Integrating cumulative plots and probe vehicle for travel time estimation on signalized urban network

Ashish Bhaskar, EPFL-LAVOC, Switzerland
Edward Chung, QUT, Brisbane, Australia
André-Gilles Dumont, EPFL- LAVOC, Switzerland

Conference paper STRC 2009

STRC | **9th Swiss Transport Research Conference**
Monte Verità / Ascona, September 9. - 11. 2009

Integrating cumulative plots and probe vehicle for travel time estimation on signalised urban network

Ashish Bhaskar
EPFL-LAVOC
Lausanne, Switzerland
(Corresponding Author)

Phone: +41 21 693 23 41
Fax: +41 21 693 23 49
ashish.bhaskar@epfl.ch

Prof. Edward Chung
Queensland University of Technology,
Brisbane, Australia

Phone: +61 7 3138 1143
Fax: +61 7 3138 1170
edward.chung@qut.edu.au

Prof. André-Gilles Dumont
EPFL-LAVOC
Lausanne, Switzerland

Phone: +41 21 693 23 45
Fax: +41 21 693 23 49
andre-gilles.dumont@epfl.ch

September 2009

Abstract

Travel time is an important network performance measure and it quantifies congestion in a manner easily understood by all transport users. In urban networks, travel time estimation is challenging due to number of reasons such as, fluctuations in traffic flow due to traffic signals, significant flow to/from mid-link sinks/sources, etc. In this research a methodology, named CUmulative plots and PRobe Integration for travel timE estimation (CUPRITE), has been developed, and validated for average travel time estimation on signalized urban network.

The basis of CUPRITE lies in the classical analytical procedure of utilizing cumulative plots at upstream and downstream locations for estimating travel time between the two locations. The classical procedure is vulnerable to detector counting error and non conservation of flow between the two locations that induces relative deviation amongst the cumulative plots (RD). The originality of CUPRITE resides integration cumulative plots and probe vehicle data to address RD issue.

CUPRITE is validated with real data collected from number plate survey at Lucerne, Switzerland. Two tailed t-test (at 0.05 level of significance) results confirm that travel time estimates from CUPRITE are statistically equivalent to real estimates from number plate survey.

Keywords

Average travel time, urban network, signalized network, cumulative plots, probe vehicle, detector error, mid-link sink, mid-link source

1. Introduction and literature review

Travel time is the time needed to travel from point upstream (*u/s*) to point downstream (*d/s*) on the network. It quantifies congestion and is an important network performance measure. It has the potential to spatial temporal dissipation of congestion. Researchers have applied different techniques for travel time estimation based on the availability of data.

A loop detector provides traffic flow characteristics at a specific location. They are the oldest and most widely used traffic data sources. Researchers have proposed a number of models, based on detector data, with various degrees of complexities ranging from simple regression (Wardrop, 1968, Young, 1988, Sisiopiku and Rouphail, 1994, Sisiopiku et al., 1994), traffic flow theory (Oh et al., 2003, Nam and Drew, 1999), pattern recognition (You and Kim, 2000, Bajwa et al., 2003, Robinson and Polak, 2005, Dailey, 1993, Coifman and Krishnamurthy, 2007), to advance neural network (Park and Rilett, 1998, Chen et al., 2001, Liu et al., 2006) based.

Data driven models such as, regression, pattern recognition and neural network based models, are non transferable. Regression based models define its parameters by best fitting the observed data and should not be applied for traffic conditions that are different from those assumed in the model's formulation. They are simple and fast to compute and hence are favorable for transport planning and policy applications. Neural network based models utilize the data to build the model structure as well as its parameters and are more robust than regression based. They should be applied well within the limits for which it is trained. Pattern recognition models, such as k-NN technique, match the current traffic pattern with historical database and can fail to estimate travel time for traffic conditions absent in the database.

Mobile sensors such as probe vehicle (e.g., taxi fleet) is a vehicle equipped with vehicle tracking equipment (e.g., GPS) and can provide data for the vehicles' trajectory (time stamp and position coordinates) and hence its travel time. In practice, only few vehicles are equipped with mobile sensors and hence they represent the random sample from the population of the vehicles traversing the link. Therefore, average travel time for all the vehicles traversing the link can be estimated by statistical sampling techniques (Hellinga and Fu, 2002). Researchers (Srinivasan and Jovanis, 1996) have shown interest to determine minimum number of probes required for statistically significant travel time estimation.

