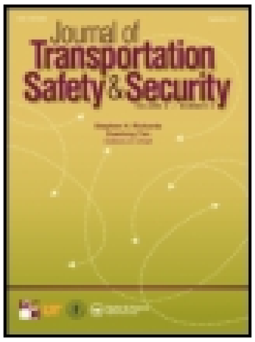

Journal of Transportation Safety & Security

ISSN: 1943-9962 (Print) 1943-9970 (Online) Journal homepage: https://www.tandfonline.com/loi/utss20

Flooding related traffic crashes: findings from association rules

Subasish Das, Xiaoduan Sun, Sarvesh Goel, Ming Sun, Ashifur Rahman & Anandi Dutta

To cite this article: Subasish Das, Xiaoduan Sun, Sarvesh Goel, Ming Sun, Ashifur Rahman & Anandi Dutta (2020): Flooding related traffic crashes: findings from association rules, Journal of Transportation Safety & Security

To link to this article: https://doi.org/10.1080/19439962.2020.1734130

Published online: 09 Mar 2020.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2020.1734130

Flooding related traffic crashes: findings from association rules

Subasish Das[a], Xiaoduan Sun[b], Sarvesh Goel[a], Ming Sun[b], Ashifur Rahman[a], and Anandi Dutta[c]

[a]Texas A&M Transportation Institute, Texas A&M University, College Station, Texas, USA; [b]Civil Enginerring Department, University of Louisiana at Lafayette, Lafayette, Louisiana, USA; [c]Department of Computer Science, University of Texas at San Antonio, San Antonio, Texas, USA

ABSTRACT

In the United States, flooding related crashes have exposed the vulnerability of transport networks. Without realizing the depth of a flooded roadway, some may attempt to cross. At night during heavy storms, determining if a road is flooded is especially difficult. Only 18–24 inches of moving water are needed to sweep away a truck, and only 6 inches are needed to carry away a small car. Louisiana's geographic location places the state in a unique position to receive both frontal tropical hurricanes and large air masses, which may result in high air moistures in almost any direction and drop rain with heavy intensity. In 2016, 12% of flooding related crashes occurred in Louisiana. During 2010–2016, flooding resulted in a total of 449 crashes in Louisiana with a total of 22 fatalities. This study collected seven years (2010–2016) of flooding related crash data to identify the key contributing factors. The findings show that two-way roadway with separation, rear-end crashes, higher average precipitation, and driver violations are frequently seen in the generated rules. This study provides a comprehensive picture of flooding related to traffic crashes; thus, targeting suitable countermeasures can be devised to reduce the number of crashes.

KEYWORDS

Association rules; countermeasures; flood-related crashes; rules mining

1. Introduction

On August 11, 2016, southern Louisiana experienced the heaviest concentration of rainfall. Baton Rouge and surrounded areas received rainfall more than 24 inches. Many roadways including the interstates (I-10 and I-12) in Baton Rouge, were submerged and closed for few days. In 2016, there were 15 fatalities involved with flooding with a total of 194 traffic crashes (National Weather Service, 2019). During 2010-2017, a total of 449 crashes occurred in Louisiana due to flooding with a total of 22 fatalities (Highway Safety Research Group, 2018). Flooding related crash events in Louisiana have exposed the vulnerability of transport networks. People might

CONTACT Subasish Das s-das@tti.tamu.edu Texas A&M Transportation Institute, Texas A&M University, College Station, TX, USA.

attempt to cross a flooded roadway not realizing its depth or, especially at night during heavy storms, when it might be difficult to see that a road is flooded. The geography of Louisiana allows the state to receive both frontal tropical hurricanes and large air masses that can bring in air moistures in almost any direction and drop rain with heavy intensity. A recent report measured that the monetary loss due to the 2016 flooding was $8.733 million (Terrell, 2017). The current research gap suggests a need to examine the key contributors of flooding-related crashes from historical crash data to exercise caution to avoid flooded areas when driving in those locations.

Traditional crash databases do not provide a filtering option when classifying flooding-related crashes. The study gathered seven years (2010–2016) of crash data from Louisiana to identify the trends of the key contributing factors. In many studies, regression models are used extensively to estimate traffic crashes. It is crucial to note that regression models typically ignore any cluster effect and evaluate the mean effect of the contributing variables. Consequently, interventions are usually tailored toward the mean effect in the absence of any subgroup effect consideration. Due to the lack of interpretability, these models are not valuable for practitioners. The present study uses the flooding-related crash data in Louisiana to exhibit the applicability of rule-based modeling techniques that can mitigate the clustering effects with heterogenous profiles without considering any previous assumptions. The results of this study can aid in the reduction of flooding related crashes and provide direction for the development of appropriate countermeasure.

2. Literature review

The current literature search showed that the state-of-the-art flooding related studies can be divided into four major sub-groups: (1) human factor related studies, (2) studies focused on roadway and crash characteristics, (3) geomorphology analysis, and (4) potential countermeasures for reducing flood-related crashes. As human perception error can contribute to flooding related crashes, several studies focused on human factors and their association with crash occurrence. Roadway geometry and environment are also key factors in traffic crashes, particularly flooding associated crashes. Some studies examined the geomorphology of roadbeds to determine the impact of flooding related crashes. Other studies focused on innovative countermeasures to reduce these crashes.

2.1. Human factors

Traffic crash is a key safety concern. It is argued that around 93% of road traffic crashes happen due to behavioral factors. The key behavioral factors include decreasing driving capability on a long-term basis, decreasing driving capability on a short-term basis, risk-taking behavior with long-term impact, and risk-taking behavior with short-term impact. Considerably, males are more prone to flood related traffic crash fatalities. Chen et al. (Chen, Baker, Braver, & Li, 2000) stated that the presence of a male (either driver or passenger) increases the per capita fatality rate. Rappaport (Rappaport, 2000), who amassed a database of tropical cyclone fatality rate using the newspaper reporting on crashes, also found a high percentage of male fatalities. It is argued that most drivers fail to perceive the dangers associated with flooded areas while driving (Maria Pregnolato, Ford, Wilkinson, & Dawson, 2017).

Studies conclude that the consumption of alcohol while driving can lead to major crashes. Drobot et al. (Drobot, Benight, & Gruntfest, 2007) showed that impaired drivers tend to drive onto flooded highways due to failure of perceived knowledge regarding the potential dangers. Proactive prevention actions can help in reducing many fatalities related to driving through flooded areas. Peoples' decision on crossing a flooded area is impacted based on both their combined impact and risk perception of the surroundings.

