



World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016

# A methodology for forecasting freeway travel time reliability using GPS data

Zun Wang<sup>a\*</sup>, Anne Goodchild<sup>a</sup>, Edward McCormack<sup>a</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, 98195, USA

---

## Abstract

The objective of this paper is to develop a methodology for forecasting freeway vehicle travel time reliability for transportation planning using probe GPS data. Travel time reliability is measured using the coefficient of variation of the GPS spot (instantaneous) speed distribution. The proposed approach establishes relationships between travel time reliability and roadway traffic density in order to forecast reliability given future traffic conditions. The travel time reliability and traffic density datasets are segmented into different homogenous groups using the K-means cluster algorithm and the corresponding reliability-density relationship of each cluster is fitted by minimizing squared errors. This paper employs a truck probe GPS dataset as an example to demonstrate the proposed approach. The approach can be applied with any GPS datasets for forecasting reliability.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

*Keywords:* Travel time reliability forecasting; GPS; speed distribution; coefficient of variation; traffic density

---

## 1. Introduction

Travel time reliability represents the level of consistency in travel times for the same trip for a time period (Lomax et al. 2003). It has been recognized as one of the key factors influencing vehicle scheduling and routing choices. Due to the importance of travel time reliability and the different information described by mean travel time and travel time reliability, travel time reliability should be integrated into project prioritization processes, for example, benefit-cost analysis to support efficient freight planning. Despite the fact that there are considerable

---

\* Corresponding author. Tel.: +352-665-1286

E-mail address: [zunwang@uw.edu](mailto:zunwang@uw.edu)

quantitative project prioritization applications, only a few of them consider travel time reliability due to the lack of vehicle movement data and the complexity of forecasting reliability. The increasing market penetration of GPS technology for vehicle navigation and commercial vehicle fleet management purposes introduces new opportunities for collecting vehicle movement data and supporting travel time reliability forecasting. This paper proposes a methodology to forecast freeway vehicle travel time reliability using GPS data. The approach is expected to be used by transportation planners to forecast travel time reliability associated with transportation investments to support project prioritization and long-term transportation planning.

## 2. Background

This section first introduces a GPS spot speed based travel time reliability measure. The second part reviews the existing travel time reliability forecasting methods employed in transportation project prioritization applications. The third part of this section describes the traffic level of service (LOS) concept to illustrate the relationship between traffic density and vehicle performance.

### 2.1. GPS spot speed based travel time reliability measurement

Many previous research efforts have discussed how to measure travel time reliability using GPS data. Most existing approaches rely upon travel time observations, e.g. the travel time standard deviation approach, the 95<sup>th</sup> percentile method, and the buffer time method (Texas Transportation Institute and Cambridge Systematics Inc, 2006). GPS devices record vehicle spot (instantaneous) speed, rather than travel times. Thus these travel time based approaches require additional data processing efforts to convert GPS spot speed to travel time estimates. In addition, the data conversion may provide opportunities for introducing errors. To overcome these limitations, Zhao et al. (2012) proposed a truck GPS spot speed distribution based approach. This work demonstrated that the mixture of two Gaussian distributions provides the best fit for the truck GPS spot speed observations (refer to Fig. 2. for the fitted GPS spot speed distribution). The probability density function of mixture of two Gaussian distributions is shown in Equation 1. The parameters are fitted based on the maximum likelihood rule.

$$f(x) = w \cdot n(x, \mu_1, \sigma_1) + (1 - w) \cdot n(x, \mu_2, \sigma_2)$$

$$n(x, \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left[-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right] \quad (1)$$

where  $w$  = the proportion of the first normal distribution,  
 $\mu_1$  and  $\mu_2$  = mean of the first and second Gaussian distribution,  
 $\sigma_1$  and  $\sigma_2$  = standard deviation of the first and second Gaussian distribution.

Furthermore, they developed a rule to classify reliability into three categories: reliably fast, reliably slow and unreliable. Travel condition is defined as unreliable if and only if  $|\mu_1 - \mu_2| \geq |\sigma_1 + \sigma_2|$ ,  $w \geq 0.2$ , and  $\mu_1 \leq 0.75 \times V_p$  ( $V_p$  is the posted speed), otherwise, it is viewed as reliable. If travel condition is defined as reliable and the average speed is less than 75% of the posted speed ( $v \leq 0.75 \times V_p$ ), it is defined as reliably slow, otherwise, it is viewed as reliably fast. The primary advantage of this methodology is that the reliability evaluation does not require a large number of travel time observations, rather only spot speed. However, this method does not provide a numerical value which would allow for a more quantitative evaluation and ranking. Thus we further improved the approach by computing the coefficient of variation (COV) of the fitted mixture of two Gaussian distributions (Wang et al. 2015). The corresponding calculation is given in Equation 2 and 3. The improved approach is able to provide a numerical value which would allow for a more quantitative evaluation and reliability ranking. We employ this method to evaluate the current travel time reliability using GPS data. Also, the proposed reliability forecasting approach presented in this paper uses the COV of GPS speed distribution to quantify travel time reliability.

