Evaluation of Real World Scenarios Involving Toll Plazas Using Simulation

2015 SAFER-SIM Symposium

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.

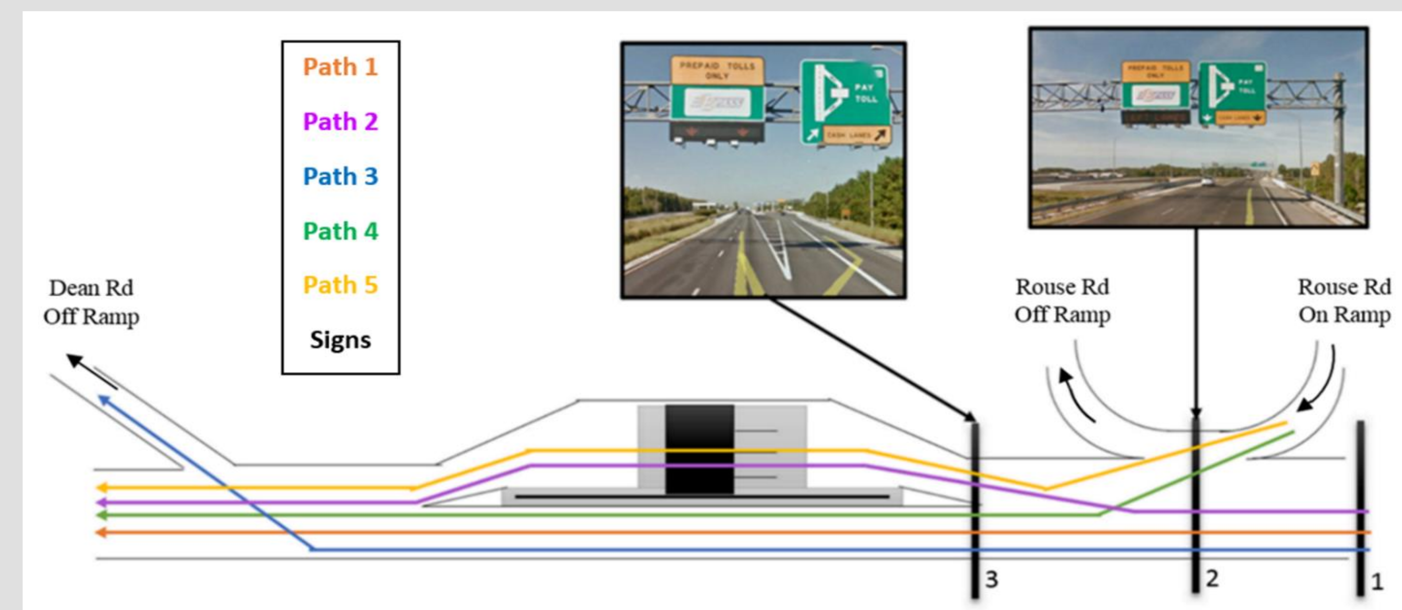
Experimental Design

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.

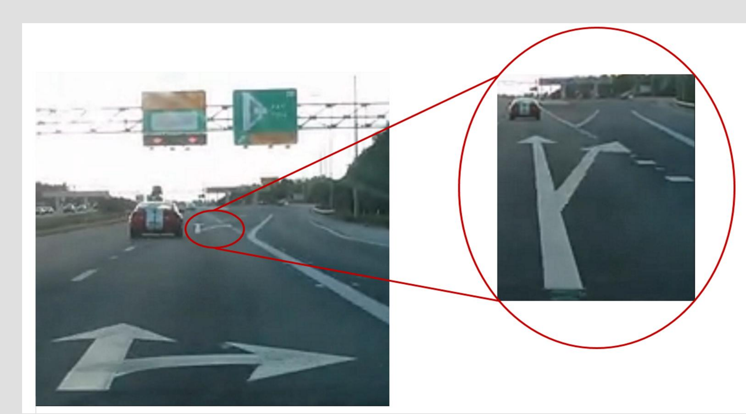
	Factor	Description	Factor Levels
X1	Path	Setting of the path	1. Mainline-Express-Mainline
			2. Mainline-Cash-Mainline
			3. Mainline-Express-Ramp
			4. Ramp-Express-Mainline
			5. Ramp-Cash-Mainline
X2	Traffic	Setting of traffic conditions	1. Peak hours/Heavy
			2. Non-peak hours/Mild
X3	Pavement Marking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

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.



The length factor will vary between the base length, a longer distance between the toll plaza and the downstream off ramp, and a longer distance between the toll plaza and the upstream on ramp. The base length will be the existing condition at the toll plaza, while a distance of 500 feet will be added for each distance change. There are three different scenarios for the signage factor (please refer to the figure displaying the driver paths):

1. The first scenario is the existing base condition that is shown.
2. Another scenario, is simply removing the sign closest to the toll plaza labeled #3.
3. The third scenario, involved:
    - Adding a DMS sign, similar to figure above
    - Removing sign #3
    - Moving sign #2 farther upstream before the on ramp

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, 19.9, 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.

Lane	Mean Speed (mph)	Standard Deviation (mph)
1	67.4	2.96
2	59.03	4.42
3	58.02	4.03
On-Ramp	45.45	2.86

Lane	Mean Speed (mph)	Standard Deviation (mph)
1	69.7	2.4
2	63.5	2.3
3	60.9	4.0
On-Ramp	45.0	5.5

	Peak	Off-Peak
Lane 1	1,162 vph	769 vph
Lane 2	1,543 vph	807 vph
Lane 3	247 vph	120 vph
Total (All Lanes)	2,952 vph	1691 vph
Expressway vs. Cashway	71:29	85:15:00
On-Ramp	559 vph	204 vph
Off-Ramp Before Toll Plaza	52 vph	24 vph
Off-Ramp After Toll Plaza	77 vph	78 vph
Truck on Cashway	6%	15%
Truck on Expressway	6%	14%

Integration of Microscopic Big Traffic Data in Simulation-Based Safety Analysis

2015 SAFER-SIM Symposium

Qi Shi, Ling Wang, Yina Wu, Binya Zhang, Mohamed Abdel-Aty , Essam Radwan
Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, Florida

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

TTI indicates the additional time spent on a trip compared to an ideal trip on the same corridor.

$$TTI = \frac{\text{Actual travel time}}{\text{Free flow travel time}}$$

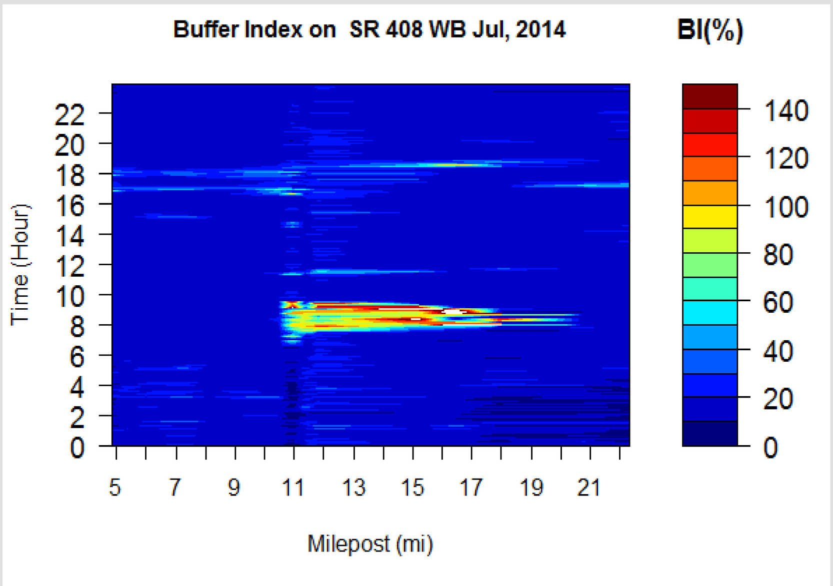
- Travel Time reliability

It measures consistency or dependability in travel times by:

- Buffer index

$$\left[\frac{\text{95th Travel Time} - \text{Average Travel Time}}{\text{Average Travel Time}}\right] \times 100\%$$

- 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.

