IATSS Research xxx (xxxx) xxx

International Association of Traffic and Safety Sciences

Contents lists available at ScienceDirect

IATSS Research

Research Article

Safety performance functions for low-volume rural minor collector two-lane roadways

Subasish Das [a,*], Ioannis Tsapakis [a], Ali Khodadadi [b]

[a] Texas A&M Transportation Institute, Texas A&M University System, 3500 NW Loop 410, San Antonio, TX 78229, United States of America
[b] Zachry Department of Civil Environmental Engineering, Texas AM University, College Station, TX 77843, United States of America

ARTICLE INFO

Article history:
Received 15 October 2020
Received in revised form 17 February 2021
Accepted 23 February 2021
Available online xxxx

Keywords:
Low-volume roads
Rural minor collectors
Safety performance functions
Safety evaluations

ABSTRACT

The Fixing America's Surface Transportation Act (FAST Act) highlights a data-driven method to improve traffic safety on all public paved roads in the U.S. The first edition of the Highway Safety Manual (HSM) is a widely used tool that provides crash predictive models in the form of safety performance functions (SPFs). There are no specific SPFs for low-volume roadways in the HSM. It is important to know that low-volume roadways are the major roadway types in terms of total mileage. This study used 2015–2019 crash data from Texas, incorporating with other relevant geometric and traffic variables, to develop SPFs for a specific low-volume roadway type (rural minor collector two-lane roadways). This study proposed a rules-based SPF developed approach that makes the prediction accuracies higher compared to the full model. The R^2 values range from 0.18 to 0.22 for all data (without splitting) for different injury level models. The prediction accuracies are improved in the decision tree-based models. For different class specific models (based on injury levels), the R^2 values range from 0.25 to 0.41. Three SPF groups are developed based on crash injury types. The SPFs can provide guidance in refining the prediction accuracies of rural minor collectors.

1. Introduction

The Fixing America's Surface Transportation Act (FAST Act) highlights a data-driven method to improve traffic safety on all public paved roads in the U.S. [1]. The first edition of the Highway Safety Manual (HSM) is considered the most popular and widely used tool to predict traffic crashes on different facility types [2]. Safety Performance Function (SPF) is a crash prediction equation, which is developed to estimate or predict the expected average crash frequency per year at a location as a function of geometric and operational variables. Part C of the HSM provides a list of predictive models, in the form of SPFs, that use different traffic and geometric data inputs. The HSM SPFs focus on four major facility types: 1) rural two-lane two-way roadways; 2) rural multilane highway; 3) urban and suburban highways; and 4) rural and urban interstates and freeways. For each of these facility types, analysts can use SPFs to estimate the total crash frequencies (or crash frequencies based on different injury levels) under certain base conditions.

Low-volume roads (LVRs) are defined as roadways with daily traffic volumes less than the conventional values of other higher functional roadways. Three roadway functional classes are generally considered LVRs: 1) rural local (7R), 2) urban local (7U), and 3) rural minor collector (6R). Due to budgetary constraints, data collection on LVRs is usually sporadic. Additionally, stations with count data are not based on year-long traffic counts. In many cases, short-term traffic counts are usually converted to annual average daily traffic (AADT). This study acquired traffic count, roadway network, and crash data on rural minor collectors to develop SPFs with an aim to improve the accuracy of crash predictions. This study proposed rules-based methods for SPF development to improve prediction accuracies.

The rest of the paper is organized as follows. In the next section, a brief overview of relevant studies is provided. The next section is methodology, which is comprised of three sub-sections: 1) concepts of predictive modeling, 2) rules-based modeling, and 3) data collection and analysis. The following section is 'results and discussions.' The last section, 'conclusions,' provides a broad overview of this study, general findings, limitations, and future needs.

2. Literature review

The improvement of roadway safety is a top priority in transportation safety planning. Historical crash data analytics is the most popular approach in conducting transportation safety research. The general

* Corresponding author.
E-mail addresses: s-das@tti.tamu.edu (S. Das), i-tsapakis@tti.tamu.edu (I. Tsapakis), a.khodadadi1994@tamu.edu (A. Khodadadi).
Peer review under responsibility of International Association of Traffic and Safety Sciences.

https://doi.org/10.1016/j.iatssr.2021.02.004

approach is to develop crash prediction models using the historical crash data and other co-variates such as roadway, traffic, and environmental variables. A crash prediction model provides a simple interpretation of the complex structure of crash data. The overall crash data analysis is based on two basic categories: frequency analysis (crash counts), and classification analysis (crash injury types). Three extensive survey papers provide a grand picture of the latest crash modeling techniques [3–5]. Interested readers can consult these papers for additional insights on crash data modeling approaches.

Many SPFs and related crash modification factors (CMFs) have been developed for roadways with moderate to high traffic volume. However, limited studies have developed SPFs for LVRs, mostly due to the lack of traffic volume data on these roadways. As traffic volume is one of the key components of SPFs, the absence of this measure will make the models biased and less powerful in producing accurate estimates. This section provides a brief overview on LVR related safety studies.

Zegeer et al. [6] performed a comprehensive safety analysis on LVRs by collecting around 4100 mi of rural two-lane roadway segments (AADT ≤2000 vehicles per day or vpd) from seven states. To validate the SPFs, this study used independent databases of 54,000 mi roadways from three states (Minnesota, Illinois, and North Carolina). The results show that shoulder type (paved or unpaved) was not statistically significant. Studies on unpaved roadway safety is limited in number. Caldwell and Wilson [7] compared crash injury rates on unpaved county road segments to crash injury rates on all roads in Wyoming. Crash occurrence on county roads were found to be five times more likely than crashes on all roads. This study provided a systematic way of segment prioritization for safety improvement. Using crash data from Kentucky and North Carolina for several years (1993–1995), Stamatiadis et al. [8] identified several driver-related contributing factors on LVRs (AADT ≤ 1000 vpd). Considering data from rural areas, Achwan and Rudjito [9] depicted the extent of traffic characteristics on LVRs. The study demonstrated that the key vulnerable groups were motorcyclists, pedestrians, and truck associated causalities. In his study, Madsen [10] determined that 60% of those killed and 75% of the injured persons in LVR crashes were occupants of non-farm vehicles. Some of the key contributing factors are the lack of retroreflective signs and taillights on slow-moving vehicles. Using data from gravel roads in Kansas, Liu and Dissanayake [11] surveyed county agencies and performed an analysis by incorporating speed. The findings demonstrate that traffic speeds are not significantly affected by the segment width, surface type, and percentage of large vehicles in traffic, but not with posted speed limit. In a follow-up study, Liu and Dissanayake [12] developed logistic regression models to find factors associated with injury crashes on gravel roads in Kansas. The findings indicate that failure to use safety equipment, intoxication, speeding, distraction, aging drivers, failure to yield, and ruts/potholes on surfaces increased the likelihood of more severe crash occurrence. The NCHRP synthesis report provided a handful of effective safety tools and strategies to improve safety on LVRs [13]. A report, published in 2006, showed the effectiveness of several key countermeasures that are appropriate for LVRs [14]. Al-Kaisy et al. [15] developed a prioritization scheme as a way to produce crash risk index that can be used in ranking candidate sites for safety improvements on low-volume roads in Oregon. To predict total crashes, as well as fatal and injury crashes, Das et al. [16] used 2013–2017 traffic count along with roadway network and crash data from North Carolina to develop six SPFs for three LVRs (rural local, urban local, and rural minor collector roadways). Additionally, a sensitivity analysis was performed to show how traffic volumes influence expected crash frequencies. Farid et al. [17] conducted a crash severity analysis of motorcycle crashes on low-volume roadways. If all other conditions are the same, speed and impairment showed high likelihood of motorcycle crashes with severe injuries. Stapleton et al. [18] developed several SPFs for rural two-lane county roadway segments in Michigan. Three separate models were developed: 1) paved non-federal-aid segments (NFAS), 2) paved federal-aid segments, and 3) low-volume NFAS (both paved and unpaved). The results show that county NFAS paved roadways showed almost similar crash rates to the HSM-base rural two-lane roadway model.

