



The Open Transportation Journal

Content list available at: <https://opentransportationjournal.com>



RESEARCH ARTICLE

Using Generalized Linear Mixed Models to Predict the Number of Roadway Accidents: A Case Study in Hamilton County, Tennessee

Eric M. Laflamme^{1*}, Peter Way², Jeremiah Roland² and Mina Sartipi²

¹ Department of Mathematics, Plymouth State University, Plymouth, NH 03264, England

² College of Engineering and Computer Science, University of Tennessee at Chattanooga, Chattanooga, TN 37403, USA

Abstract:

Introduction:

A method for identifying significant predictors of roadway accident counts has been presented. This process is applied to real-world accident data collected from roadways in Hamilton County, TN.

Methods:

In preprocessing, an aggregation procedure based on segmenting roadways into fixed lengths has been introduced, and then accident counts within each segment have been observed according to predefined weather conditions. Based on the physical roadway characteristics associated with each individual accident record, a collection of roadway features is assigned to each segment. A mixed-effects Negative Binomial regression form is assumed to approximate the relationship between accident counts and several explanatory variables including roadway characteristics, weather conditions, and several interactions between them. Standard diagnostics and a validation procedure show that our model form is properly specified and suitably fits the data.

Results:

Interpreting interaction terms leads to the follow findings: 1) rural roads with cloudy conditions are associated with relative increases in accident frequency; 2) lower/moderate AADT and rainy weather are associated with relative decreases in accident frequency, while high AADT and rain are associated with relative increases in accident frequency; 3) higher AADT and wider pavements are associated with relative increases in accident frequency; and 4) higher speed limits in residential areas are associated with relative increases in accident frequency.

Conclusion:

Results illustrate the complicated relationship between accident frequency and both roadway features and weather. Therefore, it is not sufficient to observe the effects of weather and roadway features independently as these variables interact with one another.

Keywords: Accident frequency model, Generalized linear mixed model, Negative binomial regression, Roadway features, Weather, Railways.

Article History

Received: October 16, 2019

Revised: January 16, 2020

Accepted: January 24, 2020

1. INTRODUCTION

Roadways in the United States are among the busiest in the world, with more than 240 million vehicles registered and more than 210 million registered drivers [1]. With such busy roadways, we naturally see a high number of motor vehicle accidents. According to the US Department of Transportation and the National Highway Traffic Safety Administration, there were over 7 million motor vehicle crashes in 2016 on US road-

-ways [2]. With these accidents come serious costs to Americans. According to 2015 statistics, motor vehicle crashes were the leading cause of death in the first three decades of Americans' lives [3]. Furthermore, roadway accidents are responsible for \$77.4 billion in loss of productivity, \$76.1 billion in property damage, \$31.5 billion in congestion/delay, and \$23.4 billion in medical costs [4].

While roadway crashes are an unavoidable consequence of busy roadways, researchers in the field aim to find ways to mitigate roadway accidents. To this end, we ask: "why do some roadways observe more accidents than others?" If we can identify the factors or conditions most closely associated with

* Address correspondence to this author at the Department of Mathematics, Plymouth State University, Plymouth, NH 03264, United States; Tel: 603-724-5336; E-mail: emlaflamme@plymouth.edu

roadway accidents, then our findings may inform transportation administrators of the nature of roadway crashes, ultimately helping these decision-makers to implement policy changes to improve safety.

The primary objective of this work is to find the roadway characteristics, weather conditions, and potential interactions between roadway characteristics and weather conditions associated with the variability in observed accident counts. This will be achieved via a Negative Binomial regression model that identifies relationships between accident counts and combinations of weather conditions and roadway features. We feel this approach is novel in that it 1) combines diverse sources of data, weather and roadway characteristics, that are critically important to explaining accident frequency, and 2) uses a flexible model structure that allows us to consider both random effects as well as a variety of interactions between variables that may potentially shed light on the complicated nature of accidents. To the best of our knowledge, currently, no work uses a mixed-model in conjunction with combinations of weather and roadway features. As an application, we consider real-world accident data collected from Hamilton County, Tennessee. Hamilton County is an ideal location for a case study as it contains busy, metropolitan areas surrounding Chattanooga (Tennessee’s fourth largest city) as well as a

number of residential and rural areas. As such, the county has a variety of roadway types (from major interstate highways to rural backroads) and our results are likely to be transferable/generalized.

The remainder of this paper is organized as follows: Section 2 (Literature Review) presents previous research related to this work; Section 3 (Data and Preprocessing) presents the available data used in the analysis, how these data sources are merged together, and ultimately how the data are aggregated into accident counts for each roadway segment; Section 4 (Statistical Modeling) presents the statistical model to be used in this analysis and the theoretical justification for the model choice; Section 5 (Model-fitting) discusses details of the model-fitting procedure and further justification of the model; Section 6 (Results) presents interpretation of model-fitting results as well as diagnostic tests for the model; Section 7 (Validation) provides evidence that our model specification is appropriate for the data; and Section 8 (Discussion and Conclusion) provides a summary of the procedure, potential application, and potential refinements for future studies. As this work involves a complicated process using multiple sources of data and substantial data preparation, a workflow of the analysis is presented below in Fig. (1) to assist the reader.

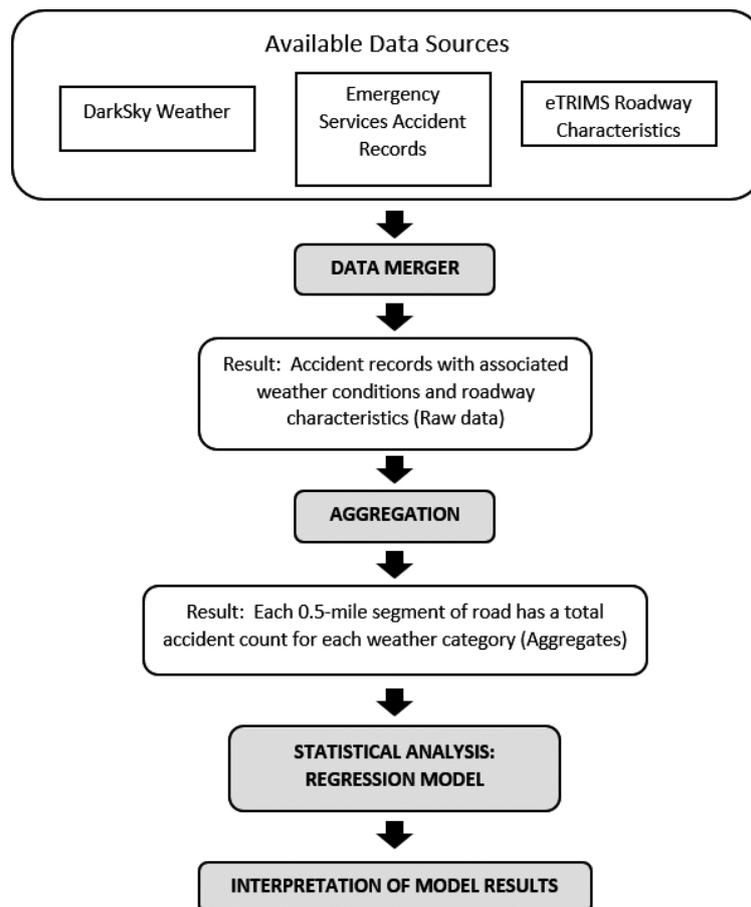


Fig. (1). Workflow of the ensuing analysis presented in this manuscript.

2. LITERATURE REVIEW

As stated above, our work uses a regression model to identify relationships between accident frequency and combinations of weather conditions and roadway features. Previous research has shown that many roadway features, particularly those available to us, are strongly associated with roadway accidents. Specifically, AADT/DHV/ traffic volume [5 - 7], roadway type [8], terrain/grade [9 - 11], land use /type of road/urban versus rural roadways [12, 13], illumination [9], posted speed limit [9, 14, 15], pavement widths [11, 16], pavement type [17], presence of on-ramps/off-ramps [18 - 20], and number of lanes [10, 12, 21], have all been linked to roadway accidents.