Real-time Safety Evaluation Model for Mainlines

Parameter	Estimate	Std. Error	Wald Chi-Square	P-value
Intercept	-3.1420	0.1318	567.9733	<.0001
Peak	0.1659	0.0888	3.4933	0.0616
U1_lanevol	0.0130	0.000891	212.6196	<.0001
U1_spddiff	0.0228	0.00598	14.5063	0.0001
D1_trkpct	1.2891	0.2388	29.1463	<.0001
D1_ci	4.6351	0.3374	188.7165	<.0001
Lane45	0.3196	0.0906	12.4456	0.0004
Median	-0.00505	0.00178	8.0038	0.0047
Shoulder	-0.5613	0.0900	38.9195	<.0001
AUC	0.7095			

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

Volume (veh/h)	Speed Limit (mph)	Conflict number		
		Lane-change	Rear-end	Total
4000	50	25	3	28
	70	134	48	182
8000	50	104	56	160
	70	292	271	563
12000	50	198	131	329
	70	309	270	579

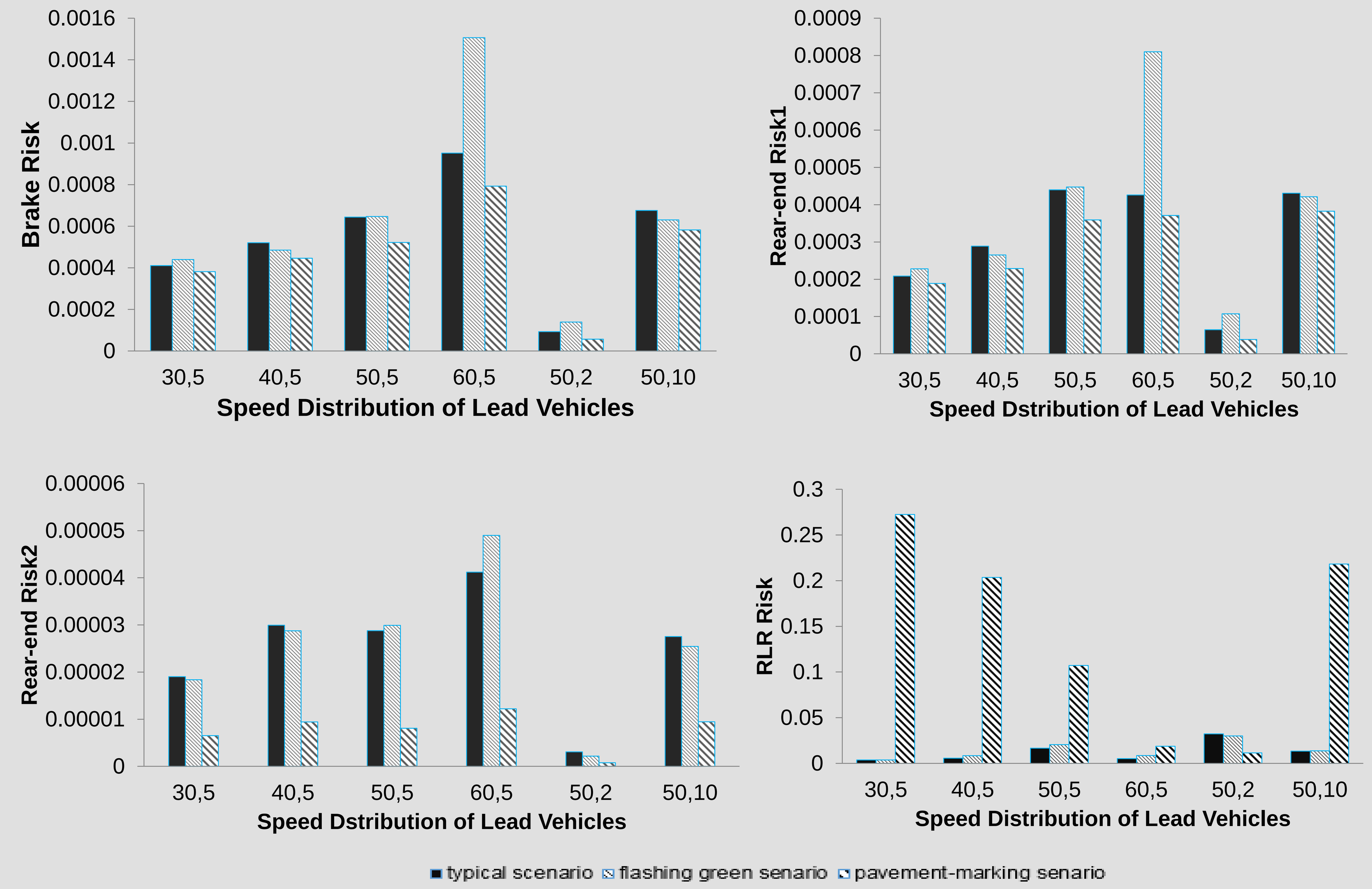
The results are as follows:

- 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

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 the lead/follow 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 risky 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



IMPACT OF DYNAMIC MESSAGE SIGNS ON DRIVER BEHAVIOR UNDER REDUCED VISIBILITY CONDITIONS USING SIMULATION

2015 SAFER-SIM Symposium

Ryan Selby, Yina Wu, Qing Cai, Qi Shi, Mohamed Abdel-Aty

Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, Florida

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 participants driving behavior will be observed on: Overall Speed, Breaking, Acceleration/Decleration, 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)

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

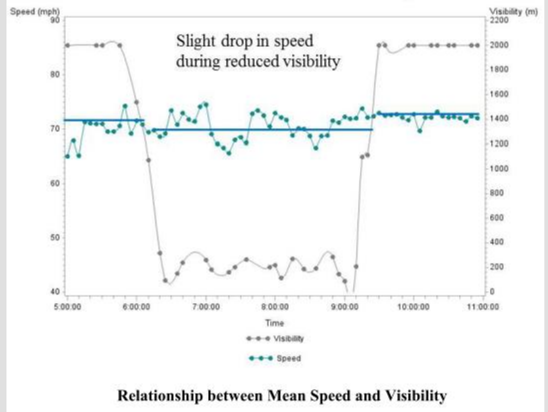
A balanced block factorial design is chosen to break the variable of interest into scenarios to be tested. To simplify, restrictions are established to eliminate unusable scenarios; further, 12 random scenarios are chosen for each roadway type. Using these 24 scenarios the block design is established.

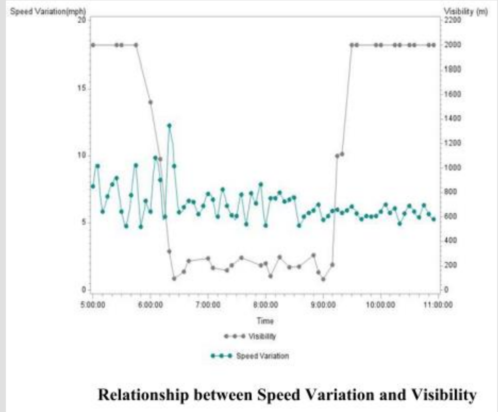
- Testing order is broken into 9 blocks with 8 groups.
- Each 'scenario' pair will be encountered 3 times.
- Total of 72 participants needed to complete test scheme.
- Age distribution of participants based on FDOT and local crash data.

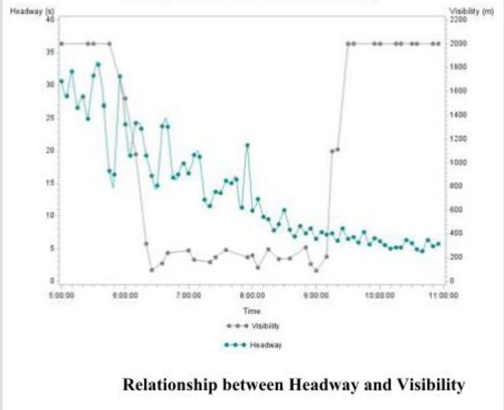
Balanced Block Design

Block	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24
1	23	4	1	18	10	13	11	20	24	16	17	6	8	7	12	19	21	3	15	9	14	2	5	22
2	11	24	9	1	8	19	23	13	5	7	21	12	20	17	16	18	3	10	4	22	2	6	14	15
3	21	5	23	7	12	3	16	20	8	6	24	2	19	9	13	22	10	18	11	4	15	17	1	14
4	4	10	1	17	13	6	18	12	5	24	15	2	20	8	23	14	22	11	7	16	21	19	9	3
5	7	8	1	5	2	13	15	9	12	6	24	19	10	3	11	16	18	20	14	21	22	17	4	23
6	12	14	23	2	24	5	8	4	16	6	20	21	19	7	10	9	3	18	22	1	11	17	13	15
7	2	16	9	6	23	4	18	24	3	17	20	1	19	11	7	10	12	13	22	14	15	5	8	21
8	11	10	9	19	5	24	3	20	6	16	1	18	14	21	4	13	22	17	12	15	2	7	23	8
9	16	18	23	6	20	5	7	13	22	1	14	21	3	2	4	10	15	24	12	8	9	17	11	19

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







Abdel-Aty, M., Oloufa, A., Peng, Y., et al. "Real Time Monitoring and Prediction of Reduced Visibility Events on Florida's Highways." (2014)

