

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Final Reports & Technical Briefs from Mid-
America Transportation Center

Mid-America Transportation Center

2020

Analysis of Component Errors in the Highway Capacity Manual Travel Time Reliability Estimations for Urban Streets

Ernest Tufuor

Laurence Rilett

Follow this and additional works at: <https://digitalcommons.unl.edu/matcreports>



Part of the [Civil Engineering Commons](#), and the [Transportation Engineering Commons](#)

This Article is brought to you for free and open access by the Mid-America Transportation Center at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Final Reports & Technical Briefs from Mid-America Transportation Center by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Analysis of Component Errors in the Highway Capacity Manual Travel Time Reliability Estimations for Urban Streets

Ernest O. A. Tufuor and Laurence R. Rilett

Department of Civil and Environmental Engineering, University of Nebraska-Lincoln, Lincoln, NE

Corresponding author — Ernest O. A. Tufuor, ernest.tufuor@huskers.unl.edu

Abstract

The Highway Capacity Manual 6th edition (HCM6) includes a new methodology to estimate and predict the distribution of average travel times (TTD) for urban streets. The TTD can then be used to estimate travel time reliability (TTR) metrics. Previous research on a 0.5-mi testbed showed statistically significant differences between the HCM6 estimated TTD and the corresponding empirical TTD. The difference in average travel time was 4 s that, while statistically significant, is not important from a practical perspective. More importantly, the TTD variance was underestimated by 70%. In other words, the HCM6 results reflected a more reliable testbed than field measurement. This paper expands the analysis on a longer testbed. It identifies the sources and magnitude of travel time variability that contribute to the HCM6 error. Understanding the potential sources of error, and their quantitative values, are the first steps in improving the HCM6 model to better reflect actual conditions. Empirical Bluetooth travel times were collected on a 1.16-mi testbed in Lincoln, Nebraska. The HCM6 methodology was used to model the testbed, and the estimated TTD by source of travel time variability was compared statistically to the corresponding empirical TTD. It was found that the HCM6 underestimated the TTD variability on the longer testbed by 67%. The demand component, missing variable(s), or both, which were not explicitly considered in the HCM6, were found to be the main source of the error in the HCM6 TTD. A focus on the demand estimators as the first step in improving the HCM6 TTR model was recommended.

Published in *Transportation Research Record* 2020, Vol. 2674(6) 85–97

DOI: 10.1177/0361198120917977

Copyright © 2020 National Academy of Sciences: Transportation Research Board.

Published by SAGE Publications. Used by permission.

The Highway Capacity Manual (HCM) is used widely by U.S. transportation engineers for evaluating roadway performance, including travel time, congestion, delay, and so on (1). It has also been used by other nations to develop HCM estimation and prediction methodologies for their national conditions (2).

The previous editions of the HCM expresses arterial roadway performance in relation to the level of service (LOS) (2). The HCM defines LOS as “a quantitative stratification of a performance measure or measures that represent quality of service” (3). The intent of the LOS is to simplify decision making on the general acceptance of roadway performance (3). However, much of the complexity in roadway performance is hidden in the LOS (2). Arterial roadway users, such as commercial entities and commuters, are interested not only in the aggregation/ average conditions but also in measures of dispersion (e.g., variance) of the roadway performance, because both affect their daily travel (4).

In most of the literature, reliability is expressed as a function of a measure of central tendency and a measure of dispersion (5). There are many performance metrics used on arterial roadways (e.g., travel time, speed, delay) that can be examined from a reliability perspective. This paper will focus on arterial roadway travel time reliability (TTR).

In recent years, TTR has grown in popularity among transportation agencies (6). TTR is identified as a key roadway mobility measure by the U.S. Federal Highway Administration (7–9). TTR is an inherent part of travelers’ route choice decisions and is used by traffic managers to better quantify operations rather than simply using average travel times (10). Commercial truck operators also desire TTR information for effective trip planning.

It has been shown that random events are a prime contributor to unreliable roads. For example, about 60% of road congestion causes are from random events. This includes 25% from traffic incidents, 15% from bad weather, 10% from work zones, 5% from poor signal timings, and 5% is because of special events and other effects (6). Intuitively, reliability estimation and forecasting methods should explicitly account for these random events.

The latest and 6th edition of the Highway Capacity Manual (henceforth referred to as HCM6) has introduced a new methodology to estimate and predict TTR metrics for urban arterials under interrupted flow conditions. Specifically, the HCM6 methodology estimates and predicts the distribution of average travel times (TTD) by explicitly accounting for the effect of five sources of travel time variability: weather events, demand variations, traffic incidents, the presence of work zones, and special events (e.g., festivals and football game days). The estimated TTD is then used to estimate common TTR metrics, such as the Travel Time Index (TTI) and the Planning Time Index (PTI).

The inclusion of a TTR procedure in the HCM6 is a very important and impactful step for urban arterial road analysis. An overview and application of the HCM6 methodology will be provided in subsequent sections. A complete description is provided in the Strategic Highway Research Program Report 2 project L08 (7).

A literature search showed that an output from a corridor simulation model, CORSIM (version 5.1), was used to validate the through-vehicle delay model utilized in the HCM6 TTR model (7). In the 2019 meeting of the Interrupted Flow Group of the Transportation Research Board Committee on Highway Capacity and Quality of Service (AHB40), the developers of the HCM6 methodology confirmed to the authors that the HCM6 TTR model was not calibrated or validated with empirical travel time data.

A recent study by Tufuor and Rilett (11) compared the TTD estimated by using the HCM6 methodology to an empirical Bluetooth TTD on a 0.5-mi testbed in Lincoln, Nebraska. The authors found that there was a statistically significant difference between the HCM6 methodology estimated mean travel time and the empirical mean travel time. Because the difference was only 4 s, over an average travel time of 158 s, the authors concluded that the difference, while statistically significant, was not important from a practical perspective. More importantly, the authors found that the HCM6 methodology underestimated the variability in the TTD by 70%. Not surprisingly, the HCM6 TTR measures also underestimated field measurements. In other words, the HCM6 results reflected a much more reliable testbed than was measured in the field.

It is hypothesized that because of the central limit theorem, the HCM6 methodology may perform better on a longer testbed. Also, it was decided that the component errors within the HCM6 TTR model should be analyzed as this would provide insight into where the considerable differences in the HCM6 and the empirical TTD variance originated.

This paper first compares the HCM6 TTD and the empirical TTD on a 1.16-mi testbed in Lincoln, Nebraska. Subsequently, the sources of travel time variability that contribute to the error in the HCM6 methodology are quantified. Understanding the potential sources of error and their quantitative values are the first step to improving the HCM6 TTR model so that it better reflects actual conditions.

