Evaluation of Real World Scenarios Involving Toll Plazas Using Simulation

Kali Carroll, Mohamed Abdel-Aty, Qi Shi, Yina Wu, Qing Cai
Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, Florida

Overview

Toll plazas are becoming an essential part of the highway systems, especially within the state of Florida. A primary reason for many vehicle collisions happening at these facilities, is the fact that each toll plaza agency has different designs and even signs. This, in turn, causes driver confusion and possible last minute weaving. Even though the varying design of toll plazas is a clear highway safety factor, research in the field is very limited but expanding. This study focuses on one toll plaza in particular, the Dean Mainline Toll Plaza, located in Orlando, Florida. Using the NADS MiniSim Simulator, seventy-two subjects are needed to complete this study. Five factors will be tested throughout twenty-four scenarios by means of factorial experimental design.

Experimental Design

As previously mentioned, a factorial experimental design was used for this research and five factors were analyzed. These factors and their levels are shown in the figure below. With these factors, there were a total of one hundred and eighty-eight scenarios. However, with one restriction, the scenarios could be reduced to one hundred and forty-four scenarios. From these, twenty-four scenarios were randomly chosen due to the experiment being limited to seventy-two participants.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
<th>Factor Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1 Path Setting of the path</td>
<td>1. Mainline-Express-Mainline</td>
<td></td>
</tr>
<tr>
<td>2. Mainline-Cash-Mainline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Mainline-Express-Ramp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Ramp-Express-Mainline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Ramp-Cash-Mainline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X2 Traffic Setting of traffic conditions</td>
<td>1. Peak hours/Heavy</td>
<td></td>
</tr>
<tr>
<td>2. Non-peak hours/Mild</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X3 Pavement Marking Whether there will be pavement marking or not</td>
<td>1. Yes</td>
<td></td>
</tr>
<tr>
<td>2. No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X4 Length Segment Length</td>
<td>1. Default (current)</td>
<td></td>
</tr>
<tr>
<td>2. Adding length before toll plaza</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Adding length after toll plaza</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X5 Signage The allocation of signs</td>
<td>1. Default (current)</td>
<td></td>
</tr>
<tr>
<td>2. Remove 3rd sign</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Remove 3rd sign, move 2nd sign and add sign on ramp</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor Descriptions

Five of the eight possible paths are used for this design. The figure below shows the five paths that will be taken.

The traffic conditions will vary between peak and off-peak hour. Real-world traffic data were analyzed and entered into the driving simulator to formulate realistic scenarios (to be explained in detail in the next section). The pavement markings that are being considered in this study are shown in the next figure. Some participants will be given scenarios with the markings that show where the lane splits and other participants will be given scenarios without these markings.

Traffic Data Preparation

In order to create realistic traffic volumes for the toll plaza driving simulator study, real traffic data from the Dean Mainline Toll Plaza was analyzed. Data was collected from six separate detectors located at the following mileposts on SR-408 Westbound: 18.8, 19.0, 19.4, 19.7, 19.8, 20.7. The locations of the detectors are shown in the figure below. To be more specific, the detectors located at miles 18.8, 19.7, and 19.9 are located in the gore areas. These are the merging and diverging areas for the ramp and mainline.

