Urban Freeway Fatality Analysis

Subasish Das, Ph.D. (Co-PI)
Karen Dixon, Ph.D., P.E. (Co-PI)
September 1, 2016

Table of Contents

Page

List of Tables

List of Figures

EXECUTIVE SUMMARY

Traffic safety is a major concern because of the economic and societal costs of traffic crashes. Improving the safety of road users remains a top priority of transportation and roadway safety planners. The vehicle miles traveled (VMT) on urban freeways contributes to 15 percent of the national VMT with a contribution of 5 percent of the fatal crashes. Texas and California are the second and third most extensive states by area respectively. Although the population of the licensed drivers and the VMT of California are higher than Texas, the urban freeway fatality rate of Texas is nearly double of the fatality rate of California in 2014. The research question of this study is straightforward: why are Texas urban freeways more prone to fatal crashes than California urban freeways? This study used Fatality Analysis Reporting System (FARS) data to develop a generalized logistic regression model to determine the safety outcomes of urban freeways in Texas. The research team also used speed detector data to reveal the incident clearance time for both of the states. The findings from this study are: odds of involvement of a fatal crash on Texas urban freeway is lower with increase of speed, increase of lanes is beneficial for the safety improvement on Texas urban freeways, and odds of peak hour, and straightly aligned roadways on Texas urban roadways are higher. The analysis on speed detector shows that Texas usually maintains lower incident clearance time than California. The findings of this study shed some lights on the research question.

CHAPTER 1

Background

Texas and California are the second and third most extensive states by area respectively. Although the population of the licensed drivers and the VMT of California are higher than Texas, the urban freeway fatality rate of Texas is significantly higher than California. Hence, there comes the obvious research question: why are Texas urban freeways more prone to fatal crashes than California urban freeways? This study attempts to answer this research question. This study aims at identifying patterns in the urban freeway fatal crash data from California and Texas that potentially precede the potential contributing factors.

To answer the research question, the research team used FARS crash data to develop a crash outcome prediction model. As crashes on the urban freeways usually associate with speed, the research team also used speed detector data to explore speed difference, queue time, and incident clearance time.

Report Organization

The reports is organized as follows: Chapter 1 describes the background of the study, Chapter 2 presents a detailed literature review on urban freeway crash analysis, Chapter 3 describes the methodology developed in this research, and Chapter 4 presents the research findings.

CHAPTER 2

This chapter summarizes the literature review on urban freeway crash analysis. As limited research conducted on the fatality analysis on urban freeways, the research team conducted a broader literature review by considering crash analysis on urban freeways.

Literature Review

CRASH ANALYSIS ON OPERATIONAL FEATURES

There has been an abundance of research that encounters crash analysis on urban freeways. Ullman and Ogden *(1)* studied the major freeway incidents between 1986 and 1992 in Houston, Texas. Their study found that the majority of the major freeway incidents involved trucks. In their study Clark and Cushing *(2)* studied the effect of population density, VMT per capita, presence of a trauma system and the southern location of a state on the fatality rate in rural and urban areas. Their study found that the urban mortality rates were not dependent on the population density but urban fatality rates were higher in the south. Kweon *(3)* highlighted the shortcomings of the existing methods of directly comparing the raw fatality rates between the states. The study considered various influential factors like percentages of rural VMT, poverty, interstate highway lane miles, seat belt use, consumption of beer and wine. The random intercept models were developed and the factors for all of the states were tuned using Virginia as reference.

Another group of researchers focused on the effects of the changes in operational, geometric characteristics and the effects of the weather on the crashes on the highways. Haselton et al. *(4)* studied the effect of the speed limit changes on the traffic collisions in California due to the speed limit changes after 1995. This analysis used observational before-after methodologies developed by Ezra Hauer. The study found that there is an increase in the nighttime collisions for the 55-65 mph group. Bauer et al. *(5)* analyzed the projects involving narrower lanes or shoulder conversions on existing urban freeways in California with four or five lanes in one direction of travel. The safety impacts of these projects were analyzed by observational before and after study using Empirical Bayes (EB) method. The evaluation found that converting the four lanes to five lanes resulted in increase of 10% to 11% in crash frequency. The projects converting five lanes to six lanes resulted in a lower increase in the percentage of accidents.

Kopelias et al. *(6)* evaluated the relationship between the number of crashes and the severity of crashes with various geometric, operational and weather effects on the urban freeway. The study concluded that majority of the crashes were attributable to the driver behavior. The study found a negative correlation of both number as well as severity of crashes with speed, existence of merge-diverge pavement markings, downhill grade sections, and curves. The study also reported that the fixed and temporal roadway characteristics can only explain about 5 to 10% of the crashes and severity of crashes. Kononov et al. *(7)* used the neural networks to find the relationship between safety and exposure on urban freeways. The study found that the roadway safety decreases with the degrading level of service (LOS). The study found that the addition of lanes to highway only results in a temporary safety improvement and as the congestion increases there is no net safety benefit. The study also indicates that as AADT increases the number of accidents increase at a faster rate when compared to a freeway with fewer lanes. Fitzpatrick et al. (8) developed the crash modification factors (CMFs) for median characteristics on urban and rural freeways and on rural multilane highways. A negative binomial regression model was used to determine the effects of the independent variables on crashes. The variables considered in the base model included ADT, left-shoulder width, barrier offset, median (with shoulder) width, and pole density.

Park et al. *(9)* evaluated the effect of two geometric elements: the ramp density and horizontal curve on the roadway safety. The negative binomial regression model was used to study the effects of the independent variables on crashes. The study found that crashes on freeway segments were associated with average daily traffic, on-ramp density, degree of curvature, median width with inside shoulder, the number of lanes (for urban freeways), and whether the freeway is an urban or rural area. Kononov et al. *(10)* studied the relationship between flow density, speed and crash rate on selected freeways in Colorado. The study found that the crash rate initially remains constant until a critical threshold combination of speed and density is reached and increases rapidly beyond this point. The paper calibrates the performance functions for corridor-specific safety that relate crash rate to hourly volume-density and speed. Harwood et al. *(11)* developed the relationship between safety and congestion using the traffic volume, speed and crash data by 15-min periods on the urban freeways in Seattle and Minnesota. The data was collected on 564 urban freeway segments with the traffic volume and speed data for the individual lanes. Their study showed a U-shaped relationship between the crash rate per million vehicle miles of travel and traffic density. The highest crash rates were found at low and high traffic densities and the lowest crash rates in the middle range of traffic densities. The study found that the high crash rate at lower traffic densities represent predominantly single-vehicle crashes whereas the high crash rates at higher traffic densities represent multiple-vehicle crashes. Jackson and Sharif *(12)* examined the temporal and spatial distribution patterns of the rain-related fatal crashes in Texas from 1982 to 2011. The data used for the study was obtained from Fatality Analysis Reporting System (FARS). The study used Getis-Ord statistic to identify spatial clustering patterns of rain-related fatal crashes and their correlation with rainfall and compare them to spatial patterns of other crashes. The results suggest that rain is a contributing factor to the crashes in a few counties but at less than 95% confidence interval in some wetter counties. Imprialou et al. *(13)* used a new data aggregation called the condition-based approach where the crashes are grouped together based on the pre-crash traffic and geometric conditions. The study compared the results of the new approach with the traditional link-based approach. Mulokozi and Teng *(14)* studied the effect of the geometric features on the freeway crashes by accounting for the effects of unobserved factors. Their study found that more number of vehicles and short segments increased crash frequency while a wider right shoulder decreased crash frequency. The study also found that the weaving segments have decreased crash frequency compared to the non –weaving segments. Yu et al. (15) studied the effect of roadway lighting facilities on the roadway safety on all the freeways and expressways in Dane county, Wisconsin. The night time crash data for a 5 year period on all the freeways, expressways and interchanges was collected along with the light pole, traffic and intersection data. The study revealed that number of lanes was significant factor for the interchange ramps while the roughness index was a significant factor for all segments. The study found that the entry ramps had higher number of crashes than exit ramps under unlighted conditions. The study recommends installation of lighting on interchange segments, entry ramps or exit ramps with AADT greater than 10,000 vehicles per day.

ANALYSIS ON SPEED DETECTORS

The speed detector data has been used by several researchers for the crash prediction research. Xu et al. *(16)* used the loop detector data to predict the likelihood of crash at different levels of severity. Their study used the sequential logit model to link the likelihood of crashes at different severity levels to the traffic flow characteristics derived from detector data collected on I-880 freeway in California. Their study showed that PDO crashes were more likely to occur under congested condition with highly variable speed and frequent lane changes while the other type of crashes KA & BC occur under less congested traffic flow conditions. The KA crashes were more likely to occur due to high speed coupled with large speed difference between adjacent lanes under uncongested conditions. Yu et al. *(17)* used the speed detector data and the crash data to study the crash occurrence mechanism. Their study used the Bayesian semi-parametric inference technique to do the crash risk analysis. Their study considered three different study periods namely weekday peak hour crashes, weekday non-peak hour crashes, and weekend non-peak hour crashes to investigate different crash occurrence scenarios. Wang et al. *(18)* identified the secondary crashes on the freeways that occur due to a primary accident. The speed detector data was used to identify the traffic shock waves and define the spatiotemporal boundaries of a primary accident. Their study showed that secondary accidents accounts for 1.08% of California interstate freeway accidents. Kwak and Kho *(19)* developed real time crash risk prediction models for different segment types and traffic flow states on the expressways. This study showed that the traffic characteristics leading to the crash varied by the segment type and traffic flow state. Levine *(21)* reviewed the Houston-Galveston area council metropolitan traffic safety-planning program. The main goals of this program are to identify and monitor safety, implement safety roadway improvements and to support other safety efforts.

INCORPORATING TRANSPORTATION PLANNING PROCESS

Several researchers focused to incorporate urban freeway crash analysis with transportation planning process. In their study, Kononov et al. (*22)* pointed out the shortcomings of using the crash rates directly in the transportation planning process. This study introduced a two phase process consisting of safety assessment of a facility and evaluating the multiple design alternatives to enhance safety. Voigt et al. *(23)* studied the impacts of the usage of dual-advisory speeds for providing different advisory speeds for the trucks and passenger vehicles. This research found that the average and 85th percentile speed at the midpoint of each study curve, the dual-advisory warning sign had a positive effect on reducing the speeds at the point of curvature on the curve and had an accompanying reduction in speed related crashes. Geedipally et al. *(24)*

evaluated the TxDOT's Safety Improvement Index for identifying, ranking and selecting the eligible projects. Park et al. *(25)* evaluated the safety effectiveness of Super 2 highways in Texas. The before-after study was done using the empirical Bayes (EB) method. The data used for this project was the roadway inventory and crash history data for 12 years period (1997 to 2001 and 2003 to 2009). Their study found a statistically significant reduction of crashes on the installation of super 2 highways. Viallon *(26)* studied the trends of the fatal crashes caused due to speeding in France. Their study found that there is a decrease in the number of crashes due to speeding from 2001 to 2010. This shows that the crash countermeasures have been very effective. Burger et al. *(27)* evaluated the effect of the cell phone ban in California on the crashes. The study used high-frequency data and a regression discontinuity design. The study concluded that there was no effect of the ban on hand-held cell phone use on the accidents.

The current study contributes to the existing urban freeway fatality analysis by comparing the results from two major states: California and Texas. This study used both FARS crash data and speed detector data to investigate the safety outcome of urban freeways and the incident clearance time.

CHAPTER 3

This chapter summarizes the methodologies applied in this study. The study collected urban freeway fatal crash data from 2005-2014 for California and Texas from the Fatality Analysis Reporting System (FARS) *(28)*. The research team applied generalized mixed effect logistic regression on the final dataset to investigate the safety outcome. This study also collected speed data from INRIX® *(29)* speed detectors from California and Texas to perform an analysis on incident clearance.

Analysis on FARS data

DATA COLLECTION

Crash data from urban freeway highways in California and Texas were collected for a 10-year period (2005–2014 inclusive) from the FARS to perform this analysis. Figure 1 shows the fatal crashes on urban arterial interstate roadways in California and Texas.

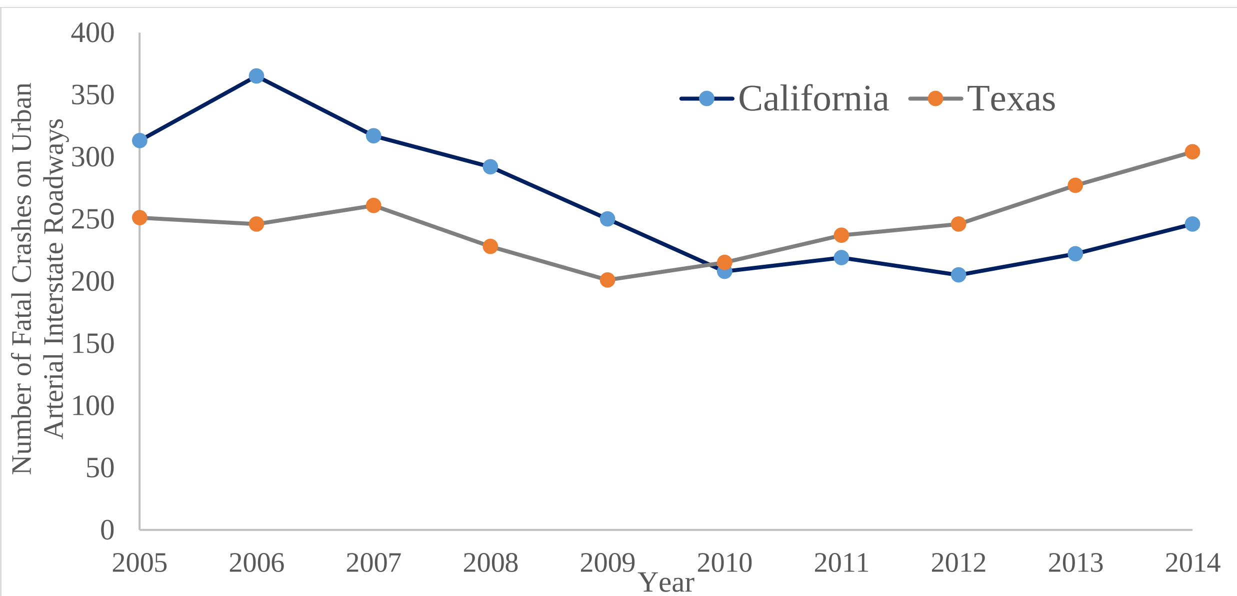

Figure 1. Fatal crashes on urban freeways in California and Texas

Figure 1 shows an upward trend of fatal crashes in Texas from 2009. Another way to illustrate potential trends is to calculate fatal crash rates by incorporating some exposures like number of licensed drivers and vehicle miles traveled (VMT) on urban freeways. As noted by Hauer *(30)*, this normalization equalizes differences in intensity of use, thus making comparisons more meaningful, and can help identify differences between different populations' characteristic crash rates. In addition to the FARS data, two other secondary data sources were used to obtain the total number of licensed drivers and VMT on urban freeways *(31-32)*. Using both sources of data, the research team calculated fatal crashes on urban freeways per 100,000 licensed drivers (shown in Figure 2), and fatal crashes on urban freeways per 100 million VMT (shown in Figure 3). These plots clearly indicate that Texas urban freeways have significant safety problems.

