 8: Tree for crossing at 608311K using Static Method 15

Figure 9: Dynamic Tree using Modified Method B for 608311K 15

Figure 10: Map and Sketch for 811479J 16

Figure 11: Static tree for 811479J 17

Figure 12: Dynamic tree using Modified Method B for 811479J 17

Figure 13: Map and Sketch of 028380R 18

Figure 14: Static Tree for 028380R 18

Figure 15: Tree using Modified Method B for 028380R 19

Figure 16: Map and Sketch of 026517B 20

Figure 17: Static Tree for 026517B 20

Figure 18: Tree using Modified Method B for 026517B 21

Figure 19: Map and sketch of crossings 522646H 22

Figure 20: Tree using Static method for 522646H 23

Figure 21: Tree using modified method B for 522646H 23

Figure 22: Map and Sketch of 879204S 24

Figure 23: Tree using modified method B for 879204S 25

Figure 24: Tree using modified method B for 879204S 26

Figure 25: Six Accidents where Motorists were stopped on the Crossing 27

Figure 26: Selected Crossings along Corridor1 near Chicago 28

Figure 27: Tree for Corridor 1 using Static Method 29

Figure 28: Dynamic Tree analysis for Corridor 1 using Modified Method B... 30

Figure 29: Selected Crossings along Corridor 2 near Chicago... 30

Figure 30: Tree for Corridor 2 using Static Method 31

Figure 31: Dynamic Tree using Modified Method B for Corridor 2 32

Figure 32: Dynamic tree using Modified Method B for Corridor without pedestrian accidents 33

Figure 33: Dynamic Tree using Modified Method B for all Single Locations 34

Figure 34: Dynamic Tree Analysis using Modified Method B for all Single Locations with Cross bucks 35

Figure 35: Dynamic Tree Analysis using Modified Method B for single accident locations with Flashing Lights 36

Figure 36: Dynamic Tree Analysis using Modified Method B for single accident locations with Gates 37

Figure 37: Accidents at Gated Crossings with Crossing Angle ≥ 60 degrees 39

Figure 38: Accidents at Gated Crossings with Crossing Angle < 60 degrees... 40

Figure 39: Accidents at Crossings with Flashing Lights with crossing angle ≥60 degrees 41

Figure 40: Accidents at Crossings with Flashing Lights with crossing angle <60 degrees 42

Figure 41: Accidents at Crossings with Crossbucks with Angle < 60 degrees.. 43

Figure 42: Accidents at Crossings with Crossbucks with Angle ≥ 60 degrees.. 44

Figure 43: Accidents at Crossings with Gates with Highway intersection <75 feet.................................. 46

Figure 44: Accidents at Crossings with Gates with Highway intersection between 75-500 feet 47

Figure 45: Accidents at Crossings with Gates with Highway intersection ≥ 500 feet................................ 48

Figure 46: Accidents at Crossings with Flashing Lights with Highway < 75 feet.. 49

Figure 47: Accidents at Crossings with Flashing Lights with nearby Highway ≥ 500 feet 50

Figure 48: Accidents at Crossings with Crossbucks with Highway < 75 feet 51

Figure 49: Accidents at Crossings with Crossbucks with Highway ≥ 500 feet 52

Figure 50: Cumulative number of accidents at all locations for Crossings with Gates (Model Development Dataset) ... 71

Figure 51: Cumulative number of accidents at top locations for crossings with Gates (Model Development Dataset) ... 72

Figure 52: Cumulative number of accidents at all locations for Crossings with Gates (Model Validation Dataset) ... 73

Figure 53: Cumulative number of accidents at top Locations for Crossings with Gates (Model Validation Dataset) ... 73

Figure 54: Cumulative number of Accidents for all locations for Crossings with Flashing Lights (Model Development Dataset) ... 74

Figure 55: Cumulative number of accidents at top locations for Crossings with Flashing Lights (Model Development Dataset) ... 75

Figure 56: Cumulative number of accidents at all locations for Crossings with Flashing Lights (Model Validation Dataset) ... 76

Figure 57: Cumulative number of accidents at top locations for Crossings with Flashing Lights (Model Validation Dataset) ... 77

Figure 58: Cumulative number of accidents at all locations with Crossbucks (Model Development Dataset) ... 78

Figure 59: Cumulative number of accidents at top locations for Crossings with Crossbucks (Model Development Dataset) ... 79

Figure 60: Cumulative number of accidents at all locations for Crossings with Crossbucks (Model Validation Dataset) ... 80

Figure 61: Cumulative number of accidents at top locations Crossings with Crossbucks (Model Validation Dataset) ... 81

Figure 62: Predicted Vs Observed Accident Count for Gated Crossings (Model Development Dataset) ... 82

Figure 63: Predicted Vs Observed Accident Count for Gated Crossings (Model Validation Dataset) ... 82

Figure 64: Predicted Vs Observed Accident Count for Crossings with Flashing Lights (Model Development Dataset) ... 83

Figure 65: Predicted Vs Observed Accident Count for Crossings with Flashing Lights (Model Validation Dataset) ... 83

Figure 66: Predicted Vs Observed Accident Count for Crossings with Crossbucks (Model Development Dataset) 84

Figure 67: Predicted Vs Observed Accident Count for Crossings with Crossbucks (Model Validation Dataset) 84

LIST OF TABLES

Table 1. Attributes used in the Static Method 3

Table 2: Attributes used in the Dynamic Method 5

Table 3: Accident Details for Hypothetical Crossing 7

Table 4: Hypothetical Database to Explain the Procedure to Handle Ties 11

Table 5: Number of Accidents for Each Angle Group and Warning Device 38

Table 6: Number of Accidents Split by HwyNDist and Warning Device 45

Table 7: Filters Applied to Database 55

Table 8: Number of Crossings and Accidents in Development Data (Illinois) 57

Table 9: Number of Crossings and Accidents for the Validation Data 57

Table 10: USDOT Accident Prediction Formula Equations 58

Table 11: Accident Prediction Normalizing Constants 59

Table 12: Variables used in Model Development 62

Table 13: Variables used in ZINB models and AIC values for Crossings with Gates 64

Table 14: Variables used in ZINB models and AIC values for Crossings with Flashing Lights 66

Table 15: Variables used in ZINB models and AIC values for Crossings with Crossbucks 67

Table 16: Number of Accidents Observed in top ‘n’ Crossings in Model Development Dataset 85

Table 17: Number of Accidents Observed in top ‘n’ Crossings in Model Validation Dataset 86

EXECUTIVE SUMMARY

There were 10,501 accidents at highway rail grade crossing locations nationwide in five years (2012-2016), resulting in 1,215 fatalities and 4,709 injuries *(1)*. Among these locations, the higher risk crossings should be identified for efficient allocation of scarce resources towards operational and safety improvements. The higher risk crossings can be identified by analyzing the accidents history and the crossing characteristics.

Accident analysis at highway rail grade crossings may be done at a microscopic level or at a macroscopic level. At the microscopic level, accidents at a crossing are analyzed to explore if there are common characteristics. At the macroscopic level, accident counts at a group of crossings are analyzed to explore relationships with the crossing related variables. These two types of analyses complement each other and provide information that can be used to identify higher risk crossings.

Chapter 2 of this report discusses accident analysis at the microscopic level. The researchers had previously developed a tree-based tool (Static Method) to visualize the trends related to the attributes of a crossing and its accidents *(2, 3)*. This type of analysis gave positive results, showing that such an approach had the potential for accident analysis.

This project explores the implementation of improvements to the Static Method and proposes a micro-level data-driven approach – called a “Dynamic Method” – with the goal of highlighting not only the trends in attributes related to grade crossing crashes, but also outliers and unexpected characteristics of groups of crashes. The researchers implemented the Dynamic Method in a computer program to reduce analysis time and frequency of errors. The Dynamic Method can also be used for corridor analysis and to discover trends on specific groups of crashes of interest: for example, accidents involving older drivers or taking place in specific weather conditions.

Analysis of multiple accident locations using the dynamic tree provides information about the accidents that otherwise would be difficult to visualize. The findings from the microscopic method were also incorporated in the macroscopic models to improve the overall accident predictions from a macroscale point of view.

Chapter 3 of this report discusses accident analysis on a macroscopic scale and the development of new statistical accident prediction models. The objective was to improve predictions over the USDOT accident prediction formula *(7)*. The most recent grade crossing accident and inventory datasets were downloaded from the Federal Railroad Administration’s (FRA) website so that the most up-to-date data was used for model development.

In this study the researchers developed Zero Inflated Negative Binomial models (ZINB) for each category of warning devices (gates, flashing lights and crossbucks). These three categories reflect those used by the USDOT model. The researchers used ZINB models to fit accident counts to account for the high number of crossings with no accidents during the analysis period. The dynamic tree method identified new variables used in the model. Two new variables (crossing angle and distance to nearby highway intersection) were chosen. The researchers adjusted the predictions using the Empirical Bayes approach

to account for the accident history at the crossings *(8)*. The research team explored ZINB models previously *(3)*. The main differences between the two models include the number of variables used in the model development and the accident history adjustment applied to the model.

The researchers applied various data cleaning filters to the dataset to ensure erroneous or missing data is not used. The new ZINB models were developed using data from Illinois and validated using data from four other states (Iowa, Pennsylvania, Texas and South Carolina). The EB-adjusted ZINB models are also compared with the USDOT accident predictions.

The comparisons show that the newly developed ZINB models with EB adjustment resulted in cumulative predicted accident distributions that closely represent the field data. Also, plots of actual accident counts against predicted accident counts by the two models showed that the new accident prediction values were significantly closer to the actual accident counts than the USDOT models. More accurate predictions from the EB-adjusted ZINB model were obtained for the top 10, 20, 30, 40 and 50 locations with highest accident frequency for all three warning devices.

The authors recommend that data from external datasets be incorporated to fine-tune the accident prediction models. The Empirical Bayes approach should be explored further and considered with a “better” count model as an alternative to USDOT models to make accident count predictions at grade crossings.

CHAPTER 1: INTRODUCTION

Between the years of 2012 and 2016, there were 10,501 accidents at highway rail grade crossing locations nationwide, resulting in 1,215 fatalities and 4,709 injuries*(1)*. Among these locations, the higher risk crossings should be identified for efficient allocation of scarce resources towards operational and safety improvements. The higher risk crossings can be identified by analyzing the accidents history and the crossing characteristics.

Accident analysis at highway rail grade crossings may be done at a microscopic level (individual accidents) or at a macroscopic level (modeling accident counts with respect to crossing characteristics). At the microscopic level, accidents at a crossing are analyzed to explore if there are common characteristics (for example, accidents involving older drivers or specific weather conditions). At the macroscopic level, accident counts at a group of crossings (for example, all the crossing in a particular region or state) are analyzed to explore any relationships that may exist with the crossing related variables. These two types of analyses complement each other and provide information that can be used to identify higher risk crossings. This report contains four chapters. Chapter 1 is the introduction.

Chapter 2 discusses the microscopic analysis of accidents at highway rail grade crossings. Microscopic analysis of accidents was conducted using a tree-based structure to identify trends in accident characteristics happening at an individual crossing (or a group of crossings). Two different methods are discussed in this chapter: Static Method and the Dynamic Method. An overview of the Static Method, developed previously, along with its identified shortcomings, is presented in this chapter. A discussion of the development of the dynamic tree method for microscopic accident analysis at highway rail grade crossings is also provided. The static tree method and the newly developed dynamic tree method were used on accidents on crossings from the states of Illinois, Indiana, and California, and the resulting trees are shown in this report.

Chapter 3 presents the macroscopic analysis developed to model accident counts at highway rail grade crossings. This chapter discusses data cleaning, model development for accident count estimation, and model validation. A discussion of the Empirical Bayes (EB) method to improve the accuracy of the accident count prediction is given. Models developed in this chapter used data from the state of Illinois collected over five years (2012-2016). The developed models are validated using data from four other states. Models are developed for each of the three warning device categories in the USDOT accident prediction formulae. A comparison of the newly developed model to the USDOT formulae is also given in this chapter.

Chapter 4 is the conclusion and recommendation of the study. The findings of chapters 2 and 3 are summarized in this chapter.

CHAPTER 2: DYNAMIC TREE FOR ACCIDENT VISUALIZATION AT HIGHWAY RAIL GRADE CROSSINGS

2.1 INTRODUCTION TO CHAPTER 2

This project explores accidents at railroad-highway grade crossings from a microscopic scale (individual accident level) and in combination with newly developed models at a macroscopic scale (aggregated level). Previous efforts were focused on developing a microscopic methodology to provide a tree structure that helps to visualize the trends related to the attributes of the subject crossing and its accidents *(2, 3)*. This method used nine attributes from the FRA accident and inventory database *(4, 5)*, and these attributes were selected and sorted in a fixed order to form a tree structure (this will be referred hereafter as the “Static Method”). Results from using the tree structure were positive and showed that such approach had potential for accident analysis. However, the Static Method could be improved by adding the following capabilities: 1) including additional attributes available in the FRA database, 2) removing the fixed hierarchy in which the attributes were ordered, and 3) computerizing of the process to reduce time requirements and potential for human errors.

This project explores the implementation of such improvements to the Static Method and proposes a micro-level data-driven approach – called a “Dynamic Method” – with the goal of highlighting not only the trends in attributes related to grade crossing crashes but also outliers and unexpected characteristics of groups of crashes. The Dynamic Method used data obtained from the FRA database, from which a total of 22 attributes of an accident and its location were considered for building the tree structure. The dynamic method was implemented in a computer program to incorporate a routine that efficiently sorts the 22 potential attributes (listed in **Table 2**) based on the trends for the particular crossing being analyzed. A computer program to run the Dynamic Method also allows for reduced analysis time and the frequency of errors in data processing. Lastly, the Dynamic Method is also suited for corridor analysis and to discover trends on specific groups of crashes of interest, for example: older drivers or weather conditions.

The Dynamic Method is expected to provide the analysis with a new perspective on accident trends not only at individual crossings but also for groups of crossings. The findings from the microscopic method, when it is possible to generalize, will be incorporated in the macroscopic models to improve the overall accident predictions from a multi-scale point of view. In turn, improved accident predictions will allow for more efficient resource allocation for crossing upgrades, enhancing the investment of resources to maximize the crash reductions.

2.2 OVERVIEW OF STATIC METHOD: A TREE TO VISUALIZE ACCIDENTS

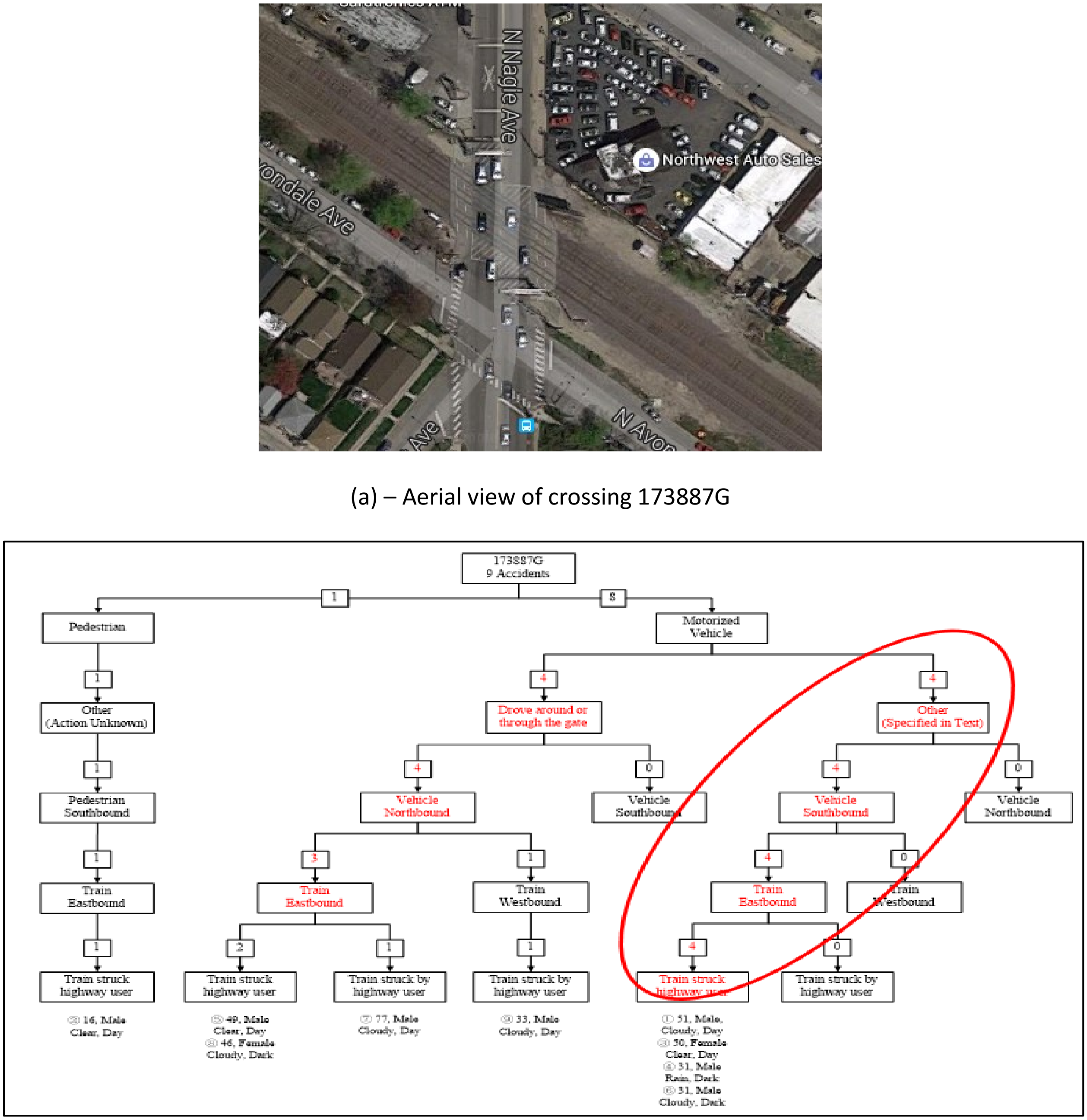
In previous studies *(2, 6)*, a tree structure was proposed to visualize potential accident trends using a pre-defined arrangement of specific attributes available from the FRA database. The tree structure is referred hereafter as the Static Method and the variables considered are shown in **Table 1**. The attribute in **Table 1** are listed in the order they are applied. The order was fixed for all locations. The data used for the Static Method is 10 years data from 2003-2012

Table 1. Attributes used in the Static Method

Attribute Number	Attribute	Number of Sub-Categories	Remarks
1	TYPVEH	3	Vehicle Type: Pedestrian, Motorist and Others
2	MOTORIST	5	Action of Motorist
3	VEHDIR	4	Direction of Highway User
4	TRNDIR	4	Direction of the Train
5	TYPACC	2	Accident Type: Train struck HW user OR Train struck by HW user
6	DRIVAGE	3	0-30, 30-60, >60
7	DRIVGEN	2	Gender of Driver
8	WEATHER	6	Clear, Cloudy, Rain, Fog, Sleet, Snow
9	VISIBLTY	4	Dawn, Day, Dusk Dark

The main advantage of the Static Method is that it is relatively simple to implement. It also has a predetermined hierarchy based on our general knowledge of rail-highway crossing accidents that keeps related (or potentially related) attributes adjacent to each other. However, the main limitation of the Static Method is that the order the attributes are used is fixed and it is possible that at a given crossing an attribute that is kept at a lower level of the tree could present a more focused accident concentration than the attribute kept at a higher level. Another disadvantage of the Static Method is that it limits the number attributes that make up the tree structure.

An example of the Static Method for the crossing with an ID of 173887G in Illinois (Figure 1a) is given in Figure 1b. The crossing had 9 accidents in the 10-year analysis period (2003-2012), and the tree could highlight four accidents involving a motorized vehicle traveling in the southbound direction and a train traveling eastbound.



(a) – Aerial view of crossing 173887G

(b) Tree using Static Method

Figure 1: Crossing 173887G and Tree using Static Method

2.3 IMPROVED ACCIDENT VISUALIZATION: DYNAMIC METHOD

In light of the limitations of the Static Method, a dynamic data-driven tree structure was explored to select and prioritize the most desirable attributes to visualize accident trends, it is called Dynamic Method. Given the increased number of variables in the database, and the high number of possible combinations to sort them, the process was automated using a computer program. In addition to reducing the time needed to find and sort the most significant attributes for a set of accidents, the computer program reduced the likelihood of human error and opened a window to expand the process to greater collection of accidents from corridors or special crossings of interest.

Thus, with the "Dynamic Method" user can consider a greater number of attributes depending on the database they have. Currently, we are using a pool of 22 attributes and the program selects from the following 22 variables in the FRA database:

Table 2: Attributes used in the Dynamic Method

Attribute Number	Attribute	Number of Sub-Categories	Remarks
1	MONTH	4	Depending on season
2	AMPM	2	AM or PM
3	VEHSPD	5	Vehicle Speed: <20, 20-40, 40-60, 40-60, >60
4	TYPVEH	3	Vehicle Type: Pedestrian, Motorist and Others
5	VEHDIR	4	Direction of Highway User
6	POSITION	4	Stalled, Stopped, Moving, Trapped
7	TYPACC	2	Accident Type: Train struck HW user OR Train struck by HW user
8	VISIBLTY	4	Dawn, Day, Dusk Dark
9	WEATHER	6	Clear, Cloudy, Rain, Fog, Sleet, Snow

10	TYPTRK	4	Track Type: Main, Yard, Siding, Industry
11	TRKCLAS	10	FRA Track Class
12	TRNSPD	5	Train Speed: <20, 20-40, 40-60, 40-60, >60
13	TRNDIR	4	Direction of the Train
14	LOCWARN	3	Location of the warning device *wrt* HW user: Both Sides, Side of Vehicle Approach, Opp. Side
15	WARNSIG	3	Interconnection of warning device to HW signal: Yes, No, Unknown
16	LIGHTS	3	Lights at Crossing: Yes, No, Unknown
17	MOTORIST	5	Action of Motorist
18	VIEW	8	Primary obstruction of track view: 8 categories
19	CROSSING	2	Some warning device or No warning device
20	PUBLIC	2	Public or Private
21	DRIVAGE	3	0-30, 30-60, >60
22	DRIVGEN	2	Gender of Driver

Three methods to prioritize the attributes were considered and tried and they are explained below. The meaning of each of the attributes mentioned in **Table 2** is detailed in the FRA file structure for the accident database (specifically, form 6180.57) that is available at: (http://safetydata.fra.dot.gov/officeofsafety/publicsite/downloadfstructure.aspx). The order in which the methods are presented is from simple to more improved ones.

2.3.1 Method A (Absolute Sorting)

In Method A, accidents were distributed into the subcategories of each of the 22 different attributes. Attributes were sorted based on the largest number of accidents that was kept in a subcategory. Through this process the order of the attributes (levels) in the tree structure is determined. Then, the main branch is constructed from the subcategory with the most accidents in the previous level. A step by step procedure for this method is given below:

1. Accidents are divided into the sub-categories of the 22 different attributes considered.
2. For each attribute, the subcategory which holds the highest number of accidents is identified.
3. The largest subcategories (found in step 2) are sorted in the descending order. This order gives the hierarchy of attributes.
4. The tree is built following the hierarchy by dividing the largest subcategory from the top level into subcategories in the second level. This process is continued as long as the number of accident can be divided into subcategories.

For explanation, consider three attributes A1, A2, A3 are used for the formation of the tree. Each of the three attributes has two subcategories, namely A11 and A12. Assume that the total number of accidents at the crossing is 12. The details of this hypothetical crossing are given in **Table 3**.

Table 3: Accident Details for Hypothetical Crossing

Accident Number	Attribute 1	Attribute 2	Attribute 3
1	A11	A21	A31
2	A11	A22	A31
3	A11	A21	A31
4	A11	A21	A32
5	A11	A21	A31
6	A11	A21	A32
7	A11	A22	A31
8	A11	A21	A32

9	A11	A21	A31
10	A11	A22	A32
11	A12	A22	A31
12	A12	A22	A31

Per method A, the accidents are first classified into sub-categories of all the attributes used. Attribute A1 groups 10 accidents into the subcategory A11 and the remaining accidents into A12. Attribute A2 groups 7 accidents into A21 and the remaining into A22, and attribute A3 groups 8 accidents into A31 and the remaining into A32. The highest subcategory from each of the group is selected, thus A11, A21 and A31 are selected and these are used to sort the attributes in the descending order. The resulting order of the attributes is A1, A3, and finally A2. The tree is formed by dividing the highest node in each level of the tree as per the above determined hierarchy of attributes, as shown in **Figure 2**.

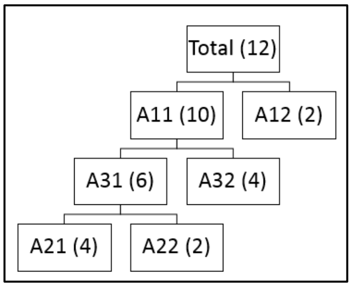


Figure 2: Tree using Dynamic Method A for hypothetical crossing

The advantage of Method A is that it is relatively simple and the hierarchy determined by this method is dependent on the distribution of the accidents into subcategories of 22 attributes. An illustration of the dynamic tree formed using Method A is given in **Figure 3**. For the dynamic tree, we used the same crossing with 9 accidents (173887G) that was used for the static tree (the same time period). The VEHDIR recorded in the database were interpreted as: 1,3 (NB) and 2,4(SB), and TRNDIR recorded in the database were interpreted as: 2,3 (EB) and 1,4(WB). The column to the right in **Figure 3** shows the attribute name, and the tree on the left shows the categories of the attributes and the number of accidents in that category (in parenthesis). The tree reveals that in 8 of the 9 accidents vehicle speeds were lower than 20 mph, in five of those occasions the rail equipment struck the highway user (TYPACC="Rail->HW"), and in all of these cases the train was traveling in the eastbound direction (TRNDIR="East").

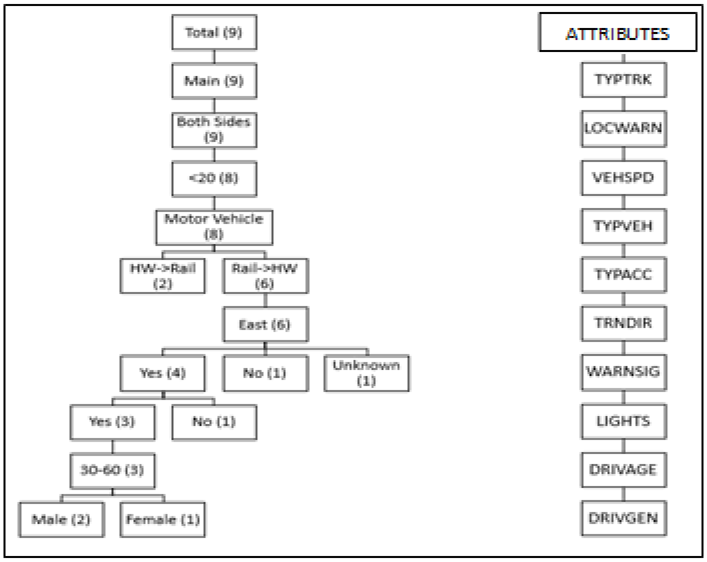


Figure 3: Tree using Dynamic Method A for crossing 173887G

Although this is an improvement on the Static Method, which uses pre-determined hierarchy, it is possible that the resulting tree may not identify a main branch with the highest possible number of accidents (and therefore, with the most prominent trend).

2.3.2 Method B (Nested Sorting)

As an improvement over Method A, Method B includes a stepwise procedure to decide the hierarchy for the attributes. The hierarchy of an attribute is decided based on the accidents in the main branch of the preceding level. This procedure is also simple to implement and visualize and is described as follows:

1. The highest-ranking attribute is selected as the attribute at the top level of the tree using the procedure described in Method A.
2. The accidents in the largest subcategory in the attribute selected in the previous step are further divided into sub-categories using the attributes which are not selected yet.
3. The attribute which gives the highest concentration of accidents in a subcategory is selected as 2nd attribute in the hierarchy.
4. Steps 2 and 3 are repeated until all the attributes are selected and the dynamic tree is formed.

This procedure is able to keep a trend in the main branch that is likely to be the most prominent trend for the selected set of accidents, and thus it is more likely to highlight common attributes of the accidents occurring at the crossing or a corridor.

To explain the prioritizing Method B, consider the hypothetical crossing from **Table 3**. Attribute A1 would be selected at the top of the hierarchy as the subcategory A11 holds the highest number of accidents out of all the subcategories in all the attributes. Out of the remaining two attributes, A2 is able to cluster a higher number of accidents than A3 can. Therefore, A2 is selected as the second attribute in the hierarchy. The tree obtained in this procedure for the hypothetical crossing is shown in **Figure 4**. Note that this procedure is able to keep a higher number of accidents in the main branch by placing attribute A2 at a higher level higher than the attribute A3 Larger accident clusters are expected to highlight stronger accidents trends, so this is a desired aspect of Method B.

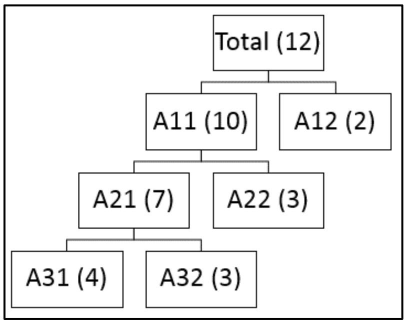


Figure 4: Tree for Hypothetical Crossing using Dynamic Tree Method B.

The tree formed using Method B for the same grade crossing described in the previous section for Method A is given in **Figure 5**. The VEHDIR recorded in the database were interpreted as: 1,3 (NB) and 2,4(SB), and TRNDIR recorded in the database were interpreted as: 2,3 (EB) and 1,4(WB). A comparison between the two trees shows that a greater number of accidents are maintained in the main branch using Method B than Method A. For example, in 7 of the accidents trains were traveling southbound and in 5 of them the train hit a motor vehicle that was in the stopped position with a driver between 30 and 60 years old.

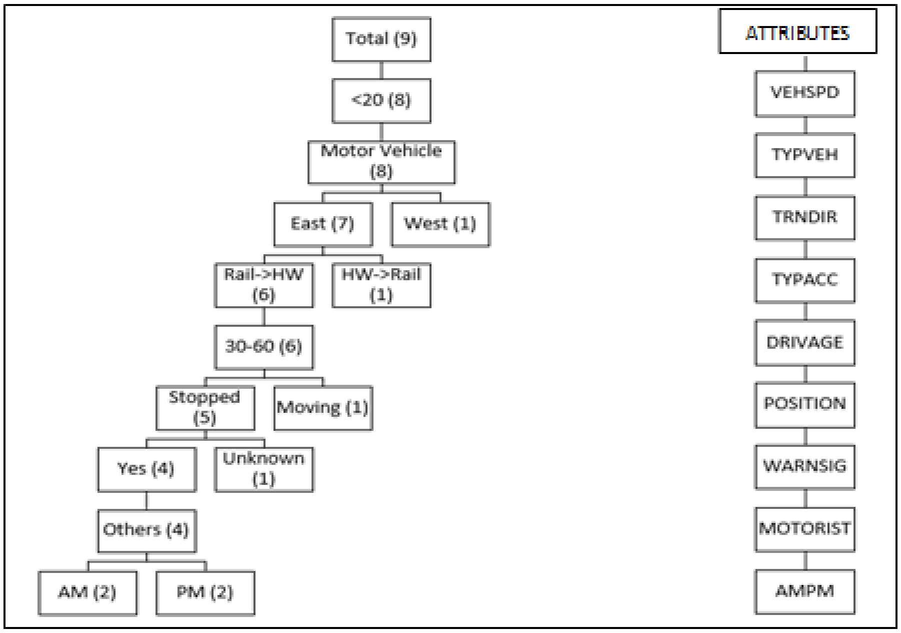


Figure 5: Tree for crossing 173887G using Method B

The Method B kept TRNDIR above TYPACC in the hierarchy. This is because TRNDIR could cluster more number of accidents (into East) than TYPACC could. Therefore, method B could reveal that higher number of EB trains were involved in accidents at the crossing and hence suggest that the directionality of the train could be a more important attribute to look into than the type of the accident.

One of the limitations of Method B is that it when there is a tie in ranking of the attributes, it requires a pre-determined order for the attributes. We used the order the attributes are listed in **Table 2** to resolve the tie issue. The order of the attributes in the table is the order in which they appear the in the FRA database. Another limitation of Method B is that the analysis of the attributes on the main branch takes away the focus of the investigator from the accidents in other branches. To address these limitations, a Modified Method B was developed.

2.3.3 Modified Method B (Modified Nested Sorting)

Modified Method B address the limitations of the Method B. The first limitation regarding the ties is resolved using historic data from the database. A variable score is defined in the computer program for the resolution of the tie. The score for an attribute is the sum of the number of accidents in the largest subcategory of that attribute in all the crossings present in the database analyzed. A higher score for an attribute is an indication of that attribute being able to cluster more accidents together.

To understand the logic used to break the ties, consider a database with 10 locations and three attributes, each with two subdivisions, as shown in **Table 4**.

Table 4: Hypothetical Database to Explain the Procedure to Handle Ties

Locations	No of Accidents at location	Attribute A1		Attribute A2		Attribute A3	
		A11	A12	A21	A22	A31	A32
1	10	10	0	8	2	6	4
2	5	5	0	4	1	3	2
3	3	2	1	2	1	1	2
4	1	1	0	1	0	1	0
5	7	6	1	4	3	3	4

6	3	2	1	2	1	1	2
7	6	6	0	4	2	3	3
8	1	1	0	1	0	1	0
9	1	0	1	1	0	1	0
10	1	1	0	0	1	1	0

The score of attribute A1 is calculated as the sum of the largest subcategories in all the locations in the database. Thus, for A1 this number is the sum of 10, 5, 2, 1, 6, 2, 6, 1, 1 and 1 equaling 35. The scores of attributes A2 and A3 are calculated similarly, totaling 28 and 24, respectively. An attribute with a higher score will appear in the tree above an attribute with a lower score in case of a tie.