This paper is divided into five sections. In the first part, a description of the study testbed and the empirical TTD is provided. Next, the HCM6 methodology is briefly discussed and applied to model the testbed. Subsequently, the HCM6 estimated TTD for the testbed is analyzed and the TTR metrics compared with the empirical TTR metrics. This is followed by a statistical comparative analysis of the results. Finally, the differences are quantified and discussed, and the findings are presented.

Testbed and Field Data Description

A 1.16-mi section of N 27th Street in Lincoln, Nebraska, which is a principal urban arterial, was used as the testbed. This arterial is a main North–South corridor that links Interstate 80 with South Lincoln. The testbed is part of the Nebraska Transportation Center’s arterial monitoring corridor. The detailed operational performance of the testbed is described elsewhere (11).

Bluetooth detectors have been installed in the traffic cabinets for collecting empirical point-to-point Bluetooth travel time (BT) data. A comprehensive analysis of the BT data collection system, its filtering algorithm, and a validation process are provided elsewhere (11). The BT sampling rate was from 4% to 6%, which falls within the detection range of similar travel time studies (12–14). Public holidays, weekends, and football game days were removed from the BT data to eliminate periods that may cause outliers in the TTD. The filtering and elimination of these days reduced the sample size of the empirical BT data.

A total of 5893 individual vehicles’ BT data was collected within the p.m. peak hour (4:30–5:30 p.m.) for all weekdays from January 2016 to December 2016. The BT travel times were aggregated at a 15-min interval, which was the analysis period chosen for this paper. The mean number of samples per 15-min interval was six, with a standard deviation of about three samples.

The resultant BT TTD, which represents the TTD across all conditions, was disaggregated into the same categories considered by the HCM6: (a) normal conditions (no weather, work zone, or traffic incident); (b) weather (rain/snow) days; (c) work zone days; and (d) traffic incident periods. Figure 1, *a–d*, shows the TTDs based on normal, rain/snow, work zone, and combined conditions, respectively. Note that the weather and the work zone conditions are not mutually exclusive. For example, there are eight 15-min periods where there was a work zone active on the testbed and there was rain.

It may be seen from **Figure 1** that, there were 663 periods of normal conditions, 29 periods of snow and rain, and 152 periods when a work zone was active. It should be noted that it did not make sense to show the traffic incident histogram because it only consists of two data points. The BT for these periods was 189 and 178 s, respectively, with an average of approximately 15% greater than the average travel time under the normal condition.

Figure 1 shows that all four BT TTDs are positively skewed and range from a minimum of 130 s to a maximum of 291 s. The standard deviations range from 13.9 to 26.6 s. It may be seen in Figure 1*a* that the normal conditions constitute about 78% of the sample size of the combined conditions (Figure 1*d*). Not surprisingly, the characteristics of the BT TTDs for the normal and combined conditions are similar.

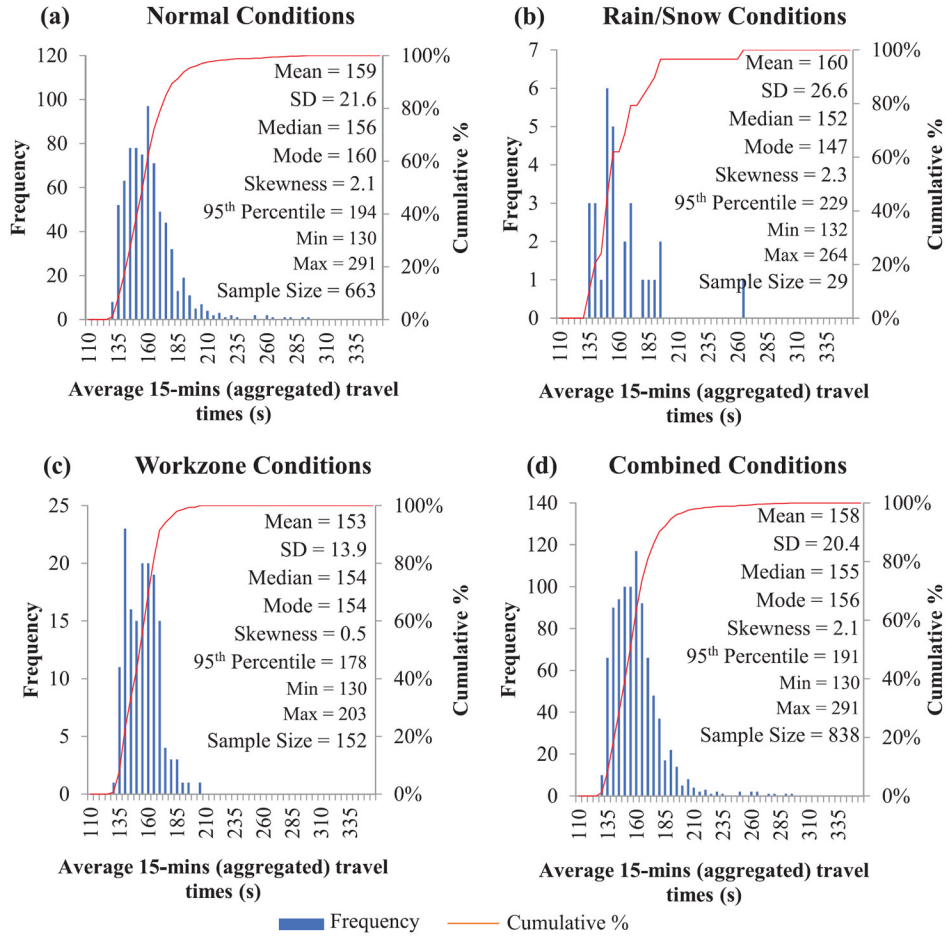


Figure 1. Empirical travel time distribution for weekdays in 2016 (4:30–5:30 p.m.). SD = standard deviation.

The mean of the BT TTD for the rain/snow conditions, as shown in Figure 1*b*, is higher than that for the normal conditions. More importantly, the standard deviation is 23% more than the normal conditions. This signifies that the testbed reliability will be lower during rain/snow conditions.

The work zone BT TTD in Figure 1*c* shows that the mean BT is 6 s faster than the mean BT of the normal conditions. Interestingly the variability is also lower, indicating the corridor is more reliable during work zone situations. It is hypothesized that drivers may have used alternative routes, which would have reduced the demand. Also, most of the work zones occur in the summer months when school is out, and demand is comparatively low, resulting in faster travel times.

Overview of the HCM6 TTR Methodology

To discuss the HCM6 TTR methodology, the following terms are useful.