In addition to roadway features, weather conditions have been shown to play a role in explaining the variability of accident counts. In an analysis of accident frequencies in Great Britain, Edwards [22] presented temporal trends for 9 broad weather categories as defined by the Department of Transport. More generally, other works suggest weather conditions, especially precipitation events, are important components of accident prediction [18, 20, 23 - 29], and we know intuitively, as drivers, that roadways become more dangerous in inclement weather.

For this analysis, we use a mixed-effects Negative Binomial (NB) model. There is extensive literature supporting the use of the NB form to model roadway accidents, yet these works vary considerably in terms of scope. Chengye and Ranjitkar [20] used NB models and data collected from a motorway in Auckland, NZ to link accident frequency to ‘non-behavioral’ factors such as traffic conditions, roadway characteristics, and weather conditions. The authors performed three NB analyses: one on the entire motorway, one separating urban and rural sections, and one separating segments by on- and off-ramps. Shankar *et al.* [26] used a NB model on data collected from rural highways to identify relationships between accident frequency and roadway geometry, weather factors, and various interactions between them. Hall and Tarko [30] showed that the NB form was suitable to model accident frequency on low-volume, rural roads. Milton and Mannering used the NB model and data collected from Washington highways to isolated the effects of geometric and traffic characteristics on annual accident frequency [31]. Abdel-Aty and Radwan [21] used the NB form to model the frequency of Florida highway accident occurrence based on roadway geometry, urban/rural designations, and section length (the authors also developed NB models to account for age and gender of the driver). Eisenberg [27] used both monthly and daily precipitation as explanatory variables, and found a negative correlation between monthly precipitation and fatal accidents. Chin and Quddus [32] used a random-effect NB model on data collected from signalized intersections in Singapore to identify relationships between accident occurrence and the geometric, traffic, and control characteristics. Shankar *et al.* explored NB and random-effect

NB forms to model median crossover accident frequencies for several routes in Washington [33]. Using roadway geometry, traffic volume, and a random effect to account for different routes, they found the random-effect model to be superior [33]. In a related study, Ulfarsson and Shankar used the same explanatory variables to compare NB and random-effects NB models to a Negative Multinomial model form for accident frequency [34].

As seen, there are numerous works using the NB regression form and/or random-effects models to approximate roadway accidents, and there are works that consider roadway features and/or weather conditions in their statistical models. However, to our knowledge, no current work exists that implements both a random-effect model and this combination of roadway characteristics/features and weather conditions, including interactions between them. Herein lies the novelty of our work.

3. DATA AND PREPROCESSING

3.1. Available Data

A hallmark of this work is the use of detailed and accurate datasets from three distinct sources: Hamilton County emergency services, eTRIMS, and DarkSky.

3.1.1. Hamilton County Emergency Services

The Hamilton County Emergency Communications Center provides emergency dispatch service for 105 Police, Fire, and EMS departments covering 42 political jurisdictions in the county. This service, which collects reports from first responders, was used to collect information for traffic accidents in Hamilton County between January 1, 2017 and December 31, 2018 (the collection period). Importantly, each accident record has an associated physical location measured in terms of route mile marker (log-mile).

Over the collection period of two years, 55,324 total accidents were reported in Hamilton County. As seen in Fig. (2), accidents occurred in highly populated areas within the city limits of Chattanooga as well as in less populated, rural areas in the county.

3.1.2. eTRIMS

Since the 1970s, the Tennessee Department of Transportation (TDOT) has maintained location descriptions and inventory data on all local roads. This information was integrated into the Tennessee Roadway Information Management System (TRIMS) that serves as a referencing database for State and local roadway information. In 2007, TDOT began developing eTRIMS, a map-centric, web-based version of TRIMS with the purpose of encouraging wider use of the TRIMS database. Based on previous literature (Literature Review) and the data available for Hamilton County via eTRIMS, we consider the following roadway and pavement characteristics for our analysis.

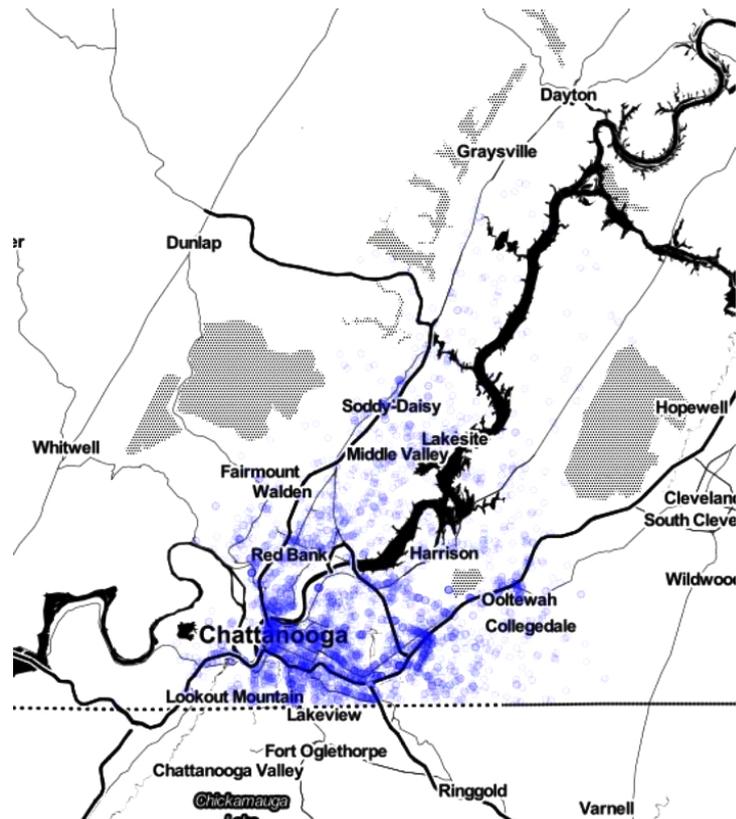


Fig. (2). Locations of accidents (blue dots) collected in Hamilton County, Tennessee, between January 1, 2017, and December 31, 2018.

- Average Annual Daily Traffic or AADT (numerical variable)
- Administrative System (categorical variable representing road system designation; *i.e.*, city highway, city road, interstate, rural highway, or rural road)
- Terrain (categorical variable; *i.e.*, flat, rolling, mountainous)
- Land Use (categorical variable representing designation of the surrounding land; *i.e.*, commercial, industrial, public, residential, rural, or a combination)
- Illumination (indicator variable designating if street lamps are present)
- Posted Speed Limit (numerical variable, measured in miles per hour)
- Pavement Width (numerical variable, measured in feet)
- Pavement Type (categorical variable; *i.e.*, asphalt, cement)
- Design Hour Volume or DHV (numerical variable)
- Number of Lanes (numerical variable)
- Presence of on- and off-ramps (indicator variable)

Since accident records have precise log-mile locations (a designation amenable to the eTRIMS database), the roadway characteristics listed above are easily assigned to each of these accident records.

3.1.3. DarkSky Weather

DarkSky [35] is an API for Python that collects weather

data from several weather resources and makes available information associated with specified locations at precise time periods. Since our accident records have both a time and latitude/longitude designation (designations amenable to the DarkSky data), Darksky weather conditions are easily assigned to each.

For our analysis, to assist the model-fitting procedure and make results more interpretable, we simplify the available weather information. Specifically, we consider a continuous variable for visibility (measured in miles) and a categorical variable that distinguishes between the following broad weather categories.

- *Clear* - no precipitation, no clouds
- *Start rain* - currently raining, no rain in the previous time step
- *Continue rain* - currently raining, raining in previous time step
- *Start cloudy* - currently cloudy, not raining/cloudy in previous time step
- *Continue cloudy* - currently cloudy, raining in previous time step
- *Fog*

In their study of the effects of weather on accident risk, Malin *et al.* [23] used similar, discretized categories for weather conditions.

3.2. Descriptive/Summary Statistics of Data

Table 1. Summary of weather conditions (as defined in section 3.1.3 above) for accidents observed during collection period.

