Safety Performance Functions for Two-Lane Urban Arterial Segments

Bekir Bartin[1], Kaan Ozbay, Chuan Xu
Ozyegin University, Istanbul, Turkey, bekir.bartin@ozyegin.edu.tr
New York University, New York USA, kaan.ozbay@nyu.edu
Southwest Jiaotong University, Sichuan, China, xuchuan@swjtu.edu.cn

1. Introduction

The predictive method provided by the Highway Safety Manual (HSM) is based on regression models named Safety Performance Functions (SPFs) that estimate the predicted average crash frequency N_b for certain base geometric and operational conditions. To account for the differences in geometric design and traffic control features between the specific base conditions and the site conditions, crash modification factors (CMFs) are utilized to adjust N_b, as follows.

$$N_e^i = N_b^i \prod_{\forall k} CMF_k^i \tag{1}$$

Where, at site *i*, N_b^i is the predicted crash frequency for base conditions, N_e^i is the expected crash frequency, and CMF_k^i is calculated for specific geometric or operational feature *k*. When the base condition at site *i* is met for a given feature k, then CMF_k^i is equal to 1. The HSM provides separate SPFs for the segments and intersections of rural two-lane roads, rural multilane highways, urban and suburban arterials and freeway facilities. To make the SPFs better accommodate the local data, two strategies are usually employed. One is calibrating the SPFs provided in the HSM so that the contents of the HSM can be fully leveraged. The other is developing location-specific SPFs.

Since the first edition of the Highway Safety Manual (HSM), published in 2010, there have been numerous studies that either estimated calibration factors to utilize its SPFs and/or developed jurisdiction-specific ones. A review of these studies shows that SPF calibration and development is highly data driven. In the calibration process, a calibration factor C is calculated by using 30 to 50 independent sites with a total of minimum 100 crashes per year, as per the HSM's suggestion. In essence, C is a straightforward ratio of the total observed (N_o) to the total estimated ($N_e = \sum_{\forall i} N_e^i$) number of crashes at the selected sites, yet the difficulty lies in the calculation of N_e. Whether the objective is calibration or development of SPFs, collecting or extracting these data is labor intensive, and therefore it is crucial to automatically acquire as much data as possible from existing sources.

This paper presents the SPF calibration and development process for the undivided two-lane urban and suburban arterial (U2) segments in New Jersey (NJ). U2 segments are defined as a roadway consisting of two lanes with a continuous cross-section providing two directions of travel in which the lanes are not physically separated by either distance or a barrier. Data requirements, the availability of required data, and the data processing and extraction methods are presented, along with detailed results of the calibration and development process. This paper also shows the impact of crash location information on analyses results, and underlines that efforts made to manually extract the missing required data can easily be offset by the inaccuracies in crash frequency databases, and the thresholds used to identify intersection related crashes.

2. Methodology

2.1. Data Availability

The available data sources are grouped into three categories: (1) traffic volume data, (2) roadway features data, and (3) crash data. Traffic volume data are compiled from the continuous and short-term traffic count databases and turning movement counts database maintained by the New Jersey Department of Transportation (NJDOT).

[1] * Corresponding author. Tel.: +90-216-564-9080;
E-mail address: bekir.bartin@ozyegin.edu.tr

The key source for roadway features data is the Straight Line Diagrams (SLD) database, maintained by the NJDOT. SLD includes various tables for different geometric and operational features of NJ roadways. The secondary source is the NJ roads centerlines GIS dataset (NJ GIS Map). Motor vehicle crash data come from Safety Voyager crash database, provided by NJDOT for 2011 to 2015. The relevant data elements include a standard route identifier (SRI) i.e. route number, milepost and coordinates of crash location, data, time, severity, collision type, crash type, number of vehicles, fatalities, injuries, pedestrian fatalities and injuries.

The information gathered from these three data sources can be used to generate the data required for the calibration and development of SPFs for R2 segments and intersections. However, before generating these required datasets, the compiled data need be cleaned and corrected. The procedure used to generate the required U2 segments database was implemented in C programming language.

In order to process an analysis ready database both for calibration and development, it is necessary to identify homogeneous road segments. Homogeneity means the geometric, operational characteristics and the AADT along a segment do not vary over the study period. Thus, homogeneous segments are determined by first splitting road segments at intersections, interchanges or any other locations where vehicles are allowed to make turns, and then at each point where there are any changes in geometric or operational characteristics. Following this segmentation procedure, a total of 36,008 homogeneous U2 segments were identified. The HSM suggests using segments of 0.1 mile or longer for calibration and development purposes. It was determined that 11,610 segments were longer than 0.1 mile. It was assumed that the validity of AADT counts assigned to each segment increases with its proximity to the detector used to calculate its AADT value. Of the 11,610 segments, 1,639 were found to include a detector present within the segment.