$$\mu = \sum_{i=1}^n w_i \mu_i \quad (2)$$

$$\sigma^2 = \sum_{i=1}^n w_i ((\mu_i - \mu)^2 + \sigma_i^2)$$

$$\text{Coefficient of Variation (COV)} = \frac{\sigma}{\mu} \quad (3)$$

where  $\mu$  = mean of the mixture of Gaussian distributions,

$w_i$  = weight of the  $i$ th Gaussian distribution,

$\mu_i$  = mean of the  $i$ th Gaussian distribution,

$\sigma$  = standard deviation of the mixture of Gaussian distributions,

$\sigma_i$  = standard deviation of the  $i$ th Gaussian distribution,

$n$  = number of Gaussian distributions,  $n=2$  since it has been proved that spot speed follows of the mixture of two Gaussian distributions.

## 2.2. Travel time reliability forecasting review

Puget Sound Regional Council (2009) benefit-cost analysis tool forecasts travel time reliability using the “certainty equivalent” approach. The certainty equivalent concept was defined as the decline of average speed that travelers are willing to accept to ensure speed reliability. The certainty equivalent was estimated based on the historical five-minute vehicle counts and speeds collected at multiple locations for hundred time periods by comparing the lowest 1 percent speeds with the mean speeds. However, this value is limited for Puget Sound region for mixed traffic. The values for a specific type of vehicle, e.g. passenger car or truck in this region, or other locations are unknown.

Peer et al. (2009) developed an economic model that can be used to predict travel time variability for benefit-cost analysis. Travel time reliability was defined as standard deviation of travel time observations. The authors investigated different variables’ impact on travel time reliability, including delay, segment volume-capacity ratio, weather, day-of-week, speed limit, segment length, shoulder width and etc. It is found that delay and segment volume-capacity ratio are the two most significant variables in improving the predictive power of the model. Similarly to the regression model developed by Peer et al., Eliasson (2006) developed models to quantify relationships between travel time reliability and traffic condition. Travel time reliability was represented by the relative standard deviation calculated by dividing standard deviation by travel time. Traffic condition was quantified by the relative travel time computed by dividing travel time by free-flow travel time. It was found that the relationships vary depends on traffic condition.

The above discussed regression models developed based on data collected by loop detectors and camera system. Black et al. (2008) developed an equation to establish a relationship between travel time reliability and congestion index using GPS data collected along 34 routes. Reliability in this equation was evaluated using the coefficient of variation of travel time. The analysis revealed that reliability is determined by travel distance and trip congestion index, which was computed by dividing average travel time by free flow travel time. Longer distance and higher congestion index lead to lower reliability. The authors retrieved travel times from GPS data instead of using GPS speed information directly for reliability forecasting, which requires additional data processing efforts and may lose information of GPS data due to data conversion.

## 2.3. Level of service

Level of Service is discussed in Highway Capacity Manual (HCM) (2010) as an indicator of traffic condition. It categorizes the traffic performance into 6 levels, denoted from A to F. The categorization of freeway traffic condition is determined by traffic density, as shown in Fig.1. When segment density is less than 11 vehicle/mi/ln,

travelers can travel at desired speed and are not affected by other vehicles, and the LOS is defined as A. The increase of traffic density leads to lower and less reliable speed. When density is greater than 45 veh/mi/ln, travelers experience considerable delay and the LOS is denoted as F.