- Reliability reporting period (I)—This is the number of days (I) over which TTR is to be estimated. The HCM6 allows (I) to span from 30 to 365 days and the user decides which days (weekdays, weekends, holidays, etc.) should be included. For the study testbed, (I) represents all weekdays in the year 2016 (e.g., $I = 261$ days).
- Analysis period (T_a)—This is the time interval evaluated by a single application of the HCM6 TTR methodology. The HCM6 allows for only 15- or 60-min intervals (15). In this paper, a 15- min aggregation period was chosen so $T_a = 15$ min.
- Study period (T_s)—This is the time interval that is studied within each day of the reliability reporting period. The HCM6 recommends that the study period be 60–360 min in duration (15). In this paper, the study period was from 4:30 to 5:30 p.m. Therefore, $T_s = 60$ min.
- Estimation periods in each day (J)—This is the number of T_a periods examined for each day and it is defined by Equation 1.

$$J = \frac{T_s}{T_a} \quad (1)$$

The value of J must be an integer. Therefore, T_s must be chosen so that it is evenly divided by T_a . For the testbed $T_s = 60$ min and $T_a = 15$ min, so $J = 4$.

- Every time period that is evaluated is referred to as a scenario. The number of scenarios is denoted by N and it is the product of I and J , as shown in Equation 2.

$$N = IJ \quad (2)$$

For the testbed, there are 261 days examined ($I = 261$) and there are four periods per day ($J = 4$), resulting in a total of 1044 scenarios examined. In other words, the resulting TTD estimated using the HCM6 procedure will consist of 1044 average 15-min travel times.

Figure 2 shows the flow chart of the HCM6 TTR methodology. A discussion of the four steps in the HCM6 TTR methodology is provided below. A comprehensive explanation of the HCM6 methodology can be found in the HCM (15) and the Strategic Highway Research Project 2 report L08 (1).

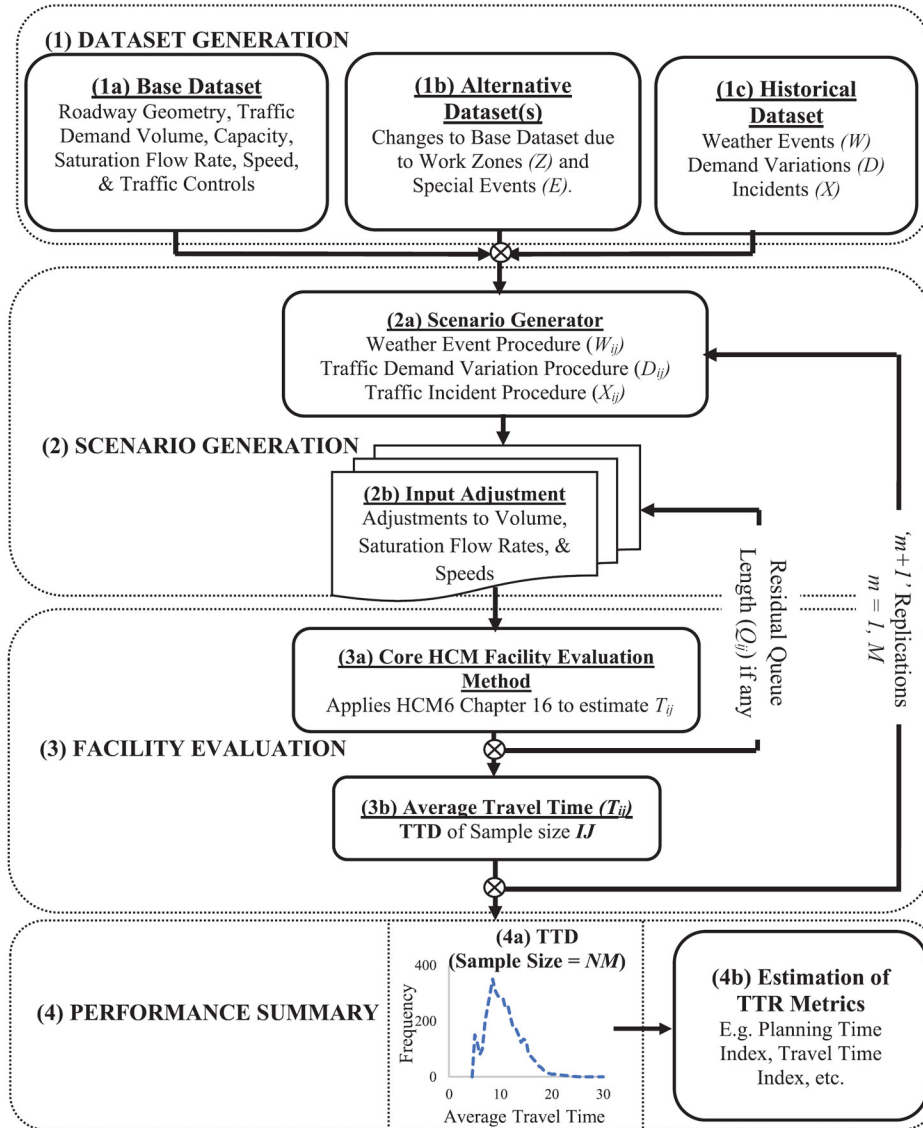


Figure 2. HCM6 travel time reliability methodology framework [leveraged from Exhibit 17-13 (15)]. HCM = Highway Capacity Manual; HCM6 = Highway Capacity Manual 6th edition; TTD = distribution of average travel times; TTR = travel time reliability.

Step 1: Dataset Generation Procedure

As shown in step 1 of Figure 2, the input dataset can be divided into three categories as follows.

- (1a) Base dataset—This dataset is used to describe the base conditions of the urban street. It consists of the supply features, traffic demand,

roadway capacity, saturation flow rate, and estimated free-flow speed for a specific day during the reliability reporting period where no rain/snow, incidents, work zones, and special events occur.

The testbed traffic demand data was collected by using the HCM6 protocol. The average annual daily traffic is approximately 26,500 (16). There were four signalized intersections on the testbed and each northbound and southbound movement had four driveways. The traffic demand volumes at all signalized intersections and all driveways were collected on March 30, 2016.

Traffic signal data from 2016 was obtained from the City of Lincoln. Traffic signal settings can vary by time of day and this may affect the variability in the performance of the arterial. The traffic signals on the testbed are operated in semi-actuated mode. However, because of the high traffic demand volume in the p.m. peak, the traffic signal essentially operates in fixed-time mode. Also, it was confirmed by the City of Lincoln that the signal timings were not optimized or changed in 2016. Consequently, the signals were assumed to operate under fixed-time control for the analysis conducted in this paper.

(1b) Alternative dataset(s) (Figure 2)—These datasets are used to describe the conditions when specific work zones (*Z*), special events (*E*), or both, occur. The HCM6 user must specify the changes in the base dataset (e.g., available lanes, traffic controls) associated with the work zone or special event. A maximum of seven alternative data files can be created and the number of days *I* that the work zone or special events occur is specified by the user.