Researchers have also applied data fusion techniques (Berka et al., 1995, Westerman et al., 1996, Choi and Chung, 2002, El Faouzi, 2004) to fuse data from different sources, specifically detector and probe vehicles, with the aim to improve the accuracy and reliability of the estimates.

Majority of the above research is limited to freeways, and cannot be applied to urban networks, where problem is rather more challenging due to number of reasons such as intersections, mid-link sources and sinks etc.

In this paper we present a methodology named, CUmulative plots and PRobe Integration for travel timE estimation (CUPRITE) (Bhaskar, 2009). The methodology is based on classical analytical procedure for travel time estimation using cumulative plots. Analytical modeling is performed to integrate cumulative plots with probe vehicle data for accurate estimation of average travel time.

CUPRITE can be applied for performance evaluation and Level Of Service (LOS) for different intersections. The performance of the system can act as a feeback to the signal controller to optimise its parameters. It can also be used for ITS application such as advance traveller information system and public transport priority systems.

2. Methodology

2.1 Classical procedure for average travel time estimation

The classical analytical procedure for travel time estimation is based on considering cumulative plots U(t) and D(t) at upstream and downstream locations, respectively (Daganzo, 1997). Refer to Figure 1, if the vehicles represented from time t_1 to t_2 in U(t) and t_3 to t_4 in D(t) are same then the area between the plots is the total travel time from upstream to downstream. Average travel time is the total travel time divided by the number of vehicles departing.

For travel time estimation the plots should be based on only those vehicles that traverse from upstream to downstream. But cumulative plots are defined based on the detector counts at a specific location. Due to detector counting error or loss or gain of vehicles between plots location, there is relative deviation (RD) amongst the plots (also termed as "drift"). The RD issue is critical in the application of classical procedure. For instance, in Figure 2 illustrates the application when upstream detector is overcounting, the error in the travel time estimation is represented as the shaded region, and if left unchecked can exponentially grow with time.

Example of loss or gain of vehicles includes parking or side-street between upstream and downstream location. Detectors are not always perfect and one can easily observe 5% error in detector counting.