The literature review indicates that a handful of studies performed safety analysis focusing on LVRs. One of the common finding indicates that many important and influential variables are missing on low-volume roadways. For example, traffic volume count is mostly missing on low-volume roadways. Consideration of traffic volume is a critical component for SPF development. Most the studies focused on developing SPFs with limited number of sites with available variables. It is found that none of the studies focused on the improvement of the SPF prediction accuracy. The current study aims to mitigate this research gap by applying an innovative rules-based SPF development approach. The key objective of this study is develop precision based SPFs for Texas rural minor collector two-lane roadways.

3. Methodology

3.1. Concepts of predictive modeling

Most of the traffic safety analysis studies use the empirical Bayes (EB) estimation method to evaluate expected yearly crashes (or other temporal durations) before and after site improvement with countermeasure or a group of countermeasures. The EB method uses both predicted number of crashes (from the SPFs) and observed number of crashes to provide the measures of expected number of crashes.

An SPF is an equation that can predict the mean crash frequencies for a temporal duration at a location as a function of geometric properties of the roadway segment or intersection and traffic volume measures. There are two major uses of the SPF: (1) to develop a localized SPF for the facility and certain crash types for a temporal duration, or (2) to calibrate the SPFs in the first version of the HSM. A baseline SPF can be developed using segment length and annual average daily traffic (AADT):

$$C_{Predicted} = \exp\left[\beta_0 + \beta_1 \times \ln(L) + \beta_2 \times \ln(AADT)\right] \quad (1)$$

where:

$C_{Predicted}$ = predicted crash count

$\beta_0, \beta_1, \beta_2$ = coefficients

L = segment length

$AADT$ = annual average daily traffic

One can predict crashes (N), by multiplying three components: baseline SPF ($C_{Predicted}$), a series of CMFs and a calibration factor, C:

$$N = C_{Predicted} \times \prod CMF \times C \quad (2)$$

It can be noted that the EB approach is based on a weighted average concept. Many studies used this method to develop localized SPFs for different facility and crash types [19–31].

3.2. Rules based modeling

New modeling approaches are needed to tackle the complexities of crash data and the associated improvement of estimation accuracy. Conventional SPFs generally examine the mean effects of key contributing factors and ignore subgroups with different factors. Due to the generalization, this approach fails to inquire the specific subgroup effects within the population of roadway segments or intersections. Rule-based modeling is one of several emerging approaches that avoid these consequences. These methods can identify subgroups effects without imposing any prior assumption or group of assumptions [32]. The rules provide a subset of SPFs that represent subsets of roadway segments or intersections by not only considering the interactions between the contributing factors but also their ranges. Recursive partitioning is one of the simplest rules-based modeling techniques. This procedure led to the models with different boxes (see Fig. 4) and the splitting variables and their ranges. This study used two open source R (rpart, rattle) packages to perform the decision tree modeling [33,34].

3.3. Data description

This study assembled a comprehensive database of roadway, traffic volume, and crash data for rural minor collectors in Texas. Two major Texas specific data sources (Crash Record Information Systems [CRIS] and Road-Highway Inventory Network Offloads [RHiNO]) were used for the five-year period 2015–2019. SPF development requires a comprehensive crash database with geocode crash location information or relevant route information, roadway type, injury type, collision type, and other relevant information. The injury classification system (known as KABCO) divides crash severity into five major groups: 1) fatal injury (K), 2) incapacitating or severe injury (A), 3) non-incapacitating or moderate injury (B), 4) minor or complaint injury (C), and 5) no injury or property damage only (PDO) or O. It is important to note that the new second edition of its low-volume road guidance includes design advice for local and minor collector roads carrying AADT volumes of 2000 vpd or less, by replacing the previous threshold of 400 vpd or less [35]. This study used a threshold of 4000 vpd or less for the AADT measures. However, around 98% of the used roadways have AADT 2000 vpd or less. This study kept 4000 vpd as the threshold because throwing out 2% of these sites are involved with 10% of all crashes. As the definition of low-volume roadways has been changing over the years (previously 400 vpd roadways were defined as low-volume roadways, which has been increased to 2000 vod in the recent years), it is important to examine impact of a larger threshold of AADT range for the low-volume roadways.

To develop the crash database, the following steps are taken (see Fig. 1). Software usage for each step in shown in parenthesis.

- Step 1: Collect five (2015–2019) years of crash data. Use filter 'INTER_RELAT_ID' to remove intersection and intersection related crashes. Add several filters (AADT < 4001 vpd, number of lanes = 2, segment length > 0.099 mi.) to prepare the database of interest. (R)
- Step 2: Geolocate crashes and create a geodatabase or a shapefile. (ArcGIS)
- Step 3: Create a geodatabase for rural minor collector two-lane roadways from the comprehensive roadway inventory data. (ArcGIS)
- Step 4: By using the 'Near' function, assign crash locations to roads. (ArcGIS)
- Step 5: Match roadway names in both crash data and roadway inventory data and remove the erroneous entries. (R)
- Step 6: Assign number of crashes based on injury type and year to develop the final dataset. (R)

The primary dataset contains several variables. This study selected variables such as type and width of shoulders and medians for the preliminary analysis. After performing a correlation analysis and determined variable importance measures (using open sour R software 'vip' package [36]), only two variables (segment length and AADT) were found as the best explanatory variables. Table A (see the Appendix) lists descriptive statistics of key measures (including measures such as mean, maximum, minimum, inter quantile range, kurtosis, and skewness) based on the classes determined from the decision tree rules modeling. Fig. 2(a–d) shows the mean measures of the key variables by different datasets based on crash injury types. The bar plots show that thresholds of different variables widely vary from one class to another class.

Fig. 3 shows the correlation plot for 6R roadways for KABCO model. The correlation measures indicate that only segment length and AADT are statistically significant factors. The other correlation analysis for other models show also same finding.

To determine the optimum number of classes (or clusters), different optimized complexity parameters have been used by performing validation efforts. To explain the decision tree procedure, a dataset for 6R roadways for KABCO model has been selected. Using count of KABCO crashes as the response variable, the decision rules algorithm (classification and regression tree or CART) was applied to the dataset determine the appropriate number of clusters. A threshold of three branches have been used to limit the number of clusters. Fig. 4 (a) provides annotation of the measures shown in the decision tree developed for KABCO models. All of the decision trees used in this study have been validated by splitting data into training and test data. For example, the decision rules generated for KABCO crashes on rural minor collector two-lane roadways are:

- Class 1 rule: LEN_SEC (Segment Length) < 1.3 & ADT_CUR (AADT Current or the latest year AADT) < 612 (5963 segments with mean crash frequency = 0.28 crash/year per segment) [Note: the variable codes are described in the notes of Table A]
- Class 2 rule: LEN_SEC < 1.3 & ADT_CUR > 611 (1458 segments with mean crash frequency = 1 crash /year per segment)
- Class 3 rule: LEN_SEC > 1.2 & ADT_CUR < 331 (2622 segments with

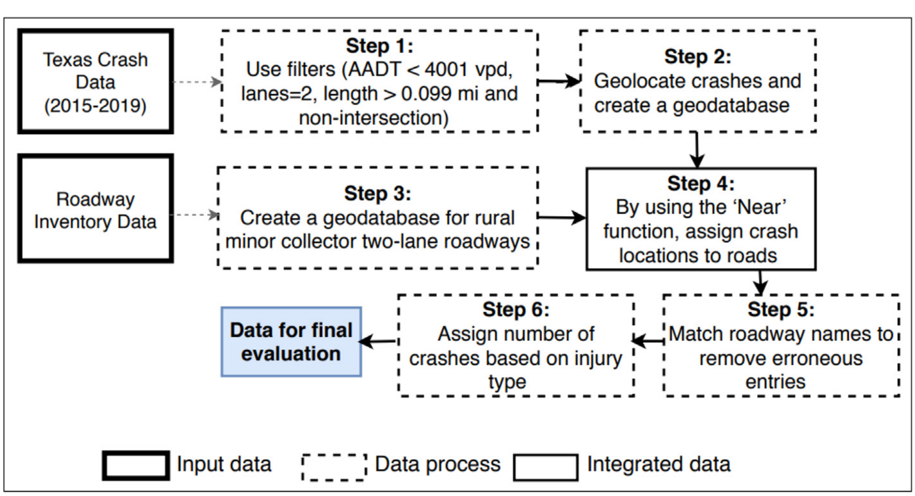

Fig. 1. Flowchart of database development.