The peak data and off-peak data were collected and analyzed in a similar manner. The data was collected between the hours of 7 and 8 AM on October 1, 8, 15, 22, and 29 of 2014. It was found that there was no significant difference in speeds due to the date, time, and location of the data taken. However, the speed of each lane to be slightly different. The results of the speed data of the peak hour are shown below and the off-peak hour speed data is shown below the peak hour results. Lane 1 is the inner most lane and lane 3 is the outer most lane. The volumes are also shown below the speed data results, with peak hour on the left and off-peak hour on the right.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Mean Speed (mph)</th>
<th>Standard Deviation (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Speed (mph) Standard Deviation (mph)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane</td>
<td>Mean Speed (mph)</td>
<td>Standard Deviation (mph)</td>
</tr>
<tr>
<td>Lane 1</td>
<td>67.4</td>
<td>2.96</td>
</tr>
<tr>
<td>2</td>
<td>59.03</td>
<td>4.42</td>
</tr>
<tr>
<td>3</td>
<td>58.02</td>
<td>4.03</td>
</tr>
<tr>
<td>On-Ramp</td>
<td>45.45</td>
<td>2.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lane</th>
<th>Mean Speed (mph)</th>
<th>Standard Deviation (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Speed (mph) Standard Deviation (mph)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane</td>
<td>Mean Speed (mph)</td>
<td>Standard Deviation (mph)</td>
</tr>
<tr>
<td>Lane 1</td>
<td>69.7</td>
<td>2.4</td>
</tr>
<tr>
<td>2</td>
<td>63.5</td>
<td>2.3</td>
</tr>
<tr>
<td>3</td>
<td>66.9</td>
<td>4.0</td>
</tr>
<tr>
<td>On-Ramp</td>
<td>45.0</td>
<td>5.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lane</th>
<th>Mean Speed (mph)</th>
<th>Standard Deviation (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Speed (mph) Standard Deviation (mph)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane</td>
<td>Mean Speed (mph)</td>
<td>Standard Deviation (mph)</td>
</tr>
<tr>
<td>Lane 1</td>
<td>1,162 vph</td>
<td>769 vph</td>
</tr>
<tr>
<td>Lane 2</td>
<td>1,543 vph</td>
<td>807 vph</td>
</tr>
<tr>
<td>Lane 3</td>
<td>247 vph</td>
<td>120 vph</td>
</tr>
<tr>
<td>Total (All Lanes)</td>
<td>2,952 vph</td>
<td>1691 vph</td>
</tr>
<tr>
<td>Expressway vs. Cashway</td>
<td>71.29</td>
<td>85.15</td>
</tr>
<tr>
<td>On-Ramp</td>
<td>559 vph</td>
<td>204 vph</td>
</tr>
<tr>
<td>Off-Ramp Before Toll Plaza</td>
<td>52 vph</td>
<td>24 vph</td>
</tr>
<tr>
<td>Off-Ramp After Toll Plaza</td>
<td>77 vph</td>
<td>78 vph</td>
</tr>
<tr>
<td>Truck on Cashway</td>
<td>6%</td>
<td>15%</td>
</tr>
<tr>
<td>Truck on Expressway</td>
<td>6%</td>
<td>14%</td>
</tr>
</tbody>
</table>
Overview
The Big Traffic Data which are collected from various ITS traffic detection systems provide insights about the facilities at microscopic level in real-time. Consequently, efficient integration and utilization of such data for better performance of transportation system become a critical issue for traffic operators. In this project, different applications of the real-time microscopic traffic data were explored with a focus on operation efficiency and traffic safety:
- Evaluation traffic operation
- Real-time traffic safety evaluation
- Traffic data in Micro-simulation
- Dilemma zone analysis

Data Collection
- Automatic Vehicle Identification (AVI) Traffic Data
AVI was installed at toll plazas for Electronic Toll Collection (ETC) and at other locations for travel time estimation.
- Microwave Vehicle Detection System (MVDS) Traffic Data
MVDS is point-based roadway detection system. It was used to collect traffic flow parameters on each lane at one minute interval.

Evaluation Traffic Operation
- Congestion measurement
T(1) = \frac{\text{Actual travel time}}{\text{Free flow travel time}}
- Travel Time reliability
It measures consistency or dependability in travel times by:
  - Buffer index
  - Planning time index
  - Percent variation
  - Misery index