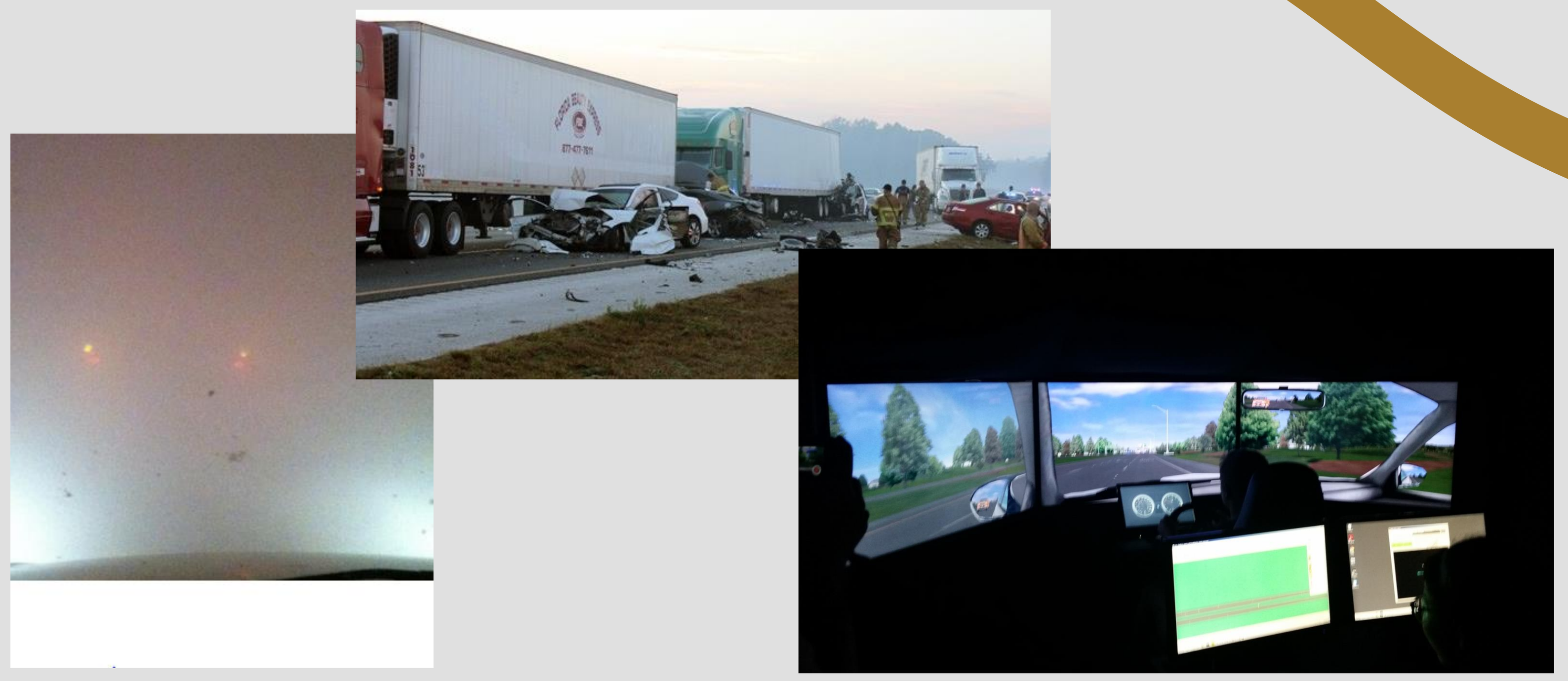
Scenario Structure

The scenario itself is based on I-75 and SR441 located in Polk county near Gainesville Florida. Back in 2012, a smog incident occurred leading to a major vehicle pileup leading to several fatalities and injuries. This has led to the desire of an early warning system to alter drivers of upcoming hazards.

The design of the scenario follows:

1) Clear Zone; allows drivers to adjust to the scenario and allow the study of initial driver behavior.
2) Variable Zone; DMS and beacons are present and driver reaction is observed.
3) Transition Zone; approximately 0.75mi to the study zone, the visibility distance decreases to desired level.
4) Fog Zone; allows for observation of driver behavior to the reduced visibility fog condition

Plaines Prairie Incident and MiniSim Device

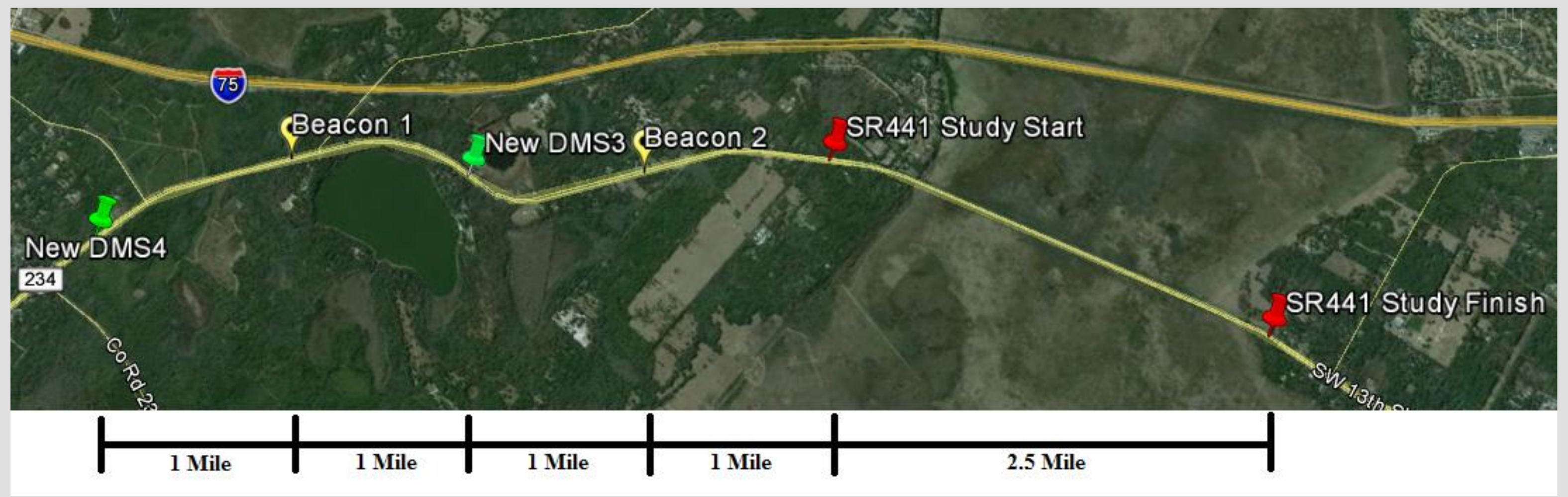


Summary and Future Analysis

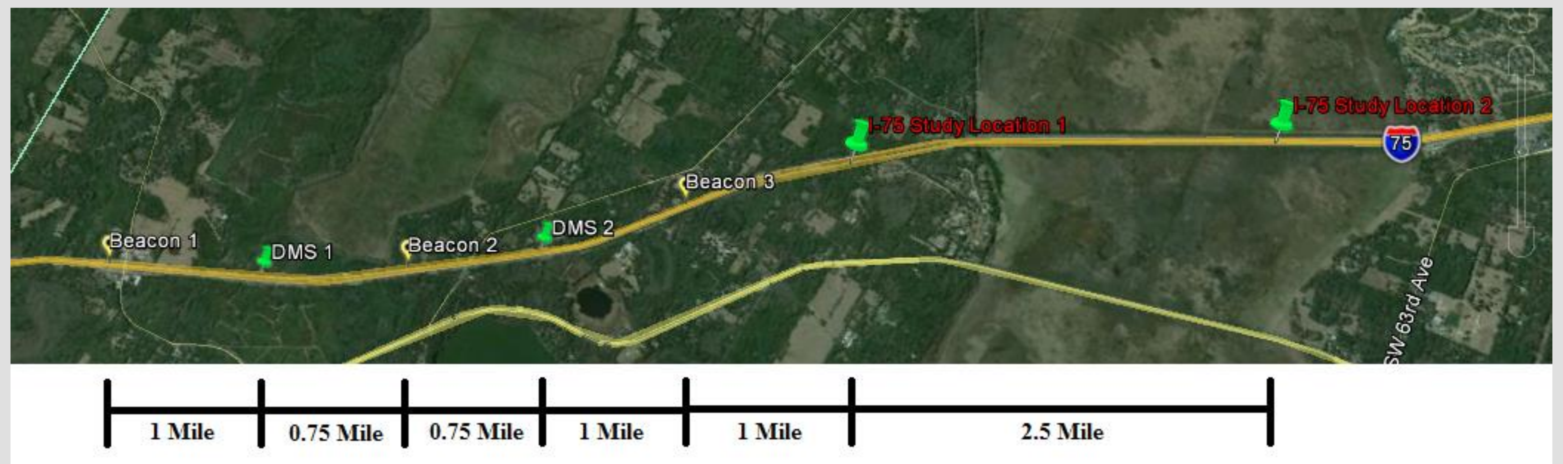
Ultimately it is expected that driver behavior compared between the clear and foggy segments of roadway will show different trends based on the presence of DMS, beacons, the message provided, and the traffic setting present.

