THE UNIVERSITY OF TEXAS AT EL PASO

Comparisons of Discretionary Lane Changing Behavior

Matthew Vechione[1] E.I.T., Ruey Cheu, Ph.D., P.E., M.I.T.E.

[1]Department of Civil Engineering, The University of Texas at El Paso,
500 W. University Ave., El Paso, TX 79968, U.S.A.
Tel: (915) 443-3333
Email: mmvechione@miners.utep.edu

March 31, 2017

TABLE OF CONTENTS

1 ABSTRACT

This research studies the statistical properties of four parameters that affect driver's lane changing decision, using data from the Next Generation SIMulation database. The results show that (i) there is statistical evidence to indicate that the population averages for each parameter differ based on time-of-day; and (ii) the gap parameters are best described by the log-normal distribution. This implies that autonomous vehicles should be programmed to behave differently at different times of the day.

2 KEYWORDS

Lane Change, Gap, Autonomous Vehicles, Safety, Driving Behavior

3 INTRODUCTION

A lane changing event involves up to five vehicles, the subject vehicle (*S*), the preceding vehicle before the lane change (*PB*), the following vehicle before the lane change (*FB*), the preceding vehicle after the lane change (*PA*), and the following vehicle after the lane change (*FA*). These vehicles may be seen in Figure 1 below. In coding the vehicles, *P* represents a preceding vehicle; *F* represents a following vehicle; *B* is before the lane change occurs; and *A* is after the lane change occurs. The longitudinal position of each vehicle may be seen in Figure 1 as the *Y* value, using the front center of each vehicle as the reference point.

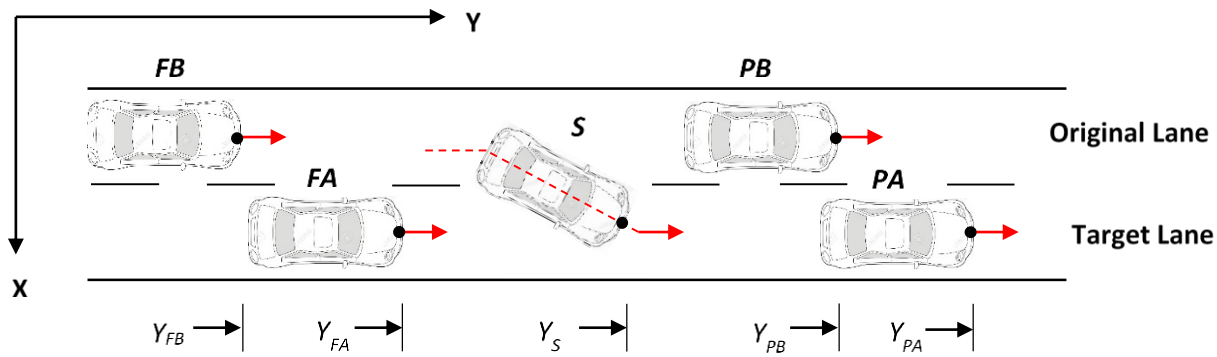


Figure 1. Vehicles and Their Positions During a Typical Lane Change

There are two types of lane changes: mandatory and discretionary. A mandatory lane change occurs when a driver must change lanes in order to exit a freeway or make a turn at a downstream intersection. A discretionary lane change occurs at the driver's own discretion and speed is typically a factor (i.e. the preceding vehicle is travelling too slowly or the following vehicle is travelling too fast). A driver is expected to have different decision rules and/or risk taking behavior for these two types of lane changes, on different highway facilities (i.e. freeways vs. arterial streets), and under different traffic congestion levels, which typically correlate with different times during a day.

A lane change may be modeled as a four-step process[1]:

(1) Motivation;
(2) Selection of target lane;
(3) Checking for the opportunity to move; and
(4) The actual move.

This research focuses on step (3), checking for the opportunity to move, which a wrong decision could lead to a crash in step (4).

The objectives of this research are to (i) use statistical tests to determine if drivers use the same thresholds when making discretionary lane change decisions under different driving conditions (i.e. time of day; and subsequently, congestion level), and (ii) determine the distribution types for discretionary lane change parameters for arterial streets. The thresholds, once determined, may be programmed into the lane changing decision rule in autonomous vehicles, while the distribution of thresholds may be used to represent the aggressiveness (i.e., risk taking behavior) of different drivers (including autonomous vehicles) in traffic simulation models.