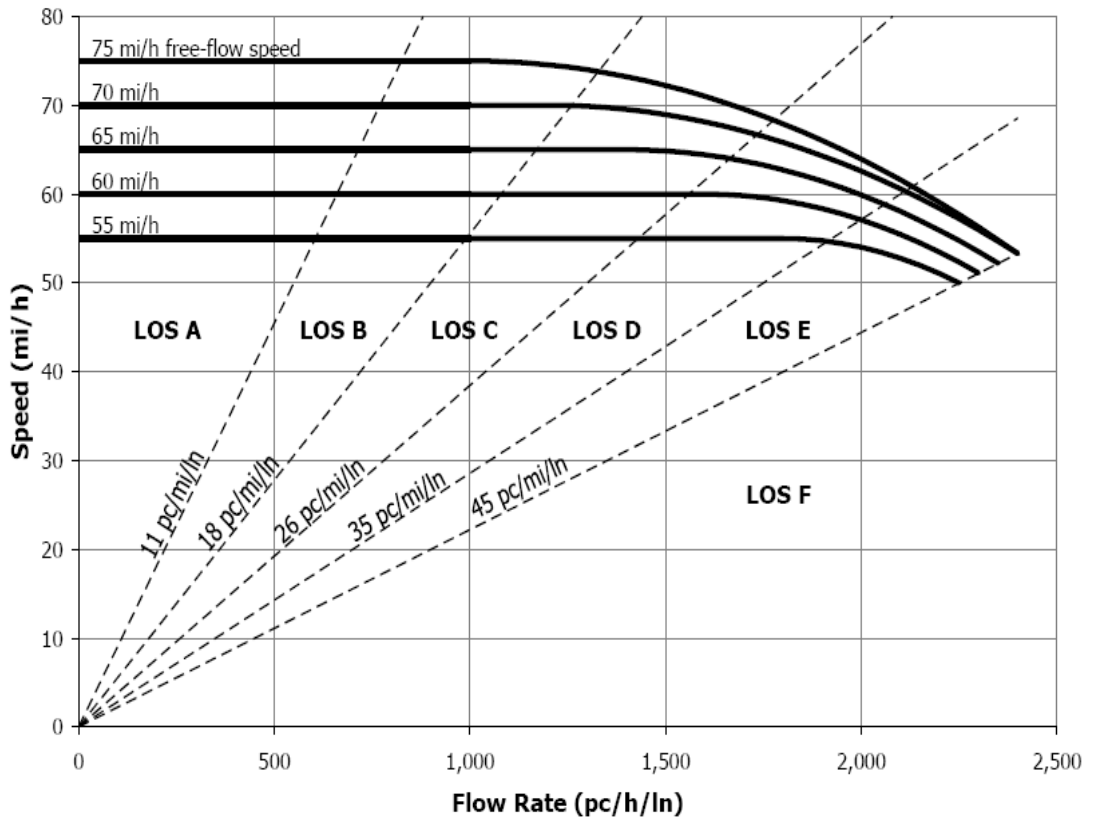


Fig. 1. Definition of Level of Service for Freeway Segment (Highway Capacity Manual, 2010)

The LOS definition indicates that, for a given freeway, there exist relationships between roadway density and traffic performance. Thus we employ this idea to forecast travel time reliability based on the relationships between segment density and spot speed distribution COV under different traffic regimes.

For forecasting travel time reliability, understanding how COV changes in response to different traffic conditions is the key since speed and reliability are associated with traffic condition closely. Under free-flow condition, traffic volume is low and most of the vehicles travel at a reliable and desired speed, which is close to free-flow speed. During congestion phase, vehicles stop and go due to influence of other vehicles, which results in unreliable travel time.

The following section uses a case study to demonstrate the changes of speed distribution and distribution COV in response to different traffic densities along a selected segment and consequently introduces the proposed reliability forecasting approach.

### 3. Data and analytical process

The purpose of forecasting reliability is to obtain reliability estimates for different traffic conditions. Future traffic conditions can be obtained from engineering experience and established models, e.g. the reduction in traffic volume/density by adding one lane.

To demonstrate the approach, a case study segment is identified. The segment is a stretch of 3.5 miles of southbound Interstate 5 (I-5) through the city of Seattle, Washington. Both GPS data and loop data were collected between January 2012 and December 2012. The GPS data was collected by GPS devices installed in commercial

vehicles traveling along the segment of interest. The data was cleaned to remove duplicated and problematic records (e.g. heading direction is greater than 360 degree or value of GPS spot speed is null). The cleaned data was geocoded to the corresponding network to reflect truck travel speed along the segment within ArcGIS environment. Details of the data cleaning and preparation can be found in (Zhao et al. 2012 and McCormack et al. 2011). The monthly GPS data was aggregated for one hour interval to generate hourly truck spot speed distributions. For each one hour dataset, the spot speed data was fitted using the mixture of two Gaussian distributions. The distribution fitting was accomplished using the R software, using the package called “mixdist” developed by Du (2002). The fitting process generates the mean values and standard deviations of the two normal distributions, and the probability of each distribution.

Traffic conditions are represented using traffic density computed based on traffic volume and speed information. Both traffic volume and speed were collected by Washington State Department of Transportation using the loop detectors deployed along the segment being studied. The data was recorded at every 20-second interval. Similar to the GPS data, monthly loop data was aggregated for one hour intervals to generate the average hourly traffic volume and hourly speed. Traffic conditions were quantified using traffic density obtained by dividing traffic volume by speed.

As demonstrated in Zhao et al. (2012), truck spot speed distribution follows a mixture of two Gaussian distributions, and it is either a unimodal or bimodal distribution. Truck travel time is classified as unreliable if spot speed distribution follows a bimodal distribution; otherwise travel time is reliable and the spot speed distribution follows a unimodal distribution. When travel time is reliable, most of the trucks travel at a constant and desired speed, and therefore the speed distribution follows a unimodal distribution. However, during the traffic congestion, trucks can be observed at much lower speeds, and therefore the presence of another speed distribution emerges. This generates a bimodal speed distribution representing two traffic regimes, the low speed regime and high speed one. The more unreliable the system is, the wider the speed distributions and the greater possibility of spot speed falling within the low speed regime.

Fig. 2 displays the hourly truck spot speed distributions between 5 AM and 11 AM in May 2012 along the segment being studied. The fitted distribution parameters, the speed distribution coefficient of variation (reliability) and the corresponding calculated hourly traffic density values are given in Table 1.