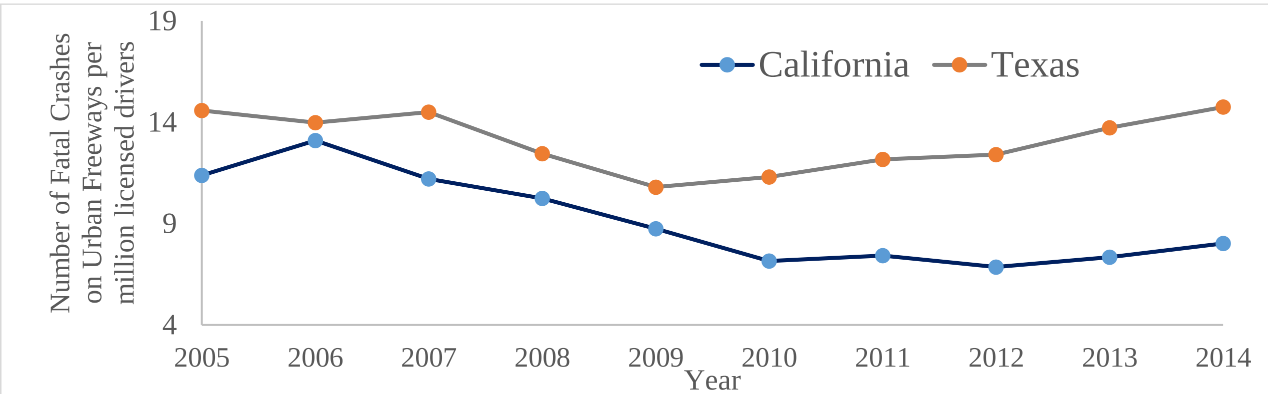

Figure 2. Fatal crashes on urban freeways per million licensed drivers

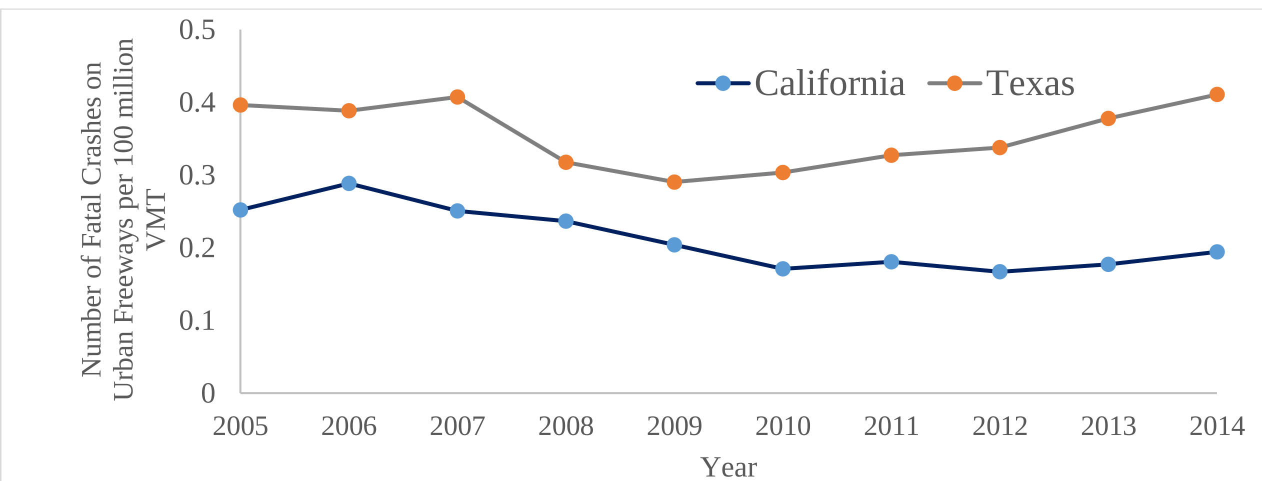

Figure 3. Fatal crashes on urban freeways per 100 million VMT

DESCRIPTIVE STATISTICS

Crash data from urban freeway highways in California and Texas were collected for a 10-year period (2005-2014) from the FARS to perform the analysis. This study emphasizes on the roadway geometrics and crash related information. The outcome of a crash occurrence involves interactions between many factors, including the roadway, the environment, driver, and traffic characteristics. A preliminary data exploration was first conducted to examine the contributing factors that may contribute to crash occurrence on urban freeways. Table 1 lists the percentage distribution of the attributes. The percentage distribution analysis seems to suggest that significant differences were visible for the variables like traffic flow condition, number of lanes, speed limit, and pavement types. Nighttime collisions were higher in percentage for both states, which is consistent with the research findings of Haselton et al. *(4).*

Table 1. Percentages of the attributes

Category	Percentage		Category	Percentage	
Traffic Flow Condition	**California**	**Texas**	**Lighting Condition**	**California**	**Texas**
Divided	86.04%	89.74%	Daylight	32.99%	32.27%
Undivided	0.20%	2.01%	Dark	63.44%	65.65%
Ramp	13.76%	8.25%	Dawn/Dusk	3.57%	2.08%
Number of Lanes			**Weather Condition**		
One lane	6.30%	6.29%	Clear	87.01%	86.05%
Two lanes	14.03%	29.76%	Rain	5.31%	4.91%
Three lanes	14.59%	28.91%	Snow/Sleet	0.00%	0.56%
Four lanes	32.04%	22.78%	Fog	0.87%	0.48%
Five lanes	21.35%	9.67%	Others	6.81%	8.00%
Six or more lanes	11.69%	2.60%	**Hour Category**		
Speed Limit (mph)			Peak	22.18%	22.70%
40-60	10.80%	68.89%	Off-Peak	50.50%	51.54%
60-65	78.80%	20.72%	Moderate	27.32%	25.76%
65-70	10.40%	6.35%	**Manners of Collision**		
70-75	0.00%	3.36%	Angle	5.70%	8.55%
75-80	0.00%	0.68%	Head-On	2.88%	4.74%
Alignment			Rear-end	19.75%	18.00%
Curve	18.40%	15.29%	Sideswipe	8.11%	3.30%
Straight	81.60%	84.71%	Single Vehicle	63.56%	65.41%
Pavement Type			**Number of Occupants**		
Blacktop	39.89%	24.97%	Single	63.13%	66.13%
Concrete	56.67%	29.28%	Two	19.15%	19.26%
Others	3.44%	45.75%	Three	7.13%	5.90%
			Four or more	10.59%	8.71%

Figure 4 shows the speed distribution of urban freeways in California and Texas. It shows that Texas urban roadways involved higher percentage of crashes on roadways with speed limit lower than 60 mph, which requires more investigation.

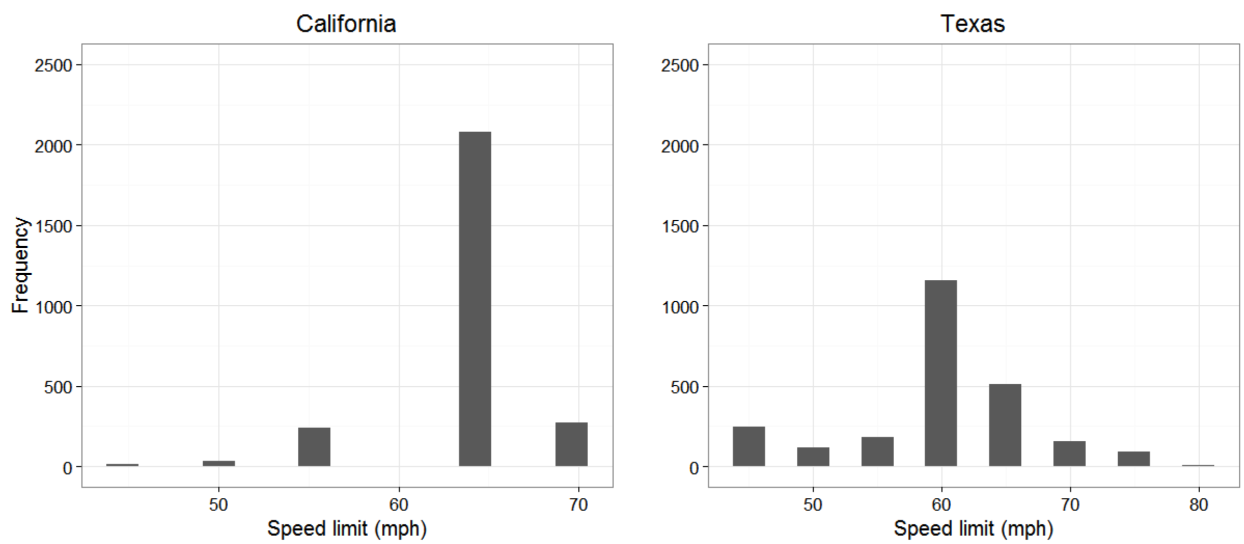

Figure 4. Speed distribution on urban freeways in California and Texas

From Figure 5, it is seen that higher number of fatal crashes happened on low speed Texas urban roadways (speed in between 40 mph to 60 mph) with two lanes, three lanes, four lanes, and six or more lanes.

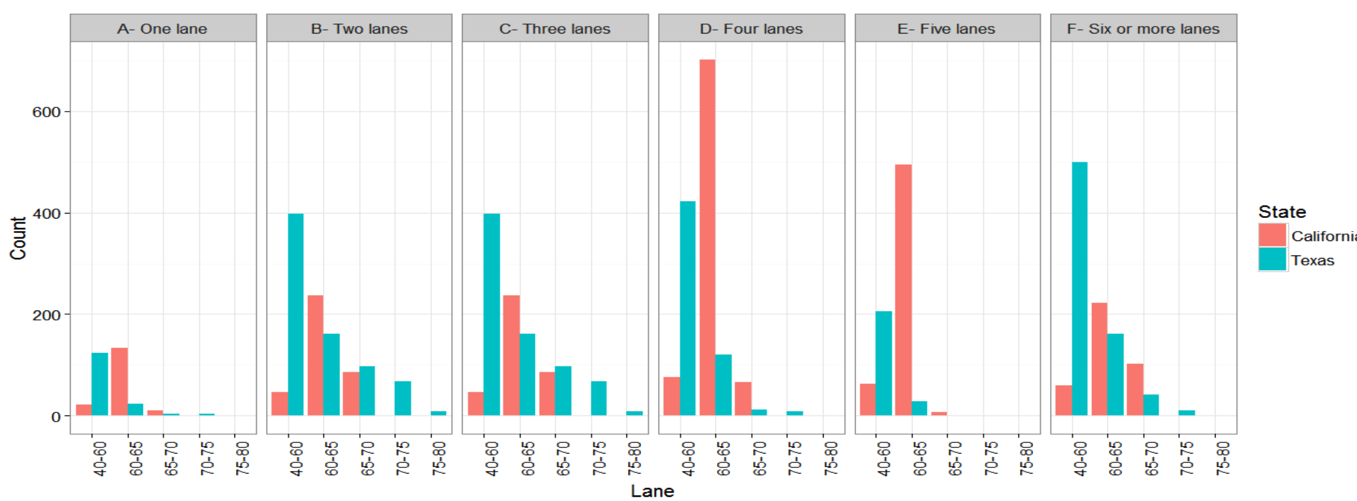

Figure 5. Speed distribution for urban freeways with different number of lanes

Figure 6 illustrates the involvement of number of lanes for two facility types: ramp, and undivided roadways. It shows that California was involved with more ramp-based crashes than Texas. On the other hand, undivided two lane urban roadways in Texas showed higher number of crash involvement than California.