2.2. Assigning Crashes to Homogeneous Segments

The number of observed crashes at homogeneous segment i, N_o^i, is determined using the available Safety Voyager crash database. N_b^i and CMF_k^i, used in Equation 1, are calculated using the AADT, geometric and operational feature data for each site. States' crash databases are not comprehensive enough to make it possible to differentiate whether a crash is intersection related or not. Using the coordinates of crashes and the available coordinates of all intersections included in the NJ SLD database information, all intersection-related crashes were identified based on the widely accepted 250-ft threshold. The remaining non intersection-related crashes were presumed segment-related, and the observed number crashes at the automatically identified homogeneous segments were determined.

2.3. Final Datasets for Analyses

As mentioned earlier, the automatically identified dataset included 1,639 homogeneous U2 segments of 0.1 mile or longer that included a detector within its bounds, and that the data required by the HSM were extracted manually for 372 segments out of the 1,639 due to time and resource constraints.

Henceforth, the dataset comprised of the 372 segments including the data required by the HSM will be referred to as the *Test Dataset*, and the dataset including remaining 1,267 segments as the *Development Dataset*. Since the crash data were available for five years between 2011 and 2015, the initial sample sizes were 1,860 and 6,335 for test and development datasets, respectively. However, AADT data were not available for all five years as the AADT values are usually collected at every two to three years on each segment. Although the HSM procedure suggests interpolation of the available AADT values to fill in the missing years' values, it was decided to only include the years when detector data were available. With that, the final sample size reduced to 486 and 1,596 for test and development datasets, respectively, corresponding to a training/test sample size split of 77/23 percent. It should be noted that test and development datasets do not vary significantly with respect to geometric and operational features, and crash frequency.

3. Analysis and Results

The main objective of the analyses presented here was to demonstrate the robustness of NJ-specific SPFs developed using the development dataset based on its prediction accuracy on the test dataset, and to compare with the calibrated HSM SPFs. To that end, the analyses were structured as follows:

(1)The test dataset was used to compute the calibration factor for the U2 segments in NJ. The jurisdiction-specific SPFs for U2 segments were developed using the development data.

(2) The development dataset was used to estimate U2 segments SPFs specific to NJ. Four different count regression models, namely negative binomial, Poisson, zero inflated negative binomial (ZINB) and Hurdle models were developed and compared.

(3) The prediction accuracy of the SPFs developed using the development dataset were then compared to the ones of the calibrated HSM SPFs using absolute residual statistics.

3.1. Calibration Results

The crash prediction for urban and suburban segments in the HSM is conducted for 5 different crash types. These are multi-vehicle non-driveway collisions (N_{brmv}), single-vehicle crashes (N_{brsv}), multi-vehicle driveway-related collisions (N_{brdwy}), vehicle-pedestrian collisions (N_{pedr}), and vehicle-bicycle collisions (N_{biker}). CMFs are applied only to the first three collision types.

The Calibrator tool developed by the Federal Highway Administration (FHWA) is used to calculate the calibration factor and measure its goodness of fit. Using the compiled dataset, the calibration factor for U2 segments was found to be **1.35** with a coefficient of variation of 0.11. It was suggested in the literature that a reasonable upper threshold for the coefficient of variation was 0.10 to 0.15. In addition, to assess the validity of the calculated calibration factor, the cumulative residual (CURE) plots with respect to AADT and segment length were generated, as shown in Figure 1.