The second limitation of Method B was eliminated by determining the division of the total number of accidents at the crossing into sub-categories for each attribute considered. This is done so that the accidents which do not appear on the main branch of the tree are also considered. All the accidents appear as the crossing cluster in the result as shown in **Figure 6.**

An additional improvement to the accident analysis is the detection of unexpected clustering of accidents. The expected number of accidents in a node is determined using the database as the ratio of the number of accidents in the subcategory to the total number of accidents available in the database. Any overrepresentation or underrepresentation of accidents is detected by comparing the number of accidents occurred with the number of accidents expected. A node is said to be over represented if the number of accidents in the node exceeds the expected value by 50%, although this threshold is a user-defined value. Any node which shows over-representation is automatically flagged in the crossing cluster.

A stepwise procedure for Modified Method B is given below:

1. The highest-ranking attribute is selected using the procedure described in Method A
2. The accidents in the largest sub category in the attribute selected in the previous step are further divided into sub-categories using the attributes which are not selected yet.
3. The attribute which gives the highest concentration of accidents in a subcategory is selected as the following attribute in the hierarchy.
4. If two or more attributes gives the same number of accidents clustered into a subcategory, the score of the attribute is used to break this tie.

5. The procedure is repeated until all the attributes are selected and the dynamic tree is formed.
6. The crossing cluster for each attribute is calculated for each attribute and printed along the side of the tree, noting flags due to clustering overrepresentation

The analysis of accidents at crossing 173887G using the Modified Method B is given in **Figure 6**, from which the following facts are highlighted: Almost all the accidents (7 out of 9) involved eastbound trains traveling at low speeds (lower than 20mph). In 5 of such accidents vehicles were stopped at the crossing, and 4 of them were traveling southbound. The action of these drivers was described as stopping at the crossing before the gates were lowered (this was supported with the narrative in the accident record part of the FRA database), which could provide an indication of the circumstances of these accidents.

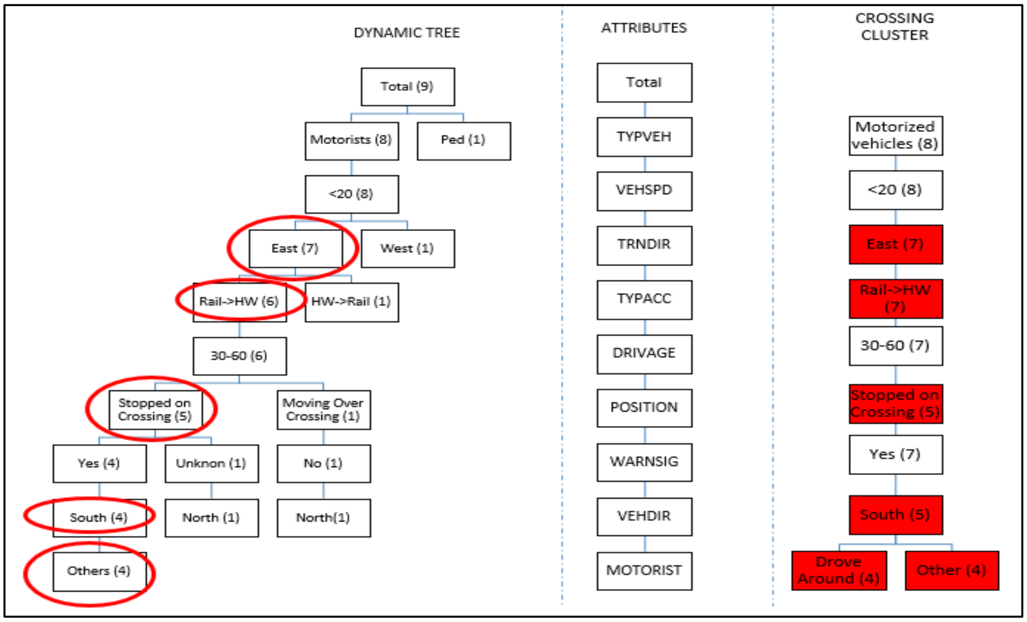


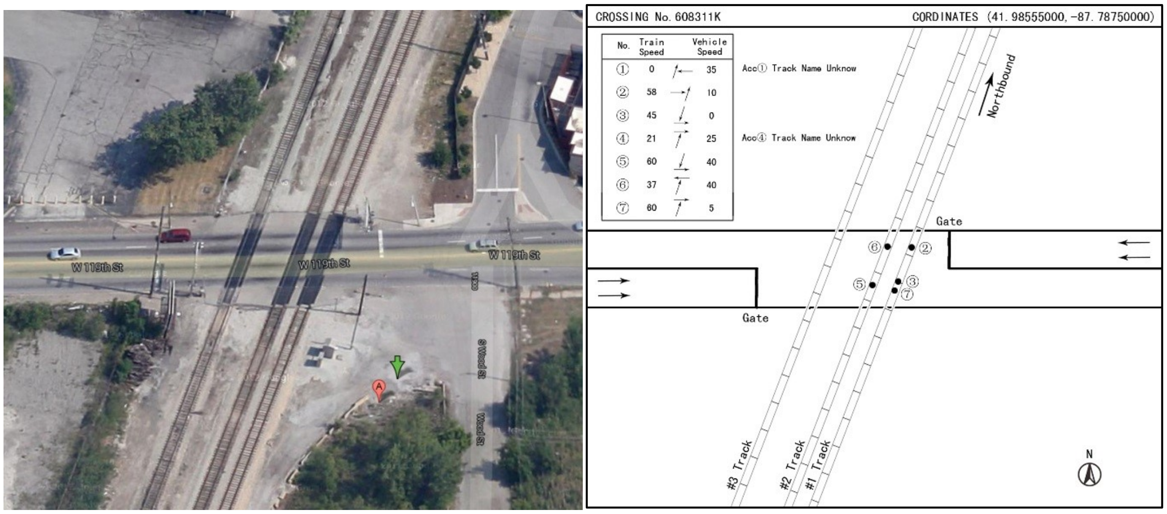
Figure 6: Modified Method B for Crossing 173887G.

2.4 STATIC AND DYNAMIC METHODS USING RECENT DATA FROM THREE STATES

In this section, we used more recent data (2005-2014) from Illinois and applied the Static and Dynamic Methods. To validate the approach and show it the methods are not restricted to Illinois data, we used data sets from California and Indiana, as described later in this section. In addition, considering crossings from multiple states provide access to data for an increased sample of crossings with high accident frequencies. The tree developed using the Static Method is compared to the tree developed using Modified Method B for five different crossings. Three crossings from California and two crossings from Indiana with high accident frequencies between the years 2005 to 2014 were considered for this analysis.

2.4.1 Selected Crossings from Illinois (Recent Data)

The crossing selected from Illinois was the one that had the second highest number of accidents (in 2003-2012 as well as in 2005-2014 time periods). It should be noted that, the location with the highest number of accident was analyzed in the previous section. The selected crossing had an ID of 608311K and had 7 accidents between 2005 and 2014. It is located at the crossing of rail line and W 119th Street in Blue Island. The map and sketch of the crossing is given in Figure 7.



CROSSING No. 608311K

CORDINATES (41. 98555000, -87. 78750000)

No.	Train Speed	Vehicle Speed
①	0	35
②	58	10
③	45	0
④	21	25
⑤	60	40
⑥	37	40
⑦	60	5

Figure 7: Map and Sketch of 608311K

The VEHDIR recorded in the database were interpreted as: 2,3 (EB) and 1,4(WB) and TRNDIR recorded in the database were interpreted as: 1,3 (NB) and 2,4(SB). As shown in **Figures 8 and 9**, more information is extracted from the dynamic tree using Modified Method B as compared to the static tree. Five accidents involved the highway user going around the crossing was seen both in the static tree as well the tree using Modified Method B. The dynamic tree also reveals that 4 out of 7 accidents occurred during the AM hours; and 4 out of 7 accidents occurred on FRA track class 1 giving us information about the train speed limits.

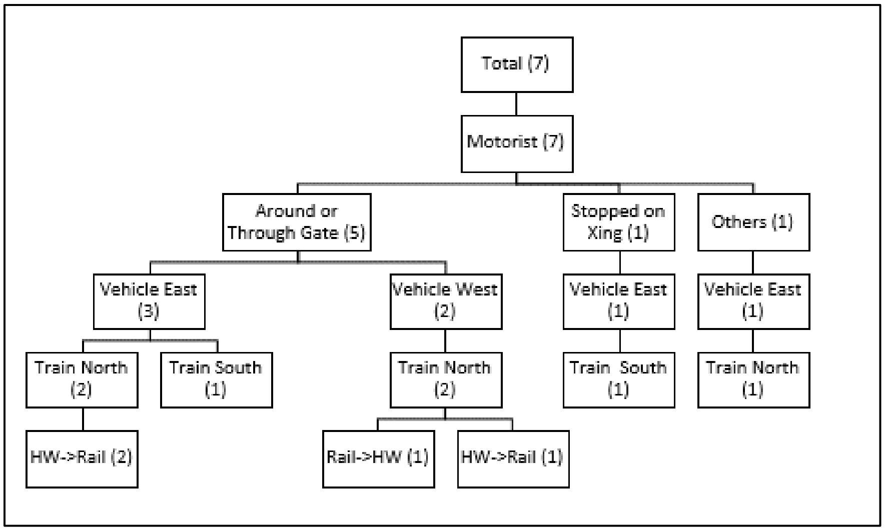


Figure 8: Tree for crossing at 608311K using Static Method

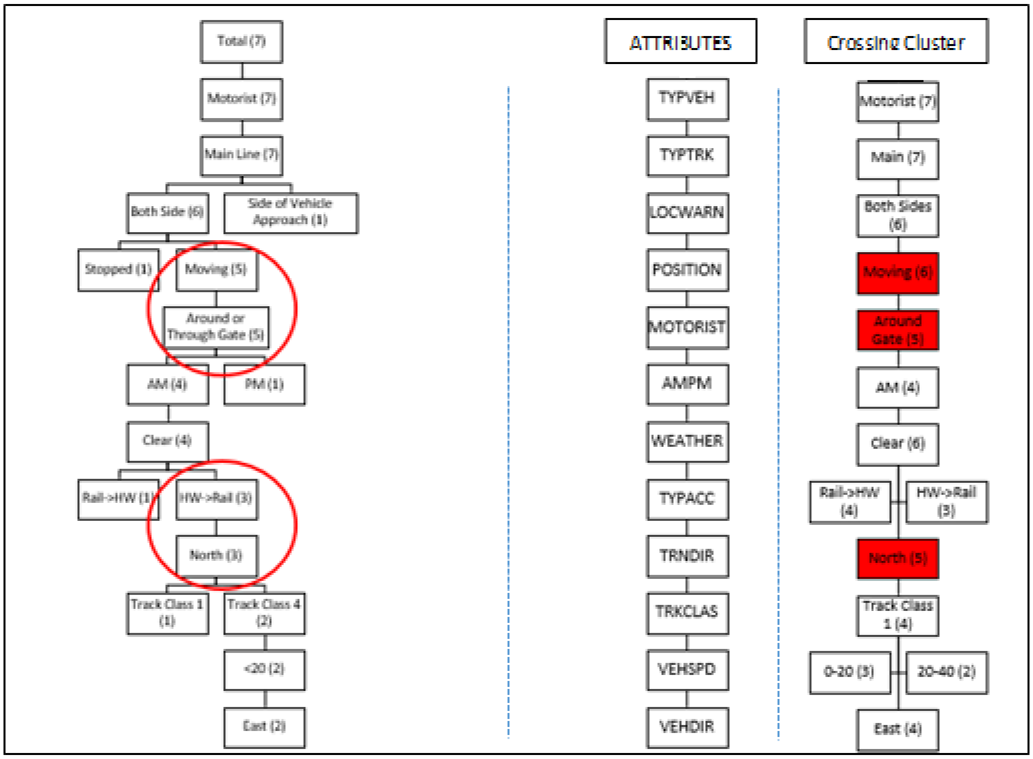


Figure 9: Dynamic Tree using Modified Method B for 608311K

These two examples are illustrations as to why the new method using more attributes and a data driven approach to prioritize such attributes provides a clearer picture of the commonalities of the accidents in the analysis.

2.4.2 Selected Crossings from California

The locations considered from California were 811479J, 028380R and 026517B with 9, 6 and 6 accidents, respectively, in the 10 year time period (2005-2014).

Crossing 811479J is located at Nogales St and UP rail line in Los Angeles, where 8 accidents involving motor vehicles occurred in the analysis period. The map and sketch for the crossing is shown in **Figure 10.**

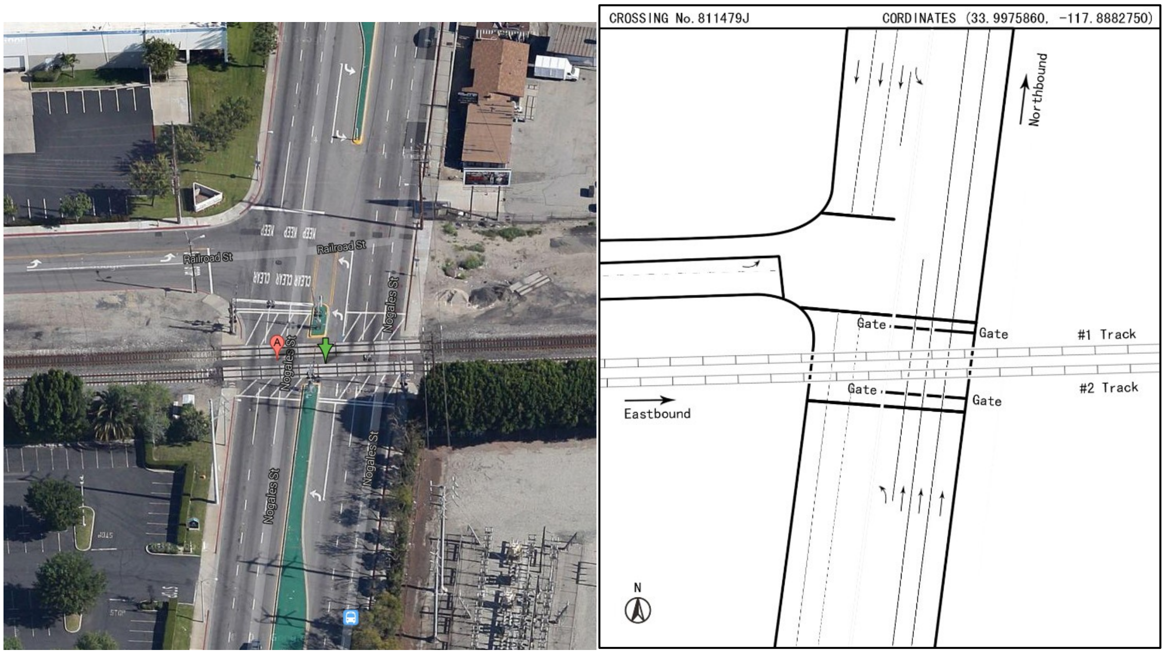


Figure 10: Map and Sketch for 811479J

For comparison purposes, the dynamic and the static trees for this crossing were created as described in the previous sections, and are shown in **Figures 11 and 12**. From the two figures a more in-depth perspective on the common attributes of the accidents can be obtained from the dynamic tree. From the static tree, we gathered that 8 accidents involved motorized vehicles, most of them were stopped on the crossing (5), and all of them involved southbound vehicles (in addition to the 3 accidents when vehicles were not stopped on the tracks). This information may be useful, but it was certainly enhanced by the dynamic tree. From the dynamic tree, we observed that not only all accidents involved a southbound vehicle, but also they occurred in the PM hours and the visibility was unobstructed, except for one case where vegetation obstruction was suspected (potentially an issue that can be fixed). Also, in all 8 cases the drivers were male, and in 3 occasions the vehicles were trapped and in the remained the vehicles were stopped (this information was taken from the crossing cluster on the right side). The ability of observing not only accidents on the main branch, but also the overall degree of concentration on a given subcategory by looking at the cluster may prove useful as in this case.

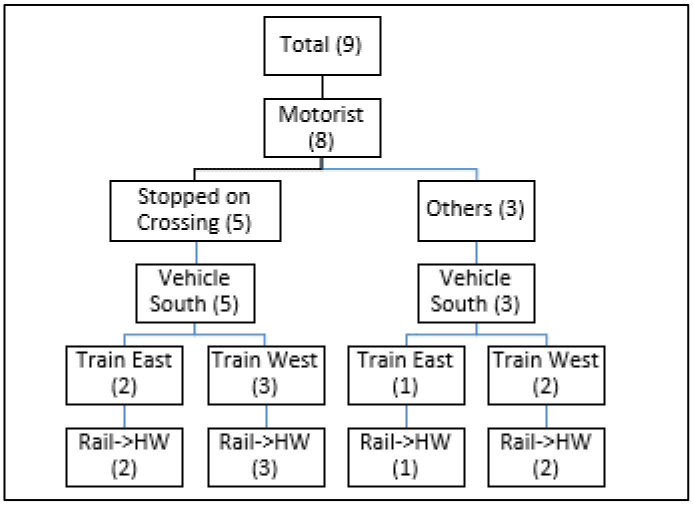


Figure 11: Static tree for 811479J

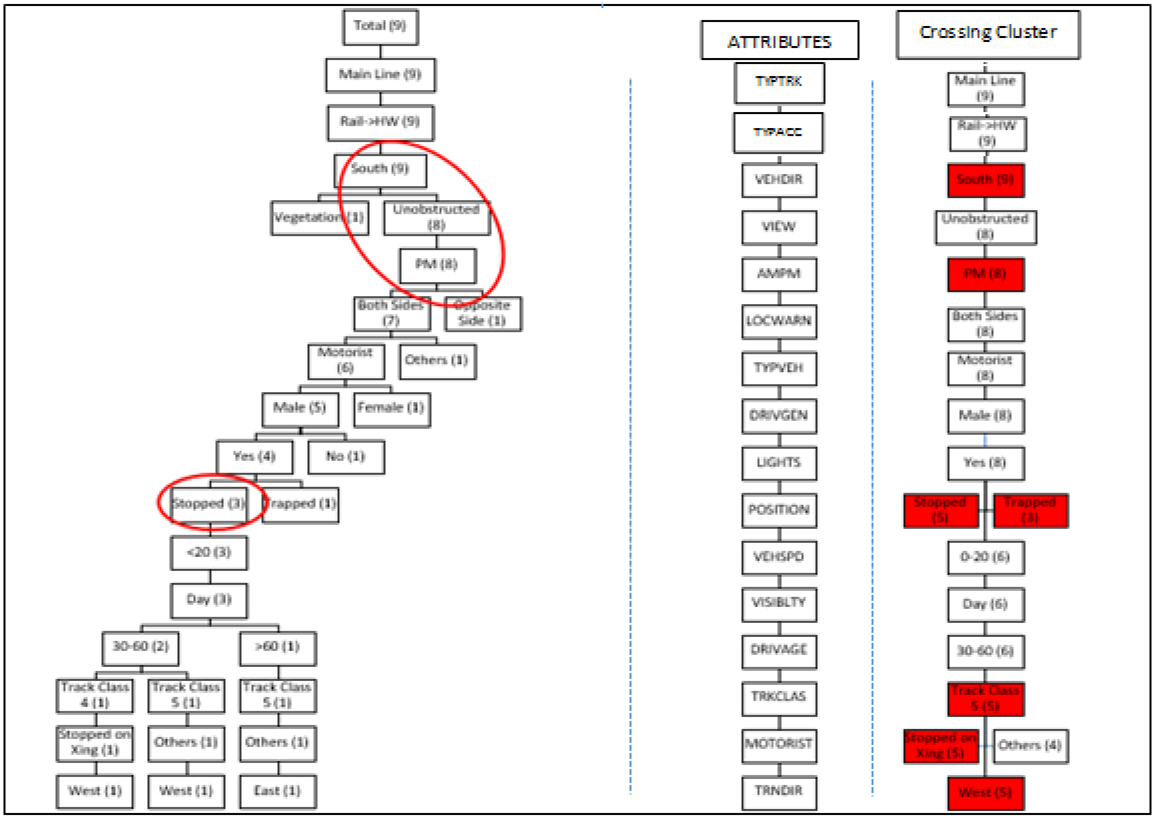


Figure 12: Dynamic tree using Modified Method B for 811479J

The second crossing analyzed was 028380R which is located at the crossing of Kratzmeyer Road and BNSF rail line in Bakersfield. Six accidents happened here in the 10 year time period (from 2005 to 2014). The map and sketch of the crossing are given in **Figure 13**.

Both the static and dynamic trees for this crossing were created and showed information that could be useful (**Figures 14 and 15**). Both analyses showed that 5 out of 6 accidents involved westbound vehicles and westbound trains, although this is easier to see in the dynamic tree. In addition, the dynamic tree provided valuable information that was not in the static tree, for example that 5 of the vehicles involved are truck, and 4 of the accidents involved trains traveling at more than 60 mph during daytime and in two of them the view could have been obstructed by other standing RR equipment. It is easy to see that this additional information can be of interest for diagnostic teams.

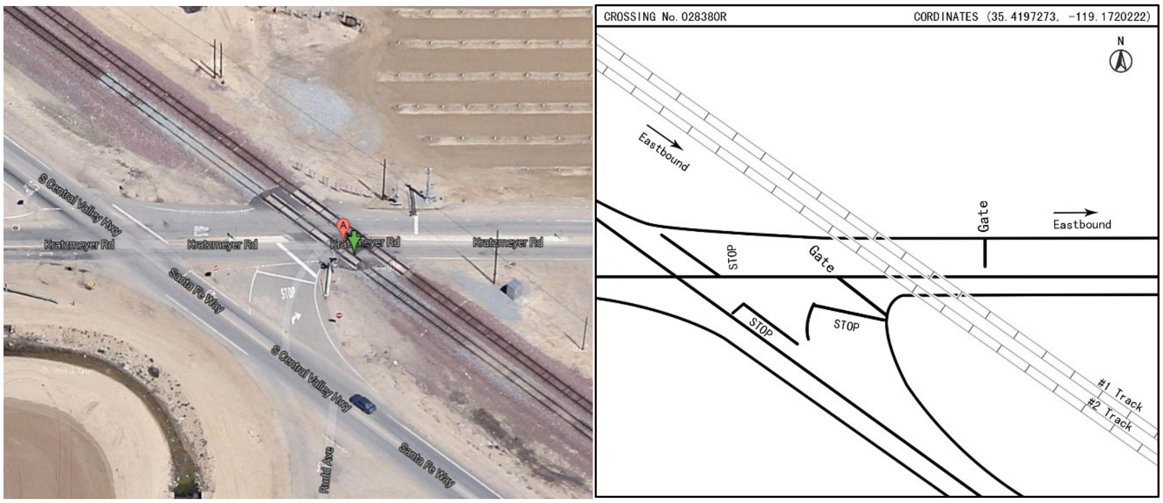


Figure 13: Map and Sketch of 028380R

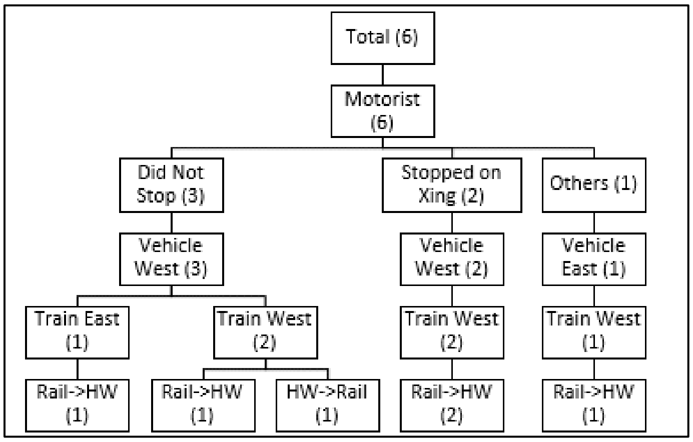


Figure 14: Static Tree for 028380R

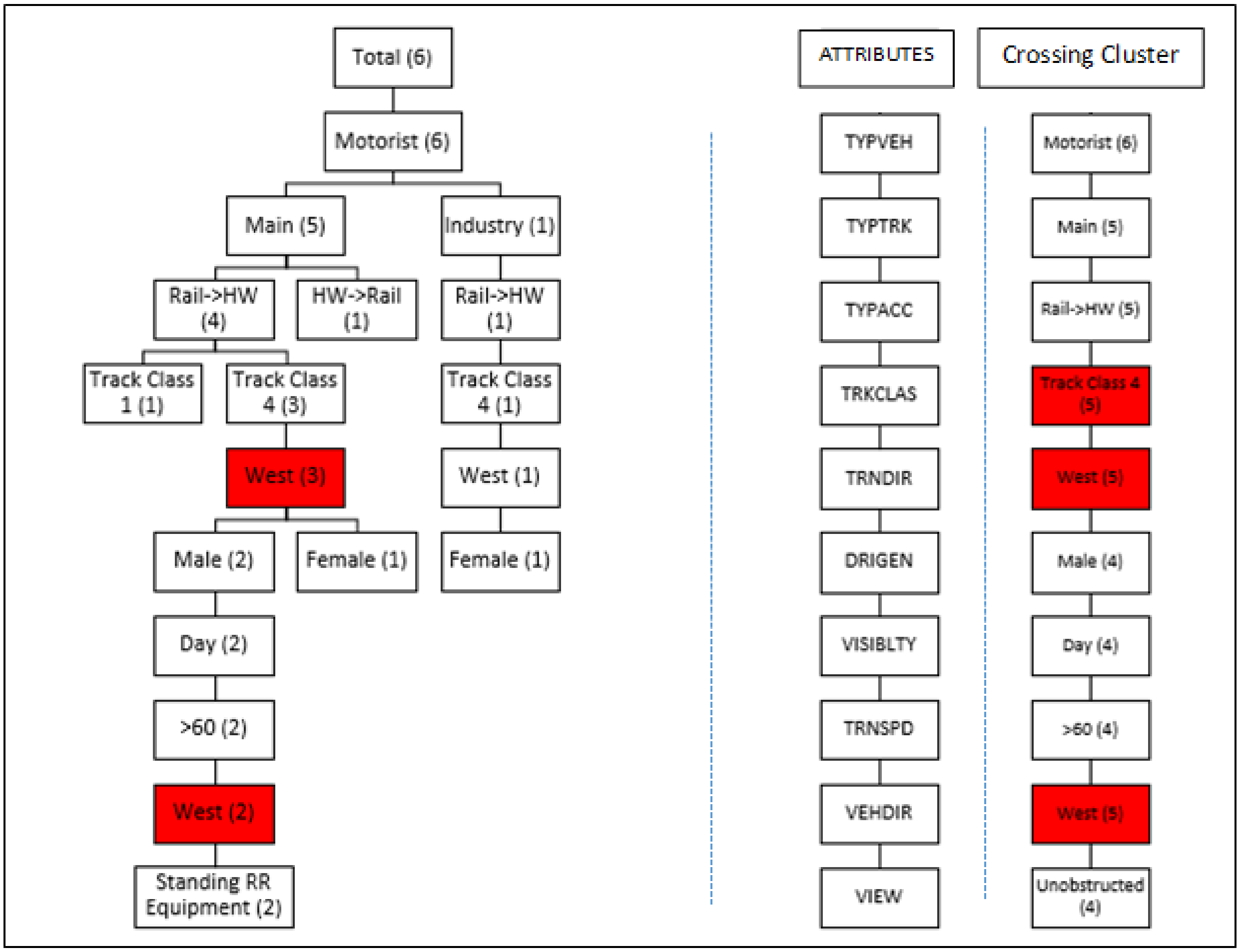


Figure 15: Tree using Modified Method B for 028380R

A third crossing from California was also considered for the analysis (ID= 026517B). It is located at the crossing of Magnolia Ave and BNSF rail line in San Bernardino, where six accidents were reported from 2005 to 2014. The map and sketch for the crossing are given in **Figure 16**.

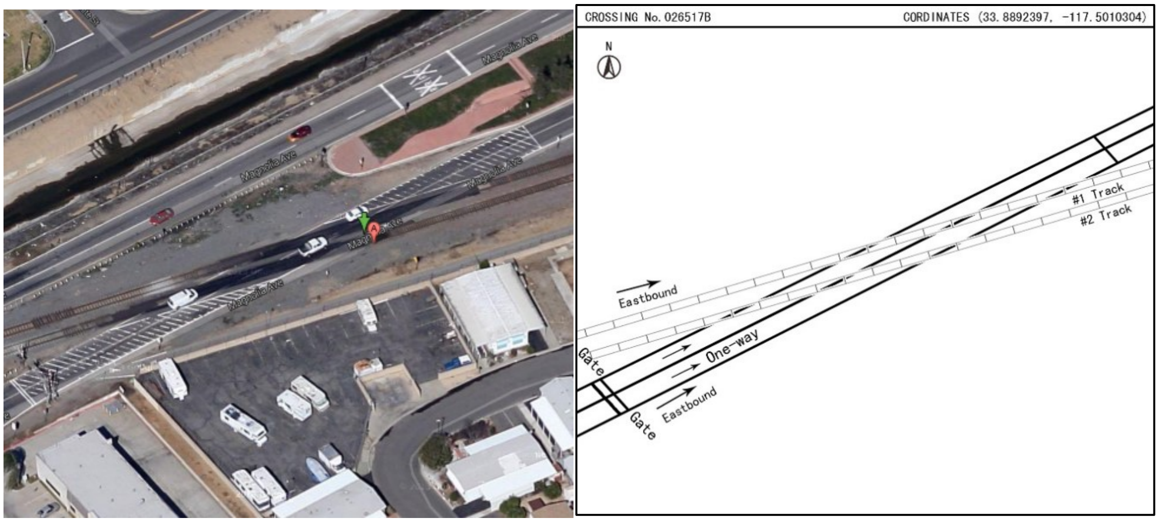


Figure 16: Map and Sketch of 026517B

From both the static and dynamic trees (**Figure 17 and 18**) it is observed that all accidents involved eastbound vehicles and eastbound trains (although only 5 could be seen in the static tree), and in 5 occasions the vehicles were stopped on the crossings. However, the dynamic tree also highlights that all accidents happened in the PM hours and in 5 cases the drivers were male. In this case, trends regarding the train directionality were very clear with both trees, which can arguably be related to the small angle between the traffic lanes and the tracks (vehicle trends were expected given that the roadway is one way). Eastbound is the train direction with reduced visibility form a driver's point of view.

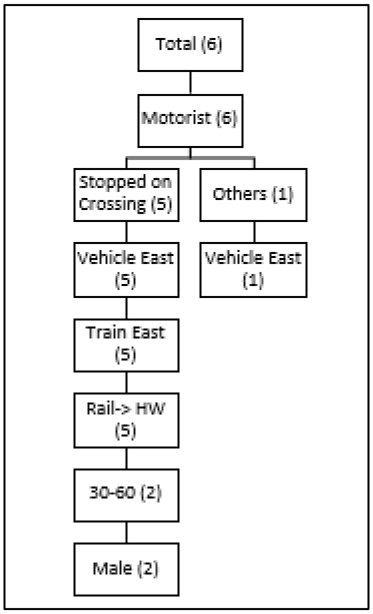


Figure 17: Static Tree for 026517B

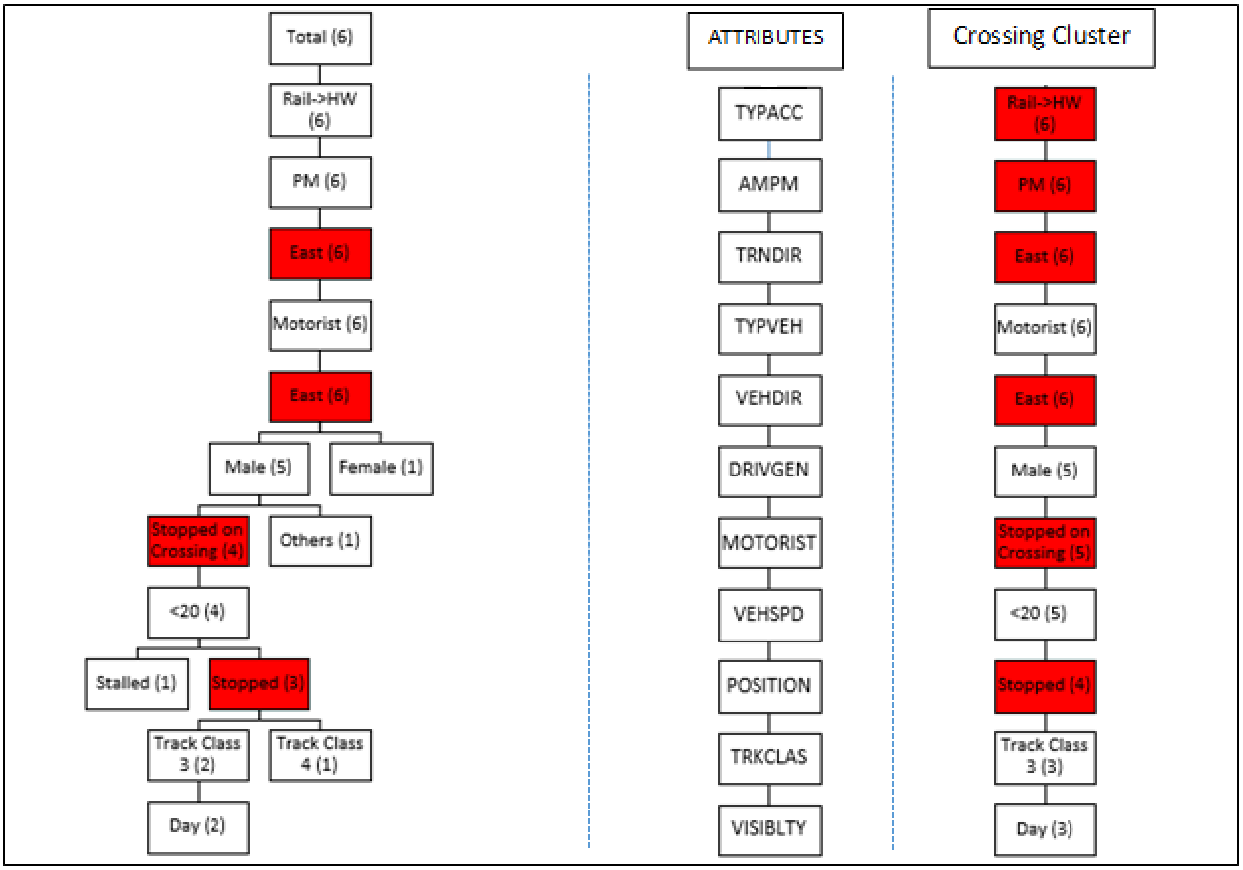


Figure 18: Tree using Modified Method B for 026517B

2.4.3 Selected Crossings from Indiana

The locations considered from Indiana were 522646H and 879204S with 6 and 15 accidents, respectively during the time period of 2005-2014.