4 LITERATURE REVIEW

This research made use of the Next Generation Simulation (NGSIM) data to analyze discretionary lane changing behavior on arterial streets. The Federal Highway Administration (FHWA) funded the NGSIM Project to collect vehicle trajectory data via eight cameras mounted on top of tall buildings for arterial streets and freeways[2,3].

Published articles that use NGSIM data were reviewed to determine if other research has been conducted regarding lane changing behavior at similar highway facilities. The NGSIM data have been used

to analyze the probability distribution of discretional lane change decision parameters in freeway driving[4]; traffic relaxation, anticipation, and hysteresis[5]; and the calibration of nonlinear car-following laws for traffic oscillation prediction[6]. The NGSIM databases have also been used to validate models such as the kinematic equation-based vehicle queue location estimation method for signalized intersections using mobile sensor data[7]; and for artificial intelligence car-following model (i.e. connected vehicles)[8]. None of the above studies have considered the potential difference in gap acceptance behavior with regards to highway facilities and traffic congestion level.

A survey from 443 drivers in El Paso, Texas has found that the top four input parameters used in making lane change decisions by at least 81% of the respondents are gaps, primarily G_{PB}, G_{PA}, G_{FA}, and D [4]. These gaps will be described in next section.

5 METHODOLOGY

The NGSIM used was collected at Peachtree Street in Atlanta, Georgia. The data collected between 4:00 and 4:15 p.m. (with traffic exposure[a] of 1,725) is denoted as dataset P1. The data collected between 12:45 and 1:00 p.m. (with traffic exposure of 1,458) is named dataset P2.

Each dataset was filtered to only include vehicles travelling along Peachtree St. without making a turn, as to eliminate a possible mandatory lane change maneuver. Each dataset was then filtered based on vehicle class, as to only include passenger vehicles; and then filtered based on lane identification number, as to eliminate vehicles in turn-bays. There were 134 and 135 subject vehicles after data filtration for datasets P1 and P2, respectively

The remaining vehicles were then analyzed to determine which vehicles made exactly one lane change and which vehicles made more than one lane change. Any lane change that occurred in an intersection was omitted. This was because the NGSIM database tracks the lane identification for each vehicle at each time interval (i.e. every tenth of a second); however, when a vehicle is in an intersection, the NGSIM database shows "0" as the lane identification, and therefore no parameters could be derived. After this round of data filtration, there were 47 lane change occurrences in dataset P1, and 51 lane change occurrences in dataset P2.

The formula used for calculating G_{PB} and G_{PA} was the longitudinal distance between the rear bumper of the preceding vehicle (before and after the lane change, respectively) and front bumper of the subject vehicle. Likewise, the formula used for calculating G_{FA} was the longitudinal distance between the rear bumper of the subject vehicle and the front bumper of the following vehicle after the lane change. The formula used for calculating D was the summation of G_{PA} and G_{FA}, which does not include the length of the subject vehicle.

The NGSIM data consisted of vehicle positions at 0.1 second intervals. For each lane change, the instant t, when the subject vehicle (the front center of a vehicle, as seen in Figure 1) crossed the lane marker was identified. The four lane change parameters were calculated at t-0.4, t-0.3, t-0.2, t-0.1, and t seconds and the average values from t-0.4 to t seconds were used at the representative values of the "accepted" parameters. This method of averaging data to 0.5 second resolution was done for three reasons (i) to reduce error caused by instantaneous values in NGSIM data; (ii) to be more consistent with human perception

[a] The traffic exposure was calculated by the summation of all vehicles entering each intersection for all five intersections within the study area.

time; and (iii) to be consistent with other researches that used NGSIM data. Instances where there was any missing surrounding vehicle (i.e. when G_{PB}, G_{PA}, G_{FA}, or D equals to ∞), the lane change occurrence was omitted for correlation analysis. The final dataset contained 29 lane change events for dataset P1, and 32 lane change events for dataset P2.

6 RESULTS AND DISCUSSIONS

Table 1 lists the descriptive statistics of the four parameters analyzed for both the P1 and P2 datasets. Each gap and distance parameter is processed to 0.01 feet (0.003 meter) precision.