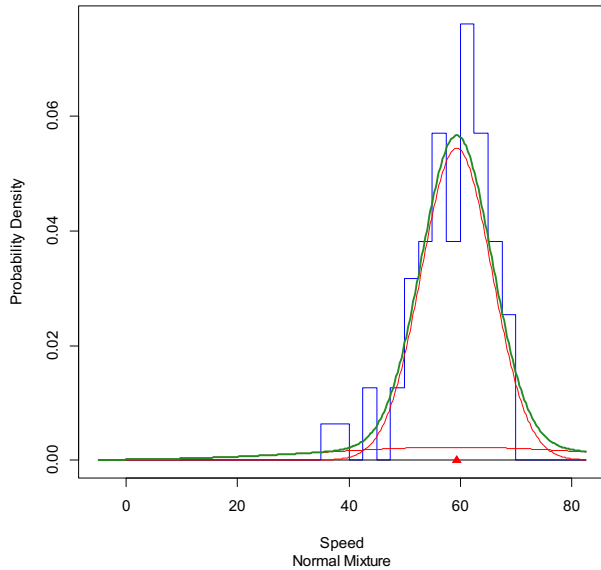
As shown in Fig. 2 (a) and Table 1, between 5 AM and 6 AM, the average segment density was 11.4 veh/mi/ln, and therefore it was classified as level of service (LOS) A according to HCM (2010). The description of LOS A indicates that this is a free flow traffic condition. Traffic density is low, with uninterrupted flow speeds controlled by driver desires, speed limits, and physical roadway conditions. Drivers can maintain their desired speeds with little or no delay. This description is consistent with the information conveyed by the spot speed distribution. According to Fig. 2 (a) the spot speed distribution followed a unimodal distribution and the reliability was defined as reliably fast.

Traffic density increased to 17.75 veh/mi/ln between 6 AM and 7 AM, and therefore fell into LOS B. The presence of the second distribution emerged. Namely, not all trucks were able to travel at the desired speed, but some of them were influenced by others and traveled at a lower speed. The spot speed distribution was composed of two normal distributions. According to Table 1, the average truck speed of the first distribution was 39.99 mph, with standard deviation of 13.31 mph; the average speed of the second distribution was 58.55 mph, with standard deviation of 3.85 mph. The probability of truck travel speed falling within the low speed traffic regime (the first distribution) was 17%. The average truck travel speed was 58.55 mph. Travel time was still classified as reliable based on the predefined rule. Meanwhile, according to the definition of LOS B, drivers still have reasonable freedom to travel at their desired speeds, but this is a low probability that traffic flow being restricted. These changes of speed distribution and truck performance are consistent with the transition from LOS A to LOS B.

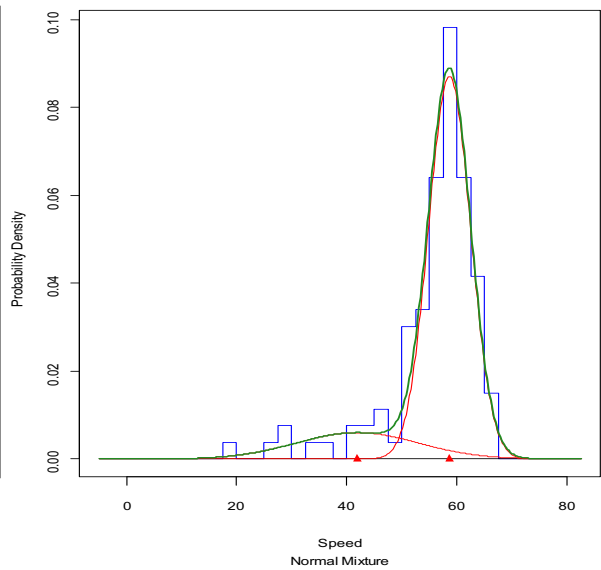
From Fig. 2 (b) to Fig. 2 (c), the probability of falling within the low speed regime increased to 47% with the increased traffic density of 22.55 veh/mi/ln, and a clear bimodal distribution was presented. Meanwhile, the average speed of the first traffic regime dropped to 20.76 mph, with standard deviation of 9.92 mph; the average speed of the second traffic regime decreased to 51.63 mph, with standard deviation of 7.87 mph. Travel time was changed to unreliable as evaluated by the predefined rules. The travel time reliability was reduced with the growth of traffic density. The density of 22.55 veh/mi/ln indicates LOS C. The definition of LOS C suggests that most drivers are restricted in selecting their own speed, but closely controlled by the higher volume.

Traffic density did not change significantly between 8 AM and 9 AM and a similar speed distribution was observed. The probability of falling within the low speed regime is 58%. Travel time was identified as unreliable based on the predefined reliability evaluation rules.