Crossing 522646H is located at the intersection of Clark Rd and NS rail line in New Castle, providing entrance to Olympic Steel, Inc, an industrial supplier. Six accidents happened here in the 10 years from 2005 to 2014. For the analysis, TRNDIR coded as 1 and 4 were considered westbound trains. The map and sketch of the crossing is shown in **Figure 19.**

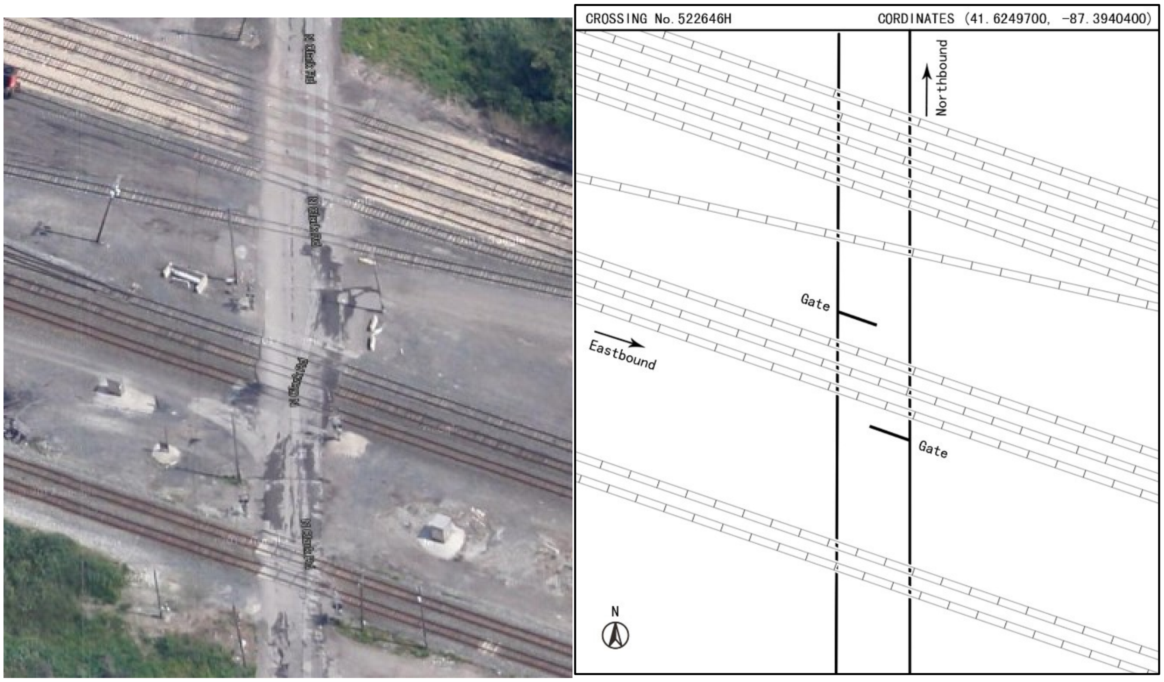


Figure 19: Map and sketch of crossings 522646H

In this crossing, as shown in the static and dynamic trees (**Figures 20 and 21**), all accidents involved southbound vehicles that were stopped on the crossing, and in 5 of the 6 occasions trains were traveling eastbound during the PM hours. Given the clear trends, both the static and the dynamic trees provide a similar picture, with some added details by the dynamic tree.

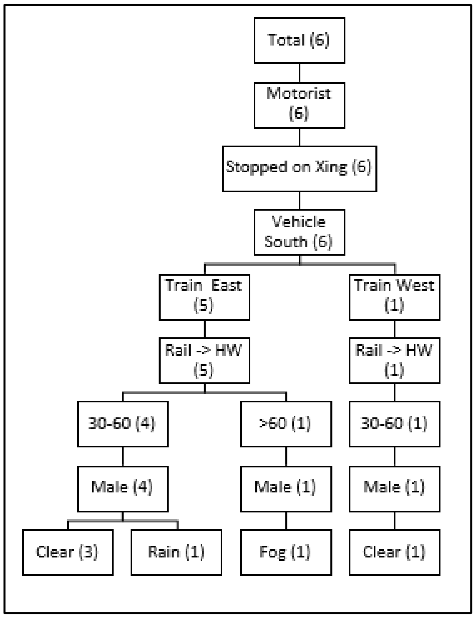


Figure 20: Tree using Static method for 522646H

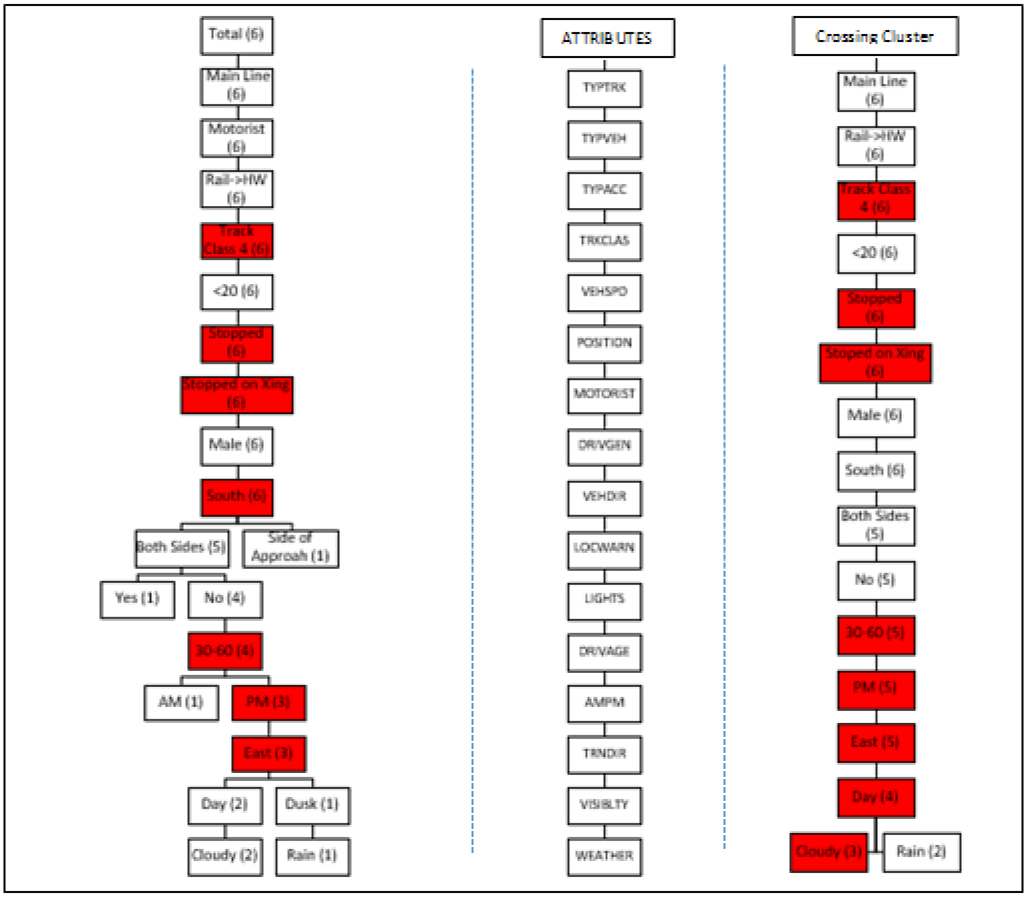


Figure 21: Tree using modified method B for 522646H

The second crossing considered in Indiana was 879204S. This crossing is located at the intersection of Mcgalliard Rd and NS rail line in the town of Muncie. Fifteen accidents happened here in the 10 years from 2005 to 2014. The map and sketch of the crossing are given in **Figure 22.**

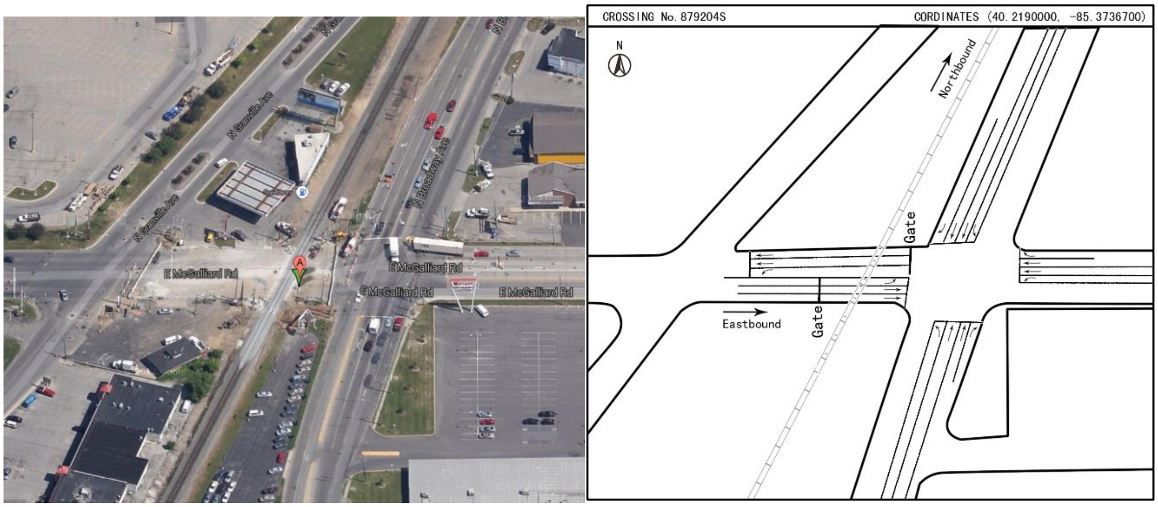


Figure 22: Map and Sketch of 879204S

This crossing is an example where the static tree provided some informative about direction of stopped vehicles than the dynamic tree did not present it as clearly. VEHDIR coded as 2 and 3 are eastbound while VEHDIR coded as 1 and 4 are westbound. TRNDIR coded as 1 and 3 are northbound trains while others are southbound. The static tree is shown in **Figure 23** and pictures the accidents as distributed mostly between those where the vehicles did not stop before the crossing (8 out of 15) and those where vehicles were actually stopped at the crossing when the accident happened (5 out of 15). Accidents do not show any specific trend except for the direction of the stopped vehicles, with 5 out of 6 stopped while traveling eastbound. On the other hand, dynamic tree presented information that was not revealed with the static tree: for example 14 out of 15 accidents were in PM hours, and that 8 out of 15 accidents involved vehicles that did not stop.

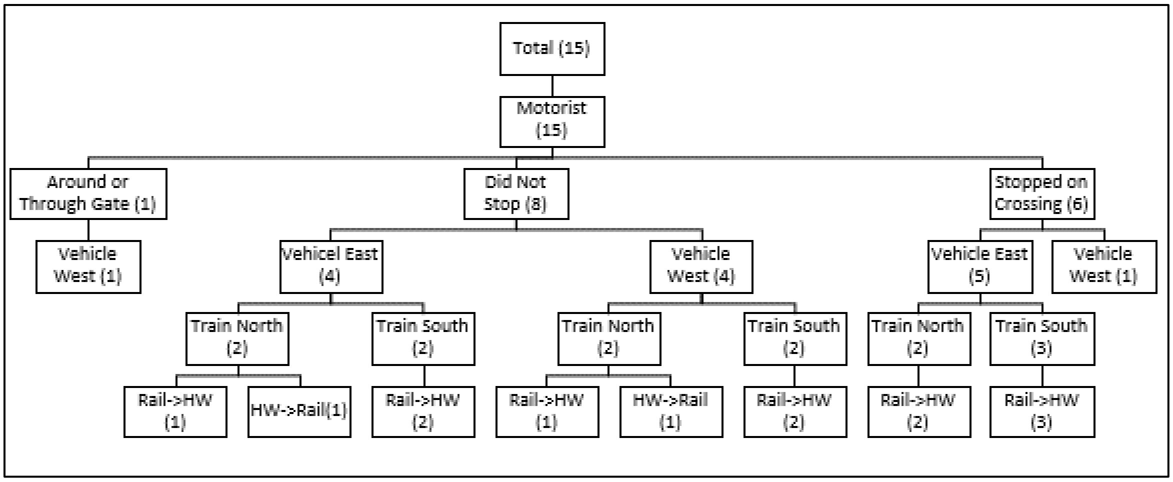


Figure 23: Tree using modified method B for 879204S

On the other hand, the dynamic tree is shown in **Figure 24**. Trends in the dynamic tree were centered on attributes most common to the 15 accidents, highlighting that in 14 cases the accidents happened in the PM hours, the train speed was between 20 and 40 miles per hour (i.e. trains were not almost stopped), and vehicle speeds were lower than 20 miles per hour in 10 of those occasions. However, other attributes that were not common to most of the 15 accidents, such as train and vehicle directions were left out of the top levels of the tree and were not visible.

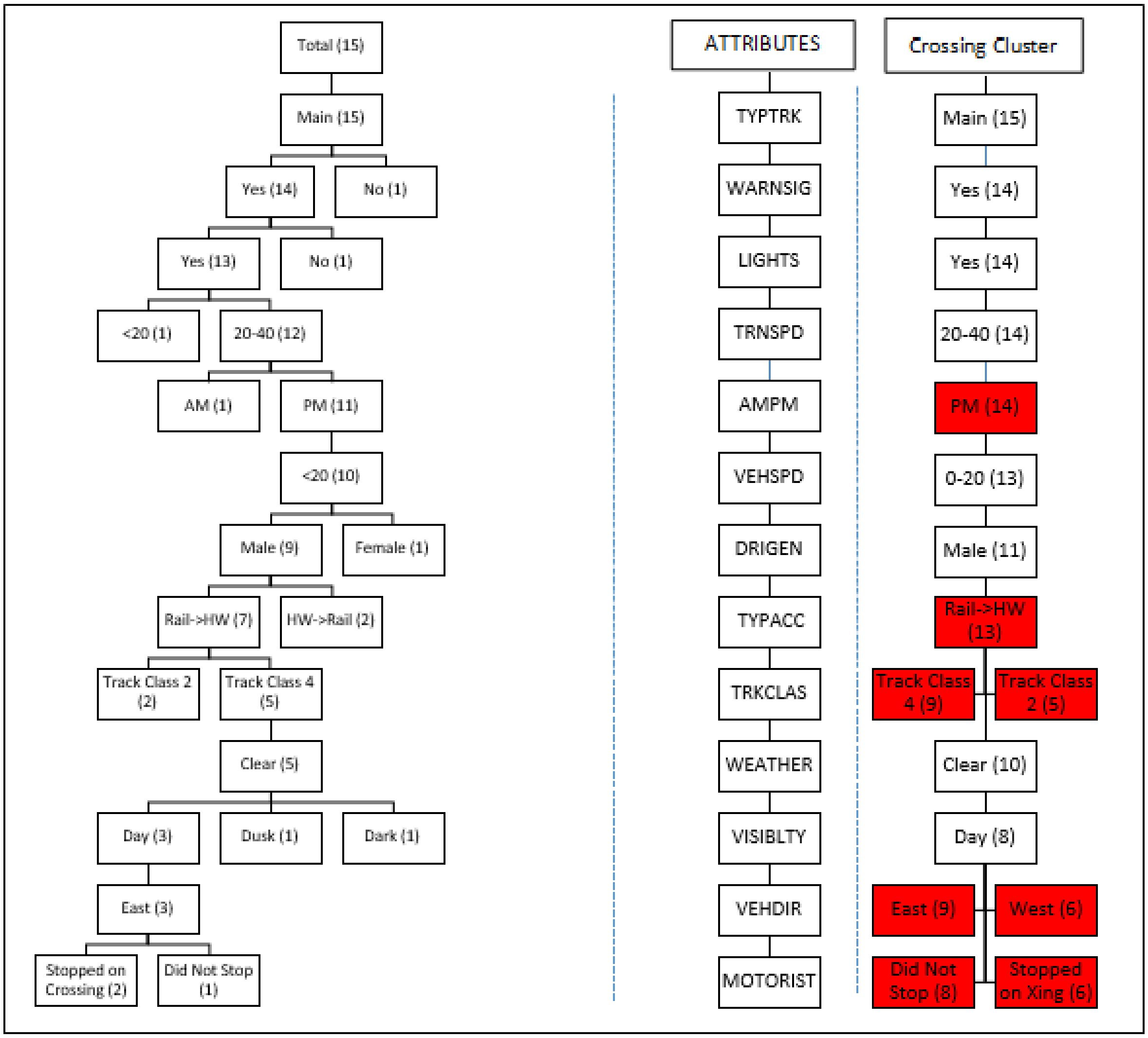


Figure 24: Tree using modified method B for 879204S

Thus, the dynamic tree did not capture a trend that was highlighted in the static tree, related to the travel direction of the vehicles stopped on the tracks when the accidents happened. It rather hinted at towards further exploration of the 6 accidents involving vehicles stopped on the crossing as observed from the crossing cluster, but the tree stopped at that level without highlighting the vehicles' travel direction. In this case, further analysis may be required because the trend isn't present in the main branch of the tree. This is discussed in the next section as a special case.

2.5 HANDLING SPECIAL CASES USING THE MODIFIED METHOD B

As mentioned above, the Modified Method B identifies over representation (outliers) based on the expectation value calculated from the database, and highlights them in color red on the accident cluster. If accidents in any subcategory are grouped such that they exceed their expected value by more than 1.50 times, it is considered over representation (outlier).

Based on extended testing and analysis of several crossings, it was noted that outliers could be interesting starting points to start a new dynamic tree when the number of accidents grouped is in the order of more than 4. For example, in the second crossing selected from Indiana in the previous section, the last level of the dynamic tree indicated that vehicles were stopped at the crossing in six of the accidents and this grouping was higher than expected (highlighted in red). Therefore, this grouping would be a candidate for a new dynamic tree itself.

If this recommendation is followed, a new tree such as the one shown in **Figure 25** will be obtained. In the new tree, trends that were initially not observed are now clear, emphasizing that out of the 6 accidents, all of them occurred in the PM hours and involved eastbound vehicles stopped in the tracks when trains traveling at 20-40 miles per hour hit them.

This extension of the Modified Method B becomes more useful as the number of crossings analyzed increases, for example for a corridor analysis.

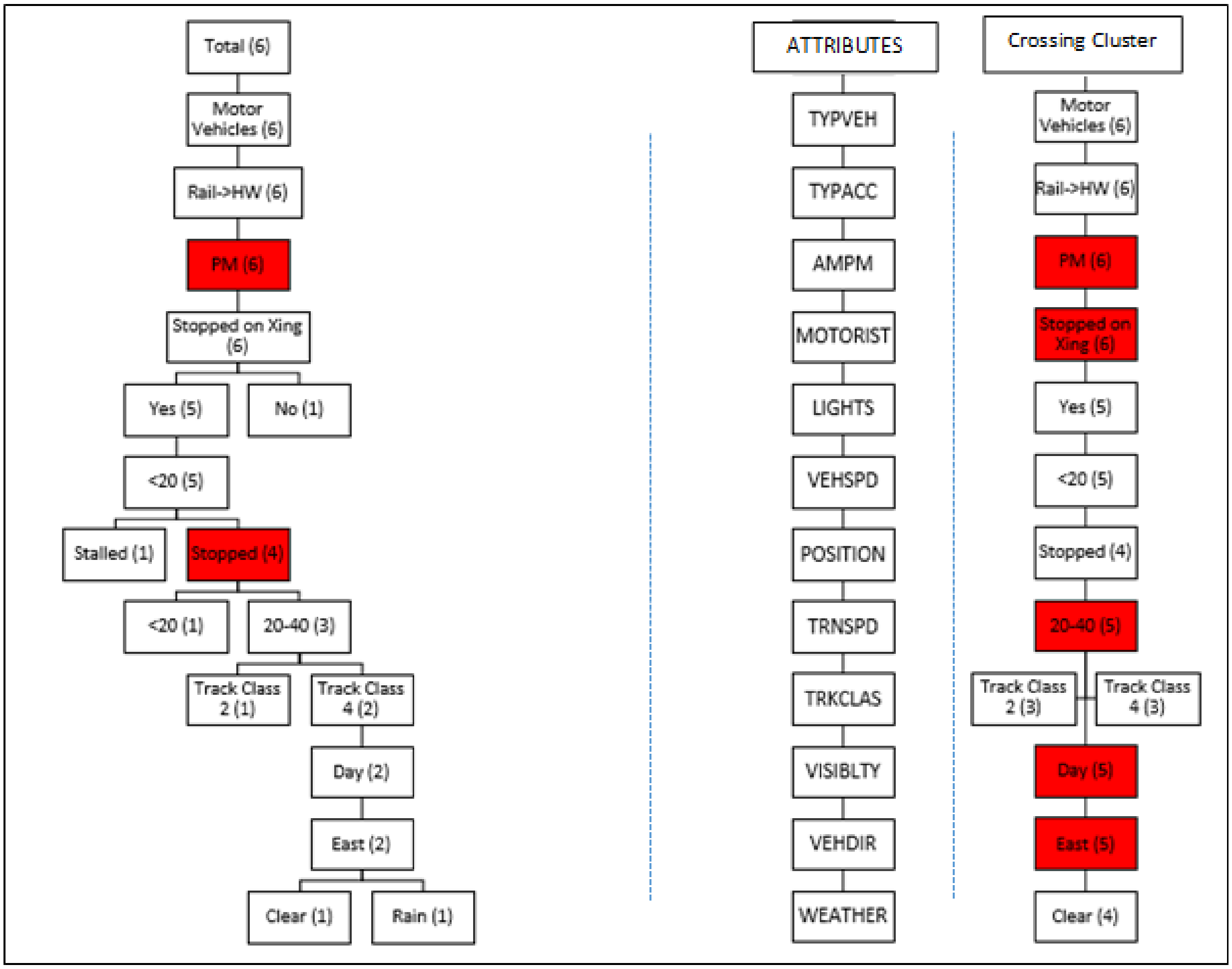


Figure 25: Six Accidents where Motorists were stopped on the Crossing

2.6 DYNAMIC TREE FOR CORRIDORS

The dynamic tree using Modified Method B could be used not only at a specific crossing, but also for a group of locations simultaneously, for example to analyze crossings along a corridor. An alternative use of the procedure could also involve multiple crossings with particular characteristics of interest, for example to study commonalities of locations with a single accident.

The procedure for crossings along a corridor is expected to follow similar steps as described for a single location. To illustrate this process, a group of crossings was selected along the Northeast Illinois Regional Commuter Corridor. Among the 8 selected crossings, a total of 23 accidents were recorded over 10 years spanning from 2002 to 2011. The general location of the corridor and the selected crossings is shown in **Figure 26**.

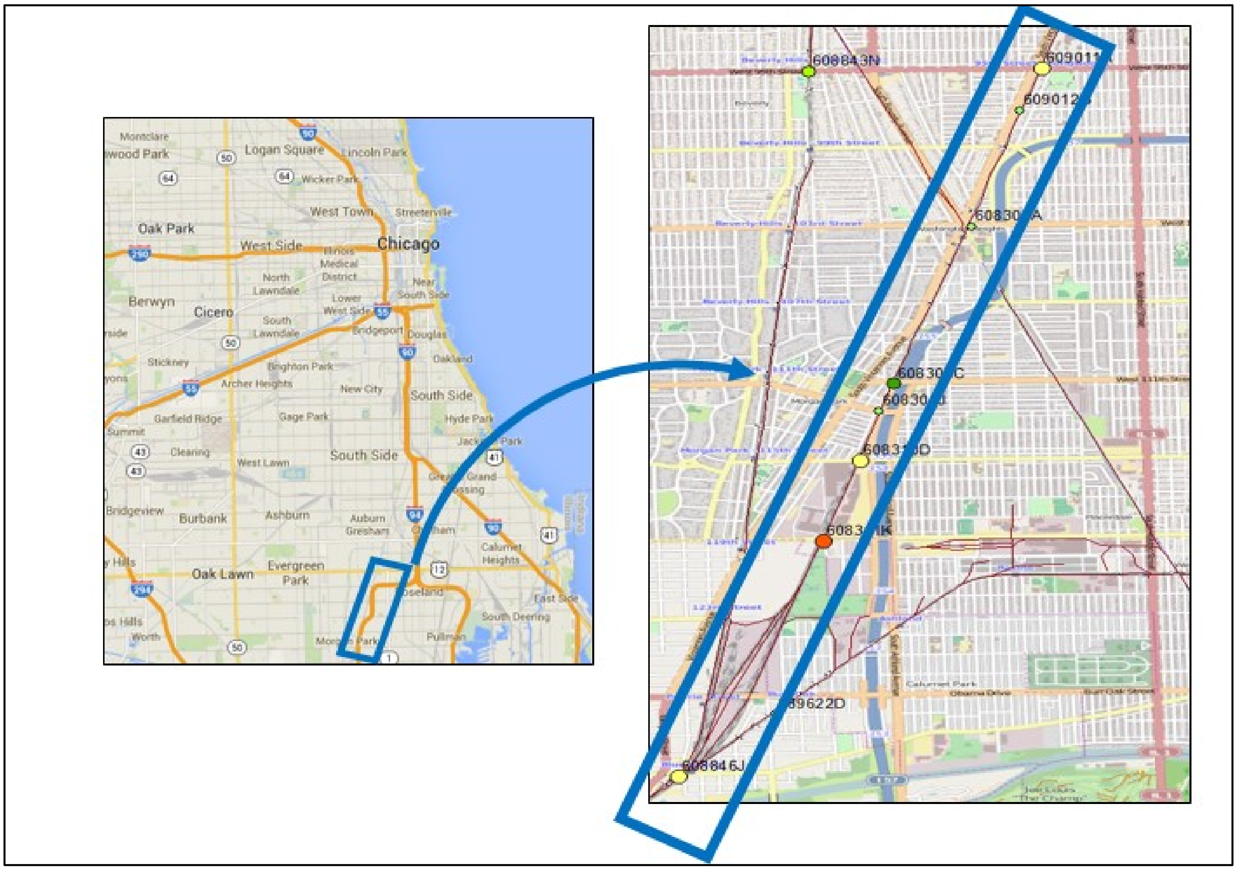


Figure 26: Selected Crossings along Corridor1 near Chicago

First, a tree using the Static Method was created for the corridor. It is noted that VEHDIR with values 1, 3 were code as eastbound while the others were considered westbound, and TRNDIR values 1, 3 were coded as northbound while 2, 3 were southbound. The static tree is shown in **Figure 27** and indicated that accidents were fairly distributed among the pre-selected attributes, with a big incidence of highway users driving around or through the gates when the train was approaching (15 out of 23) and 5 of the remaining cases with vehicles stopping at the crossing.

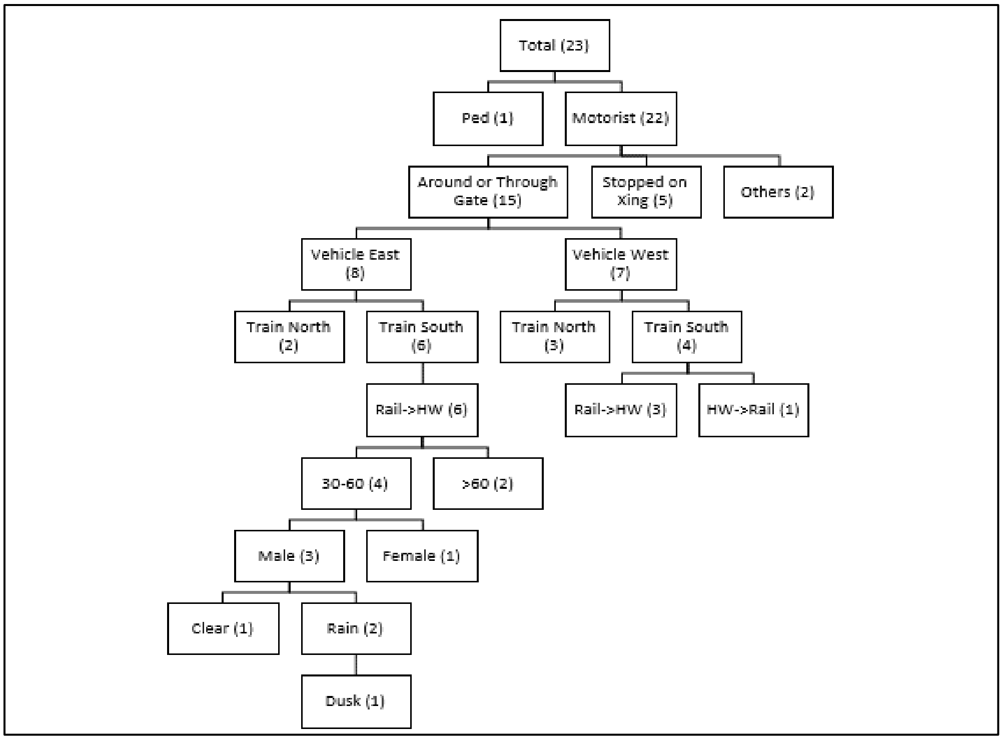


Figure 27: Tree for Corridor 1 using Static Method

On the other hand, the dynamic tree using modified method B is shown in **Figure 28**. From this tree, it is observed that in 19 of the 23 accidents trains struck highway users. Most of the trains involved were traveling southbound and the vehicles were traveling at low speeds (<20 mph) during the PM hours. Lastly, the most common maneuver of the highway user was to drive around the gates, as recorded in 15 of the 23 accidents. This information could be valuable at the corridor level and can point out to some issues that may require further investigation at the individual crossing level. For example, analysis at individual crossing with 2 or 3 accidents may not reveal the train directionality issue that the corridor level analysis revealed.

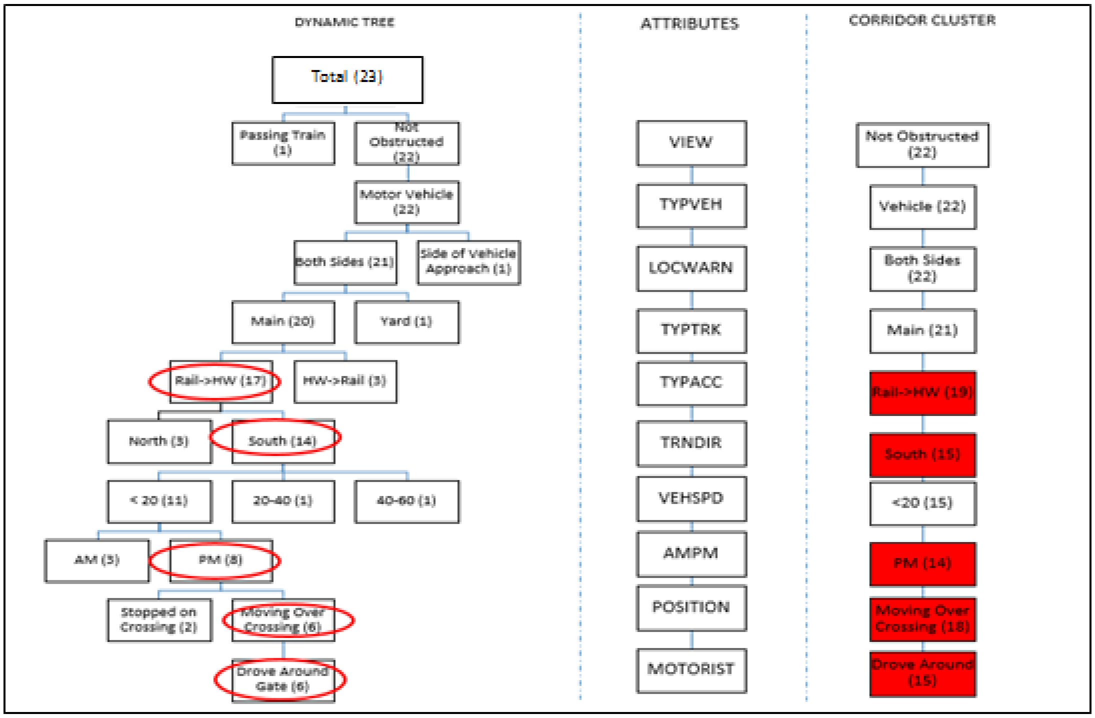


Figure 28: Dynamic Tree analysis for Corridor 1 using Modified Method B

A second corridor (Corridor2) analyzed is located in the Chicago area between Union Station and Aurora along a BNSF rail line and it is used by both passenger and freight trains. Corridor2 contains 8 crossings, where the trains can be easily identified to run eastbound and westbound (as opposed to northbound and southbound). Consequently, vehicles run in the northbound and southbound directions. A total of 19 accidents occurred at the 8 crossings together in the analysis period. The location of the crossings along the corridor is shown in **Figure 29**.