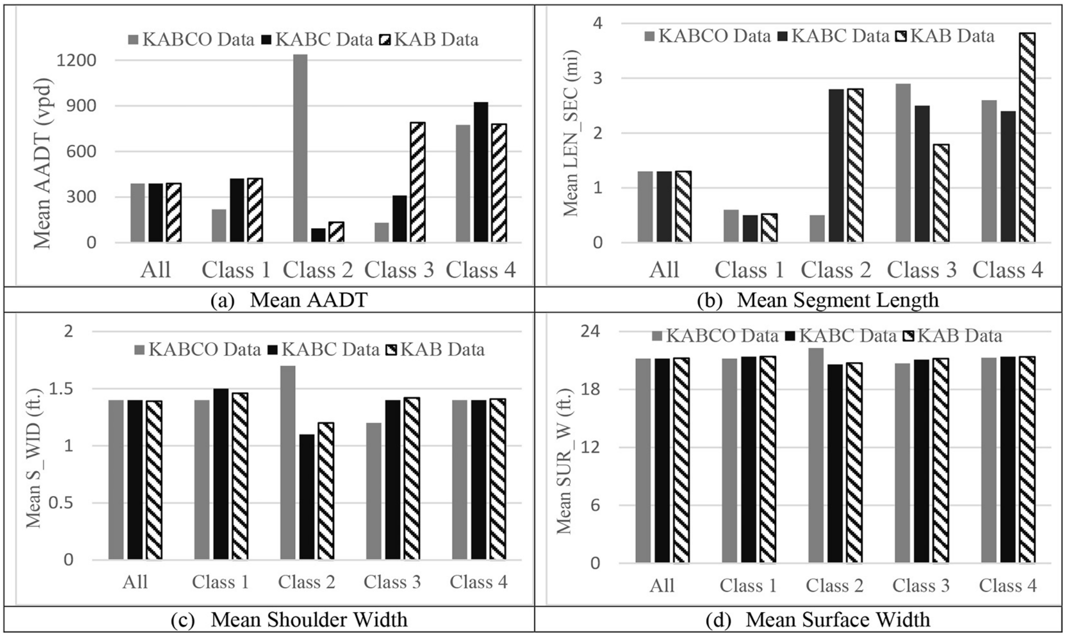

Fig. 2. Mean measures of key variables.

mean crash frequency = 0.8 crash/year per segment)
- Class 4 rule: LEN_SEC > 1.2 & ADT_CUR > 330 (1169 segments with mean crash frequency = 3.3 crashes/year per segment)

A parallel coordinate plot was used to understand sub-group effects in the data. Fig. 5 shows a parallel coordinate plot for four key variables used in the KABCO modeling framework. The colors of the links are based on the class category. Compared to other classes, class 2 shows higher AADT values (many light green links are associated with the high value of AADT). Similarly, class 3 and class 4 shows higher segment length values based on the visualization of the links. It is obvious that class 3 and class 4 are associated with rules of segment length greater than 1.2 miles.

4. Results and discussions

4.1. Rule-based SPFs

For Texas rural minor collector two-lane roadways (shown as 6R in Table 1), there are 11,212 segments with available AADT data. Table 1

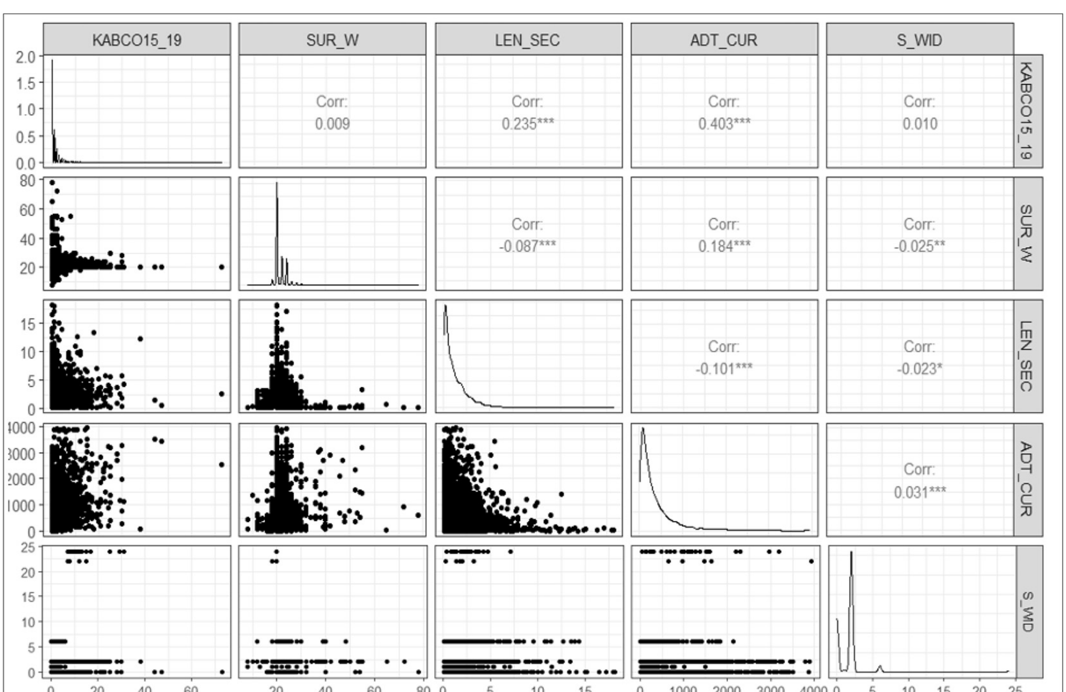

Fig. 3. Correlation analysis plot.