2.2. Crash and geometric characteristics

Many studies investigated inclement weather condition as a key contributing factor in road traffic crash occurrences (Mannering & Bhat, 2014; Peng, Abdel-Aty, Shi, & Yu, 2017; Tamerius, Zhou, Mantilla, & Greenfield-Huitt, 2016). Studies have examined the safety concerns of different weather-related factors such as temperature (Yu, Abdel-Aty, & Ahmed, 2013), precipitation (Black, Villarini, & Mote, 2017; Jung, Qin, & Noyce, 2010; Tamerius et al., 2016), fog and smog (Kouchaki, Roshani, Prozzi, & Bernardo, 2017). The effect of weather condition is associated with compounding factors including lower roadway friction, and visibility issues (Black et al., 2017; Kouchaki et al., 2017; Tamerius et al., 2016).

French et al. (French, Ing, Allmen, & Wood, 1983) found that most of the flash floods occurred during the warmer season (July–September). In the U.S., flood-related mortalities have been an important topic of several studies. Vehicle-associated deaths were the leading cause in both studies conducted by Mooney (Mooney, 1983) and French et al. (Ashley & Ashley, 2008), constituting 49 percent and 42 percent of all logged fatalities respectively. Studies showed that flood is the leading cause of drowning related deaths worldwide (Ashley & Ashley, 2008; Berz et al., 2001). Research suggests that driving through floodwaters is a common safety concern in the flood prone localities; however, little is known about the danger issues of motor vehicle-related drowning (Yale, Cole, Garrison, Runyan, & Ruback, 2003).

2.3. Geomorphology

Geomorphologic and hydraulic analysis in the Edwards Plateau provided enhanced understanding to bedload transport and deposition at LWC (Yale et al., 2003). Arnaud-Fassetta et al. (2009) focused on the contribution of fluvial geomorphology to flood management. This study detailed how fluvial geomorphology can present innovative approaches to flood prevention, river maintenance and floodplain restoration.

In their book, Baker, Kochel, and Patton (1988) analyzed global causes, effects and dynamics of floods and includes methods for related environmental management. The book describes the use of morphometric parameters of drainage basins and establishes typical procedures for calculating geomorphically important attributes following a major flood. In his paper, Baker (1994) noted that floods are the most globally pervasive, environmentally diverse and continually destructive of all natural hazards and commented that geomorphological flood studies are needed as a complement to conventional hydrological approaches. Snead, McCulloh, & Heinrich (2019) provided detailed contexts of Louisiana's widespread riverine and coastal floodplain, which experience episodic stream flooding and hurricane storm surges.

2.4. Potential countermeasures

To reduce flooding related crashes, several countermeasures can be proposed such as indicators along the verge of hardened surfaces, warning signs to the approach of a low water crossing or frequently flooded locations, and water depth gauges. Many drivers have issues in understanding the depth and speed of water over the roadway and enter the flooded road. In the nonexistence of any helpful signs, drivers set their standards to determine whether the road is drivable, which may lead to loss of property and life. For the safety of the commuters in the flooded areas, it is suggested to use one regulatory two warning sign in advance of LWC. Barricades and signs can be vague because they are not usually present at all crossings and can remain in position when the location is dry. For example, Ashley & Ashley (2008) showed that future structural modifications of flood control designs are not significantly associated with fatality reduction.

It was evident that the trust of information is an important aspect in inspiring drivers to avoid flooded roadways. Many fatalities related to floods can be evaded. In Queensland, Australia, strategy makers adopted a campaign with the slogan "If it's Flooded, Forget it" after January 2011 floods. However, vehicle-related flooding losses are still high. To reverse this trend, it is always necessary to create empirical evidence on significant associated factors that guide precautions while entering a flooded zone (Clarkin, Keller, Warhol, & Hixson, 2006). To better understand individual behavior, innovative analytical models will help in the growth of more effective interference programs (Ajzen, 1991; Michie & Johnston, 2012; Sorock, Ranney, & Lehto, 1996) to battle risky driving behavior.

Meeting road management objectives while fulfilling site-specific biological and geomorphic goals requires a truly interdisciplinary approach in which a biologist and hydro geomorphologist work with the design engineer. A strong structure must integrate the engineering requirements with hydrologic and biological factors. Between the years of 2005–2011, at least four concrete-cable LWCs were installed across Camp Atterbury Joint Maneuver Training Center or CAJMTC (Clarkin, Keller, Warhol, & Hixson, 2006). Most crossings were orientated using the existing trail path. With respect to the stream, the orientations are now under consideration as LWCs have been found as sources of sediment deposition, especially as vehicles traverse.

The literature review showed that human perception issues on the risk of flood water depth contributed to flood involved crashes. Lower roadway friction and visibility issues are found as the key geometric and crash characteristics in several relevant studies. Few studies showed the positive effectiveness of innovative countermeasures to reduce flooding associated crashes. The literature review indicated a need for an in-depth study of flooding related crashes. The current study collected data using natural language processing (NLP) and designed a study using association rules mining to achieve a better understanding of flooding related crashes.

3. Methodology

Data mining methods aim to extract undetermined knowledge patterns by exploring an extensive number of patterns. Data mining involves statistical learning, modeling concepts, algorithm capability, and efficient database management to identify patterns and trends without any prior assumptions and hypothesis design.

3.1. Overview of association rules mining

There is an abundance of earlier research on developing algorithms to solve the itemset associated problem. Agrawal and Srikant (1994) developed *a priori* algorithm to tally transactions and mine recurrent itemsets. The concept of *a priori* presumes that subsets of recurrent itemsets are frequent. The algorithm works by mining reoccurring itemsets or subsequences from a comprehensively large dataset. This process allows users to determine relations between various items. In recent years, multiple studies applied association rules mining techniques to identify the latent patterns in the complex crash databases (Das et al., 2018; Das, Kong, & Tsapakis, 2019; Das, Dutta, & Sun, 2019; Weng, Zhu, Yan, & Liu, 2016). The authors conferred Das et al. (2019) study to provide a short synopsis on the association rules mining.