Figure 1 Classical analytical procedure for average travel time estimation

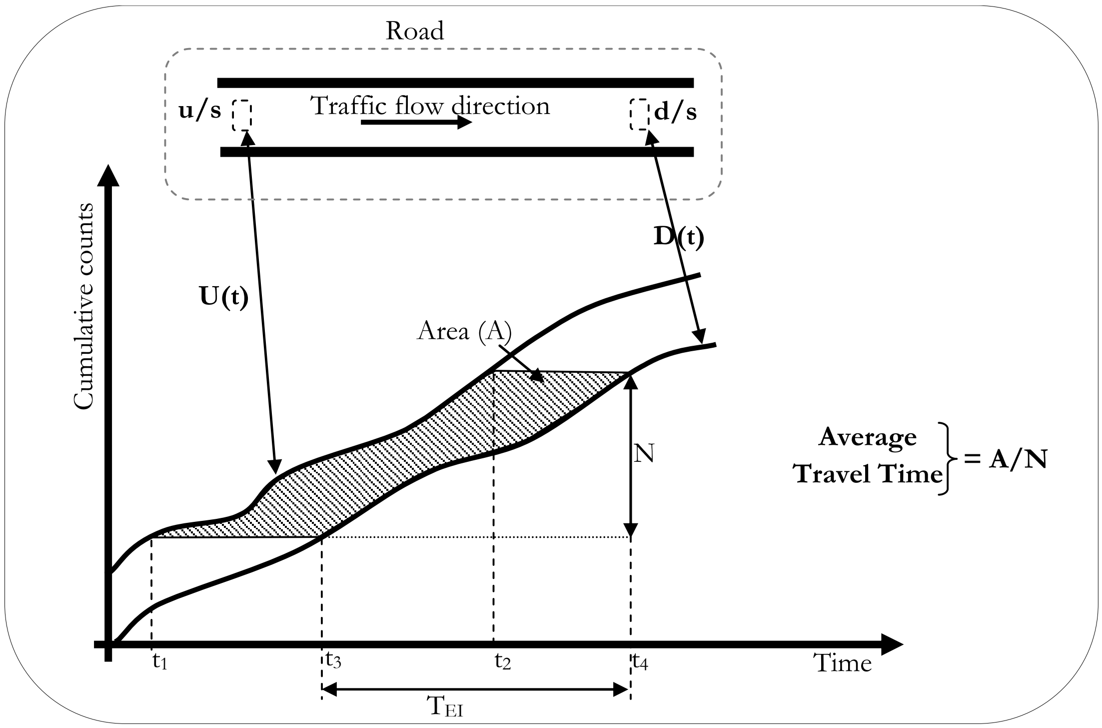
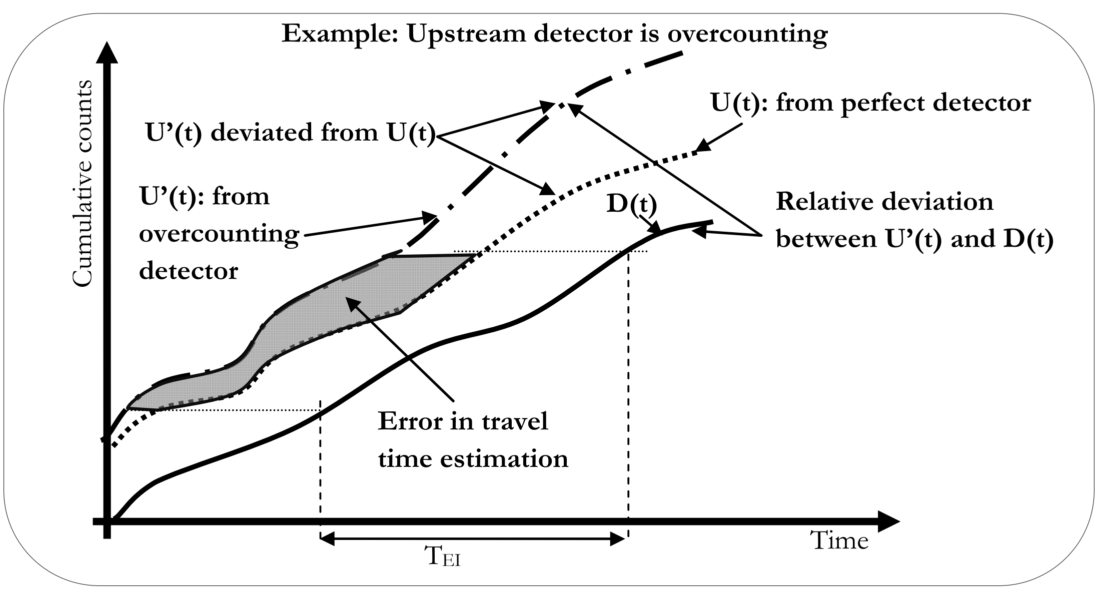


Figure 2 Vulnerability of classical analytical procedure to relative deviation (RD) amongst the plots.



2.2 Cumulative plots and probes

Here probe vehicles are vehicles equipped with vehicle tracking equipments. We assume that the time when the probe vehicle is at upstream (t_u) and downstream (t_d) locations is accurately obtained. The travel time of this vehicle is $t_d - t_u$.

We define the rank of the probe vehicle in the cumulative plots as $D(t_d)$ and define a parameter Δt (1), which is the difference in time when the probe is represented in U(t) (given that we fix its rank as $D(t_d)$) to the time when it is actually at upstream location.

$$\Delta t = U^{-1}(D(t_d)) - t_u \tag{1}$$

If all the vehicles in U(t) and D(t) are same then $\sum \Delta t$ from all the vehicles should be zero. This is an important property and is the explanation for the area between the plots is the total travel time.

Probe is random sample from the population of vehicles, and we make a hypothesis that we can reduce the RD by redefining U(t) such that $\sum \Delta t$ from all the probes is zero.

2.3 Virtual probe

Virtual probe is defined as a virtual vehicle that, during undersaturated traffic flow, departs from the downstream at the end of signal green phase and its travel time is free-flow travel time of the link. The probe is not real and is defined with the aim of reducing RD.

For undersaturated traffic conditions vehicle queue should vanish at the end of each signal green phase (t_{GE}). Travel time for the vehicle entering the intersection at time t_{GE} should be close to free-flow travel time (t_{ff}) of the link. Therefore, during undersaturated traffic conditions we can define virtual probe such that it is observed at upstream and downstream at time t_{GE} - t_{ff} and t_{GE}, respectively (i.e. for virtual probe $t_u = t_{GE} - t_{ff}$ and $t_d = t_{GE}$.). Note: virtual probe is only defined if the following *conditions for virtual probe* are satisfied:

2.3.1 Conditions for virtual probe

i. As the travel time of a virtual probe is defined as free-flow travel time of the link, therefore on the study link the sources for significant mid-link delay such as, mid-link intersections and on-street bus stop should be absent

ii. There should not be any leftover queue at the end of signal green phase.

iii. There should be presence of RD i.e., the following equation should be satisfied:

$$U^{-1}(D(t_{GE})) - t_{GE} \notin [t_{ff} - \delta, t_{ff} + \delta] \tag{2}$$

Where δ is a calibration parameter taking into account the variation in the estimation of t_{ff}. It can be considered equal to the standard deviation of the estimate of t_{ff}.