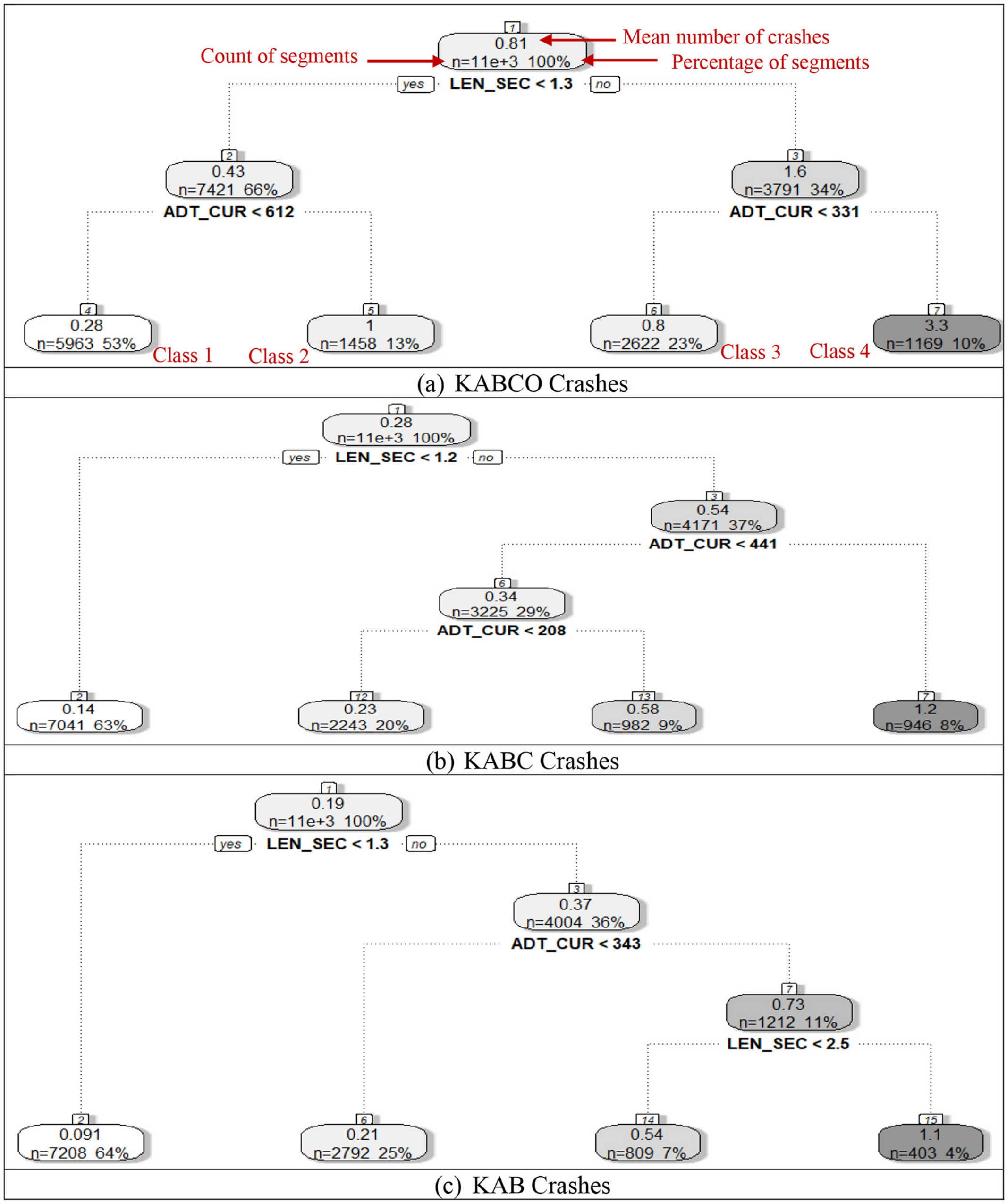

Fig. 4. Decision trees for three datasets based on crash counts by severity groups.

lists the developed SPFs for all data and subset data (based on decision-rule criteria) for there injury levels: 1) KABCO, 2) KABC, and 3) KAB. The table shows the model equations along with the overdispersion parameter (b) and the loglikelihood of each model. As these measures are based on different clusters of datasets, explanations of these measures are not needed to be detailed. Four decision trees have been generated by exploring different complexity parameter thresholds. For example, 6R KABCO dataset shows five models: 1) All ($N_{6R,\ tot,\ all}$) represents data before performing decision tree, 2) Case 1 to Case 4 ($N_{6R,\ tot,\ class1}$, $N_{6R,\ tot,\ class2}$, $N_{6R,\ tot,\ class3}$, $N_{6R,\ tot,\ class4}$): four classes of data based on the decision tree filters. This study developed negative binomial (NB) models for each of these five cases. Similar actions are taken for other injury levels (KABC and KAB). As this study contains 15 SPFs, validation of 15 SPFs will be extensive. KABCO Class 1 data has been divided into train (75%) data and validation (25%) data. The validation model results are similar to the train model.

As mentioned earlier, regression models examine the mean effects of the explanatory variables and ignore subclass effect in the overall population of all roadway segments. This study performed decision tree algorithms to determine the subclass effect in the dataset. As the current model is completely based on the rural minor collector two-lane roadways in Texas, transferability of these models to other states should be carefully considered.

4.2. Model validation

The R^2 values range from 0.18 to 0.22 for all data (without splitting) for different injury level models. The prediction accuracies are improved in the decision tree-based models. For different class specific models (based on injury levels), the R^2 values range from 0.25 to 0.41. The adjusted R^2 values are same with R^2 values for the developed models. To understand the goodness-of-fit, another

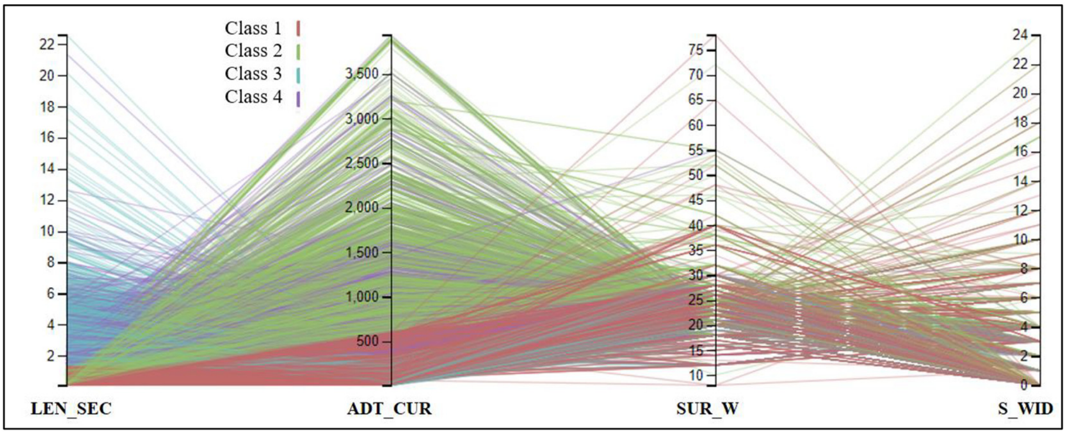

Fig. 5. Parallel coordinate plots for four key variables (for KABCO severity group data).

quick diagnostic is the development of Cumulative Residual (CURE) plots. Residuals indicate the disparities between historical crash frequencies and predicted crash frequencies. Modeling fitting can be performed by examining the residuals. If the surrounding residuals of a model are close to zero, the model can be considered as a good-fit model. CURE plot is a good visualization tool to examine the SPF predictions based on the individual explanatory variables used in the model. A horizontal stretch of the CURE plot infers to a region of the variable where the estimates are unbiased [37]. On the contrary, in locations where the CURE plot drifts up or down significantly, the estimates are not considered to be unbiased. The CURE plot for an unbiased SPF must be within the boundaries of two standard deviations [37].

Fig. 6, Fig. 7, and Fig. 8 show the CURE plots based on the modeling outcomes of three SPF groups based on the injury type: 1) KABCO, 2) KABC, and 3) KBC. The left side of each of these three figures show CURE plots based on segment length. The right side of each of these three figures show CURE plots based on AADT. There are five CURE plots in each side of Fig. 6–8. Investigation on the projections of the residuals of each plot will show the improved performance measures of the class-based models. The CURE plots of the main models and rules based models clearly indicate that rules based model are mostly inside the confidence boundary. For example, in Fig. 6(a), CURE plots for main KABCO model and rules-based KABCO models for 6R roadways are shown. The residual plot (in red) in outside of the confidence boundary in two zones. However, the residual plots for the rule-based models (except Class 4 shows that the residual plot is outside the bound when segment length is above 12.5 mile) are mostly within the confidence boundary. The CURE plots for other models show similar trends. The example clearly shows the effectiveness of rules-based models in generating more robust crash estimates.