Table 1. Descriptive Statistics of Lane Changing Parameters for Both Datasets

Dataset	P1			
Parameters	G_{PB}	G_{PA}	G_{FA}	D
Unit	ft. (m)	ft. (m)	ft. (m)	ft. (m)
Sample Size	29	29	29	29
Min Value	14.34 (4.37)	26.15 (7.97)	29.79 (9.08)	78.51 (23.93)
Max Value	1,593.11 (485.58)	1,685.50 (513.74)	2,894.13 (882.13)	3,832.38 (1,168.11)
Mean	251.71 (76.72)	351.38 (107.10)	676.87 (206.31)	1,028.25 (313.41)
Std. Deviation	332.25 (101.27)	449.02 (136.86)	762.57 (232.43)	898.33 (273.81)
Skewness	2.93	2.03	1.62	1.49
Dataset	P2			
Parameters	G_{PB}	G_{PA}	G_{FA}	D
Unit	ft. (m)	ft. (m)	ft. (m)	ft. (m)
Sample Size	32	32	32	32
Min Value	38.32 (11.68)	54.79 (16.70)	90.29 (27.52)	270.24 (82.37)
Max Value	2,241.37 (683.17)	2,153.12 (656.27)	2,782.48 (848.10)	4,586.19 (1,397.87)
Mean	687.30 (209.49)	804.63 (245.25)	1,127.89 (343.78)	1,932.51 (589.03)
Std. Deviation	670.37 (204.33)	636.39 (193.97)	753.54 (229.68)	1,096.36 (334.17)
Skewness	1.05	0.72	0.39	0.41

6.1 Hypothesis Testing for Datasets P1 and P2

Hypothesis tests were conducted to determine if the average value for each parameter was statistically similar between the two datasets (i.e. late afternoon vs. mid-day). For each hypothesis test, the null hypothesis was that the population averages (μ_1 and μ_2) for each parameter were the same; and the alternate hypothesis was that the population averages for each parameter differ. The test statistic followed

the *t*-distribution with small sample sizes; and it was assumed that the variances were unknown and not equal[9]. Each hypothesis test was two-sided with an alpha value of 0.05 (i.e. $\frac{\alpha}{2} = 0.025$). Table 2 presents the results of hypothesis tests.

