

Modeling Crash Risk at Rail-Highway Grade Crossings by Track Class

Soumya Sharma, Srinivas S. Pulugurtha*

Department of Civil & Environmental Engineering, The University of North Carolina at Charlotte, Charlotte, NC, USA

Email: soumyasharmajec@gmail.com, *sspulugurtha@uncc.edu

How to cite this paper: Sharma, S. and Pulugurtha, S.S. (2019) Modeling Crash Risk at Rail-Highway Grade Crossings by Track Class. *Journal of Transportation Technologies*, 9, 261-275.

<https://doi.org/10.4236/jtts.2019.93016>

Received: March 22, 2019

Accepted: May 3, 2019

Published: May 6, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Federal Railroad Administration (FRA)'s Web Based Accident Prediction System (WBAPS) is used by federal, state and local agencies to get a preliminary idea on safety at a rail-highway grade crossing. It is an interactive and user-friendly tool used to make funding decisions. WBAPS is almost three decades old and involves a three-step approach making it difficult to interpret the contribution of the variables included in the model. It also does not directly account for regional/local developments and technological advancements pertaining to signals and signs implemented at rail-highway grade crossings. Further, characteristics of a rail-highway grade crossing vary by track class which is not explicitly considered by WBAPS. This research, therefore, examines and develops a method and models to estimate crashes at rail-highway grade crossings by track class using regional/local level data. The method and models developed for each track class as well as considering all track classes together are based on data for the state of North Carolina. Linear, as well as count models based on Poisson and Negative Binomial (NB) distributions, was tested for applicability. Negative binomial models were found to be the best fit for the data used in this research. Models for each track class have better goodness of fit statistics compared to the model considering data for all track classes together. This is primarily because traffic, design, and operational characteristics at rail-highway grade crossings are different for each track class. The findings from statistical models in this research are supported by model validation.

Keywords

Rail-Highway Grade Crossing, Crash, Risk, Modeling, Track Class

1. Introduction

A rail-highway grade crossing works as an at-grade junction to allow for traffic

movement between railroads and highways. These are one of the most vulnerable spatial locations in a rail transportation system. The United States has about 129,644 public crossings [1]. Data for the year 2010 show that there were 11,555 incidents at public rail-highway crossings in the United States. These incidents resulted in 746 deaths and 8307 injuries [2]. The estimated cost of a rail-highway crash is about \$2.6 million [3].

The Federal Railway Administration (FRA) has undertaken many measures to improve the safety at rail-highway grade crossings. The installation and maintenance of these measures are expensive. The cost of a warning device installation at each rail-highway grade crossing exceeds \$250,000 in today's dollars [4]. Therefore, it is crucial for the agencies to make intelligent as well as informed engineering decisions to effectively prioritize the implementation plans and improve safety at such conflicting locations. The successful implementation of such engineering decisions depends on current techniques to model and assess risk at rail-highway grade crossings.

The FRA provides users with an analytical tool called the Web Based Accident Prediction System (WBAPS). The intent of the tool is to help each individual state agency, railroad agency, and local highway authority with allocating funds for safety improvement. The WBAPS is based on the United States Department of Transportation (USDOT) FRA accident (hereafter referred to as "crash") prediction formula. This formula was developed in April 1968 and revised in June 1987 to address the shortcomings of previous models-Peabody Dimmick Formula, New Hampshire Index, and the National Cooperative Highway Research Program (NCHRP) Hazard Index [4].

The FRA crash prediction formula uses: 1) basic data about a rail-highway grade crossing's physical and operating characteristics, and, 2) five years of crash history data at the rail-highway grade crossing. A basic formula was first developed, using non-linear regression methods, based on the physical characteristics of the rail-highway grade crossing [4]. The basic formula was updated considering the un-normalized crash prediction based on the crash history [4]. The constants for each category of the warning device are multiplied with the un-normalized crashes to obtain the final prediction. The normalization constants are updated on a yearly basis. This is done to adjust for the change in the number of crashes and any warning device changes at the rail-highway grade crossings [4].

Approximately, two-thirds of rail-highway grade crossings have not had a crash in the last 5 years, while 93% have had two or fewer crashes [5]. The weighted formula described previously uses the crash history of rare events in order to predict the number of crashes in the future. Mutabazi and Berg [6] tested the various versions of FRA rail-highway grade crossing crash prediction formula for their accuracy. Their findings indicate that the basic formula performed better when compared to the five-year crash history adjusted model [6]. Except for the aforementioned study, literature documents no research on a comparison of the FRA rail-highway crash predictions of a year and the actual

number of crashes at a rail-highway grade crossing.