By considering I= $\{i_1, i_2, \ldots, i_m\}$ as a set of items (for example, a set of crash categories for flood related crashes) and $C = \{c_1, c_2, \ldots, c_n\}$ as a set of database crash information or transaction where each crash record, c_i, contains a subset of items chosen from I, a k-itemset is a itemset with k items. An association rule can be written as Antecedent (A) → Consequent (B). Here, A and B are thought of as disjoint itemsets. The scales of significance are deduced by three parameters (support, confidence, and lift). The equations of support are listed below.

$$S(A) = \frac{\sigma(A)}{N} \tag{1}$$

$$S(B) = \frac{\sigma(B)}{N} \tag{2}$$

$$S(A \rightarrow B) = \frac{\sigma(A \cap B)}{N} \tag{3}$$

where,

$\sigma(A)$ = frequency of occurrences with A
$\sigma(B)$ = frequency of occurrences with B
$\sigma(A \cap B)$ = frequency of occurrences with both A and B
N = total frequency of occurrences
S(A) = support of A
S(B) = support of B
S(A → B) = support of the association rule (A → B)

Confidence is known as the measure of reliability for the inference of a generated rule. High confidence for a given A → B indicates that strong presence of B in transactions with having A. The lift depicts the association with the frequency and the expected frequency of co-occurrence of the antecedent-consequent.

$$C(A \rightarrow B) = \frac{S(A \rightarrow B)}{S(A)} \tag{4}$$

$$L(A \rightarrow B) = \frac{S(A \rightarrow B)}{S(A).S(B)} \tag{5}$$

where,

C(A → B) = Confidence of the association rule (A → B)

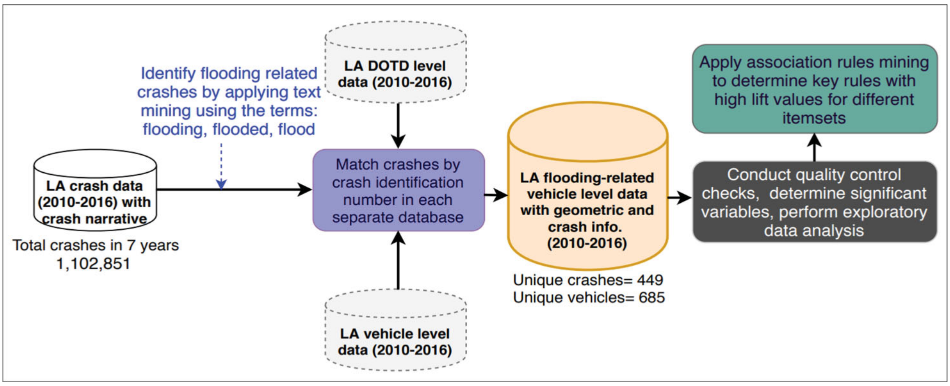

Figure 1. Flowchart of data integration.

$L(A \rightarrow B)$ = Lift of the association rule $(A \rightarrow B)$

While 'lift $< 1'$ signifies negative interdependence between A and B, a value larger than one signifies positive interdependence. A value of one indicates independence. Based on the number of consequents and antecedents, rules are defined as n-itemset rule. If there is one consequent and one antecedent, then it is a 2-product rule. It is significant to note that generated rules imply only association and not causation.

3.2. Data integration

The dataset of the current study is police-reported crashes in Louisiana from 2010 to 2016. Louisiana crash datasets also include crash narrative in a database format. There is no filter in the crash database that can identify flood-related crashes from the crash database. To identify the flooding-related crash database, we applied a text mining algorithm to distinguish the flooding related crashes. Three terms were used to identify the flooding-related crashes: *flood*, *flooding*, and *flooded*. After extracting these crashes, a manual approach of reading the crash reports was conducted to remove the redundant crashes. In seven years, 449 crashes in Louisiana occurred due to the flooding. From 2015 to 2016, the state of Louisiana experienced a 234% increase in reported flooding-related crashes. Prolonged rainfall in August 2016 resulted in catastrophic flooding for Louisiana in which thousands of houses and businesses were submerged. Many rivers and waterways overflooded, and rainfall exceeded 20 inches in multiple parishes. The flowchart of data integration and analysis is shown in Figure 1.

A majority of the parishes of District 3, 61, and 62 were designated as federal disaster areas by the Federal Emergency Management Agency (FEMA) in the aftermath of the floods. Figure 2 illustrates the heatmap of flood-related crashes by the Louisiana Department of Transportation and

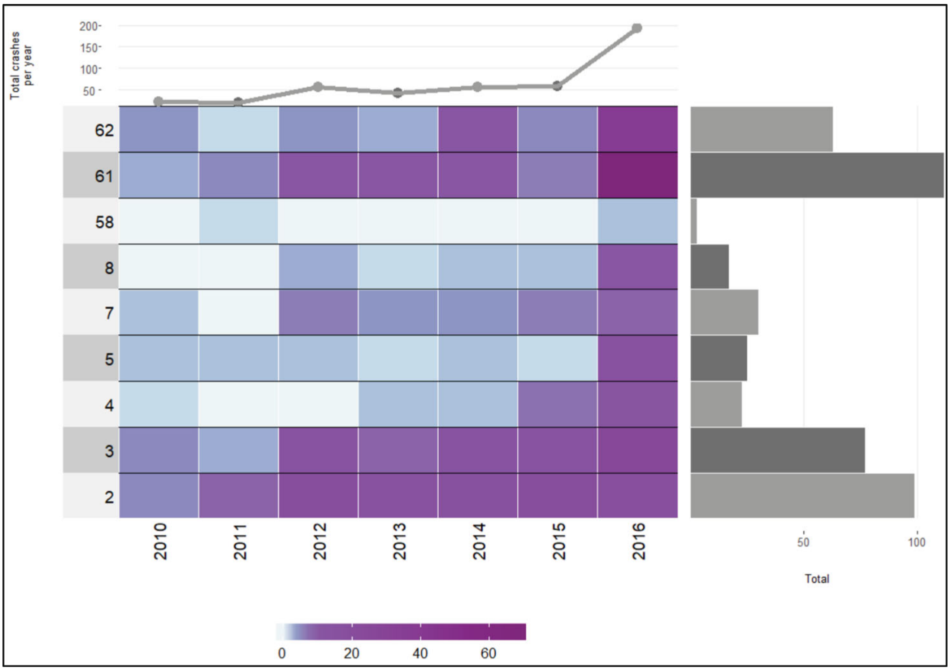

Figure 2. Flood-related crashes by DOTD districts.

Development (DOTD) Districts and year. Both East and West Baton Rouge were in District 61. District 61 has been ranked one due to the high number of flooding-related crashes. The number of crashes by year are illustrated in the top section of Figure 2. The line chart clearly indicated the sudden rise of flooding-related crashes in Louisiana in 2016.

The NLP or text mining algorithm examined to identify the level of water depth from the crash data. Majority of the crash narratives lack information about water depth of the flooding related crashes. For example, the following crash report, tagged as flood related crash, does not provide sufficient information about water depth: "*Vehicle #1 driver states that, during a severe rain with street flooding, she was driving west on Government and attempted to make a left turn onto St. Phillips St. She told me that her car was turning left and then started sliding sideways. She told me that she could not see out of her windshield very well and then suddenly hit a fire hydrant. Vehicle #1 was driving too fast for conditions. no damage was observed to the fire hydrant.*"

For flooding related crashes, it is also important to have additional information about the rainfall intensity. The rainfall data from National Oceanic and Atmospheric Administration (NOAA) usually provides information on a broad geographic area or area surrounding the weather stations.

Table 1 exhibits the attributes leading to road traffic crashes and the frequency distribution among each characteristic. Attributes such as age,

Table 1. Descriptive statistics.