2.4 Architecture of CUPRITE

The summary of the algorithm (*see* Figure 3) is as follows:

Step 1 Cumulative plots are defined by integrating signal controller data with detector data (Refer to Bhaskar et al., (2008)).

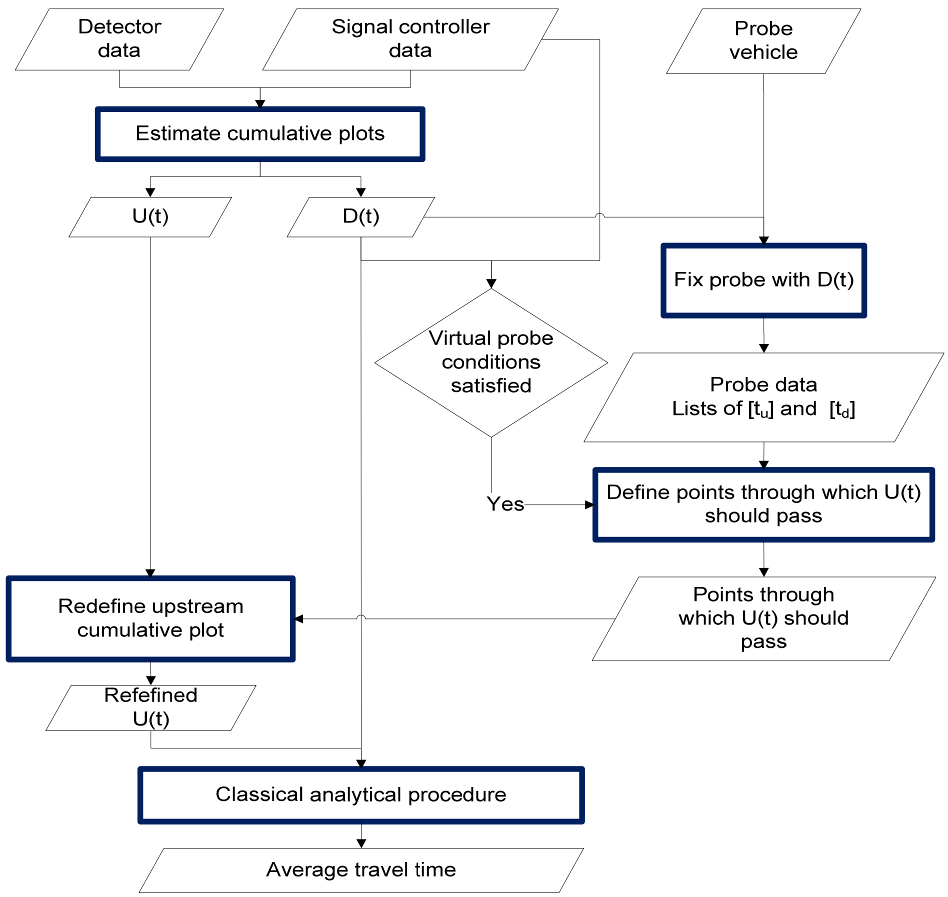
Step 2 Probe vehicle data (list of [t_u] and [t_d]) is defined. If the *conditions for virtual probe* are satisfied then the list [t_u] and [t_d] is appended with additional elements corresponding to the virtual probe i.e., $t_u = t_{GE} - t_{ff}$; $t_d = t_{GE}$, where t_{GE} is the time corresponding to the end of signal green interval.

Step 3 Points through which U(t) should pass are defined.

Step 4 U(t) is redefined by a) first vertical scaling and shifting the plots so that it passes through the above defined points (Step 3).

Step 5 Finally, for each estimation interval average travel time is estimated using classical analytical procedure.

Figure 3 Architecture for CUPRITE.