The three-step process of the FRA crash prediction formula includes 128 explanatory variables but has its own shortcomings. The formula was developed almost three decades ago and has been used since without much improvement (apart from updating coefficients). The formula only gives a preliminary idea to the decision makers to allocate resources. It is quite complex and difficult to interpret in terms of the most influencing factors of safety at a rail-highway grade crossing. It does not take into consideration regional/local level geographic and other site-specific data such as sight-distance, highway congestion, local topography, and passenger exposure (train or vehicle).

The causes of crashes, driver behavior, geometric features, topographical conditions, and the presence of safety devices at rail-highway grade crossings vary for one state to another in the United States. As an example, North Carolina has active warning devices at more than 50% of its public rail-highway grade crossings. Therefore, the warning device criteria as in the FRA rail-highway grade crossing crash prediction formula may be of little use to identify hazardous rail-highway grade crossings in North Carolina. The formula could also be simplified if an analysis is performed and a method/model developed using state or regional-level data.

WBAPS does not explicitly consider track class. Since design and operational characteristics vary by track class, developing models by track class may yield more meaningful results and assist rail practitioners. This research, therefore, focuses on the development of rail-highway grade crossing crash prediction models, using regional/local level data, by track class as well as considering data for all track classes.

2. Literature Review

Researchers have adopted various methods to develop crash prediction formulas for rail-highway grade crossing safety improvement. Negative Binomial (NB) crash prediction models were developed for rail-highway grade crossings using a simple one-step process [7]. NB distribution-based model was also found to be the best fit for the data to identify rail-highway grade crossing blackspots for three categories (passive, flashing lights, and gates) [8].

The relation between the number of crashes and characteristics of rail-highway grade crossings was also observed through the use of a gamma distribution-based model [9]. The results from their study showed that crashes would increase with an increase in the total traffic volume and the average daily train volume. Further, the proximity of an industrial area and the time between signal and gate activation was observed to be associated with higher crash frequencies [9].

Zero-inflated models were also developed, to examine the role of factors affecting rail-highway grade crossing crashes, to tackle data scarcity even with a large number of rail-highway grade crossings [10] [11]. The literature also documents the application of logistic regression models to observe the trends in the

number of crashes at rail-highway grade crossings over time [5]. Stepwise regression analysis was also adopted to develop the rail-highway grade crossing crash prediction formula that aids in prioritizing signal improvements [12].

The above discussion on models to estimate crash risk at rail-highway grade crossings clearly indicates that a single statistical distribution may not be applicable to all datasets or locations. It also emphasizes on the development of a method and models that best fit the data, accounting for factors at the regional/local level.

Literature also documents research to develop methods or examine the effect of the countermeasures on rail-highway grade crossing safety. Park and Saccomanno [13] examined the interactions between various countermeasures (such as warning devices and the posted speed limit) on safety at rail-highway grade crossings. They also studied the effect of a less explored combination of countermeasures and control measures (highway class) on crash frequency using a sequential analytical strategy. This strategy combines the tree-based regression stratification of data with generalized linear regression models [14].

Sacomanno and Lai [15] categorized the rail-highway grade crossing inventory variables into non-linear factors and assigned scores. The scores were used to cluster rail-highway grade crossings and then develop a separate model for each cluster using explanatory variables relevant to that cluster [15]. Bayesian data fusion method was used to tackle the problem of sparse crash data when evaluating countermeasure effectiveness. The method used previous research inferences for countermeasure effectiveness along with a calibrated model of the study area to finally generate a collision response and probability distribution for each countermeasure [16].

The type of countermeasure also plays a role in safety at rail-highway grade crossings. As an example, Yan *et al.* [17] showed that stop-sign treatment is an effective countermeasure to improve safety at rail-highway grade crossings. Likewise, upgrading flashing lights to gates on single track may be more effective than at a rail-highway grade crossing along multiple tracks. However, the train speed variation did not have much influence on the effectiveness of the upgrade [18].

While some new models for rail-highway crash prediction were developed, past research did not primarily account for regional/local factors, which influence the crash trends to a great extent. The States (California, Texas, Illinois, Georgia, and New York) that have been chosen for research in the past are usually the ones with high train traffic. States such as North Carolina with relatively less train traffic may have different types of challenges. Considering such diverse geographic patterns is important. Therefore, this research aims at developing an approach to predict crashes at a rail-highway grade crossing based on regional/local data. Also, unlike most of the prediction models developed in the past, the models developed in this research do not make use of crash history information. This is mainly because crashes at rail-highway grade crossings in the study area are rare events, which makes the crash history of little use when pre-

dicting crashes in the future. Such an approach will also help when planning, designing and building new tracks with rail-highway grade crossings.