Table 1
Developed SPFs.

Roadway and severity group	Class	Safety performance functions	Over-dispersion parameter	Log-likelihood
6R (KABCO)	All	Rule: All Data		
		$N_{6R,\ tot,\ all} = exp\ (-4.759) \times Length^{0.900} \times AADT^{0.766}$	1.3075	−22,215.077
	Class 1	Rule: LEN_SEC < 1.3 & ADT_CUR < 612		
		$N_{6R,\ tot,\ class1} = exp\ (-4.170) \times Length^{0.898} \times AADT^{0.658}$	0.8020	−7317.370
	Class 2	Rule: LEN_SEC < 1.3 & ADT_CUR > 611		
		$N_{6R,\ tot,\ class2} = exp\ (-4.298) \times Length^{0.958} \times AADT^{0.699}$	1.3710	−3676.017
	Class 3	Rule: LEN_SEC > 1.2 & ADT_CUR < 331		
		$N_{6R,\ tot,\ class3} = exp\ (-4.627) \times Length^{0.764} \times AADT^{0.756}$	0.9225	−6036.538
	Class 4	Rule: LEN_SEC > 1.2 & ADT_CUR > 330		
		$N_{6R,\ tot,\ class4} = exp\ (-4.606) \times Length^{0.832} \times AADT^{0.763}$	2.129	−5071.186
6R (KABC)	All	Rule: All Data		
		$N_{6R,\ kabc,\ all} = exp\ (-5.636) \times Length^{0.940} \times AADT^{0.736}$	1.2247	−12,386.275
	Class 1	Rule: LEN_SEC < 1.2		
		$N_{6R,\ kabc,\ class1} = exp\ (-5.335) \times Length^{0.953} \times AADT^{0.686}$	1.035	−5141.116
	Class 2	Rule: LEN_SEC > 1.1 & ADT_CUR < 208		
		$N_{6R,\ kabc,\ class2} = exp\ (-5.323) \times Length^{0.649937} \times AADT^{0.649}$	0.4449	−2459.781
	Class 3	Rule: LEN_SEC > 1.1 & 207 < ADT_CUR < 441		
		$N_{6R,\ kabc,\ class3} = exp\ (-4.043) \times Length^{1.023} \times AADT^{0.446}$	1.191	−1954.343
	Class 4	Rule: LEN_SEC > 1.1 & ADT_CUR > 440		
		$N_{6R,\ kabc,\ class4} = exp\ (-4.707) \times Length^{0.688} \times AADT^{0.639}$	2.518	−2738.366
6R (KAB)	All	Rule: All Data		
		$N_{6R,\ kab,\ all} = exp\ (-5.949) \times Length^{0.960} \times AADT^{0.719}$	1.259	−9630.336
	Class 1	Rule: LEN_SEC < 1.3		
		$N_{6R,\ kab,\ class1} = exp\ (-5.846) \times Length^{0.945} \times AADT^{0.699}$	0.961	−4002.588
	Class 2	Rule: LEN_SEC > 1.2 & ADT_CUR < 343		
		$N_{6R,\ kab,\ class2} = exp\ (-6.169) \times Length^{0.963} \times AADT^{0.756}$	0.654	−2893.847
	Class 3	Rule: 1.2 < LEN_SEC < 2.5 & ADT_CUR > 342		
		$N_{6R,\ kab,\ class3} = exp\ (-5.401) \times Length^{0.858} \times AADT^{0.649}$	3.420	−1552.351
	Class 4	Rule: LEN_SEC > 2.5 & ADT_CUR > 342		
		$N_{6R,\ kab,\ class4} = exp\ (-3.999) \times Length^{0.619} \times AADT^{0.501}$	1.935	−1137.224

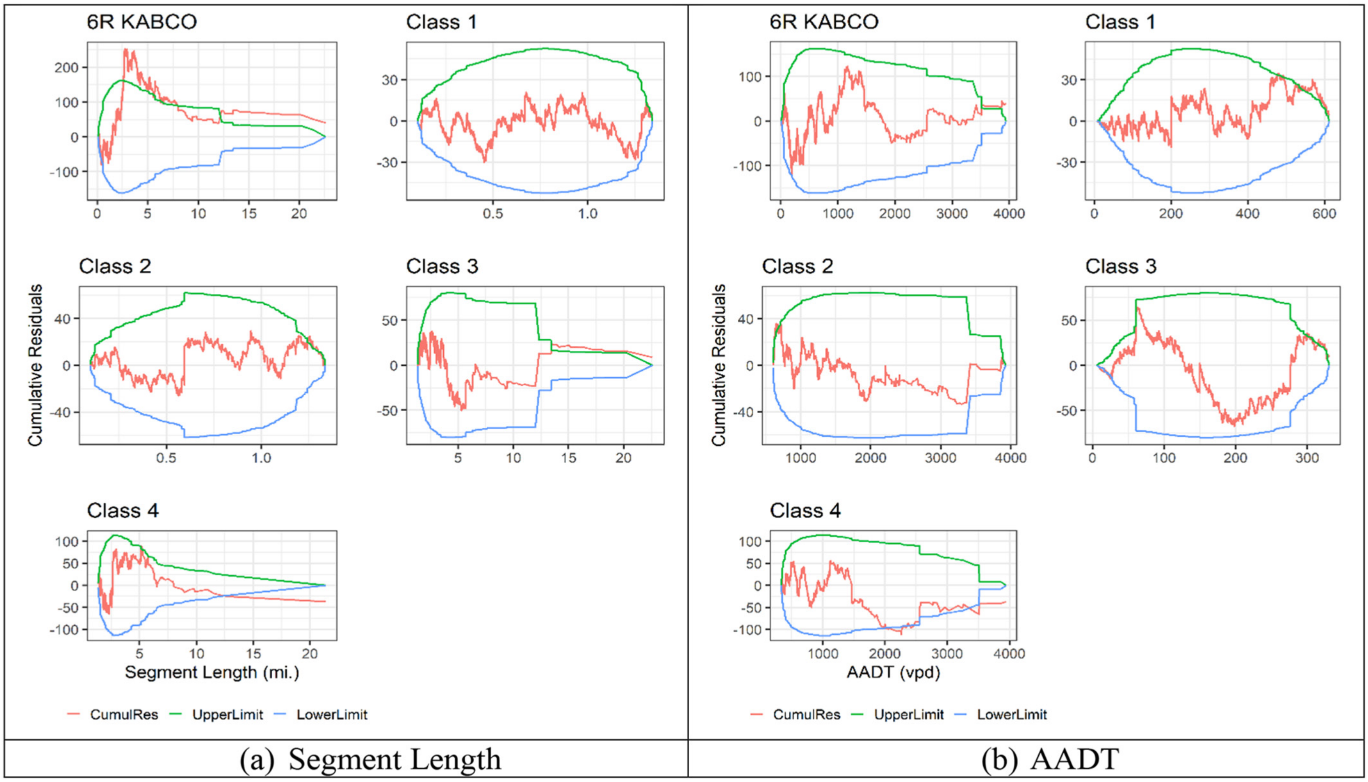

Fig. 6. CURE plots for KABCO models.