2.4.1 How to define the points from where U(t) should pass

Say, we have n probe vehicles and the database for the probe is defined as list of $[t_u]$ and list of $[t_d]$. These lists are appended with additional elements satisfying the *conditions for virtual probe*. If the conditions are satisfied, then t_{GE} is appended to the list $[t_d]$; and $(t_{GE} - t_{ff})$ is appended to the list $[t_u]$.

Following are the steps to be followed to define the points from where U(t) should pass:

Step 1 Sort list $[t_d]$ in ascending order of its values.

Step 2 Sort list $[t_u]$ in ascending order of its values.

Step 3 The required points through which U(t) should pass are $(t_{uj}, D(t_{dj}))$; where t_{uj} and t_{dj} are jth value in the sorted list of $[t_u]$ and $[t_d]$, respectively.

2.4.2 How to redefine U(t)

Say, we have: a) a reference point (t_{Ref}, U(t_{Ref})), i.e., the point in which we have confidence that it is a correct point on the plot; and b) point (t_p, Y_p) through which U(t) should pass. Then, refer to equations (3), (4) and (5); we redefine U(t) by applying correction on it such that all points on the plot:

i. Before time t_{Ref} have no correction;

ii. Between t_{Ref} to t_p are scaled vertically; and

iii. Beyond t_p are shifted vertically so that the redefined curve is parallel to U(t) and is continuous with the points before time t_p.

$$U(t) = U(t) + Correction \tag{3}$$

$$Correction = \begin{cases} 0 & \forall\ t \le t_{Ref} \\ (scale-1)*(U(t)-U(t_{\mathrm{Ref}})) & \forall\ t_{Ref} < t < t_p \\ (scale-1)*(U(t_p)-U(t_{\mathrm{Ref}})) & \forall\ t \ge t_p \end{cases} \tag{4}$$

$$scale = \begin{cases} \dfrac{Y_p - U(t_{\mathrm{Ref}})}{U(t_p) - U(t_{\mathrm{Ref}})} & if\, U(t_p) \ne U(t_{\mathrm{Ref}}) \\ 1 & if\, U(t_p) = U(t_{\mathrm{Ref}}) \end{cases} \tag{5}$$

Reference point

U(t) and D(t) are initially two independent cumulative plots. When the traffic condition is free-flow (for instance during night) then counts for cumulative plots can be initialized to zero. This is the initial reference point (P_0). Say [P_1, P_2, P_3, ..., P_n] is the list of n points from where U(t) should pass then for redefining U(t) for point P_i, the reference point is P_{i-1}.

2.4.3 Average travel time estimation

The classical procedure (see section 2.1) is applied between redefined U(t) and D(t) to estimate average travel time.