Work zone data on the testbed was also provided by the City of Lincoln. A total of 38 days within the summer months had lane closure on two segments for road repair and maintenance work.

Because there were no weekday special events in 2016 and data from national holidays were not used, the special event source of variability was not studied.

(1c) Historical data (Figure 2): This dataset represents weather events (rain, snow, and temperature data by month), the traffic demand variation factors (e.g., hour-of-day, day-of-week, and month-of-year), and crash frequencies and crash adjustment factors. Exhibit 17-3 of the HCM6 (15) provides a comprehensive description of the general data required.

The HCM6 methodology requires weather data for the testbed. The following weather data from the Nebraska Mesonet database (17) was used in the analysis.

- Total normal precipitation (inches)—In 2016, there was a total of 28.9 in. of precipitation.
- Total normal snowfall (inches)—There was a total of 14.9 in. of snowfall in 2016.
- Number of days of precipitation (days)—The number of days when the rainfall or liquidequivalent of snowfall amount was greater than or equal to 0.01 in. There were 85 days of precipitation on the testbed in 2016.
- The normal daily mean temperature (degrees Fahrenheit)—In 2016, the average 24-h temperature was 55°F.

The HCM6 methodology requires demand variation factors for the testbed. There was no continuous traffic count data on the testbed. Therefore, demand variation factors on a similar urban principal arterial near the testbed, provided by the Nebraska Department of Transportation (18), were used. These factors account for the systematic traffic demand volume variation by hour-of-day (d_h), day-of-week (d_w), and month-of-year (d_m). For the testbed, there were four time periods, five weekdays, and 12 months, resulting in 240 unique combinations of demand factors.

The traffic demand volume for each scenario is estimated using a two-step process. First the demand modification factor for scenario ij is estimated using Equation 3. The numerator is the product of the hour-of-day, day-of-week, and month-of-year factor corresponding to the given scenario. The quotient is a similar product but for the base scenario that corresponds to the conditions when the base traffic demand volume was collected.

$$F_{ij} = \frac{d_{ij}^h d_{ij}^w d_{ij}^m}{d_b^h d_b^w d_b^m} \quad \forall i=1, I, \forall j=1, J \quad (3)$$

where

F_{ij} = demand modification factor for scenario ij .

d_{ij}^h = hour-of-day demand factor for scenario ij .

d_{ij}^w = day-of-week demand factor for scenario ij .

d_{ij}^m = month-of-year demand factor for scenario ij .

d_b^h = hour-of-day demand factor for base volume in step 1 of Figure 2.

d_b^w = day-of-week demand factor for base volume in step 1 of Figure 2.

d_b^m = month-of-year demand factor for base volume in step 1 of Figure 2.

In the second step, the traffic demand volume on the testbed intersections and segments is then calculated as the product of the demand modification factor (F_{ij}) and the base traffic demand volume for the network, as shown in Equation 4.

$$D_{ij} = F_{ij} D_b \tag{4}$$

where

D_{ij} = intersection and segment traffic demand volumes in scenario ij (vector).

D_b = intersection and segment traffic demand volumes in step 1 of Figure 2 (vector).

Figure 3, *a* and *b*, show the demand factors and a histogram of F_{ij} from Equation 3.

It may be seen from Figure 3*b* that the demand modification factor distribution has a relatively small range from 0.75 to 1.35 with 95% of the demand factors lying between 0.95 and 1.15. Intuitively, the traffic demand volume variability across all scenarios will also be relatively narrow.

The HCM6 TTR methodology requires the mean crash frequency for all segment-related and intersection-related crashes. In addition, crash frequency adjustment factors (CFAFs) for four weather events, active work zones, special events, or any combination of these, are required.

Traffic incident data for 2015–2017 on the testbed were provided by the City of Lincoln and used to estimate the mean crash frequency for all three segments and four intersections. Default CFAF values provided by the HCM6 were used because specific CFAFs were not available for the testbed conditions.

(a)

d^m		d^w		d^h	
J	0.829	M	0.980	16:00	0.072*
F	1.019	T	0.980	17:00	0.077
M	1.029*	W	1.000*		
A	0.980	T	1.030		
M	1.010	F	1.150		
J	1.047				
J	0.989				
A	1.052				
S	1.089				
O	0.948				
N	0.991				
D	0.939				

Note: (*) is the base demand factors

(b)

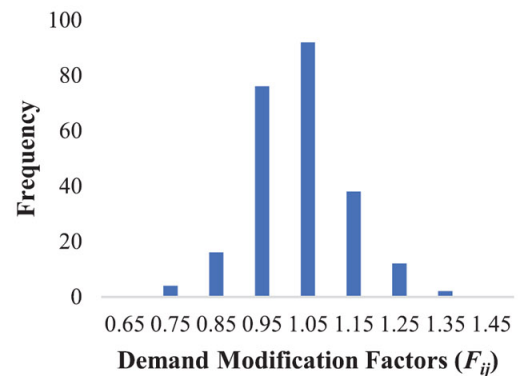


Figure 3. Distribution of demand factors and demand modification factors.

Step 2: Scenario Generation Procedure

The Scenario Generation procedure shown in step 2 of Figure 2 estimates values to the five sources of variability in travel times for each scenario ij . There are two main components of the scenario generation procedure—the deterministic component and the stochastic component.

The deterministic component uses the schedule of the work zone, special events provided by the analyst, or both, in the alternative dataset in step 1b of Figure 2. The output is a dummy variable indicating whether an active work zone occurs in each scenario ij ($Z_{ij}=1$). A dummy variable indicating whether a special event ($E_{ij}=1$) occurs for a scenario S_{ij} is also output.

Each of the N scenarios will have stochastic weather and incident values. If the analyst chooses a 15-min evaluation period, then the traffic demand volume values will also be stochastic. These values are obtained by sampling an underlying distribution using a Monte Carlo simulation procedure.

It is important to note that the stochastic component in step 2a of Figure 2 consists of three sequential procedures that begin with the prediction of weather events for all scenarios, then predicts demand variations (if a 15-min evaluation period is selected), and ends with predicting traffic incidents. The detailed description of all underlying distributions for the weather events, demand variations, and traffic incidents are provided elsewhere (7).

In summary, each scenario S_{ij} will randomly be assigned a weather event W_{ij} (rain, snow, neither rain nor snow), traffic demand volume D_{ij} (if 15-min analysis period is selected), and incident Y_{ij} (incident or no incident). In other words, if a scenario has a weather event (W_{ij}), there will be information on the type of precipitation, the intensity, and the length of time the pavement remains wet after the event. Similarly, a scenario with a predicted incident (Y_{ij}) will have information on the type of incident and the location on the subject facility.