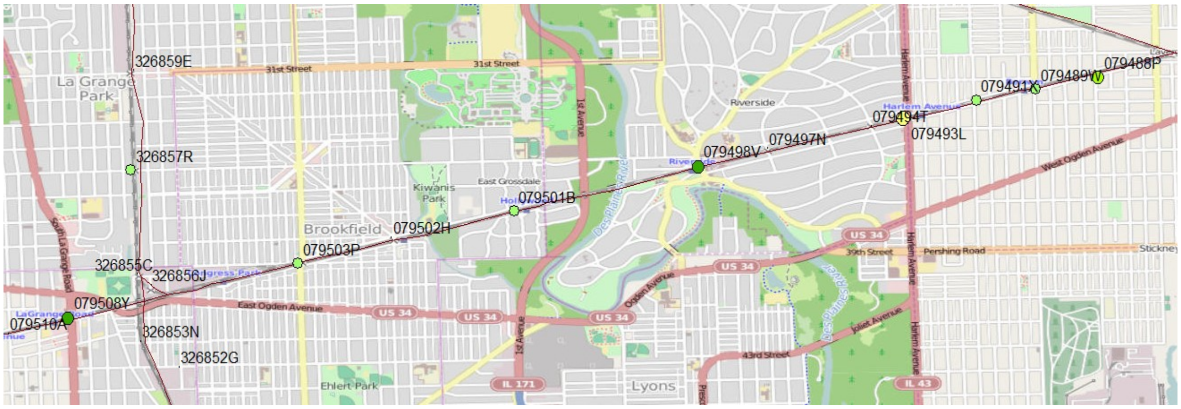


Figure 29: Selected Crossings along Corridor 2 near Chicago

Similar to Corridor 1, the static tree was build first for comparison with the dynamic tree (**Figure 30**). A first observation is the high number of accidents involving pedestrians compared to the crossings previously analyzed. Passenger trains in this corridor are part of a Metra commuter line that connects the suburbs with the city of Chicago, and thus pedestrian traffic is heavy along the area. However, no other trends stand out from the tree, which focused mainly on motorist accidents.

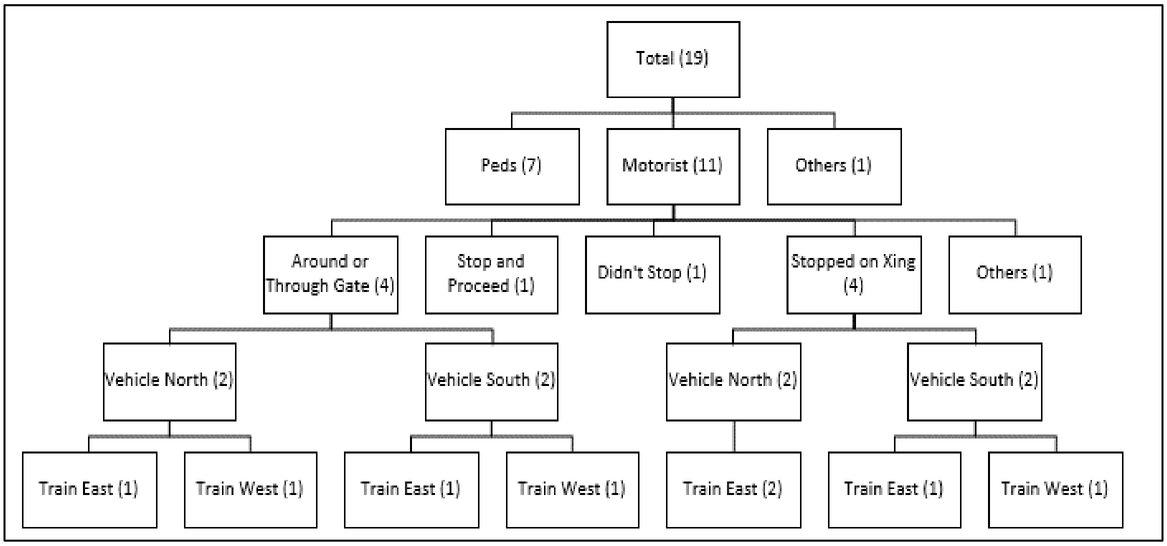


Figure 30: Tree for Corridor 2 using Static Method

Then, a dynamic tree using Modified Method B was also created for Corridor2, as shown in **Figure 31**. The main attributes highlighted by the tree were the type and time of accidents: trains hitting highway users constituted 17 out of the 19 accidents, and accident occurring daytime were 15 of the 19 accidents. A few levels below we observe the division between pedestrian and motorized accidents, but not conclusive trends. This observation indicates the need to analyze the motorized accidents from the pedestrian accidents. This idea was tested by building a new tree without pedestrian accidents, as seen in **Figure 32**. This approach did not lose any of the trends that were observed before. The separate analysis also revealed that 7 out 12 accidents happened in cloudy days and 7 accidents happened when the vehicle was moving over the crossing.

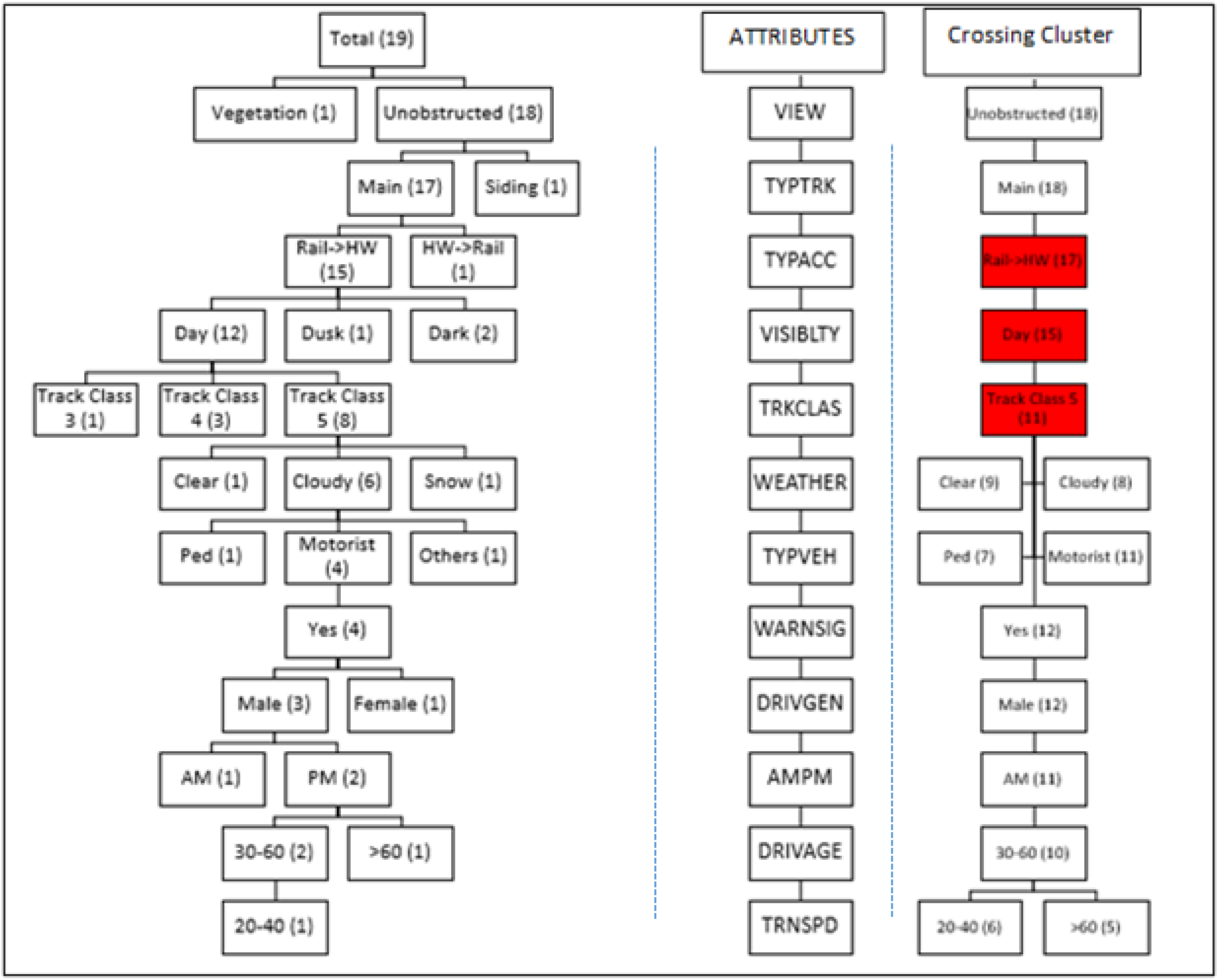


Figure 31: Dynamic Tree using Modified Method B for Corridor 2

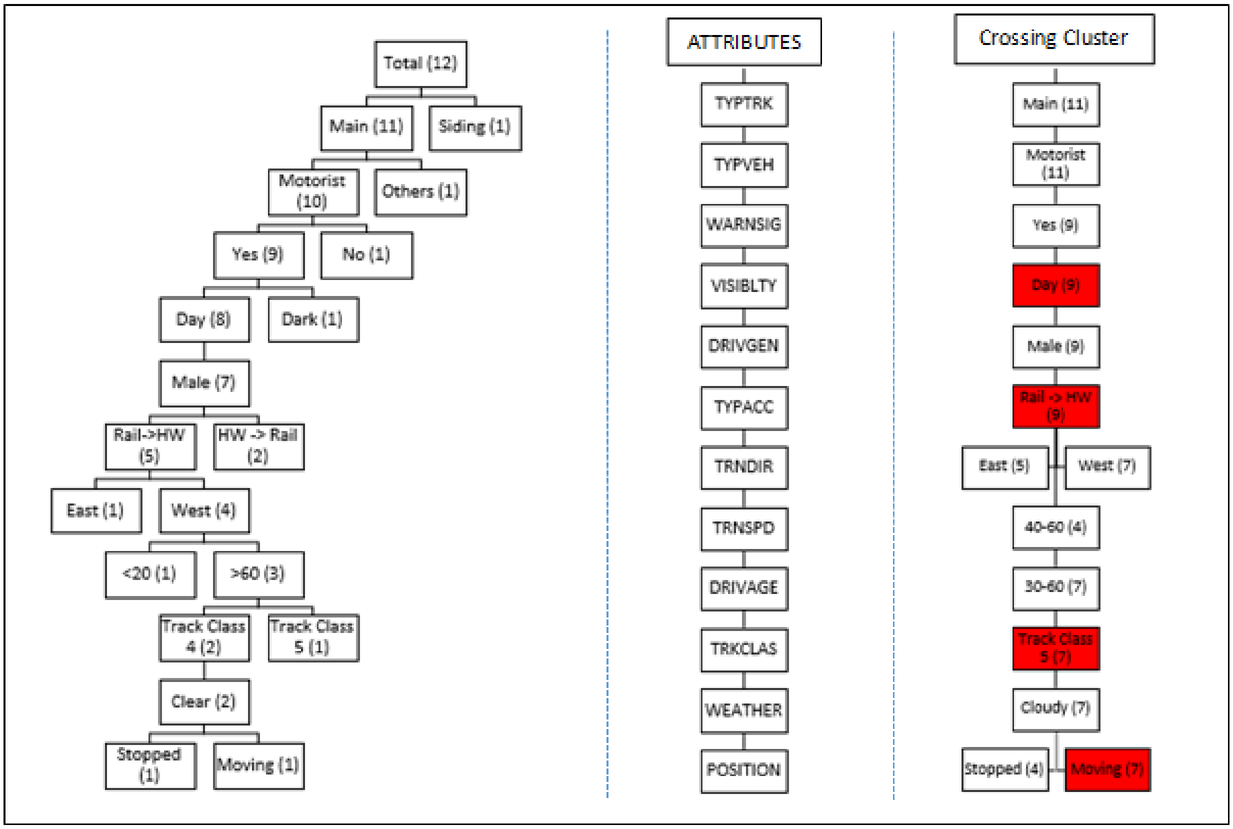


Figure 32: Dynamic tree using Modified Method B for Corridor without pedestrian accidents

2.7 DYNAMIC TREE ANALYSIS FOR A GROUP OF SINGLE ACCIDENT LOCATIONS

As an example to use the dynamic tree for a group of crossings, it was decided to select the crossings with only one accident in a 10-year time period (2002-2011). It has been expressed to the research team by the Illinois Commerce Commission that the identification of trends for low accident locations is a difficult task, particularly for prioritization and selection of improvements. Thus, crossings with only one accident were selected, resulting in 720 locations in Illinois, and a single dynamic tree was constructed for all of them. The resulting tree is shown in **Figure 33**, where the bulk of accidents were observed to happen at locations with warning devices. However, no distinction between these devices is made in the CROSSING variable. Overall, the distribution of the accidents among the variables in the tree was expected and no clear trends were observed by analyzing all 720 crossings at once. A few observations from the dynamic tree for all 720 accidents include 60 pedestrian accidents, a high proportion of train striking vehicle accidents (72%), a high proportion of vehicles (60%) were travelling at speeds under 20 miles per hour. It would be interesting to explore these observed trends and how it could vary if the data is divided based on traffic control devices. The next logical step was to divide the crossings by type of warning device into: gates, active, and crossbucks. Out of the 720 crossings, 354 crossing have gates, 148 crossings have flashing lights, 215 crossings have cross bucks, 1 crossing has a stop sign as a warning device and 2 crossings didn't have any warning devices.

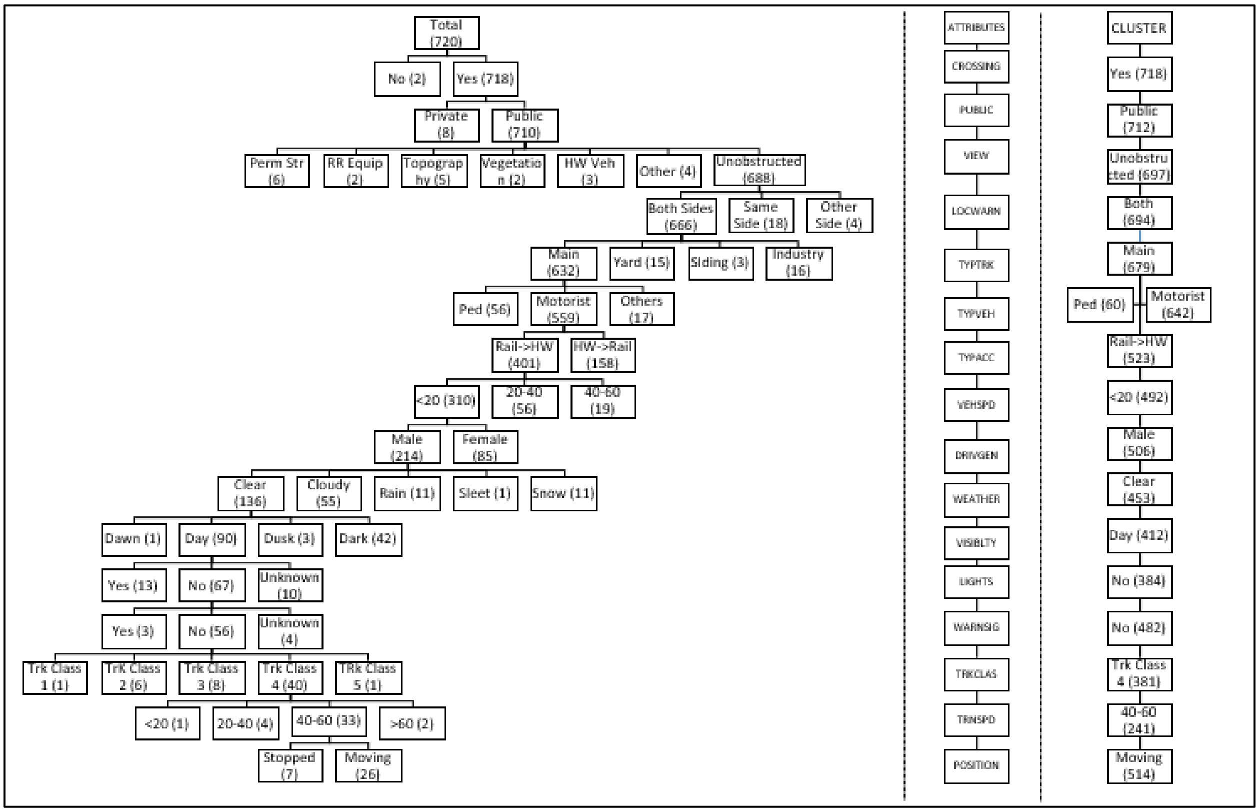


Figure 33: Dynamic Tree using Modified Method B for all Single Locations

The first group analyzed included crossings with crossbucks and contained 215 locations with each having only 1 accident in the last 10 years. The dynamic tree for these crossings is shown in **Figure 34**, and highlights some items that are interesting. First, only one accident involved pedestrians; second, about ¾ of the accidents occurred during day-time and third, close to 1/3 of the accidents (64 out of 215) involved vehicles with speed over 20mph. Even at this general level and for locations not seemingly related, these observations could serve as pointers to focus on specific areas. For example, efforts to reduce pedestrian risks or improve approach visibility may not be as effective in preventing accidents compared to measures to reduce approaching vehicle speeds.

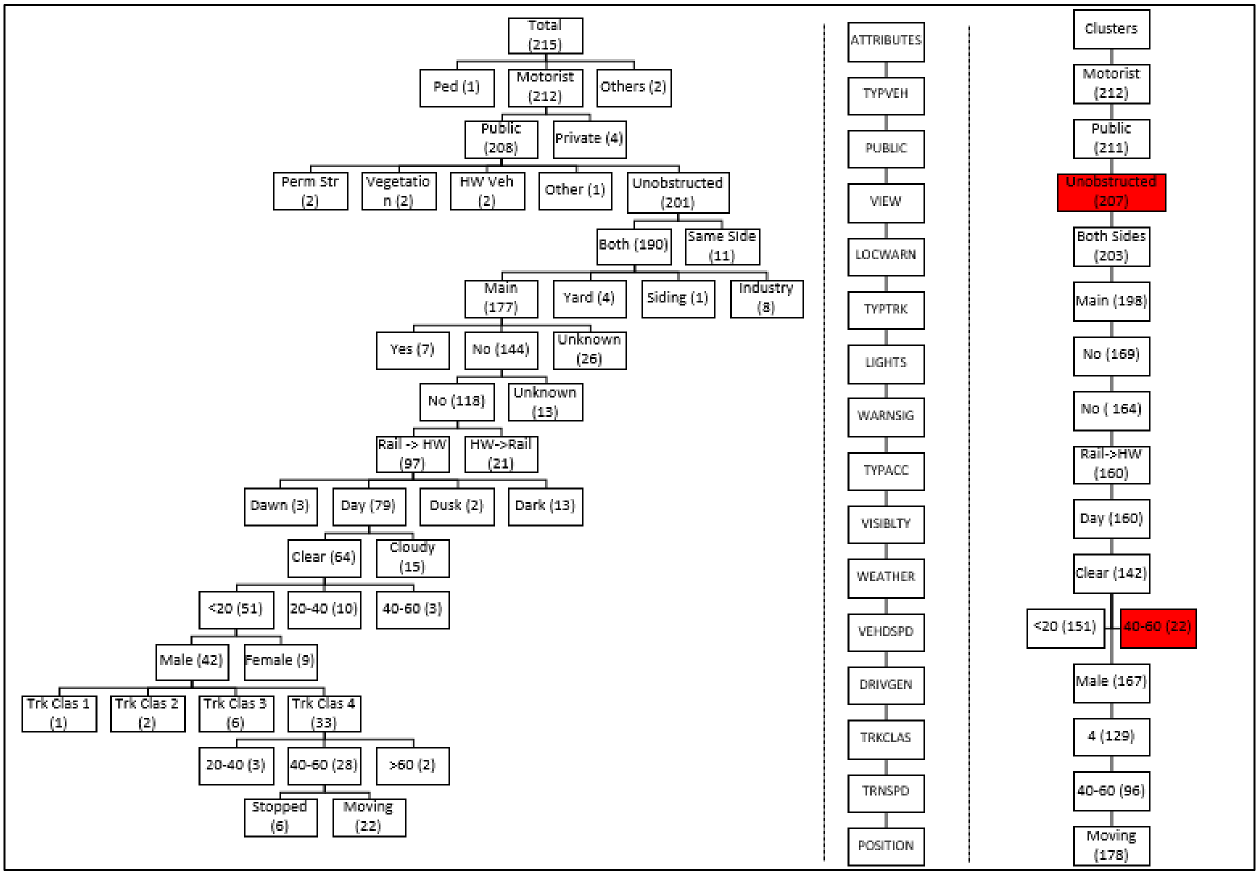


Figure 34: Dynamic Tree Analysis using Modified Method B for all Single Locations with Cross bucks

The second group included crossings with flashing lights but without gates, and was composed of 148 single accident locations in the analysis period. General characteristics of these accidents are shown in the dynamic tree shown in **Figure 35**. The accident distribution among the variables showed that pedestrians had a low incidence in the total accident frequency (similar to passive crossings), close to 2/3 of the accidents involved moving vehicles, and about half of the vehicles did not stop prior to entering the crossing. This basic information can also be used to expand the analysis and point to directions for future investigation of countermeasures.

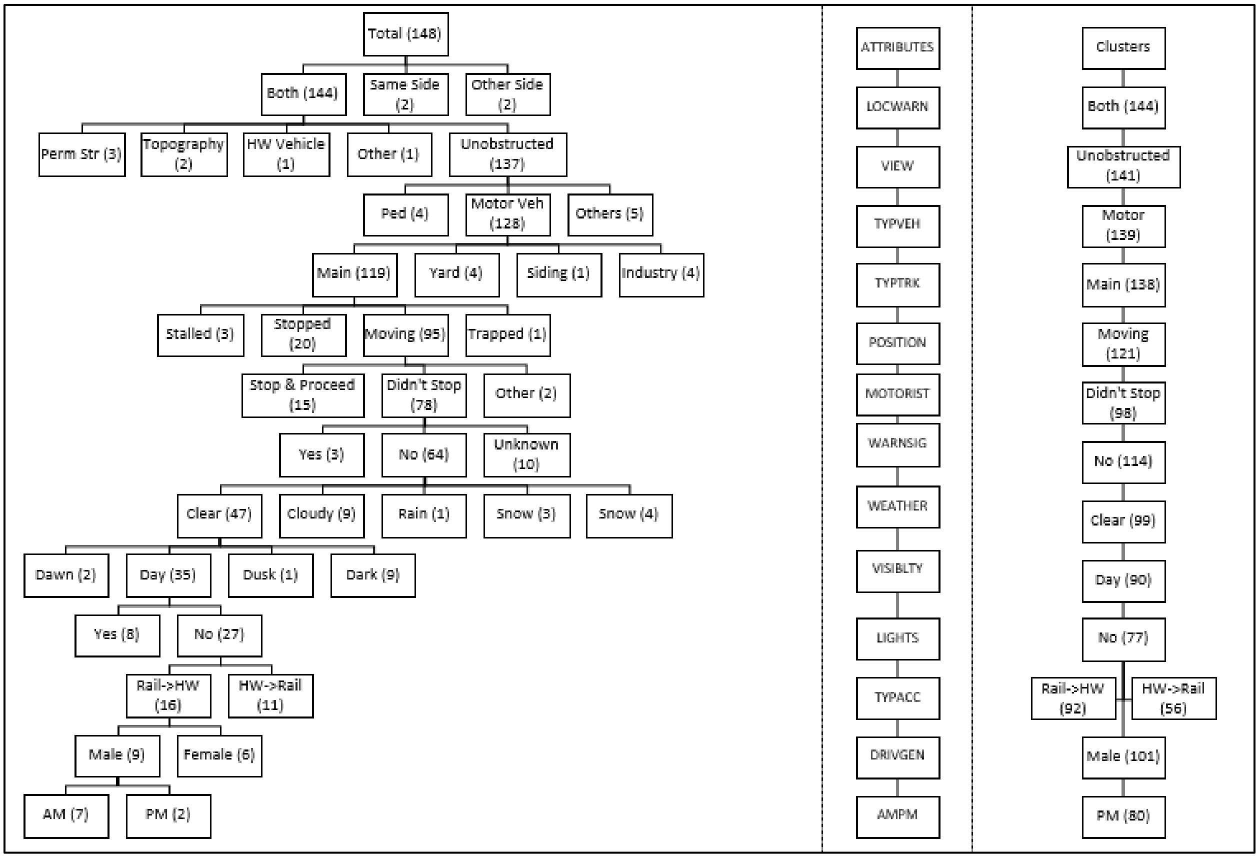


Figure 35: Dynamic Tree Analysis using Modified Method B for single accident locations with Flashing Lights

The third group of crossings included those that were equipped with gates and had 1 accident in the analysis period. A total of 354 crossings were in this category, as shown in **Figure 36**. Similar to the previous trees, valuable information can be extracted from this group, as follows: First, the incidence of pedestrian accidents was significant, with 55 from the total of 60 pedestrian accidents happened at gated locations. Second, the speed of the great majority of vehicles was lower than 20 mph, contrary to the findings for crossings with cross bucks. Third, the fraction of the accidents where trains struck vehicles (268/354=0.76) was the same as that from crossings with cross bucks (160/215=0.75) despite the differences in the vehicle approaching speeds. Fourth about half of the accidents happened at night.

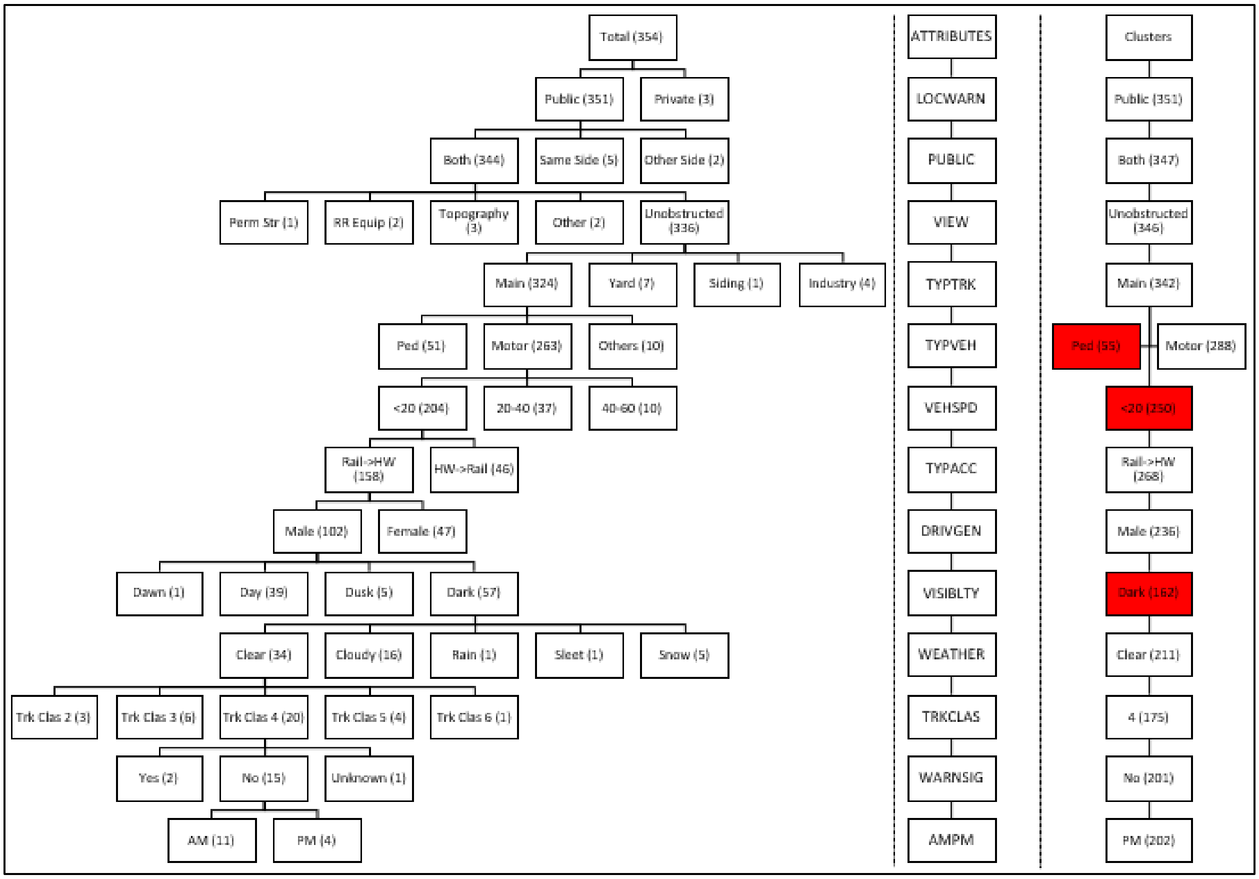


Figure 36: Dynamic Tree Analysis using Modified Method B for single accident locations with Gates

2.8 DYNAMIC TREE ANALYSIS USING VARIABLES FROM INVENTORY DATABASE

The dynamic tree analysis for a group of crossings could also be used to identify new attributes to be used in the accident prediction models. Two variables from the FRA inventory database *(4)*: XAngle and HwyNDist, were explored by the researchers. Accidents in Illinois between the years of 2005 and 2014 was used in this study.

2.8.1 XAngle

This variable in the inventory database categorizes the angle between the highway and the railroad at the crossing into 3 groups: <30 degrees, between 30 and 60 degrees and 60 degrees or greater. **Table 5** shows the number of accidents in each angle group for all three warning device.

Table 5: Number of Accidents for Each Angle Group and Warning Device

XAngle	Xbucks	Flashing Lights	Gates
<30	6	12	37
30-60	26	24	132
≥60	132	119	615

Dynamic tree analysis was done for all crossings after categorizing the angle groups only into two groups: with angle of at least 60 degrees and under 60 degrees. The angle categories <30 and 30-60 degrees were combined because of the limited number of accidents in the <30 category in all the three warning devices. This analysis was done for crossings with crossbucks, flashing lights and gates separately.

Figures 37 and 38 show the dynamic tree for gated crossings with crossings angles ≥60 and <60, respectively.

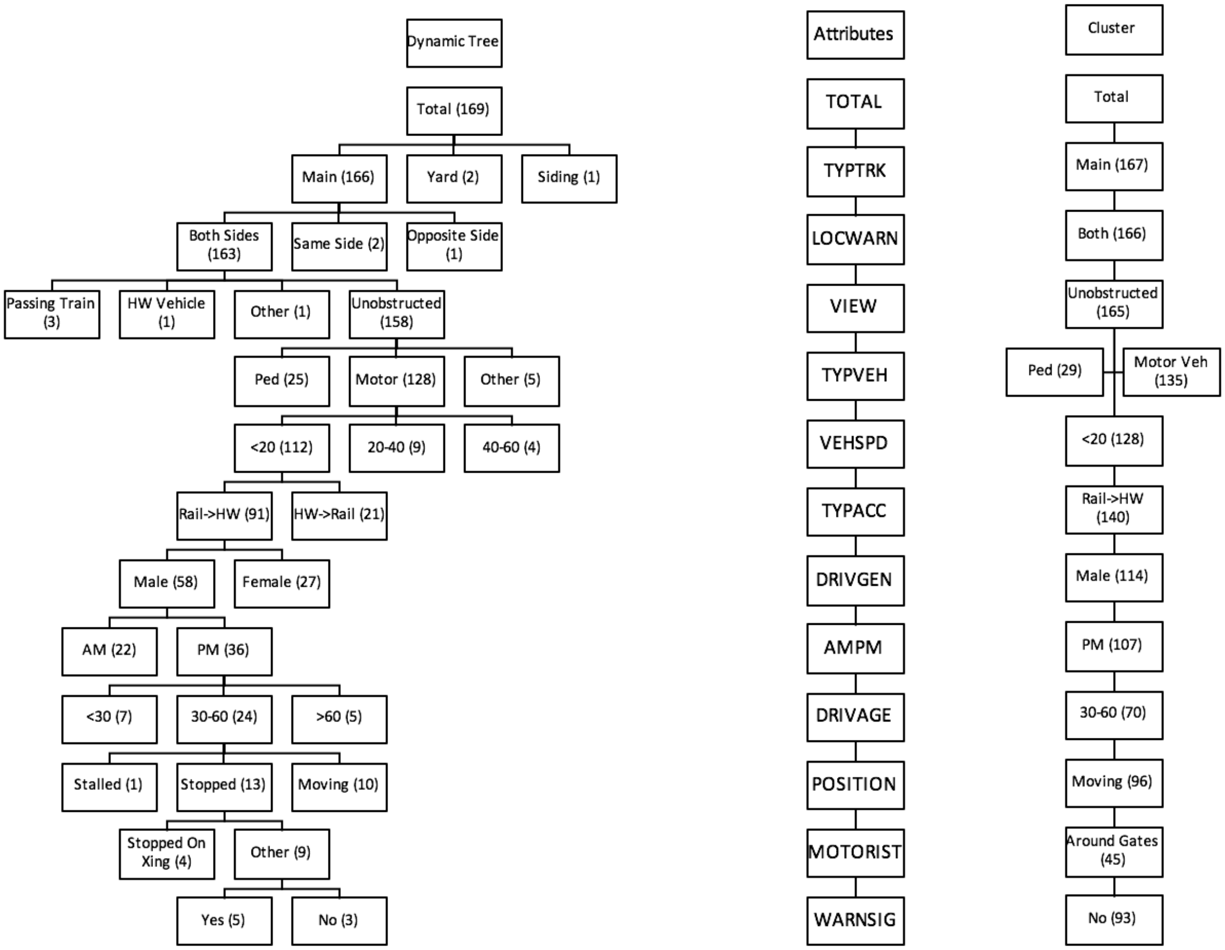


Figure 37: Accidents at Gated Crossings with Crossing Angle ≥ 60 degrees

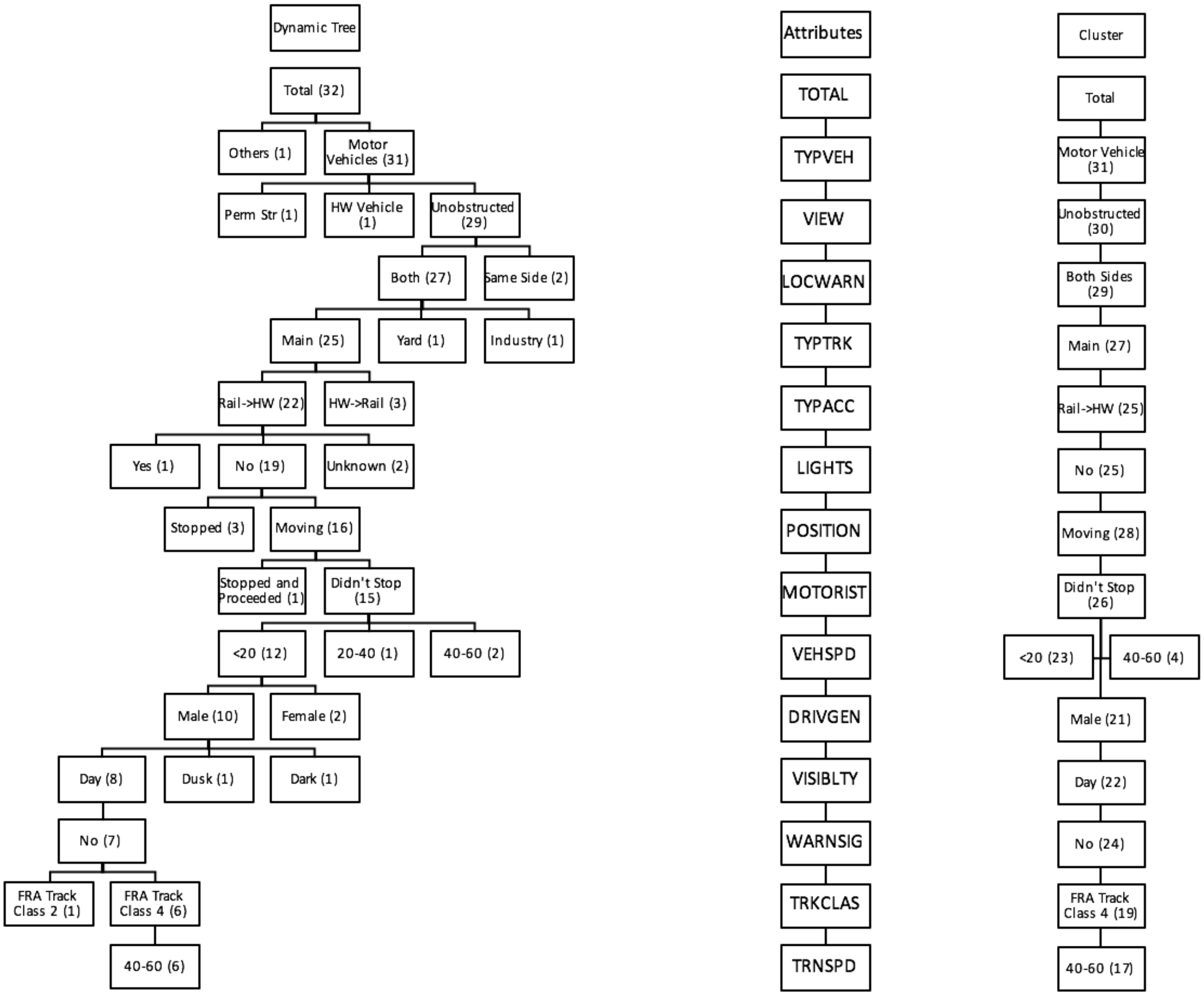


Figure 38: Accidents at Gated Crossings with Crossing Angle < 60 degrees

From **Figures 37 and 38** the following observations could be made.