Intersection (Int)		Season (Season)	
Attribute	Percent	Attribute	Percent
No	68.76	Fall	24.82
Yes	31.24	Spring	27.59
Posted Speed Limit (PSL)		Summer	35.18
20 or less (mph)	11.68	Winter	12.41
25–35 (mph)	44.23	**Driver Condition (Cond)**	
40–50 (mph)	30.51	Normal	52.41
55–65 (mph)	11.82	Inattentive	26.72
70 or above (mph)	1.75	Impaired	2.63
Collision Type (CollTyp)		Others	18.25
Rear End	31.68	**Driver Age (Age)**	
Single	23.94	15–25	20.88
Sideswipe	14.74	25–35	24.38
Right Angle	7.01	35–45	20.58
Head-On	3.07	45–55	14.01
Turning	4.82	55–65	13.43
Others	14.74	65–75	5.69
Major Contributing Factor (Cont)		$>$ 74	1.02
Violations	61.02	**Vehicle Type (VehTyp)**	
Movement Prior To Crash	17.08	Car	46.42
Roadway Condition	10.07	Lt. Truck	24.53
Condition of Driver	4.82	Suv	15.47
Weather	7.01	Truck/Bus	6.57
Roadway Type (RdTyp)		Van	2.63
2-way no Sep.	68.18	Others	4.38
2-way with Sep	20.73	**Average Precipitation (Avg_Precip)**	
2-way with Barrier	3.21	35–45 in.	0.87
Others	7.88	46–55 in.	8.32
Severity (Seve)		56–60 in.	9.63
K (Fatal)	0.45	61–63 in.	56.93
ABC (Injury)	20.58	64–65 in.	13.72
PDO (No Injury)	78.98	$>$ 65 in.	10.51

Note: [1]KABC indicates Fatal and Injury Crashes (K = Fatal, A = Incapacitating Injury, B = Non-Incapacitating Injury, C = Minor Injury);
[2]PDO (Property Damage Only) indicates No injury.

vehicle type, season, and more are listed below. Approximately 70 percent of the flooding crashes are segment related. Areas with posted speed limits (PSL) from 25 to 50 mph are more likely to be involved in a road traffic crash than areas with PSL of 20 mph or less and 55 mph or more. Summer months experience more crashes than any other season. As winter has less flood related events, the seasonal distribution shows lower number of crashes during winter. Rear-end crashes have the highest frequency in flooding related crashes. Two-lane roadways with no physical separation display high frequencies in flooding related crashes. It is important to note that some events of interest in traffic safety analysis occur rarely (for example, the occurrence of a fatal crash). For example, due to the lower frequency in the overall crash data, the support for some rules associated with fatal crashes could be quite low. Keeping the support threshold very low makes the number of rules to be enormously large. According to the Highway Safety Manual (HSM), crash modeling can be conducted based on major severity types (for example, fatal and injury crashes can be defined

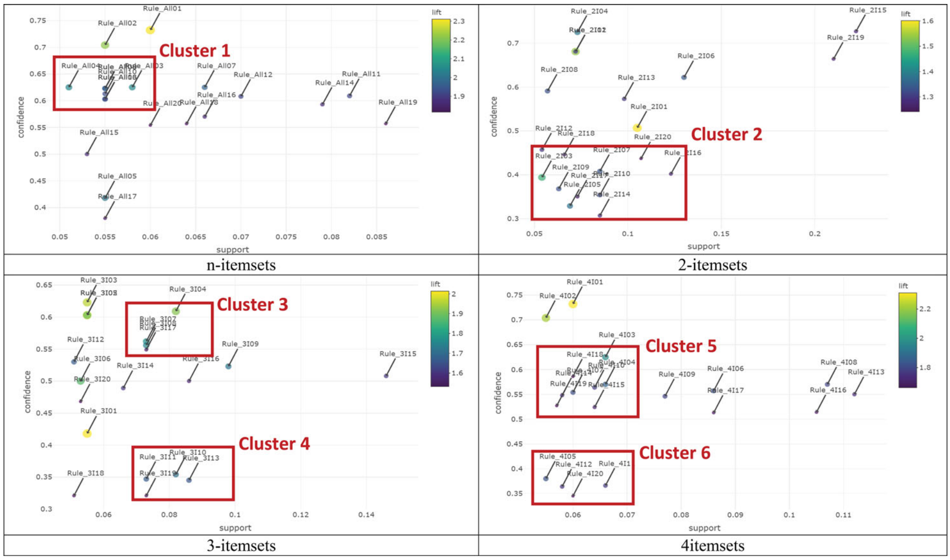

Figure 3. Scatter plots based on support, confidence, and lift values.

as KABC crash or non-PDO crash). This study divided the crash severity into three major groups: fatal (K), injury (ABC) and no-injury (PDO) crashes. The research team collected precipitation data on the roadway segment on the crash occurrence season from the NOAA. The annual average precipitation on these roadways is calculated for the final dataset. Approximately 80 percent of these roadways experience over 60 inches annual average precipitation.

4. Results and discussions

Association rules mining is capable of managing large and complex data. Any prior assumption is unnecessary. The present study used 'a priori' algorithm to conduct the analysis. The open-source R Software package 'arules' was used to perform the analysis (Hahsler, Grün, & Hornik, 2015). For intuitive and effective rules, determining the threshold of confidence and support values is crucial. Based on a recommendation of a study, we set multiple trail thresholds to minimum values of support and confidence (Das et al., 2019). This particular approach has been embraced by other studies (Das et al., 2017, 2019). The minimum support for this study is considered to be 0.05, and after numerous trial and errors, the minimum confidence was determined to be 0.30.

The research team generated 'id numbers' to define the rules. As this study developed a unique graphics (Figure 3) based on the measures (support, confidence, and lift), it is important to replace the clutters of texts in the rules by providing 'id numbers' for the rules. In the following section, the research team used 'Rule_All' for n-itemsets, 'Rule_2I' for 2-itemsets, 'Rule_3I' for 3-temsets, and 'Rule-4I' for 4-itemsets.