Note that for a given day i , each subsequent scenario j is dependent on the scenarios that preceded it. For example, if S_{i1} is assigned a 2-h snow event then S_{i2} , S_{i3} , and S_{i4} will also be assigned a snow event.

The scenario generation procedure creates one file for each scenario in the reliability reporting period. Each file is an adjustment of the base or alternative dataset based on the W_{ij} and Y_{ij} values for each scenario. If there are weather or incidents during a given scenario, adjustments are made to saturation flow rates at intersections and segment speeds. For example, if in scenario ij the road pavement is modeled as being wet with no rain falling, then the speed in the base or alternative dataset is reduced by a factor of 0.95 (7).

In summary, the effect of weather W_{ij} and incidents Y_{ij} on saturation flow rates and speeds are quantified for each scenario S_{ij} . A file is generated for

each scenario S_{ij} and each file will have traffic demand volumes for each movement, saturation flow rate for each signalized intersection approach, and speed for each segment.

Step 3: Facility Evaluation Procedure

In the Facility Evaluation process shown in step 3 of Figure 2, two tasks are performed in sequence for each S_{ij} . Firstly, the core HCM facility methodology, which may be found in HCM6 Chapter 16 (15), is used to evaluate the first scenario S_{i1} by assuming that there are no initial queues. An estimated average travel time T_{i1} and estimated residual queue length (Q_{i1}) (if any) are output.

Secondly, the next scenario S_{i2} is evaluated. Specifically, the residual queue at the end of the first evaluation (Q_{i1}) becomes the initial queue input value for scenario S_{i2} . The corresponding estimated average travel time T_{i2} and the estimated residual queue length (Q_{i2}) are calculated for every intersection.

This procedure is done in sequence for each period j on a given day. The process is repeated for each day l . This results in a total of N average travel times, as shown in step 3b of Figure 2.

Step 4: Performance Summary

The Performance Summary is shown in step 4 of Figure 2. In this step, two operations are undertaken. Firstly, it is essential to note that because the results of each scenario are from simulated data, changing the random seed results in different scenario values for the weather (W_{ij}), traffic demand volume (D_{ij}), and incidents (Y_{ij}). The developers of the HCM6 TTR methodology recommend that the scenario generation (step 2) and the evaluation process (step 3) should be repeated M times with different seed numbers. This is meant to obtain robust TTR estimates (1, 15). All the travel time results are compiled into the TTD of sample size NM , as shown in step 4a of Figure 2.

Once the TTD has been determined, the TTR metrics can be readily estimated from the TTD. This is done as the second operation of the performance summary process, which is shown in step 4b of Figure 2.

It should be noted that the HCM6 methodology shown as Figure 2 has been automated. The only exception is the M loop and the performance summary (step 4). Specifically, the scenario generation (step 2) and the facility evaluation (step 3) are done automatically using the HCM6 computational software known as Street eValuation (StreetVal) (19). The user then runs this software M times and uses the NM travel times to form the final TTD. In this paper, four replications were undertaken ($M = 4$). The estimated average travel time for each scenario was randomly selected from these four to form the final TTD.

HCM6 Estimated TTD and TTR Metrics

Figure 4, a–d, shows the HCM6 estimated TTD for the testbed under normal conditions, rain/snow conditions, active work zone conditions, and the combined conditions respectively.

It may be seen in Figure 4 that all the HCM6 TTDs have a relatively small range that varies from approximately 146 s to approximately 173 s. In addition, there are very few outliers. The maximum average travel time is 332 s. There were 828 periods of normal conditions, 24 periods of snow and rain, and 188 periods of an active work zone conditions. It should be noted that the HCM6 had a total of two scenarios with traffic incidents. Similar to the empirical BT TTD, it did not make sense to show a histogram with two data points representing the traffic incidents. The estimated average travel times for these periods were 176 and 218 s respectively, which are approximately

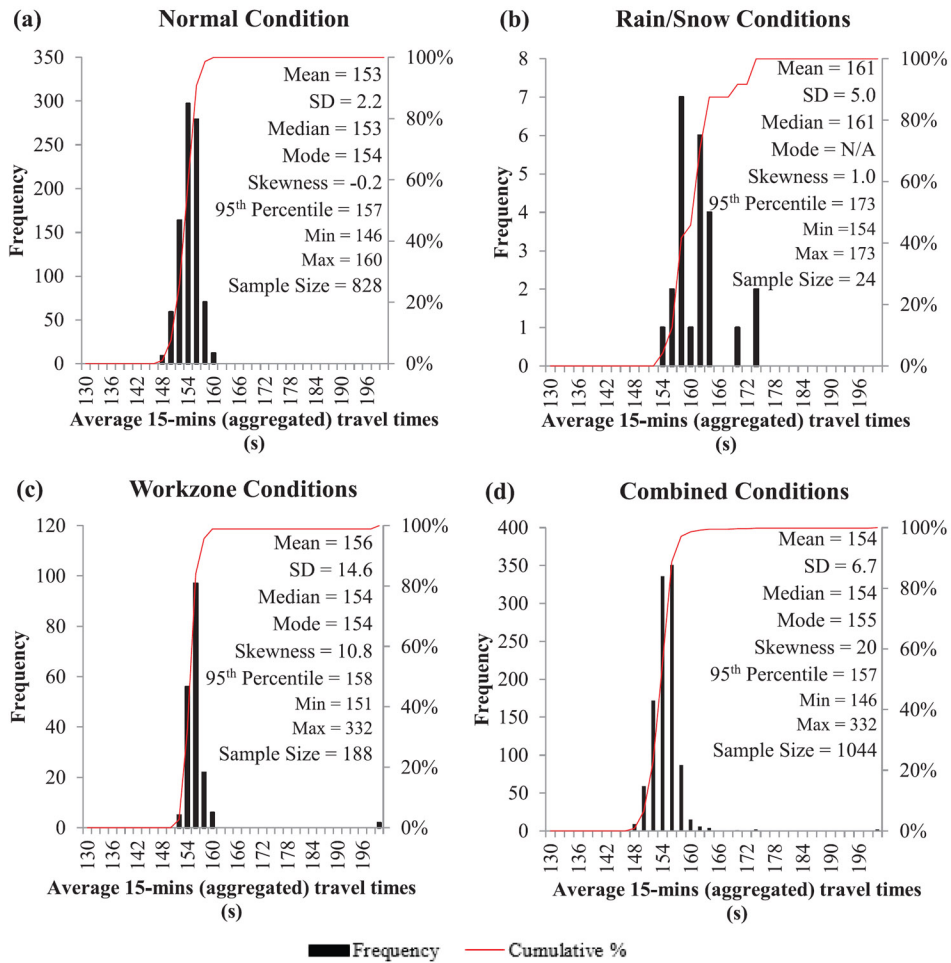


Figure 4. HCM6 estimated travel time distributions. SD = standard deviation.