Real-time Traffic Safety Evaluation
The evaluation was conducted for crashes on expressway mainlines and ramps respectively. Traffic data which were 10 to 5 minutes prior to crash and non-crash events were extracted to estimate crash risk.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald Chi-Square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.1420</td>
<td>0.1318</td>
<td>567.9733</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Peak</td>
<td>0.1659</td>
<td>0.0888</td>
<td>3.4933</td>
<td>0.0616</td>
</tr>
<tr>
<td>U1_lanevol</td>
<td>0.0130</td>
<td>0.000691</td>
<td>212.6196</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>U1_spddiff</td>
<td>0.0228</td>
<td>0.00598</td>
<td>14.5063</td>
<td>0.0001</td>
</tr>
<tr>
<td>D1_trkpct</td>
<td>1.2891</td>
<td>0.2388</td>
<td>29.1463</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>D1_ci</td>
<td>4.6351</td>
<td>0.3374</td>
<td>188.7165</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Lane45</td>
<td>0.3196</td>
<td>0.0906</td>
<td>12.4456</td>
<td>0.0004</td>
</tr>
<tr>
<td>Median</td>
<td>-0.00505</td>
<td>0.00178</td>
<td>8.0038</td>
<td>0.0047</td>
</tr>
<tr>
<td>Shoulder</td>
<td>-0.5613</td>
<td>0.0900</td>
<td>38.9195</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

AUC = 0.7095

Dilemma Zone Analysis
Driver behavior during the yellow interval at signalized intersections was evaluated. Based on field data, a logistic regression model, which was a function of speed, distance to the stop line and stop/go position of the vehicle, was developed to predict driver stop/go decisions during simulation. The Cellular Automata (CA) model was employed to simulate the traffic flow. The four scenarios are listed as follows:
- Typical scenario
- Mean speed and speed standard deviation played a significant role in rear-end crash risk situations
- Flashing green scenario
- Had little influence on rear-end risk reduction, and could not reduce the percentage of false go decisions
- Pavement marking scenario
- Effectively reduced the RLR risk in some situations
- Effectively decreased rear-end crash risk and improve safety in most situations
- New countermeasure scenario (adding a flashing green signal next to the pavement marking)
- Lowest rear-end crash
- Rare RLR violation

Traffic Data in Micro-simulation
MVDS traffic data were utilized to calibrate and validate VISSIM network under poor visibility conditions. Then Surrogate Safety Assessment Model (SSAM) was used to measure the safety.

Conflict Number under Different Situation

<table>
<thead>
<tr>
<th>Volume (veh/h)</th>
<th>Speed Limit (mph)</th>
<th>Lane-change</th>
<th>Rear-end</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>4000</td>
<td>50</td>
<td>25</td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>134</td>
<td>48</td>
<td>182</td>
</tr>
<tr>
<td>8000</td>
<td>50</td>
<td>104</td>
<td>56</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>292</td>
<td>271</td>
<td>563</td>
</tr>
<tr>
<td>12000</td>
<td>50</td>
<td>198</td>
<td>131</td>
<td>329</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>309</td>
<td>270</td>
<td>579</td>
</tr>
</tbody>
</table>

The results are as follows:
- Conflict number increased along with the traffic volume
- Less conflicts in the fog area when the speed limit was lower
- Speed limit had more impact on conflict number under low volume condition

Real-time Safety Evaluation Model for Mainlines

Speed Distribution of Lead Vehicles

Distribution of Lead Vehicles
Overview
Early warning systems along roadways are an excellent method of dealing with hazardous conditions along roadways. However, research into different designs is quite limited in terms of their effectiveness. This study presents an experimental analysis of a dynamic message sign (DMS) and beacons’ effect on a drivers behavior when dealing with a reduced visibility scenario due to fog. The experimental design of this study follows six variables of interest to generate multiple scenarios using NADS-MiniSim Driving Simulator. Through this simulator, driver speed, braking, steering, and vehicle following behavior can be observed and analyzed while only test variables are present and constant. These variables, once collected, are then analyzed via ANOVA, regression, and crosstabs to observe significance on the driver as well as each other.

Data Description and Preparation
All data of interest for this research are collected via the simulation tests as well as demographic data collected from the participants themselves. This demographic data of interest includes the drivers: Gender/Age, Experience, Driving Frequency, Crash History, and so on. Each participant driving behavior will be observed on: Overall Speed, Breaking, Acceleration/Declaration, Vehicle Following Distance, and Sign/Vehicle Recognition. The variables that these drivers will encounter are what make up the experimental design of this study and are as follows:

1) Roadway Type (Freeway / Arterial)
2) Visibility Distance (500ft / 300ft / 150ft)
3) Number of DMS Present (0 / 1 / 2)
4) DMS Message (Null / Warning / Advised)
5) Traffic Setting (Heavy / Light)
6) Beacon Presence (0 / 1 / 1)

The first variable represents the location of the scenario; a 3-lane 70MPH freeway and a 2-lane 65MPH arterial. The visibility distance represents the thickness of the fog. The chosen values for the fog are studied at extreme conditions as it shows more potential to observe a change in driving behavior. Each roadway has options for up to two DMS present along the roadway. Studying different instances of the DMS presence could produce findings. The message of the DMS is set to display either a ‘warning’ of fog presents, or a ‘advisement’ informing the driver of fog ahead and to reduce speed. The traffic setting and beacon presence are to test the effects under different traffic volumes and beacon usage.

Experimental Design
A balanced block factorial design is chosen to break the variable of interest into scenarios to be tested. To simplify, restrictions are established to eliminate unusable scenarios; further, 12 random scenarios are chosen for each roadway type. Using these 24 scenarios the block design is established.

- Testing order is broken into 9 blocks with 8 groups.
- Each ‘scenario’ pair will be encountered 3 times.
- Total of 72 participants needed to complete test scheme.
- Age distribution of participants based on FDOT and local crash data.

Balanced Block Design

<table>
<thead>
<tr>
<th>Block</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
<th>Scenario 5</th>
<th>Scenario 6</th>
<th>Scenario 7</th>
<th>Scenario 8</th>
<th>Scenario 9</th>
<th>Scenario 10</th>
<th>Scenario 11</th>
<th>Scenario 12</th>
<th>Scenario 13</th>
<th>Scenario 14</th>
<th>Scenario 15</th>
<th>Scenario 16</th>
<th>Scenario 17</th>
<th>Scenario 18</th>
<th>Scenario 19</th>
<th>Scenario 20</th>
<th>Scenario 21</th>
<th>Scenario 22</th>
<th>Scenario 23</th>
<th>Scenario 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

In order to validate the simulation data, weather and traffic data from the real world location is used for comparison.

SR441 Scenario Plan

Plaines Prairie Incident and MiniSim Device

Summary and Future Analysis
Ultimately it is expected that driver behavior compared between the clear and foggy segments of roadway will show different trends based on the presence of DMS, beacons, the message provided, and the traffic setting present. Future studies are also possible, where additional testing can be done in terms of the DMS message and how it is presented. Additionally, once more sensors become available along the real-world study location, further validation can be performed with the simulation findings. No matter the case, the goal of the study is to find an effective early warning system to protect drivers from hazardous weather conditions.
Overview

Compared to micro scale safety studies, microscopic-focused research is more efficient at integrating zone-level features into crash prediction models and identifying hot zones in large study areas. However, few studies have focused on the limitations of current hotspot/hotzone identification methods (HSID) applied at the macro level. This study applied six common HSID methods and compared their consistency in identifying hot zones. The crash data was based on five years of crash records from Central Florida (Orange, Seminole, and Osceola Counties). The results showed that the hot zones identified by the crash frequency, Empirical Bayes, and Potential for Safety Improvement methods all had high consistency and stability over time, followed by the crash rate and Equivalent Property Damage Only methods. The Propensity method had the lowest consistency. Other possible factors related to the methods’ performance were also examined, which included the time length of the before period, the time length of the after period, the time gap, hot zone threshold (\( \alpha \)), and different crash types. However, these factors affected the performance of the methods only slightly. Also, the main problem of the crash frequency method, regression-to-the-mean, was not found to affect the performance of the method at the macro level because the consistency stayed high even in cases where the time length of the before period was as low as one year.

Study Objectives

1. This paper compared the performance of six common HSID methods by the site consistency test at the macro level.
2. The limitations of current HSID methods were examined at the macro level.