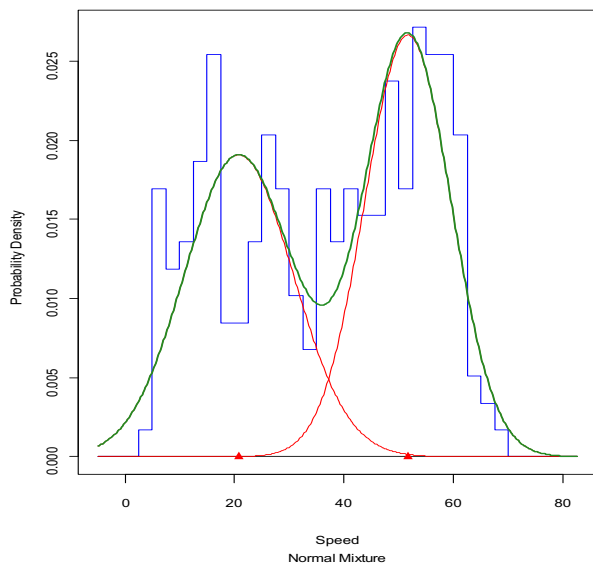
The low speed traffic regime started to wane about 9 AM while average traffic density decreased to 18.91 veh/mi/ln. The spot speed distribution showed the trend to change back to a unimodal distribution. Travel time between 9 AM and 10 AM and 10 AM to 11 AM were reliable.