29% more than the average travel time estimated during the normal condition. The two incident results are included in the combined TTD in Figure 4*d*.

Figure 4*a* shows the HCM6 TTD for normal conditions (i.e., no weather, incidents, or work zones). This condition shows a relatively low variance in average travel times. Specifically, the standard deviation is approximately 67% less than the combined condition. Also, the travel times have a shorter range from 146 to 160 s, which is 92% shorter than the combined condition. This indicates that the normal condition is predicted by the HCM6 to be relatively more reliable.

From Figure 4, *b* and *c*, the weather events and the work zone conditions have slightly higher mean travel times (e.g., 2–5% greater) than the normal conditions. More importantly, the standard deviation of the weather and work zone events are respectively 1.3 and 5.6 times greater than the standard deviation of the normal condition. This indicates that the reliability of the arterial testbed will be lower during the weather and the work zone events.

Quantification of TTR Metrics

This section compares common TTR metrics calculated using the HCM6 TTD and the empirical BT TTD. The FHWA (7) defines the following commonly used TTR metrics:

- (a) the TTI is the ratio of the mean travel time to the free-flow travel time;
- (b) the PTI represents the ratio of the 95th percentile time to the free-flow travel time; and
- (c) the Level of Travel Time Reliability (LOTTR) is the ratio between the 80th percentile travel time and the 50th percentile travel time.

Table 1 shows these TTR metrics, which were obtained from the estimated TTD.

Not surprisingly, the HCM6 tends to underestimate the TTR metrics. The only exceptions were for the TTI for rain/snow and work zone conditions, where the difference was less than 2%. The results in Table 1 show that the testbed is more reliable than the empirical data would suggest. For example, the HCM6 PTI estimation of 1.55 implies that for a trip of 100 s the trip maker must plan a total time of 155 s as compared to 190 s for the empirical case.

The percentage difference between the HCM6 and the empirical data was less than 5% for the TTI, ranges between 11% and 25% for the PTI, and ranges between 6% and 12% for the LOTTR.

Generally, the empirical TTIs are below 2.5, which indicates that congestion is not very high on the study facility and confirms the findings from a

Table 1. Reliability Performance Metrics

Testbed conditions	<i>TTR performance metrics</i>								
	<i>TTI</i>			<i>PTI</i>			<i>LOTTR</i>		
	<i>HCM6</i>	<i>BT</i>	<i>Diff.</i>	<i>HCM6</i>	<i>BT</i>	<i>Diff.</i>	<i>HCM6</i>	<i>BT</i>	<i>Diff.</i>
Demand variations	1.52	1.57	-3.6%	1.55	1.92	-19.3%	1.01	1.11	-8.4%
Rain/Snow	1.59	1.58	0.5%	1.71	2.27	-24.6%	1.02	1.16	-12.3
Work zone	1.54	1.52	1.7%	1.56	1.77	-11.4%	1.01	1.07	-6.0%
Combined	1.52	1.56	-2.5%	1.55	1.89	-17.8%	1.01	1.10	-7.8%

TTR = travel time reliability; TTI = Travel Time Index; PTI = Planning Time Index; LOTTR = Level of Travel Time Reliability; HCM6 = Highway Capacity Manual 6th edition; BT = Blue-tooth travel time; Diff. = difference.

previous study (11). Also, the percentage error of the TTR metrics under the combined conditions produced similar results for the 0.5-mi testbed analysis.

Statistical Comparison of Distributions

Different functional forms have been proposed in the literature to represent the distribution of link and corridor travel times. These include lognormal, gamma, and Weibull distributions (20–22). For the testbed condition, the quality of the best-fit distribution was determined by using the Kolmogorov–Smirnov (KS), Anderson–Darling, and Cramer–von Mises goodness-of-fit (GOF) tests. The lognormal distribution was determined to best model the HCM6 TTD and the empirical BT TTD on the corridor. This can inform changes in the underlying distributions used in the scenario generation procedure in step 2 of Figure 2.

The Welch's t -test was used to test the differences between the mean values of the simulated and empirical TTDs. This test was selected because, compared to the Student's t -test, it controls the Type I error when comparing unequal variance and unequal sample size datasets (23). The Mann–Whitney–Wilcoxon rank-sum test was used to test the differences between the median values. The Mann–Whitney test is one of the powerful nonparametric tests; its statistical power corresponds to the likelihood of rejecting a false null hypothesis (24). The KS test was used to test the differences between the cumulative distribution functions (CDFs) of the HCM6 TTDs (CDF_{HCM}) and the empirical BT TTDs (CDF_{BT}).

The following hypotheses at a 95% confidence level were tested.

- (a) Welch's t -test: The null hypothesis is that the mean of the population of the HCM6 simulated travel times (m_{HCM}) is equal to the mean of the

- population of the empirical BT travel times (m_{BT}). The alternative hypothesis is $m_{HCM6} \neq m_{BT}$.
- (b) Mann–Whitney–Wilcoxon tests: The null hypothesis is that the median of the population of the HCM6 simulated travel times (Med_{HCM6}) is equal to the median of the population of the empirical BT travel times (Med_{BT}). The alternative hypothesis is $Med_{HCM6} \neq Med_{BT}$.
- (c) KS test: The null hypothesis is that the CDF_{HCM6} = CDF_{BT}: The alternative hypothesis is that the CDF_{HCM6} \neq CDF_{BT}:

The statistical test results are shown in **Table 2**.

It was found that there were statistically significant differences between the population means and the population medians of the two datasets for the normal conditions. The differences were 6 and 3 s between sample means and medians, respectively as shown in Figures 1a and 4a.

For weather (snow and rain) conditions it may be seen in Table 2 that there are no statistically significant differences between the population means and the population medians of the two datasets. The differences were a single second and 9 s between the sample means and medians, respectively, as shown in Figures 1b and 4b.

It can be found in Table 2 that for the work zone conditions there were no statistically significant differences between the mean and the median values of the two population TTDs. The difference in the sample mean was 3 s and there were no differences in the sample medians, as shown in Figures 1c and 4c.