Study Data

- Study Area: Orange, Seminole, & Osceola Counties
- Target: Crashes (2005-2010)
- Independent variables: Roadway/traffic, and socio-economic data

Hotspot Identification Methods

Hotzones are the areas having high crash risk over the defined threshold.
1. Crash frequency: Each study unit (e.g., TSAZ in this study) is ranked by its total crash frequency.
2. Crash rate: the total crash frequency divided by the overall exposure, such as VMT for each TSAZ.
3. Equivalent Property Damage Only Crash Frequency: Different weights were developed to combine frequency and severity based on the approach of willing to pay (Fatal injury: PDO = 771.35; 1).
4. Proportion method: Define parameters regarding one target crash type, and then an estimate of the probability of this specific crash type occurring among all crashes.
5. Empirical Bayesian method (EB): a weighted combination of the predictions obtained from an SPF and the observed crash frequency
6. Potential for Safety Improvement (PSI): the difference between the expected crash count and the predicted crash count

\[ P(X_i \leq x -1, n; p_i) = \sum_{n=0}^{\infty} \frac{(n-i)!i!}{n!} p_i^{n-i} (1-p_i)^{-i} \]

\[ EB = w \times E(Y) + (1 - W)N_i \]

\[ PSI = EB - E(Y) \]

Performance Evaluation Criteria

Site consistency test (SCT): a high-risk hot-zone repeated during a study period

\[ SCT_i = \sum_{k=n-mu+1}^{n} \frac{C_{k,i+j} \times y_{i+k}}{\sum_{k=n-mu+1}^{n} \sum_{i=n-mu+1}^{n} \times y_{i+k}} \]

Results

Six scenarios were used to examine possible factors related to method consistency.

Scenario 1: Different HSID Methods

Overall, crash frequency, EB, and PSI method all have high consistency, followed by the crash rate and EPDO method. The proportion method has the lowest consistency.

Scenario 2: The Length of the Before Period

When the length of before period increases, the consistency of EPDO method increased while that of the crash rate method decreased. Other results are similar.

Scenario 3: The Length of the After Period

No significant trend change when the length of the after period is extended.

Scenario 4: Time Gap

The use of historical crash data to identify hot-zones does not change the consistency of the method in use.

Scenario 5: Hotspot Threshold

No clear trend of the consistency when the hotspot threshold changes (reduced from 95 % to 90%).

Scenario 6: Different Crash Types (FI, Pedestrian crashes)

For fatal and injury crashes, the crash frequency and PSI method still showed high consistency, although at slightly lower values than for total crash data (90 % → 80 %). For pedestrian crashes, the crash frequency and EB methods showed high consistency, with only slightly lower values than for total crash data (90 % → 70 %).

Discussion and Conclusions

2. Other possible factors related to the methods’ performance were also examined, and these factors affected the performance of the methods only slightly.
3. Also, regression-to-the-mean, was not found to affect the performance of the method at the macro level.

Acknowledgment

The authors wish to thank the Florida Department of Transportation for funding this study. Some of the research for this paper was conducted as part of the efforts of the Southeastern Transportation Center (STC) at the University of Tennessee. The primary sponsor for the STC is the United States Department of Transportation through grant number DTRT13-G-UTC34.
Overview
This study aims at identifying two zonal levels factors. The first is to identify hot zones at which pedestrian crashes occurs, while the second are zones where crash-involved pedestrians came from. Bayesian Poisson Lognormal Simultaneous Equation Spatial Error Model (BPLSESEM) was estimated and revealed significant factors for the two target variables. Then, PSIs (Potential for Safety Improvements) were computed using the model. Subsequently, a novel hot zone identification method was suggested to combine both hot zones from where vulnerable pedestrians originated with hot zones where many pedestrian crashes occur. For the former zones, targeted safety education and awareness campaigns can be provided as countermeasures whereas area-wide engineering treatments and enforcement may be effective safety treatments for the latter ones. Thus, it is expected that practitioners are able to suggest appropriate safety treatments for pedestrian crashes using the method and results from this study.

Data Preparation
Data from 983 ZIP areas in Florida were used for the analysis. Pedestrian crashes occurring between 2009 and 2011 were collected from Florida Department of Transportation (FDOT). Demographic, commute pattern, and socio-economic data were obtained from the U.S. Census Bureau and the roadway/traffic data were acquired from FDOT Roadway Characteristics Inventory. Lastly, the facility/attribution data were obtained from FDOT Unified Basemap Repository. Overall 40 candidate explanatory variables and 2 target variables were processed.