3. Validation

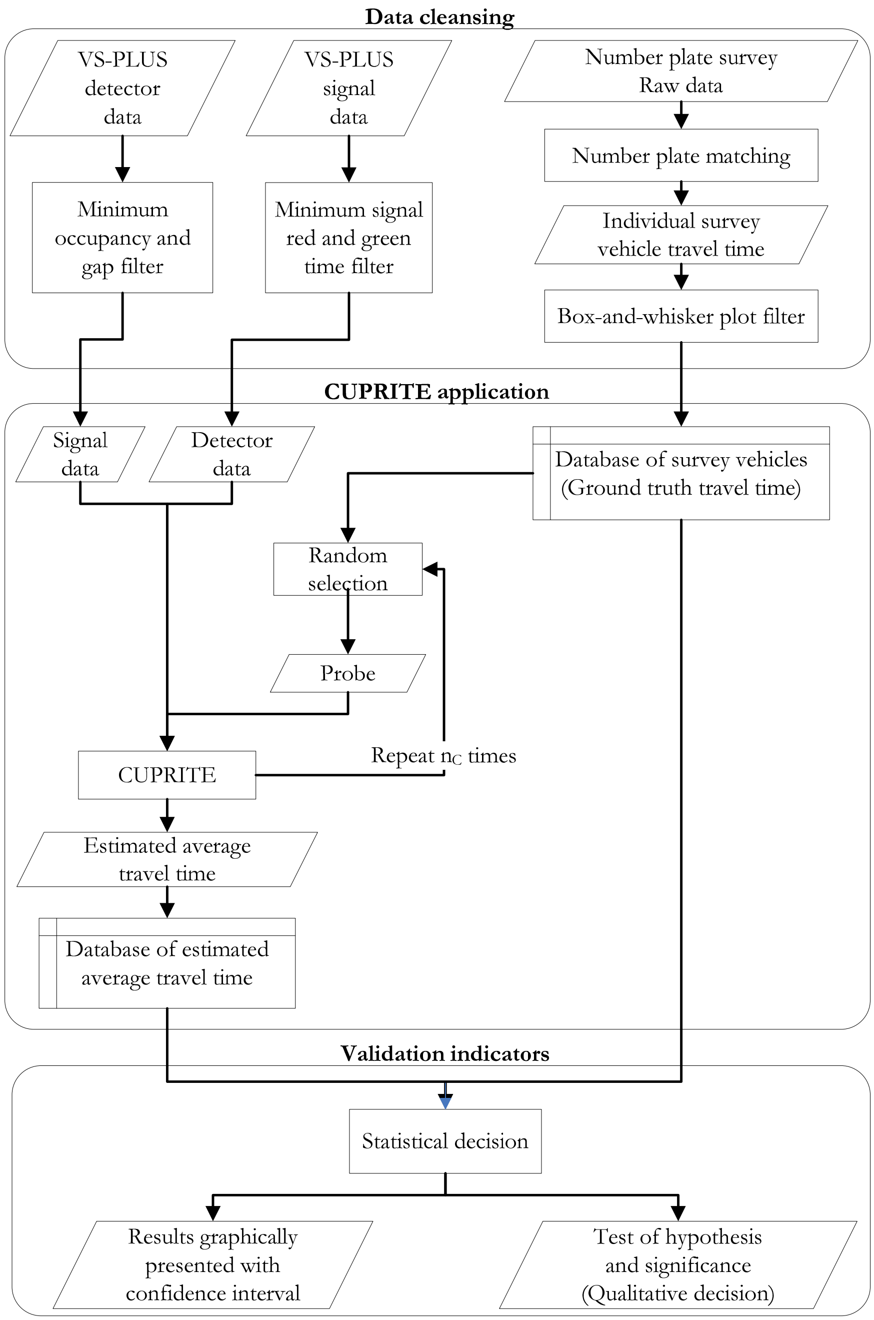
CUPRITE is validated on real data collected at Lucerne city, Switzerland. The signal control at the site is equipped with VS-PLUS signal controller (VS-PLUS). The signals are controlled centrally and the data from the controller is logged and stored by the Lucerne City Transport

Authority (StadtLuzern). The detector counts and signal timings for CUPRITE are obtained from VS-PLUS data.

Ground truth, individual vehicle travel time, is obtained from manual number plate (license plate) survey. It was performed on 15th April, 2008 (Tuesday, working day) from 3:00 p.m. to 6:00 p.m. The survey period captures both undersaturated and oversaturated traffic conditions. The required probe vehicles for CUPRITE were randomly selected from the survey data.

Figure 4 systematically illustrates the steps involved in the validation procedure. Prior to the application of the CUPRITE, both VS-PLUS data and number plate survey data need to be cleansed (Section 3.1). The cleaned data is the input to CUPRITE and it provides estimated average travel time (Section 3.2) which is finally, statistically validated with ground truth average travel time obtained through survey (Section 3.4).

Figure 4 Framework for CUPRITE validation.



3.1 Data cleansing

3.1.1 Number plate survey data

A manual number plate survey was performed and first four digits of the vehicle number plate and the corresponding time stamp when the vehicle enters the intersection were obtained. The number plate at upstream and downstream stations is matched and individual vehicle travel time is obtained. Due to human error or two vehicles having similar first four digits of the number plate or other reasons, there may be observed travel time much different from the neighbouring traversing vehicles. These deviant travel time values are considered as outliers and are not be considered for the validation procedure. Here, the *box-and-whisker plot technique* is employed to filter the outlier travel time values.

In the *box-and-whisker plot technique* a set of data is represented in: *a)* median; *b)* lower quartile (LQ) i.e., 25th percentile; and *c)* upper quartile (UQ) i.e., 75th percentile. The difference between the upper quartile and lower quartile is Inter Quartile Range (IQR) and it defines the scatter of the data. The Lower Bound Value (LBV) and Upper Bound Value (UBV) are:

$$LBV = LQ - 1.5 * IQR \tag{6}$$

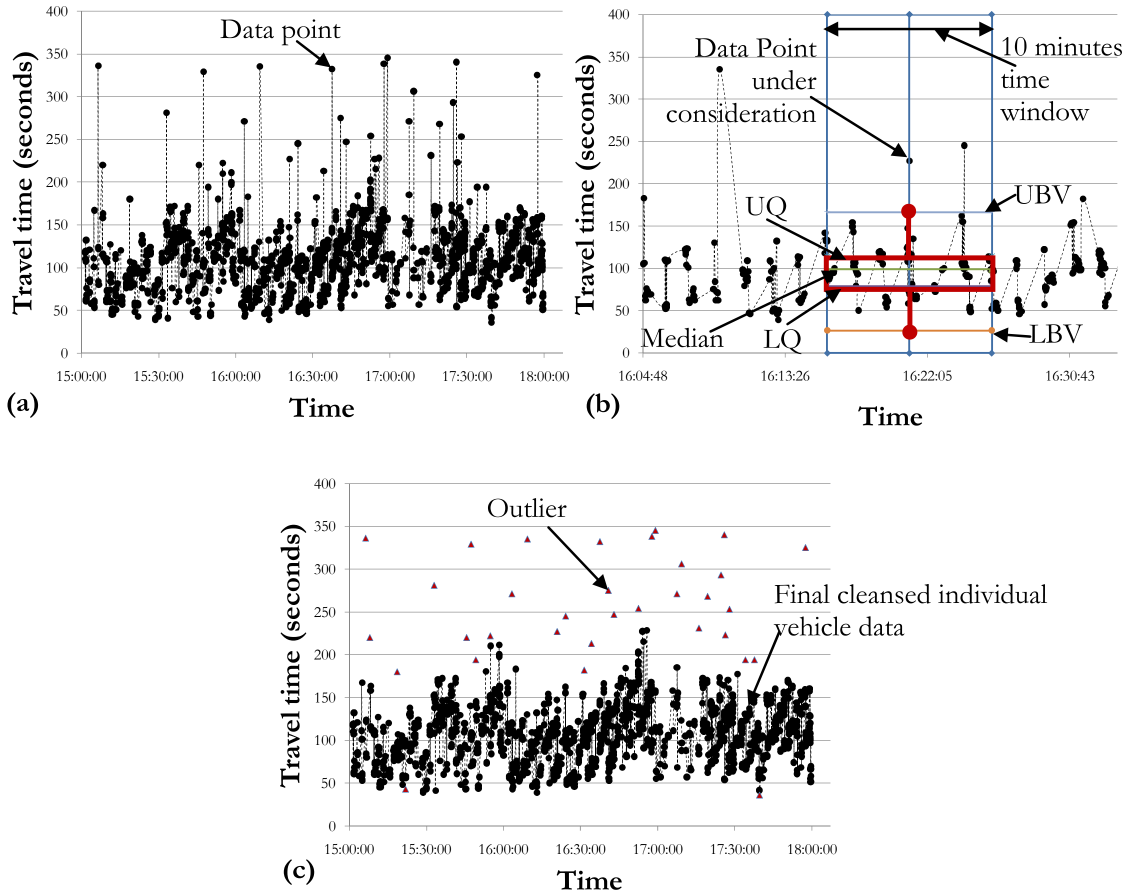
$$UBV = UQ + 1.5 * IQR \tag{7}$$

$$IQR = UQ - LQ \tag{8}$$

Any point lying below LBV or above UBV is regarded as an outlier and is disregarded.

Figure 5 represents an example. Figure 5*a* represents the raw date. To filter the outlier, a 10 min time window (5 min before and 5 min after) around the *data point under consideration* is defined. Box-and-whisker plot is obtained for all the data points within the time window. If the *data point under consideration* (*see* Figure 5*b*) is below LBV or above UBV then it is defined as outlier. The process is repeated for all the data points. Note: all the points (including those earlier defined as outliers) within the time window are considered for defining box-and-whisker plot. Figure 5*c* represents the final cleansed data with outliers removed.

Figure 5 Example of filtering the outlier using box-and-whisker plot.