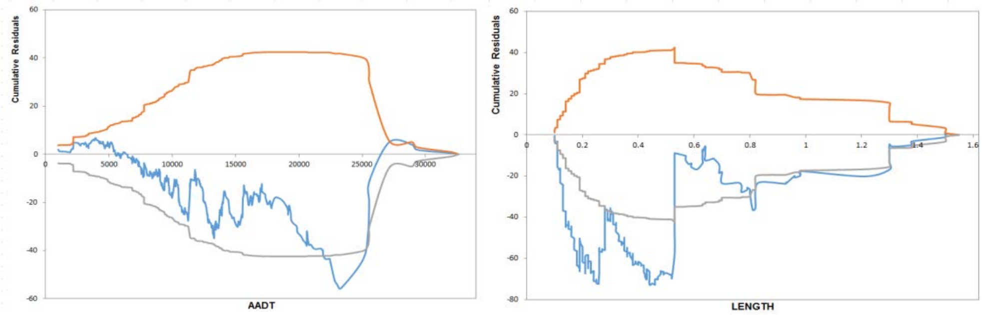

Figure 1: CURE plots for U2 Segments

It is presumed that the CURE plots should be within the expected limits of an unbiased random walk, i.e., plus/minus two standard deviations. In that respect, it can be seen from the CURE plot that the cumulative residuals deviate significantly from the allowable upper and lower bounds. This signifies that despite the calibration factor is close to 1.0 and that coefficient of variation is within acceptable bounds, the calibrated SPF for U2 segments is not statistically acceptable based on the CURE plots. This result warrants the NJ-specific SPF for U2 segments.

3.2. Development Results

The base SPFs for multi vehicle and single vehicle crashes for U2 segments in HSM have the following functional form.

$$N = \exp[a_0 + a_1 . ln(AADT) + a_2 . ln(L)] \qquad (2)$$

Where, $AADT$ is the annual average daily traffic, L is length in miles, and a_0, a_1 and a_2 are model parameters. Note that HSM's predictive model follows this functional form only for multi-vehicle and single vehicle crashes. Multi-vehicle driveway related crash counts are estimated by a simple power function with AADT as a covariate.

The model estimation was performed in R statistical package. The results shown in the paper are the best fitting model parameters after experimenting with models that included shoulder width and speed limit on each segment. Only AADT and length variables came out statistically significant in the count models except in the Poisson model. In the ZINB model, only the length variable came out significant in the zero-inflation component. Therefore, NB, Poisson and hurdle models were selected for further exploration.

The results showed that the hurdle model has a slightly lower Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) values than those of the NB model, which is also significantly lower than the Poisson model, as expected. This was also evidenced from the rootogram plots presented in Figure 2, which compare the observed and expected values graphically by plotting histogram-like rectangles for the observed frequencies and a curve for the theoretical fit.

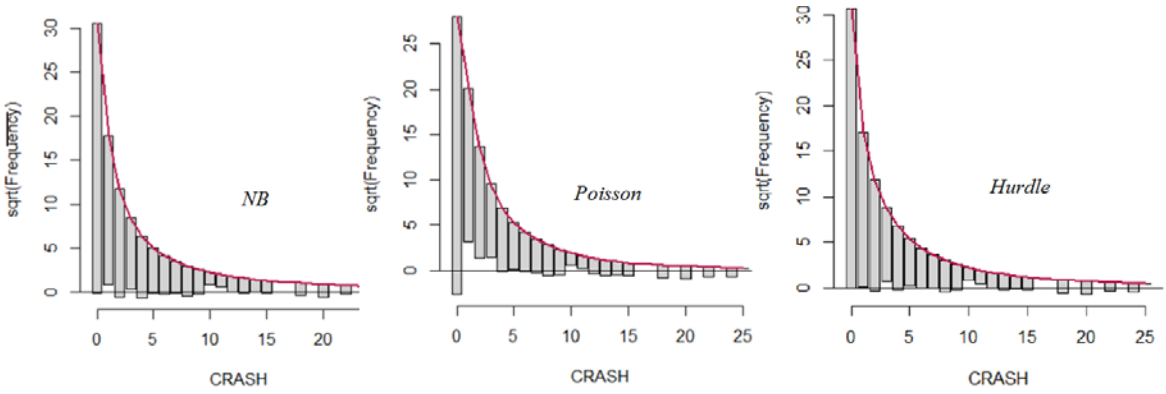

Figure 2: Rootogram plots of developed models

Hanging from each point on the curved line is a bar, the height of which represents the difference between expected and observed counts. A bar hanging below zero indicates underfitting. A bar hanging above zero indicates overfitting.

3.3. Validation Using Test Dataset

The test dataset was used to test the prediction accuracy of the SPFs generated using the development dataset, , and to compare with the HSM SPFs. The histograms of the absolute value of residuals of the SPFs' predicted values and those of calibrated HSM SPFs are plotted in Fig. 3. The red line indicates the histogram of absolute residuals obtained from the NJ-specific SPFs.