Weather Condition	No. Accidents	% Total
Start cloudy	34,447	62%
Clear	10,667	19%
Start rain	4,355	8%
Continued cloudy	2,707	5%
Continued rain	1,842	3%
Fog	1,283	2%
Other (including missing)	23	.04%

Table 2. Summary of roadway type (Administrative System designation) for accidents observed during collection period.

Rd. Type (Admin. Sys.)	No. Accidents	% Total
City road	33,120	60%
City highway	11,515	21%
Interstate highway	8,073	15%
Rural road	1,958	4%
Other (including missing)	490	1%
Rural highway	168	.3%

3.3. Combining Data/Creating Roadway Aggregates

3.3.1. Defining Roadway Segments

We observe roadways in terms of short, non-overlapping segments. This serves two purposes. First, segments act as interval in which to collect counts, since it is unlikely that more than one accident would occur at any one precise location. Second, by setting the segments to some short, fixed lengths, we are defining a fine gradient over which we view our roadways. Then, by observing roadway features associated with accidents within these short segments, we aim to capture the heterogeneity/variability of roadways (Table 1).

Our choice of segment length is based on the level of precision used by traffic engineers when observing roadways. Traditionally, actual locations on a roadway, or 'spots,' are lengths greater than 0.1 miles as this is believed to be a minimum threshold to discern differences in locations [36]. Others have suggested spot lengths to be at least 0.3 miles long [37] or between 0.3 and 0.5 miles [36]. Segments, then, should be longer than spot distances. With the aim of capturing as much variability in roadway characteristics as possible, we set our segment length at the upper limit for spot length, and we thus define segments as half-mile sections of roadways. Note: subsequent results varied little when choosing segments of slightly shorter (0.3 and 0.4 miles) or longer length (0.6 miles).

For simplicity, our segment length is held constant (fixed-length segmentation) for all roadways in the collection area. Certainly, there are more complicated approaches to identifying roadway segments (many of which are beyond the scope of this analysis); such as those defining segments based

on changes in AADT [38], curvature and roadside danger [39], or a variety of roadway geometries [40]; but there is no universally agreed-upon approach. In their comparison of the effect of roadway segmentation criteria on the performance of safety performance functions, Cafiso *et al.* [7] found that fixed-length segmentation (such as we have) was a flexible and statistically consistent technique. Also, in their study focusing on the relationship between crashes, road geometry, and segmentation, Cenek *et al.* [41] used fixed-length segments (Table 2).

3.3.2. Creating Aggregates

Aggregation, in our case, involves combing accident data, roadway characteristics, and weather conditions for each of the half-mile segments across the county. This task is conceptually straight-forward since all accident records have been previously linked to corresponding roadway characteristics and weather conditions, and each has a log-mile designation.

Note: As we do not want to average weather within segments (and potentially wash away any weather signal), for each segment we create an accident count for each weather category. This allows us to introduce the weather as an explanatory variable in our regression model.

Note: For continuous roadway features (AADT, for example), each segment is assigned the average of values associated with accidents in that segment. For categorical features (terrain, for example), the most commonly observed value is assigned to each segment. Lastly, two features are measured in terms of presence/absence: illumination and change in a number of lanes. Since most segments were homogenous in terms of illumination, we identify a segment as having illumination if most observations had that designation. For the presence of a change in the number of lanes, since we believe the effects of exiting/entering/merging/weaving extend beyond the locations where change occurs, we designate a segment as having this feature if any of the records in the segment had this feature (Table 3).

3.3.3. Final Comments Regarding Segments/Aggregates

After some data cleansing to remove incomplete records, we were left with 5,505 aggregated segments, which will provide the basis for our model-fitting procedure. For each of these, we retain the original route identification to be used as a random effect, or to account for potential roadway variation (literature supports the need for such a designation; see, for example, [15, 42, 43]). Lastly, we note that the segment/aggregate will not be used in our modeling procedure as accident count variability between aggregates is believed to be attributable to segment/aggregate characteristics, weather, or some combination of the two. Also, in an initial data investigation, no autocorrelation was found to exist between counts from adjacent segments/aggregates.

4. Statistical Modeling

For our data, we initially pursued a Poisson regression model to approximate our accident counts by roadway segment. Such models are theoretically justified as they are designed to model count data. A somewhat restrictive

drawback of this model, however, is the Poisson assumption of equi-dispersion (equality of mean and variance). In our case, an initial investigation of the data revealed our count data is over-dispersed (thus, not equi-dispersed) and thus does not follow a Poisson distribution. For such cases, to model count data that does not necessarily follow a Poisson distribution, the Negative Binomial regression model form is typically recommended as a remedial measure. The Negative Binomial model form is considered a variant of the Poisson form and includes an additional parameter to account for over-dispersion. Also, as there is likely a roadway-specific effect on the accident counts, or rather a natural heterogeneity between accident counts collected at different routes, we introduce a random effect component for the roadway. As stated by Ulfarsson and Shankar [34], in the case of accident frequency, it is likely that section-specific effects will be important. Our model is:

Table 3. Summary of selected roadway characteristics for accidents observed during the collection period.

Roadway Characteristic	No. Accidents	% Total
Land Use		
Commercial	21,455	39%
Residential	14,306	26%
Rural	12,298	22%
Industrial	1,472	3%
Other	5,793	10%
AADT		
<50,000	43,034	78%
≥50,000	12,290	22%
Posted Speed Limit		
≥60	2,494	5%
[50, 60)	9,111	16%
[40, 50)	14,722	27%
[30, 40)	15,840	29%
<30	2,948	5%
missing	10,209	18%
Pavement Width		
<8 ft.	27,469	50%
≥8 ft.	27,769	50%

$$Y_{i,j} | \mathbf{u}_j \sim \text{NegBin}(\alpha, \mu_{i,j})$$

with α the NB dispersion parameter, and

$$\mu_{i,j} = E(y_{i,j} | \mathbf{u}_j) = \exp\{\mathbf{x}'_{i,j} \boldsymbol{\beta} + \mathbf{u}_j\}$$

where y_{ij} is the accident count from the i^{th} segment of the j^{th} roadway, \mathbf{u}_j is the **random effect** for the j^{th} roadway, $\boldsymbol{\beta}$ is the vector of fixed-effects model parameters, $\mathbf{x}_{i,j}$ is the design-vector for the i^{th} segment of the j^{th} roadway, and $u \sim N(0, \sigma^2)$.

Very often, count-based regression models like the Negative Binomial include an ‘offset variable’ to account for disparate ‘exposures’ in the data. A common source of varying exposures is changes in collection periods. This, however, is not the case with our data. That is, our data consist of all roadway accidents reported in the county for a two-year period, so each record represents the total number of accidents

occurring in that segment over that two-year period. On the other hand, our data does consist of very diverse roadways that experience varying degrees (exposures) of travel and volume. Thus, an argument could be made to use AADT, a measure of traffic volume, as an offset. We have decided against using AADT (or any other measure of traffic volume, demand, etc.), however, as we wish to directly quantify the relationship between AADT and accident totals, an analysis that would not be possible if we use it as an offset. In their analysis of crashes caused by the running of red lights using Poisson and NB models, Mohamedshah *et al.* [44] similarly omitted AADT as an exposure/offset variable. In the end, no offset was used in our model.

5. MODEL-FITTING

All model-fitting and statistical analyses were performed using the R statistical software system [45]. In addition to the base packages in R, the ‘lme4’ and ‘MASS’ packages [46, 47] were used extensively. Model-fitting was performed via maximum likelihood estimation.

5.1. Independent/Dependent Variables

For our model, we define the response/dependent variable to be the accident count. For the independent variables, we consider fixed-effect terms for average AADT, average speed limit, average pavement width, average DHV, average visibility, average number of lanes, most occurring administrative system, most occurring terrain, most occurring land use, most occurring pavement type, the presence/absence of illumination, and the presence/absence of change in number of lanes; a fixed-effect, categorical term to distinguish accident counts observed under different weather conditions; and a random-effect term for the different roadways (routes). Also, since we assume that the relationship between roadway features/weather and accident frequency is potentially complicated, several logical interactions will be considered. Related to this, in their analysis of safety performance functions, Islam *et al.* [48] identified the importance of incorporating interactions between variables (in particular, those between speed limit and geometric variables) for estimation of crash prediction models.