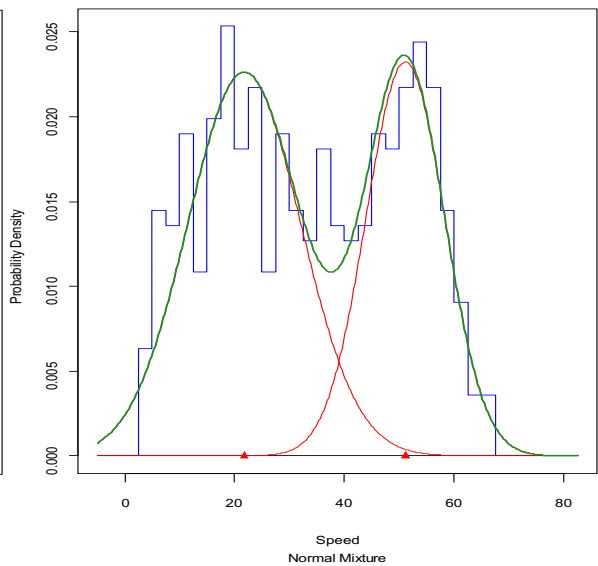
(a) 5 AM - 6 AM



(b) 6 AM - 7 AM



(c) 7 AM - 8 AM



(d) 8 AM - 9 AM

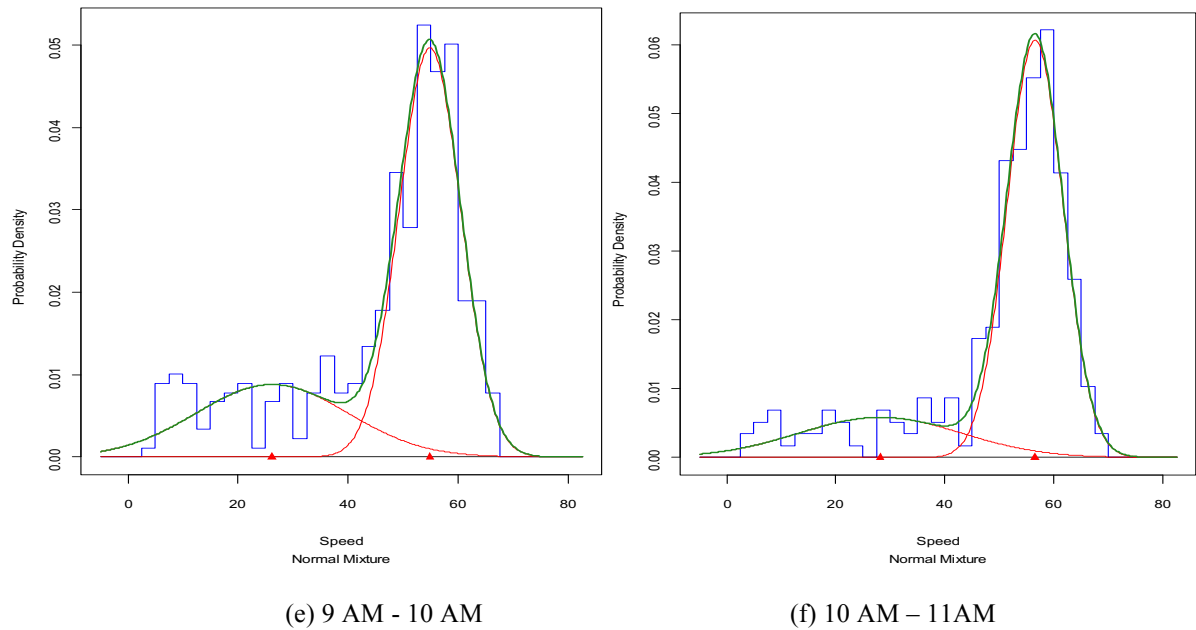


Fig. 2. Truck spot speed distributions during different time-of-day

Table 1. Truck spot speed distributions fitted results

	5 AM	6 AM	7 AM	8 AM	9 AM	10 AM
Traffic density (veh/mi/ln)	11.40	17.75	22.55	22.51	18.91	19.21
w	0.14	0.17	0.47	0.58	0.30	0.22
$\mu_1$	59.34	39.99	20.76	21.75	26.15	28.13
$\sigma_1$	24.33	13.31	9.92	10.28	13.77	14.93
$\mu_2$	59.34	58.55	51.63	51.08	54.82	56.52
$\sigma_2$	6.32	3.85	7.87	7.16	5.59	5.14
Spot speed distribution COV	0.18	0.17	0.48	0.50	0.34	0.29
$V_p$	60	60	60	60	60	60
Average speed	59.35	55.36	37.00	33.98	46.10	50.33
<i>if</i> $ \mu_1 - \mu_2  \geq  \sigma_1 + \sigma_2 , w \geq 0.2$	No	No	Yes	Yes	No	No
<i>and</i> $\mu_1 \leq 0.75 \times V_p$						
<i>if</i> average speed $\leq 0.75 \times V_p$	No	No	Yes	Yes	No	No
Reliability category	Reliably Fast	Reliably Fast	Unreliable	Unreliable	Reliably Fast	Reliably Fast

The above analyses indicate that truck spot speed distributions and travel time reliability, which is quantified by GPS spot speed distribution coefficient of variation, vary in response to different traffic conditions. Greater roadway density is associated with lower travel speed and at times, lower reliability.

This analysis also reveals that the distribution COV is strongly associated with segment density. As a result, we propose that travel time reliability can be forecasted based on the relationship between COV and density when future roadway density is available or predictable.

We further plotted the hourly segment density and hourly speed distribution COV based on one year truck GPS observations between January 2012 and December 2012, as displayed in Fig. 3. The figure illustrates that higher

traffic density is associated with higher speed distribution COV and lower travel reliability. The impacts of density on COV and reliability vary depending on the value of density. When density was between 15 veh/mi/ln and 28 veh/mi/ln, COV increased considerably with the growth of density. COV did not change substantially when density was greater than 28 veh/mi/ln. To partition data into homogeneous groups to identify relationships between density and COV, the K-means cluster algorithm was employed. The k-means cluster algorithm is a centroid-based clustering algorithm, which aims to find the k cluster centers and assign the data to the nearest cluster center whose mean yields the least within-cluster sum of squares (Hartigan, 1975).

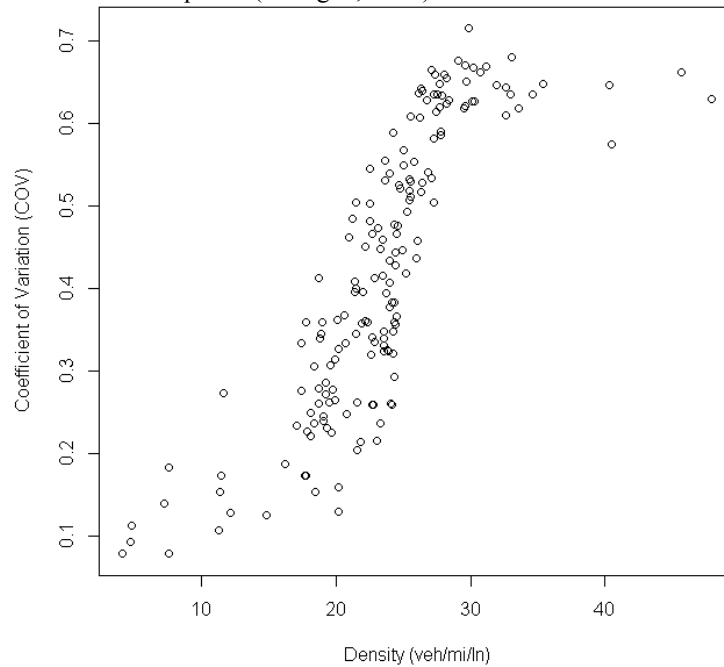


Fig. 3. Segment Density and Spot Speed Distribution COV Plot

For each identified group, the relationship between density and COV was regressed based on the least squares rule. Three commonly applied speed-density relationships were tested: linear, log-linear, and exponential relationship (Sun and Zhou, 2005). The fitted model that generates the greatest R squared value was chosen to fit the empirical data. The selected model represents the relationship between density and travel time reliability, and can be used to forecast reliability when future density is available or predictable. Fig. 4 illustrates the proposed reliability forecasting approach.