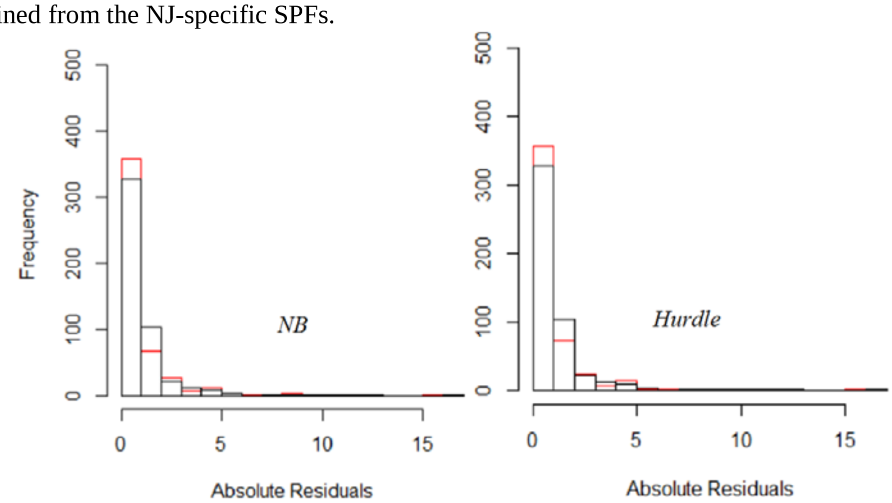

Figure 3: Histogram of absolute residuals – NB and Hurdle models vs. Calibrated HSM SPF

It can be observed that both developed models follow a very similar pattern, and that they outperform the crash predictions of the calibrated HSM SPFs, as their histograms are more skewed to the right, indicating lower absolute residuals. Overall, the average absolute residuals of NB, hurdle and HSM SPFs predictions in the test data were 1.02, 1.01 and 1.11, respectively. For segments with zero crashes, the absolute residual for NB and hurdle was 0.63 whereas it was 0.76 for HSM SPF. It can be seen that hurdle model slightly outperforms the NB model in the test dataset.

4. Discussion

It should be mentioned that the results of the calibration process, and the development process for that matter, strongly relies on N_o^i, the number of observed crashes at each segment. The validity of N_o^i depends on (1) the results of the geocoding post-process performed by NJDOT, and (2) the 250-ft threshold used to identify intersection-related crashes. As to the effect of geocoding threshold, in 2019, the NJDOT updated its post-processing procedure based on a tighter threshold used in geocoding process to increase the accuracy of crash coordinates, which resulted in 14.7 % less number of crashes state-wide compared to its previous version. The calibration process was performed again using the test dataset, but with the previously estimated crashes, and the calibration factor was calculated as **1.74**, a significant deviation from the current value of 1.35, stated before.

Also, correct identification of intersection-related crashes is of upmost importance yet this distinction is not possible in most crash databases, including the Safety Voyager data. A detailed investigation of the developed model residuals showed that many crashes, identified as segment-related as per the 250 ft. threshold, appeared to be intersection-related based on crash characteristics (e.g. cluster of rear-end crashes in the peak periods). When the calibration process was repeated for the test dataset with the 550 ft. threshold, for example, the calibration factor was calculated as **0.71**. The significance range of fluctuation of the calibration factor, from **0.71 to 1.74**, when certain assumptions are modified, sheds light on the fact that efforts made to manually extract the required roadway geometry and operational features data not included in available data repositories can easily be offset by the inaccurate or incomplete entries in crash databases.

5. Conclusions

This paper presented the SPF calibration and development process for the U2 segments in NJ. Data requirements, the availability of required data, and the data processing and extraction methods were presented. The available datasets were grouped into development and test datasets. Four generalized linear models, specific to NJ, were generated using the development database. These were negative binomial, Poisson, zero-inflated Poisson and Hurdle models. The best model fit were based on likelihood ratio test, AIC and BIC statistics, and rootograms. The test database was used to calculate the calibration factor for U2 segments, following the calibration process presented in the HSM. The prediction of the generated models were then evaluated and compared to those of calibrated HSM model, using the test dataset. The results showed that the negative binomial and hurdle models yield nearly 10 percent improvement in average absolute residual statistic. In addition, the impact of crash location on calibration factors was investigated. It was shown that calibration factor varies significantly with the crash location assumptions. Future work will investigate how the generated SPFs change with the crash location assumptions.

Acknowledgment

The study is supported by the NJDOT (FHWA-NJ-2017-001) and partially by C2SMART, a Tier 1 UTC at New York University funded by the USDOT. The contents of this paper only reflect views of the authors who are responsible for the facts and do not represent any official views of any sponsoring organizations or agencies.