5.2. Justification for Model Specification/Preliminary Analyses

Since the Poisson model form is nested within the NB model form (the NB model has one additional parameter to account for the over-dispersion), we performed a likelihood ratio test to assess whether or not the use of the more complicated model, the NB model, is justified. As suspected, we found that the NB model form provided a statistically significant improvement over a comparable Poisson model (Likelihood ratio $X^2 = 27002$, $df = 1$, $p\text{-value} < 0.0001$). Thus, the NB model form was pursued for this analysis.

Next, to assess if the inclusion of a random effect is necessary, we compared our NB model form with a simpler NB form with only the fixed effects. As above, the models were compared via likelihood ratio test, and it was found that the NB model with a random effect for the route was a

statistically significant improvement over the NB model without the random effect (Likelihood ratio $X^2= 1198$, $df = 1$, p -value < 0.0001). Thus, the random intercept term was retained for all future analyses.

Finally, we investigated potential multicollinearity between the explanatory variables since severe multicollinearity can result in coefficient estimates that are unstable and difficult to interpret. Using the ‘car’ package in R [49], because we have a combination of continuous and categorical variables, we calculated the Generalized Variance Inflation Factors (GVIFs) for our NB model with only main effects and a random-effect for the route. None of the squared GVIF $\frac{1}{2 \cdot df}$ quantities [50] are larger than 4, suggesting that multicollinearity is not an issue with this set of explanatory variables. Although we expected average AADT and an average number of lanes to be related, for example, the correlation between them does not appear to be troublesome. Also, for these two variables specifically, a likelihood ratio test indicated that both variables should be included in the model (Likelihood ratio $X^2= 116.36$, $df = 1$, p -value < 0.0001).

6. RESULTS

6.1. Initial Results

Using the ‘MuMin’ package in R [51], we calculate the conditional pseudo-R-squared or an estimate of the percentage of the observed variance in the response explained by the mixed-effects model. For our best model, we have a pseudo-R-squared of about 0.80 (0.7959), an indication that our model is highly successful in explaining the variability observed in our accident frequencies. In their NB analysis of accidents on roadway segments with on-ramps, Chengwe and Ranjitkar [20] achieved a likelihood ratio index (somewhat comparable to R-squared) of about 0.19. In their analysis of the frequency of accidents at intersection approaches, Poch and Mannering [52] used several NB models on different types of accidents, and their best model achieved a likelihood ratio index around 0.54.

6.2. Interpretation of Parameter Estimates

For several reasons, interpreting the results of our fitted model is not straight-forward. First, since we assume an NB form, our explanatory variables have a log-linear relationship with our response variable. This means that our estimated parameters must be ‘exponentiated’ to assess their relative effect on the response variable (that is, calculate an ‘incidence rate ratio’). Second, as recommended by Bates *et al.* [46], our continuous explanatory variables were scaled to facilitate the convergence of our model. Thus, these variables are converted to standard deviation units from their mean values (this means continuous variables can have negative values which correspond to values less than the average). Third, we consider a variety of interactions, so relationships between explanatory variables and the response become more complicated and nuanced. Fourth, we have a number of categorical variables, and results for those variables are presented as changes from the ‘base case’ category. Lastly, we have a large number of variables and interactions and it is not reasonable to present and interpret all parameter estimates. However, in the following sub-sections, we will discuss several estimates that

are statistically significant and noteworthy.

Note: a beneficial byproduct of centering variables is that, for some cases, it allows for interpretation of main effects when an interaction is present. Typically, the main effects are not interpreted in the presence of interaction since, in such cases, the main effect represents the effect of one variable when the other is set at a value of zero. This is often an unrealistic scenario as it would be unlikely to observe a variable at a value of zero. However, for centered variables, zero now represents the average, which is realistic.

6.2.1. Administrative System (Roadway Type) and Weather

From the analysis of variance of our best-fitted model, we find that the interaction between administrative system and weather condition (both categorical variables) to be very significant (Type II Wald Chi-square test, $X^2 = 91.2118$, $d.f. = 22$, p -value = $2.130e-10$). Among the specific levels of these two variables, we find the parameter estimate associated with the interaction between “rural road” type and the “start cloudy” weather condition to be significant.

Table (4) below gives the regression model output associated with these variables.

Table 4. Regression estimates associated with rural roads and the start of cloudy weather conditions.

Variable	Regression coefficient	Tests statistic (Wald Z value)	p-value
Rural road	0.278	0.341	0.733
Start cloudy	0.77	10.625	~ 0
(Rural road)*(Start cloudy)	-0.53	-3.586	0.000335

Table 5. Regression estimates associated with rural roads and the start of rainy weather conditions.

Variable	Regression coefficient	Tests statistic (Wald Z value)	p-value
Rural road	0.278	0.341	0.733
Start rain	-1.15	-14.596	~ 0
(Rural road)*(Start rain)	0.515	2.145	0.0319

Analyzing the main effect for rural roads, we set start cloudy to zero (or for ‘clear conditions,’ the base case for weather) and find rural roads affect the number of roadway accidents by a factor of $e^{0.278+0.770-0.53(0)} = 1.32$, or are associated with a 32% increase in accident frequency as compared to being on a ‘city highways’ (base cases for the administrative system). Now, considering the interaction term, rural roads combined with the onset of cloudy conditions affect the number of roadway accidents by a factor of $e^{0.278+0.77-0.53} = e^{0.518} = 1.69$, or are associated with a 69% increase in accident frequency when compared to clear conditions on city highway roads. This highlights the effect of cloudiness on the rural roads, especially when compared to the effect of rural roads in clear conditions, and makes intuitive sense as the onset of cloudy weather is commonly associated with more dangerous

driving conditions. Hermans *et al.* [53] studied the effects of weather variables on accidents at different locations, and they found that for the majority of cases, cloudiness had a significant effect on accident frequency, yet they did not distinguish between roadway types.

Additionally, we find the parameter estimate associated with the interaction between “rural road” and the “start rain” weather condition to be significant. Table (5) above gives the regression model output associated with these variables.

Analyzing the main effect for start rain conditions, the onset of rainy conditions on city highways (base case for road type) affects the number of roadway accidents by a factor of $e^{0.278(0)-1.15+0.515(0)} = 0.317$, or is associated with a 68% decrease in accident frequency as compared to clear weather. Along these lines, Chengye and Ranjitkar [20] found a similar decrease in city highway accidents with the onset of rainy weather. Now, considering the interaction term, the onset of rainy weather combined with rural roads affects the number of roadway accidents by a factor of $e^{0.278-1.15+0.515} = e^{-0.357} = 0.70$, or is associated with a 30% decrease in accident frequency as compared to city highways and clear conditions.

Generally, in summary, we find 1) the effect of rural roads on accident frequency changes depending on clear or cloudy weather conditions, and 2) the effect of the onset of rain on accident frequency changes depending on rural or city highway designations. This illustrates the complicated relationship between weather and road type. Other works found both positive associations between accidents and precipitation [27] or no significant association [54] but most did not consider interactions. As stated by Edwards [22], “...the relationship between adverse weather and road accidents is far from straight-forward,” but investigating interactions between rain and other variables is likely a step closer to understanding this relationship.

6.2.2. AADT and Weather

We find the interaction between AADT and weather conditions (respectively, a continuous and categorical variable) to be very significant (Type II Wald Chi-square tests, $X^2 = 44.3169$, d.f. = 6, p-value = 6.396e-08). Similarly, in a study of relative accident risks across locations in Iowa, Tamerius [55] attributed variations to an interaction effect between traffic volume/density and precipitation. Among the levels of weather conditions, we find the parameter estimate associated with the interaction between AADT and “continued rain” to be significant. Table (6) below gives the regression model output associated with these variables.