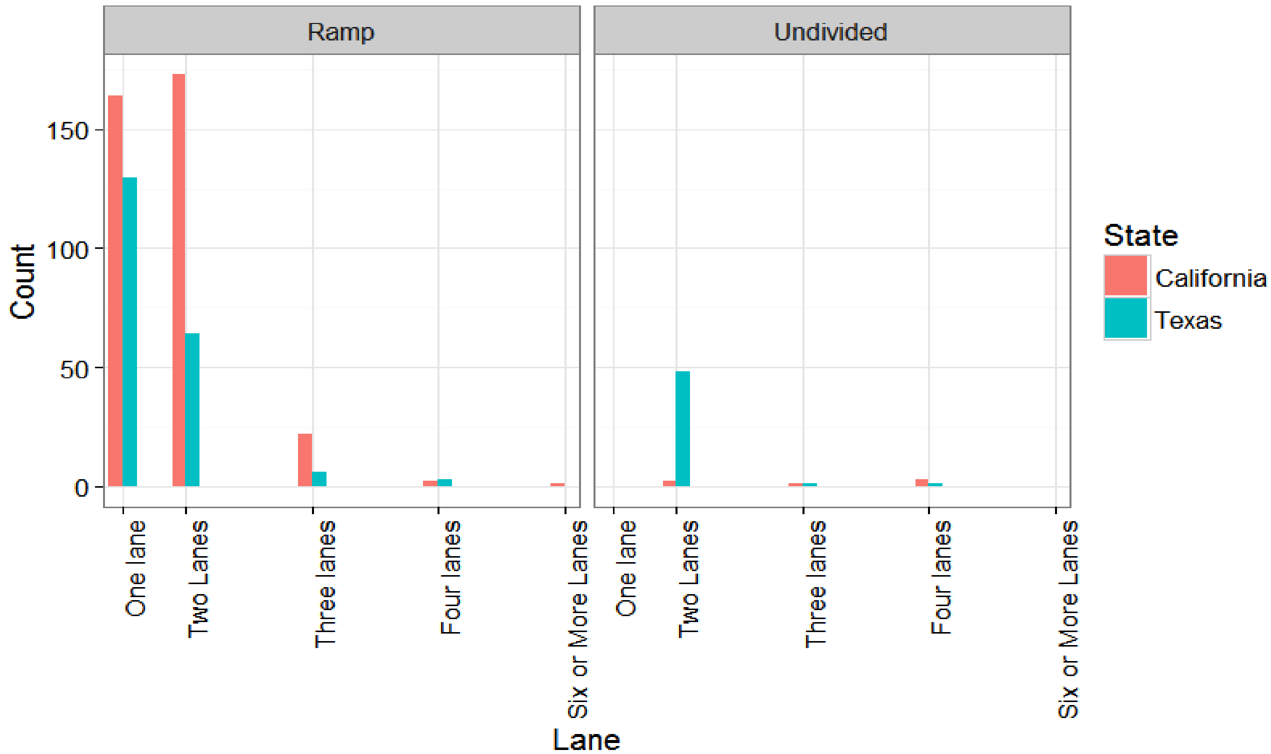

Figure 6. Involvement of number of lanes for different facility types

Mosaic plots are helpful to further depict the relationship of different variables to the probability that a fatal crash on urban freeway will involve California or Texas. Figure 8 depicts the relationship of traffic flow conditions. Note that the width of each bar represents the number of observations in a mosaic plot. For example, ramps were associated with more crashes than undivided roadways, but the probability that those crashes involving California were more than for divided or undivided roadways. Note that small sample sizes at the undivided roadways result in a less stable observation.

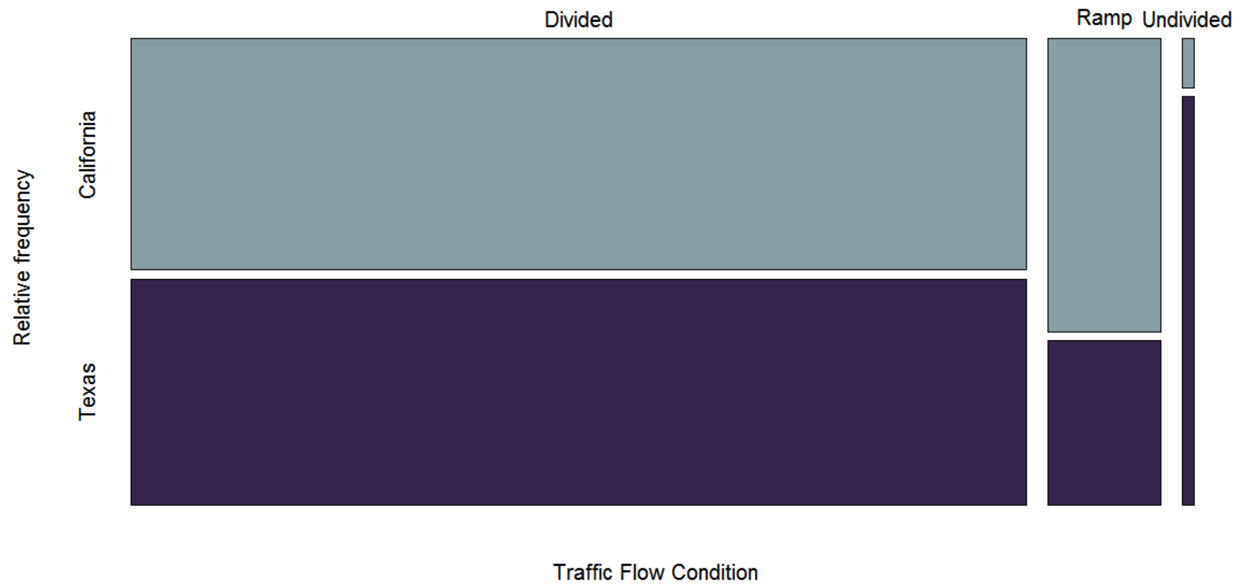

Figure 7. Mosaic plot for traffic flow condition

Figure 8 illustrates the mosaic plot for manners of collision. Single vehicle or run-off-road (ROR) crashes were associated with more counts than other manners of collisions. The probability of sideswipe crashes involving California was higher than other manners of collisions. Figure 9 illustrates the mosaic plot for hours of the day. Off-peak crashes were associated with more counts than other hours of the day.

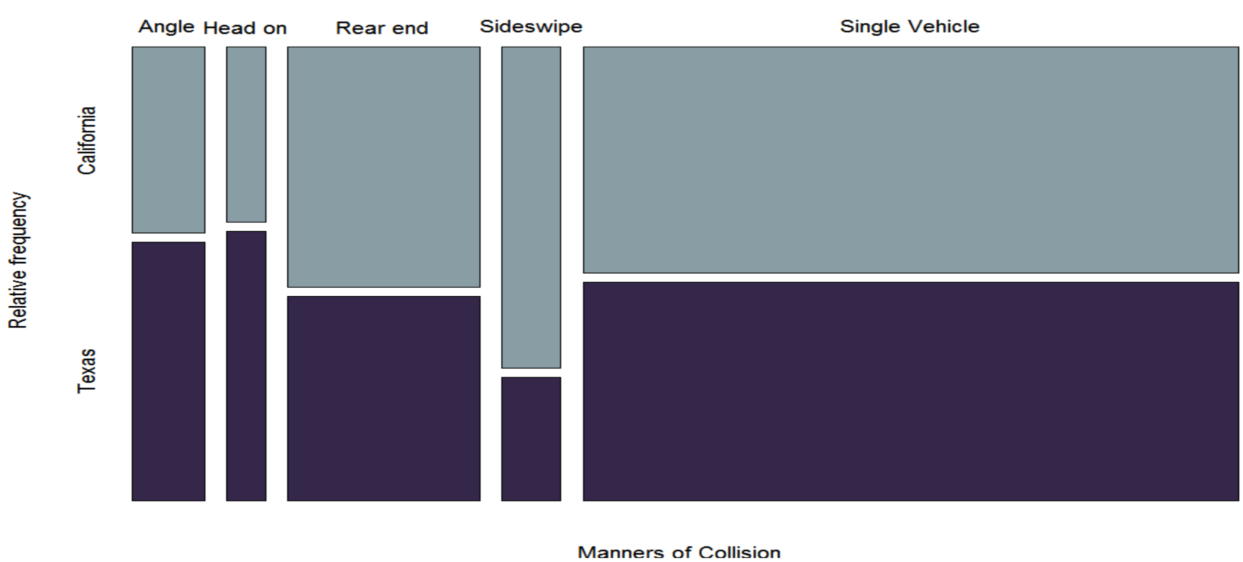

Figure 8. Mosaic plot for manners of collision

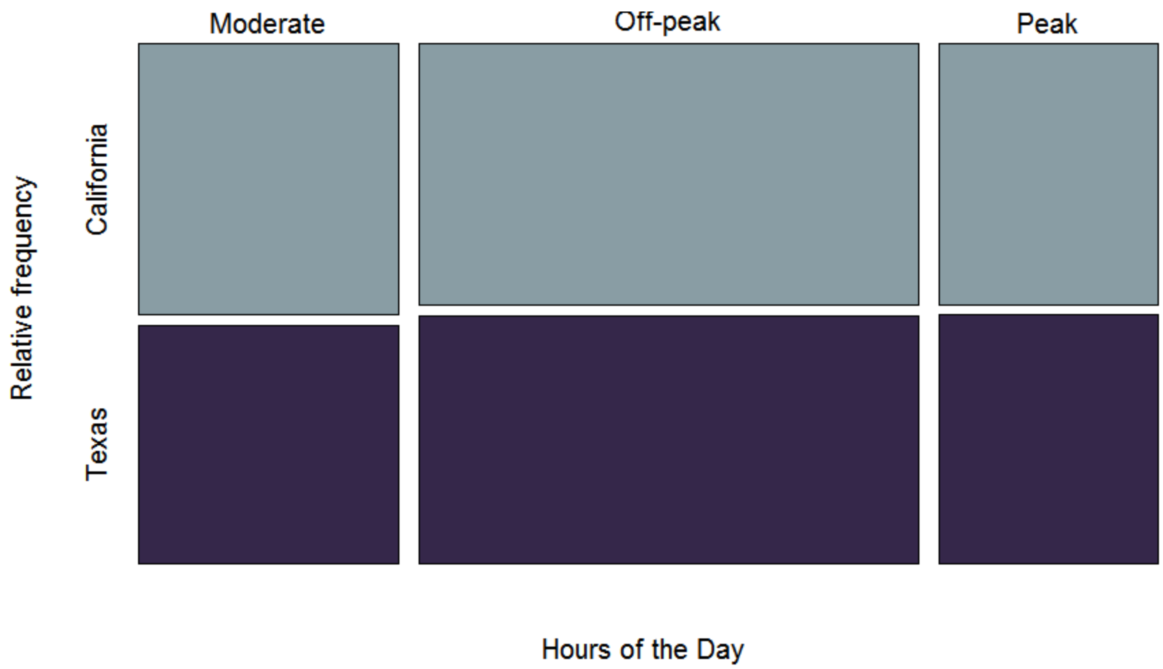

Figure 9. Mosaic plot for hours of the day

Figure 10 shows the mosaic plot for different lighting condition. Dark lighting conditions were associated with more crashes than other lighting conditions. The probability of dawn/dusk crashes in California was higher than the lighting condition either dark or daylight.

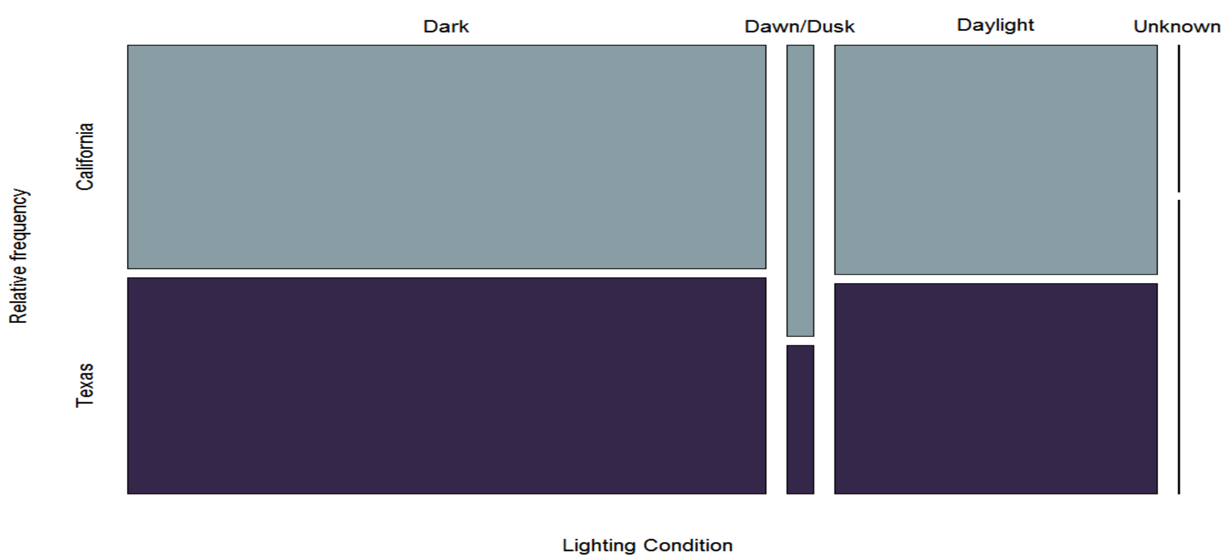

Figure 10. Mosaic plot for lighting condition

VARIABLE IMPORTANCE

The importance of the variables can be ranked by using random forest (RF) algorithms. RF method is based on the bagging principle *(33)* and random subspace method *(34)* that relies on constructing a collection of decision trees with random predictors. The general architecture of the RF is described in Figure 11 and explained as follows:

i. Generate cluster sample of size N_C from the overall data N, to grow a $tree_B$ by randomly selecting the predictors $X = \{x_i, i = 1, \dots, I\}$.

ii. Use the predictor x_i at the node n of the $tree_B$ to vote for class label k_B at this node. At each node only one predictor providing the best split is selected.

iii. Run the out-of-bag data $N - N_C$ down the $tree_B$ to obtain the misclassification rate, OOB_B

iv. Repeat i.-iii. for large number of trees till the minimum out-of-bag (OOB) error rate, is obtained.

v. Assign each observation to a final class $k$by a majority vote by averaging over the set of trees.

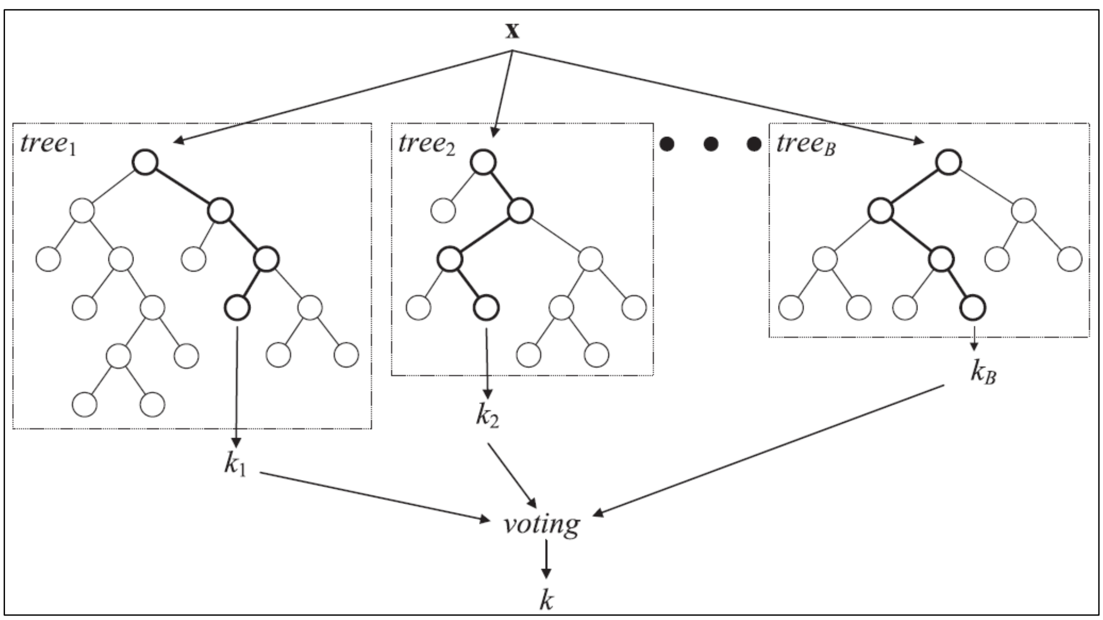

Figure 11. Forest architecture (Source: 33)

The two important byproducts of the RF method are OOB, and variable importance measures. OOB is the misclassification rate, which decreases as the number of trees increase. The number of trees does not affect to overfitting the data, therefore large number of trees can be used. To decrease the bias and the correlation, the trees are grown to the maximum depth.