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.

SR441 Scenario Plan



I-75 Scenario Plan



Comparative Analysis of Hot Zone Identification Methods At the Macroscopic Level

2015 Road Safety & Simulation International Conference

Jaeyoung Lee[1], Pei-Fen Kuo[2], Mohamed Abdel-Aty[1]

[1] Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, Florida

[2] Department of Crime Prevention and Correction, Central Police University, Republic of China (Taiwan)

Overview

Compared to micro scale safety studies, macroscopic-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/hot-zone 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 Proportion 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 (α), 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

Hot-zones 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.

$$P(X_{ij} \le x-1, n; p_j) = \sum_{i=0}^{x-1} \frac{(n)!}{(n-1)!(i)!} p_j^{\,i} (1-p_j)^{n-i}$$

5. Empirical Bayesian method (EB): a weighted combination of the predictions obtained from an SPF and the observed crash frequency

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

6. Potential for Safety Improvement (PSI): the difference between the expected crash count and the predicted crash count

$$PSI = EB - E(Y)$$

Performance Evaluation Criteria

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

$$SCT_i = \sum_{k=n-n\alpha+1}^{n} C_{k,j,i+1} \Big/ \Big(\sum_{k=n-n\alpha+1}^{n} L_{k,j} \times y_{i+1} \Big)$$

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

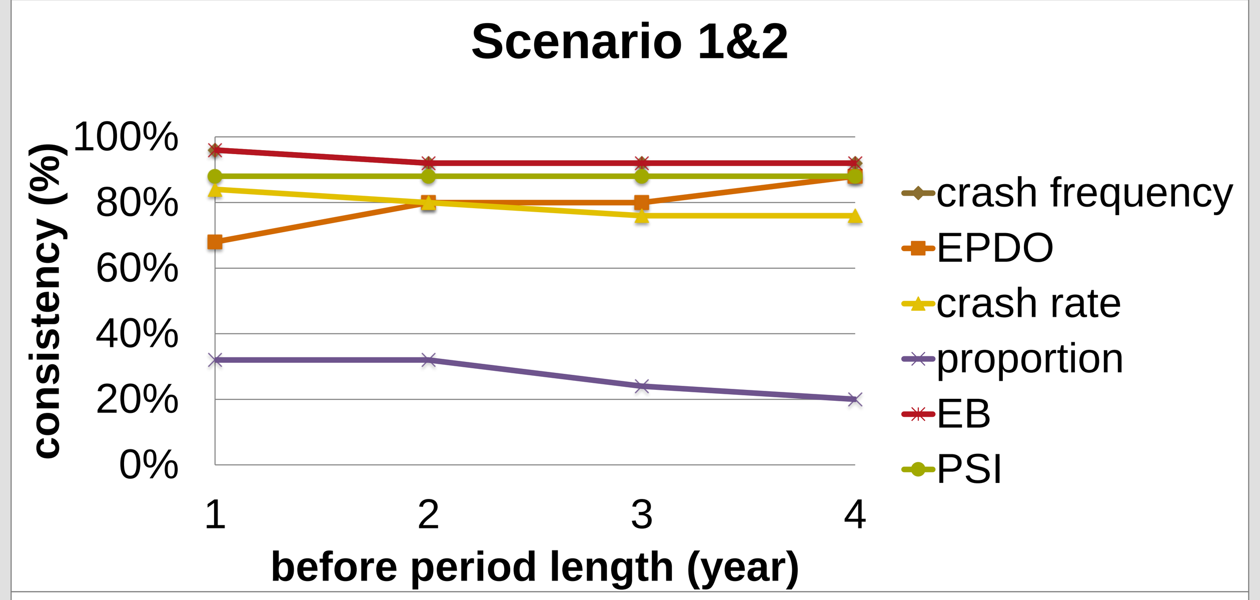
The use of historical crash data to identify hot-zones does not change the consistency of the method in use.

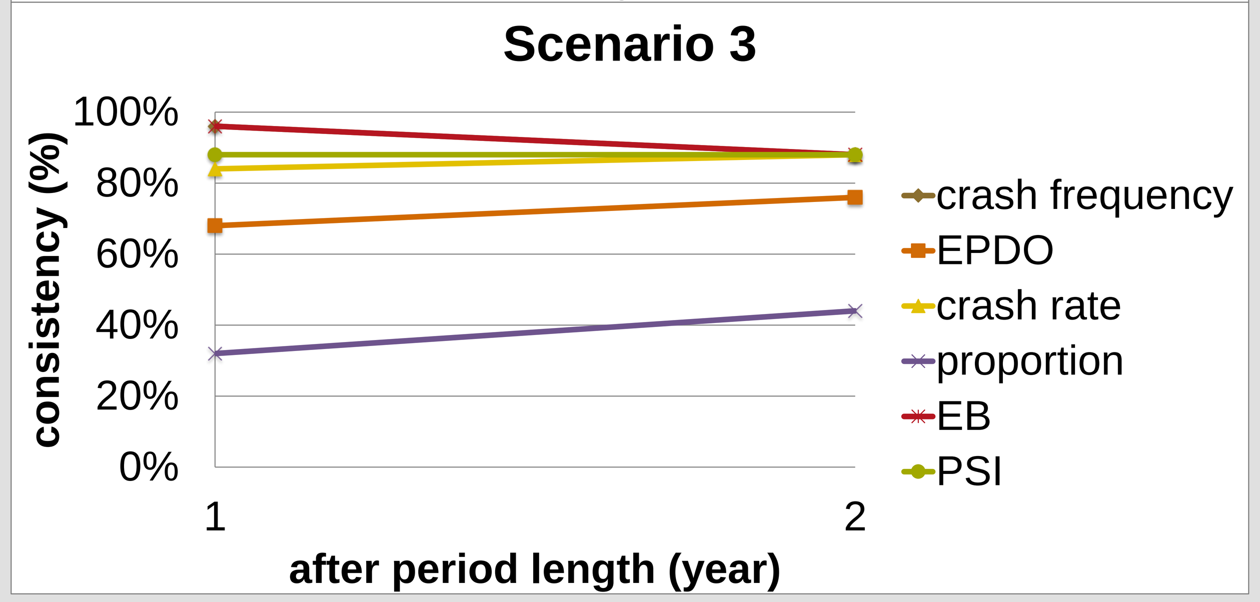
Scenario 5: Hotspot Threshold

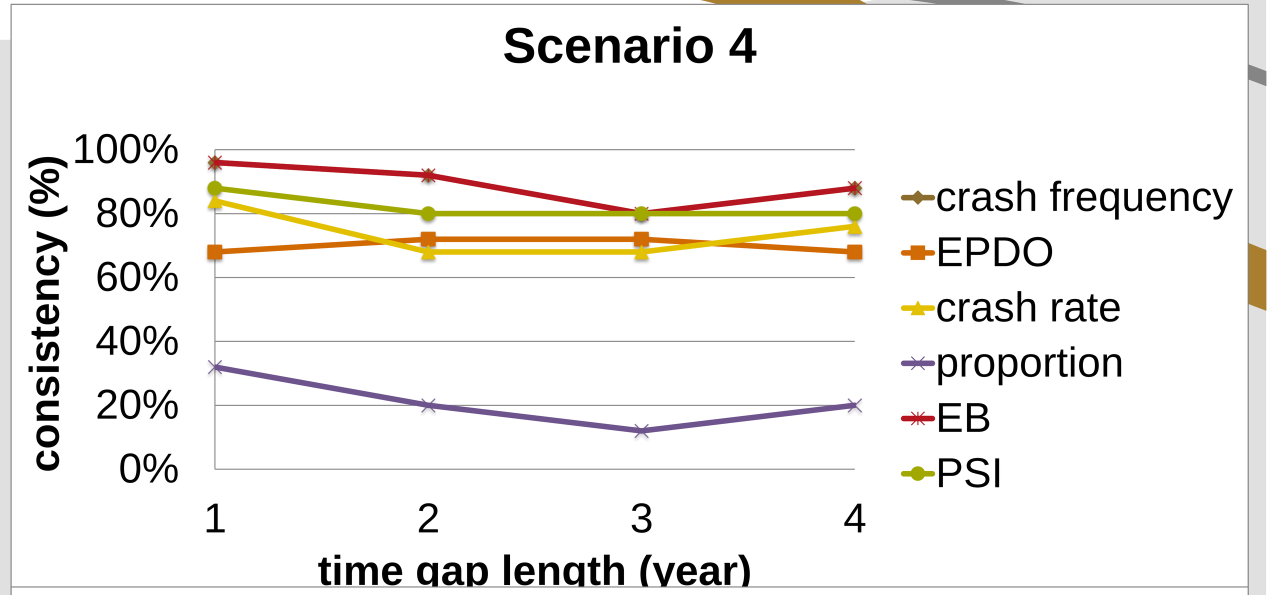
No clear trend of the consistency when the hotspot threshold changes (reduced from 95 % to 90%).