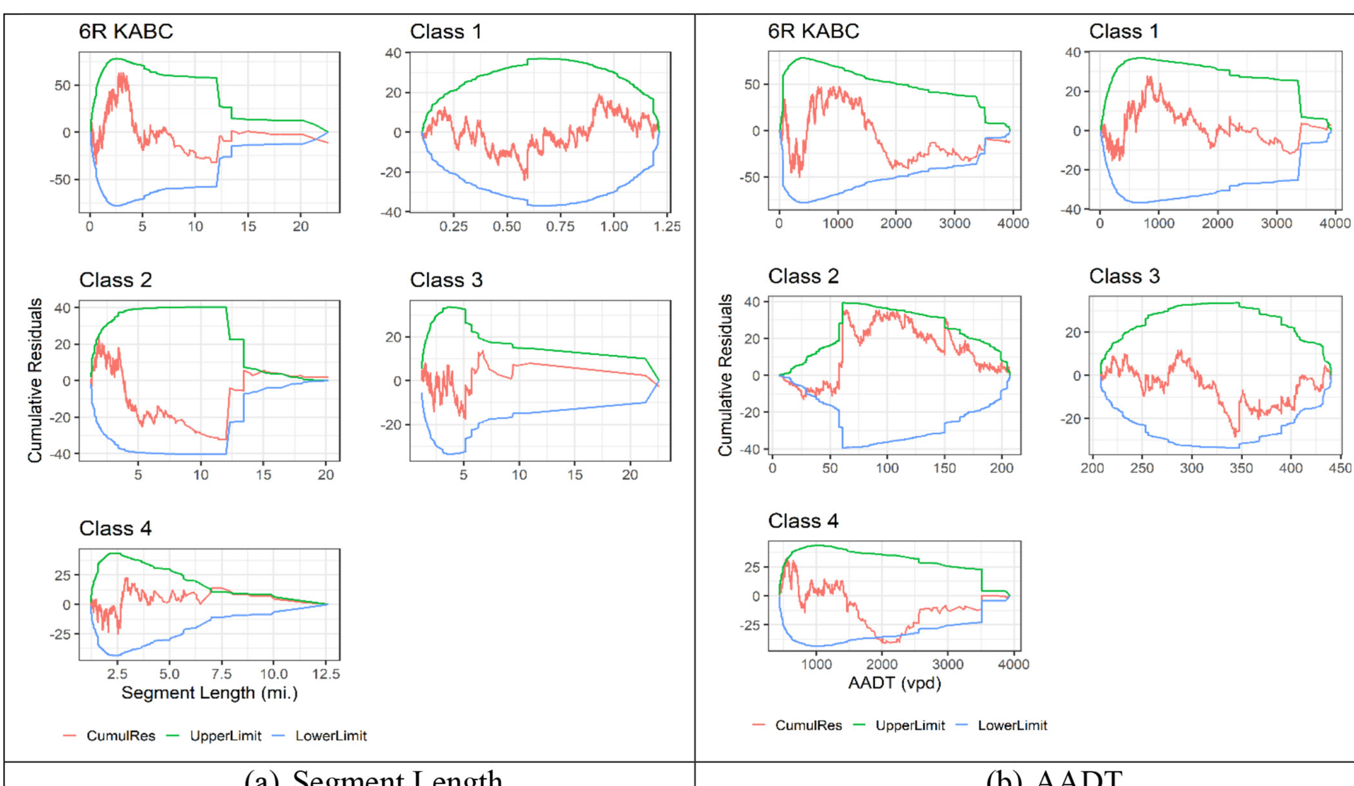

Fig. 7. CURE plots for KABC models.

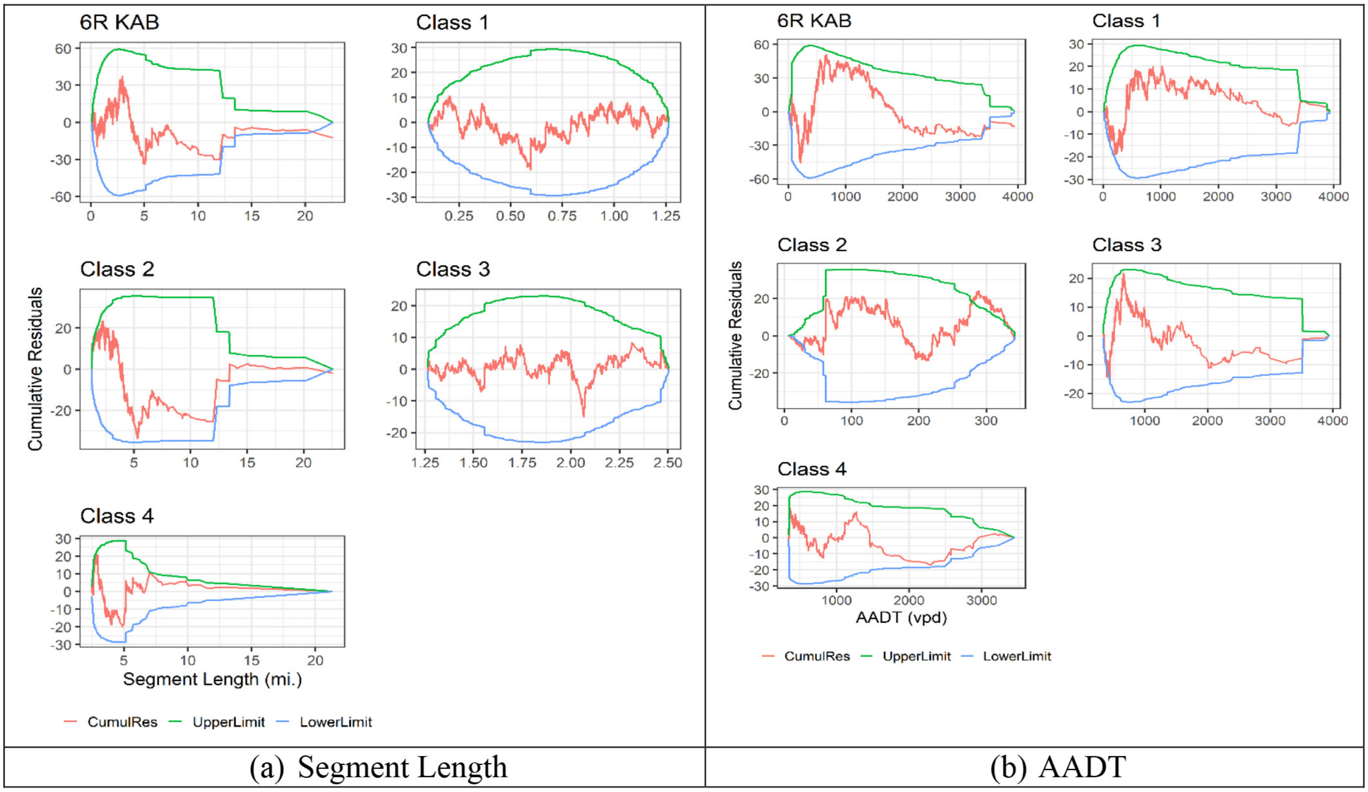

Fig. 8. CURE plots for KAB models.

5. Conclusions

The safety studies on the LVRs are limited in number. Additionally, no prior LVR SPF modeling approached considered the improvement of prediction accuracies. This study acquired a comprehensive database of rural minor collector two-lane roadways in Texas. This study has some unique contributions. First, this study developed a procedure of using local roadway network data in estimating crash frequencies. This study developed a reproducible database preparation framework, which can be replicated to other states. Correlation analysis and variable importance were performed to determine the key contributing factors. Segment length and AADT are found as the key factors for the model development. Second, it developed SPFs for rural minor collector two-lane roadways in Texas for three different injury levels (KABCO, KABC, and KAB). The models are validated by using the interpretations of the developed CURE plots. The goodness-of-fit measures showed that the decision rule-based subclass models perform better than full models.

This study is not without limitations. First, the current analysis is limited to only one LVR facility type with 4000 vpd as the maximum traffic volume. Second, the current study has not performed sensitivity analysis. As the current study is more focused towards the precision accuracy of the SPFs, sensitivity analysis for all models (15 models) will be comprehensive and can be considered as a scope for future studies. Third, the SPFs are localized as the models are based on the available data on rural minor collector two-lane roadways in Texas. Transferability of these models requires careful attention in terms of the thresholds of the explanatory variables and sample size. Fourth, the goal of this paper is to show the performance improvement in terms of R^2 and CURE plots. Omission of other variables may be the reason. However, road inventory data of low-volume roadways are not well maintained. Many important geometric variables such as lane width., shoulder width are not readily available. Future studies can perform an in-depth investigation by exploring different performance measures (for example, silhouette value) of decision tree outcomes.