Funds available and countermeasures implemented at rail-highway grade crossings vary based on train activity-levels and track design characteristics in addition to the risk. These characteristics differ for each track class. On a different note, analyzing and modeling by track class could yield better results rather than developing models considering data for all track classes in a region. This research addresses the aforementioned aspects to add to the current state of knowledge on safety at rail-highway grade crossings.

3. Methodology

The methodology to model crash risk at rail-highway grade crossings is comprised of five steps. Each of those steps is discussed next in detail.

3.1. Selection of Rail-Highway Grade Crossings

The selection of rail-highway grade crossings needs to be performed so as to have the best representative sample of the population of all the rail-highway grade crossings in the study region. The selection should comprise of rail-highway grade crossings with zero as well as more than zero crashes. Likewise, the representative sample should have a fair distribution of rail-highway grade crossings pertaining to all the track classes.

3.2. Selection of Explanatory Variables

The explanatory variables considered should represent the characteristics of the highway, rail-track and the types of warning devices at the rail-highway grade crossings. This research tried not to use minimal warning device variables so as to avoid endogeneity, which means that the cause of crashes is the reason a particular warning device is installed at a rail-highway grade crossing. The selection of the variables in this research is mainly based on the correlations between the variables and the dependent variable (“crashes per five years”) and amongst the other variables considered for the analysis.

3.3. Development of Crash Risk Estimation Models

The dependent variable for all the models is the “number of crashes per five years” at a rail-highway grade crossing. Crash count models were primarily explored in this research. Poisson, NB and Gamma log-link distribution-based models are the popular count models. While count models provide a sensible output, they suffer from certain limitations. The Poisson model assumes the mean and variance to be equal, while the NB model is capable of handling data with variance greater than the mean (over-dispersed). The Gamma model, however, is capable of dealing with both over-dispersed and under-dispersed data.

In this research, the analysis was conducted using SPSS software [19], in which the Gamma model excludes the zeroes in the dependent variable while

modeling. As both the zero as well as non-zero values of the dependent variable are crucial, the use of a Gamma log-link distribution-based model has been excluded in this research. Researchers have considered zero-inflated models when studying crash data in the past. The zero-inflated NB model could be a special case of the NB model, and the difference in performance might be trivial [20]. For this reason, only Poisson and NB distributions are only discussed in this research.

The Akaike's Information Criterion (AIC) was used to assess the quality of various statistical models developed from the same data. The statistic provides an estimate of the information that has been lost as a result of using a particular model that generates the data. Given a set of candidate models for the data, the best model is the one with the minimum AIC value.

The Corrected Akaike's Information Criterion (AICC) was also checked to ensure that the model does not tend to over-fit the results. In general, the difference between AIC and AICC should be as low as possible.

In addition, the likelihood ratio Chi-Square and Deviance values were also computed and considered to assess the goodness-of-fit of the developed models.

The probability value of the selected explanatory variables was also tested at a 95% confidence level (significance value ≤ 0.05).

3.4. Validation of the Models

The best-fitting model was then validated using data set aside for model validation (not used for the model development). The number of crashes at each selected rail-highway grade crossing is computed and compared with the actual number of crashes at the rail-highway grade crossing. To test the predictability of models compared to WBAPS, the number of crashes were compared to the analogous term "number of collisions per year" from the WBAPS output.

A t-test was then conducted in order to check if the two groups of data belong to the same population or not. The null hypothesis is that the two groups being tested are statistically different while the alternate hypothesis is that the two groups are not statistically different. The null hypothesis cannot be rejected if the P-value is less than 0.05 (at a 95% confidence level).

4. Data

The data collected for this research includes two databases: 1) The FRA Office of Safety rail-highway grade crossing inventory, and, 2) The FRA Office of Safety crash/incident database, both for the state of North Carolina. The rail-highway grade crossing inventory provides site-specific details of the rail-highway crossing and highway characteristics—the number of daily through trains, warning devices, annual average daily traffic (AADT), and the posted highway speed limit. The crash history data is available for each year. This database includes details of each incident at any of the operational rail-highway grade crossing in that year. The database also includes the type of railway equipment involved in the

crash (freight train, passenger train, and inspection car) and the circumstance of the crash (if the rail-user was struck by the train or vice-versa). Crash history from the year 2009 to the year 2013 was considered to develop models in this research. Only rail-highway grade crossing where conditions remained same over this five-year period were selected for analysis and modeling.