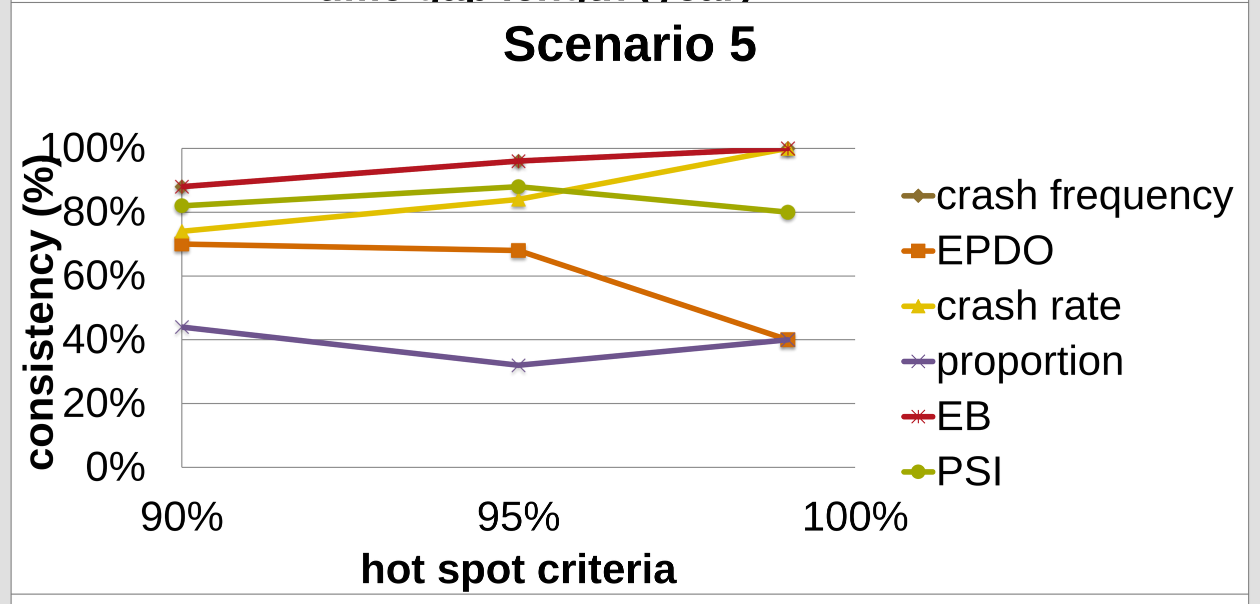
Scenario 6: Different Crash Types(FI, Pedestrian crashes)

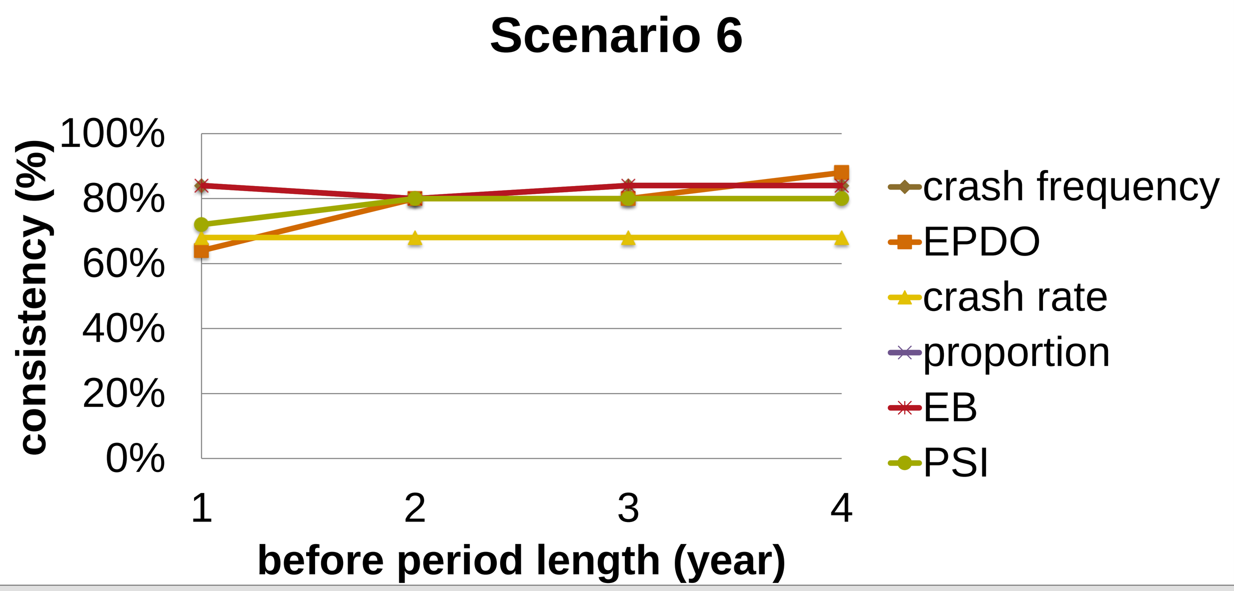
For fatal and injury crashes, the crash frequency and PSI method still showed high consistency, although at slightly lower values than for total crash data (90 % → 80 %). For pedestrian crashes, the crash frequency and EB methods showed high consistency, with only slightly lower values than for total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











Discussion and Conclusions

1. Consistency: the crash frequency, Empirical Bayesian, and PSI> EPDO> Proportion Method.
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.

Pedestrian Safety Data Analytics: Modeling and Screening for Residence and Crash Zones

2015 Road Safety & Simulation International Conference

University of Central Florida

Jaeyoung Lee[1], Mohamed Abdel-Aty[1], Helai Huang[2]

[1] Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, Florida

[2] Urban Transportation Research Center, School of Traffic and Transportation Engineering Central South University, PR China

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/attraction 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):

$$\lambda_{i1} = exp(\beta_1 X_{i1} + \delta_1 u_{i1} + \varphi_i)$$

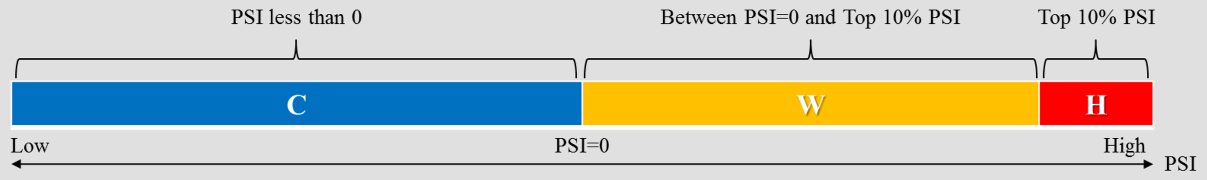
Equation (2):

$$\lambda_{i2} = exp(\beta_2 X_{i2} + \delta_2 u_{i1} + \delta_3 u_{i2} + \varphi_i)$$

where, λ_{ik} is the expected number of pedestrian crashes per crash location ZIP *i* (*k*=1) or the expected number of crash-involved crashes per residence ZIP *i* (*k*=2), X_{ik} is a row vector of explanatory variables showing characteristics of ZIP *i*, for target k, β_k is a coefficient estimate of model covariates X_{ik}, θ_{ik} is a random error term representing normal heterogeneity of ZIP *i*, for target k, u_{ik} follows normal distribution (0, τ_θ) for ZIP *i* and target k, τ_θ is the precision parameter that is the inverse of the variance; it follows prior gamma (0.5, 0.005), δ_1 is the coefficient for u_{i1} in Equation (1), while δ_2 and δ_3 are the coefficients for u_{i1} and u_{i2} in Equation (2), respectively, and φ_i is a shared spatial autocorrelation error term (CAR).