Statistical Modeling
Bayesian Poisson Lognormal Simultaneous Equations Spatial Error Model (BPLSESEM) was adopted in this study.

Equation (1):
\[ \lambda_{ik} = \exp(\beta X_{ik} + \delta u_k + \psi) \]

Equation (2):
\[ \lambda_{ik} = \exp(\beta X_{ik} + \delta u_k + \delta_{2} u_{2} + \psi) \]

where, \( \lambda_{ik} \) is the expected number of pedestrian crashes per crash location ZIP (\( k = 1 \)) or the expected number of crash-involved crashes per residence ZIP (\( k = 2 \), \( X_{ik} \) is a row vector of explanatory variables showing characteristics of ZIP \( i \) for target \( k \), \( \beta \) is a coefficient estimate of model covariates \( X_{ik} \), \( \theta \) is a random error term representing normal heterogeneity of ZIP \( i \) for target \( k \), \( u_k \) follows normal distribution \((0, \tau_k^2)\) for ZIP \( i \) and target \( k \), \( \tau_k^2 \) is the precision parameter that is the inverse of the variance; it follows prior gamma \((0.5, 0.005)\), \( \delta \) is the coefficient for \( u_k \) in Equation (1), while \( \delta_{2} \) and \( \delta_{3} \) are the coefficients for \( u_{2} \) in Equation (2), respectively, and \( \psi \) is a shared spatial autocorrelation error term (CAR).

Zonal-level Screening
In this study PSI (Potential for Safety Improvement) was selected as the performance measure. PSI, or excess crash frequency, shows whether a zone is experiencing more or less number of crashes compared to other zones with similar characteristics. PSI is calculated by the difference between the expected and predicted number of crashes.

In the preceding section, hot zones for two targets: ‘Pedestrian crashes per crash location ZIP’ and ‘Crash-involved pedestrians per residence ZIP’ are identified individually. In this section, the hot zone identification results of the two targets are combined to provide a broad spectrum perspective for both locations with higher risk for pedestrians and residences with many pedestrians vulnerable to crashes.


Conclusion
A novel hot zone identification method was suggested to combine both hot zones with many pedestrian crash occurrences and hot zones with many crash-involved pedestrians in the residence. For the former zones, area-wide engineering treatments and enforcement can be provided as general countermeasures whereas targeted safety education and campaigns may be effective safety treatments for the latter ones.

Acknowledgment
The authors wish to thank the FDOT for funding this study. Some of the research for this paper was conducted as part of the efforts of the Southeastern Transportation Center (STC) at the University of Tennessee. The primary sponsor for the STC is the United States Department of Transportation through grant number DTRT13-G-UTC34.
Urban Expressway Traffic Safety and Operation Evaluation and Improvement Using Big Data

Qi Shi and Mohamed Abdel-Aty
Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, Florida

Background
The advent of Big Data era

• Since 2010
• Key words: Data Information Big Predictions

What is Big Data?
• Volume: increasing size of data
• Velocity: unprecedented streaming speed of data
• Variety: wide range of data formats
• Other dimensions of Big Data: Veracity, Variability, Complexity.

Big Data in transportation arena

Data sources
• Intelligent Transportation System (ITS) facilities
• Roadway geometric data
• Crash database
• Socio-demographic database
• Web traffic and social network database

Data types
• Structured data
• Unstructured data

Applications of Big Data in Transportation arena
• Operation efficiency (congestion improvement)
• Traffic safety (crash prevention)

Research Objectives
1) Evaluation of the relationship between traffic operation and safety
2) Improving traffic operation and traffic safety simultaneously

Tools
Three expressways in Central Florida area:
• 75 miles in length located in urban area

ITS system:
• Microwave Vehicle Detection System
• Data collection at 1-minute interval from Jul, 2013 to Feb, 2014 from 275 detectors

Crash reports:
• 243 rear-end crashes on the expressways during the time period
• Crash cases vs non-crash control cases: 1:4