Table 6. Regression estimates associated with AADT and continued (sustained) rainy conditions.

Variable	Regression coefficient	Tests statistic (Wald Z value)	p-value
AADT	1.31	13.252	~ 0
Continued rain	-1.72	-17.114	~ 0
(AADT)*(Continued rain)	-0.297	-4.031	~ 0

Analyzing the main effect for continued rain, we set

AADT to zero (average AADT setting) and find continued rain affects the number of roadway accidents by a factor of $e^{1.31(0)-1.72-0.297(0)} = 0.179$, or is associated with an 82% decrease in accident frequency relative to clear weather (base case). When considering the interaction term, for AADT of -1 (one standard deviation below average AADT), the continued rain affects the number of roadway accidents by a factor of $e^{1.31(-1)-1.72-0.297(-1)} = e^{-2.733} = 0.065$, or is associated with a 94% decrease in accident frequency relative to average AADT in clear conditions; and for AADT of 1, the continued rain is associated with a 50% decrease in accident frequency relative to average AADT in clear conditions. However, for AADT of 2, continued rain affects the number of roadway accidents by a factor of $e^{1.31(2)-1.72-0.297(2)} = e^{0.306} = 1.36$, or is associated with a 36% increase in accident frequency relative to average AADT in clear conditions. In summary: for low, average, and moderately high AADT, continued rain is associated with lower accident frequency; for high AADT, continued rain is associated with higher accident frequency. This result is somewhat similar to that of Black *et al.* [56] who concluded that most significant increases in accident risk occur in areas with major highways (locations with higher average AADT). Also, from their study of major roads in Finland, Malin *et al.* [23] concluded that accident risk in poor weather conditions was higher on motorways than on lower volume roads.

6.2.3. AADT and Pavement Width

We find the interaction between AADT and pavement width (both continuous variables) to be very significant (Type II Wald Chi-square tests, $X^2 = 61.5266$, d.f. = 1, p-value = 4.368e-15). Table (7) below gives relevant regression model output associated with AADT and pavement width.

Table 7. Regression estimates associated with AADT and pavement width.

Variable	Regression coefficient	Tests statistic (Wald Z value)	p-value
AADT	1.31	13.252	~ 0
Pavement width	-0.113	4.937	~ 0
(AADT)*(Pavement width)	0.16	7.844	~ 0

Analyzing the main effect for pavement width, we set AADT to zero (average AADT setting) and find a one standard deviation increase in pavement width affects the number of roadway accidents by a factor of $e^{-0.113} = 0.893$, or is associated with an 11% decrease in accident frequency. This agrees with the findings of Zegeer *et al.* [11], for example, who found that certain types of accident rates decreased with increasing lane width. However, considering the interaction term, when both pavement width and AADT are set to 1, we observe an increase in accident frequency by a factor of $e^{1.31(1)-.113(1)+0.16(1)} = e^{1.357} = 3.88$. This result seems to be counterintuitive, and it seems to contradict traditional findings ([11], for example), but recent work supports our result and generally that pavement/roadway width has a more complicated relationship with accident rates. In the very comprehensive analysis of crash data collected from both Tokyo and Toronto, Karim [57] found wider lanes both adversely affect overall side-impact collisions and are

associated with higher crash rates on busy, high AADT roads. As stated by the author, “given...the empirical evidence..., the 'wider is safer' approach based on personal or intuitional opinion should be discarded once and for all.” The author further suggests a logical reason for this finding: narrower lanes result in less aggressive driving and more ability to slow or stop a vehicle to avoid collision.

6.2.4. Land Use and Speed Limit

Lastly, we find the interaction between land use and posted speed limit (respectively, a categorical and a continuous variable) to be very significant (Type II Wald Chi-square tests, $X^2 = 95.9517$, d.f. = 5, p-value = $2e-16$). Among the levels of land use, we find the parameter estimate associated with the interaction between speed limit and “residential area” to be significant. Table (8) below gives the regression model output associated with these variables.

Table 8. Regression estimates associated with residential area and posted speed limit.

Variable	Regression coefficient	Tests statistic (Wald Z value)	p-value
Residential area	-0.04	-0.895	0.371
Speed limit	-0.083	-2.246	0.025
(Sp. limit)*(Residential area)	0.41	8.662	~ 0

Analyzing the main effect for speed limit, for residential area set to 0 (or ‘commercial area,’ the base case), a one standard deviation increase in speed limit affects the number of roadway accidents by a factor of $e^{-0.083} = 0.92$, or is associated with an 8% decrease in accident frequency. Now, considering the interaction term, a one standard deviation increase in speed limit in residential areas affects the frequency of accidents by a factor of $e^{-0.04 \cdot 0.083 + 0.41} = 1.34$, or is associated with a 34% increase in accident frequency compared to commercial locations. In a study of residential areas in Switzerland, Lindenmann [58] found that higher posted speed limits were associated with considerably higher accident occurrence, a result that agrees with our findings. Many other works give opposing results, either positive associations between speed and accident frequency [59 - 61], or negative associations, [62, 63]. As stated by Qudus [42], “...the relationship between the change in traffic speed and the probability of accident occurrence is not as straightforward as one may think.” In response to one such contradictory result, Aarts and van Schagen [64] stated that the “discrepancy may be explained by the fact that the factors that were examined...do not stand alone, but [rather] interact with each other.” Our findings support this claim, that the relationship between speed/speed limit and accident frequency is not absolute, but rather conditional on other factors. Furthermore, if we consider land use as a proxy for general differences in roadway structure, our result highlights the dependency between geometry and speed and their combined influence on accident frequency.

6.3. Model Diagnostics

The key assumptions for NB models, and GLMs in general, are homogenous, Normal, and independent deviations centered on zero [65]. For our case, these assumptions are met and we present diagnostic plots for support (see Fig. (3)). From the histogram (left) and Normal quantile-quantile plot (center) of the model residuals, we have evidence that deviations are both Normal and centered at zero. From the autocorrelation plot of residuals, we find no evidence of dependence. Thus, our model form is valid.

7. VALIDATION

To validate our procedure and model specification, we perform a Cross-Validation (CV) procedure. Specifically, the CV technique used here is a repeated stratified random subsampling procedure where the data is tested against itself via training and testing subsets. Simple random sub-sampling to create training and test sets is not appropriate in our particular case since our model is estimating a random effect for route location. Thus, we need to ensure both training and testing sets contain the same routes, the random effect. This also means that some data are ‘lost’ since we have to omit data corresponding to routes with only one aggregate (only 7 routes had one aggregate). For each of the remaining routes (242), half of the associated aggregates were randomly assigned to a training set and the other half were assigned to a testing set (50% split).

Next, our NB model is fitted to the training set. The fitted regression parameters (based on the training set) and the observed values of the variables in the testing set are then used to predict the number of accidents in the testing set. This process was repeated 100 times, and for each repetition, for each route, 50% of the aggregates are randomly selected for training and testing sets.

To assess our model, we observe the root mean squared error, or RMSE, given by

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

where \hat{y}_i is the predicted response (number of accidents), y_i is the observed response (number of accidents), and n is the sample size. When we compare the RMSE from both the training and test sets, a good model will yield similar results. Otherwise, if the RMSE calculated from the training and test sets differ, we have evidence that our model is over-fitting or under-fitting the data. Since we produce an RMSE for the training and test set for each of the 100 iterations of our CV procedure, we have paired RMSEs and we used a paired test to identify potential differences. Assuming Normality in the differences, we have no evidence to suggest that the training and testing sets yield different RMSEs ($t = 0.542$, $df = 99$, p-value = 0.5887). We can conclude that our model is suitable and neither over- nor under-fitting the data.