Zonal-level Screening

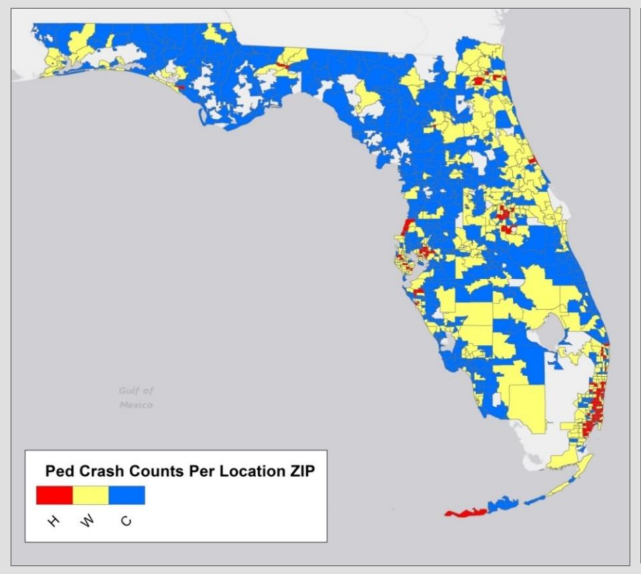
In this study PSI (Potential for Safety Improvement) was selected as the performance measure. PSI, or excess crash frequency, shows whether a zone is experiencing more or less number of crashes compared to other zones with similar characteristics. PSI is calculated by the difference between the expected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

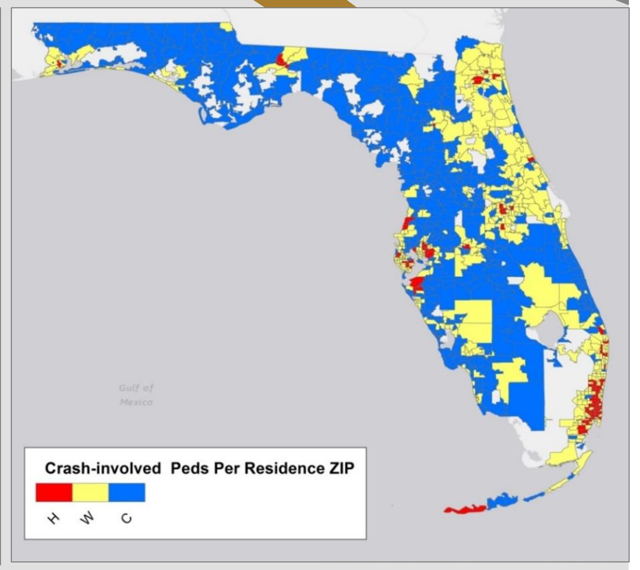


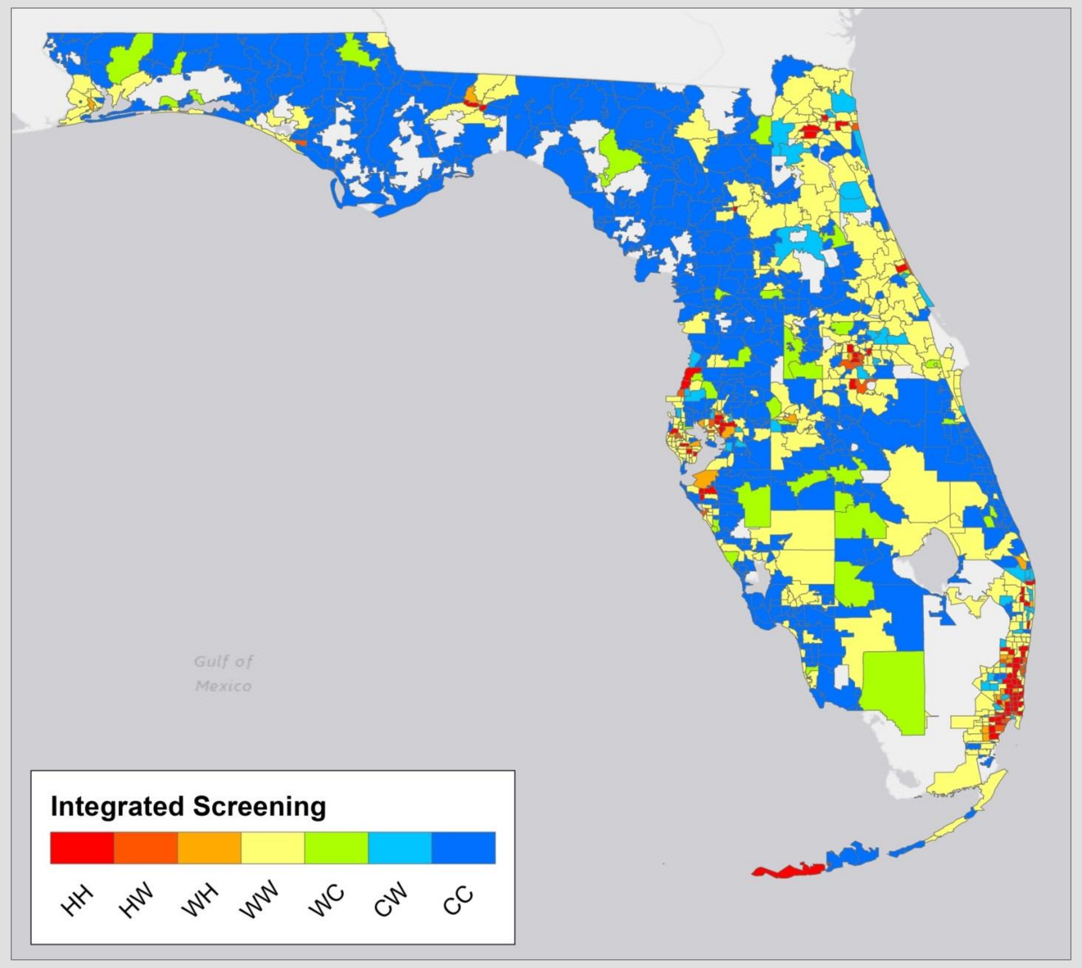
In the preceding section, hot zones for two targets: 'Pedestrian crashes per crash location ZIP' and 'Crash-involved pedestrians per residence ZIP' are identified individually. In this section, the hot zone identification results of the two targets are combined to provide a broad spectrum perspective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

All zones were again categorized according to the two scopes: location and residence, and 3 traffic safety levels: 'H', 'W', and 'C'. Therefore, there are overall 9 combination classifications: 'HH', 'HW', 'HC', 'WH', 'WW', 'WC', 'CH', 'CW', and 'CC'. The initial letter of the classifications represents the location-based pedestrian safety risk, and the latter character symbolizes the residence-based pedestrian safety risk.

Variables	Pedestrian crashes per crash location ZIP				Crash-involved pedestrians per residence ZIP			
	mean	s.d.	BCI 2.5%	BCI 97.5%	mean	s.d.	BCI 2.5%	BCI 97.5%
Intercept	-2.210	0.254	-2.599	-1.733	-3.902	0.630	-4.860	-2.560
(Log of population) × (Log of VMT)	0.036	0.002	0.032	0.038				
Log of population					0.760	0.062	0.634	0.854
Proportion of children (5-14 years)					1.804	0.495	0.982	2.877
Proportion of people working at home					-1.930	0.492	-2.940	-1.012
Proportion of households without available vehicle					1.849	0.529	0.719	2.760
Proportion of households below poverty level	2.820	0.192	2.479	3.174				
Median household income (in $1,000)					-0.013	0.001	-0.016	-0.011
Proportion of high-speed roads (55 mph or higher)	-1.161	0.089	-1.329	-0.989	-0.598	0.088	-0.767	-0.441
Number of rail and bus stations per mi²	0.035	0.017	0.001	0.068				
Number of hotels, motels, and guest houses per mi²	0.022	0.004	0.013	0.029				
Number of marina/ferry terminals per mi²	0.222	0.062	0.093	0.332				
Number of K-12 schools per mi²	0.084	0.017	0.047	0.115				
δ_1, δ_2	2.018	0.739	1.090	3.764	0.810	0.342	0.354	1.538
δ_3					-0.645	0.356	-1.540	-0.192
s.d. of φ_i	0.544	0.060	0.424	0.664	same			
DIC	11101.8							







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

2015 Road Safety & Simulation International Conference

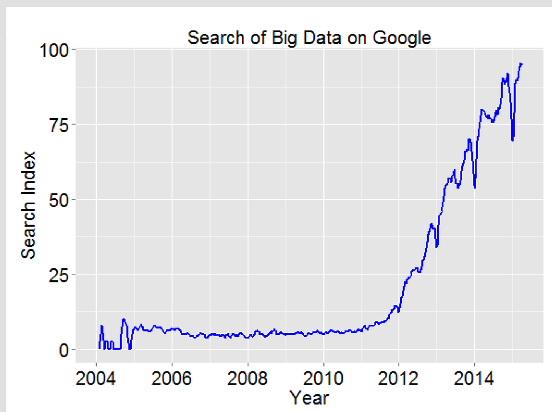
Background

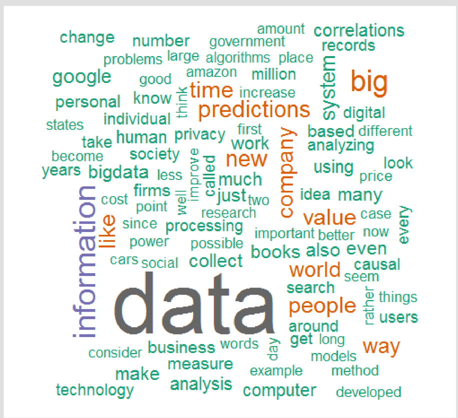
The advent of Big Data era

- Since 2010
- Key words:

  Data Information

  Big Predictions





What is Big Data?