Variable importance ranking is measured by the classification accuracy and Gini impurity. This importance measure shows how much the mean squared error or the “impurity” increases when the specified variable is randomly permuted. If prediction error does not change by permuting the variable then the importance measures will not be altered significantly which in turn will change the mean squared error (MSE) of the variable only slightly (low values). This implies that the specified variable is not important. On the contrary if the MSE significantly decreases during the permutation of the variable then the variable is deemed as important. The classification accuracy measure, MDA of the variable x_i is averaged over the number of trees, B, used to construct the random forest:

$$MDA(x_i) = \frac{\sum_{tree=1}^{B} MDA^{tree}(x_i)}{B} \tag{1}$$

Where:

$MDA^{tree}(x_i)$ =the importance rate of the variable x_i at $tree = tree_{b=\{1,\dots,B\}}$.

The mean decrease in Gini impurity computes the contribution of the variable to the homogeneity of the nodes and leaves in the resulting RF. The Gini coefficient is a measure of homogeneity from 0 (homogeneous) to 1 (heterogeneous). Gini impurity of the variable x_i at the node n is:

$$MDG^n(x_i) = 1 - \sum_{k=1}^{K} p^2(k|n) \tag{2}$$

Where:

$p^2(k|n)$ = the probability of class k at the node n (weights)

K = the number of classes.

Each time a specified variable is used to split a node, the Gini coefficient for the child nodes are calculated and compared to that of the parent node. Usually, after the split of a node, the impurity in the child node becomes smaller than the parent node. The changes in Gini are summed for each variable and normalized at the end of the calculation. Summing up the Gini impurity measures for each variable all over the trees gives the importance rate, which is often consistent with the permutation importance measure *(33)*, thus the variable with the higher impurity is deemed as more important.

Figure 12 shows the variable-importance plot for the final dataset by keeping the probability of an urban freeway crash end in either Texas or California as the response variable. It is seen that pavement type, number of lanes, speed limit, and traffic flow condition are four major important variables.

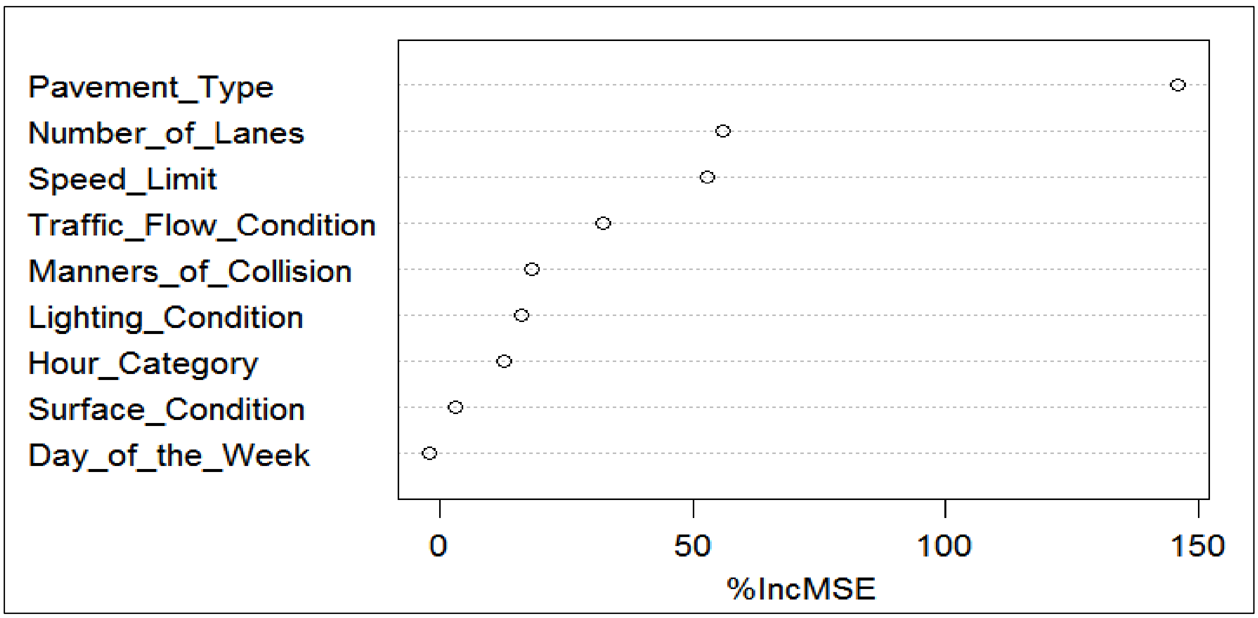

Figure 12. Variable importance plot

GENERALIZED LOGISTIC REGRESSION MODEL

To model binary outcome variables (involvement of a fatal urban freeway crash in either California or Texas), the research team used a mixed effect modeling for the model development as the data is considered to be clustered by having both fixed and random effects. The generalized logistic model (shown in Figure 13) has been widely applied to analyze crash involvement because it has advantage on accommodating discrete unobserved heterogeneity by allowing parameters to differ across observations by providing more reliable parameter estimates. The primary objective of developing a logistic regression model is to elucidate the association between California and Texas urban freeway fatal crashes and their potential contributory factors to formulate efficient and targeted crash mitigating measures.

In the generalized linear models (GLM), the response variable is assumed to follow an exponential family distribution with a mean that is assumed to be some (often nonlinear) function of $x_i^T\beta$. There are three components in a GLM (35):

- *Random Component*: It specifies the probability distribution of the response variable; e.g. binomial distribution for *Y* in the binary logistic regression.

- *Fixed Component*: It refers to the explanatory variables (X_1, X_2, ... X_k) in the model, more specifically their linear combination in creating the so called linear predictor.

- *Link Function*, *η or $g(\mu)$*: It refers to the link between random and fixed components. It says how the expected value of the response relates to the linear predictor of explanatory variables; e.g., $\eta = logit(\pi)$ for logistic regression.

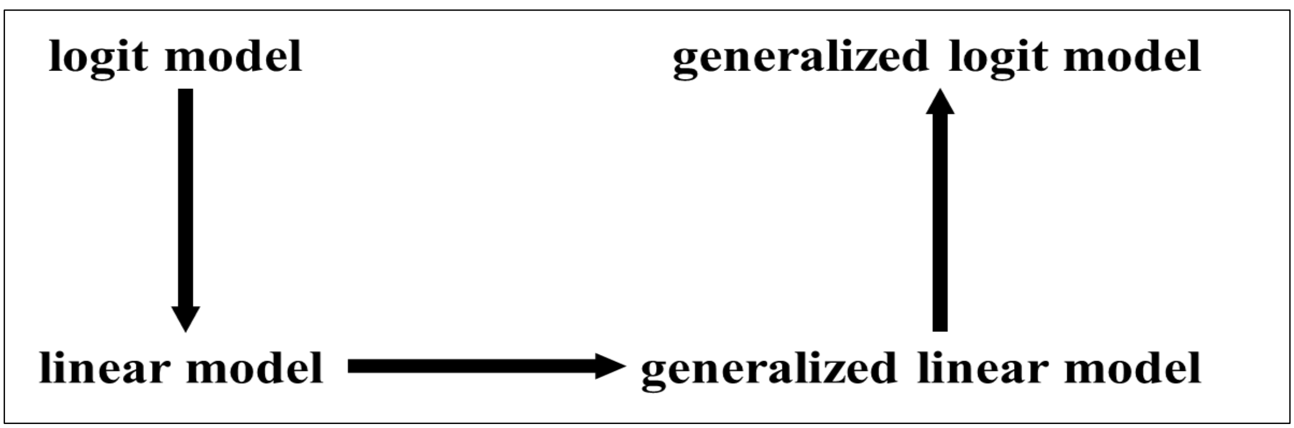

Figure 13. Generalization of the logistic model (Source: 36)

In a Generalized Logit model, the probability π_{ij} of response category *j* at subpopulation *i* is:

$$\pi_{ij} = \frac{\exp(x'_i\beta_j)}{1 + \sum_{k=1}^{J-1}\exp(x'_i\beta_j)} \quad (3)$$

where the last category *J* is assumed to be the reference category.

In terms of logits, the model can be expressed as

$$\log(\frac{\pi_{ij}}{\pi_{iJ}}) = x'_i\beta_j \quad (4)$$

for $j = 1, \ldots, J-1$.

When $J = 2$, this model is equivalent to the binary Logistic Regression model.

The outcomes of the final model are shown in Table 2.

Table 3 lists the estimates and 95% of the estimates. Odds ratio (OR) is a good measure to understand the model outcomes (listed in Table 4). An odds ratio is a relative measure of effect, which allows the comparison of the defined group (urban freeway fatal crashes in Texas) relative to the other group (urban freeway fatal crashes in California).

- OR=1 Exposure (urban freeway fatal crashes in Texas) does not affect odds of outcome (effect of urban freeway fatal crashes in Texas compared with urban freeway fatal crashes in California)
- OR>1 Exposure associated with higher odds of outcome
- OR<1 Exposure associated with lower odds of outcome

In this model, speed limit, number of lanes, lighting condition, and alignment of the roadways were found significant in the level of 95% confidence interval. For example, the negative sign for speed limit can be interpreted that for each additional five unit of speed limit, there was a reduction in the odds of fatal crash on Texas urban roadways for factor of approximately 0.842 for the combined model. For categorical variables, an attribute (from each variable) is considered as base. For number of lanes, the base attribute was urban freeways with five lanes. The odds ratio for one lane urban freeway is above 25, which means that there was an increase in the odds of fatal crash on Texas urban roadways for factor of approximately 25.55 when the base attribute was five lane roadways. The estimates suggest that addition of lanes had some safety benefits on urban freeways of Texas. This finding is consistent with the findings of Kononov et al. *(7)*. For alignment, the base attribute was curve. The odds ratio for straight urban freeway is 1.317, which means that there was an increase in the odds of fatal crash on Texas urban roadways for factor of approximately 1.317 when the base attribute was roadways with curved alignment. For hours of the day, the base attribute was moderate. The odds ratio for peak hours is 1.198, which means that there was an increase in the odds of fatal crash on Texas urban roadways for factor of approximately 1.198 when the base attribute was moderate hours of the day.

Table 2. Estimates from the Generalized Logistic Regression Model

Variables and Statistical Measures	Estimate	St. Err.	z	Pr(>|z|)
Fixed Effect				
(Intercept)	8.426	0.801	10.516	< 2.00E-16
Speed Limit	-0.172	0.008	-22.596	< 2.00E-16
Number of Lanes: Four lanes	0.536	0.100	5.364	0.000
Number of Lanes: One lane	3.241	0.231	14.017	< 2.00E-16
Number of Lanes: Six or More Lanes	-0.564	0.164	-3.445	0.001
Number of Lanes: Three Lanes	1.536	0.108	14.226	< 2.00E-16
Number of Lanes: Two Lanes	2.557	0.128	20.020	< 2.00E-16
Lighting Condition: Dawn/Dusk	-0.607	0.222	-2.734	0.006
Lighting Condition: Daylight	-0.096	0.106	-0.907	0.364
Alignment: Straight	0.275	0.096	2.859	0.004
Hours of the Day: Off-peak	0.041	0.099	0.419	0.675
Hours of the Day: Peak	0.181	0.099	1.831	0.067
Random Effect				
Traffic Flow Condition (Intercept): Variance	1.24			
Traffic Flow Condition (Intercept): Std. Err.	1.114			
Statistical Measures				
AIC	5540.9			
BIC	5632.4			
Log likelihood	-2756.5			
Deviance	5512.9			
DF Residual	5089			

Table 3. Upper and Lower limits of the model estimates

Variables	Estimate	Lower Limit	Upper Limit
(Intercept)	8.426	6.856	9.997
Speed Limit	-0.172	-0.187	-0.157
Number of Lanes: Four lanes	0.536	0.340	0.732
Number of Lanes: One lane	3.241	2.787	3.694
Number of Lanes: Six or More Lanes	-0.564	-0.885	-0.243
Number of Lanes: Three Lanes	1.536	1.324	1.748
Number of Lanes: Two Lanes	2.557	2.307	2.808
Lighting Condition: Dawn/Dusk	-0.607	-1.042	-0.172
Lighting Condition: Daylight	-0.096	-0.303	0.111
Alignment: Straight	0.275	0.087	0.464
Hours of the Day: Off-peak	0.041	-0.152	0.235
Hours of the Day: Peak	0.181	-0.013	0.374

Table 4. Odds ratio of the variables

Variables	Odds Ratio
(Intercept)	4564.262
Speed Limit	0.842
Number of Lanes: Four lanes	1.709
Number of Lanes: One lane	25.549
Number of Lanes: Six or More Lanes	0.569
Number of Lanes: Three Lanes	4.646
Number of Lanes: Two Lanes	12.900
Lighting Condition: Dawn/Dusk	0.545
Lighting Condition: Daylight	0.909
Alignment: Straight	1.317
Hours of the Day: Off-peak	1.042
Hours of the Day: Peak	1.198

Analysis on INRIX® Speed data

DATA COLLECTION

The research team extracted speed data from INRIX® that includes speed information obtained by speed detectors for the selected 26 sites. Interstate freeway crashes that occurred in California in 2012-2014 collected from the California Statewide Integrated Traffic Records System (SWITRS) *(37)* were used in this study. The research team used Texas CRIS *(38)* data for 2012-2014 to select potential crash sites. After determining some key criteria (number of lanes, Annual Average Daily Traffic (AADT), number of High Occupancy Vehicle (HOV) lanes, and HOV types), the research team randomly selected 13 urban freeway crash sites from each of the states. Table 5 lists the final selected sites.