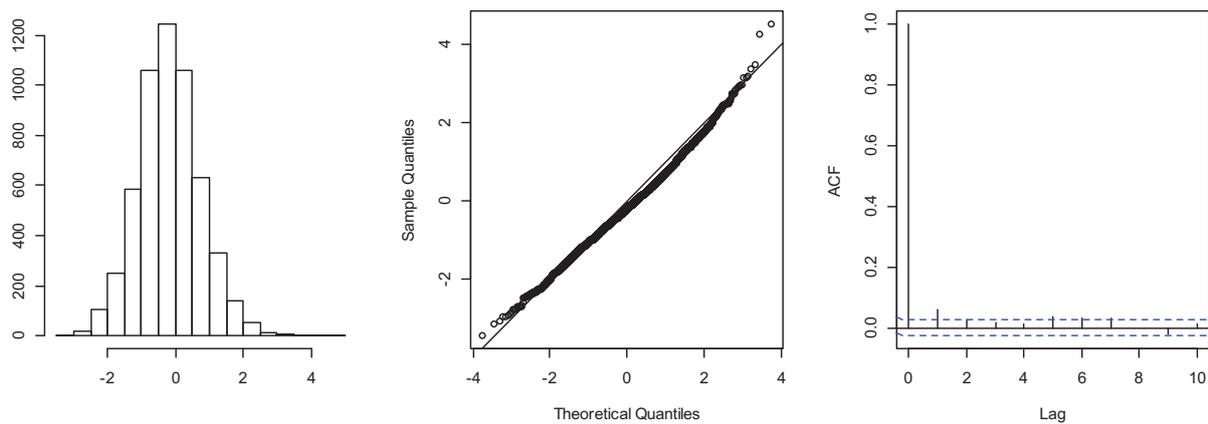


Fig. (3). Model diagnostics for NB with random effect: histogram of residuals (left), Normal quantile-quantile plot of residuals (center), and auto-correlation (ACF) plot of residuals. These plots provide evidence that the assumptions of our GLM have been appropriately met.

8. DISCUSSION

We acknowledge that ‘human error’ plays an important role in roadway accidents [66, 67]. While we cannot detect this effect directly, we can quantify it indirectly. That is, we assume that human error will ‘emerge’ in the presence of certain roadway features, weather conditions, *etc.*, and consequently, accidents will occur more frequently. Thus, we can consider our variables as catalysts, in some way, for human error.

In this work, we present a procedure aimed at identifying the relationships between roadway accident frequency and a variety of variables, and in so doing, we introduce several novel approaches. We present a segmentation/aggregation procedure that first breaks roadways into half-mile sections, then for accidents within each segment, a variety of roadway characteristics and features are identified. Next, we introduce a weather variable into our analysis by distinguishing accident totals by broad weather condition categories. Finally, we introduce a Negative Binomial (NB) regression model, a form that is theoretically justified, to identify the log-linear relationship between accident counts and roadway features, weather conditions, several interactions between weather and roadway features, and a random effect for the particular roadway.

Our procedure described above is applied to real-world data collected in Hamilton County, TN. Based on variation explained, model diagnostics, and a validation procedure, our methodology and model specification is justified. A novel result from our fitted model is the significance of several interactions between variables that are traditionally associated with accident frequency. Interpretations of coefficient estimates associated with these interactions are often logical or agree with current literature and the work of researchers in the field. Furthermore, statistically significant interactions suggest that the relationships between variables and accident frequency are nuanced, complicated, and cannot be described via main effects only. Understanding such relationships is critical to understanding the conditions associated with roadway accidents so that, ultimately, these conditions can be corrected [52].

Interpreting interaction terms from our fitted regression model, though sometimes complicated by categorical variables with several levels, leads to several interesting findings. First, when observing rural roads, cloudy conditions are associated with relative increases in accident frequency; and when observing the onset of rainy conditions, rural roads are associated with relative increases in accident frequency. Second, we find an interesting result related to the combined effect of AADT and continued, sustained rain. For low, average, and moderately high AADT, continued rain is associated with relative decreases in accident frequency; and for high AADT, continued rain is associated with relative increases in accident frequency. Third, considering the interaction between pavement width and AADT, we find higher AADT and wider pavements are associated with relative increases in accident frequency, whereas average AADT and wider pavements width are associated with relative decreases in accident frequency. Fourth, for higher speed limits, being in a residential area is associated with relative increases in accident frequency. These results illustrate the complicated relationship between accident frequency and both roadway features and weather. Furthermore, it is not sufficient to observe the effects of weather and roadway features independently as these variables interact with one another.

Like any work, there are a number of refinements that can be made to this work. First, our accident records have precise date and time designations, and it would be feasible to introduce a time-of-day, day-of-week, or season variable to our work. Related to this, Clarke *et al.* [68] showed how time-of-day (along with other variables) affects the crash frequency of young drivers, Qin *et al.* [69] revealed how the relationship between crashes and hourly traffic varies by time of day, and Laflamme and Ossebruggen [70] provided evidence of time-of-day and day-of-week effects on highway congestion which itself has been linked to accident occurrence [71]. We, however, opted for a more parsimonious approach, focusing on roadway characteristics and weather conditions. Second, our procedure does not consider the severity of accidents. Although severity information (such as fatality, *etc.*) was not available at

the time of this analysis, including such information in the data aggregation procedure would be straight-forward. Numerous published works have performed analyses distinguishing between accidents of varying severity [12, 15]. Lastly, it has been shown that additional factors such as vehicle type [72, 73]; driver demographics such as age, gender, *etc.* [21, 74]; and driver state [75] play important roles in accident frequency/severity. Future work will pursue these variables.

CONCLUSION

Lastly, we acknowledge that our procedure makes several simplifications that could overlook important features. For example, among our weather variable categories, we do not distinguish between rain events of different intensity. Such data is available, and future work could introduce additional rain categories, or possibly create a continuous variable for the rain to account for varying intensity. Several works have investigated the effect of rain intensity on roadway accidents [23, 76].

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The authors confirm that the data supporting the findings of this research are available within the article.

FUNDING

None.

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The authors would like to thank the University of Tennessee at Chattanooga's SimCenter. This work would not have been possible without their resources and support.