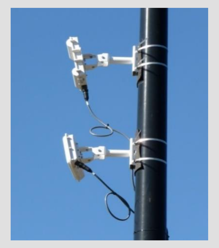
- Volume: increasing size of data
- Velocity: unprecedented streaming speed of data
- Variety: wide range of data formats
- Other dimensions of Big Data: Veracity, Variability, Complexity.

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

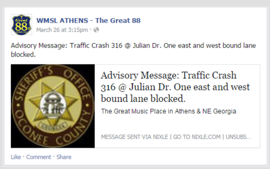
*Automatic Vehicle Detection System



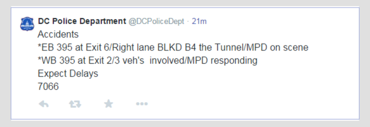
*Microwave Vehicle Detection System

Data types

- Structured data
- Unstructured data



*facebook



*twitter

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

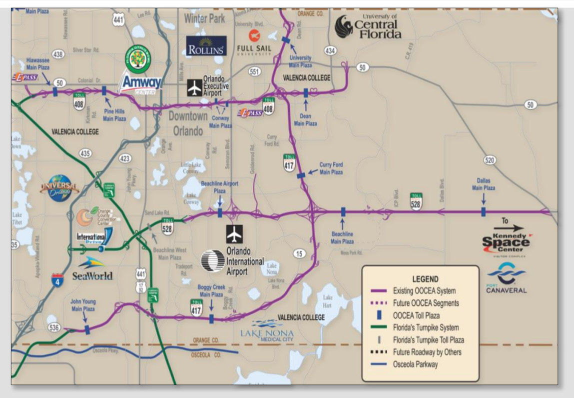
Three urban expressways;
Real-time ITS traffic data;
Crash report;

Route	Length (mi)	Direction	Mainline Detectors	Mean Distance
SR 408	21.4	EB	55	0.38
		WB	55	0.39
SR 417	31.5	NB	55	0.58
		SB	55	0.58
SR 528	22.4	EB	26	0.84
		WB	29	0.84

Data Preparation

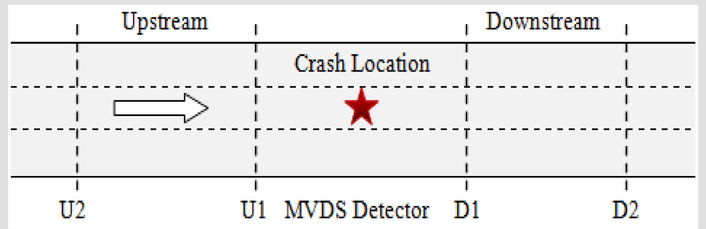
Three expressways in Central Florida area:

- 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:

$$CI = \frac{\text{free flow speed} - \text{actual speed}}{\text{free flow speed}} \text{ when } CI > 0; \quad (1)$$
$$= 0 \text{ when } CI \leq 0$$

Real-time traffic safety (rear-end crashes) evaluation:

- Random forest: importance of crash contributing factors
- Bayesian logit model: effects of crash contributing factors

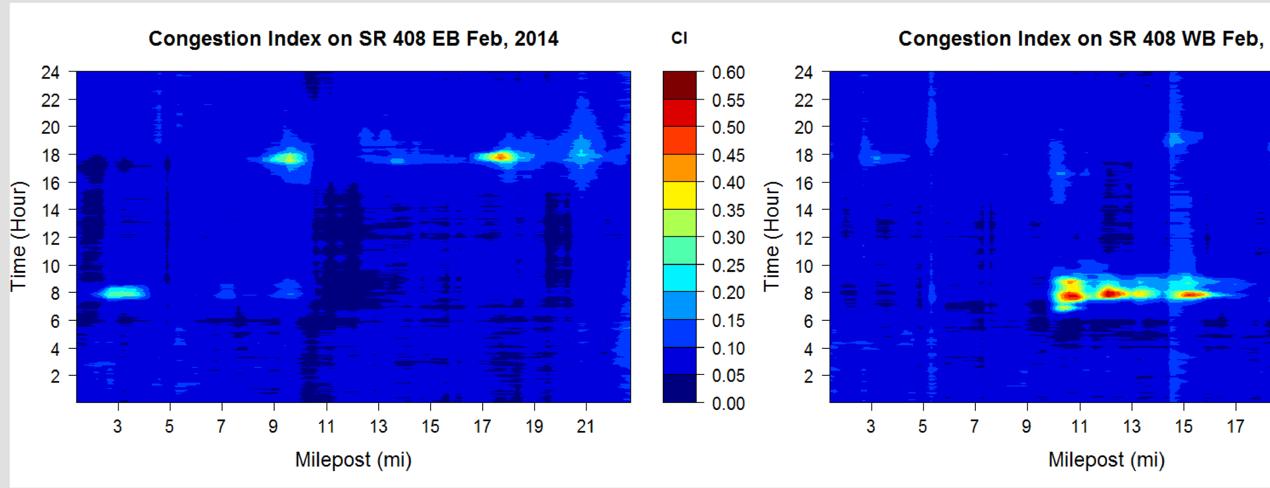
Simultaneous improvement of traffic operation and safety:

- First Order Reliability Method (FORM) Analysis

Analysis Results

Operation evaluation

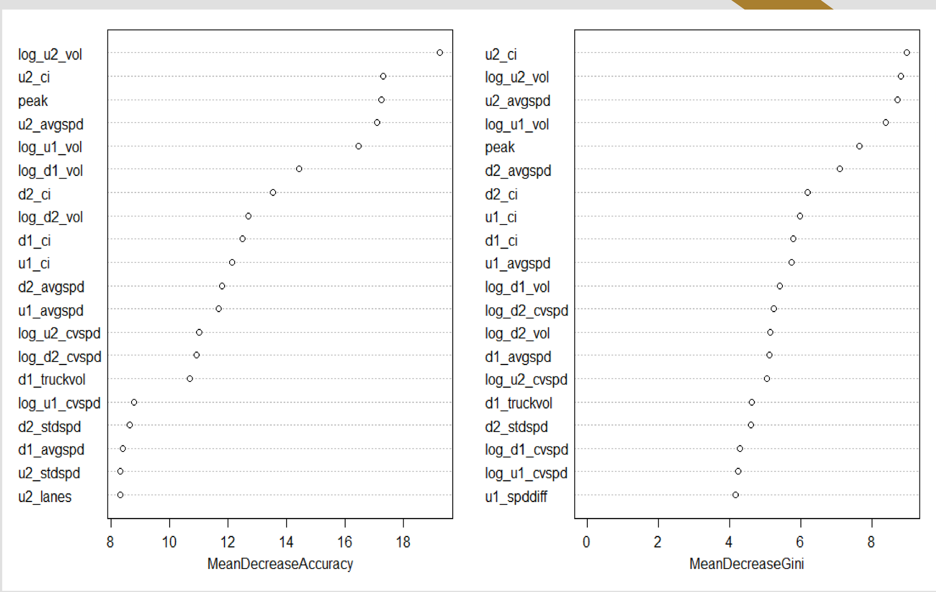
- Congestion occurrence: highly localized and time specific
- Urban area: congestion in rush hours
- Congestion intensity: varies on spatial-temporal dimensions



Real-time traffic safety evaluation

Variable selection:

- Peak hour
- Logarithm of volume
- Average speed
- Congestion index



Real-time logit model:

	Random Effect		Fixed Effect		Random Parameter	
	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI
Intercept	--	--	-1.505	(-3.487, 0.541)	-1.031[1]	(-4.176, 2.589)
					-3.315[2]	(-5.268, -0.092)
peak	1.905	(1.435,2.419)	1.857	(1.373, 2.363)	--	--
					--	--
log_u2_vol	0.275	(0.116, 0.435)	0.382	(0.171, 0.596)	0.338[1]	(0.117, 0.554)
					0.823[2]	(0.291, 1.374)
u2_avgspd	-0.057	(-0.070, -0.045)	-0.042	(-0.066, -0.016)	-0.032[1]	(-0.059, -0.002)
					-0.048[2]	(-0.087, -0.017)
d1_ci	6.053	(3.253, 9.546)	6.809	(3.658, 10.920)	7.288[1]	(3.428, 12.160)
					6.190[2]	(6.200, 10.630)
Model Estimation						
$\bar{D}$	634.211		632.975		629.562	
p_D	4.879		5.171		6.652	
DIC	639.090		638.146		636.214	
Training Data ROC	0.774		0.779		0.781	
Validation Data ROC	0.755		0.755		0.755	
[1] non-peak hours; [2] peak hours.						