Table 5. Selected sites for speed data collection

No.	ID	Latitude	Longitude	Highway	TIME	Crash Date	Control Site Date
1	CA001	38.77375	-121.24296	RT 80	8:11:00 AM	12/17/13	12/10/13
2	CA002	37.196	-121.07583	RT 5	8:12:00 AM	06/23/13	06/30/13
3	CA003	36.55754	-120.52294	RT 5	8:35:00 AM	04/06/13	04/20/13
4	CA004	32.67212	-117.08253	RT 805	4:49:00 PM	02/10/13	02/03/13
5	CA005	34.00935	-118.05759	RT 605	5:07:00 PM	09/08/13	09/01/13
6	CA007	33.66095	-115.826	RT 10	5:05:00 PM	04/06/14	04/13/14
7	CA008	36.06729	-120.08641	RT 5	5:10:00 PM	03/17/14	03/10/14
8	CA009	36.66879	-121.63455	RT 101	5:15:00 PM	08/15/14	08/22/14
9	CA010	38.59362	-121.28268	RT 50	5:17:00 PM	11/14/14	11/07/14
10	CA011	37.806621	-122.278271	RT980	8:35:00 PM	01/23/13	01/16/13
11	CA012	38.358114	-121.974753	RT80	8:22:00 AM	01/21/13	01/14/13
12	CA013	33.92578	-118.21238	RT 105	4:56:00 PM	07/08/13	07/15/13
13	CA014	38.531776	-121.51775	RT 5	9:20:00 AM	02/14/14	02/21/14
14	TX001	31.62080729	-106.2096755	IH0010	8:20:00 AM	01/11/13	01/18/13
15	TX002	32.76582726	-96.85499213	IH0030	5:47:00 PM	02/23/13	02/16/13
16	TX003	30.36058775	-97.68656056	IH0035	8:35:00 PM	04/18/13	04/25/13
17	TX004	29.68097004	-95.45814572	IH0610	8:30:00 AM	08/25/13	08/18/13
18	TX005	31.60305046	-97.10705084	IH0035	5:24:00 PM	10/21/13	10/28/13
19	TX006	29.70536673	-95.27402301	IH0610	8:14:00 AM	01/13/14	01/20/14
20	TX007	29.34810357	-98.62771119	IH0410	5:50:00 PM	02/16/14	02/23/14
21	TX008	32.73289988	-97.38408321	IH0030	5:00:00 PM	04/08/14	04/15/14
22	TX009	32.7045175	-96.58369787	IH0020	7:58:00 AM	09/16/14	09/09/14
23	TX010	31.68626924	-106.2672505	IH0010	5:44:00 PM	10/19/14	10/12/14
24	TX011	32.979874	-96.928645	IH35E	7:39:00 PM	07/06/13	07/13/13
25	TX012	29.784614	-95.572427	IH10 (KATY)	11:30:00 PM	10/03/14	09/26/14
26	TX013	32.922056	-96.899851	IH35E	10:30:00 AM	10/28/13	11/04/13

The measurement interval is 5 min resulting to a dataset of approximately 112,827 observations. The preliminary dataset was divided into two broader sections:

- Dataset Type 1: Site based loop detector information
- Dataset Type 2: Speed data for 5 minute interval for the loop-detectors (3 or 4 hours before and after the crash incidence)

The variables in each of the dataset types are listed in Table 6.

Table 6. Variables in speed detector data

Dataset Type 1	Dataset Type 2
Code of Speed Detector	Code of Speed Detector
Roadway	Measurement Timestamp
Direction	5 Minute interval Speed (mph)
State	Average Speed (mph)
Start Latitude	Reference Speed (mph)
Start Longitude	Travel Time (in minutes)
End Latitude	Confidence Score
End Longitude	C Value
Distance (in miles)	
Road Order	

The main objective of the speed data collection is to visualize the patterns of the changes of the speed for different spatio-temporal patterns in both of the states. Table 7 lists the geometric characteristics of the selected sites. The research team determined ten different groups (Group 1 to Group 10) using both of the states based on the similar nature of the geometric features. The other six sites have distinct geometric features, which kept them left alone as separate single state groups.

Table 7. Geometric characteristics of the selected sites

Group	State	ID	AADT	Number of Lanes	City	HOV Type	Speed Limit (mph)
Group01	California	CA007	25,025	2	Palm Springs	NA	70
	Texas	TX010	26,590	2	El Paso	NA	75
Group02	California	CA008	34,625	2	Fresno	NA	70
	Texas	TX001	29,900	2	El Paso	NA	75
Group03	California	CA009	57,750	2	Salinas	NA	65
	Texas	TX007	50,200	2	San An	NA	70
Group04	California	CA011	73,000	3	Oakland	NA	65
	Texas	TX005	89,100	3	Waco	NA	70
Group05	California	CA014	161,500	4	Sacramento	NA	65
	Texas	TX008	130,570	4	Fort	NA	60
Group06	California	CA012	155,500	4	Vacaville	NA	65
	Texas	TX004	138,740	4	Houston	NA	60
Group07	California	CA013	233,000	4	Lynwood	Concurrent-Flow	65
	Texas	TX011	159,430	4	Carrollton	Concurrent-Flow	70
Group08	California	CA001	126,000	5	Sacramento	Concurrent-Flow	65
	Texas	TX006	130,570	5	Houston	NA	60
Group09	California	CA004	211,750	6	San Diego	Concurrent-Flow	55
	Texas	TX012	286,110	6	Houston	Pylon Concurrent-Flow	
Group10	California	CA010	146,000	5	Sacramento	Concurrent-Flow	65
	Texas	TX002	129,470	5	Dallas	Separated	60
Group11	California	CA002	36,250	2	San Jose	NA	70
Group12	California	CA003	35,500	2	Fresno	NA	70
Group13	California	CA005	250,000	5	Los Angeles	Concurrent-Flow	65
Group14	Texas	TX013	159,430	4	Dallas	Concurrent-Flow	65
Group15	Texas	TX003	122,990	4	Austin	NA	70
Group16	Texas	TX009	44,660	3	Dallas	NA	65

DATA ANALYSIS

Speed detector speed data was collected from both directions of the roadways (direction where crash happened, and opposite direction). To understand corridor-specific time series analysis of the speed data, the research team also collected speed data for all 26 sites one week before or after the crash incident as control sites. The corridors for similar spatio-temporal (location and crash time is fixed; only dates are one week before or after) control sites do not involve any crashes throughout the collection time of speed data.

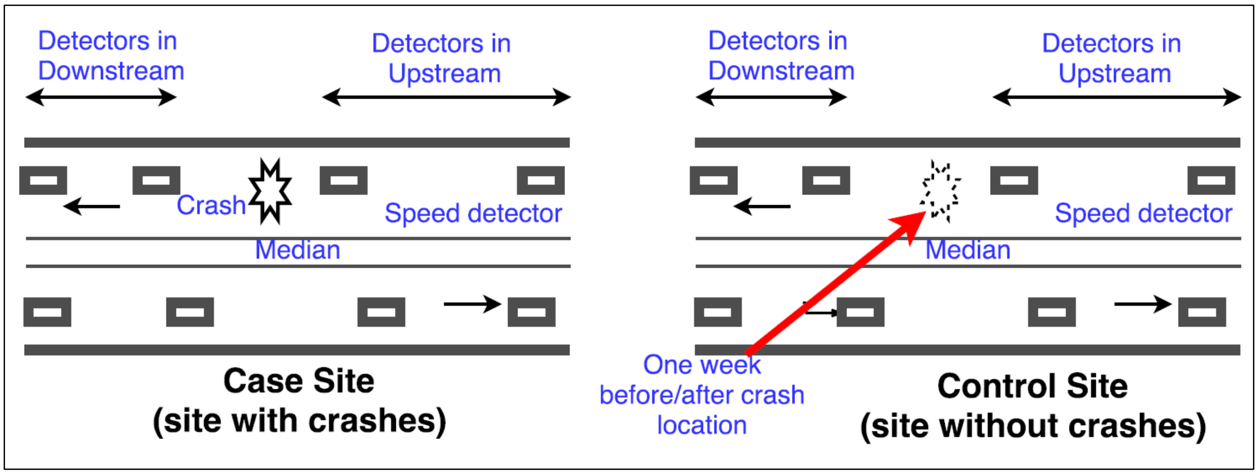

Figure 14. Schematic diagram of the case site (sites with crashes) and control sites (sites without crashes)

Two different diagrams (Figure 15 and Figure 16) are plotted to visualize the difference of the average speeds of the speed detectors. Figure 15 illustrates schematic diagram for a crash site showing average speeds of the speed detectors. The arrows in the figure show the locations of the speed detectors. This particular crash took place on the side of the Northbound of the highway. The downstream speed detectors showed lower average speeds than the upstream speed detectors on the Northbound direction. For southbound direction, significant difference in speeds were not visible.

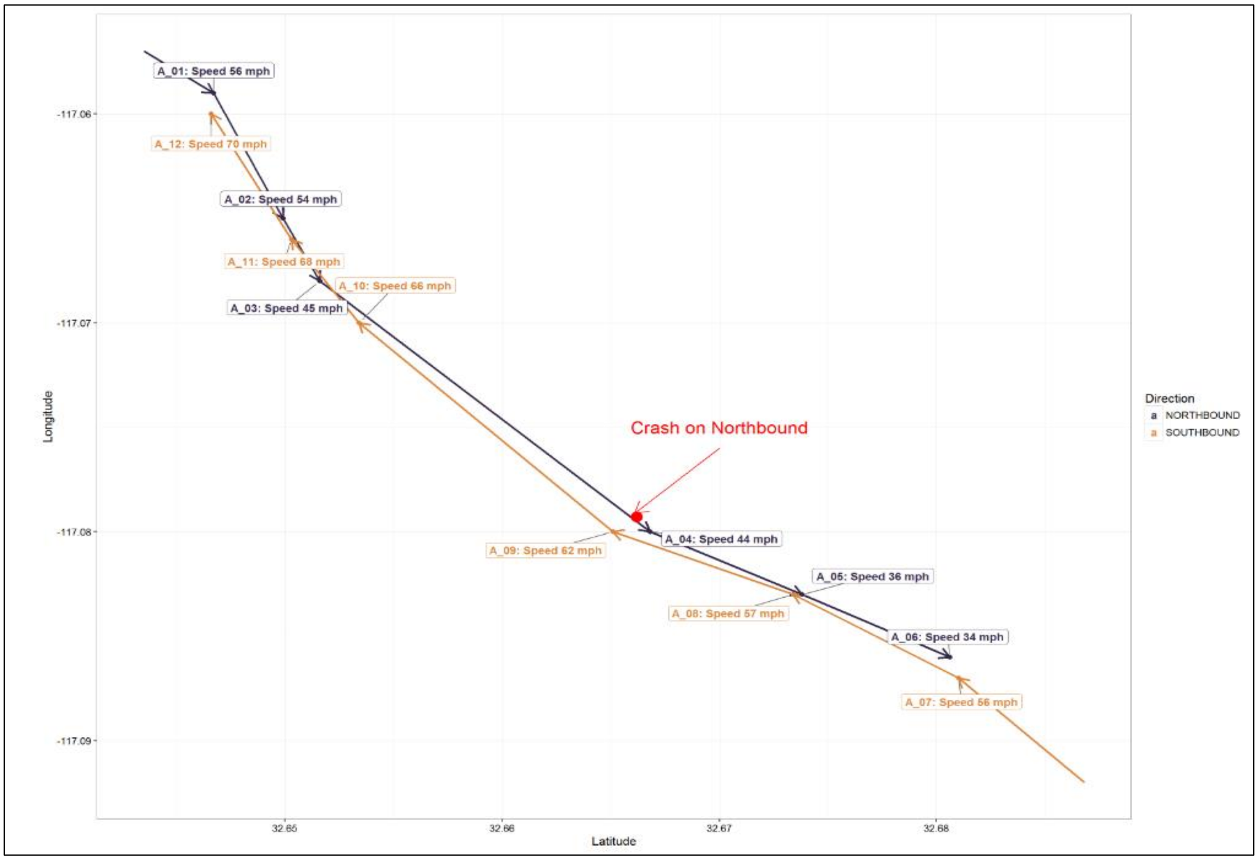

Figure 15. Schematic diagram for a crash site showing average speeds of the speed detectors

Figure 16 shows the schematic diagram for a crash site showing average speeds of the speed detectors in before and after crash condition. In Northbound direction, A_01, A_02, and A_03 are the upstream detectors and A_04, A_05, and A_06 were the downstream detectors. In the after crash periods, the average speeds of the detectors were lower in values. The opposite direction of the roadway did not involve any crash. The difference between before and after period speeds was not significant for the opposite direction.