REFERENCES

- [1] US Department of Transportation, Federal Highway Administration, *Policy and Governmental Affairs, Office of Highway Policy Information, Highway Statistics [online]*, 2017. Available from: <https://www.fhwa.dot.gov/policyinformation/statistics/2017/mv1.cfm> [Accessed July 2019].
- [2] US Department of Transportation, National Highway Traffic Safety Administration, *Traffic Safety Facts, Summary of Motor Vehicle Crashes. [online]*. Available from: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812580> [Accessed July 2019].
- [3] Centers for Disease Control and Prevention, *Injury Prevention and Control, Leading Causes of Death Charts, Causes of Injury Death: Highlighting Unintentional Injury. [online]*. Available from: https://www.cdc.gov/injury/wisqars/pdf/leading_causes_of_injury_dea_ths_highlighting_unintentional_injury_2015-a.pdf [Accessed July 2019].
- [4] L. Blincoe, T.R. Miller, E. Zaloshnja, and B.A. Lawrence, *The economic and societal impact of motor vehicle crashes, 2010*, Revised, 2015.
- [5] F. Mustakim, I. Yusof, I. Rahman, A.A.A. Samad, and N.E.M. Salleh, "Blackspot study and accident prediction model using multiple liner regression", *First Int. Conf. Constr. Dev. Ctries.*, 2008 pp. 121-36
- [6] L. Lin, Q. Wang, and A.W. Sadek, "A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction", *Transp. Res., Part C Emerg. Technol.*, vol. 55, pp. 444-459, 2015. [http://dx.doi.org/10.1016/j.trc.2015.03.015]
- [7] S. Cafiso, C. D'Agostino, and B. Persaud, "Investigating the influence of segmentation in estimating safety performance functions for roadway sections", *J Traffic Transp Eng.*, vol. 5, pp. 129-36, 2018.
- [8] KR Agent, and RC Deen, *Relationships between roadway geometrics and accidents*, 1974.
- [9] C Dong, C Shao, J Li, and Z Xiong, "An improved deep learning model for traffic crash prediction", *J Adv Transp*, vol. 2018, 2018.
- [10] K.R. Kalokota, and P.N. Seneviratne, *Accident prediction models for two-lane rural highways.*, Mountain-Plains Consortium, 1994.
- [11] C.V. Zegeer, R. Stewart, F. Council, and T.R. Neuman, "Accident relationships of roadway width on low-volume roads", *Transp. Res. Rec.*, p. 160, 1994.
- [12] X. Yan, B. Wang, M. An, and C. Zhang, "Distinguishing between rural and urban road segment traffic safety based on zero-inflated negative binomial regression models", *Discret Dyn Nat Soc*, vol. 2012, 2012. [http://dx.doi.org/10.1155/2012/789140]
- [13] T. Lu, Z. Donyao, Y. Lixin, and Z. Pan, "The traffic accident hotspot prediction: Based on the logistic regression method", *2015 Int. Conf. Transp. Inf. Saf.*, , 2015pp. 107-10 [http://dx.doi.org/10.1109/ICTIS.2015.7232194]
- [14] N.V. Malyskhina, and F. Mannering, "Effect of increases in speed limits on severities of injuries in accidents", *Transp. Res. Rec.*, no. 2083, pp. 122-127, 2008. [http://dx.doi.org/10.3141/2083-14]
- [15] J. Ma, and K. Kockelman, "Crash frequency and severity modeling using clustered data from Washington state", *2006 IEEE Intell. Transp. Syst. Conf.*, , 2006pp. 1621-6
- [16] L.I. Griffin, and K.K. Mak, *Benefits to be achieved from widening rural, two-lane, farm-to-market roads in Texas.*, National Research Council, Transportation Research Board, 1988.
- [17] SN Baskara, H Yaacob, MR Hainin, and SA Hassan, "Accident due to pavement condition-a review.", *J Teknol*, vol. 78, .
- [18] L. Wang, Q. Shi, and M. Abdel-Aty, "Predicting crashes on expressway ramps with real-time traffic and weather data", *Transp. Res. Rec.*, no. 2514, pp. 32-38, 2015. [http://dx.doi.org/10.3141/2514-04]
- [19] A.T. McCart, V.S. Northrup, and R.A. Retting, "Types and characteristics of ramp-related motor vehicle crashes on urban interstate roadways in Northern Virginia", *J. Safety Res.*, vol. 35, no. 1, pp. 107-114, 2004. [http://dx.doi.org/10.1016/j.jsr.2003.09.019] [PMID: 14992851]
- [20] P. Chengye, and P. Ranjitkar, "Modelling motorway accidents using negative binomial regression", *J. East. Asia Soc. Transp. Stud.*, vol. 10, pp. 1946-1963, 2013.
- [21] M.A. Abdel-Aty, and A.E. Radwan, "Modeling traffic accident occurrence and involvement", *Accid. Anal. Prev.*, vol. 32, no. 5, pp. 633-642, 2000. [http://dx.doi.org/10.1016/S0001-4575(99)00094-9] [PMID: 10908135]
- [22] J.B. Edwards, "The temporal distribution of road accidents in adverse weather", *Meteorol. Appl.*, vol. 6, pp. 59-68, 1999. [http://dx.doi.org/10.1017/S1350482799001139]
- [23] F. Malin, I. Norros, and S. Innamaa, "Accident risk of road and weather conditions on different road types", *Accid. Anal. Prev.*, vol. 122, pp. 181-188, 2019. [http://dx.doi.org/10.1016/j.aap.2018.10.014] [PMID: 30384088]
- [24] G. Yannis, and M.G. Karlaftis, "Weather effects on daily traffic accidents and fatalities: A time series count data approach", *Proc. 89th Annu. Meet. Transp. Res. Board*, vol. 10, 2010 p. 14
- [25] X Zou, and WL Yue, "A bayesian network approach to causation analysis of road accidents using netica", *J Adv Transp*, vol. 2017, .
- [26] V. Shankar, F. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies", *Accid. Anal. Prev.*, vol. 27, no. 3, pp. 371-389, 1995. [http://dx.doi.org/10.1016/0001-4575(94)00078-Z] [PMID: 7639921]
- [27] D. Eisenberg, "The mixed effects of precipitation on traffic crashes", *Accid. Anal. Prev.*, vol. 36, no. 4, pp. 637-647, 2004. [http://dx.doi.org/10.1016/S0001-4575(03)00085-X] [PMID: 15094418]
- [28] L-Y. Chang, and W-C. Chen, "Data mining of tree-based models to analyze freeway accident frequency", *J. Safety Res.*, vol. 36, no. 4, pp. 365-375, 2005. [http://dx.doi.org/10.1016/j.jsr.2005.06.013] [PMID: 16253276]