The rail-highway grade crossings were identified using the unique rail-highway grade crossing ID number. This number is a common element in both the databases. The rail-highway grade crossing ID was used to merge the rail-highway grade crossing inventory data with the crash frequency information from the crash/incident database to generate a database that was used for further analysis. Almost 97% of rail-highway grade crossings in the study area have warning devices installed at them. In such a case, including variables related to warning devices may pose endogeneity issues. In this study, it would imply that the presence of warning devices may result in zero crashes (which are the frequency of crashes found in abundance at rail-highway grade crossings) and the zero crashes are caused at a rail-highway grade crossing due to the warning devices installed at these locations. Hence, warning device variables were not included in the models as far as possible. They also were observed to be correlated to other variables considered for modeling in this research.

All rail-highway grade crossings without data for a five-year period were removed from the database and further analysis. In addition, only public and at-grade rail-highway grade crossings were retained in the database.

The data had certain variables that were categorical in nature. For example, “highway near crossing” had four fields—less than 75 ft, 75 to 200 ft, 200 to 500 ft, and no highway nearby. These variables were reduced to indicator variables *i.e.*, one variable for each of the four fields. Also, AADT was converted to a rate of per 10,000 vehicles. All other continuous variables were used in the analysis without any changes.

Based on the FRA guidelines [21], the following range of train time table speed was used for track classification: 0 - 10 mph—Track class 1; 10 - 25 mph—Track class 2; 25 - 40 mph—Track class 3; 40 - 60 mph—Track class 4; 60 - 80 mph—Track class 5. Data was segregated based on each track class, forming five subsets; one for each class.

Overall, the dataset considered had 681 rail-highway grade crossings in track class 1; 1432 rail-highway grade crossings in track class 2; 870 rail-highway grade crossings in track class 3; 656 rail-highway grade crossings in track class 4; and 133 rail-highway grade crossings in track class 5. About 20% of the rail-high grade crossings were randomly selected for each track class and set aside for the model validation.

The data from each track classes were combined for comparing the results for each track class model with a model for all track classes data taken together.

5. Variable Selection to Develop Models

Table 1 summarizes the list of variables considered in this research. The correlation

Table 1. Variables considered for analysis and modeling.

| Variable | Description |
|----------|--|
| CROSSING | Crossing ID |
| Crashes | # of Crashes |
| LTM | Less than one train movement per day; 1 = yes, 0 = no |
| MTS | Maximum time table speed |
| TRKCLS | Track class (1 - 5) |
| MNS | Minimum train speed |
| MTK | # of main tracks |
| OTK | # of other tracks |
| WDGD | Warning device code 1 = no signs or signals, 2 = other signs or signals, 3 = crossbucks, 4 = stop signs, 5 = special active warning devices, 6 = wigwags, bells 7 = flashing lights, 8 = all other gates (two and three quadrant gates*), 9 = four quadrant (full barrier) gates |
| BL | # of bells |
| NS | 1 = no signs or signals; 0 = at least one sign or signal |
| SGEQ | Is track equipped with train signals? 1 = yes, 0 = no |
| OS | Development type open; 1 = yes, 0 = no |
| RS | Development type residential; 1 = yes, 0 = no |
| COM | Development type commercial; 1 = yes, 0 = no |
| INDUS | Development type industrial; 1 = yes, 0 = no |
| INST | Development type institutional; 1 = yes, 0 = no |
| STPL | If stop lines are present; 1 = yes, 0 = no |
| RRX | If rail road crossing symbol is present; 1 = yes, 0 = no |
| NMK | If there are no pavement markings; 1 = yes, 0 = no |
| STPL | If there are stop lines and rail-road crossing signals; 1 = yes, 0 = no |
| L75 | If the highway is less than 75 ft away; 1 = yes, 0 = no |
| B200-500 | If the highway is in the vicinity of 200 to 500 feet; 1 = yes, 0 = no |
| B75-200 | If the highway is in the vicinity of 75 ft to 200 feet; 1 = yes, 0 = no |
| NHWY | If there is no highway nearby; 1 = yes, 0 = no |
| TRFLN | # of trafficlanes |
| STHWY | Is crossing on state highway? 1 = yes, 0 = no |
| AADT | Average annual daily traffic |
| PCTRK | % of truck traffic |
| SCHLB | Average number of school buses passing through the crossing on a school day |
| WHISTB | If there is a whistle ban; 1 = 24 hr, 0 = no ban |
| TTRN | Total number of trains |
| TSWT | Total number of switching trains |
| XBUCK | # of crossbucks |
| GT | # of gates |
| FLP | # of flashing light pairs |
| HWYSP | Posted highway speed limit |

*Three quadrant gates: gates at a rail-highway grade crossing along with a median on the approach to the rail-highway grade crossing that only has a gate on the entrance lane.

between these variables as well as with the number of observed crashes during the five-year period was examined by constructing a Pearson correlation coefficient matrix. The examination was done for Pearson correlation coefficient matrix for the all track class dataset as well as a dataset for each track class.