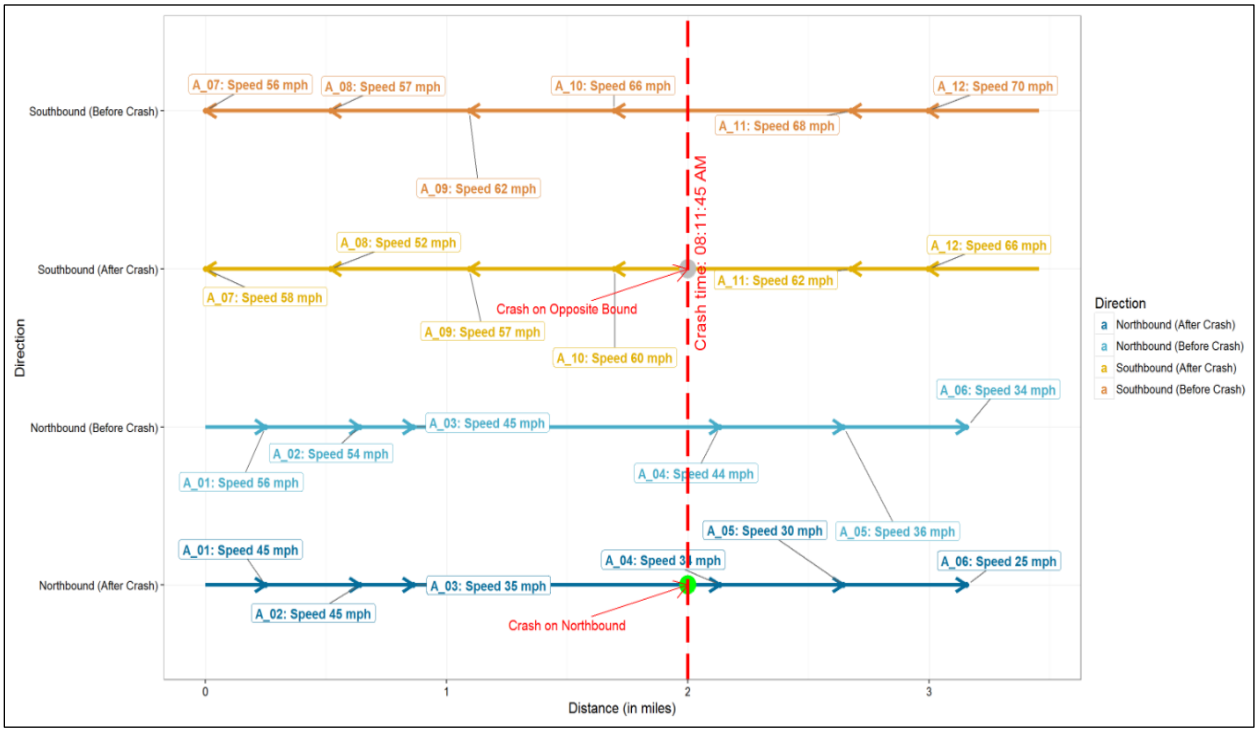

Figure 16. Schematic diagram for a crash site showing average speeds of the speed detectors in before and after crash condition

Another objective of this particular task is to observe the incident clearance time. Figure 17 and Figure 18 show the heat maps of the speed detectors for California and Texas sites in Group 3, respectively. In Figure 17, the light blue colors show the lower speeds and brown and dark brown depict the higher speeds. Crash location is the intersecting point of the dotted lines. For this particular case, crash happened on the northbound direction. After the crash occurrence, a significant drop of speeds was visible in both upstream and downstream detectors. It is interesting that similar situation was seen in the opposite bound where no crash happened. For this particular site, it can be said that this speed reduction was not for peak hour congestion, as the control sites did not not show any peak hour speed reduction. For Texas site of the same group, speed reduction was visible in the after crash periods. The heat maps of the two other groups (Group 4 and Group 5) are plotted in Figure 19 to Figure 22.