- [29] D. Jaroszowski, and T. McNamara, "The influence of rainfall on road accidents in urban areas: A weather radar approach", *Travel Behav. Soc.*, vol. 1, pp. 15-21, 2014.
[http://dx.doi.org/10.1016/j.tbs.2013.10.005]
- [30] T. Hall, and A.P. Tarko, "Adequacy of negative binomial models for managing safety on rural local roads", *Accid. Anal. Prev.*, vol. 128, pp. 148-158, 2019.
[http://dx.doi.org/10.1016/j.aap.2019.03.001] [PMID: 31022630]
- [31] J. Milton, and F. Mannering, "The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies", *Transportation (Amst)*, vol. 25, pp. 395-413, 1998.
[http://dx.doi.org/10.1023/A:1005095725001]
- [32] H.C. Chin, and M.A. Quddus, "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections", *Accid. Anal. Prev.*, vol. 35, no. 2, pp. 253-259, 2003.
[http://dx.doi.org/10.1016/S0001-4575(02)00003-9] [PMID: 12504146]
- [33] V.N. Shankar, R.B. Albin, J.C. Milton, and F.L. Mannering, "Evaluating median crossover likelihoods with clustered accident counts: An empirical inquiry using the random effects negative binomial model", *Transp. Res. Rec.*, no. 1635, pp. 44-48, 1998.
[http://dx.doi.org/10.3141/1635-06]
- [34] G.F. Ulfarsson, and V.N. Shankar, "Accident count model based on multiyear cross-sectional roadway data with serial correlation", *Transp. Res. Rec.*, no. 1840, pp. 193-197, 2003.
[http://dx.doi.org/10.3141/1840-22]
- [35] Darksy n.d., www.darksy.net
- [36] JA Deacon, C V Zegeer, and RC Deen, *Identification of hazardous rural highway locations*, 1974.
- [37] J.K. Kihlberg, and K.J. Tharp, *Accident rates as related to design elements of rural highways.*, NCHRP Rep, 1968.
- [38] S. Cafiso, and G. Di Silvestro, "Performance of safety indicators in identification of black spots on two-lane rural roads", *Transp. Res. Rec.*, no. 2237, pp. 78-87, 2011.
[http://dx.doi.org/10.3141/2237-09]
- [39] S. Cafiso, A. Di Graziano, G. Di Silvestro, G. La Cava, and B. Persaud, "Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables", *Accid. Anal. Prev.*, vol. 42, no. 4, pp. 1072-1079, 2010.
[http://dx.doi.org/10.1016/j.aap.2009.12.015] [PMID: 20441815]
- [40] *Officials AA of SH and T. Highway safety manual.*, vol. 1. AASHTO Washington: DC, 2010.
- [41] P.D. Cenek, R.B. Davies, M.W. McLarin, G. Griffith-Jones, and N.J. Locke, *Road environment and traffic crashes.*, Transfund New Zeal Res Rep, 1997.
- [42] M. Quddus, "Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS", *J. Transp. Saf. Secur.*, vol. 5, pp. 27-45, 2013.
[http://dx.doi.org/10.1080/19439962.2012.705232]
- [43] X. Wang, Q. Zhou, M. Quddus, T. Fan, and S. Fang, "Speed, speed variation and crash relationships for urban arterials", *Accid. Anal. Prev.*, vol. 113, pp. 236-243, 2018.
[http://dx.doi.org/10.1016/j.aap.2018.01.032] [PMID: 29433070]
- [44] Y.M. Mohamedshah, *Association of Selected Intersection Factors with Red-light-running Crashes.*, Summary Report. Highway Safety Information System. US Department of Transportation, Federal Highway Administration, Research, 2000.
- [45] R.C. Team, *R Core Team. A language and environment for statistical computing.*, R Foundation for Statistical Computing: Vienna, Austria, 2015.
- [46] D Bates, M Mächler, B Bolker, and S Walker, *Fitting linear mixed-effects models using lme4.* *ArXiv Prepr ArXiv14065823*, 2014.
- [47] W.N. Venables, and B.D. Ripley, *Modern applied statistics.*, New York, 2002.
[http://dx.doi.org/10.1007/978-0-387-21706-2]
- [48] M.S. Islam, J.N. Ivan, N.E. Lownes, R.A. Ammar, and S. Rajasekaran, "Developing safety performance function for freeways by considering interactions between speed limit and geometric variables", *Transp. Res. Rec.*, no. 2435, pp. 72-81, 2014.
[http://dx.doi.org/10.3141/2435-09]
- [49] J. Fox, and S. Weisberg, *An R Companion to Applied Regression.*, 2011.<http://socserv.socsci.mcmaster.ca/jfox/Books>
- [50] J. Fox, and G. Monette, "Generalized collinearity diagnostics", *J. Am. Stat. Assoc.*, vol. 87, pp. 178-183, 1992.
[http://dx.doi.org/10.1080/01621459.1992.10475190]
- [51] K. Barton, *MuMIn: multi-model inference, R package version 0.12. 0.*, 2009.<http://R-Forge-r-Project.Org/Projects/MuMIn>
- [52] M. Poch, and F. Mannering, "Negative binomial analysis of intersection-accident frequencies", *J. Transp. Eng.*, vol. 122, pp. 105-113, 1996.
[http://dx.doi.org/10.1061/(ASCE)0733-947X(1996)122:2(105)]
- [53] E. Hermans, T. Brijs, T. Stiers, and C. Offermans, "The impact of weather conditions on road safety investigated on an hourly basis", *Transportation Research Board*, 2006.
- [54] M.M. Hasan, "Investigation of the effect of traffic parameters on road hazard using classification tree model", *IJTTE Int. J. Traffic Transp. Eng.*, vol. 2, pp. 271-285, 2012.
[http://dx.doi.org/10.7708/ijtte.2012.2(3).08]
- [55] J.D. Tamerius, X. Zhou, R. Mantilla, and T. Greenfield-Huitt, "Precipitation effects on motor vehicle crashes vary by space, time, and environmental conditions", *Weather Clim. Soc.*, vol. 8, pp. 399-407, 2016.
[http://dx.doi.org/10.1175/WCAS-D-16-0009.1]
- [56] A.W. Black, G. Villarini, and T.L. Mote, "Effects of rainfall on vehicle crashes in six US states", *Weather Clim. Soc.*, vol. 9, pp. 53-70, 2017.
[http://dx.doi.org/10.1175/WCAS-D-16-0035.1]
- [57] D.M. Karim, "Narrower lanes, safer streets", *Proc. CITE Conf. Regina*, 2015pp. 1-21
- [58] H.P. Lindenmann, "The effects on road safety of 30 kilometer-per-hour zone signposting in residential districts", *Inst Transp Eng ITE J*, vol. 75, p. 50, 2005.
- [59] R. Fowles, and P.D. Loeb, "Speeding, coordination, and the 55-MPH limit", *Am. Econ. Rev.*, vol. 79, pp. 916-921, 1989.
- [60] C.N. Kloeden, G. Ponte, and A.J. McLean, *Travelling speed and the risk of crash involvement on rural roads, report CR 204.*, Aust Transp Saf Bur Canberra, 2001.
- [61] M.C. Taylor, A. Baruya, and J.V. Kennedy, *The relationship between speed and accidents on rural single-carriageway roads.*, vol. 511. TRL, 2002.
- [62] A. Baruya, "Speed-accident relationship on European roads. Work. Pap. R 1.1. 3, Deliv", *D7, Technical Research Centre of Finland VTT Espoo*, 1998.
- [63] N.J. Garber, and R. Gadiraju, "Factors affecting speed variance and its influence on accidents", *Transp. Res. Rec.*, no. 1213, pp. 64-71, 1989.
- [64] L. Aarts, and I. van Schagen, "Driving speed and the risk of road crashes: A review", *Accid. Anal. Prev.*, vol. 38, no. 2, pp. 215-224, 2006.
[http://dx.doi.org/10.1016/j.aap.2005.07.004] [PMID: 16256932]
- [65] A.J. Dobson, and A.G. Barnett, *Exponential family and generalized linear models. An Intro to Gen Linear Model Boca Raton.*, Chapman Hall/CRC: FL, 2002.
- [66] L. Evans, *Traffic safety and the driver.*, Science Serving Society, 1991.
- [67] R. West, D. French, R. Kemp, and J. Elander, "Direct observation of driving, self reports of driver behaviour, and accident involvement", *Ergonomics*, vol. 36, no. 5, pp. 557-567, 1993.
[http://dx.doi.org/10.1080/00140139308967912] [PMID: 8500475]
- [68] D.D. Clarke, P. Ward, C. Bartle, and W. Truman, "Young driver accidents in the UK: the influence of age, experience, and time of day", *Accid. Anal. Prev.*, vol. 38, no. 5, pp. 871-878, 2006.
[http://dx.doi.org/10.1016/j.aap.2006.04.012] [PMID: 16600166]
- [69] X. Qin, J.N. Ivan, N. Ravishanker, J. Liu, and D. Tepas, "Bayesian estimation of hourly exposure functions by crash type and time of day", *Accid. Anal. Prev.*, vol. 38, no. 6, pp. 1071-1080, 2006.
[http://dx.doi.org/10.1016/j.aap.2006.04.012] [PMID: 16782038]
- [70] E.M. Laflamme, and P.J. Ossenbruggen, "Effect of time-of-day and day-of-the-week on congestion duration and breakdown: A case study at a bottleneck in Salem, NH", *J Traffic Transp Eng*, vol. 4, .
[http://dx.doi.org/10.1016/j.jtte.2016.08.004]
- [71] M. Zhou, and V.P. Sisiopiku, "Relationship between volume-to-capacity ratios and accident rates", *Transp. Res. Rec.*, no. 1581, pp. 47-52, 1997.
[http://dx.doi.org/10.3141/1581-06]
- [72] L-Y. Chang, and H-W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques", *Accid. Anal. Prev.*, vol. 38, no. 5, pp. 1019-1027, 2006.
[http://dx.doi.org/10.1016/j.aap.2006.04.009] [PMID: 16735022]
- [73] C-W. Pai, "Motorcyclist injury severity in angle crashes at T-junctions: Identifying significant factors and analysing what made motorists fail to yield to motorcycles", *Saf. Sci.*, vol. 47, pp. 1097-1106, 2009.
[http://dx.doi.org/10.1016/j.ssci.2008.12.007]
- [74] F. Valent, F. Schiava, C. Savonitto, T. Gallo, S. Brusaferrero, and F. Barbone, "Risk factors for fatal road traffic accidents in Udine, Italy", *Accid. Anal. Prev.*, vol. 34, no. 1, pp. 71-84, 2002.

- [75] [http://dx.doi.org/10.1016/S0001-4575(00)00104-4] [PMID: 11789577] F. Meng, S. Li, L. Cao, M. Li, Q. Peng, C. Wang, and W. Zhang, "Driving fatigue in professional drivers: A survey of truck and taxi drivers", *Traffic Inj. Prev.*, vol. 16, no. 5, pp. 474-483, 2015.
- [76] [http://dx.doi.org/10.1080/15389588.2014.973945] [PMID: 25357206] LY Yuan, BY Chen, and WHK Lam, "Effects of rainfall intensity on traffic crashes in Hong Kong", *Proc. Inst. Civ. Eng.*, vol. 167, pp. 343-50. [http://dx.doi.org/10.1680/tran.12.00087]

© 2020 Laflamme *et al.*

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: (<https://creativecommons.org/licenses/by/4.0/legalcode>). This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.