Declaration of competing interest

None.

Acknowledgements

The work has been done as part of SAFE-D University Transportation Center (UTC) project Use of Disruptive Technologies to Support Safety Analysis and Meet New Federal Requirements (Grant Number: 69A3551747115).

Appendix A

Table A
Descriptive statistics.

Attribute	Group	Count	Min	Max	Mean	Median	SD	IQR	Skewness	Kurtosis
Dataset based on counts of KABCO crashes										
KABCO15_19[1]	All	11,212	0	41	0.8	0.0	1.9	1.0	6.5	77.5
	Class 1	5963	0	12	0.3	0.0	0.7	0.0	4.5	35.5
	Class 2	1458	0	36	1.0	0.0	1.8	1.0	6.8	104.4
	Class 3	2622	0	38	0.8	0.0	1.7	1.0	8.3	139.4
	Class 4	1169	0	41	3.3	2.0	3.8	3.0	3.1	16.6
LEN_SEC[2]	All	11,212	0.1	22.6	1.3	0.8	1.5	1.5	3.5	24.5
	Class 1	5963	0.1	1.34	0.6	0.4	0.4	0.6	0.6	−0.9
	Class 2	1458	0.1	1.34	0.5	0.4	0.4	0.6	0.7	−0.7
	Class 3	2622	1.35	22.6	2.9	2.4	1.8	1.5	3.7	22.4
	Class 4	1169	1.35	21.38	2.6	2.1	1.4	1.2	4.2	34.5
ADT_CUR[3]	All	11,212	6	3940	389.2	218.5	493.2	372.0	3.0	11.8
	Class 1	5963	6	611	219.3	185.0	154.3	229.0	0.7	−0.5
	Class 2	1458	612	3922	1238.8	990.0	668.8	725.3	1.7	2.7
	Class 3	2622	6	330	131.3	113.0	86.7	133.0	0.6	−0.7
	Class 4	1169	331	3940	775.2	589.0	539.8	460.0	2.5	7.1
SUR_W[4]	All	11,212	8	78	21.2	20.0	2.8	2.0	4.8	56.7
	Class 1	5963	8	78	21.2	20.0	2.8	2.0	4.5	57.8
	Class 2	1458	10	72	22.3	22.0	4.1	4.0	4.5	32.8
	Class 3	2622	12	30	20.7	20.0	1.8	1.8	1.1	4.6
	Class 4	1169	12	55	21.3	20.0	2.2	2.0	3.6	47.4
S_WID[5]	All	11,212	0	24	1.4	0.0	2.0	3.0	2.2	10.3
	Class 1	5963	0	22	1.4	0.0	2.0	3.0	2.0	8.7
	Class 2	1458	0	24	1.7	0.0	2.7	3.0	2.7	12.1
	Class 3	2622	0	8	1.2	0.0	1.7	2.0	1.2	0.1
	Class 4	1169	0	10	1.4	1.0	1.7	3.0	1.2	1.2
Dataset based on Counts of KABC Crashes										
KABC15_19[6]	All	11,212	0	29	0.3	0.0	0.8	0.0	8.4	169.9
	Class 1	7041	0	14	0.1	0.0	0.5	0.0	7.6	139.3
	Class 2	2243	0	29	0.2	0.0	0.9	0.0	18.3	544.6
	Class 3	982	0	15	0.6	0.0	1.1	1.0	4.5	37.9
	Class 4	946	0	14	1.2	1.0	1.5	2.0	2.3	9.3
LEN_SEC	All	11,212	0.1	22.6	1.3	0.8	1.5	1.5	3.5	24.5
	Class 1	7041	0.1	1.21	0.5	0.4	0.3	0.5	0.6	−0.9
	Class 2	2243	1.21	20.18	2.8	2.3	1.8	1.6	3.4	18.1
	Class 3	982	1.21	22.6	2.5	2.1	1.6	1.4	5.1	50.5
	Class 4	946	1.21	12.62	2.4	2.0	1.3	1.1	3.0	13.6
ADT_CUR	All	11,212	6	3940	389.2	218.5	493.2	372.0	3.0	11.8
	Class 1	7041	6	3922	421.9	237.0	523.2	402.0	2.9	10.4
	Class 2	2243	6	207	94.7	89.0	53.8	85.5	0.3	−1.0
	Class 3	982	208	440	310.7	303.0	67.3	114.0	0.3	−1.1
	Class 4	946	441	3940	925.4	716.5	571.6	510.8	2.2	5.7
SUR_W	All	11,212	8	78	21.2	20.0	2.8	2.0	4.8	56.7
	Class 1	7041	8	78	21.4	20.0	3.1	2.0	4.8	50.9
	Class 2	2243	12	32	20.6	20.0	1.7	0.0	1.4	5.2
	Class 3	982	12	30	21.1	20.0	2.1	2.0	0.3	3.6
	Class 4	946	12	55	21.4	20.0	2.3	2.0	4.1	52.7
S_WID	All	11,212	0	24	1.4	0.0	2.0	3.0	2.2	10.3
	Class 1	7041	0	24	1.5	0.0	2.1	3.0	2.4	11.8
	Class 2	2243	0	7	1.1	0.0	1.6	2.0	1.2	0.2
	Class 3	982	0	8	1.4	1.0	1.7	3.0	0.9	−0.3
	Class 4	946	0	10	1.4	1.0	1.8	3.0	1.3	1.6
Dataset based on Counts of KAB Crashes										
KAB15_19[7]	All	11,212	0	21	0.19	0	0.61	0	8.62	176.43
	Class 1	7208	0	10	0.09	0	0.36	0	7.11	107.29
	Class 2	2792	0	21	0.21	0	0.69	0	12.74	320.69
	Class 3	809	0	8	0.54	0	0.84	1	2.29	9.75
	Class 4	403	0	15	1.11	1	1.49	2	3.14	19.49
LEN_SEC	All	11,212	0.1	22.6	1.3	0.82	1.47	1.47	3.5	24.52
	Class 1	7208	0.1	1.26	0.52	0.42	0.33	0.55	0.61	−0.86
	Class 2	2792	1.27	22.6	2.8	2.27	1.78	1.52	3.67	22.55
	Class 3	809	1.27	2.5	1.79	1.75	0.33	0.53	0.29	−0.93
	Class 4	403	2.51	21.38	3.82	3.29	1.73	1.29	4.14	29.28
ADT_CUR	All	11,212	6	3940	389.22	218.5	493.19	372	3.01	11.76
	Class 1	7208	6	3922	421.32	237	522.59	402	2.87	10.46
	Class 2	2792	6	342	134.08	115	88.54	135	0.59	−0.71
	Class 3	809	343	3940	789.25	609	533.33	466	2.44	7.35
	Class 4	403	343	3449	779.54	561	561.88	474	2.34	5.89

(continued on next page)

Table A (*continued*)

Attribute	Group	Count	Min	Max	Mean	Median	SD	IQR	Skewness	Kurtosis
SUR_W	All	11,212	8	78	21.23	20	2.77	2	4.75	56.65
	Class 1	7208	8	78	21.41	20	3.12	2	4.77	50.99
	Class 2	2792	12	30	20.73	20	1.79	2	1.07	4.59
	Class 3	809	12	30	21.19	20	2	2	0.58	2.57
	Class 4	403	12	55	21.38	20	2.57	2	5.77	71.1
S_WID	All	11,212	0	24	1.39	0	1.99	3	2.19	10.27
	Class 1	7208	0	24	1.46	0	2.12	3	2.41	11.76
	Class 2	2792	0	8	1.2	0	1.69	2	1.14	0.04
	Class 3	809	0	10	1.42	1	1.78	3	1.19	1.09
	Class 4	403	0	10	1.41	1	1.66	3	1.25	1.85

[1] KABCO15_19 = total KABCO crashes during 2015–2019.
[2] LEN_SEC = segment length (mi.).
[3] ADT_CUR = Annual Average Daily Traffic.
[4] SUR_W = Surface Width (ft.).
[5] S_WID = shoulder width (ft.).
[6] KABC15_19 = total KABC crashes during 2015–2019.
[7] KAB15_19 = total KABC crashes during 2015–2019, IQR = Inter Quantile Range.