The maximum train time table speed was considered as a key variable influencing safety and risk at the rail-highway grade crossings. The maximum train time table speed was included in the analysis and modeling process for the model considering data for all rail-highway grade crossings. The number of main tracks and AADT were also forced into the models. However, the maximum train time table speed was not forced into the model for each track class as the track class is based on the maximum train time table speed. The AADT and/or the number of main tracks were selected as the key variables influencing safety risk at the rail-highway grade crossing in this case. The variables that were found not to be correlated to the key variables (at a 95% confidence level) were identified and used in the development of models (only if correlated with the observed number of crashes during the five-year period).

Table 2 summarizes the explanatory variables selected based on correlation to develop models for each track class.

6. Analysis and Results

Track class 1 has two-quadrant gates and crossbucks installed at 38.3% and 30.9% of the total rail-highway grade crossings. Similarly, track class 2 has 51.4% and 31.5% rail-highway grade crossings with two quadrant gates and crossbucks, respectively. Track class 1 has 15.6% rail-highway grade crossings with flashing lights installed, while track class 2 has a higher number *i.e.*, 121 rail-highway grade crossings with flashing lights installed. Further, for track class 3 and above, more rail-highway grade crossings have flashing lights and gates installed at them rather than just crossbucks. Track class 5 has 92.7% of its total rail-highway

Table 2. Variables considered for developing each track class model.

| Track Class | Variable |
|-------------|--|
| 1 | 1) AADT 2) Total number of trains 3) # of main tracks |
| 2 | 1) If stop lines present 2) # of traffic lanes 3) % of trucks |
| 3 | 1) # of main tracks 2) Total number of switching trains 3) No highway nearby |
| 4 | 1) # of main tracks 2) No highway nearby |
| 5 | 1) # of main tracks 2) # of traffic lanes |

grade crossings with two quadrant gates. Four quadrant gates seem to be rarely installed at rail-highway grade crossings in the study area. The warning devices across track classes are justified as track class is related to the speed of the train.

Track classes 1, 2 and 3 have rail-highway grade crossings with mostly zero or one main track while track classes 4 and 5 have a few rail-highway grade crossings with one or two main tracks. There are fairly a low number of rail-highway grade crossings in any of the classes with three or four main tracks. Also, a higher number of rail-highway grade crossings with one main track is found to have a higher number of reported crashes at them. This can be mainly due to the abundance of rail-highway grade crossings in this category (# of main tracks = 1) or may be due to some other unexplainable factor. The rail-highway grade crossings with two main tracks have more than 90% of the rail-highway grade crossings with two quadrant gates.

6.1. Modeling Based on All Track Class Data

Models were first developed considering data for all the track classes together. The model developed is shown in **Table 3**.

The number of main tracks is positively correlated to crashes, while the highway speed limit has a negative coefficient (possibly because warning devices and signals are provided at rail-highway grade crossings with the higher posted speed limit on the highway). The model also has a negative intercept. The significance value for the likelihood ratio Chi-Square is less than 0.01. The AIC and AICC are equal to each other for both the models. However, the NB distributed-based model has marginally lower AIC, AICC, and Deviance values than the Poisson distribution-based model. Further, the computed variance is greater than the mean. Therefore, NB distribution-based model is considered to better fit the data used in this research.

Table 3. All track class data model.

| Variable | Poisson (Log-link) | | Negative Binomial (Log-link) | |
|-----------------------------|--------------------|-------|------------------------------|-------|
| | C | P | C | P |
| Intercept | -3.92 | <0.01 | -4.17 | <0.01 |
| # of main tracks | 1.49 | <0.01 | 1.68 | <0.01 |
| Posted highway speed Limit | -0.02 | 0.03 | -0.02 | 0.05 |
| Likelihood Ratio Chi-Square | 90.2 | <0.01 | 84.4 | <0.01 |
| AIC | 1176.6 | | 1157.4 | |
| AICC | 1176.6 | | 1157.4 | |
| Deviance | 891.8 | | 760.7 | |

*C is coefficient and P is probability or significance value.