1. 83% of accidents at crossings with angle <60 degrees involved a train striking a highway user while this proportion was 77% at crossings with angle ≥60 degrees.

2. The speed of the highway vehicles involved in the accident were under 20 mph in most of the accidents (76% in Crossings with Angle<60 and 72% in Crossings with Angle≥60).

3. The “POSITION" attribute appears in the dynamic tree for accidents at crossings with Angle<60 but does not appear in dynamic tree for crossings with Angle≥60

4. The "MOTORIST" attribute appears towards the end of the hierarchy in both the figures and this attribute clusters around 30% of the accidents in both the cases into the category "Drove around or thru gate".

Figures 39 and 40 show the dynamic tree for crossings with flashing lights with crossings angles ≥60 and <60 respectively.

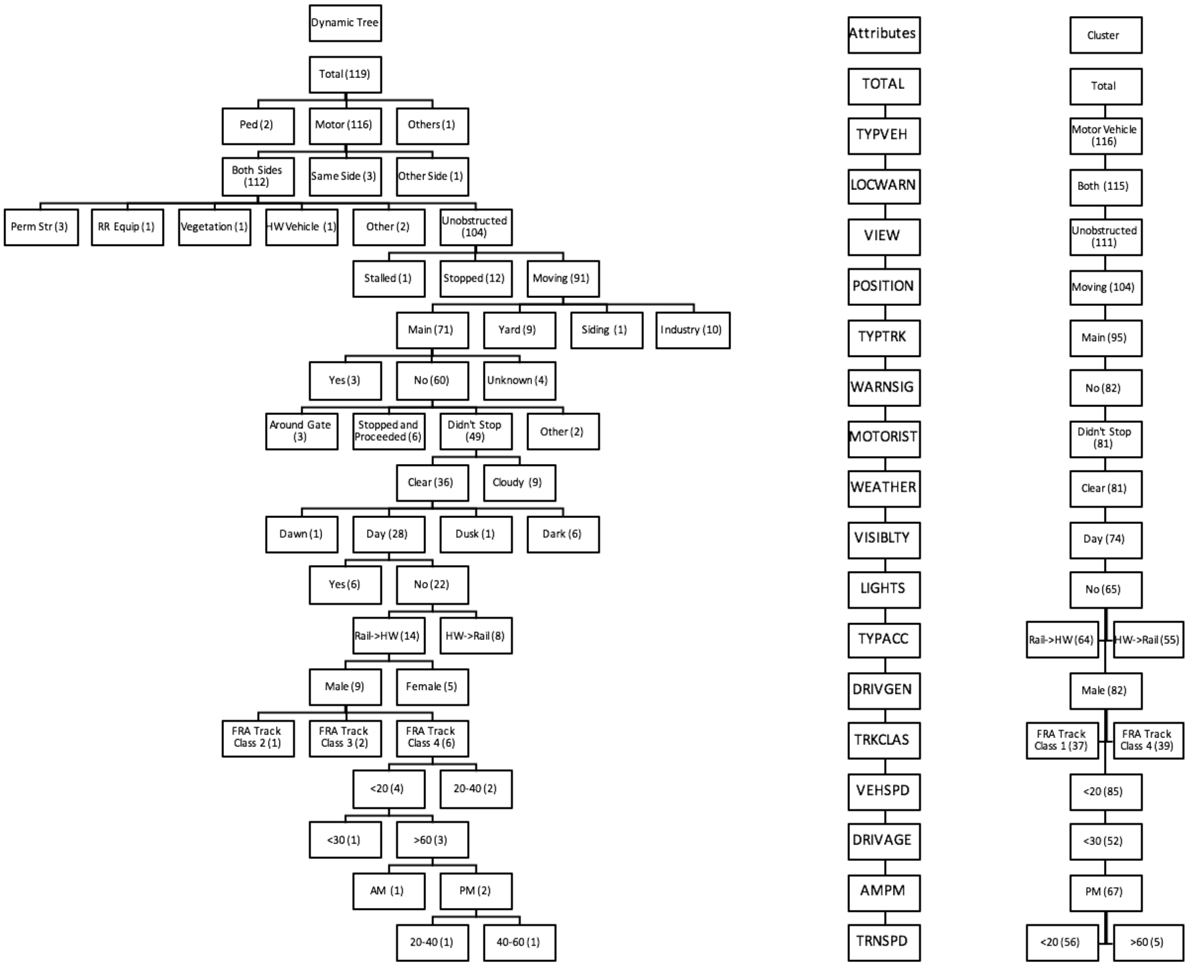


Figure 39: Accidents at Crossings with Flashing Lights with crossing angle ≥60 degrees

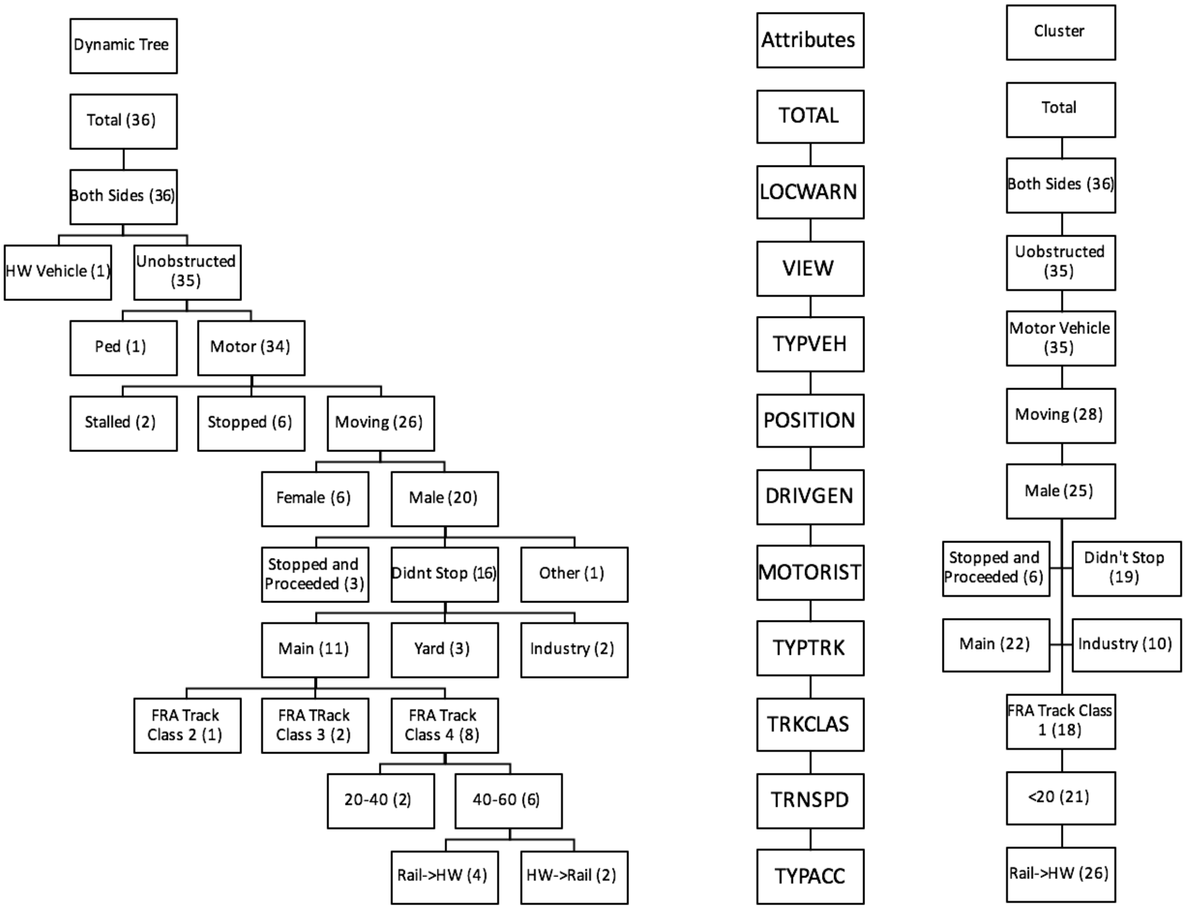


Figure 40: Accidents at Crossings with Flashing Lights with crossing angle <60 degrees

From the **Figures 39 and 40** the following observations could be made.

1. Around 77% of the accidents at crossings with angle <60 degrees involved the highway user moving while over 87% of the accidents at crossings with angle ≥60 degrees involved the highway user moving (attribute "POSITION”).

2. Around 53% of the highway users involved in accidents at crossings with angle <60 degrees didn't stop at the crossing while 68% of the highway users didn't stop at crossing where the crossing angle was ≥60 degrees (attribute "MOTORIST”)

3. Around 72% of the accidents at crossings with angle <60 degrees involved a train striking a highway user while only 54% of the accidents at crossings with angle ≥ 60 degrees involved a train striking a highway user.

Figures 41 and 42 show the dynamic tree for crossings with crossbucks with crossings angles ≥60 and <60 respectively.

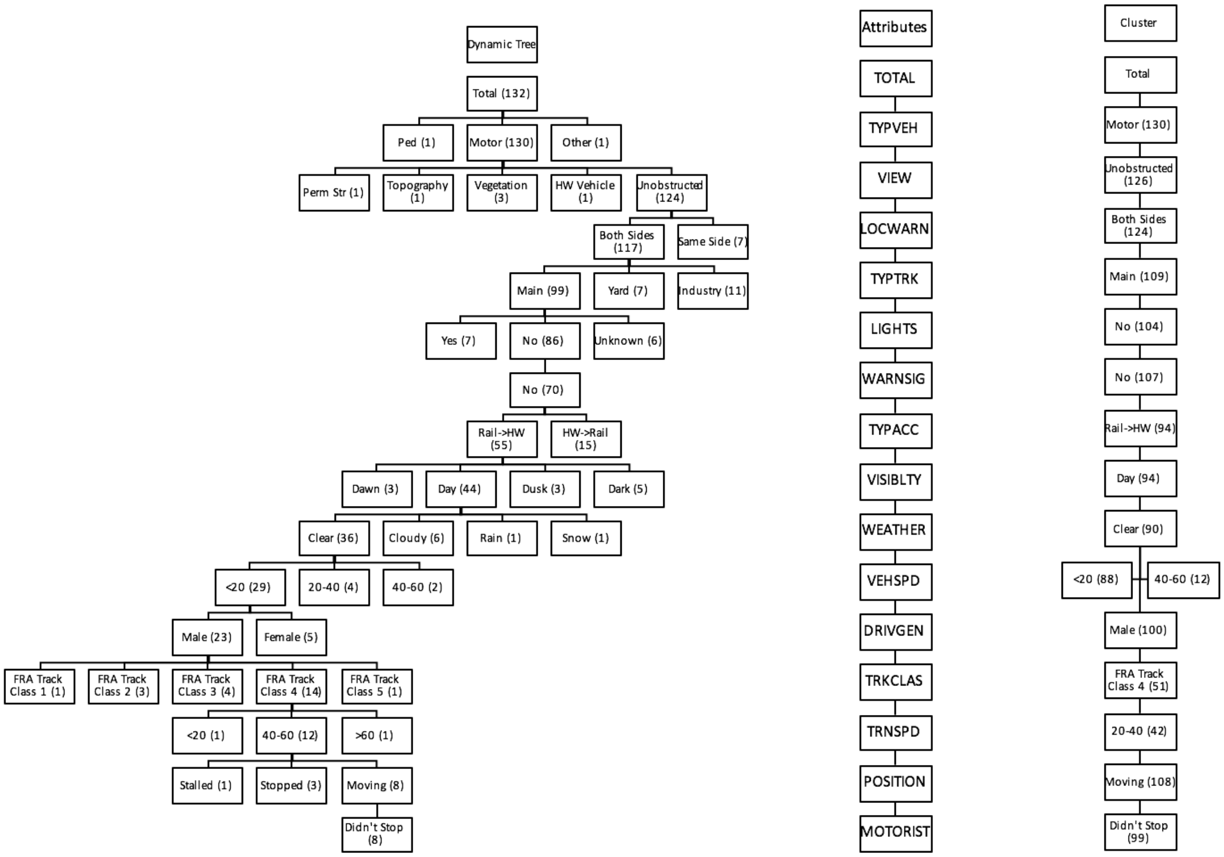


Figure 41: Accidents at Crossings with Crossbucks with Angle < 60 degrees

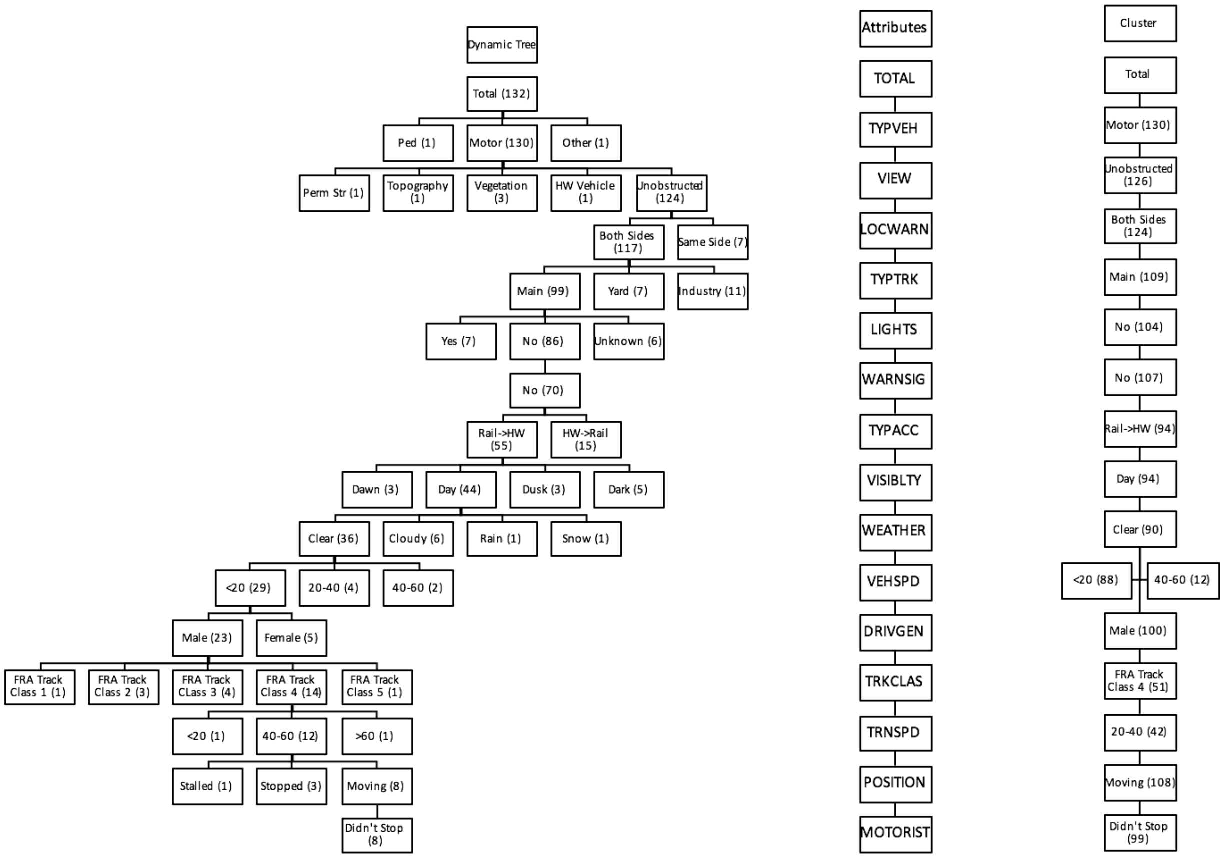


Figure 42: Accidents at Crossings with Crossbucks with Angle ≥ 60 degrees

From the above two figures the following observations could be made.

1. Around 78% of the accidents at crossings with angle <60 degrees involved a train striking a highway user while only 71% of the accidents at crossings with angle ≥ 60 degrees involved a train striking a highway user.

2. The attributes "POSITION" and "MOTORIST" appear much lower in the hierarchy in Crossings with Angle ≥ 60, but appears earlier in the hierarchy in Angle<60.

From the observations made for each warning device category, it can be said that there are differences in accident characteristics at crossings with angle <60 degrees and ≥ 60 degrees. This indicates that the variable could play a role in macroscopic modelling of accidents. The researchers explored the role of the crossing angle while developing the ZINB model.

2.8.2 HwyNDist

The HwyNDist variable in the inventory database gives the approximate intersecting roadway distance from the grade crossing. This variable was categorized into four: <75 feet, between 75 and 200 feet, between 200 and 500 feet and over 500 feet from the crossing. The **Table 6** gives the number of accidents at each group of locations analyzed.

Table 6: Number of Accidents Split by HwyNDist and Warning Device

HwyNDist	Xbucks	Flashing Lights	Gates
d<75	59	62	373
75≤d<200	0	2	51
200≤d<500	1	0	27
d≥500	104	91	333

*d is the distance from the grade crossing to a nearby HW intersection

Due to the low number of accidents, the dynamic tree analysis was not carried out for crossings with crossbucks and flashing lights for the middle two categories of HwyNDist, while it was combined for crossings with gates.

Figures 43, 44 and 45 gives the dynamic tree visualization of accidents at gated crossings for different categories of distance to nearby HW intersection.

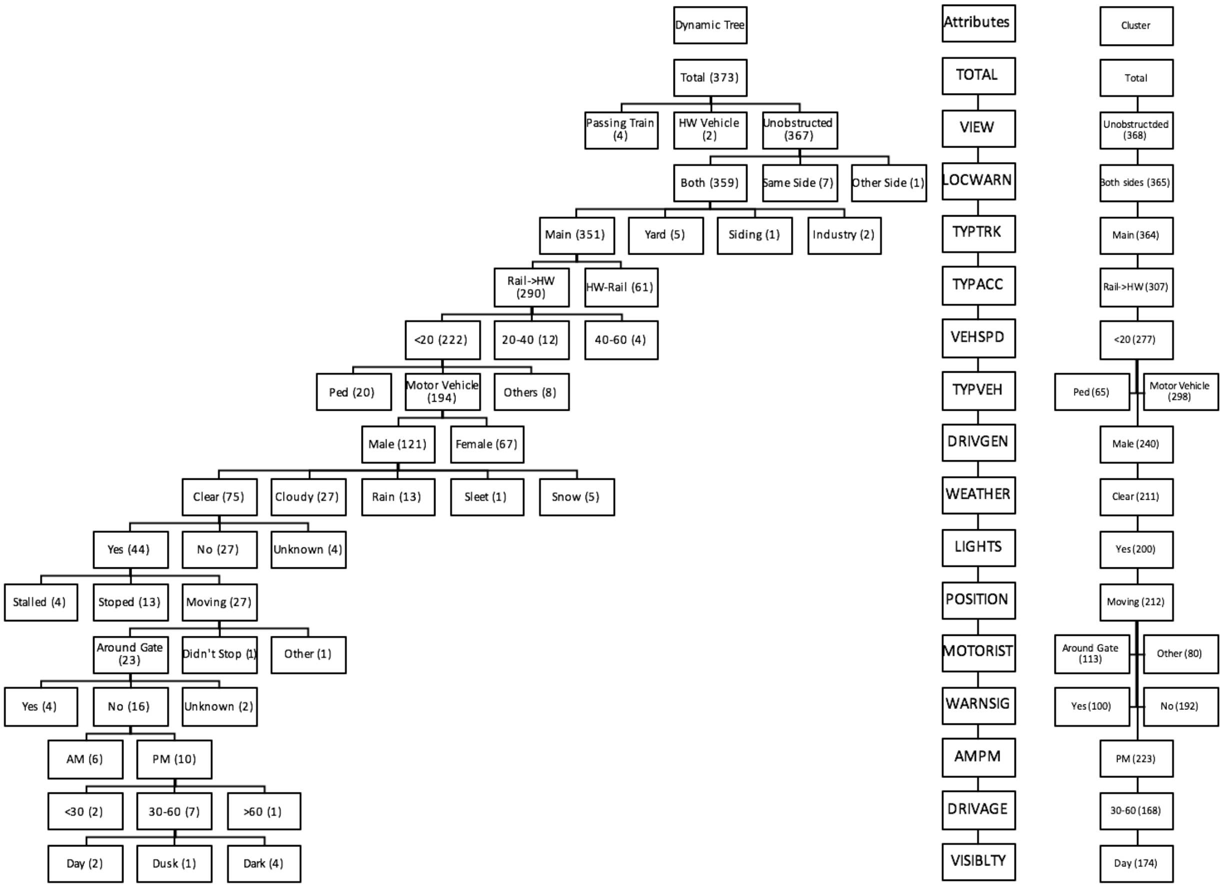


Figure 43: Accidents at Crossings with Gates with Highway intersection <75 feet

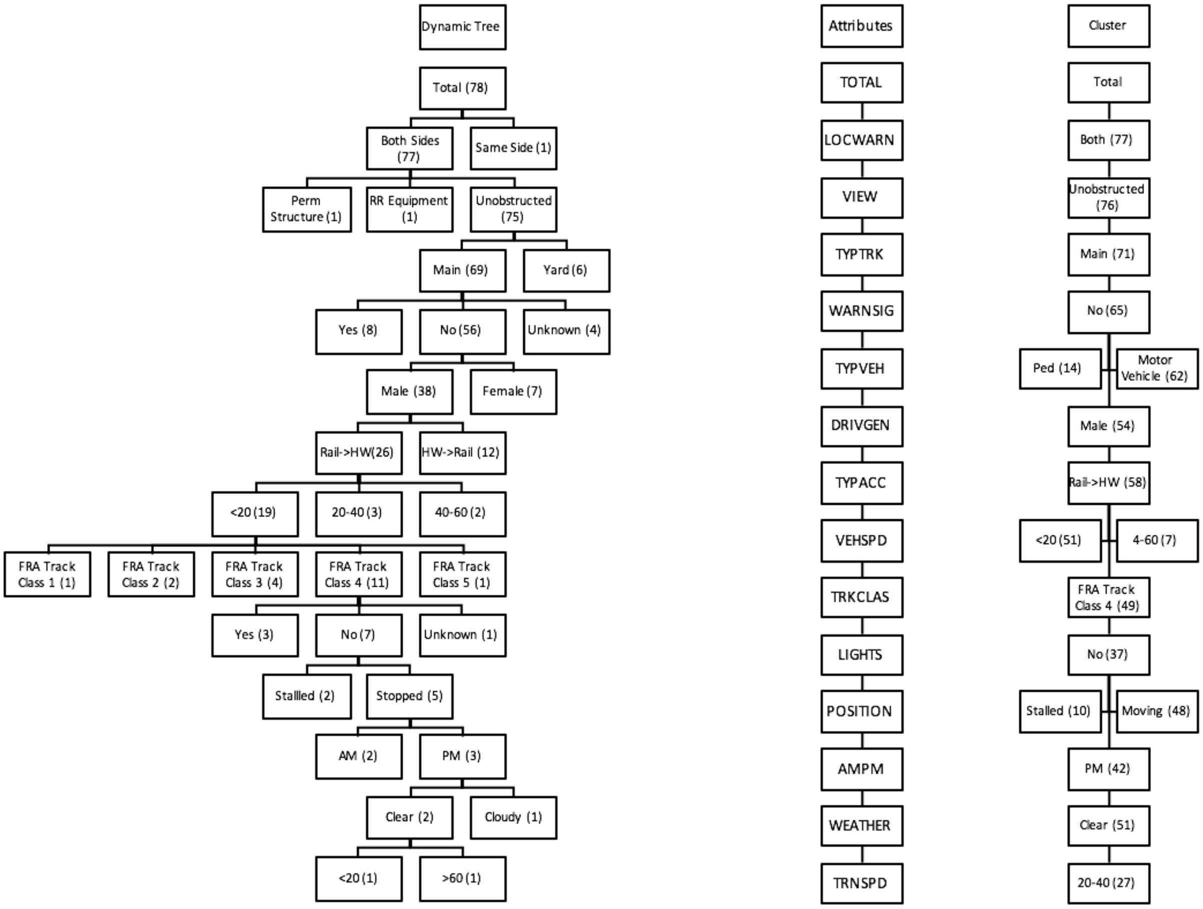


Figure 44: Accidents at Crossings with Gates with Highway intersection between 75-500 feet

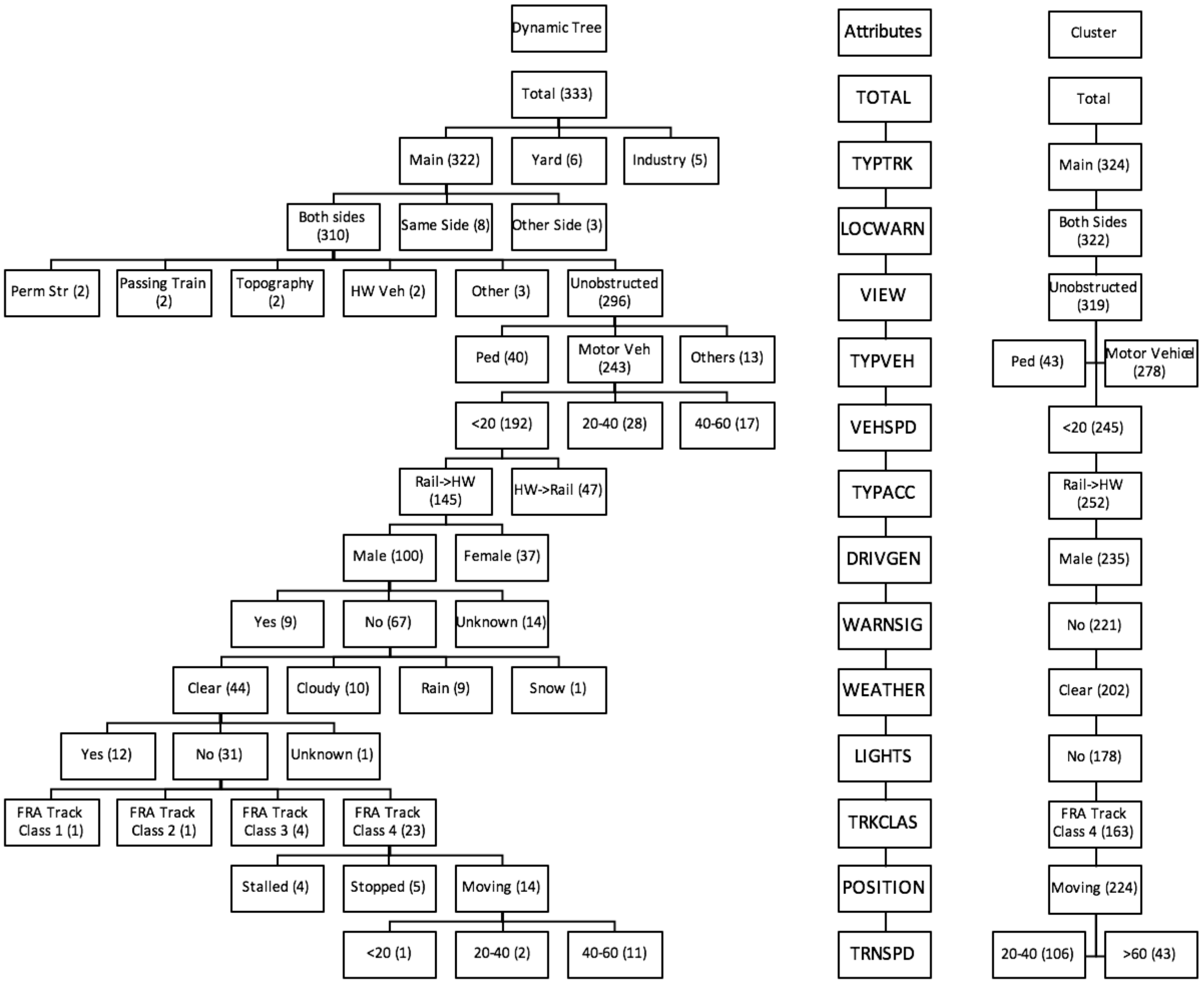


Figure 45: Accidents at Crossings with Gates with Highway intersection ≥ 500 feet

From **Figures 43, 44 and 45** the following observations could be made.

1. The variable "MOTORIST" doesn't appear in **Figures 44 and 45**, but appears in **Figure 43** showing that around 30 percent of the highway users violated the gate.
2. The attribute indicating the speed of the train ("TRNSPD") doesn't show in **Figure 43** but does in **Figures 44 and 45**.
3. The number of accidents involving a train striking the highway user is nearly 75% in **Figures 44 and 45** and 86% in **Figure 43**.

These observations indicate a difference in accident characteristics at gated crossings for crossings with intersecting highway <75 and ≥200 feet. This indicates that the HwyNDist should be explored in the macroscopic analysis for accident prediction for gated crossings.

Figures 46 and 47 gives the dynamic tree visualization of accidents at crossings with Flashing Lights for different categories of distance to nearby HW intersection.