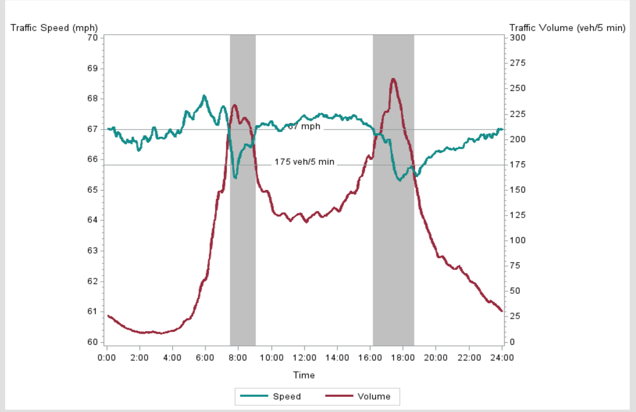
First Order Reliability Method (FORM)

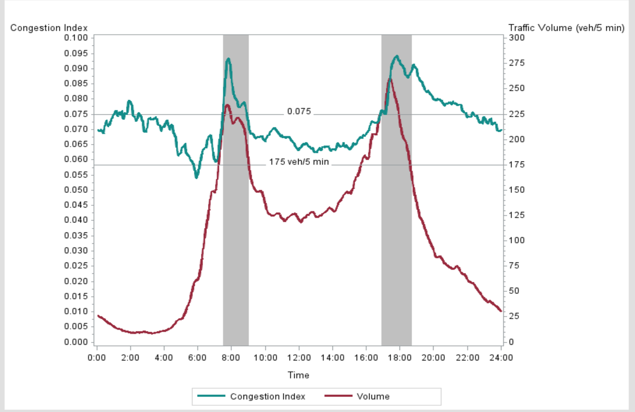
Distribution	log_u2_vol			u2_avgspd			d1_ci		
	Converged	-2 LL	Selected	Converged	-2 LL	Selected	Converged	-2 LL	Selected
Normal	Yes	496	Yes	Yes	1663	No	Yes	-76	No
Lognormal	Yes	1053	No	Yes	1757	No	Yes	-443	No
Exponential	Yes	1098	No	Yes	2058	No	Yes	-389	No
Weibull	Yes	555	No	Yes	1645	Yes	Yes	-458	Yes
Gamma	Yes	769	No	Yes	1719	No	Yes	-452	No

	Basic Statistics		Pearson Correlations			FORM Critical Point
	Mean	Std. Dev	log_u2_vol	u2_avgspd	d1_ci	
log_u2_vol	5.007	0.561	1	-0.366	0.259	5.17
u2_avgspd	67.448	4.785	-0.366	1	-0.371	67.0
d1_ci	0.071	0.034	0.259	-0.371	1	0.075

Conclusions:

		Individual detector Congestion Index (CI)	
		< 0.2	≥ 0.2
System Congestion Index (CI)	< 0.075	Safe and no congestion	congestion warning
	≥ 0.075	Prepare for safety warning	congestion and safety Warning





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

University of Central Florida

2015 Road Safety & Simulation International Conference

Background

Providing motorists with efficient and safe traffic system has long been considered a priority of traffic professionals. With the growth in traffic demand outpacing the construction of road infrastructure, congestion and safety concerns arise. In urban areas, many traffic authorities have turned to toll/turnpike 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

SR 408

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:

$$\text{TTI} = \frac{\text{actual travel time}}{\text{free flow travel time}}$$

- MVDS -- Travel speed based:

$$\text{CI} = \frac{\text{free flow speed} - \text{actual speed}}{\text{free flow speed}} \text{ if CI} > 0; \text{CI} = 0 \text{ if CI} \le 0$$

- 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: $R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$
  - Tolerance (TOL): $TOL_k = 1 - R_k^2$
  - Variance Inflation Factor (VIF): $VIF_k = \frac{1}{TOL_k}$
- Bayesian ridge regression
  - Crash frequency model

$$Y_{ijt} \sim Poisson(\lambda_{ijt})$$

  - Hierarchical data structure and random effects

$$log(\lambda_{ijt}) = a_{jt[i]} + \mathbf{X_i \beta} + \delta_1 \epsilon_{it} + \delta_2 \epsilon_i$$

$$a_{jt} = \mathbf{U_{jt} \gamma_t}$$

  - Ridge regression

$$\begin{cases} z_{jt} = \dfrac{(u_{jt} - \bar{u}_{jt})}{\text{sd}(u_{jt})} \\ \gamma_t = \dfrac{b_t}{\text{sd}(u_{jt})} \end{cases}$$

$$a_{jt} = \mathbf{Z_{jt} b_t}$$

Results and Discussion

Presence of multicollinearity:

$$VIFs < 10$$

$$R^2_{auxiliary} > R^2_{general}$$

Response: Multicollinearity should be taken into account

Peak Hours	Correlation between variables			Coefficient of Determination	TOL	VIF
	MVDS volume	Speed	CI			
MVDS volume	1.000	-0.315	0.468	0.744	0.256	3.906
Speed	-0.315	1.000	-0.669	0.826	0.174	5.736
CI	0.468	-0.669	1.000	0.847	0.153	6.544
VIF mean						1.845
R^2 general				0.458		
Non-peak Hours	Correlation between variables			Coefficient of Determination	TOL	VIF
	MVDS volume	Speed	CI			
MVDS volume	1.000	-0.172	0.009	0.693	0.307	3.253
Speed	-0.172	1.000	-0.404	0.743	0.257	3.887
CI	0.009	-0.404	1.000	0.738	0.262	3.819
VIF mean						1.620
R^2 general				0.383		

Time-specific nature of congestion on crash frequency

Variables	(1) Uncorrelated random effects			(2) Random parameter with uncorrelated random effects			(3) Random parameter with correlated random effects		
	Mean	SD	95% BCI	Mean	SD	95% BCI	Mean	SD	95% BCI
RCI level									
Intercept	0.873	0.196	(0.466, 1.250)	0.970	0.187	(0.601, 1.344)	0.947	0.241	(0.487, 1.406)
log(Length)	0.769	0.137	(0.495, 0.1.042)	0.801	0.134	(0.541, 1.082)	0.821	0.179	(0.490, 1.180)
Auxiliary lane	0.268	0.168	(-0.061, 0.595)	0.211	0.167	(-0.120, 0.535)	0.343	0.194	(-0.012,0.749)*
MVDS level									
log(MVDS volume)	0.646	0.151	(0.349, 0.948)	0.580[1]	0.352	(-0.074,1.269)	0.482[1]	0.313	(-0.097,1.120)
				0.149[2]	0.037	(0.077,0.225)	0.128[2]	0.028	(0.078,0.183)
CI	0.162	0.026	(0.112, 0.215)	0.791[1]	0.312	(0.151, 1.384)	0.671[1]	0.278	(0.129,1.211)
				0.209[2]	0.145	(-0.067,0.478)	0.092[2]	0.127	(-0.138,0.336)
Model performance									
$\bar{D}$	755.196			763.963			749.718		
p_D	124.000			113.099			90.698		
DIC	879.195			877.062			840.416		
*significant at 90% BCI									
[1] peak hours; [2] non-peak hours									

Comparison of congestion measures

Congestion Measures	Occupancy			Congestion Index (%)			Travel Time Index		
Variables	Mean	SD	95% BCI	Mean	SD	95% BCI	Mean	SD	95% BCI
RCI level									
Intercept	0.964	0.256	(0.479, 1.472)	0.947	0.241	(0.487, 1.406)	0.993	0.204	(0.580, 1.372)
log(Length)	0.842	0.180	(0.483, 1.192)	0.821	0.179	(0.490, 1.180)	0.852	0.146	(0.563, 1.143)
Auxiliary lane	0.411	0.233	(-0.059, 0.786)*	0.343	0.194	(-0.012,0.749)*	0.303	0.175	(-0.012, 0.656)*
AVI/MVDS level									
log(MVDS volume)	0.572[1]	0.342	(-0.045, 1.228)	0.482[1]	0.313	(-0.097,1.120)	--	--	--
	0.157[2]	0.045	(0.067, 0.245)*	0.128[2]	0.028	(0.078,0.183)	--	--	--
log(AVI volume)	--	--	--	--	--	--	1.015[1]	0.522	(0.025, 1.987)
	--	--	--	--	--	--	0.869[2]	1.350	(-1.742, 3.602)
Occupancy	0.572[1]	0.334	(0.024, 1.216)	--	--	--	--	--	--
	-0.031[2]	0.081	(-0.146, 0.187)	--	--	--	--	--	--
CI	--	--	--	0.671[1]	0.278	(0.129,1.211)	--	--	--
	--	--	--	0.092[2]	0.127	(-0.138,0.336)	--	--	--
TTI	--	--	--	--	--	--	0.634[1]	0.355	(0.026, 1.352)
	--	--	--	--	--	--	-1.232[2]	1.295	(-3.818, 1.216)
Model performance									
$\bar{D}$	751.326			749.718			754.224		
p_D	94.147			90.698			93.212		
DIC	845.473			840.416			847.436		
*significant at 90% BCI									
[1]peak hours; [2]non-peak hours									

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