6.2. Modeling Based on Each Track Class Data

The crash distribution, the mean and the variance for each track class are shown in **Table 4**. The data for track classes 1 and 5 were found to be under-dispersed, while the data for track classes 2, 3, and 4 were found to be over-dispersed. Since the model based on all track classes data and three out of five track classes have variance greater and then the mean, NB distribution-based models are developed for each class and summarized in **Table 5**.

From **Table 5**, the explanatory variables that have an effect on crashes at rail-highway grade crossing vary by the track class. The AIC and AICC are reasonably close to each other for each track class model. The significance value for the likelihood ratio Chi-Square is less than or equal to 0.01 for each track class model. The AIC, AICC, and Deviance values are lower than the corresponding value for the model based on all track class data. This indicates that developing models for each track class may lower prediction errors and improve accuracy than compared to all track class data model.

Table 4. Descriptive statistics of data based on track class.

| Track Class | 1 | 2 | 3 | 4 | 5 | |
|--------------------------|---------|---------|---------|---------|---------|-----|
| Mean | 0.02202 | 0.02269 | 0.0546 | 0.14095 | 0.06024 | |
| Variance | 0.02153 | 0.02392 | 0.06024 | 0.20147 | 0.05661 | |
| Frequency (# of Crashes) | 0 | 97.80% | 97.90% | 95.40% | 89.30% | 94% |
| | 1 | 2.20% | 2.09% | 4.60% | 7.80% | 6% |
| | 2 | 0% | ~0% | ~0% | 2.30% | 0% |
| | 3 | 0% | 0% | 0% | 0.60% | 0% |

Table 5. Models by track class.

| Variable | C | | P | | C | | P | | C | | P | |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|--|
| Track Class | 1 | | 2 | | 3 | | 4 | | 5 | | | |
| Intercept | -4.77 | <0.01 | -4.99 | <0.01 | -5.35 | <0.01 | -3.37 | <0.01 | | | | |
| Total # of trains | 0.06 | <0.01 | | | | | | | | | | |
| # of main tracks | 0.61 | 0.41 | | | 2.24 | <0.01 | 1.23 | <0.01 | -1.89 | <0.01 | | |
| No highway nearby | | | | | -1.23 | 0.03 | -0.86 | 0.02 | | | | |
| Total # of switching trains | | | | | 0.20 | 0.02 | | | | | | |
| Stop lines present | | | 1.47 | 0.01 | | | | | | | | |
| # of Traffic Lanes | | | 0.37 | 0.04 | | | | | | | | |
| % of trucks | | | 0.04 | 0.01 | | | | | | | | |
| Likelihood Ratio Chi-Square | 8.9 | 0.01 | 12.9 | 0.01 | 30.1 | <0.01 | 34.0 | <0.01 | 70.3 | <0.01 | | |
| AIC | 112.9 | | 244.6 | | 276.9 | | 420.0 | | 53.7 | | | |
| AICC | 113.0 | | 244.6 | | 276.9 | | 420.0 | | 53.7 | | | |
| Deviance | 73.7 | | 166.2 | | 168.7 | | 241.0 | | 37.8 | | | |

The total number of trains, the number of main tracks, the total number of switching trains, the percent of heavy vehicles, and the number of traffic lanes have generally, a positive coefficient, while no highway near the rail-highway grade crossing has a negative coefficient. All the models have a negative intercept indicating that the number of crashes per year would be very low (almost zero).

The negative coefficient for the number of main tracks in the case of track class 5 could be attributed to the warning devices and signals implemented at such rail-highway grade crossings.

6.3. Model Validation

The model validation was performed using data set aside for each track class. The computed WBAPS collision per year was converted to a five-year scale by multiplying the value with 5 (assuming conditions remain constant over the five-year period) to assist with the comparison. The T-test was performed by comparing the difference between the predictions from the developed model and the observed number of crashes with the difference between predictions from WBAPS and the observed number of crashes.

Table 6 shows the mean, standard deviation, significance value, and the absolute value of T-statistic comparing the model output from this research and WBAPS output for each track class. The mean difference for the developed model is lower than WBAPS for track class 1 and track class 2 (also shown in **Figure 1**), while the standard deviation is lower than WBAPS for track classes 1, 2 and 3. The significance value is less than 0.05 and the absolute value of T-statistic is greater than 2.0 for track class models 1, 2, 4, and 5. Overall, the results obtained from the model validation indicate that the predictions vary by track class and are comparable or better than those obtained from WBAPS.