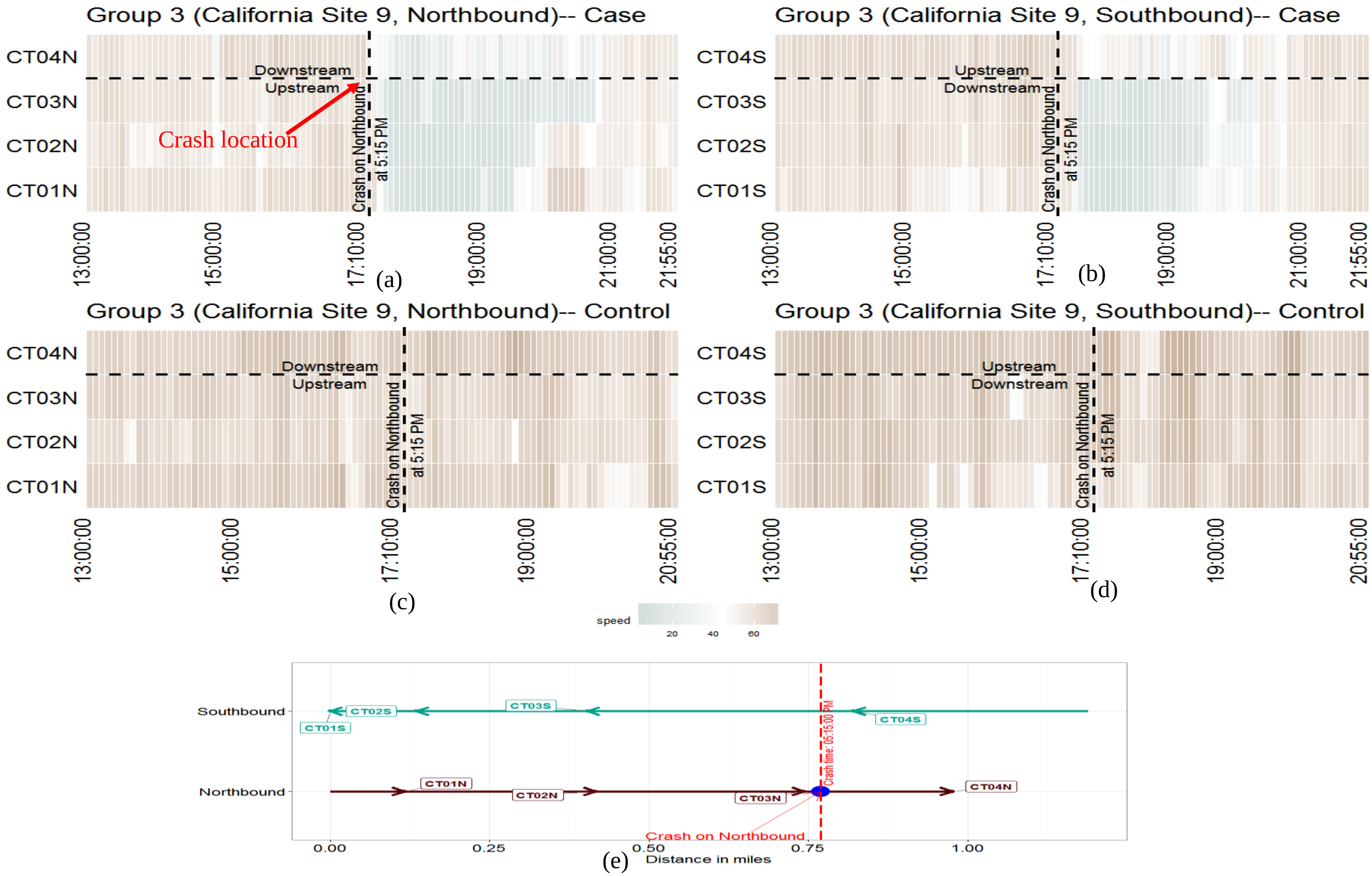

Figure 17. Heat map of the speed detectors for California site in Group 3

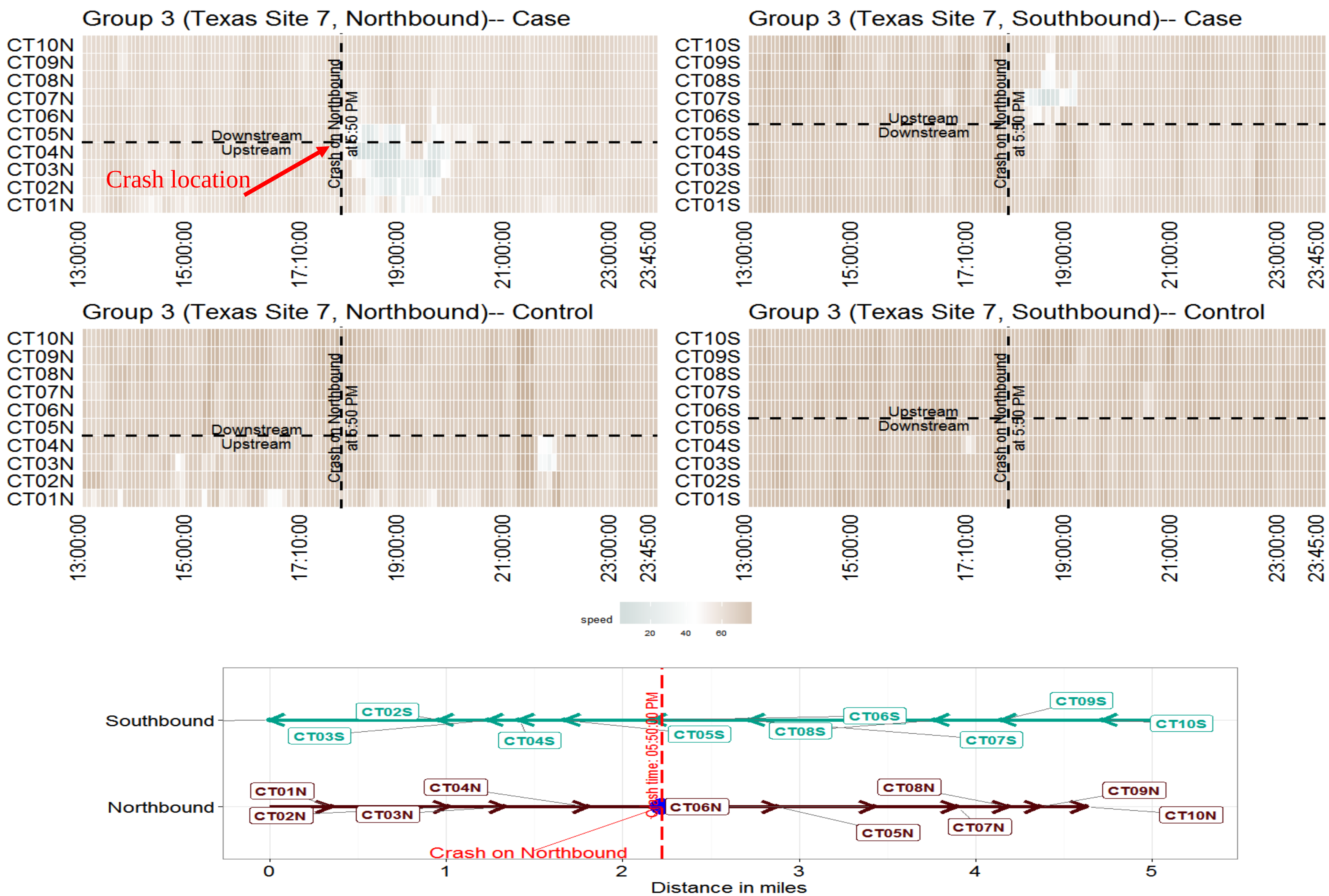

Figure 18. Heat map of the speed detectors for Texas site in Group 3

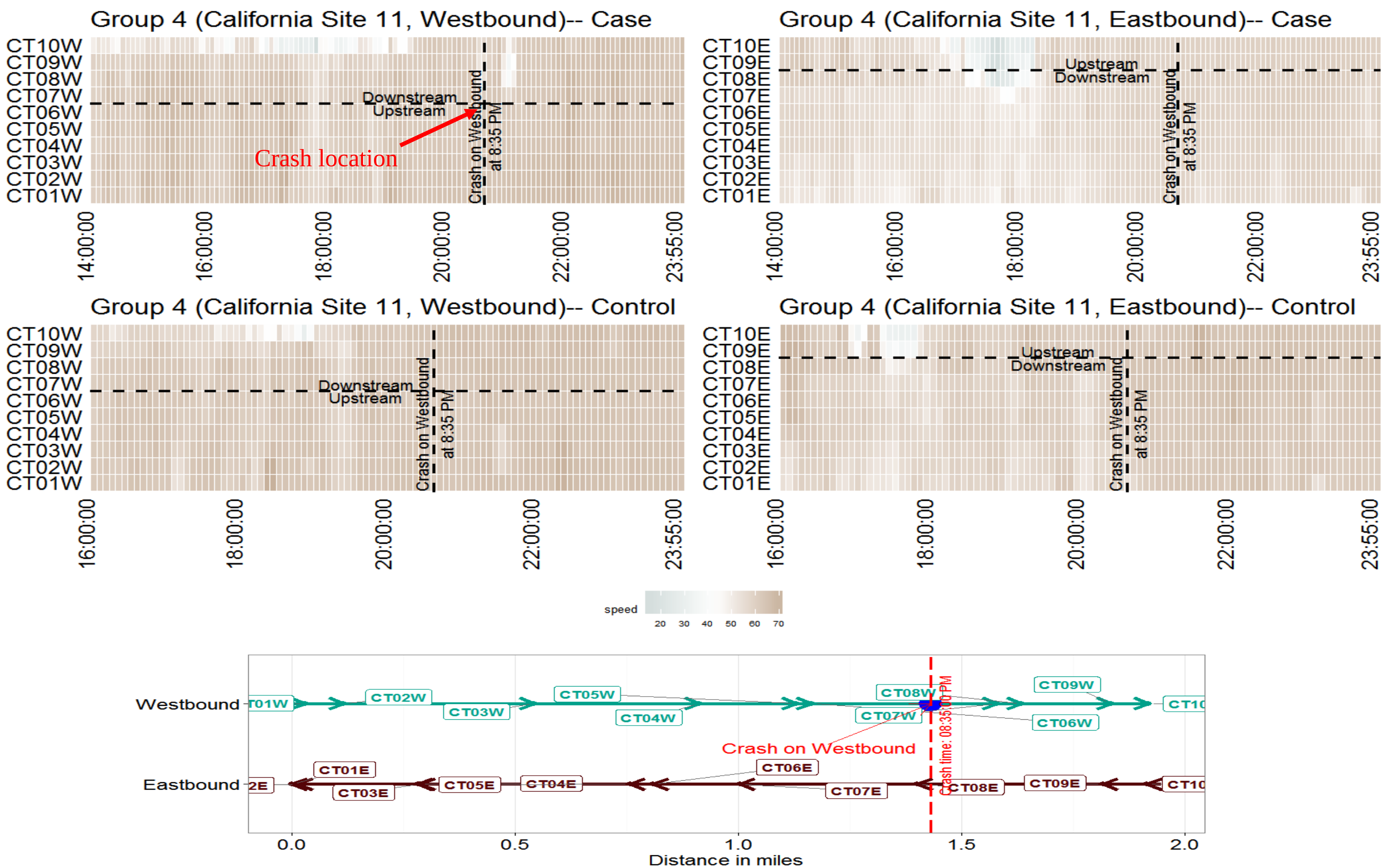

Figure 19. Heat map of the speed detectors for California site in Group 4

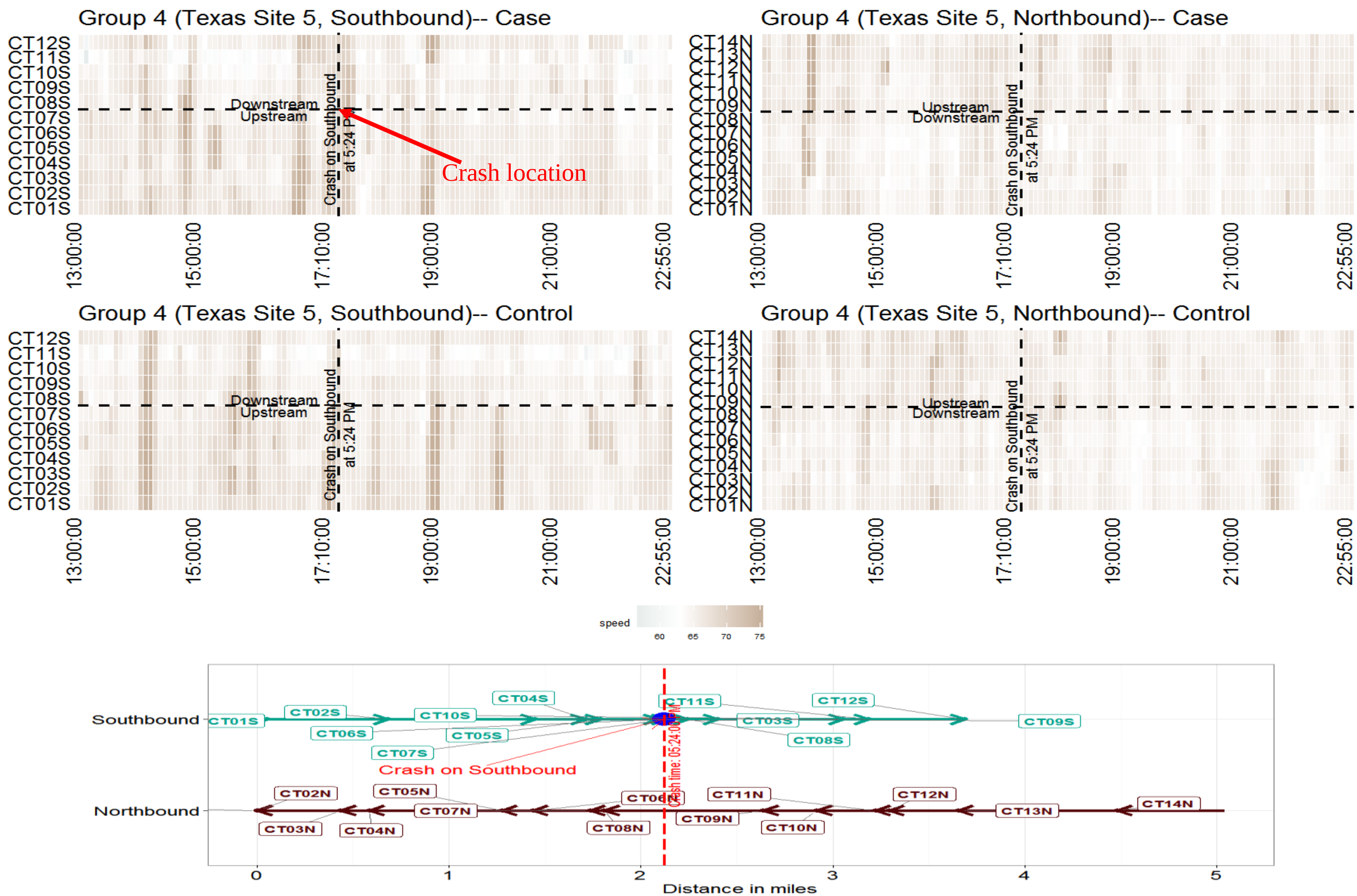

Figure 20. Heat map of the speed detectors for Texas site in Group 4

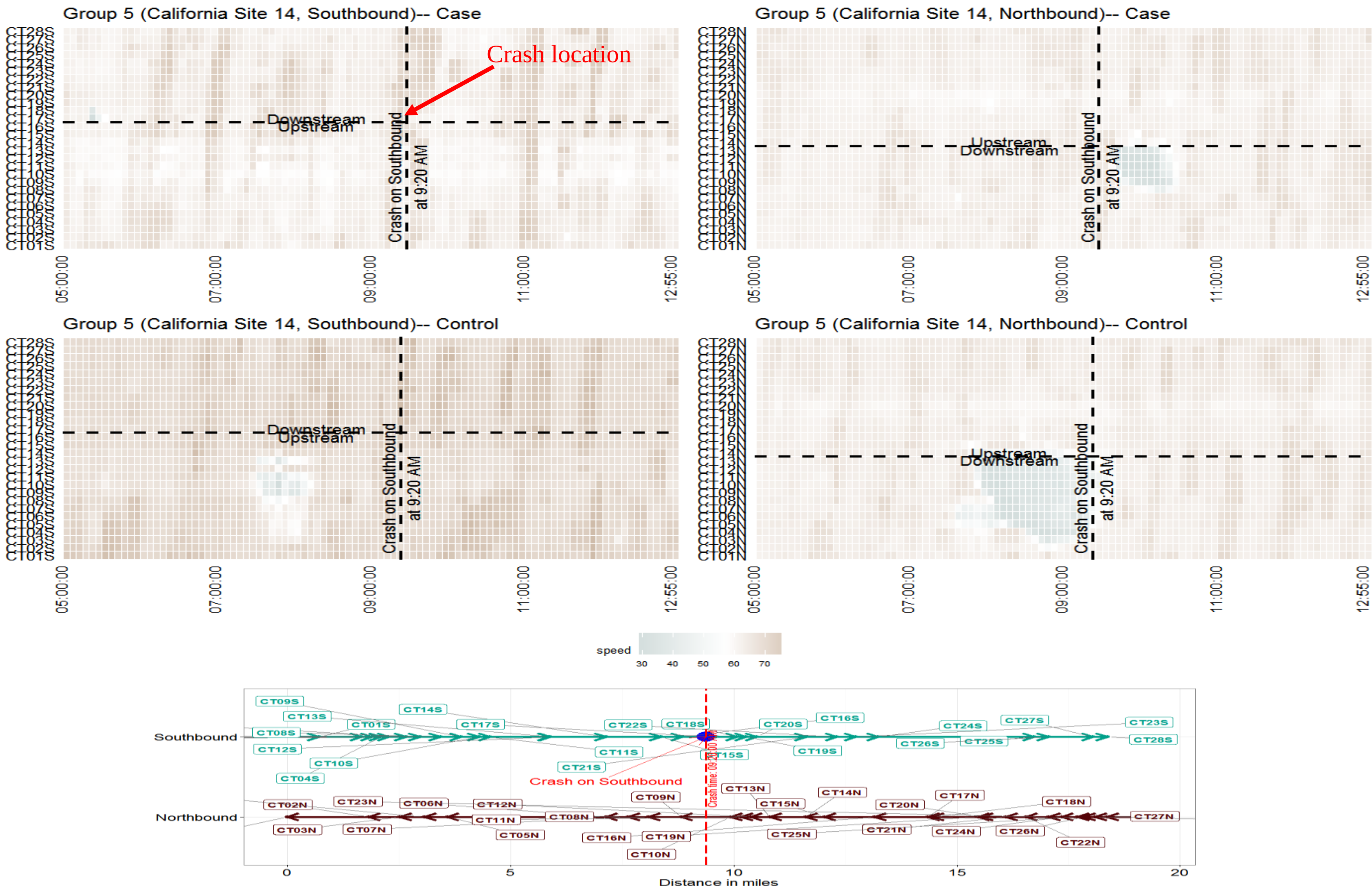

Figure 21. Heat map of the speed detectors for California site in Group 5

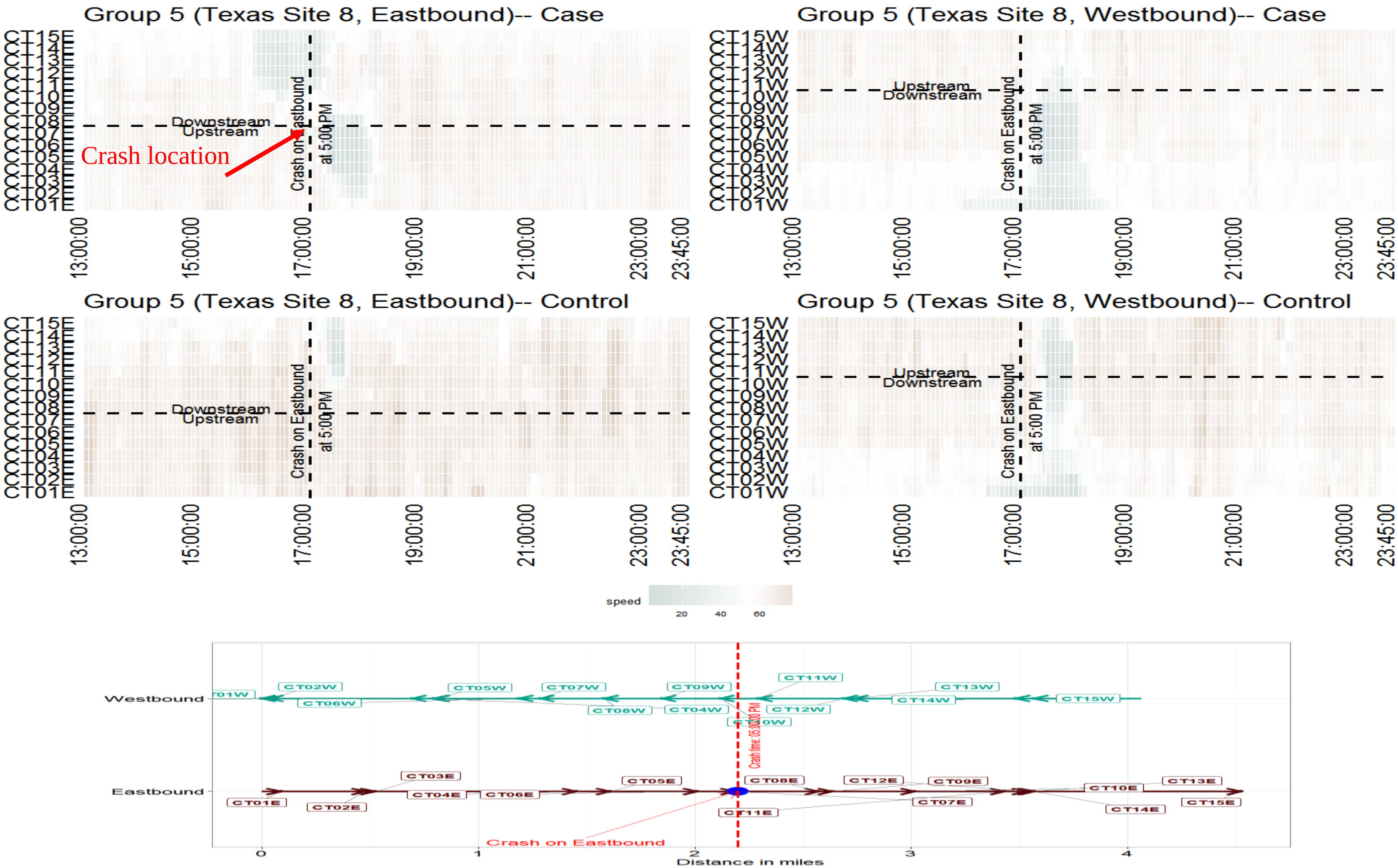

Figure 22. Heat map of the speed detectors for Texas site in Group 5

Fatal Crash Clearance Time

Incident management requires coordination and planning approach to restore freeway traffic to normal operation after a crash or incident by using available resources. To make incident management effective, the authority needs to develop a coordinated, cooperative, systematic approach to reduce time to detect and verify, and to act promptly. Effective incident management requires several items like the following:

- Traffic and incident management teams
- Freeway surveillance team
- Fast vehicle removal policies
- Motorist assistance patrols
- Tow truck contracts
- Road clearance equipment resources
- Alternative routing plan
- Ordinances allowing travel or shoulders around incidents

The Caltrans performance report documents that the average incident clearance time for a fatal crash is 3.91 hours *(39).* The report does not provide information for the clearance time for a fatal crash on urban freeways. TxDOT Project 7-4907 states that the average incident clearance time for a crash or incident is 36.3 minutes *(40).* The report lacks to provide information for the clearance time for a fatal crash on urban freeways.

Figure 23 shows the box and whisker plot for the clearance time of fatal crashes for the urban freeways of California and Texas. For the crash sites (case), the mean clearance times for downstream traffic are 100 minutes and 86 minutes for California and Texas, respectively. For upstream traffic, Texas takes slightly higher time (36 minutes) than California (32 minutes).