Matching ITS and crash data:
• 10-5 minutes prior to each crash case

Methodology
Operation efficiency (congestion) evaluation:
\[
CI = \frac{\text{free flow speed} - \text{actual speed}}{\text{free flow speed}} \quad \text{when } CI \leq 0 \quad \text{ when } CI > 0
\]

Real-time traffic safety (rear-end crashes) evaluation:
• Random forest: importance of crash contributing factors
• Bayesian logit model: effects of crash contributing factors

Simultaneous improvement of traffic operation and safety:
• First Order Reliability Method (FORM) Analysis

Analysis Results
Operation evaluation
• Congestion occurrence: highly localized and time specific
• Urban area: congestion in rush hours
• Congestion intensity: varies on spatial-temporal dimensions

Conclusions:
A Bayesian Ridge Regression Analysis of Congestion’s Impact on Urban Expressway Safety
Qi Shi, Mohamed Abdel-Aty, Jaeyoung Lee
Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, Florida

Background
Providing motorists with efficient and safe traffic system has long been considered a priority of traffic professionals. With the growth in traffic demand, the construction of road infrastructure, and congestion and safety concerns arise. In urban areas, many traffic authorities have turned to toll/tunpike facilities and efficient use of Intelligent Transportation Systems (ITS) techniques as remedies for congestion and to improve safety.

Challenges in studies on congestion-safety relationship
1) How congestion is measured could affect the conclusion
2) Congestion could be time specific
3) Multicollinearity could alter the estimation of contributing factors

Objective
Identifying the relationship between congestion and crashes on urban expressways

Data Preparation
Urban expressway:
- State Road (SR) 408, Orlando
- Length: 21.4 miles
ITS Traffic Detection systems:
- Automatic Vehicle Identification (AVI) System
- Microwave Vehicle Detection System (MVDS)

Traffic flow data from AVI and MVDS systems
Development of congestion measures:
- AVI -- Travel time based:
  \[ \text{TTT} = \text{actual travel time} \] free travel time
- MVDS -- Travel speed based:
  \[ \text{Cl} = \frac{\text{free flow speed - actual speed}}{\text{free flow speed}} \] if \( \text{Cl} > 0; \text{Cl} = 0 \text{ if } \text{Cl} \leq 0 \]
- MVDS -- Density based:
  Lane occupancy: percent of time a point on the road is occupied by vehicles

Roadway geometric characteristics data
- Geometric elements: number of lanes, existence of auxiliary lanes, horizontal degree of curvature, speed limit, etc.
- Homogeneous segments: 75 segments on Eastbound (EB) and 76 segments on Westbound (WB)

Crash data
Selection of crash data for congestion-safety analysis
1) Data should reflect traffic conditions for the days when recurrent congestion occurs
2) Crashes should be more likely to be influenced by traffic flow
- 06:00 to 21:00 on weekdays, Sep. 2012 – Dec. 2013
- 472 crashes

Methodology
- Diagnostics of multicollinearity
  - Correlation test: Pearson’s correlation test
  - Coefficients of determination: \( R^2 = \frac{\sum(y_i - \bar{y}_i)^2}{\sum(y_i - \bar{y})^2} \)
  - Tolerance (TOL): \( TOL_k = 1 - R_k^2 \)
  - Variance Inflation Factor (VIF): \( VIF_k = \frac{1}{TOL_k} \)

- Bayesian ridge regression
  - Crash frequency model
    \[ \log(\lambda_{ij}) = \gamma + X_i\beta + X_1e_i + X_2e_i \]
  - Ridge regression
    \[ z_j = \frac{(y_j - \bar{y}_j)}{sd(y_j)} \]
    \[ y_j = \frac{b_j}{sd(y_j)} \]
    \[ a_j = z_jb_j \]

Results and Discussion
Presence of multicollinearity:
\[ VIFs < 10 \]
\[ R_{\text{Auxiliary}}^2 > R_{\text{general}}^2 \]
Response: Multicollinearity should be taken into account

Conclusions
- The impact of congestion on traffic safety is significant during peak hours
- Different congestion measures might alter its impact on safety
- Considering multicollinearity helps clarify the congestion-safety relationship