Table 2. Statistical Test Results

Conditions	Test statistic (p-value)			Remarks
	Mann–Whitney– Wilcoxon	Welch t-test	KS test	
Normal HCM6 versus BT	262440 (2.3e-05)	−7.05 (4.2e-12)	0.46 (2.2e-16)	Statistically significant evidence to reject the null hypotheses.
Weather HCM6 versus BT	451 (0.07)	0.20 (0.845)	0.58 (3.0e-4)	Statistically significant evidence not to reject the null hypotheses of the mean and median tests.
Work zone HCM6 versus BT	15567 (0.191)	1.68 (0.094)	0.43 (4.65e-14)	However, there are statistically significant differences in the CDFs.
Combined HCM6 versus BT	404940 (0.0055)	−5.72 (1.4e-08)	0.41 (2.2e-16)	Statistically significant evidence to reject the null hypotheses.

KS = Kolmogorov-Smirnov; HCM6 = Highway Capacity Manual 6th edition; BT = Bluetooth travel time; CDFs = cumulative distribution functions.

It may also be seen in Table 2 that there are statistically significant differences between the population mean and the median of the HCM6 TTD and the BT TTD for the combined conditions. The difference in average travel time was 4 s and the difference in median values was a single second, as shown in Figures 1*d* and 4*d*. While the differences and means and medians were statistically significant, they are not important from a practical perspective.

Error Estimation

The pairwise comparison of the differences between the CDFs and the frequency distributions of the HCM6 and BT TTDs are shown in **Figure 5**. The root mean square error (RMSE), defined in Equation 5, is also shown in Figure 5.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{HCM}_i - \text{BT}_i)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} \quad (5)$$

where

HCM_i = The frequency of class i of the HCM6 travel times.

BT_i = The frequency of class i of the BT travel times.

n = The number of bins. In this paper, the bin size was 5 s and there were 32 bins.

For each RMSE calculation, the bin width was 5 s within the range of 130–290 s. The number of bins is 32.

It may be seen from Figure 5 that the HCM6 TTDs have considerably less variability as compared to the observed data. The KS test results in Table 2 confirmed that there were statistically significant differences between the HCM6 and empirical TTDs. Specifically, the HCM6 TTD combined condition standard deviation was determined to be 67% less than that of the empirical BT TTD. More importantly, the HCM6 methodology indicates that the arterial is more reliable than would be indicated by field measurements.

These results conform to the findings of the previous studies on a shorter testbed (11). Therefore, doubling the length of the testbed did not substantially change the performance of the HCM6 predictions.

The RMSE quantitatively measures the differences between the TTDs, as shown in Equation 5. From Figure 5*a*, the differences in the normal condition TTDs have the largest RMSE. Specifically, the normal conditions RMSE was 82% of the combined condition RMSE. It can be posited that the TTD of normal conditions (no rain/snow, incident, work zone, or special events) represents the effect of one or more combination of factors such as the demand

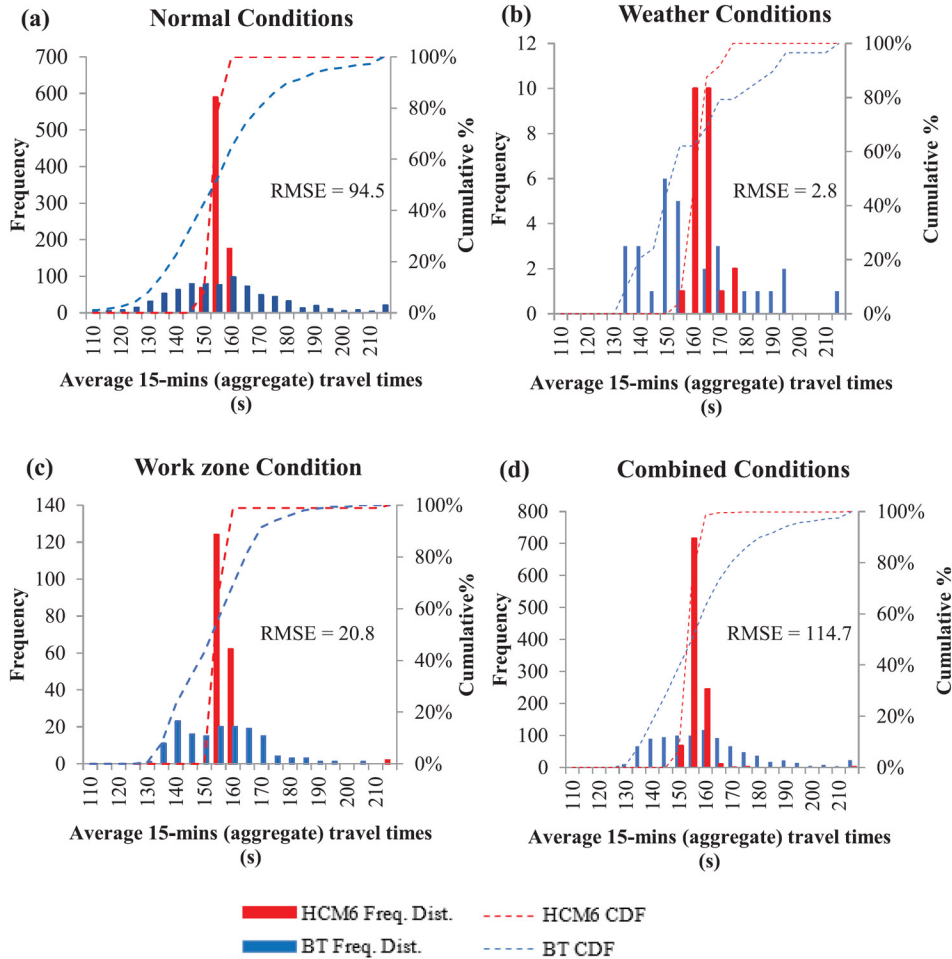


Figure 5. Pairwise comparison of travel time distributions. RMSE = root mean square error; HCM6 = Highway Capacity Manual 6th edition; BT = Bluetooth travel time; CDF = cumulative distribution function.

variations or not having detailed volume data (e.g., by day) or other missing variable(s) that were not explicitly considered in the HCM6 TTR model.

Concluding Remarks

This paper expands the analysis of the HCM6 estimated TTDs on a longer testbed. More importantly, it identifies the sources and magnitude of travel time variability that contribute to the HCM6 error. These sources were weather events, demand variations, traffic incidents, and work zones. Understanding the potential sources of error and their quantitative values are the

first steps in improving the HCM6 TTR model to better reflect actual conditions. The following was found.