References

[1] Federal Highway Administration (FHWA), Moving Ahead for Progress in the 21st Century (MAP-21), https://www.fhwa.dot.gov/map21/ Accessed July 27, 2018.
[2] AASHTO, Highway Safety Manual, 1st Edition, Washington D.C. 2010.
[3] D. Lord, F. Mannering, The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives, Transp. Res. A Policy Pract. 44 (2010) 291–305.
[4] P. Savolainen, F. Mannering, D. Lord, M. Quddus, The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives, Accid. Anal. Prev. 43 (2011) 1666–1676.
[5] F. Mannering, C. Bhat, Analytic methods in accident research: methodological frontier and future directions, Analytic Methods in Accident Research, 1, 2014, pp. 1–22.
[6] C.V. Zegeer, R. Stewart, F. Council, T.R. Neuman, Accident relationship of roadway width on low-volume roads, Transp. Res. Rec. 1445 (1994) 160–168.
[7] R.C. Caldwell, E.M. Wilson, A safety improvement program for rural unpaved roads, MPC Report No. 97-70, January 1997.
[8] N. Stamatiadis, S. Jones, L. Aultman-Hall, Causal factors for accidents on southeastern low-volume rural roads, Transp. Res. Rec. 1652 (1999) 111–117.
[9] N. Achwan, D. Rudjito, Accident characteristics on low-volume roads in Indonesia, Transp. Res. Rec. 1652 (1999) 103–110.
[10] M. Madsen, Farm equipment crashes on upper midwest roads, Midwest Rural Agricultural Safety and Health Forum (MRASH), 2008.
[11] L. Liu, S. Dissanayake, Speed Limit-Related Issues on Gravel Roads, Kansas State University, August 2007.
[12] L. Liu, S. Dissanayake, Examination of Factors Affecting Crash Severity on Gravel Roads, Transportation Research Board 88th Annual Meeting, Washington DC2008.
[13] Roadway Safety Tools for Local Agencies -A Synthesis of Highway Practice, NCHRP Synthesis, 321, 2003.
[14] Low-Cost Local Roads Safety Solutions, American Traffic Safety Services Association, Fredericksburg, Virginia, 2006.
[15] A. Al-Kaisy, L. Ewan, F. Hossain, Identifying candidate locations for safety improvements on low-volume rural roads: the Oregon experience, Transp. Res. Rec. 2673 (12) (2019).
[16] S. Das, I. Tsapakis, S. Datta, Safety performance functions of low-volume roadways, Transp. Res. Rec. 2673 (12) (2019) 798–810, https://doi.org/10.1177/0361198119853559.
[17] A. Farid, M. Rezapour, K. Ksaibati, Modeling Severities of Motorcycle Crashes on Low-Volume Roadways. Presented at the 12th International Conference on Low-Volume Roads. Transportation Research Board, MT, 2019.
[18] S.Y. Stapleton, A.J. Ingle, M. Chakraborty, T.J. Gates, P.T. Savolainen, Safety performance functions for rural two-lane county road segments, Transp. Res. Rec. 2672 (52) (2018).
[19] X. Sun, S. Das, A Comprehensive Study on Pavement Edge Line Implementation. Report No. FHWA/LA.13/508, 2013.
[20] X. Sun, S. Das, Developing Louisiana Crash Reduction Factors. Report No. FHWA/LA.12/506, 2013.
[21] X. Sun, S. Das, Z. Zhang, F. Wang, C. Leboeuf, Investigating safety impact of edgelines on narrow, rural two-lane highways by empirical bayes method, Transp. Res. Rec. 2433 (2014) 121–128.
[22] S. Das, X. Sun, K. Dixon, M. Rahman, Safety effectiveness of roadway conversion with a two-way left turn lane, J. Traffic Transp. Eng. 5 (4) (2018) 309–317.
[23] Y. Zou, J. Ash, B. Park, D. Lord, L. Wu, Empirical Bayes estimates of finite mixture of negative binomial regression models and its application to highway safety, J. Appl. Stat. 45 (9) (2017).
[24] S. Das, L. White, RuralSpeedSafetyX: Interactive Decision Support Tool to Improve Safety, SoftwareX, 2020.
[25] S. Das, M. Le, M.P. Pratt, C. Morgan, Safety effectiveness of truck lane restrictions: a case study on Texas urban corridors, Int. J. Urban Sci. 24 (2020) 35–49.
[26] S. Geedipally, S. Das, M. Pratt, D. Lord, Determining skid resistance needs on horizontal curves for different levels of precipitation, Transp. Res. Rec. 2674 (9) (2020) 358–370.
[27] I. Tsapakis, S. Sharma, B. Dadashova, S. Geedipally, A. Sanchez, M. Le, L. Cornejo, S. Das, K. Dixon, Evaluation of Highway Safety Improvement Projects and Countermeasures: Technical Report. Publication FHWA/TX-19/0–6961-R1, Texas Department of Transportation, 2019.
[28] S. Das, S.R. Geedipally, K. Fitzpatrick, Inclusion of Speed and Weather Measures in Safety Performance Functions for Rural Roadways, IATSS Research, 2020.
[29] M.P. Pratt, S.R. Geedipally, B. Wilson, S. Das, M. Brewer, D. Lord, Pavement Safety-Based Guidelines for Horizontal Curve Safety. Publication FHWA/TX-18/0-6932-R1, Texas Department of Transportation, 2018.
[30] R. Avelar, S. Geedipally, S. Das, L. Wu, B. Kutela, D. Lord, I. Tsapakis, Evaluation of Roadside Treatments to Mitigate Roadway Departure Crashes: Technical Report. Publication FHWA/TX-20/0–6991-R1, Texas Department of Transportation, 2020.
[31] K. Dixon, K. Fitzpatrick, R. Avelar, S. Das, Analysis of the Shoulder Widening Need on the State Highway System. Publication FHWA/TX-15/0-6840-1, Texas Department of Transportation, 2017.
[32] M. Haghighi, S.B. Johnson, X. Qian, K.F. Lynch, K. Vehik, S. Huang, A comparison of rule-based analysis with regression methods in understanding the risk factors for study withdrawal in a pediatric study, Sci. Rep. 6 (1) (2016) 30828, https://doi.org/10.1038/srep30828.
[33] T. Therneau, B. Atkinson, rpart: Recursive Partitioning and Regression Trees. Package Version 4.1-15, 2019, 2019 https://CRAN.R-project.org/package=rpart Accessed August 1, 2020.
[34] G.J. Williams, Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, https://cran.r-project.org/web/packages/rattle/rattle.pdf.
[35] AASHTO, Guidelines for Geometric Design of Low-Volume Roads, AASHTO, Washington DC, 2019.
[36] B. Greenwell, B. Boehmke, B. Gray, vip: Variable Importance Plots, https://koalaverse.github.io/vip/index.html Accessed on August 1, 2020.
[37] E. Hauer, The Art of Regression Modeling in Road Safety, Springer, 2015.