Table 2 depicts the variables considered as antecedent-consequent rules for n-itemsets that contribute to road traffic crashes. The rules and their corresponding antecedent-consequent were ordered according to decreasing lift value. Each rule has a lift value greater than 1, indicating positive interdependence between the antecedent and the consequent. In addition to the lift value, each antecedent lists their consequent, support, confidence, and count. Table 2 has 2-itemset rules to 5-itemset rules. Two attributes *CollTyp = Rear End* and *Cont = Violations* are present in majority of the rules. The other attributes that are presented in multiple times in the rules are: *Avg_Precip = 61-63 in* (in 10 rules), *Cond = Norma*l (in 10 rules), *RdTyp = 2-way with Sep* (in 9 rules), and *Int = No* (in 9 rules). Other studies (Abdel-Aty, Ekram, Huang, & Choi, 2011; Tamerius et al., 2016; Yu et al., 2013) showed similar findings as they found the significant effects of precipitation on flooding related crashes. The roadway with physical separation usually experiences stalled water from rainfall. The rule with highest lift value is Rule_All01: *Cont = Violations, RdTyp = 2-way with Sep, Int = No→ CollTyp = Rear End* (Support = 0.060, Confidence = 0.732, Lift = 2.311). The explanation of the first rule is: 6.0% of flooding related rear-end segment crashes occurred on two-way roadways segments (Int = No means not at intersection) with separation with major contribution factor as driver violation; out of all flooding related crashes with the antecedent (*Cont = Violations, RdTyp = 2-way with Sep, Int = No*) 73.2% crashes occur when the collision type is rear-end; the proportion of rear-end flooding related crashes (with violation as the contributing factor) on rural two-lane roadway segments with separation was 2.311 times the proportion of flooding related rear-end crashes. As PDO crashes are dominating in flooding related crashes, majority of the rules with severity are associated with PDO. One rule (**Rule_All14:** *Cont = Violations, Int = No, Cond = Normal, Seve = Injury→ CollTyp = Rear End)* shows injury as an attribute in the rule.

Table 3 depicts the variables considered as antecedents that contribute to road traffic crashes. A rule is defined as a 2-item rule if there are a single antecedent and single consequent. The most frequent attributes are: *Age = 15-24* (in 3 rules), *PSL = 25-35 (*in 3 rules), *PSL = 40-50 (*in 3 rules), and *Cond = Normal* (in 3 rules). Different age groups, posted speed limits, and average precipitation measures are frequently seen in 2-itemset rules. Rule_2I15 (*PSL = 40-50 → Avg_Precip = 61-63 in*) occurred the most often with a count of 152 in the top 20 2-itemset rules.

Table 4 depicts the variables considered as antecedents that contribute to road traffic crashes. A 3-item rule indicates that there are two antecedents

Table 2. Top 20-rules from N-Itemsets.

Rule No.	Antecedent	Consequent	Supp.	Conf.	Lift	Count
Rule_All01	Cont = Violations, RdTyp = 2-way with Sep, Int = No	CollTyp = Rear End	0.060	0.732	2.311	41
Rule_All02	Cont = Violations, RdTyp = 2-way with Sep, Cond = Normal	CollTyp = Rear End	0.055	0.704	2.221	38
Rule_All03	CollTyp = Rear End, Int = No, Cond = Normal, Avg_Precip = 61–63 in	PSL = 40–50	0.058	0.625	2.048	40
Rule_All04	Season = Summer, Int = No, Cond = Normal, Avg_Precip = 61–63 in	PSL = 40–50	0.051	0.625	2.048	35
Rule_All05	CollTyp = Rear End, Season = Summer	RdTyp = 2-way with Sep	0.055	0.418	2.014	38
Rule_All06	RdTyp = 2-way with Sep, Avg_Precip = 61–63 in	PSL = 40–50	0.055	0.603	1.977	38
Rule_All07	Cont = Violations, RdTyp = 2-way with Sep, Seve = PDO	CollTyp = Rear End	0.066	0.625	1.973	45
Rule_All08	RdTyp = 2-way with Sep, Season = Summer	CollTyp = Rear End	0.055	0.623	1.966	38
Rule_All09	Cont = Violations, Int = No, Cond = Normal, Seve = PDO, Avg_Precip = 61–63 in	CollTyp = Rear End	0.055	0.623	1.966	38
Rule_All10	Cont = Violations, Int = No, Cond = Normal, VehTyp = Car	CollTyp = Rear End	0.055	0.613	1.935	38
Rule_All11	Cont = Violations, RdTyp = 2-way with Sep	CollTyp = Rear End	0.082	0.609	1.921	56
Rule_All12	Cont = Violations, Int = No, Cond = Normal, Avg_Precip = 61–63 in	CollTyp = Rear End	0.070	0.608	1.918	48
Rule_All13	RdTyp = 2-way with Sep, Avg_Precip = 61–63 in	CollTyp = Rear End	0.055	0.603	1.904	38
Rule_All14	Cont = Violations, Int = No, Cond = Normal, Seve = Injury	CollTyp = Rear End	0.079	0.593	1.873	54
Rule_All15	CollTyp = Single, Cont = Violations	Cond = Inattentive	0.053	0.500	1.872	36
Rule_All16	CollTyp = Rear End, Cond = Normal, Avg_Precip = 61–63 in	PSL = 40–50	0.066	0.570	1.867	45
Rule_All17	CollTyp = Rear End, Cont = Violations, Cond = Normal	RdTyp = 2-way with Sep	0.055	0.380	1.833	38
Rule_All18	Cont = Violations, Int = No, Cond = Normal, Avg_Precip = 61–63 in	PSL = 40–50	0.064	0.557	1.825	44
Rule_All19	CollTyp = Rear End, Int = No, Avg_Precip = 61–63 in	PSL = 40–50	0.086	0.557	1.824	59
Rule_All20	Cont = Violations, Cond = Inattentive, Avg_Precip = 61–63 in	PSL = 40–50	0.060	0.554	1.816	41

Table 3. Top 20-rules from 2-Itemsets.

Rule No.	Antecedent	Consequent	Supp.	Conf.	Lift	Count
Rule_2I01	RdTyp = 2-way with Sep	CollTyp = Rear End	0.105	0.507	1.601	72
Rule_2I02	Avg_Precip= > 65 in.	PSL = 25–35	0.072	0.681	1.539	49
Rule_2I03	Avg_Precip = 64–65 in.	Cond = Inattentive	0.054	0.394	1.473	37
Rule_2I04	Cont = Roadway Condition	Cond = Normal	0.073	0.725	1.383	50
Rule_2I05	Age = 15–24	CollTyp = Single	0.069	0.329	1.373	47
Rule_2I06	Age = 15–24	VehTyp = Car	0.130	0.622	1.341	89
Rule_2I07	RdTyp = 2-way with Sep	PSL = 40–50	0.085	0.408	1.339	58
Rule_2I08	Avg_Precip = 56–60 in	PSL = 25–35	0.057	0.591	1.336	39
Rule_2I09	Cont = Movement Prior To Crash	Season = Spring	0.063	0.368	1.332	43
Rule_2I10	CollTyp = Single	Cond = Inattentive	0.085	0.354	1.324	58
Rule_2I11	Avg_Precip= > 65 in.	Cond = Normal	0.072	0.681	1.299	49
Rule_2I12	PSL = 55–65	Season = Summer	0.054	0.457	1.298	37
Rule_2I13	Cont = Movement Prior To Crash	PSL = 25–35	0.098	0.573	1.295	67
Rule_2I14	Season = Spring	CollTyp = Single	0.085	0.307	1.282	58
Rule_2I15	PSL = 40–50	Avg_Precip = 61–63 in	0.222	0.727	1.277	152
Rule_2I16	PSL = 40–50	CollTyp = Rear End	0.123	0.402	1.269	84
Rule_2I17	Age = 15–24	Season = Spring	0.073	0.350	1.267	50
Rule_2I18	CollTyp = Sideswipe	Season = Summer	0.066	0.446	1.266	45
Rule_2I19	CollTyp = Rear End	Cond = Normal	0.210	0.664	1.266	144
Rule_2I20	Age = 25–34	Season = Summer	0.107	0.437	1.242	73