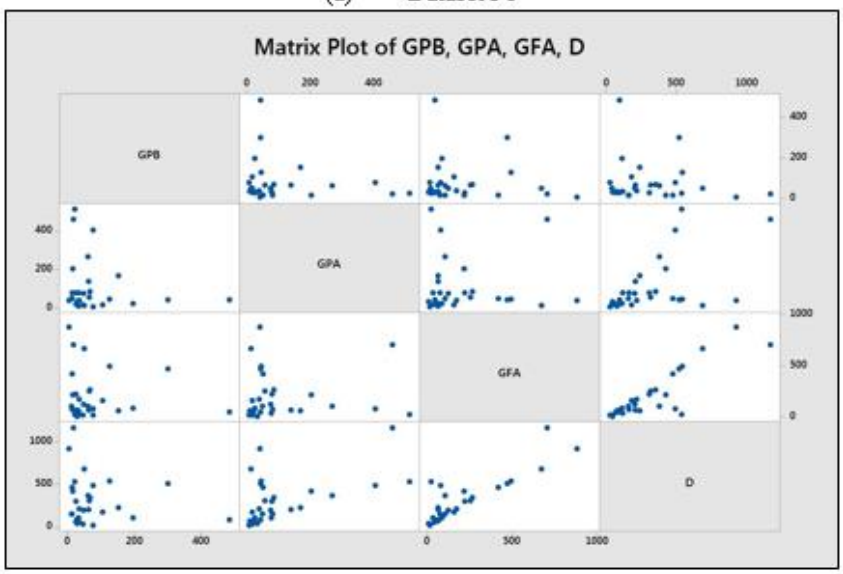
Table 2. Hypothesis Test Results

Parameters	P1	P2
G_{PB}	$H_0: \mu_1 = \mu_2$	
Sample Mean, $\bar{X}$ (ft., m)	251.71 (76.72)	687.30 (209.49
Sample Std. Dev., *s* (ft., m)	332.25 (101.27)	670.37 (204.33)
No. of Observations, *n*	29	32
t-Value	-3.26	
Conclusion	**Reject H_0**	
G_{PA}	$H_0: \mu_1 = \mu_2$	
Sample Mean, $\bar{X}$ (ft., m)	351.38 (107.10	804.63 (245.25)
Sample Std. Dev., *s* (ft., m)	449.02 (136.86)	636.39 (193.97)
No. of Observations, *n*	29	32
t-Value	-3.24	
Conclusion	**Reject H_0**	
G_{FA}	$H_0: \mu_1 = \mu_2$	
Sample Mean, $\bar{X}$ (ft., m)	676.87 (206.31)	1,127.89 (343.78)
Sample Std. Dev., *s* (ft., m)	762.57 (232.43)	753.54 (229.68)
No. of Observations, *n*	29	32
t-Value	-2.30	
Conclusion	**Reject H_0**	
D	$H_0: \mu_1 = \mu_2$	
Sample Mean, $\bar{X}$ (ft., m)	1,028.25 (313.41)	1,932.51 (589.03)
Sample Std. Dev., *s* (ft., m)	898.33 (273.81)	1,096.36 (334.17)
No. of Observations, *n*	29	32
t-Value	-3.54	
Conclusion	**Reject H_0**	

Based on the results presented in Table 2, there is statistical evidence to suggest that the population averages for each parameter at different times of the day differ at 95% confidence. This implies that drivers have different rules and/or risk taking behaviors based on time-of-day (which have different traffic congestion levels).

Correlation analysis was next performed for all the parameters in each dataset using MINITAB[10]. The purpose of correlation analysis is to determine if there is any strong relationship between any two parameters. If any strong correlation exists, one may be able to design lane changing decision model based on fewer parameters, which translates to fewer sensors and cost saving in the vehicle control system. The correlations between any two parameters are visualized in the scatter plots in Figure 2.

(a) Dataset P1



(b) Dataset P1

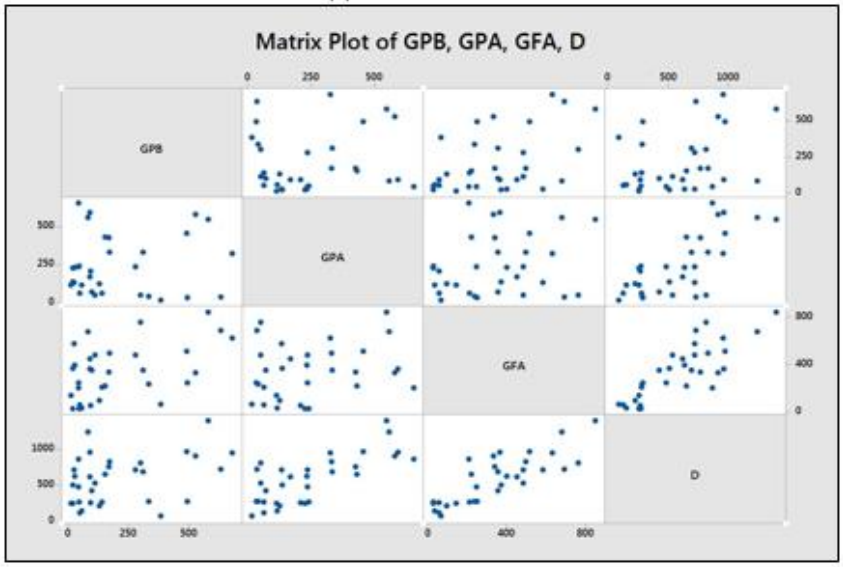


Figure 2. Correlation Analysis of Parameters

The results from Figure 2 illustrate that there is no statistical correlation between any two parameters in either data set other than G_{FA} and D. This is likely because D is the summation of G_{PA} and G_{FA}.

6.2 Distribution Types

The probability distribution for each parameter was analyzed using @RISK[11]. In Table 3, the top three fitted distributions for each parameter, among the 26 distributions tested, were chosen based on the Akaike Information Criterion (AIC) for goodness of fit. It is preferred to have one probability distribution to fit gaps and distance[4]. A numeric scoring system was used to select one probability distribution. Distributions that provide the best, second-best, and third-best fits were assigned scores of 3, 2, and 1, respectively. The distribution with the highest total score was recommended.

Table 3. Fitted Probability Distributions of Lane Changing Parameters

Parameters	G_{PB}	G_{PA}	G_{FA}	D
Unit	ft.	ft.	ft.	ft.
P1				
Best Fit	Log-logistic	Pareto 2	Pareto 2	Exponential
2nd Best Fit	Pearson	Log-normal	Exponential	Pareto 2
3rd Best Fit	Log-normal	Log-normal 2	Erlang	Pert
Recommended	**Log-normal**			
P2				
Best Fit	Exponential	Triangle	Triangle	Triangle
2nd Best Fit	Erlang	Exponential	Pert	Rayleigh
3rd Best Fit	Pareto 2	Pareto 2	Uniform	Pert
Recommended	**Log-normal**			

The log-normal distribution was recommended for both dataset P1 and P2, as it is a more commonly known distribution. An example of the @RISK log-normal distribution fitting of G_{PA} for dataset P1 is illustrated in Figure 3.