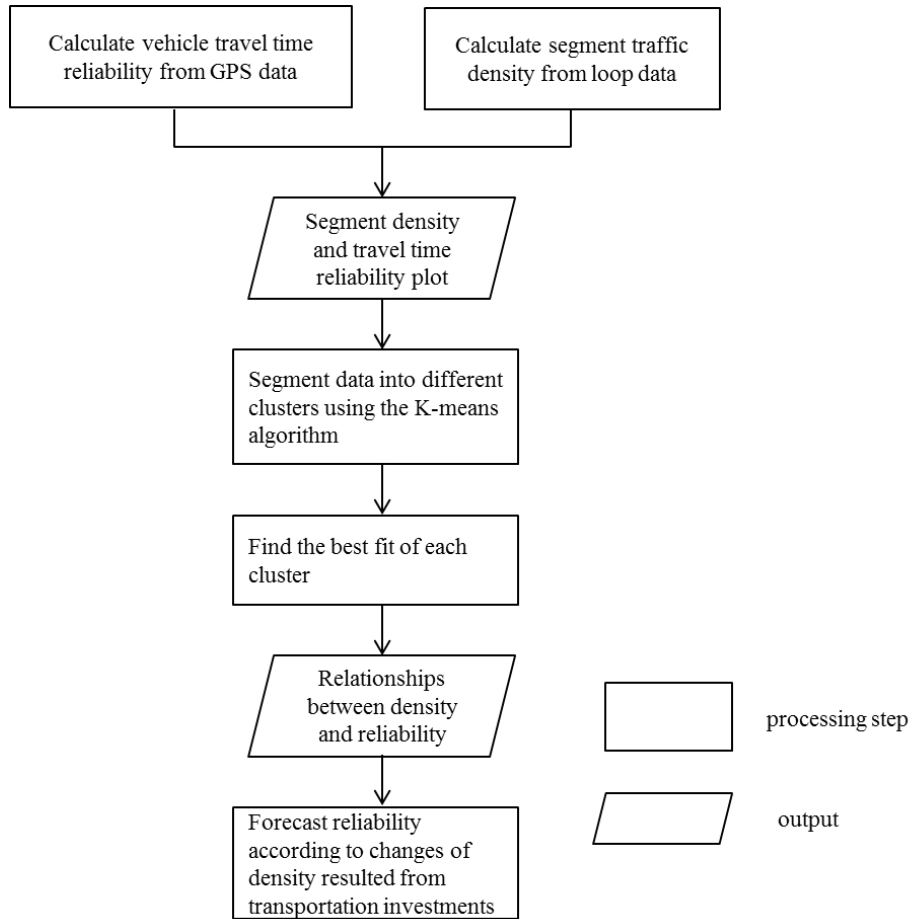


Fig. 4. Schematic of the Proposed Freeway Vehicle Travel Time Reliability Forecasting Approach

The next section presents a case study to illustrate the process of establishing the relationship between COV and density to support reliability forecasting.

#### 4. Case study

The above section implemented several reliability forecasting steps, including calculating vehicle travel time reliability from GPS data, calculating segment density from loop data and plotting reliability and density data. This section continues the forecasting analysis by implementing the rest of steps defined in Fig. 4.

##### *Segment data into different clusters*

We continue using the same data to illustrate the proposed method. By using the K-means algorithm, three clusters were identified representing different traffic regimes as well as relationships between density and travel time reliability. The three segmented clusters were denoted as 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> cluster as shown in Fig. 5. The clustering process was accomplished using the R software. The three clusters represent different traffic regimes.

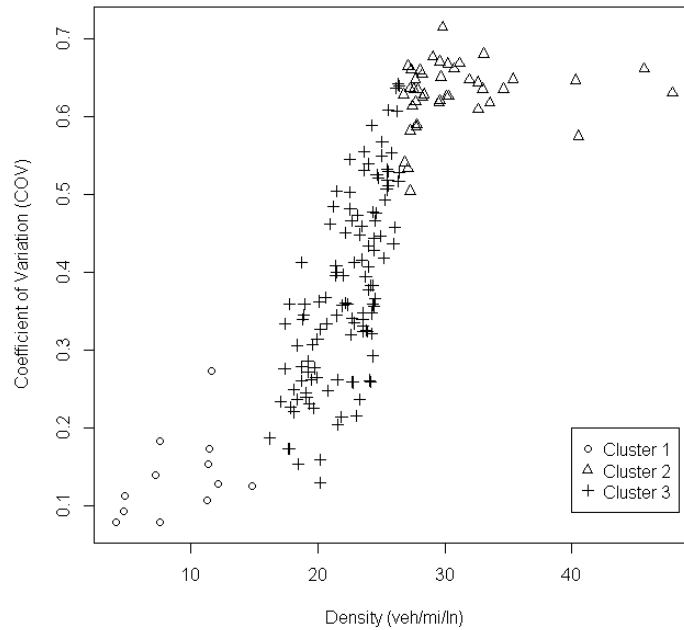


Fig. 5. K-Means Analysis Results of the Segment Density and Spot Speed Distribution COV Plot

For the first cluster, density was less than 18 veh/mi/ln and the corresponding COV was less than 0.2, (excluding one outlier). According to the rule developed to evaluate travel time reliability by Zhao et al. (2012), the truck travel time of the first cluster was assessed as reliable. This result is consistent with the LOS definition, in which LOS A and B are defined as free flow and near free flow conditions when density is less than 18 veh/mi/ln. Travelers travel at the desired or near-desired speed and are not influenced by other travelers.

When density reaches the value of 28 veh/mi/ln, it is classified as the second cluster. The COV does not change significantly with the growth of density, and stays between 0.6-0.7.