- The mean and variance of the work zone HCM6 estimation were similar to the corresponding empirical BT measurement. The number of traffic incidents was only two; therefore, using the default CFAFs will not significantly affect the variability in the final TTD. Consequently, the potential source of error from the work zone and incidents was ignored for this testbed condition.
- One potential source of the error is the use of aggregated data as input in the HCM6. For example, monthly aggregations of the weather parameters could affect the variability in the resulting TTD. However, the RMSE of the weather conditions, as shown in Figure 5*b*, was relatively small. Specifically, the weather RMSE was 2% of the RMSE of the combined conditions in Figure 5*d*. Consequently, it was concluded that the weather assumption did not overly contribute to the error for this testbed.
- The analysis indicated that there are two potential sources of error for the differences between the estimated and measured TTDs. The first is that the traffic demand volume data, demand factors, or both, do not adequately capture the volume variability in the field. The second is that there may be other variables that are not considered in the HCM6 estimations. Arguably, one of the limitations of the HCM6 TTR methodology is the use of a single-day observed traffic demand volume as the basis to determine the traffic demand volume for all scenarios. Intuitively, if the one-day volume is atypical, the final TTD may also be inaccurate.

It is important to note that the results are only valid for one corridor for one year in one city. Further study of other corridors in other locations is recommended. In addition, the authors recommend that the HCM6 TTR model be calibrated to local data. It is hypothesized that the demand estimators are good candidates for improving the HCM6 TTR model.

Acknowledgments Our sincere gratitude goes to the City of Lincoln, and especially to Andy Jenkins, for making the testbed data available. The contents of this paper reflect the views of the authors, who are responsible for the facts and accuracy of the information presented herein and are not necessarily representative of the City of Lincoln.

Author Contributions The authors confirm contribution to the paper as follows: study conception and design: E. O. A. Tufuor, L. R. Rilett; data collection: E. O. A. Tufuor; analysis and interpretation of results: E. O. A. Tufuor, L. R. Rilett; and manuscript preparation: E. O. A. Tufuor, L. R. Rilett. All authors reviewed the findings and approved the final version of the manuscript.

Conflicting Interests The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding The authors received no financial support for the research, authorship, and/or publication of this article.

References

1. Zegeer, J., J. A. Bonneson, R. G. Dowling, P. Ryus, M. Vandehey, W. Kittelson, N. M. Roupail, B. J. Schroeder, A. Hajbabaie, B. Aghdashi, and T. Chase. *SHRP 2 Report S2-LO8-RW-1: Incorporating Travel Time Reliability into the Highway Capacity Manual*. National Research Council, Transportation Research Board of the National Academies, Washington, D.C., 2014.
2. Roess, R. P., and E. S. Prassas. *The Highway Capacity Manual: A Conceptual and Research History*. Springer International Publishing, Switzerland, 2014, pp. 249–338.
3. *Highway Capacity Manual 2010*. Transportation Research Board of the National Academies, Washington, D.C., 2010.
4. Figliozzi, M. A., N. Wheeler, E. Albright, L. Walker, S. Sarkar, and D. Rice. Algorithms for Studying the Impact of Travel Time Reliability along Multisegment Trucking Freight Corridors. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2224: 26–34.
5. Wu, Z. *Measuring Reliability in Dynamic and Stochastic Transportation Networks*. PhD dissertation. Department of Civil Engineering, University of Nebraska-Lincoln, 2015.
6. Federal Highway Administration. *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation*. No. FHWA-HOP-05-064. Federal Highway Administration, Washington, D.C., 2005.
7. Federal Highway Administration. *Travel Time Reliability: Making It There on Time, All the Time*. Federal Highway Administration, U.S. Department of Transportation, Office of Operations Report, 2006. https://ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm . Accessed January 12, 2019
8. Federal Highway Administration. *Fixing America Surface Transportation Act*. Federal Highway Administration, U.S. Department of Transportation, 2015. <https://www.fhwa.dot.gov/fastact/> . Accessed December 9, 2019.
9. Federal Highway Administration. *Moving Ahead for Progress in the 21st Century*. Federal Highway Administration, U.S. Department of Transportation, 2012. <https://www.fhwa.dot.gov/map21/> . Accessed December 9, 2019.
10. Transportation System Management and Operations. *Organizing and Planning for Operations*. Federal Highway Administration. <https://ops.fhwa.dot.gov/tsmo/index.htm> . Accessed December 9, 2019.
11. Tufuor, E. O. A., and L. R. Rilett. Validation of the Highway Capacity Manual Urban Street Travel Time Reliability Methodology using Empirical Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 415–426.
12. Haghani, A., M. Hamed, K. F. Sadabadi, S. Young, and P. Tarnoff. Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, 2010. 2160: 60–68.

13. Brennan, T.M., Jr, J. M. Ernst, C. M. Day, D. M. Bullock, J. V. Krogmeier, and M. Martchouk. Influence of Vertical Sensor Placement on Data Collection Efficiency from Bluetooth MAC Address Collection Devices. *Journal of Transportation Engineering*, Vol. 136, pp. 1104–1109, 2010.
14. Hainen, A. M., J. S. Wasson, S. M. L. Hubbard, S. M. Remias, G. Farnsworth, and D. M. Bullock. Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2256: 43–50.
15. *Highway Capacity Manual: A Guide for Multimodal Mobility Analysis*, 6th ed. Transportation Research Board, Washington, D.C., 2016.
16. The City of Lincoln. *Public Works & Utility Department. Traffic Operations*. <https://lincoln.ne.gov/city/ltu/engine/traffic/adtv/pdf/2015-citywide.pdf> . Accessed May 5, 2019.
17. University of Nebraska, Lincoln. *Weather Data*. <http://awdn.unl.edu/classic/home.cgi> Accessed January 12, 2019.
18. Nebraska Department of Transportation. *2018 Continuous Traffic Count Data and Traffic Characteristics on Nebraska Streets and Highways*. <https://dot.nebraska.gov/media/3811/annual-traffic-count-data.pdf> . Accessed November 5, 2019.
19. StreetVal Beta Version. <https://sites.google.com/site/jbreportsandtools3/home/streetval-beta-version> . Accessed February 20, 2019.
20. Polus, A. A Study of Travel Time and Reliability on Arterial Routes. *Transportation*, Vol. 8, No. 2, 1979, pp. 141–151.
21. Al-Deek, H., and E. B. Emam. New Methodology for Estimating Reliability in Transportation Networks with Degraded Link Capacities. *Journal of Intelligent Transportation Systems*, Vol. 10, No. 3, 2006, pp. 117–129.
22. Tufuor, E. O. A., and L. R. Rilett. *Analysis of Low Cost Bluetooth-Plus-WiFi Device for Travel Time Research*. Presented at 97th Annual Meeting of the Transportation Research Board, Washington, D.C., 2018.
23. Derrick, B., D. Toher, and P. White. Why Welch’s Test Is Type I Error Robust. *The Quantitative Methods in Psychology*, Vol. 12, No. 1, 2016, p. 30.
24. Landers, J. *Quantification in History, Topic 4: Hypothesis Testing II-Differing Central Tendency*. All Souls College, Oxford, 1981.