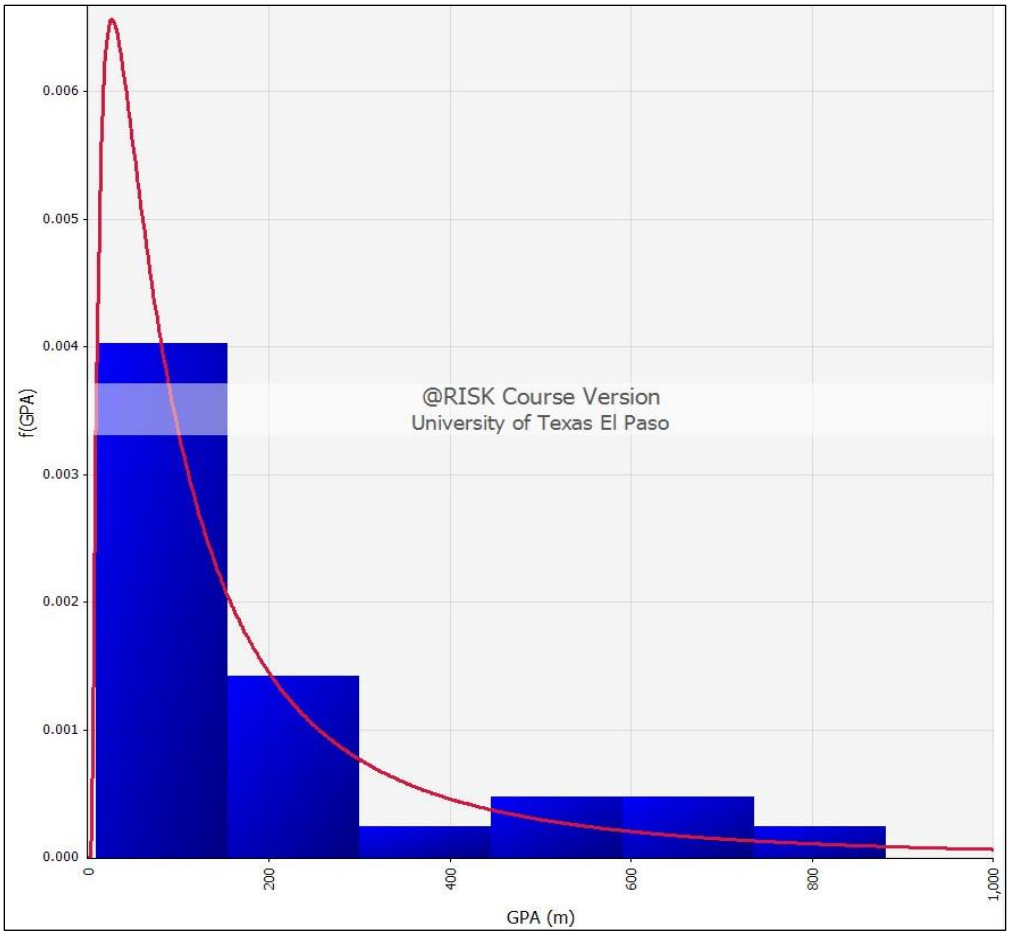


Figure 3. Log-normal Distribution Fitting of G_{PA} for Dataset P1

7 APPLICATIONS

The findings of this research have two important applications in the lane changing decision model. In step (3) of the four-step process, the driver or sensors of the autonomous vehicle estimate G_{PB}, G_{PA}, G_{FA}, or D and compare each parameter with their respective thresholds.

Microscopic traffic simulation is a commonly used approach used by transportation engineers to perform traffic impact studies. Simulation approach has also been used by researchers to investigate the impacts of gradual introduction of autonomous vehicles or connected vehicles in mixed traffic stream[12,13,14]. In traffic simulation models, the stochastic behavior of drivers of conventional vehicles are represented by using probability distributions to generate the parameter values. The recommended log-normal distributions may be used in the simulation software to represent the varied thresholds used by drivers with varying degree of aggressiveness.