Table 6. Comparison of errors.

| Track Class | Parameter | Model Developed | WBAPS | P | T-statistic |
|-------------|-----------|-----------------|-------|--------|-------------|
| 1 | Mean | 0.020 | 0.030 | 0.0003 | 3.7 |
| | St. Dev. | 0.025 | 0.032 | | |
| 2 | Mean | 0.030 | 0.040 | 0.0027 | 3.0 |
| | St. Dev. | 0.027 | 0.048 | | |
| 3 | Mean | 0.080 | 0.070 | 0.1357 | 1.5 |
| | St. Dev. | 0.043 | 0.085 | | |
| 4 | Mean | 0.160 | 0.020 | 0.0001 | 17.2 |
| | St. Dev. | 0.094 | 0.013 | | |
| 5 | Mean | 0.070 | 0.020 | 0.0001 | 6.9 |
| | St. Dev. | 0.038 | 0.013 | | |

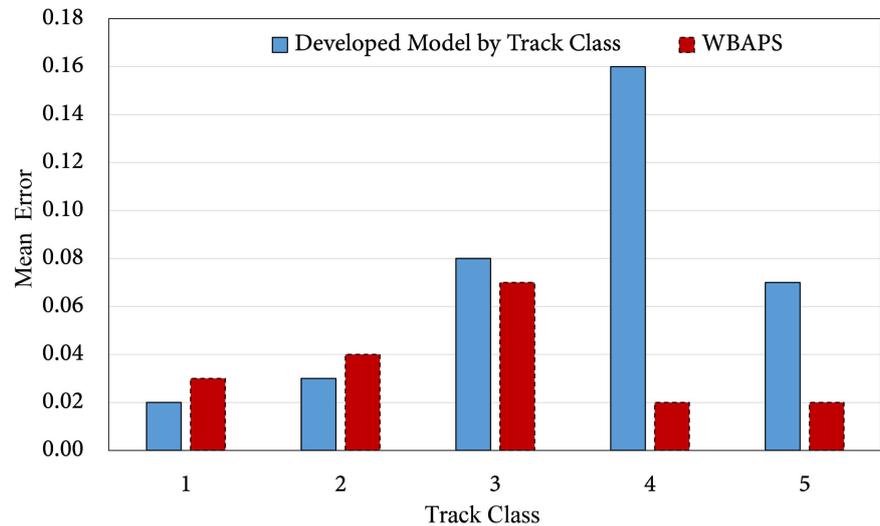


Figure 1. Comparison of mean errors.

7. Conclusions

The NB model for each track class model was found to be the best fitting model to predict the number of crashes at rail-highway grade crossings. The total number of trains, if stop lines are present, the number of traffic lanes, the percentage of trucks, the number of main tracks, the total number of switching trains and no highway near the rail-highway grade crossing are critical explanatory variables to model crash risk by track class at rail-highway grade crossings. The variables in each track class are different from one another, which support the fact that rail-highway grade crossings for each track class must be considered separately when modeling crash risk.

The comparison of WBAPS with the developed model outputs suggests that these models give a more conservative picture of the number of crashes. It also shows that track class is a critical factor related to the risk at a rail-highway grade crossing. The track class governs the number of crashes at rail-highway grade crossings largely and should thus always be considered when addressing rail-highway grade crossing safety problems.

The models suffer from certain limitations as they have been developed using data available which is very scarce in nature. In the models based on track class, there are classes in which only a marginal number of rail-highway grade crossings exist and so a very accurate estimate could not be made.

In the absence of funds or to enhance design standards, the agencies make the decision of closing some rail-highway grade crossings. This leads to an increase in the vehicular traffic and, hence, the risk at the other nearby rail-highway grade crossings. There are also other factors that contribute to crash reduction which could not be accommodated in the models developed in this research and are potential topics for future research. These include driver behavior at rail-highway grade crossings, driving under the influence of alcohol, and rail-highway safety awareness among users.