The COV increased considerably with the growth of density in the third cluster. The author fitted the data using linear regression, and obtained the relationship between COV and density as presented in Equation 4 and Table 2. Both the intercept and density are significant. The sign of density is positive, which confirms the adverse impact of density on reliability. The adjusted R-squared value is 0.68.

$$\begin{cases} COV = -0.442 + 0.037 \times density & 15veh / mi / ln < density < 28veh / mi / ln \\ COV = 0.631 & density \geq 28veh / mi / ln \end{cases} \quad (4)$$

Table 1. The second cluster fitting results

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	-0.442	0.049	-8.937	<0.0005
Density	0.037	0.002	17.624	<0.0005

For this case study, if density is less than 15 veh/mi/ln, vehicles experience reliable travel times and travel at a speed of 60 mph. The system is unreliable when density is greater than 15 veh/mi/ln. Reducing segment density can improve reliability considerably when density is between 15 veh/mi/ln and 28 veh/mi/ln. When density is greater than 28 veh/mi/ln, moderate density reduction does not affect reliability substantially; notable improvement can be observed only until density is reduced to less than 28 veh/mi/ln.

This cluster analysis based approach is expected to be applied for forecasting travel time reliability and support project prioritization, and to be conducted individually for each segment of interest using data from the segment itself. Potential changes of segment density resulting from highway investments can be estimated based on

engineering experience and established models. Since the approach is designed for supporting long-term planning, factors affect daily operation, e.g. weather and accidents, are not considered. The project prioritization process should also include sensitivity analysis to assess impacts of different density changes on future travel time reliability.

## 5. Conclusions and discussion

This paper proposes an approach to forecast freeway vehicle travel time reliability for planning purposes using GPS data. The proposal is a methodology that can be applied to any segment of interest for which GPS data is available. The authors analyze the changes in truck GPS spot speed distribution in response to different traffic conditions. The analysis reveals that traffic has considerable impact on speed distribution, and a bimodal distribution emerges with increasing traffic density. Greater speed distribution COV is closely associated with greater segment density, but the relationships are dependent on the traffic regime. A cluster analysis based approach is proposed to segment the COV and density dataset into several groups and identify the breakpoints. The authors further quantify the impact of density on COV by fitting the data using a linear model. The developed equation is able to forecast segment reliability with changes of density.

Given the simplicity of this approach and the increasingly available vehicle GPS data, we recommend users to apply this proposed approach, and develop their own equations for different locations for project prioritization and planning.

## References

- Lomax, T., D. Schrank, S. Turner, and R. Margiotta, Schrank, D., Selecting travel reliability measures, 2003, <http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/TTI-2003-3.pdf>, accessed April 2014.
- Texas Transportation Institute (TTI) and Cambridge Systematics, Inc., Travel time reliability: making it there on time, all the time, Rep. prepared for Federal Highway Administration, 2006.
- Zhao, W., E. McCormack, Dailey, D., Scharnhorst, E., Using truck probe GPS data to identify roadway bottlenecks, *J. of Transportation Engineering*, 2012, 139 (1), pp. 1-7.
- Wang, Z., A. Goodchild, E. McCormack, Measuring Truck Travel Time Reliability Using Truck Probe GPS Data, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 2015.
- Puget Sound Regional Council, Benefit-Cost Analysis: General Methods and Approach, 2009, [http://www.psrc.org/assets/2127/BCA\\_Methods\\_Report\\_Mar2010update.pdf](http://www.psrc.org/assets/2127/BCA_Methods_Report_Mar2010update.pdf), accessed May 2014.
- Peer, S., C. Koopmans, and E. Verhoef, Predicting travel time variability for cost-benefit analysis, 2009, <http://www.internationaltransportforum.org/Proceedings/reliability/Peer.pdf>, accessed May 2014
- Eliasson, J. and T. AB, Forecasting Travel Time Variability, 2006, [http://web.mit.edu/11.951/oldstuff/albacete/Other\\_Documents/Europe%20Transport%20Conference/value\\_of\\_time/forecasting\\_travel1726.pdf](http://web.mit.edu/11.951/oldstuff/albacete/Other_Documents/Europe%20Transport%20Conference/value_of_time/forecasting_travel1726.pdf), accessed March 2016.
- Black, I., Fearon, J., and Gilliam, C. (2008) Forecasting and Appraising Travel Time Variability in Urban Area, European Transport Conference. Highway Capacity Manual 2010.
- McCormack, E., Scharnhorst, E., and Zhao, W. Using GPS truck data to identify and rank bottlenecks in Washington State. Report prepared for Washington State Department of Transportation, Olympia, WA, 2011.
- Du, J., Combined algorithms for constrained estimation of finite mixture distributions with grouped data and conditional data, A Project Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science. McMaster University, Hamilton, CA, 2002
- Hartigan, J.A. Clustering algorithms. John Wiley & Sons, Inc. 1975
- Sun, L. and Zhou, J. Development of Multiregime Speed-Density Relationship by Cluster Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2005, No. 1934, 64-71