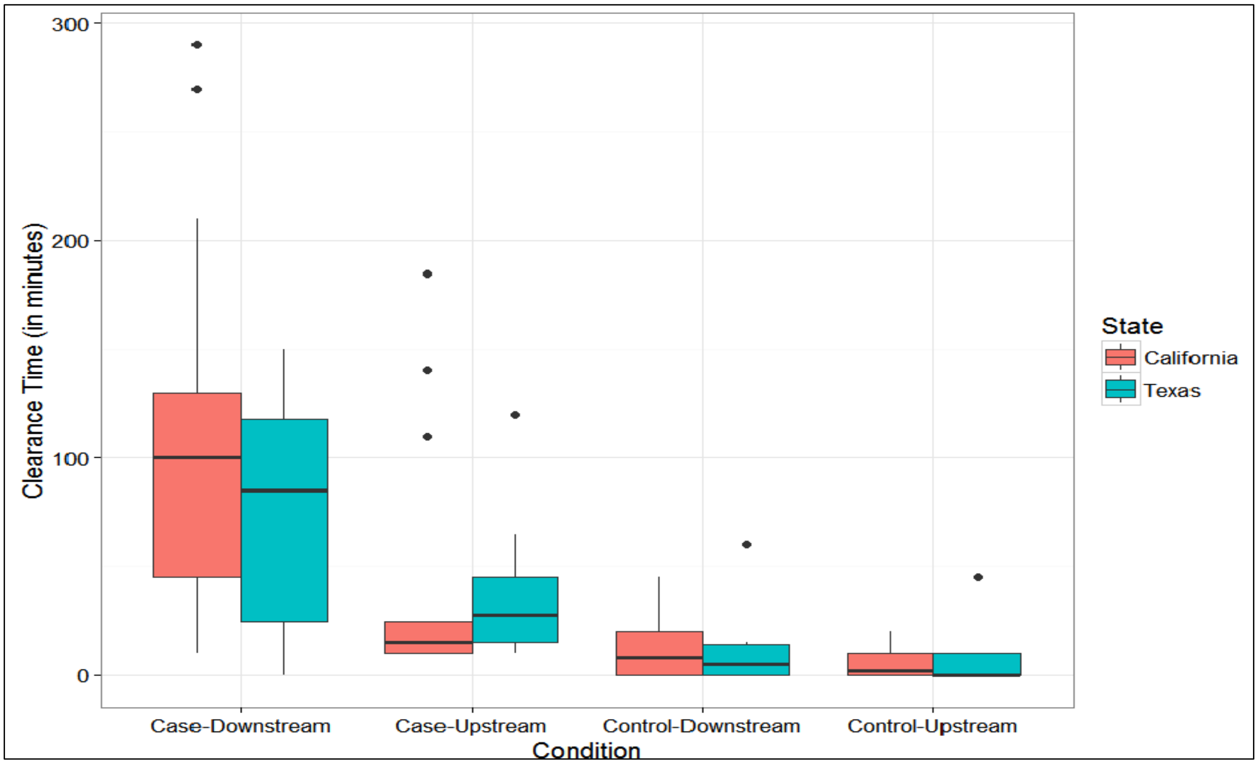

Figure 23. Box and whisker plot of the clearance time

REAL-TIME INCIDENT DETECTION

In this study, the research collected speed data at least 3 or 4 hours before and after of the crash occurrence. After performing iterations for multiple scenarios, the research team developed an overall heat map, shown in Table 8, based on a threshold of 20 minutes of before and after the crash occurrence. In most of the case sites, significant drops of average crashes were visible. It clearly shows that speed detector data can be used in determining real-time crash incident. The average reductions of speed (for 20 minutes threshold) in downstream are 3 percent and 1 percent for California and Texas, respectively. For upstream, the crash reduction on Texas urban roadways (10 percent) is higher than California (9 percent). Further exploration of the real-time crash detection from speed detector is beyond the scope of this study and would require a new project to quantify.

Table 8. Summary of the heat maps for all groups

Group	ID	Case Sites (sites with crashes)				Control Sites (sites with no crashes)			
		$Speed_1$*	$Speed_2$*	Speed3*	Speed4*	Speed1*	Speed2*	Speed3*	Speed4*
Group01	CA007	62.42	61.68	63.61	66.97	61.70	63.77	61.90	61.74
Group01	TX010	-	-	56.64	55.43	-	-	65.32	65.65
Group02	CA008	63.83	64.82	63.00	62.30	69.39	66.52	68.23	64.82
Group02	TX001	66.43	63.70	-	-	64.03	65.53	-	-
Group03	CA009	35.58	65.97	30.63	62.23	63.70	63.67	57.91	61.86
Group03	TX007	58.88	66.78	53.27	63.69	69.20	69.37	65.22	66.96
Group04	TX005	68.91	68.25	66.92	69.57	64.84	66.69	65.82	70.47
Group04	CA011	61.07	61.12	64.42	63.86	67.60	66.56	65.26	65.77
Group05	TX008	39.64	37.95	50.01	66.74	58.18	63.67	65.59	66.88
Group05	CA014	70.12	64.33	65.91	64.60	66.37	64.16	61.92	63.80
Group06	TX004	61.58	62.65	55.17	63.29	63.98	62.36	64.40	62.29
Group06	CA012	69.53	70.29	68.97	70.75	67.83	67.68	68.83	66.65
Group07	CA013	47.67	47.09	18.29	27.55	38.96	46.67	31.22	34.68
Group07	TX011	52.70	64.64	59.78	55.37	57.07	46.31	64.92	63.65
Group08	CA001	53.87	43.42	48.94	65.15	64.30	63.81	64.22	63.65
Group08	TX006	51.53	64.39	44.37	21.44	61.86	62.59	61.64	61.99
Group09	CA004	62.50	69.55	30.19	67.63	64.68	66.35	64.95	66.99
Group09	TX012	64.12	66.66	65.62	67.04	63.41	60.74	64.78	65.61
Group10	CA010	55.92	51.18	47.26	64.37	60.58	44.69	66.96	64.95
Group10	TX002	57.25	52.79	69.19	66.67	66.88	63.53	68.37	65.87
Group11	CA002	60.33	60.65	62.00	58.73	66.49	61.43	61.02	64.27
Group12	CA003	59.13	62.57	-	-	61.85	60.83	-	-
Group13	CA005	62.06	70.08	49.92	72.07	68.08	65.57	71.44	71.83
Group15	TX003	63.23	64.19	57.54	61.11	66.15	68.41	65.67	64.93
Group16	TX009	66.88	66.40	70.71	67.84	67.42	68.13	69.65	70.12
Group17	TX013	61.75	41.41	62.74	50.63	58.11	58.54	59.25	61.23
Note: (*)									
Speed1: Downstream Average Speed (mph) [After Crash (20 min)]									
Speed2: Downstream Average Speed (mph) [Before Crash (20 min)]									
Speed3: Upstream Average Speed (mph) [After Crash (20 min)]									
Speed4: Upstream Average Speed (mph) [Before Crash (20 min)]									
(-): All the speed detectors are either in upstream or downstream									

CHAPTER 4

Research Findings

This study was conducted to answer the research question: why are Texas urban freeways more prone to fatal crashes than California urban freeways? The current study contributes to the existing urban freeway fatality analysis literature by developing a generalized logistic regression to answer the research question. This study also quantifies incident clearance time for fatal crashes on urban freeways of California and Texas by using INRIX® speed detector data. The findings of this study are below:

- The distribution of geometric and traffic features (e.g. traffic flow condition, number of lanes, speed limit, and pavement types) of Texas and California urban freeways were not similar.
- Undivided urban two lane freeways were not safe for Texas.
- The fatal crashes on lower speed limit Texas urban roadways, especially with two lanes, three lanes, four lanes, and six or more lanes, were higher.
- The odds of fatal crash on Texas urban roadways were lower with unit increase of speed limit.
- The odds of fatal crash on Texas urban roadways with one or two lanes were higher. Addition of lanes had some safety benefits on the urban freeways of Texas.
- The odds of fatal crash on straightly aligned Texas urban roadways were higher.
- The odds of fatal crash on Texas urban roadways were higher in peak hours.
- For downstream, the fatal crash clearance time of Texas urban roadways was lower.
- For upstream, the reduction of average speeds for Texas urban roadways was higher.

REFERENCE

1. Ullman, G., and M. Ogden. Analysis of Major Freeway Incidents in Houston, Texas. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1554, 1996, pp. 221–227.
2. Clark, D. E., and B. M. Cushing. Rural and urban traffic fatalities, vehicle miles, and population density. *Accident Analysis & Prevention*, Vol. 36, No. 6, 2004, pp. 967–972.
3. Kweon, Y.-J. Prediction of Fatality Rates for Comparison between States. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2019, 2007, pp. 127–135.
4. Haselton, C., A. Gibby, and T. Ferrara. Methodologies Used to Analyze Collision Experience Associated with Speed Limit Changes on Selected California Highways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1784, 2002, pp. 65–72.
5. Bauer, K., D. Harwood, W. Hughes, and K. Richard. Safety Effects of Narrow Lanes and Shoulder-Use Lanes to Increase Capacity of Urban Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1897, 2004, pp. 71–80.
6. Kopelias, P., F. Papadimitriou, K. Papandreou, and P. Prevedouros. Urban Freeway Crash Analysis: Geometric, Operational, and Weather Effects on Crash Number and Severity. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2015, 2007, pp. 123–131.
7. Kononov, J., B. Bailey, and B. Allery. Relationships between Safety and Both Congestion and Number of Lanes on Urban Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2083, 2008, pp. 26–39.
8. Fitzpatrick, K., D. Lord, and B.-J. Park. Accident Modification Factors for Medians on Freeways and Multilane Rural Highways in Texas. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2083, 2008, pp. 62–71.
9. Park, B.-J., K. Fitzpatrick, and D. Lord. Evaluating the Effects of Freeway Design Elements on Safety. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2195, 2010, pp. 58–69.
10. Kononov, J., C. Durso, D. Reeves, and B. Allery. Relationship Between Traffic Density, Speed, and Safety and Its Implications for Setting Variable Speed Limits on Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2280, 2012, pp. 1–9.
11. Harwood, D., K. Bauer, and I. Potts. Development of Relationships Between Safety and Congestion for Urban Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2398, 2013, pp. 28–36.
12. Jackson, T. L., and H. O. Sharif. Rainfall impacts on traffic safety: rain-related fatal crashes in Texas. *Geomatics, Natural Hazards and Risk*, Vol. 7, No. 2, 2014, pp. 843–860.
13. Imprialou, M.-I. M., M. Quddus, D. E. Pitfield, and D. Lord. Re-visiting crash–speed relationships: A new perspective in crash modelling. *Accident Analysis & Prevention*, Vol. 86, 2016, pp. 173–185.

14. Mulokozi, E., and H. (H. Teng. Safety Analysis of Freeway Segments with Random Parameters. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2515, 2015, pp. 78–85.
15. Yu, L., Z. Li, A. R. Bill, and D. A. Noyce. Development of Freeway and Interchange Safety Performance Functions with Respect to Roadway Lighting. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2485, 2015, pp. 16–25.
16. Xu, C., A. P. Tarko, W. Wang, and P. Liu. Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis & Prevention*, Vol. 57, 2013, pp. 30–39.
17. Yu, R., X. Wang, K. Yang, and M. Abdel-Aty. Crash risk analysis for Shanghai urban expressways: A Bayesian semi-parametric modeling approach. *Accident Analysis & Prevention*, 2016.
18. Wang, J., W. Xie, B. Liu, S. E. Fang, and D. R. Ragland. Identification of freeway secondary accidents with traffic shock wave detected by loop detectors. *Safety Science*, Vol. 87, 2016, pp. 195–201.
19. Fang, S. E., W. Xie, J. Wang, and D. R. Ragland. Utilizing the eigenvectors of freeway loop data spatiotemporal schematic for real time crash prediction. *Accident Analysis & Prevention*, Vol. 94, 2016, pp. 59–64.
20. Kwak, H.-C., and S. Kho. Predicting crash risk and identifying crash precursors on Korean expressways using loop detector data. *Accident Analysis & Prevention*, Vol. 88, 2016, pp. 9–19.
21. Levine, N. Houston, Texas, Metropolitan Traffic Safety Planning Program. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1969, 2006, pp. 92–100.
22. Kononov, J., B. Allery, and Z. Znamenacek. Safety Planning Study of Urban Freeways: Proposed Methodology and Review of Case History. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2019, 2007, pp. 146–155.
23. Voigt, A., C. Stevens, and D. Borchardt. Dual-Advisory Speed Signing on Freeway-to-Freeway Connectors in Texas. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2056, 2008, pp. 87–94.
24. Geedipally, S., D. Lord, and G. Reddy. Evaluation of Texas Department of Transportation Safety Improvement Index as a Prioritization Tool. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2241, 2011, pp. 39–50.
25. Park, B.-J., K. Fitzpatrick, and M. Brewer. Safety Effectiveness of Super 2 Highways in Texas. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2280, 2012, pp. 38–50.
26. Viallon, V., and B. Laumon. Fractions of fatal crashes attributable to speeding: Evolution for the period 2001–2010 in France. *Accident Analysis & Prevention*, Vol. 52, 2013, pp. 250–256.

27. Burger, N. E., D. T. Kaffine, and B. Yu. Did California's hand-held cell phone ban reduce accidents? *Transportation Research Part A: Policy and Practice*, Vol. 66, 2014, pp. 162–172.
28. Fatality Analysis Reporting Systems (FARS). http://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed on: August, 2016.
29. INRIX. http://inrix.com/ Accessed on: August, 2016.
30. Hauer, E. On Exposure and Accident Rate. Traffic Engineering and Control, Vol. 36, No. 3, 1995, pp. 134–138.
31. Highway Statistics on Licensed Drivers. https://www.fhwa.dot.gov/policyinformation/statistics/ Accessed on: August, 2016.
32. Highway Statistics on VMT. https://www.fhwa.dot.gov/policyinformation/statistics/ Accessed on: August, 2016.
33. Breiman, L. Random forests. Machine learning, Vol. 45(1), 2001, pp. 5-32.
34. Ho, T. K. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 8, 1998, pp. 832–844.
35. McCullagh, P., and J. Nelder. Generalized Linear Models, 2nd Edition. CRC Press, Washington D.C., 1999.
36. Gerken, J. Generalized Logit Model. Transportation Research Part B: Methodological Volume 25, Issues 2–3, 1991, pp. 75–88.
37. Statewide Integrated Traffic Records System (SWITRS). http://iswitrs.chp.ca.gov/Reports/jsp/. Accessed on: August, 2016.
38. Crash Record Information System (CRIS). https://cris.dot.state.tx.us Accessed on: August, 2016.
39. The Mile Marker: A Caltrans Performance Report. 2015 Third Quarter Issue, 2015.
40. Balke, K., S. Cooner, B. Durkop, D. Fenno, and R. Arredondo. Functional Process Specifications for Improving Incident Response. TxDOT Project 7-4907, 2001.

APPENDIX

AADT	Annual Average Daily Traffic
CRIS	Crash Record Information System
EB	Empirical Bayes
FARS	Fatality Analysis Reporting System
mph	miles per hour
OOB	Out-of-bag
ROR	Run-off-road
SWITRS	Statewide Integrated Traffic Records System
VMT	Vehicle Miles Traveled
vpd	vehicles per day