Acknowledgements

The authors acknowledge FRA's website and staff of the North Carolina Department of Transportation (NCDOT) Rail Division for their help with data required for this research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Angels on Track Foundation (2013) Railroad Crossing Facts. <http://www.angelsontrack.org/cts/ctsfacts.html>
- [2] Federal Railroad Administration (FRA) (2012) Chapter 1:7 Railroad Safety Statistics—Annual Report 2010—Final. FRA, Washington, DC.
- [3] The Urban Institute (1991) The Costs of Highway Crashes. Federal Highway Administration (FHWA) Research Report# FHWA-RD-91-055, Washington, DC.
- [4] Farr, E.H. (1987) Summary of the DOT Rail-Highway Crossing Resource Allocation Procedure (Revised Edition). FRA, United States Department of Transportation (USDOT), Transportation System Center, DOT/FRA/OS-87/05, Cambridge MA.
- [5] McCollister, G.M. and Pflaum, C.C. (2007) A Model to Predict the Probability of Highway Rail Crossing Accidents. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, **221**, 321-329. <https://doi.org/10.1243/09544097JRRT84>
- [6] Mutabazi, M.I. and Berg, W.D. (1995) Evaluation of Accuracy of US DOT Rail-Highway Grade Crossing Accident Prediction Models. *Transportation Research Record: Journal of the Transportation Research Board*, **1495**, 166-170.
- [7] Austin, R.D. and Carson, J.L. (2002) An Alternative Accident Prediction Model for Highway-Rail Interfaces. *Accident Analysis & Prevention Journal*, **34**, 31-42. [https://doi.org/10.1016/S0001-4575\(00\)00100-7](https://doi.org/10.1016/S0001-4575(00)00100-7)
- [8] Saccomanno, F.F., Fu, L. and Miranda-Moreno, L.F. (2004) Risk-Based Model for Identifying Highway-Rail Grade Crossing Blackspots. *Transportation Research Record: Journal of the Transportation Research Board*, **1862**, 127-135. <https://doi.org/10.3141/1862-15>
- [9] Oh, J., Washington, S.P. and Nam, D. (2006) Accident Prediction Model for Railway-Highway Interfaces. *Accident Analysis & Prevention Journal*, **38**, 346-356. <https://doi.org/10.1016/j.aap.2005.10.004>
- [10] Nam, D. and Lee, J. (2006) Accident Frequency Model Using Zero Probability Process. *Transportation Research Record: Journal of the Transportation Research Board*, **1973**, 142-148. <https://doi.org/10.1177/0361198106197300117>
- [11] Hu, S.R. and Lee, C.K. (2008) Analysis of Accident Risk at Railroad Grade Crossing. *Proceeding of the Transportation Research Board 87th Annual Meeting Compendium of Papers*, Transportation Research Board of the National Academies, Washington, D.C.
- [12] Lavette, R.A. (1977) Development and Application of a Railroad-Highway Accident-Prediction Equation. Report # HS-022 157.
- [13] Park, Y.-J. and Saccomanno, F.F. (2005) Evaluating Factors Affecting Safety at

Highway-Railway Grade Crossings. *Transportation Research Record: Journal of the Transportation Research Board*, **1918**, 1-9.

<https://doi.org/10.1177/0361198105191800101>

- [14] Park, Y.-J. and Saccomanno, F.F. (2005) Collision Frequency Analysis Using Tree-Based Stratification. *Transportation Research Record: Journal of the Transportation Research Board*, **1908**, 121-129.
<https://doi.org/10.1177/0361198105190800115>
- [15] Saccomanno, F.F. and Lai, X. (2005) A Model for Evaluating Countermeasures at Highway-Railway Grade Crossings. *Transportation Research Record: Journal of the Transportation Research Board*, **1918**, 18-25.
<https://doi.org/10.1177/0361198105191800103>
- [16] Saccomanno, F.F., Park, P.J.Y. and Fu, L. (2007) Estimating Countermeasure Effects for Reducing Collisions at Highway-Railway Grade Crossings. *Accident Analysis & Prevention Journal*, **39**, 406-416. <https://doi.org/10.1016/j.aap.2006.08.016>
- [17] Yan, X., Richards, S. and Su, X. (2010) Using Hierarchical Tree-Based Regression Model to Predict Train-Vehicle Crashes at Passive Highway-Rail Grade Crossings. *Accident Analysis & Prevention Journal*, **42**, 64-74.
<https://doi.org/10.1016/j.aap.2009.07.003>
- [18] Eck, R.W. and Halkias, J.A. (1985) Further Investigation of the Effectiveness of Warning Devices at Rail-Highway Grade Crossings. *Transportation Research Record: Journal of the Transportation Research Board*, **1010**, 94-101.
- [19] SPSS (2008) SPSS 16.0: Command Syntax Reference. SPSS Inc., Chicago, IL.
- [20] Allison, P. (2012) Do We Really Need Zero-Inflated Models. *Statistical Horizons*.
<http://www.statisticalhorizons.com/zero-inflated-models>
- [21] Code of Federal Regulations, Part 236—Rules, Standards, and Instructions Governing the Installation, Inspection, Maintenance, and Repair of Signal And Train Control Systems, Devices, and Appliances.
<http://www.ecfr.gov/cgi-bin/text-idx?c=ecfr&sid=e3336df9eea4aced3db461fccf707192&rgn=div5&view=text&node=49:4.1.1.1.30&idno=49>