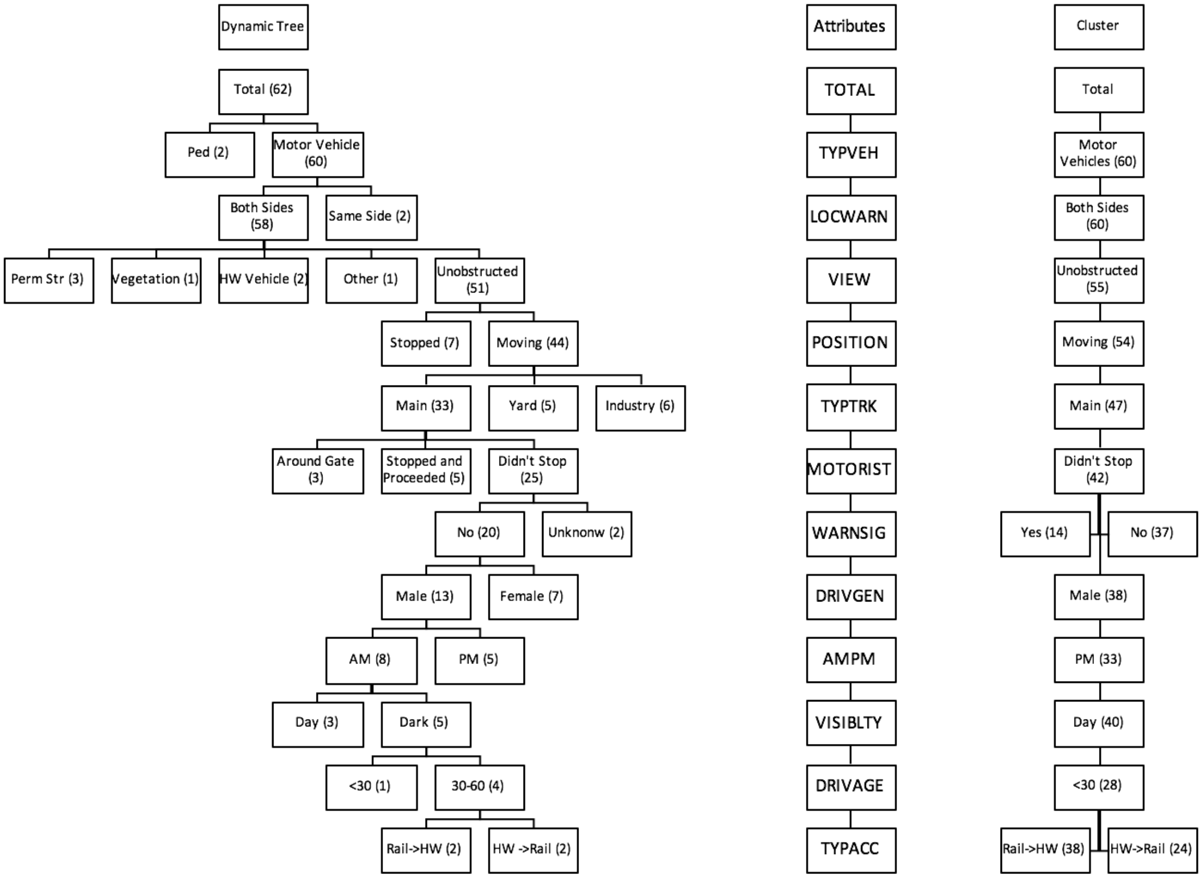


Figure 46: Accidents at Crossings with Flashing Lights with Highway < 75 feet

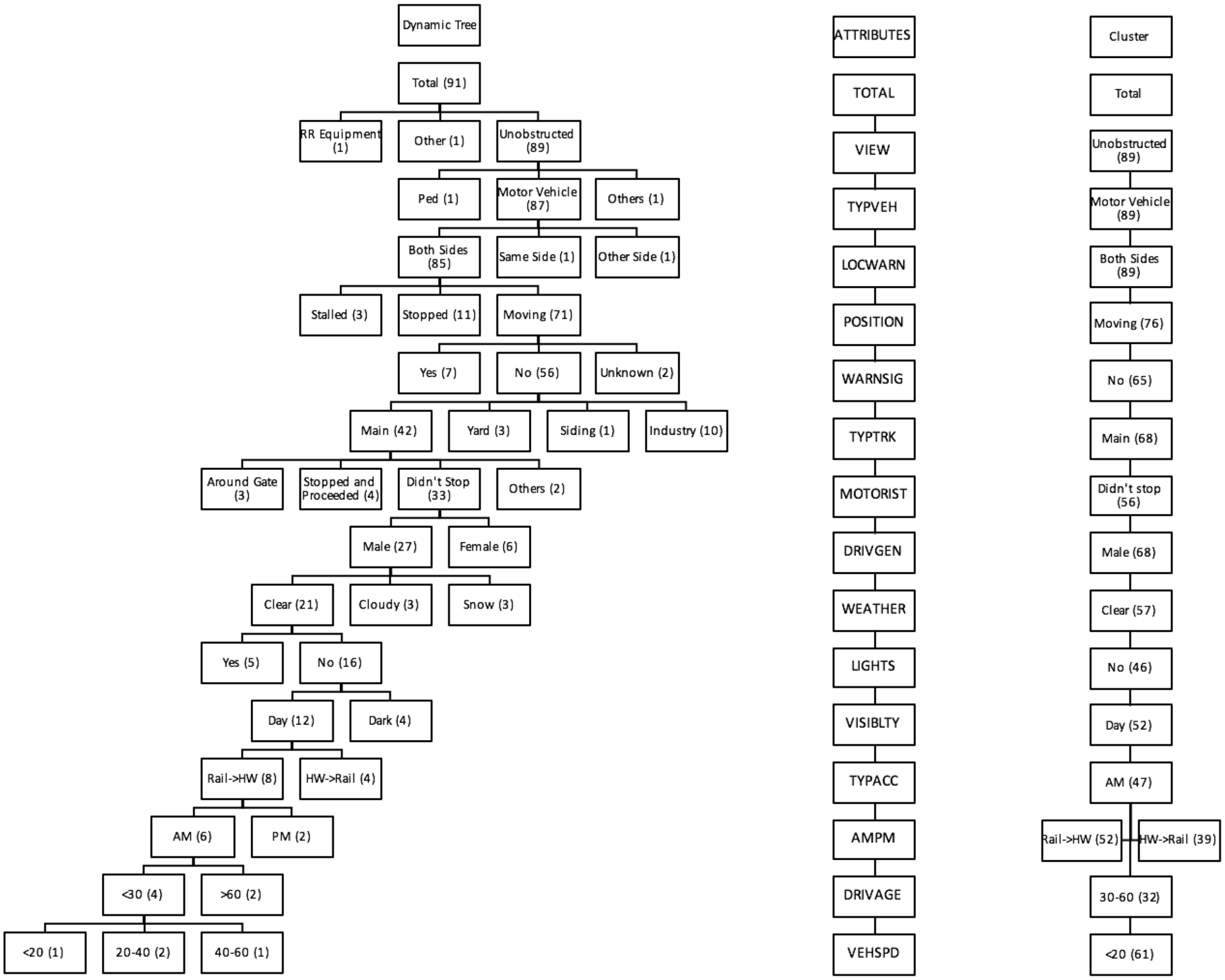


Figure 47: Accidents at Crossings with Flashing Lights with nearby Highway ≥ 500 feet

From **Figure 46 and 47** we can observe that:

1. The percentage of moving vehicles (“POSITION”) at the time of accident is around 85% in both the cases.
2. The percentage of vehicles that didn’t stop at the crossing (“MOTORIST”) is also similar in both the cases (67% in **Figure 46** and 62% in **Figure 47**).
3. The proportion of accidents involving train striking the highway vehicle is also similar in both the cases (61% in **Figure 46** and 57% in **Figure 47**).

Figures 48 and 49 gives the dynamic tree visualization of accidents at crossings with Crossbucks for different categories of distance to nearby HW intersection.

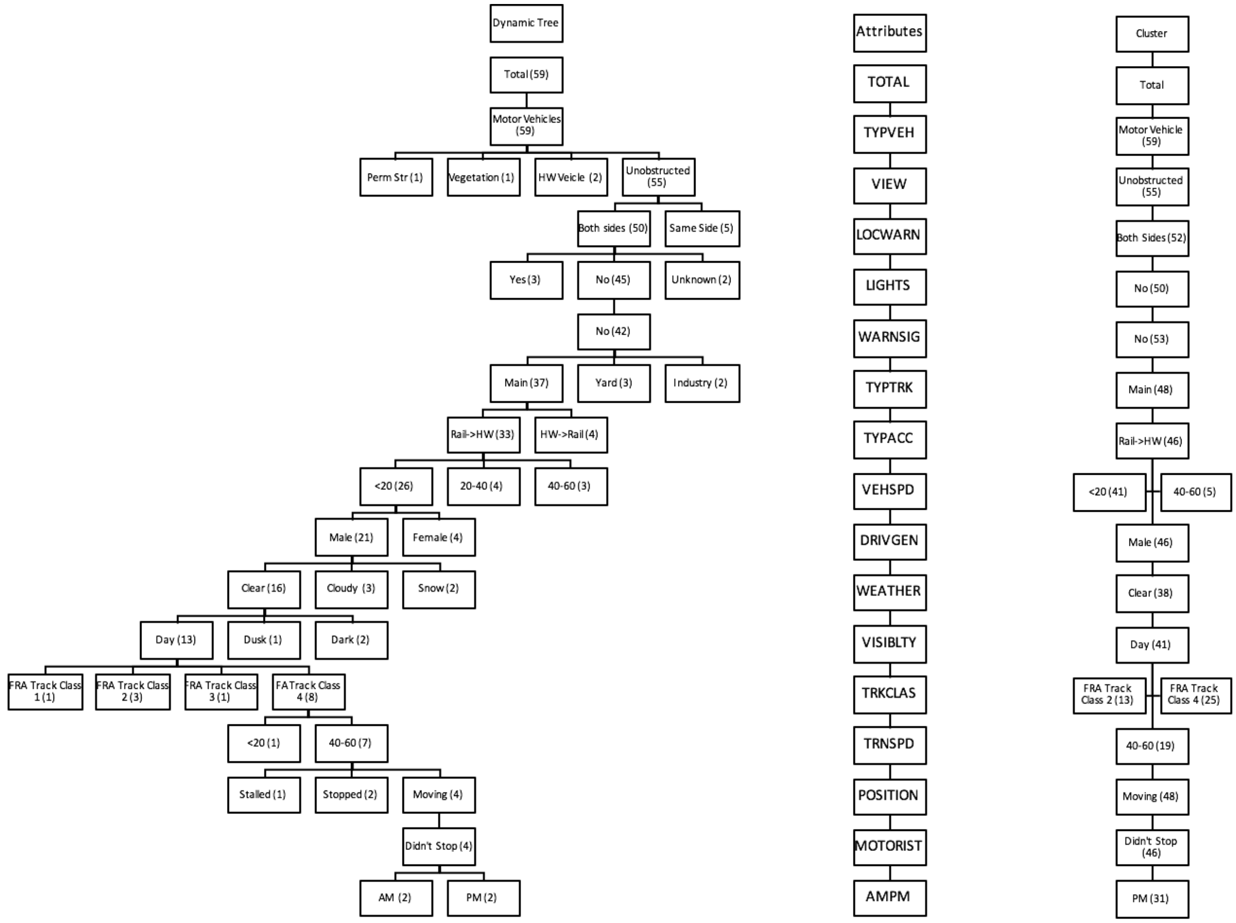


Figure 48: Accidents at Crossings with Crossbucks with Highway < 75 feet

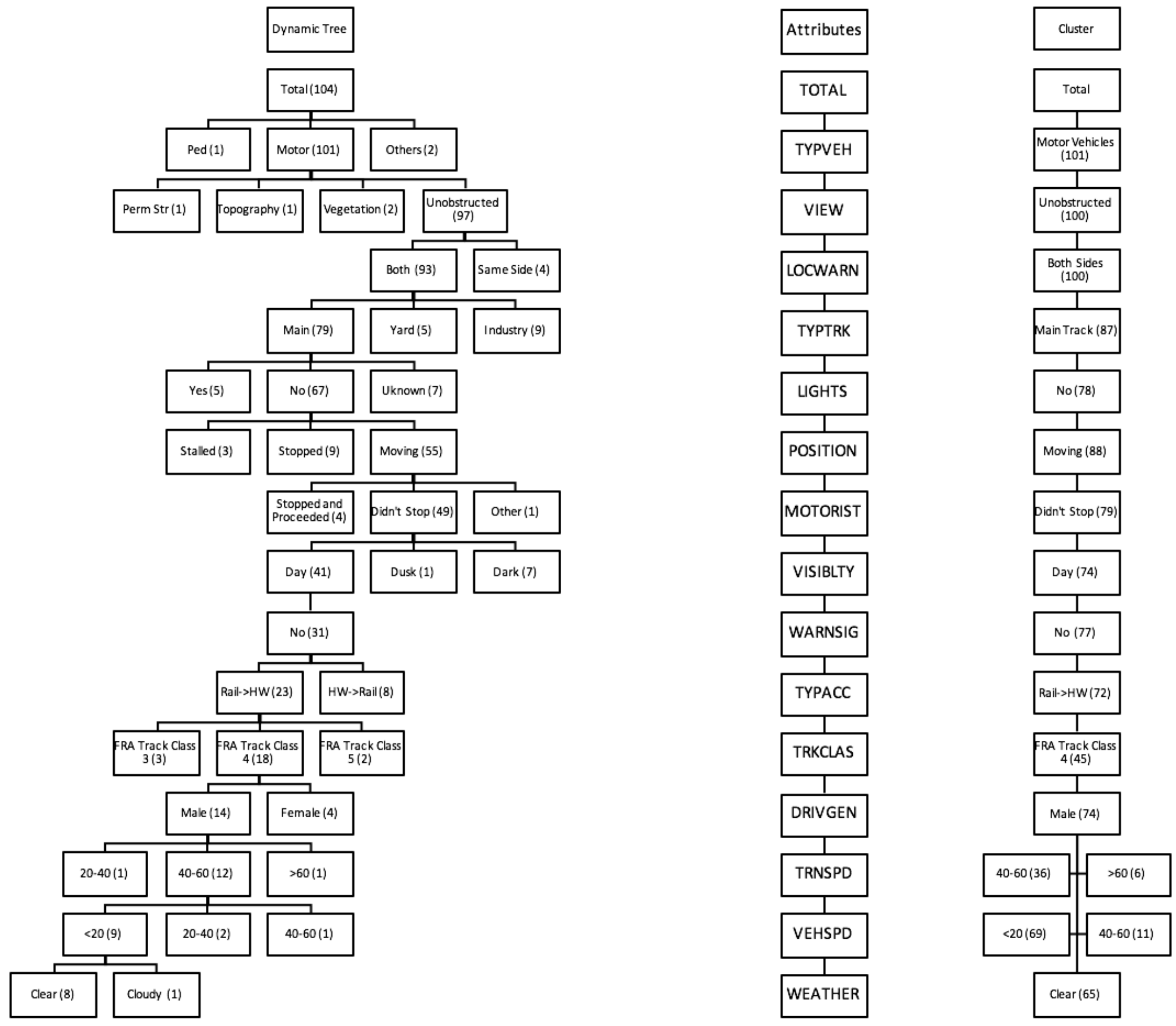


Figure 49: Accidents at Crossings with Crossbucks with Highway ≥ 500 feet

From the above two figures, the following observations could be made.

1. In both the cases, the number of accidents involving moving vehicle are very similar (81.3% in **Figure 48** and 84.6% in **Figure 49**)
2. The number of accidents which involve vehicles that didn't stop at the crossing are similar in both the cases (77.9% in **Figure 48** and 75.9% in **Figure 49**)
3. The number of highway users involved in accidents with low vehicular speeds (<20 mph) is similar in both cases (66.34% in **Figure 48** and 69.69% in **Figure 49**)

From these observations made for Flashing Lights and Crossbucks, there is little evidence to establish significance of the HwyNDist variable for these warning device types. Nevertheless, the researchers did explore its significance in the macroscopic method discussed in the next section.

In summary, the microscopic method developed in this study offers a new tool to extract useful information from the FRA accident database for safety improvements. This method creates a data—driven dynamic tree that allows the users to easily visualize any accident trends at a crossing or a group of crossings. These basic observations using dynamic trees highlighted some information that could be used as starting points for further analysis. For example, the significance of the vehicle approaching speeds in the accident frequency at crossings with crossbucks deserves more detailed analysis, as well as the high percentage of vehicles at these crossings that did not stop at all. Other findings such as the high frequency of night-time accidents at gated crossings should be further investigated as this factor itself may not be enough to draw conclusions. Overall, the analysis of multiple accident locations using Modified Method B of the dynamic tree could provide information that otherwise it would be difficult to visualize.

The dynamic tree method was also used to identify variables for macroscopic analysis. The variables XAngle and HwyNDist were chosen for macroscopic analysis to develop a Zero—Inflated Negative Binomial model to improve accident predictions over the state of practice USDOT accident prediction model.

It is also important to be able to dissect the data into meaningful groups to obtain insightful findings that could be used to direct future analysis. For example, subsets of the database with crossings from Cook County and the collar counties in Illinois could be compared against the rest of the counties in Illinois. Another example includes dataset of accidents involving freight trains compared against dataset of accidents involving commuter trains.

The next section of this report discusses the ZINB model developed to predict the accident occurrences at railroad grade crossings.

CHAPTER 3: ZERO INFLATED NEGATIVE BINOMIAL MODELS WITH EMPIRICAL BAYES ACCIDENT HISTORY ADJUSTMENT

3.1 INTRODUCTION TO CHAPTER 3

Grade crossing accident analysis on a macroscopic scale was carried out by the researchers in this study to develop new statistical accident prediction models. This was done with the objective of improving predictions over the state of practice USDOT accident prediction formula *(7)*. The most recent grade crossing accident and inventory datasets were downloaded from the Federal Railroad Administration's (FRA) website so that the most up to date data was used for model development.

Zero Inflated Negative Binomial Models (ZINB) for each category of warning devices (Gates, Flashing Lights and Cross bucks) were developed in this study. These three categories were used by the USDOT model as well. The models were developed using the statistical software R. Accident history adjustments were also made on the model based on the Empirical Bayes approach *(8)*.

ZINB models were explored by the research team previously *(3)*. The main differences between the two models include the number of variables used in the model development and the accident history adjustment applied to the model.

This report describes the datasets used to develop and validate the models. Various data cleaning filters were applied to the dataset to ensure erroneous or missing data is not used. The new ZINB model was developed using data from Illinois and was validate using data from 4 other states (Iowa, Pennsylvania, Texas and South Carolina). The adjustments due to the accident history applied to the USDOT accident prediction model and the ZINB model are described. Empirical Bayes method for accident history adjustment was applied to the ZINB model. The adjusted ZINB model is compared with the USDOT accident prediction.

3.2 DATA SOURCE AND DATA CLEANING

This study uses two databases maintained by the FRA to develop the statistical models. They are

1. Highway—Rail Crossing Inventory Database
   This database gives data regarding the characteristics of the crossings including the traffic and train volume, angle at the crossing distance to the nearby highway intersection etc.
2. Highway—Rail Crossing Accident Database

This database gives information regarding each accident that occurred at a grade crossing. This data is reported to the FRA using the form 6180.57 (Highway—Rail Grade Crossing Accident Incident Report).

Inventory data and accident data for 5 years (2012 to 2016) for the state of Illinois was downloaded. The two databases were combined using the crossing ID. The combined Illinois database was used in developing the ZINB model after it was cleaned based on the filters mentioned in **Table 7** below. The filters are an enhanced version of the filters used in the study by Medina et. al. *(6)*.

Table 7: Filters Applied to Database

	Variable Name	Variable Description	Filter	Comments
1	TypeXing	Crossing Type	Keep only '3'	'3' stands for 'Public'
2	PosXing	Crossing Position	Keep only '1'	'1' stands for 'At Grade'
3	ReasonID	Reason for update	Remove '15' Remove '16'	'15' stands for 'New Crossing' '16' stands for 'Closed'
4	DayThru	Total Daylight Thru Trains	Remove Xings: DayThru + NgthThru + TotalSwt = 0	
5	NghtThru	Total Night time Thru Trains		
6	TotalSwt	Total Switching Trains		
7	Aadt	AADT Count	Remove missing values and 1	
8	AadtYear	AADT Year	Remove every year less than or equal to 2000	
9	Xangle	Smallest Crossing Angle	Remove missing values	
10	HwySpeed	Highway Speed Limit	Remove 'NULL'	

11	SpselIDs	Train Detection	Remove missing values	
12	HwyPved	Is Roadway/Pathway Paved?	Remove missing values	
13	MaxTtSpd	Maximum Timetable Speed	Remove 'NULL' and values < 10	
14	XSurfaceIDs	Crossing Surface	Remove '17' Remove '19' Remove '20'	'17' stands for 'Metal' '19' stands for 'Composite' '20' stands for 'Other'

Further checks were also made to ensure that none of the variables had any missing values.

Furthermore, for model validation purpose similar data for 4 other states (Iowa, Pennsylvania, Texas and South Carolina) were downloaded. The same filters were applied to this dataset as well. These 4 states were selected because they were spread across the continental United States. The researchers also ensured that the filters do not wipe out most of the data points from those states. (States like California and Washington were not selected for this reason). This data was used for validating the models that were developed in this study.

3.2.1 Development Data

The dataset downloaded and filtered for Illinois is used as the development data. After the above filters were applied, the number of crossings within the dataset was reduced from 25,913 to 6,295. Over the 5-year period that was considered, there was 370 accidents at 327 locations. **Table 8** gives the division based on the warning device at each location. This dataset was used for model development.

Table 8: Number of Crossings and Accidents in Development Data (Illinois)

Warning Device	Number of crossings	Number Crossings with Accidents	Number of Accidents	% of Xings with Accidents
Gates	3286	231	268	7.03%
FL	1314	45	47	3.42%
XB	1695	51	55	3.00%

3.2.2 Validation Data

The data downloaded and filtered for the other 4 states (Iowa, Pennsylvania, Texas, and South Carolina) was used as the validation dataset in this study. After the above filters were applied, the crossing number was reduced from 66,118 to 11,492 crossings. Over the 5-year period, there were 882 accidents at 732 locations. **Table 9** gives the number of crossings and accidents for the four-state data.

Table 9: Number of Crossings and Accidents for the Validation Data

Warning Device	Number of crossings	Number of crossings with accidents	Number of accidents	% of Xings with Accidents
Gates	6051	482	604	7.96%
FL	1782	117	132	6.56%
XB	3659	133	146	3.63%

The next section describes the USDOT accident prediction model and the ZINB model developed in this study. The models which are developed are then adjusted based on the accident history at the crossing. The mathematics behind the DOT Accident history adjustment and the Empirical Bayes method for accident history adjustment is described in later sections of this report.

3.3 USDOT ACCIDENT PREDICTION FORMULA

The formula for the USDOT accident prediction value is described in the Summary of DOT Rail-Highway Crossing Resource Allocation Procedure – Revised *(9)*. The USDOT accident prediction formula involves three calculations as given below.

The basic collision prediction formula,

$$a = K * EI * DT * MS * MT * HP * HL$$

Where

a = initial accident prediction value

K = formula constant

EI = factor for exposure index based on product of highway and train traffic

DT = factor for number of thru trains per day during daylight

MS = factor for maximum timetable speed

MT = factor for number of main tracks

HP = factor for highway paved (yes or no)

HL = factor for number of highway lanes

The equations for each of the factors is given in **Table 10**.

Table 10: USDOT Accident Prediction Formula Equations

Crossing Characteristic Factors							
Device Type	K	EI	DT	MS	MT	HP	HL
Passive	0.0006938	$\left(\frac{c * t + 0.2}{0.2}\right)^{0.37}$	$\left(\frac{d + 0.2}{0.2}\right)^{0.178}$	$e^{0.0077ms}$	1	$e^{-0.5966(hp-1)}$	1
FL	0.0003351	$\left(\frac{c * t + 0.2}{0.2}\right)^{0.4106}$	$\left(\frac{d + 0.2}{0.2}\right)^{0.1131}$	1	$e^{0.1917mt}$	1	$e^{0.1826(hl-1)}$
Gates	0.0005745	$\left(\frac{c * t + 0.2}{0.2}\right)^{0.3942}$	$\left(\frac{d + 0.2}{0.2}\right)^{0.1781}$	1	$e^{0.1512mt}$	1	$e^{0.1420(hp-1)}$

Where

c = number of highway vehicles per day

t = number of trains per day

mt = number of main tracks

d = number of thru trains per day during daylight

hp = highway paved? 1=yes, 2=no

hl = number of highway lanes

The second calculation involves the consideration of collision history at the crossing. This second prediction is calculated as,

$$B = \frac{T_0}{T_0 + T}(a) + \frac{T}{T_0 + T} * \left(\frac{N}{T}\right)$$

Where

B = Second accident prediction value

T0 = Formula weighing factor (T0 = 1/(0.05+a))

N = Number of accidents recorded at the crossing in T years

T = Number of years in consideration

This procedure is called as the DOT accident history adjustment. This accident history adjusted prediction value is multiplied by the constants given in **Table 11** to reflect the current accident trends (to normalize).

Table 11: Accident Prediction Normalizing Constants

Class	April 2013 Constants
Passive	.5086
Flashing Lights	.3106
Gates	.4846

3.4 ZERO INFLATED NEGATIVE BINOMIAL (ZINB) MODEL

To model the accident count data, the zero-inflated negative binomial model is chosen. This is because of the excessive number of grade crossings with no accidents within the analysis years as seen from Table 2 and Table 3. The zero inflated models assume that the population to be modeled consists of two types of individuals: the individuals distributed based on a count distribution like Poisson or Negative Binomial and individuals with zero event counts.

Mathematically, the accident counts would have the following probability distribution per the ZINB models.

$$y_i = \begin{cases} 0 \quad with\, probability\; \varphi_i \\ g(y_i|x_i)\; with\, probability\; 1-\varphi_i \end{cases}$$

where

yi is the number of accidents at the predicted at the crossing

xi is the vector of variables input to the model

g(yi|xi) generates the count from a negative binomial process

φ_i is the probability that location would have 0 accidents

The mean and the variance of the zero-inflated negative binomial model are given as

$$\mu(y_i) = \mu_i(1-\varphi_i)$$

$$V(y_i) = \mu_i(1-\varphi_i)(1+\mu_i(\varphi_i+\alpha))$$

Where

μ_i is the mean of the negative binomial process described by g(yi|xi)

α is the over dispersion parameter of the negative binomial model.

The research team had previously explored the ZINB model as described in detail in *(3)*. The previously created ZINB models are given below.

3.4.1 Previous ZINB for Crossings with Gates

$$Accidents = \left(1 - \frac{1}{e^{-0.8641+0.0945*TotalTrn}}\right)$$
$$* \, e^{-3.3215+9.35*10^{-7}*aadt*TotalTrn+0.284*TotalTrk+f_{quad\_gate}+f_{hwy\_very\_near}}$$

$$f_{quad\_gate} = \begin{cases} 1.1149 \; if \; not \; equipped \; with \; quad \; gates \\ \quad 0 \qquad if \; equipped \; with \; quad \; gates \end{cases}$$

$$f_{hwy\_very\_near} = \begin{cases} -0.3013 \; if \; distance \; to \; nearest \; highway > 75 \; feet \\ \quad 0 \qquad\qquad otherwise \end{cases}$$

3.4.2 Previous ZINB for Crossings with Flashing Lights

$$Accidents = \left(1 - \frac{1}{e^{-1.7+0.63*TotalTrn}}\right)$$
$$* \, e^{-2.3506-1.34*10^{-6}*aadt*TotalTrn+0.3586*TotalTrk+0.2992*TrafficLane+f_{angle}}$$

$$f_{angle} = \begin{cases} \;\; 0.6071 & if \; 0 < angle < 30 \\ -0.5835 & if \; 30 \le angle < 60 \\ 0 & if \; 60 \le angle \le 90 \end{cases}$$

3.4.3 Previous ZINB for Crossings with Crossbucks

$$Accidents = \left(1 - \frac{1}{e^{-0.9006+0.1462*TotalTrn}}\right) * e^{-2.6846-1*10^{-5}*aadt*TotalTrn+0.02164*MaxTtSpd}$$

The previous ZINB model used 10 years of Illinois data from 2003 to 2012. However, the New ZINB models used the latest 5 years Illinois accident data from 2012 to 2016, as described in in the "Data Source and Data Cleaning" section of this report. The research team used the statistical software R to fit the new models.

The previous study *(3)* did not consider the use of log transformations on variables like aadt and aadt*TotalTrn. In the development of the new models, the researchers explored four different exposure measures as the base variables during the development of the model:

aadt and TotalTrn,

log(aadt) and TotalTrn,

log(aadt)*TotalTrn, and

log(aadt*TotalTrn)

Where TotalTrn is the sum of the variables DayThru,NghtThru and TotalSwt. For selection of the exposure measure the Akaike Information Criteria (AIC) is used. For all the three warning devices, the use of base variables log(Aadt) and TotalTrn resulted in a lowest AIC value.

The variables given in **Table 12** were tried along with the base variables to come up with the final model for each device type. The variables were identified by the research team via literature review of previous studies.

Table 12: Variables used in Model Development

Variable Name	Description	Comments
Aadt	Annual Average Daily Traffic (AADT) Count	
TotalTrn	Sum of “DayThru”,”NghtThru” and “TotalSwt” DayThru: Total Daylight Thru Trains NgthThru: Total Night time Thru Trains TotalSwt: Total Switching Trains	
Xangle	Crossing Angle	Categorized into three i. <29 ii. 30-59 iii. >60
HwyNDist	Approximate Intersecting Roadway Distance (feet)	Categorized into two i. <=200 feet ii. > 200 feet
TotalTrk	Sum of “MainTrk”, “YardTrk”, “SidingTrk”, “IndustryTrk” MainTrk: Number of Main tracks SidingTrk: Number of Siding tracks YardTrk: Number of Yard tracks IndustryTrk: Number of Industry tracks	Xings with >=5 tracks are collapsed into one category

MaxTtSpd	Maximum Timetable Train Speed	
HwySpeed	Posted Highway Speed Limit	
XSurfaceIDs	Crossing Surface	Categorized into five i. Timber ii. Asphalt iii. Concrete iv. Rubber v. Unconsolidated
TraficLn	Number of Traffic Lanes	Xings with >=5 lanes were collapsed into one category

The researchers decided to categorize the HwyNDist into 2 groups as shown in the table above. This is because active highway—railroad grade crossings closer than 200 feet to highway intersections with traffic signals are required to have interconnections between the traffic signal controllers and grade crossing controllers *(10).*

A lower value of AIC is desired in a model. The models described in this report are models with the lowest AIC created from a series of regression attempts.

A systematic approach was selected to identify variables for each category of warning devices as described below. A base model was created using only the exposure variables, i.e. log(Aadt) and TotalTrn. To this model, each of the variables in **Table 12** were added independently to create more models. The model which resulted in the lowest AIC value was chosen and each of the remaining 6 variables were tried to this model to explore if further reduction in AIC could be achieved. This process is repeated until any further addition of variables only resulted in an increase in AIC. All the models created for each warning device categories along with their respective AIC values are listed in **Tables 13 to 15** below. In all the cases, the log(Aadt) and TotalTrn were used to model the zero-inflated part of the model.

Table 13: Variables used in ZINB models and AIC values for Crossings with Gates

	Variables Used						AIC
	Var1	Var2	Var3	Var4	Var5	Var6	
1	TotalTrn	Log(Aadt)					1711.5
2	TotalTrn	Log(Aadt)	MaxTtSpd				1713.5
3	TotalTrn	Log(Aadt)	Xangle				1700.2
4	TotalTrn	Log(Aadt)	XSurfaceIds				1711.5
5	TotalTrn	Log(Aadt)	HwySpeed				1711.7
6	TotalTrn	Log(Aadt)	TotalTrk				1712.1
7	TotalTrn	Log(Aadt)	HwyNDist				1710.3
8	TotalTrn	Log(Aadt)	TraficLn				1708.4
9	TotalTrn	Log(Aadt)	Xangle	MaxTtSpd			1702.2
10	TotalTrn	Log(Aadt)	Xangle	XSurfaceIds			1700.6
11	TotalTrn	Log(Aadt)	Xangle	HwySpeed			1700.1

12	TotalTrn	Log(Aadt)	Xangle	TotalTrk			1699.8
13	TotalTrn	Log(Aadt)	Xangle	TraficLn			1699.2
14	TotalTrn	Log(Aadt)	Xangle	HwyNDist			1700.1
15	TotalTrn	Log(Aadt)	Xangle	HwyNDist	TotalTrk		1698.2
16	TotalTrn	Log(Aadt)	XAngle	HwyNDist	MaxTtSpd		1710.4
17	TotalTrn	Log(Aadt)	Xangle	HwyNDist	XSurfaceIds		1700.7
18	TotalTrn	Log(Aadt)	Xangle	HwyNDist	HwySpeed		1700.3
19	TotalTrn	Log(Aadt)	Xangle	HwyNDist	TraficLn		1698.5
20	TotalTrn	Log(Aadt)	Xangle	HwyNDist	TotalTrk	TraficLn	1698.6
21	TotalTrn	Log(Aadt)	Xangle	HwyNDist	TotalTrk	MaxTtSpd	1709.4
22	TotalTrn	Log(Aadt)	Xangle	HwyNDist	TotalTrk	XSurfaceIds	1698.9
23	TotalTrn	Log(Aadt)	Xangle	HwyNDist	TotalTrk	HwySpeed	1698.9

The base model had an AIC value of 1711.5 for Gates as seen from Table 13. The variables were tried systematically as shown. Model 15 (highlighted in the table) has the lowest AIC value and is chosen as the final model for crossings with gates. This model has an AIC value of 1698.2.

Table 14: Variables used in ZINB models and AIC values for Crossings with Flashing Lights

	Variables Used					AIC
	Var1	Var2	Var3	Var4	Var5	
1	TotalTrn	Log(Aadt)				394.8
2	TotalTrn	Log(Aadt)	MaxTtSpd			395.7
3	TotalTrn	Log(Aadt)	XSurfaceIds			394.5
4	TotalTrn	Log(Aadt)	HwySpeed			389.6
5	TotalTrn	Log(Aadt)	HwyNDist			398.4
6	TotalTrn	Log(Aadt)	TraficLn			399.0
7	TotalTrn	Log(Aadt)	TotalTrk			396.7
8	TotalTrn	Log(Aadt)	Xangle			398.3
9	TotalTrn	Log(Aadt)	HwySpeed	MaxTtSpd		390.1
10	TotalTrn	Log(Aadt)	HwySpeed	XSurfaceIds		388.5
11	TotalTrn	Log(Aadt)	HwySpeed	TraficLn		390.7
12	TotalTrn	Log(Aadt)	HwySpeed	HwyNDist		391.5
13	TotalTrn	Log(Aadt)	HwySpeed	totaltrk		390.1

14	TotalTrn	Log(Aadt)	HwySpeed	Xangle		391.1
15	TotalTrn	Log(Aadt)	HwySpeed	XSurfaceIds	MaxTtSpd	390.1
16	TotalTrn	Log(Aadt)	HwySpeed	XSurfaceIds	HwyNDist	390.0
17	TotalTrn	Log(Aadt)	HwySpeed	XSurfaceIds	TraficLn	389.5
18	TotalTrn	Log(Aadt)	HwySpeed	XSurfaceIds	totaltrk	389.5
19	TotalTrn	Log(Aadt)	HwySpeed	XSurfaceIds	Xangle	388.7

For crossings with Flashing Lights, the base model had an AIC value of 394.8. Addition of variables HwySpeed and XSurfaceIds (highlighted in the table) showed a reduction in the AIC value of the model to 388.5 and is chosen as the final model for flashing lights.