The authors expect that, in autonomous vehicles, such lane changing thresholds be fixed at more conservative values to ensure safety in lane change maneuvers. However, using a more conservative threshold (e.g., large gap), although improve safety, may compromise capacity, the two benefit always promoted by proponents of autonomous vehicles. The results of hypothesis tests have suggested that, the thresholds of an autonomous vehicle may be adjusted automatically depending on the time-of-day (by the clock time), or based on the traffic density (by receiving data from surrounding connected vehicles).

8 SUMMARY, LIMITATIONS, AND FUTURE RESEARCH

The NGSIM data on Peachtree St. was processed to analyze the four gap and distance parameters. Hypothesis test results indicate that there is statistical evidence to suggest that population averages for each parameter differ based on driving conditions (i.e. time-of-day and traffic congestion levels). Log-normal distribution is recommended for all four parameters for both data sets, as it is more commonly known.

This research has performed the above hypothesis tests, using real data. However, as in all research, there are limitations. Similar analysis should be conducted for NGSIM data on arterial streets in other regions of the United States, to determine if there is any local effect on the parameter values. Similar analysis should be conducted for NGSIM data on trucks (i.e. non-passenger vehicles).

9 ACKNOWLEDGEMENTS

I would like to acknowledge Mariana Mercado for all of her help to process the data. I would also like to thank Dr. Esmaeil Balal for advising me during the preliminary stage of data processing. I would also like to include my advisor, Dr. Ruey (Kelvin) Cheu, who took time to help prepare my paper to read professionally.

10 REFERENCES

1 S. Moridpour, M. Sarvi and G. Rose, "Lane changing models: a critical review," *Transportation Letters:* , vol. 2, no. 3, pp. 157-173, 2010.

2 Cambridge Systematics, Inc., "NGSIM Peachtree Street (Atlanta) Data Analysis (4:00 p.m. to 4:15 p.m.) Summary Report," 2007.

3 Cambridge Systematics, Inc., "NGSIM Peachtree Street (Atlanta) Data Analysis (12:45 p.m. to 1:00 p.m.) Summary Report," 2007.

4 E. Balal, R. L. Cheu, T. Gyan-Sarkodie and J. Miramontes, "Analysis of Discretionary Lane Changing Parameters on Freeways," *International Journal of Transportation Science and Technology,* pp. 277-296, 2014.

5 H. Deng and H. M. Zhang, "On Traffic Relaxation, Anticipation, and Hysteresis," *Traffic Flow Theory and Characteristics, Volume 2,* pp. 90-97, 2015.

6 C. Rhoades, X. Wang and Y. Ouyang, "Calibration of Nonlinear Car-Following Laws for Traffic Oscillation Prediction," *Transportation Research Part C: Emerging Technologies, Volume 69,* pp. 328-342, 2016.

7 P. Hao, X. (. Ban and J. Whon Yu, "Kinematic Equation-Based Vehicle Queue Location Estimation Method for Signalized Intersections Using Mobile Sensor Data," *Journal of Intelligent Transportation Systems, Volume 19, Issue 3,* pp. 256-272, 2015.

8 H. Hao, W. Ma and H. Xu, "A Fuzzy Logic-Based Multi-Agent Car-Following Model," *Transportation Research Part C: Emerging Technologies, Volume 69,* pp. 477-496, 2016.

9 D. C. Montgomery and G. C. Runger, "Inference on the Difference in Means of Two Normal Distributions, Variances Unknown," in *Applied Statistics and Probability for Engineers. 5th Edition*, Hoboken, NJ, John Wiley & Sons, Inc., 2011, pp. 361-372.

10 Minitab, *Minitab Statistical Software, Version 16.1.0,* State College, PA: Minitab Inc., 2010.

11 Palisade, *@Risk, Version 6.1.1,* Ithaca, NY: Palisade Corporation, 2013.

12 A. Talebpour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," *Transportation Research Part C: Emerging Technologies,* vol. 71, pp. 143-163, 2016.

13 R. Krueger, T. H. Rashidi and J. M. Rose, "Preferences for shared autonomous vehicles," *Transportatio Research Part C: Emerging Technologies,* vol. 69, pp. 343-355, 2016.

14 H. Park and B. L. Smith, "Investigating Benefits of IntelliDrive in Freeway Operations: Lane Changing Advisory Case Study," *Journal of Transportation Engineering,* vol. 138, no. 9, pp. 1113-1122, 2012.