and one consequent or one antecedent and two consequents. ***Rule_3I15:*** *Cont = Violations, Cond = Normal → CollTyp = Rear End* occurred the most often with a count of 100. Two attributes *CollTyp = Rear End* and *RdTyp = 2-way with Sep* are present in majority of the rules. The other attributes that are presented in multiple times in the rules are: *Cont = Violations* (in 5 rules), *PSL = 40-50* (in 4 rules).

Table 5 depicts the variables considered as antecedents that contribute to road traffic crashes. The frequency of each rule was evenly distributed in the 4-itemset. The attributes that are present in multiples rules are: *CollTyp = Rear End* (in 15 rules), *Cont = Violations* (in 12 rules), *Cond = Normal* (in 11 rules), *RdTyp = 2-way with Sep* (in 9 rules), *Avg_Precip = 61-63 in* (in 8 rules), *PSL = 40-50* (in 8 rules), and *Int = No* (in 6 rules).

Visualization of association rules is challenging because of the high number of rules and the generated texts for the rules. Figure 3 shows the scatter plots of the top twenty rules based on support, confidence, and lift for different itemset groups. This plot provides an overall trend of the generated top rules and the clustering patterns of the rules based on the parameter. The size and color of the circles indicate the value of the lift. The key insights from this figure are the following:

- In all itemsets (n-itemsets, 2-itemsets, 3-itemsets, and 4-itemsets), the rule with the highest lift has two attributes in common: *CollTyp = Rear End*, and *RdTyp = 2-way with Sep*.
- A limited number of rules have high support values (greater than 0.11). Higher support value is associated with higher proportion of the

Table 4. Top 20-rules from 3-Itemsets.

Rule No.	Antecedent	Consequent	Supp.	Conf.	Lift	Count
Rule_3I01	CollTyp = Rear End, Season = Summer	RdTyp = 2-way with Sep	0.055	0.418	2.014	38
Rule_3I02	RdTyp = 2-way with Sep, Avg_Precip = 61–63 in	PSL = 40–50	0.055	0.603	1.977	38
Rule_3I03	RdTyp = 2-way no Sep, Season = Summer	CollTyp = Rear End	0.055	0.623	1.966	38
Rule_3I04	Cont = Violations, RdTyp = 2-way with Sep	CollTyp = Rear End	0.082	0.609	1.921	56
Rule_3I05	RdTyp = 2-way with Sep, Avg_Precip = 61–63 in	CollTyp = Rear End	0.055	0.603	1.904	38
Rule_3I06	CollTyp = Single, Cont = Violations	Cond = Inattentive	0.053	0.500	1.872	36
Rule_3I07	RdTyp = 2-way with Sep, Int = No	CollTyp = Rear End	0.073	0.562	1.773	50
Rule_3I08	RdTyp = 2-way with Sep, Cond = Normal	CollTyp = Rear End	0.073	0.556	1.754	50
Rule_3I09	CollTyp = Rear End, Avg_Precip = 61–63 in	PSL = 40–50	0.098	0.523	1.716	67
Rule_3I10	CollTyp = Rear End, Cont = Violations	RdTyp = 2-way with Sep	0.082	0.354	1.710	56
Rule_3I11	CollTyp = Rear End, Cond = Normal	RdTyp = 2-way with Sep	0.073	0.347	1.675	50
Rule_3I12	RdTyp = 2-way with Sep, VehTyp = Car	CollTyp = Rear End	0.051	0.530	1.674	35
Rule_3I13	CollTyp = Rear End, Seve = PDO	RdTyp = 2-way with Sep	0.086	0.345	1.664	59
Rule_3I14	Cont = Violations, RdTyp = 2-way with Sep	PSL = 40–50	0.066	0.489	1.603	45
Rule_3I15	Cont = Violations, Cond = Normal	CollTyp = Rear End	0.146	0.508	1.602	100
Rule_3I16	RdTyp = 2-way with Sep, Seve = Injury	CollTyp = Rear End	0.086	0.500	1.578	59
Rule_3I17	Age = 25-34, Avg_Precip = 61–63 in	Season = Summer	0.073	0.549	1.562	50
Rule_3I18	CollTyp = Rear End, VehTyp = Car	RdTyp = 2-way with Sep	0.051	0.321	1.549	35
Rule_3I19	CollTyp = Rear End, Int = No	RdTyp = 2-way with Sep	0.073	0.321	1.546	50
Rule_3I20	Cont = Violations, Age = 15–24	PSL = 40–50	0.053	0.468	1.532	36

Table 5. Top 20-rules from 4-Itemsets.