Table 15: Variables used in ZINB models and AIC values for Crossings with Crossbucks

	Variables Used					AIC
	Var1	Var2	Var3	Var4	Var5	
1	TotalTrn	Log(Aadt)				478.91
2	TotalTrn	Log(Aadt)	MaxTtSpd			476.95
3	TotalTrn	Log(Aadt)	XSurfaceIds			474.94
4	TotalTrn	Log(Aadt)	HwySpeed			480.35
5	TotalTrn	Log(Aadt)	HwyNDist			481.77
6	TotalTrn	Log(Aadt)	XAngle			480.86
7	TotalTrn	Log(Aadt)	TotalTrk			478.72

8	TotalTrn	Log(Aadt)	TraficLn			480.78
9	TotalTrn	Log(Aadt)	XSurfaceIds	HwySpeed		478.40
10	TotalTrn	Log(Aadt)	XSurfaceIds	MaxTtSpd		473.35
11	TotalTrn	Log(Aadt)	XSurfaceIds	HwyNDist		477.80
12	TotalTrn	Log(Aadt)	XSurfaceIds	Xangle		476.94
13	TotalTrn	Log(Aadt)	XSurfaceIds	TotalTrk		475.07
14	TotalTrn	Log(Aadt)	XSurfaceIds	TraficLn		476.91
15	TotalTrn	Log(Aadt)	XSurfaceIds	MaxTtSpd	HwySpeed	474.39
16	TotalTrn	Log(Aadt)	XSurfaceIds	MaxTtSpd	HwyNDist	474.83
17	TotalTrn	Log(Aadt)	XSurfaceIds	MaxTtSpd	Xangle	475.29
18	TotalTrn	Log(Aadt)	XSurfaceIds	MaxTtSpd	TotalTrk	474.08
19	TotalTrn	Log(Aadt)	XSurfaceIds	MaxTtSpd	TraficLn	475.29

The base model for crossings with crossbucks had an AIC value of 478.91. Inclusion of variables MaxTtSpd and XSurfaceIds could reduce the AIC value to 473.35. This model (highlighted in table) is the final model chosen for crossbucks.

The final models chosen for each warning device is described in the section below.

3.4.4 Crossings with Gates

The ZINB model created for crossings with gates includes the variables TotalTrk, XAngle and HwyNDist. The model is of the form

$$Accidents\ (5\ Years) = \frac{e^{-4.41+0.004*TotalTrn+0.34*\log(Aadt)+0.13*TotalTrk+A+D}}{1+e^{4.75-0.29*TotalTrn-0.28*\log(Aadt)}}$$

$$A = \begin{cases} -0.34 & if\ 30 \leq crossing\ angle < 60 \\ -0.33 & if\ crossing\ angle\ \geq\ 60 \end{cases}$$

$$D = \begin{cases} 0 & if\ distance\ to\ nearby\ highway\ intersection\ is \leq 200\ feet \\ -0.23 & if\ distance\ to\ nearby\ highway\ intersection > 200\ feet \end{cases}$$

The over dispersion parameter (α) for this model is 0.81.

3.5 CROSSINGS WITH FLASHING LIGHTS

The ZINB model created for crossings with flashing lights include the variables HwySpeed and XSurfaceIds. The model is of the form

$$Accidents\ (5\ Years) = \frac{e^{-2.93+0.03*TotalTrn-0.16*Log(Aadt)+0.04*HwySpeed+S}}{1+e^{\ 9.72-0.11*TotalTrn-1.015*Log(Aadt)}}$$

$$S\ = \begin{cases} 0 & if\ crossing\ surface\ is\ timber \\ -0.93 & if\ crossing\ surface\ is\ asphalt \\ -0.19 & if\ crossing\ surface\ is\ concrete \\ -0.97 & if\ crossing\ surface\ is\ rubber \\ 1.18 & if\ crossing\ surface\ is\ unconsolidated \end{cases}$$

The over dispersion parameter (α) for this model is 6.70e-05

3.6 CROSSINGS WITH CROSSBUCKS

The ZINB model created for crossings with crossbucks includes the variables MaxTtSpd and XSurfaceIDs. The model is of the form

$$Accidents\ (5\ Years) = \frac{e^{-4.87-0.013*TotalTrn+0.017*Log(Aadt)+0.02*MaxTtSpd+S}}{1+e^{9.83-0.63*TotalTrn-1.67*Log(Aadt)}}$$

$$S\ = \begin{cases} 0 & if\ crossing\ surface\ is\ timber \\ 1.13 & if\ crossing\ surface\ is\ asphalt \\ 1.57 & if\ crossing\ surface\ is\ concrete \\ 1.33 & if\ crossing\ surface\ is\ rubber \\ 1.56 & if\ crossing\ surface\ is\ unconsolidated \end{cases}$$

The over dispersion parameter (α) for this model is 1.35.

Like the U.S. DOT formulas, the models created were adjusted based on the accident history at the crossing. The adjustment was based on the Empirical Bayes approach to combine two clues to safety to improve the estimator of the number of recorded accidents at a crossing.

3.7 EMPIRICAL BAYES ACCIDENT HISTORY ADJUSTMENT

The prediction using the ZINB model was then corrected using the Empirical Bayes (EB) method as described by Hauer *(8)*. The adjusted estimate of the expected number of accidents is given by the formula

$$E(y_i|Y_i) = \alpha E(y_i) + (1-\alpha)Y_i$$

$$\alpha = \frac{1}{1+\frac{V(y_i)}{E(y_i)}}$$

where Y_i is the number of accidents recorded at the crossing.

Using the above equation, the initial prediction of the number of accidents was updated based on the past accident history witnessed at the crossing.

In the next section, comparisons are made between the ZINB model adjusted using the EB procedure and the USDOT accident prediction formula.

3.8 COMPARISON BETWEEN THE MODELS

Comparisons are made between the three models for each warning device type in two datasets: the model development dataset (data from Illinois) and the model validation dataset (data from Texas, Pennsylvania, Iowa and North Carolina). The researchers compared the two models (for each type of warning device) in three different ways.

1. Cumulative distribution of predicted accidents with respect to cumulative distribution of accidents.
2. Predicted number of accidents at a location with respect to actual number of accidents at location
3. Relative ranking of crossings based on the number of accidents identified among top locations.

The following subsections describe each of the three comparisons listed above.

3.8.1 Cumulative distribution of predicted accidents with respect to cumulative distribution of accidents

3.8.1.1 Crossings with Gates

The three different models were used to plot the distribution of the cumulative number of accidents at the crossings to make comparison with the distribution of the field data. The following **Figures 50 and 51** shows the distribution within the model creation dataset and the same plot zoomed in to show the top 8% of the crossings. (From **Table 8 and Table 9**, at least 8% of crossings had an accident in all the datasets). **Figures 52 and 52** shows the same for the model validation dataset. The figures also show the cumulative distribution of the unadjusted USDOT model and ZINB models.

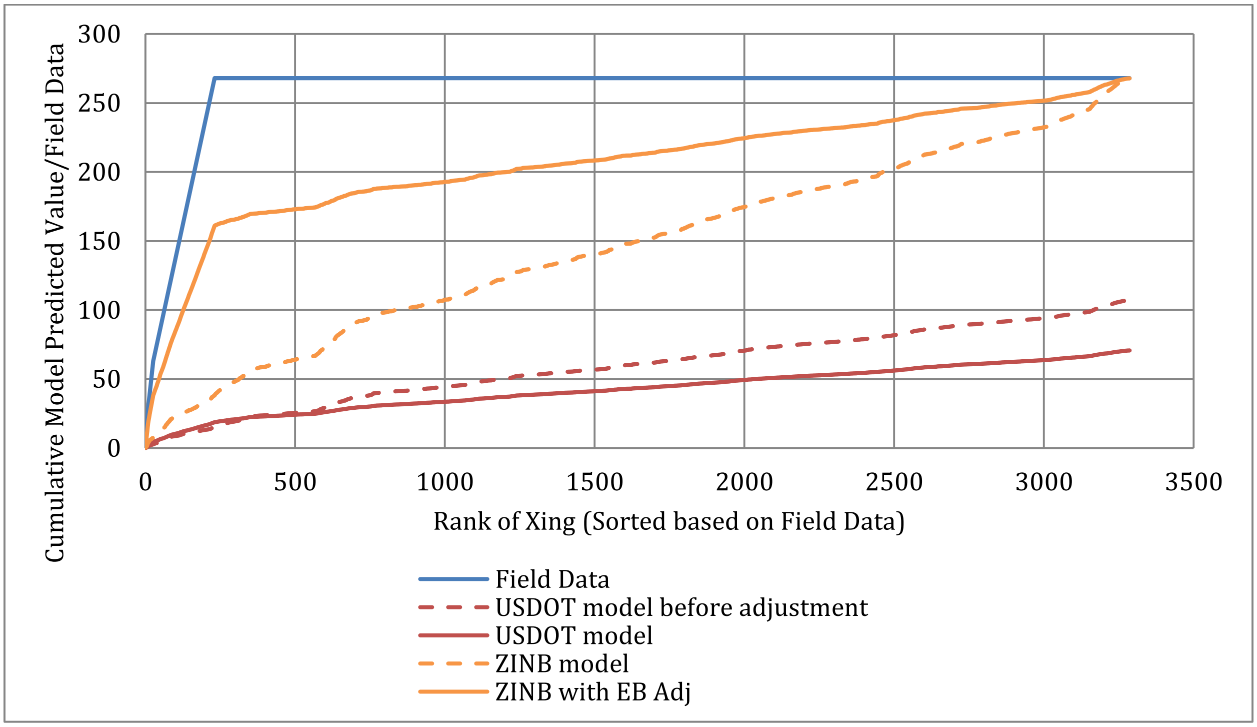


Figure 50: Cumulative number of accidents at all locations for Crossings with Gates (Model Development Dataset)

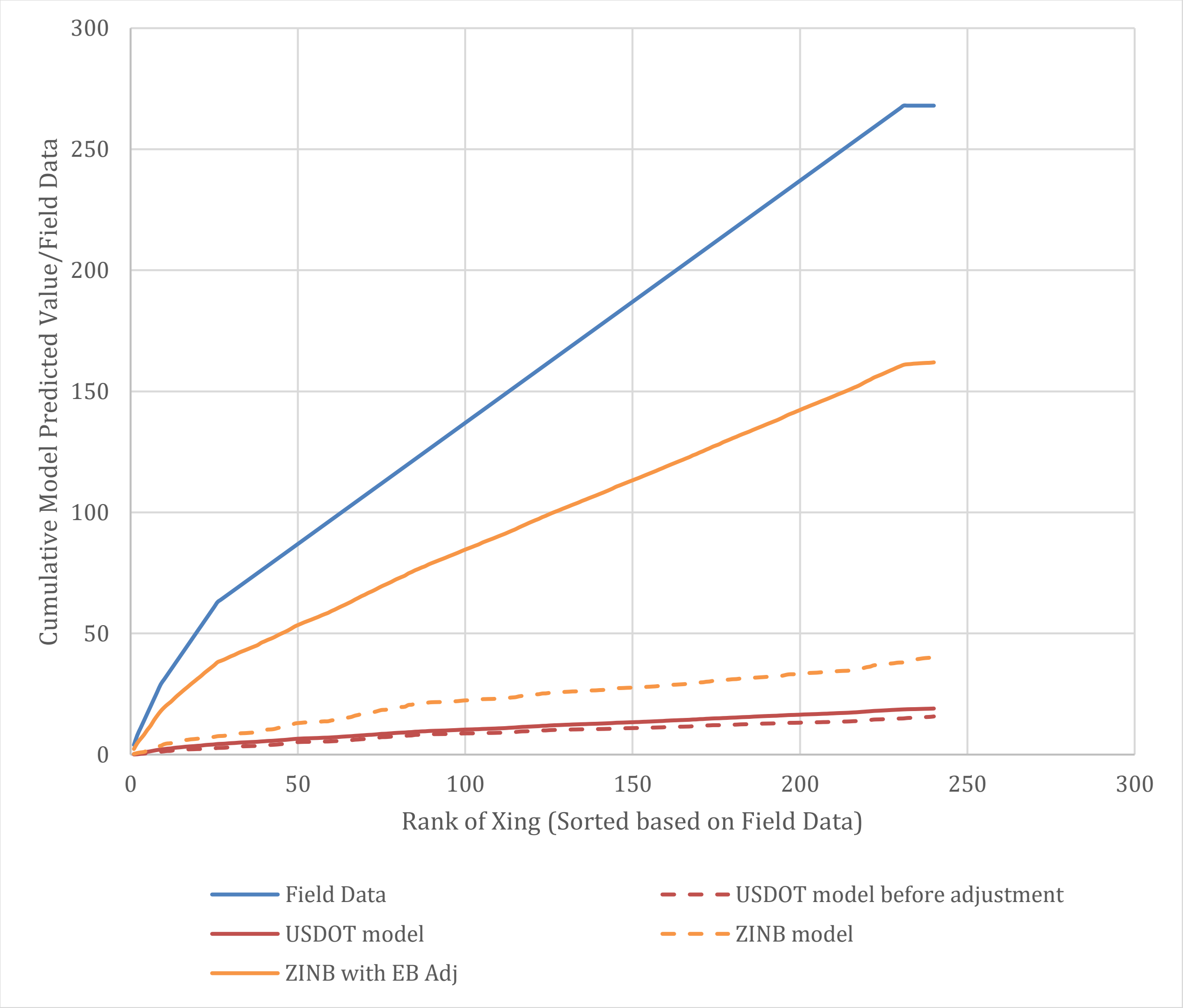


Figure 51: Cumulative number of accidents at top locations for crossings with Gates (Model Development Dataset)

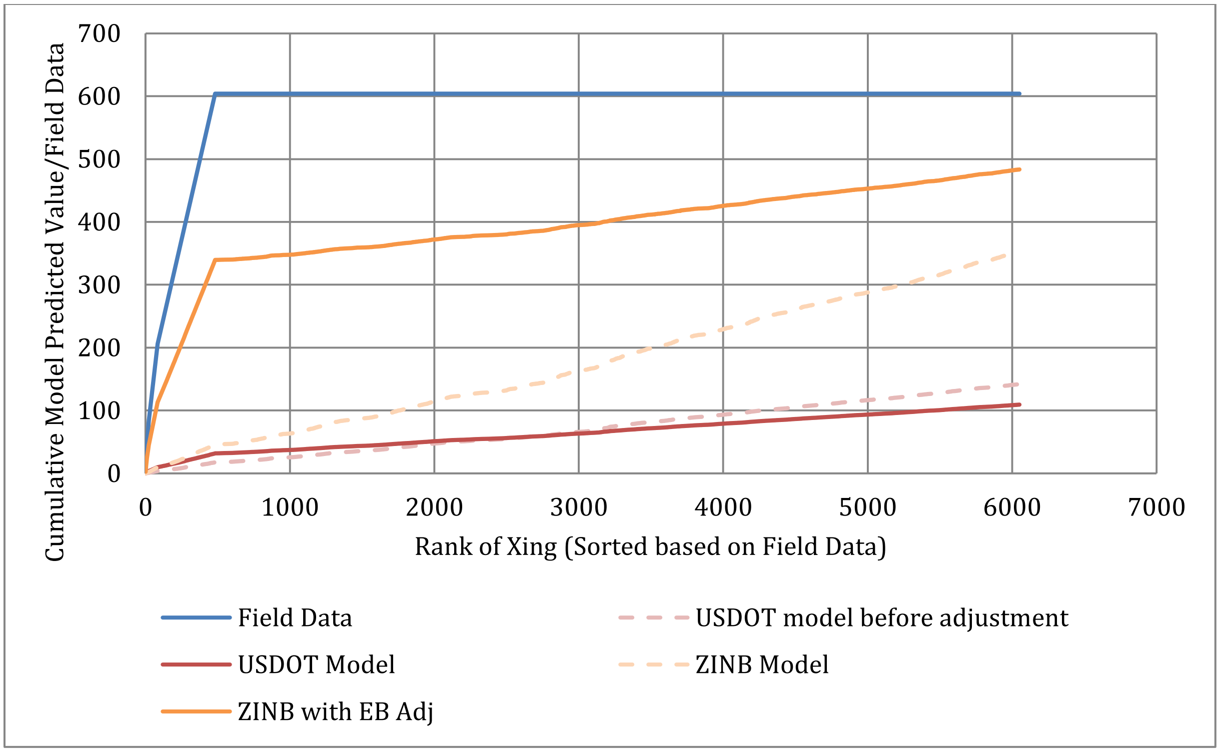


Figure 52: Cumulative number of accidents at all locations for Crossings with Gates (Model Validation Dataset)

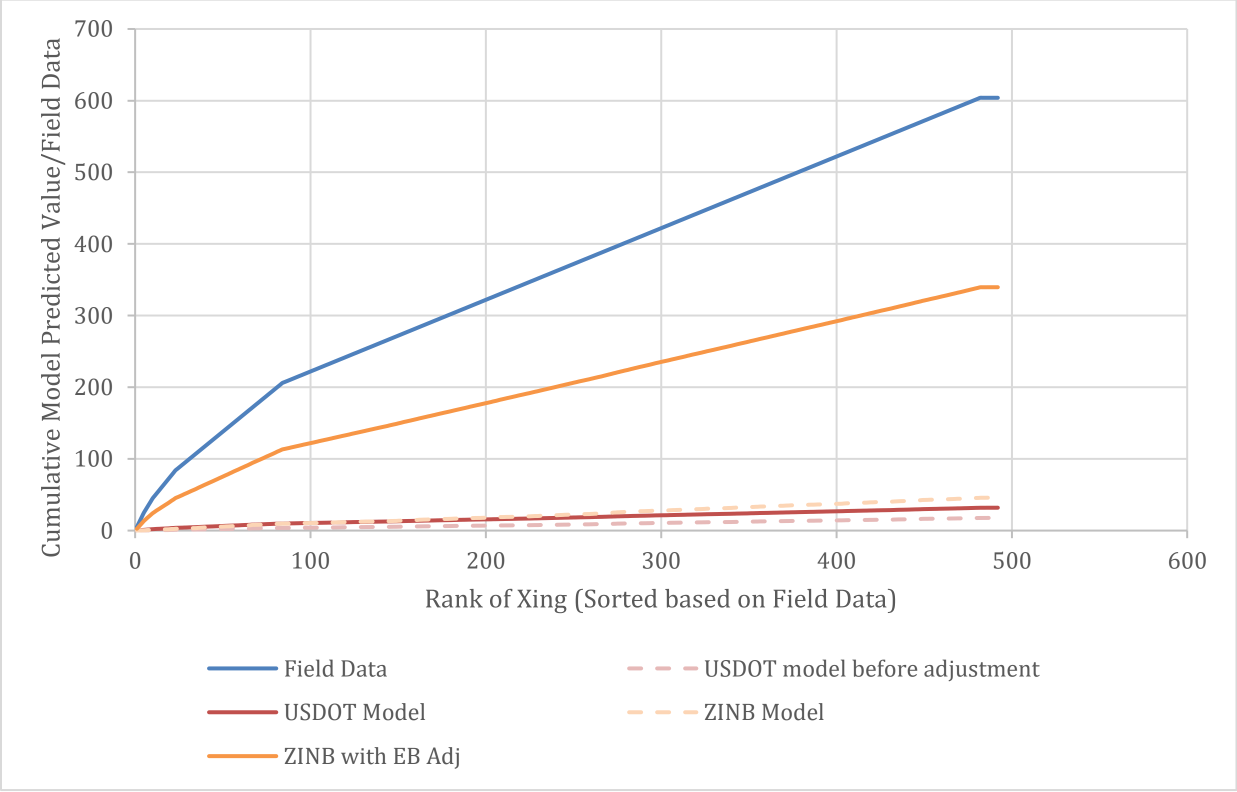


Figure 53: Cumulative number of accidents at top Locations for Crossings with Gates (Model Validation Dataset)

From the figure, the EB adjustment on the ZINB model can "pull" the model closer to the cumulative field data values. This contrasts with the behavior of the USDOT model and its adjustment where the USDOT model shows an overall deterioration in resembling the field data. The cumulative number of accidents as predicted by the USDOT model exceeds the observed number of accidents in the model development dataset. It is also noted that the ZINB model was fit to the field data, so the total sum of accidents, is expected to be total actual observed accidents, as it is seen in **Figure 50**. From **Figure 51**, at locations with accidents, the USDOT model gives a very similar performance to the unadjusted USDOT model. On the other hand, the ZINB model with EB adjustment improves the cumulative prediction of the unadjusted ZINB model at the top crossings. (Please note that the unadjusted USDOT model is calculated as: 'a', the initial USDOT prediction value, multiplied by the April 2013 normalizing constant for crossings with gates, 0.4846).

3.8.1.2 Crossings with Flashing Lights (and no Gates)

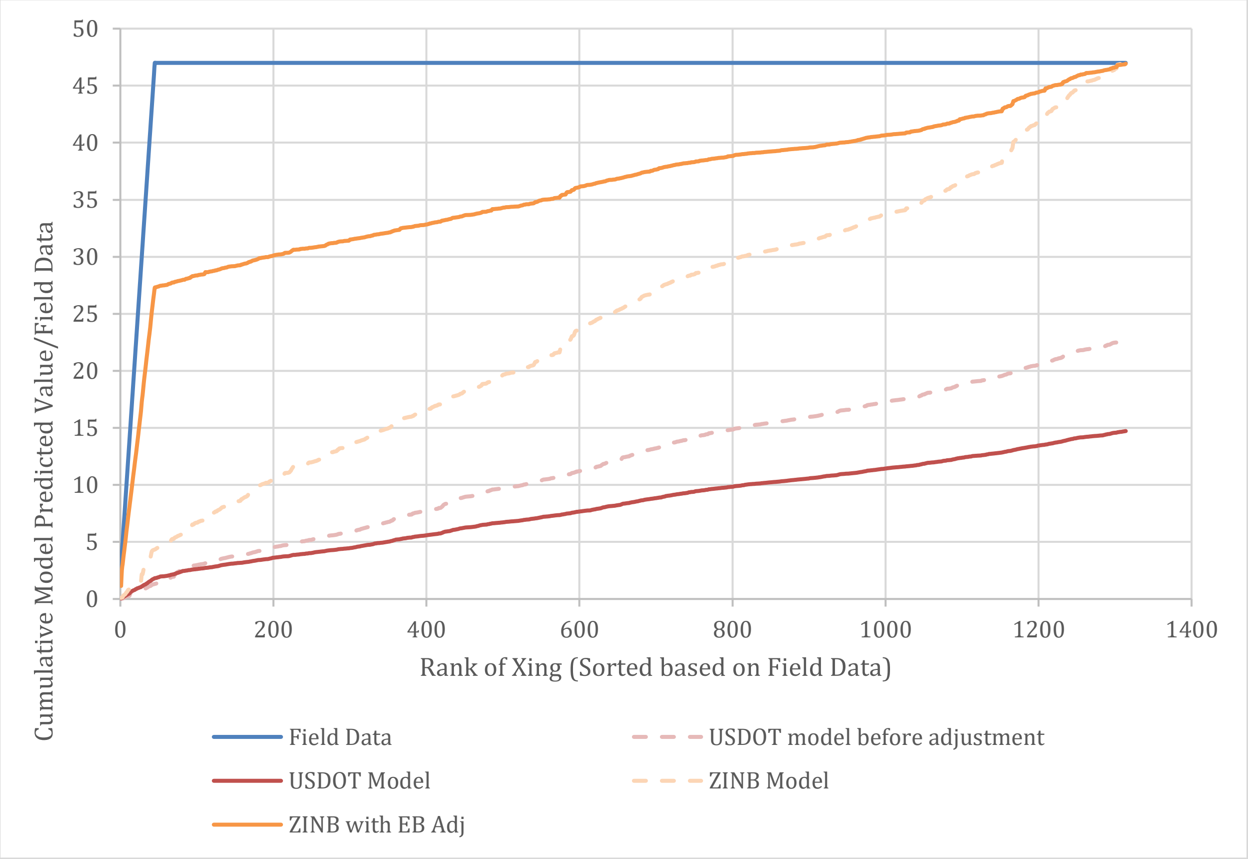


Figure 54: Cumulative number of Accidents for all locations for Crossings with Flashing Lights (Model Development Dataset)

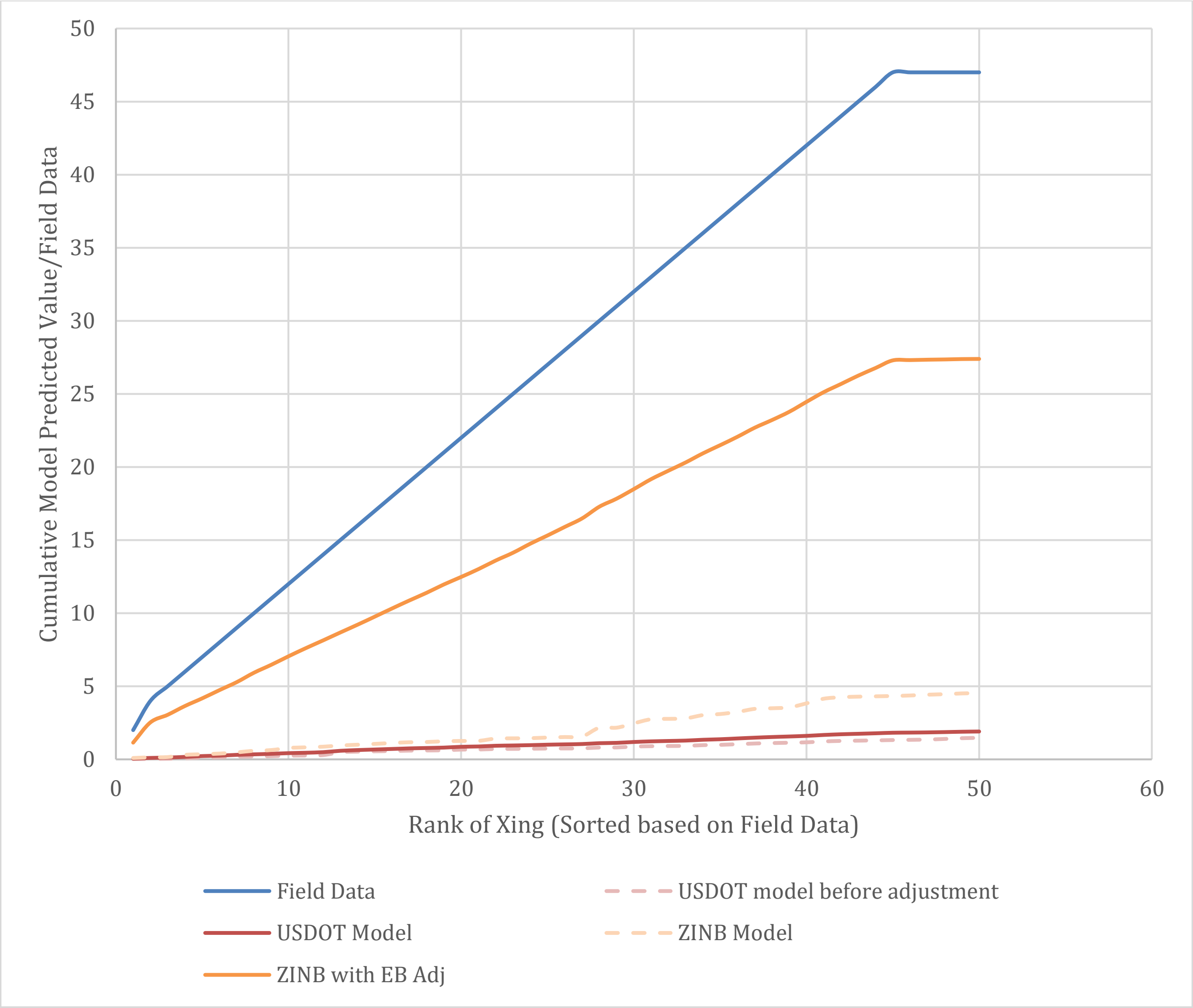


Figure 55: Cumulative number of accidents at top locations for Crossings with Flashing Lights (Model Development Dataset)

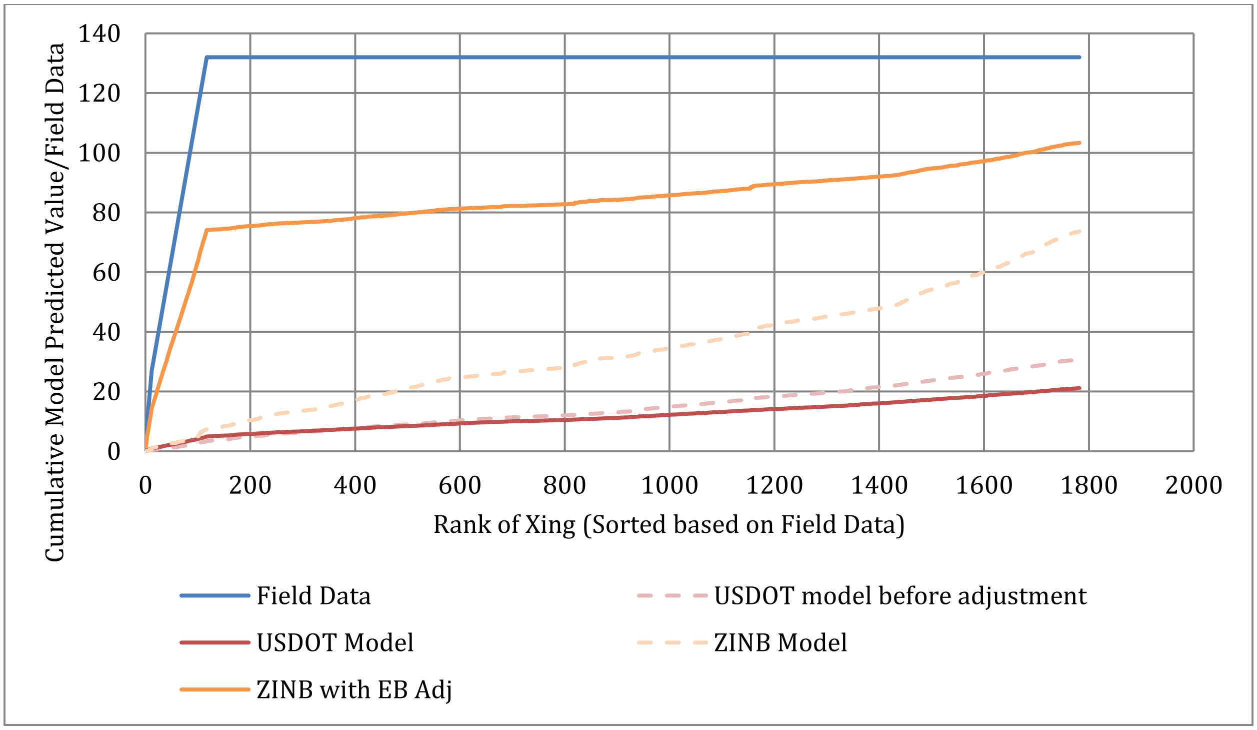


Figure 56: Cumulative number of accidents at all locations for Crossings with Flashing Lights (Model Validation Dataset)

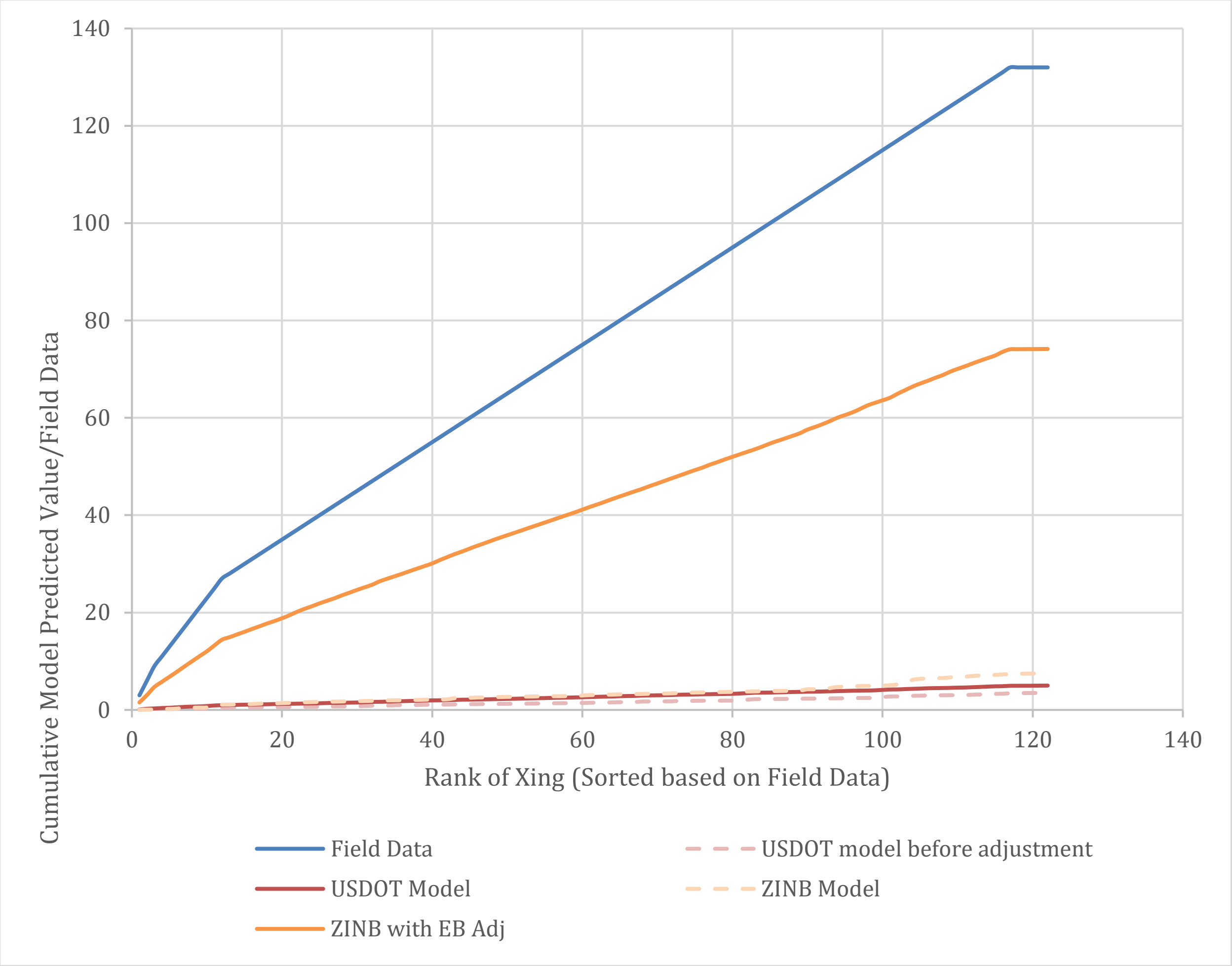


Figure 57: Cumulative number of accidents at top locations for Crossings with Flashing Lights (Model Validation Dataset)

For crossings with flashing lights, as seen from **Figures 54 to 57** above, the EB adjusted ZINB model shows the closest resemblance to the field data. The increasing trend in the cumulative number of accidents in the top Xings as shown by the EB adjusted ZINB model is comparable to the cumulative distribution of accidents in the field.