Rule No.	Antecedent	Consequent	Supp.	Conf.	Lift	Count
Rule_4I01	Cont = Violations, RdTyp = 2-way with Sep, Int = No	CollTyp = Rear End	0.060	0.732	2.311	41
Rule_4I02	Cont = Violations, RdTyp = 2-way with Sep, Cond = Normal	CollTyp = Rear End	0.055	0.704	2.221	38
Rule_4I03	Cont = Violations, RdTyp = 2-way no Sep, Seve = PDO	CollTyp = Rear End	0.066	0.625	1.973	45
Rule_4I04	CollTyp = Rear End, Cond = Normal, Avg_Precip = 61–63 in	PSL = 40–50	0.066	0.570	1.867	45
Rule_4I05	CollTyp = Rear End, Cont = Violations, Cond = Normal	RdTyp = 2-way with Sep	0.055	0.380	1.833	38
Rule_4I06	CollTyp = Rear End, Int = No, Avg_Precip = 61–63 in	PSL = 40–50	0.086	0.557	1.824	59
Rule_4I07	Cont = Violations, Cond = Inattentive, Avg_Precip = 61–63 in	PSL = 40–50	0.060	0.554	1.816	41
Rule_4I08	Cont = Violations, Int = No, Cond = Normal	CollTyp = Rear End	0.107	0.570	1.800	73
Rule_4I09	CollTyp = Rear End, Cont = Violations, Avg_Precip = 61–63 in	PSL = 40–50	0.077	0.546	1.791	53
Rule_4I10	Cont = Violations, Cond = Normal, PSL = 40–50	CollTyp = Rear End	0.064	0.564	1.781	44
Rule_4I11	CollTyp = Rear End, Cont = Violations, Seve = PDO	RdTyp = 2-way with Sep	0.066	0.366	1.765	45
Rule_4I12	CollTyp = Rear End, Cond = Normal, Seve = PDO	RdTyp = 2-way with Sep	0.058	0.364	1.754	40
Rule_4I13	Cont = Violations, Cond = Normal, Seve = Injury	CollTyp = Rear End	0.112	0.550	1.736	77
Rule_4I14	RdTyp = 2-way with Sep, Cond = Normal, Seve = PDO	CollTyp = Rear End	0.058	0.548	1.730	40
Rule_4I15	Season = Summer, Cond = Normal, Avg_Precip = 61–63 in	PSL = 40–50	0.064	0.524	1.717	44
Rule_4I16	Int = No, Cond = Normal, Avg_Precip = 61–63 in	PSL = 40–50	0.105	0.514	1.686	72
Rule_4I17	Cont = Violations, Cond = Normal, Avg_Precip = 61–63 in	PSL = 40–50	0.086	0.513	1.682	59
Rule_4I18	RdTyp = 2-way no Sep., Age = 25–34, Avg_Precip = 61–63 in	Season = Summer	0.060	0.586	1.665	41
Rule_4I19	RdTyp = 2-way with Sep, Int = No,S eve = PDO	CollTyp = Rear End	0.057	0.527	1.664	39
Rule_4I20	CollTyp = Rear End, Cont = Violations, Int = No	RdTyp = 2-way with Sep	0.060	0.345	1.662	41

attribute in the variable category. For example, *Cond = Normal* attribute is 52.41% of all attributes in driver condition (Cond) variable.

- Rule_2I15: PSL = 40-50→ Avg_Precip = 61-63 in
- Rule_2I19: CollTyp = Rear End→ Cond = Normal
- Rule_3I15: Cont = Violations, Cond = Normal→ CollTyp = Rear End
- Rule_2I06: Age = 15-24→ VehTyp = Car
- Rule_2I16: PSL = 40-50→ CollTyp = Rear End
- Rule_4I13: Cont = Violations, Cond = Normal, Seve = PDO→ CollTyp = Rear End

- Several clusters (shown as clusters) are visible in Figure 3. A common pattern of these clusters is the lower support value of the rules. The clusters are differentiated by the value of the confidence scores. For example, Cluster 1, Cluster 3, and Cluster 5 have higher confidence scores and Cluster 2, Cluster 4, and Cluster 6 have lower confidence scores. Another common pattern of these clustered rules is the low or medium score of lift values.

5. Conclusions

The current study provided a framework on how rules mining approach can determine key contributing factors that affect flooding involved crashes. Understanding the impact of flooding related crashes is crucial in identifying safety-related issues and countermeasures. One of the major contributions of this study is that more domain-specific patterns were developed in association with several factors, including geometric properties, driver characteristics, traffic and environmental factors associated with flood related crashes. This study developed several key rules that are frequent in the reported flood-related crash data. The findings can be better if the potential countermeasures can be evaluated based on the current findings. However, the current study is only limited to identify the patterns. From the generated rules, several attributes have been identified as the key contributing traits in flooding related crashes. Some of the key findings are below:

- Violation of traffic rules has been found in many of the rules. Other studies also show that many drivers underestimate the depth of the flooding water and speed of water over the roadway and enter the flooded road (Ashley & Ashley, 2008; Clarkin, Keller, Warhol, & Hixson, 2006; Maria Pregnolato et al., 2017). The researchers anticipate that safety campaigning can help in improving safety concerns among drivers. One study showed that safety campaigns such as *"If* *it's Flooded, Forget it*" helped in educating drivers about the dangers of driving in the flooding water (Clarkin, Keller, Warhol, & Hixson, 2006).

- The results show that rules with high lift values are associated rear-end crashes and 2-way roadway with separation. Two-way roadway with physical separation usually experiences stalled water from short-term rainfall. Due to the water on the roadway, vehicles fail to avoid rear-end crashes due to sudden stalling of a vehicle in front. Other studies showed that roadway marking and other retroreflective markers in the edge line can assist drivers in partially flooded areas (Ashley & Ashley, 2008; Clarkin, Keller, Warhol, & Hixson, 2006). Added investigation is needed to examine the presence of roadway signs in the two-lane roadways with high flooding related crashes.
- Another key finding is the higher presence of moderate posted speed limit (40–50 mph) roadways in flood related crash patterns.
- Majority of the severity of flood involved crashes are either PDO or injury. Fatal crashes represent only 0.45% of flooding related crashes.

This research has some limitations. One major limitation is that the current study relied on the NLP tools to select flooding related crashes. Additional tools were not used to determine flooding crashes from crash data associated with blank or insufficient crash narratives and reporting errors in the crash narratives. The second limitation is the absence of flooding related information such as roadway friction, and water depth. The current exploration examined the plausibility of using these variables. Due to the data limitation, these variables were not used in the analysis. Further research is needed to enhance current model performance and better investigate a comprehensive list of rules with added variables of interests. Limitations of the current study offer directions for future research in reducing flooding related crashes.

Disclaimer

The contents of this paper reflect the views of the authors and not the official views or policies of the Louisiana Department of Transportation & Development (LADOTD).

Acknowledgments

We like to thank two anonymous reviewers who provided critical and rigorous comments. The current version of the paper is much improved due to the suggested changes that are made to respond to the queries of the reviewers.

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753
Xiaoduan Sun http://orcid.org/0000-0001-7282-1340

References

Abdel-Aty, M., Ekram, A. A., Huang, H., & Choi, K. (2011). A study on crashes related to visibility obstruction due to fog and smoke. *Accident Analysis & Prevention*, *43*, 1730–1737. doi:10.1016/j.aap.2011.04.003

Agrawal, R., & Srikant, R. (1994). *Fast Algorithms for Mining Association Rules*. Proceedings of the 20th VLDB Conference, Santiago, Chile.

Ajzen, I. (1991). The theory of planned behaviour. *Organizational* Behavior and Human Decision Processes, *50*(2), 179–211. doi:10.1016/0749-5978(91)90020-T

Arnaud-Fassetta, G., Astrade, L., Bardou, É., Corbonnois, J., Delahaye, D., Fort, M., Gautier, E., Jacob, N., Peiry, J., Piégay, H., & Penven, M. (2009). Fluvial geomorphology and flood-risk management. *Géomorphologie :* Relief, Processus, Environnement, *15*(2), 109–128. doi:10.4000/geomorphologie.7554

Ashley, S. T., & Ashley, W. S. (2008). Flood fatalities in the United States. *Journal of Applied Meteorology and Climatology*, *47*(3), 805–818. doi:10.1175/2007JAMC1611.1

Baker, V. (1994). Geomorphological understanding of floods. *Geomorphology*, *10*(1-4), 139–156. doi:10.1016/0169-555X(94)90013-2

Baker, V., Kochel, R., & Patton, P. (1988). *Flood geomorphology* (1st ed.). New York, USA: John Wiley & Sons.