3.8.1.3 Crossings with Crossbucks

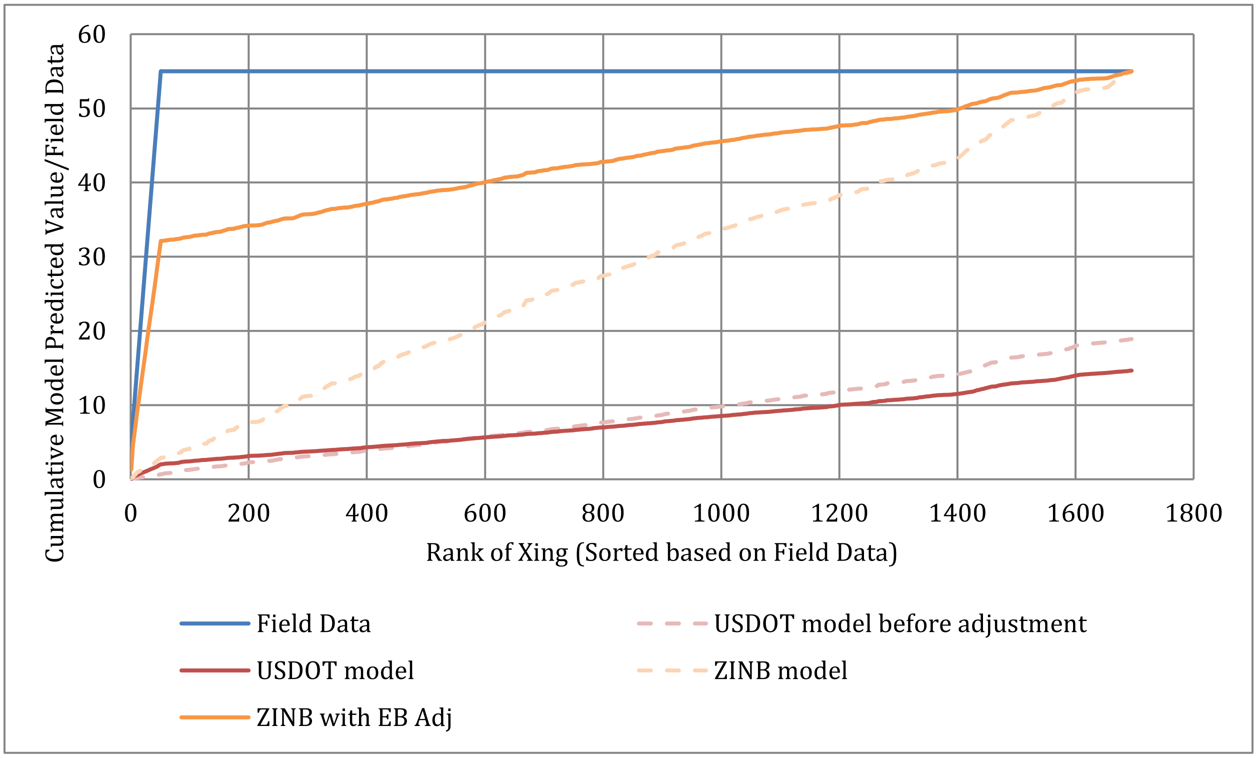


Figure 58: Cumulative number of accidents at all locations with Crossbucks (Model Development Dataset)

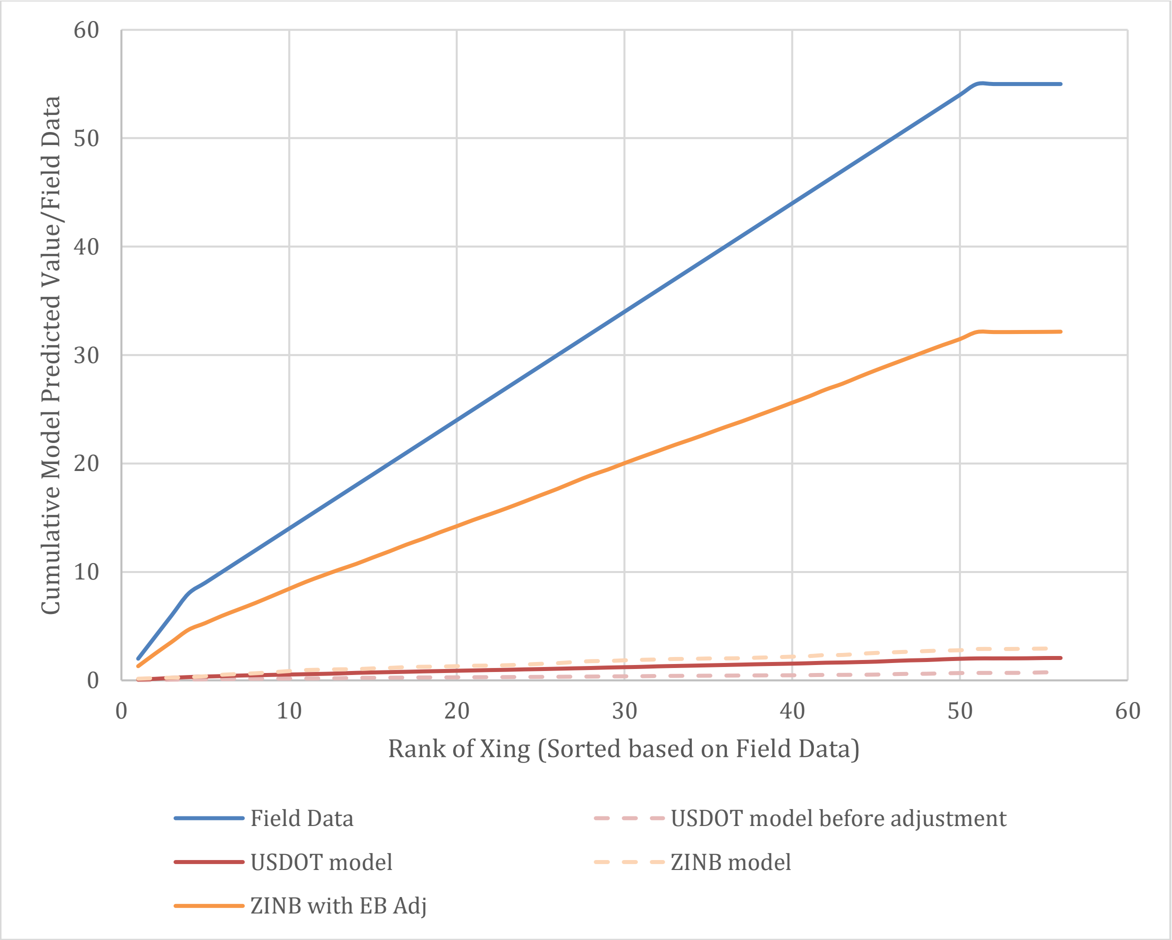


Figure 59: Cumulative number of accidents at top locations for Crossings with Crossbucks (Model Development Dataset)

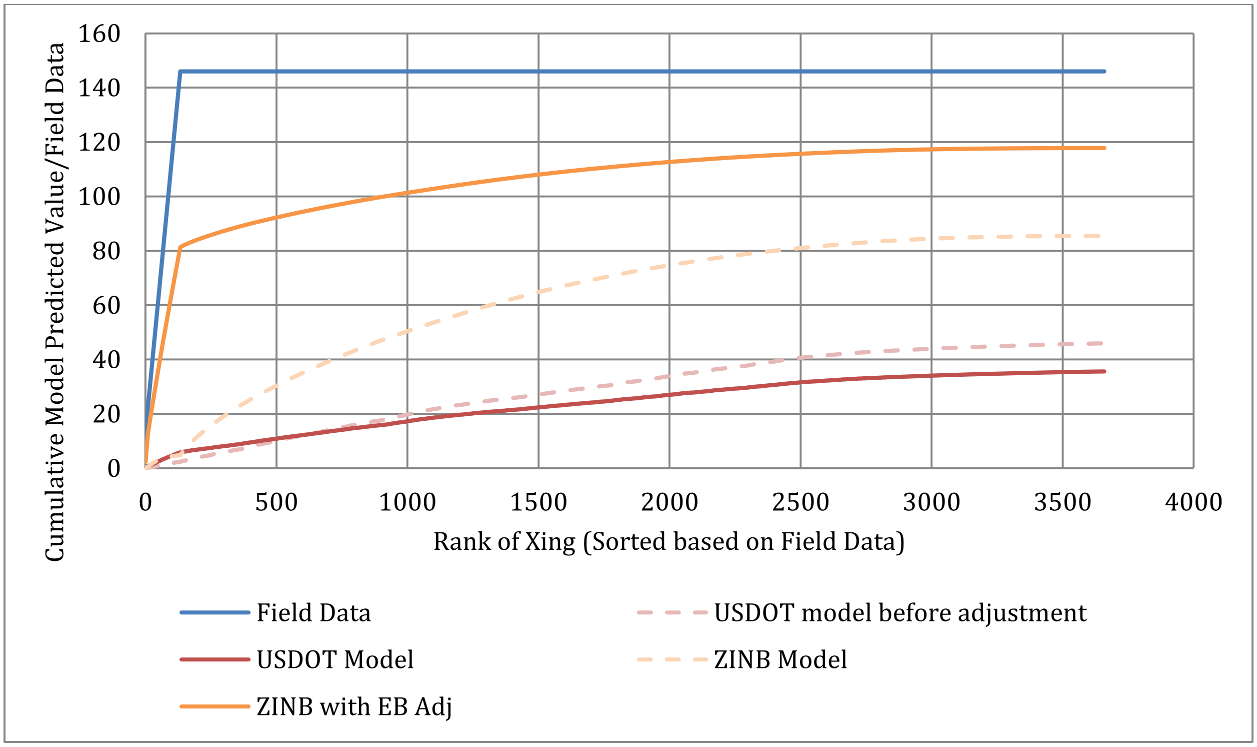


Figure 60: Cumulative number of accidents at all locations for Crossings with Crossbucks (Model Validation Dataset)

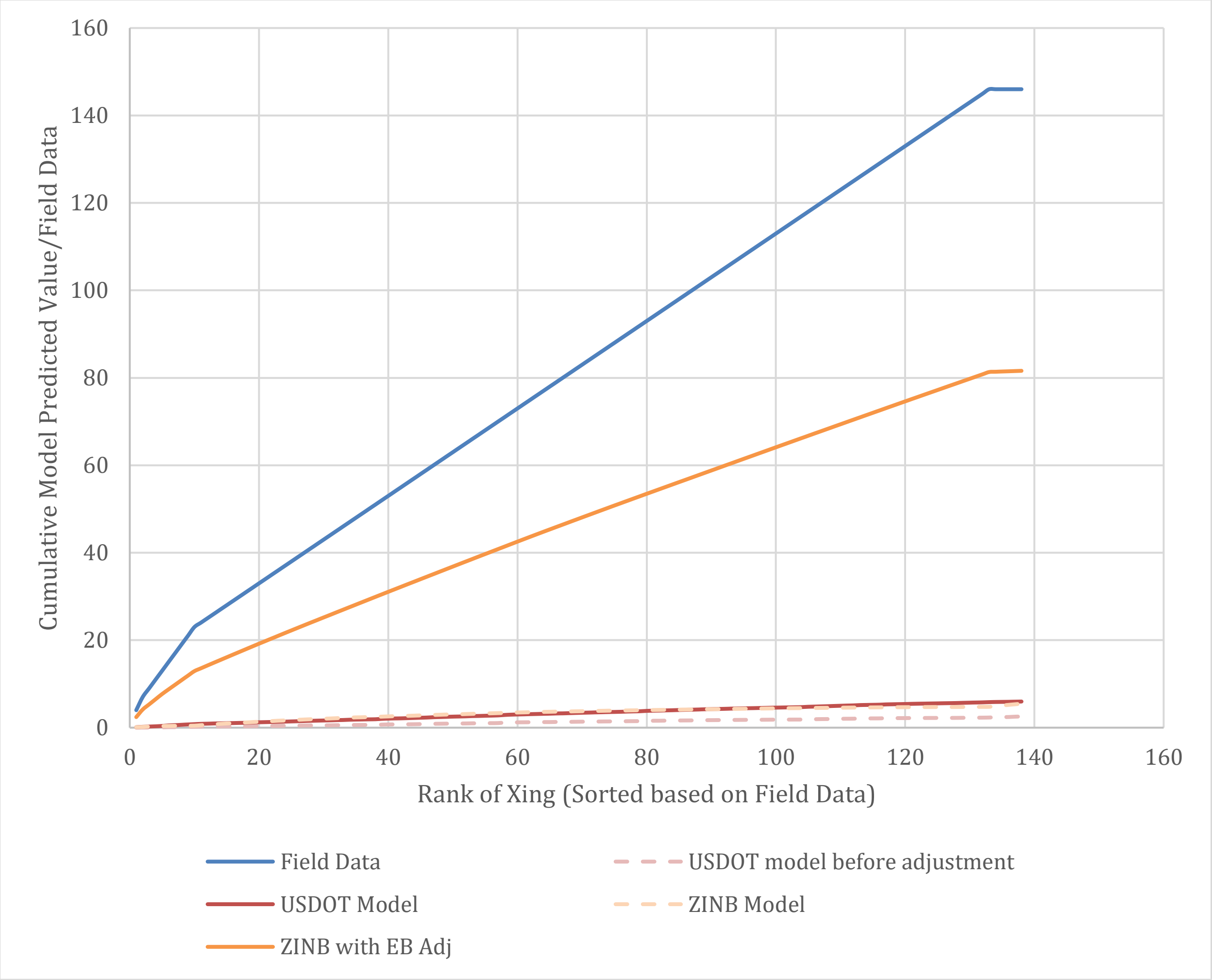


Figure 61: Cumulative number of accidents at top locations Crossings with Crossbucks (Model Validation Dataset)

Figures 58 to 61 shows the cumulative number of modelled and observed accidents for crossings with crossbucks for the model creation and model validation database respectively. For crossings with crossbucks as well, we can see that the cumulative accident distribution using EB adjusted ZINB model has the closest resemblance to the field data distribution among the models tried in both the model development and validation datasets.

In all the 6 cases (3 warning devices and 2 datasets), the EB adjusted ZINB model closely resembled the field data distribution.

3.8.2 Predicted number of accidents at a location with respect to actual number of accidents at location

3.8.2.1 Crossings with Gates

Comparisons were also made between the models with respect to the actual number of observed accidents at each crossing. **Figure 56 and Figure 57** shows the accident prediction for each crossing compared to the actual number of observed accidents using the EB adjusted ZINB models, DOT adjusted ZINB model and the USDOT models for crossings with gates in the model development and the model validation dataset.

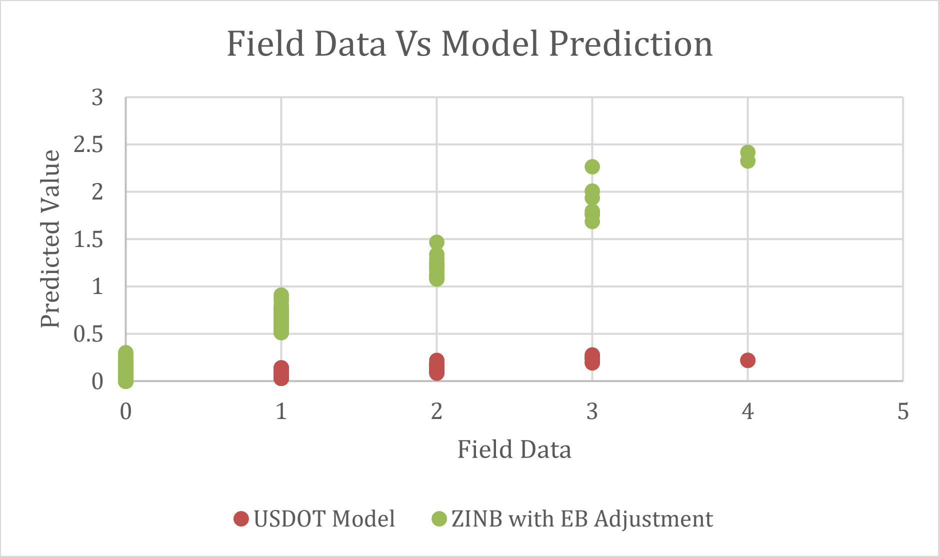


Figure 62: Predicted Vs Observed Accident Count for Gated Crossings (Model Development Dataset)

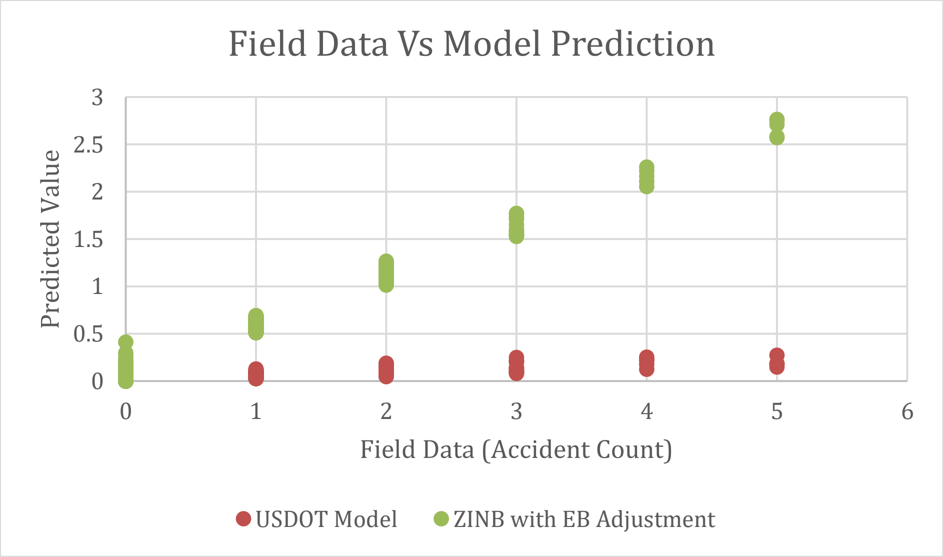


Figure 63: Predicted Vs Observed Accident Count for Gated Crossings (Model Validation Dataset)

In both **Figures 56 and 57**, the predicted accident count value using the EB adjusted ZINB model lies closer to the field data for both the model creation and validation dataset. This observation could be made for crossings with flashing lights and crossbucks in both the model creation and validation dataset as seen from **Figures 68 to 61**.

3.8.2.2 Crossings with Flashing Lights

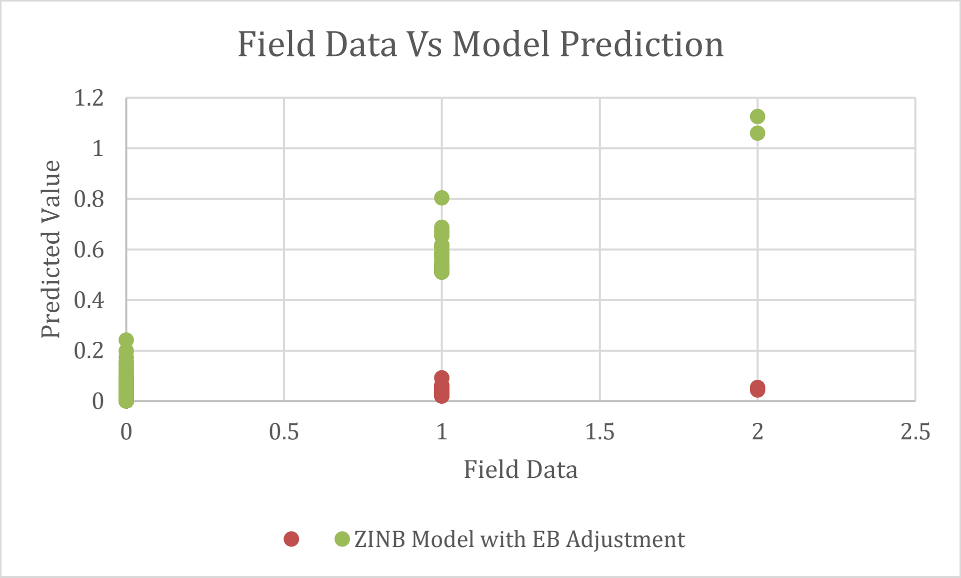


Figure 64: Predicted Vs Observed Accident Count for Crossings with Flashing Lights (Model Development Dataset)

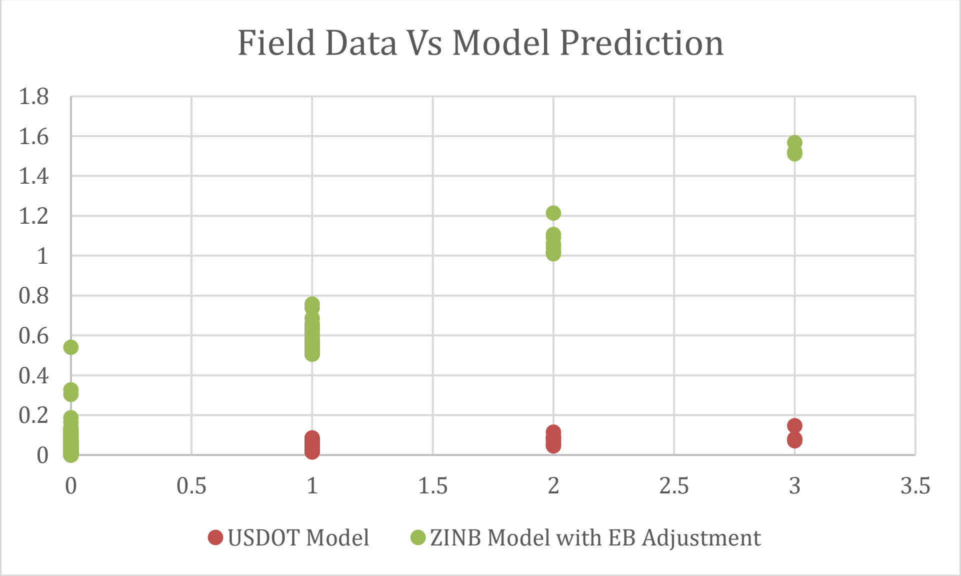


Figure 65: Predicted Vs Observed Accident Count for Crossings with Flashing Lights (Model Validation Dataset)

3.8.2.3 Crossings with Crossbucks

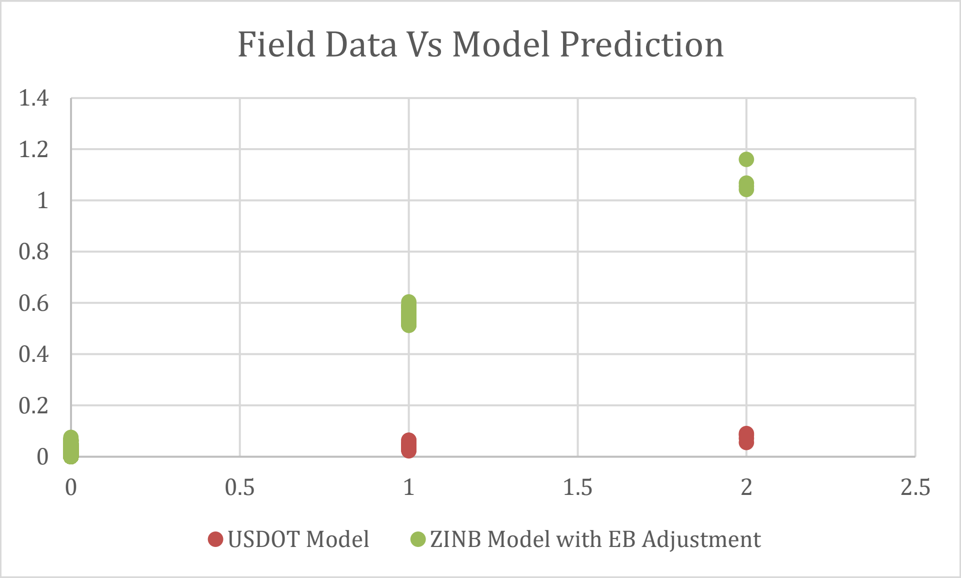


Figure 66: Predicted Vs Observed Accident Count for Crossings with Crossbucks (Model Development Dataset)

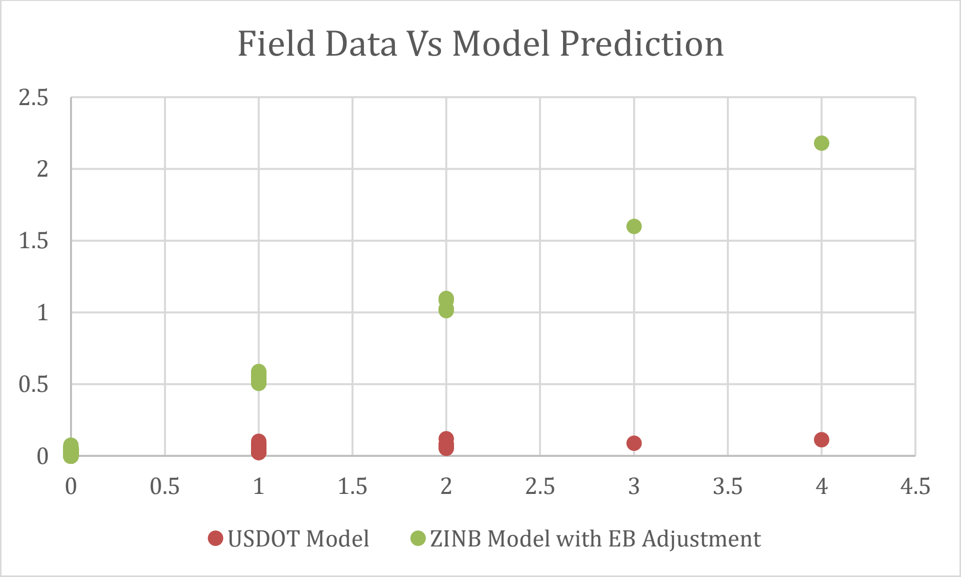


Figure 67: Predicted Vs Observed Accident Count for Crossings with Crossbucks (Model Validation Dataset)

3.8.3 Relative ranking of crossings based on the number of accidents identified among top locations.

The models were also used to rank the crossings based on the predicted accident count. The number of accidents observed at the top 'n' crossings as selected by the model are given in the following **Tables 16 and 17**. 'n' was selected in intervals of 10 and comparisons were made up to the top 50 crossings as ranked by each of the models. This was because in each dataset considered, no more than 50 crossings had observed two or more accidents in the past 5 years. An exception to this is for crossings with gates which had 85 crossings had two or more accidents.

Table 16: Number of Accidents Observed in top ‘n’ Crossings in Model Development Dataset

Warning Device	Ranking Method	Number of Crashes in Top Locations (Model Development Dataset)				
		Top 10	Top 20	Top 30	Top 40	Top 50
Gates	Field Data (Accident Count)	14	24	34	44	54
	USDOT Model	13	19	24	32	40
	ZINB Model with EB Adjustment	14	24	34	44	54
Flashing Lights	Field Data (Accident Count)	12	22	32	42	47
	USDOT Model	11	21	30	34	36
	ZINB Model with EB Adjustment	12	22	32	42	47
Crossbucks	Field Data (Accident Count)	31	51	67	77	87
	USDOT Model	31	47	59	71	81
	ZINB Model with EB Adjustment	31	51	67	77	87

Table 17: Number of Accidents Observed in top 'n' Crossings in Model Validation Dataset

Warning Device	Ranking Method	Number of Crashes in Top Locations (Model Development Dataset)				
		Top 10	Top 20	Top 30	Top 40	Top 50
Gates	Field Data (Accident Count)	45	75	98	118	138
	USDOT Model	36	68	89	110	125
	ZINB Model with EB Adjustment	45	75	98	118	138
Flashing Lights	Field Data (Accident Count)	23	35	45	55	65
	USDOT Model	18	32	42	52	61
	ZINB Model with EB Adjustment	23	35	45	54	64
Crossbucks	Field Data (Accident Count)	23	33	43	53	63
	USDOT Model	19	32	42	50	57
	ZINB Model with EB Adjustment	23	33	43	53	63

From **Table 16**, we can see that the EB adjusted ZINB model outperforms the USDOT model in identifying top crossings. Similar observations were also made for crossings with Flashing Lights and for crossings with crossbucks. Similar comparisons were also made in the model validation dataset as seen in **Table 17**. EB adjusted ZINB model could identify a higher number of accidents among the top crossings in both the datasets and for all the three warning device types. From both the datasets, the top crossings ranked by EB adjusted ZINB model had more number of accidents compared to the USDOT model. It can also be observed that the top xings identified using the ZINB model with EB adjustment had the same number of accidents as field data.

CHAPTER 4: CONCLUSIONS AND RECOMMENDATIONS

4.1 CONCLUSIONS

A microscopic method for accident analysis developed in this study offers a new tool to extract useful information from the FRA accident database for safety improvements at railroad crossings. The method creates a data-driven dynamic tree that allows the users to easily visualize any accident trends at a crossing or a group of crossings. Overall, the analysis of multiple accident locations using Modified Method B of the dynamic tree could provide information that otherwise it would be difficult to visualize.

The dynamic tree method was also used to identify variables for macroscopic analysis. The variables XAngle and HwyNDist were chosen for macroscopic analysis to develop a Zero-Inflated Negative Binomial model to improve accident predictions over the state of practice USDOT accident prediction model.

This study developed new Zero Inflated Negative Binomial model to improve the prediction of accident frequency at rail-highway grade crossings. The ZINB models were compared to the USDOT accident prediction formula. Three ZINB models were developed, one for each category of warning devices: Gates, flashing lights, and crossbucks. Like the USDOT formula, the predicted accident frequencies from the ZINB models were adjusted based on the accident history of the crossing to improve the prediction accuracy of the models. Empirical Bayes approach was implemented to adjust the crossing accident frequencies.

The models were developed using data from the state of Illinois. Then, they were validated using the data from four other states (IA, PA, SC, and TX). Different comparisons between the USDOT model and the corresponding ZINB model with EB adjustments were made. The crossing variables XAngle and HwyNDist were identified as significant through the macroscopic analysis for gated crossings. XAngle didn’t show up as a significant variable in the ZINB model for flashing lights and crossbucks even though the dynamic tree analysis indicated differences in accident characteristics.

The comparisons show that the newly developed ZINB models with EB adjustment resulted in cumulative predicted accident distributions that closely represent the field data. Also, plots of actual accident count and predicted accident counts by the two models showed that the EB adjusted ZINB accident prediction value was closer to the actual accident counts than the USDOT models. Similarly, more accurate predictions from the EB adjusted ZINB model were obtained for the top 10, 20, 30, 40 and 50 locations with highest accident frequency for all three warning devices. In summary, for all the three warning devices, the EB adjusted ZINB model outperformed the USDOT model.

4.2 RECOMMENDATIONS AND FUTURE WORK

In this study, the new multinomial models (ZINB models) were created using the most recent data from the FRA Database and all the site-related variables available in the database. Additional data from other sources should be combined with the FRA data and models should be developed. Some examples for additional sources include GIS databases, census data, etc.

The information available in the accident database remains unused in the accident prediction models. The authors are currently working on incorporating this available, unused accident information to fine tune the accident prediction models.

The researchers have validated the results using data from four other states. The researchers also recommend validating the developed model on each state separately.

Comparison between the EB model and the USDOT model for crossings showed that the EB model better represented the field data than the USDOT model. This Empirical Bayes approach should be explored further and should be considered with a “better” count model as an alternative to USDOT models to make accident count predictions at grade crossings.

REFERENCES

1. Operation Life Saver, retrieved from https://oli.org/about-us/news/collisions-casulties, October 2018.
2. Medina, J. C., Shen, S., and Benekohal, R.F. "*Microscopic analysis for accident data at railroad grade crossings*." In *T&DI Congress 2014: Planes, Trains, and Automobiles*, pp. 366-375. 2014.
3. Medina, J. C., and Benekohal, R.F., "*Macroscopic models for accident prediction at railroad grade crossings: comparisons with US Department of Transportation accident prediction formula.*", *Transportation Research Record: Journal of the Transportation Research Board* 2476, 85-93, 2015.
4. Federal Railroad Administration, 2015-January, Accident summaries as reported by the FRA, retrieved from http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/summary.aspx, January 2015.
5. Federal Railroad Administration, 2015-January, Accident and Inventory Database, retrieved from the FRA website http://safetydata.fra.dot.gov/, January 2015.
6. Medina, J., Shen, S., and Benekohal, R.F., "*Railroad Grade Crossing Micro-Level Safety Risk analysis*", University of Illinois, interim report prepared for NURail, April 2014.
7. Ogden, B.D., *Railroad-highway Grade Crossing Handbook*. No. FHWA-SA-07-010, United States. Federal Highway Administration, Washington D.C, 2007.
8. Hauer, E., *Observational before/after studies in road safety. Estimating the effect of highway and traffic engineering measures on road safety*, 1997.
9. Farr, E. H., *Summary of DOT Rail-Highway Crossing Resource Allocation Procedure-Revised*. No. DOT-TSC-FRA-86-2. United States. Federal Railroad Administration, 1987.
10. Wooldridge, M. D., Fambro, D. B., Brewer M.A., Engelbrecht, R.J., Harry, S.R., and Cho.H., *Design Guidelines for At-Grade Intersections Near Highway-Railroad Grade Crossings*. No. FHWA/TX-01/1845-3, 2000