Berz, G., Kron, W., Loster, T., Rauch, E., Schimetschek, J., Schmieder, J., … Wirtz, A. (2001). World map of natural hazards – A global view of the distribution and intensity of significant exposures. *Natural Hazards*, *23*(2/3), 443–465.

Black, A. W., Villarini, G., & Mote, T. L. (2017). Effects of rainfall on vehicle crashes in Six U.S. States. *Weather,* Climate, and Society, *9*(1), 53–70. doi:10.1175/WCAS-D-16-0035.1

Chen, L., Baker, S. P., Braver, E. R., & Li, G. (2000). Carrying passengers as a risk factor for crashes fatal to 16 and 17-year-old drivers. *JAMA*, *283*(12), 1578–1582. doi:10.1001/jama.283.12.1578

Clarkin, K., Keller, G., Warhol, T., & Hixson, S. (2006). *Low-water crossings: Geomorphic, biological, and engineering design considerations*. Washington, D.C.: U.S. Department of Agriculture.

Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., & Jalayer, M. (2018). Supervised association rules mining on pedestrian crashes in urban areas: Identifying patterns for appropriate countermeasures. *International Journal of Urban Sciences*, *23*(1), 30–48. doi:10.1080/12265934.2018.1431146

Das, S., Dutta, A., Jalayer, M., Bibeka, A., & Wu, L. (2018). Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety. *International Journal of Transportation Science and Technology*, *7*(2), 114–123. doi:10.1016/j.ijtst.2018.02.001

Das, S., Kong, X., & Tsapakis, I. (2019). Hit and run crash analysis using association rules mining. *Journal of Transportation Safety & Security*, *0*(0), 1–20. doi:10.1080/19439962.2019.1611682

Das, S., Dutta, A., and Sun, X. (2019). Patterns of rainy weather crashes: Applying rules mining. *Journal of Transportation Safety & Security*. doi: 10.1080/19439962.2019.1572681

Drobot, S. D., Benight, C., & Gruntfest, E. C. (2007). Risk factors for driving into flooded roads. *Environmental* Hazards, *7*, 227–234. doi:10.1016/j.envhaz.2007.07.003

French, J. G., Ing, R., Allmen, S. V., & Wood, R. (1983). Mortality from flash floods: A review of the National Weather Service reports, 1969–81. *Public Health Reports*, *98*(6), 584–588.

Hahsler, M., Grün, B., & Hornik, K. (2015). Introduction to arules – A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software*, *14*(15), 1–25.

Highway Safety Research Group. (2018). Louisiana Crash Data Reports. Retrieved from http://datareports.lsu.edu/Reports/SpecialReports/2017_Factbook.pdf.

Jung, S., Qin, X., & Noyce, D. A. (2010). Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis & Prevention*, *42*, 213–224. doi:10.1016/j.aap.2009.07.020

Kouchaki, S., Roshani, H., Prozzi, J. A., & Bernardo, J. H. (2017). Evaluation of aggregates surface micro-texture using spectral analysis. *Construction and Building Materials*, *156*, 944–955. doi:10.1016/j.conbuildmat.2017.08.174

Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research*, *1*, 1–22. doi:10.1016/j.amar.2013.09.001

Maria Pregnolato, M., Ford, A., Wilkinson, S., & Dawson, R. (2017). The impact of flooding on road transport: A depth-disruption function. *Transportation Research Part D: Transport and Environment*, *55*, 67–81. Volume doi:10.1016/j.trd.2017.06.020

Michie, S., & Johnston, M. M. (2012). Theories and techniques of behavior change: Developing a cumulative science of behavior change. *Health Psychology Review*, *6*(1), 1–6. doi:10.1080/17437199.2012.654964

Mooney, L. E. (1983). Applications and implications of fatality statistics to the flash flood problem. Fifth Conference on Hydrometeorology, American Meteorological Society, 127–129.

National Weather Service (NWS). (2019). NWS Preliminary US Flood Fatality Statistics. Retrieved from https://www.weather.gov/arx/usflood.

Peng, Y., Abdel-Aty, M., Shi, Q., & Yu, R. (2017). Assessing the impact of reduced visibility on traffic crash risk using microscopic data and surrogate safety measures. *Transportation Research Part C: Emerging Technologies*, *74*, 295–305. doi:10.1016/j.trc.2016.11.022

Rappaport, E. N. (2000). Loss of Life in the United States Associated with recent Atlantic tropical cyclones. *Bulletin of the American Meteorological Society*, *81*(9), 2065–2074. doi:10.1175/1520-0477(2000)081<2065:LOLITU>2.3.CO;2

Snead, J., McCulloh, R., & Heinrich, P. (2019). *Landforms of the Louisiana coastal plain*. Louisiana Geological Survey, Baton Rouge, Louisiana, USA.

Sorock, G. S., Ranney, T. A., & Lehto, M. R. (1996). Motor vehicle crashes in roadway construction workzones: An analysis using narrative text from insurance claims. *Accident Analysis & Prevention*, *28*(1), 131–138. doi:10.1016/0001-4575(95)00055-0

Tamerius, J. D., Zhou, X., Mantilla, R., & Greenfield-Huitt, T. (2016). Precipitation effects on motor vehicle crashes vary by space, time, and environmental conditions. *Weather, Climate, and Society*, *8*(4), 399–407. doi:10.1175/WCAS-D-16-0009.1

Terrell, D. (2017). *The Economic Impact of the August 2016 Floods on the State of Louisiana*. Baton Rouge, Louisiana, USA: L. T. and Associates, LLC.

Weng, J., Zhu, J. Z., Yan, X., & Liu, Z. (2016). Investigation of work zone crash casualty patterns using association rules. *Accident Analysis & Prevention*, *92*, 43–52. doi:10.1016/j.aap.2016.03.017

Yale, J. D., Cole, T. B., Garrison, H. G., Runyan, C. W., & Ruback, J. K. R. (2003). Motor vehicle-related drowning deaths associated with inland flooding after hurricane Floyd: A field investigation. *Traffic Injury Prevention*, *4*(4), 279–284. doi:10.1080/714040485

Yu, R., Abdel-Aty, M., & Ahmed, M. (2013). Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. *Accident Analysis & Prevention*, *50*, 371–376. doi:10.1016/j.aap.2012.05.011