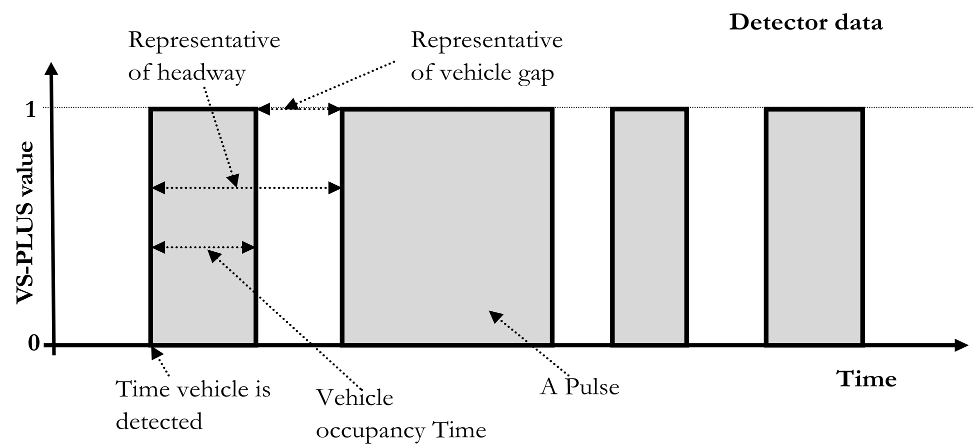
3.1.2 VS-PLUS data

VS-PLUS provides pulse data for each detector and signal phase, i.e., value '1' or '0' and corresponding time stamp. If we plot the values versus time, then a pulse can be defined as the portion of the graph represented by value of one (*see* Figure 6). Due to different reasons, sometimes there is noise in the pulses (unexpected fluctuations) which need to be filtered out. The noise can be due to pulse breakup.

Filter for VS-PLUS detector data

The values of '1' and '0' indicate the presence and non-presence of a vehicle on the detector, respectively. Therefore: a) the time length for a pulse represents the occupancy time (OT) of the vehicle on the detector; b) the time difference between the end of the leading pulse and start of the following pulse is represents of the gap (G) between the vehicles; and c) the time difference between the start of two consecutive pulses is the representative of the headway between vehicles (*see* Figure 6). Ideally, a pulse should correspond to a vehicle and hence the vehicle by vehicle count can be obtained. However, due to noise in the pulse there can be overcounting of vehicles. To avoid this we define minimum accepted occupancy time (OT_{min}) and minimum accepted gap (G_{min}). The filter is applied such that: a) if the gap between two consecutive pulses is less than G_{min} then both the pulses are merged, representing only one count for two pulses; and b) if the occupancy time is less than OT_{min} then pulse is disregarded. The value of OT_{min} and G_{min} used in the present analysis is 0.3 s, each.

Figure 6 Pulse data representation for VS-PLUS detector data.



The above filter of minimum occupancy and minimum gap can only remove noise in the pulse. This does not resolve the problem of detector counting error due to closely spaced vehicles, cross-talk etc. For instance, if the gap between vehicles is small and detector is not able to differentiate two consecutive vehicles then a long pulse, instead of two pulses is obtained. This results in undercounting. CUPRITE addresses this issue of detector counting error.

Filter for VS-PLUS signal data

The values of ‘0’ and ‘1’ indicate the start of display red light and display green light for the signal phase, respectively and hence the corresponding displayed signal red time and displayed signal green time. Ideally, a displayed green or red should be more than some minimum value but due to noise in the data there are periods where we have pulses close to each other. Analogous to the previous filter for VS-PLUS detector data, we consider the minimum red and green time to be 3 s and pulse or gaps less than 3 s are ignored.

3.2 CUPRITE application

As the survey vehicle data is available for a fixed time period and the probe data required for CUPRITE application is randomly selected from the survey vehicle data. Therefore, for each estimation interval CUPRITE is applied for n_C times (10) with different values of the seed for random number generator to randomly selecting probe vehicles. Hence, the application of CUPRITE provides different travel time estimates for a given estimation interval. Say for an estimation interval the mean and standard deviation of the estimates be $\overline{X}_C$ and S_C, respectively. Then we apply the sampling theory and confidence bounds for the travel time estimate by CUPRITE are defined by:

$$\overline{X}_C - t_{\alpha/2,n_C-1}\frac{S_C}{\sqrt{n_C}} \le \mu_C \le \overline{X}_C + t_{\alpha/2,n_C-1}\frac{S_C}{\sqrt{n_C}} \tag{9}$$

Where:

μ_C is the mean of the population of estimates from CUPRITE application;

$t_{\alpha/2,n_C-1}$ is the t-statistic at α level of significance and n_C-1 degrees of freedom;

n_C is defined as follows:

$$n_C = Min(\frac{n_s!}{N!(n_s-N)!}, 20); \qquad \text{assuming } n_s \ge N \tag{10}$$

n_s is number of survey vehicles in the estimation interval.

This means that, for an estimation interval, if N number of probe vehicles is required, then CUPRITE is applied by randomly selecting different combinations (without repetition of same combination) of N probe vehicles, or for 20 times, whichever is the minimum. For instance, say 2 (=N) probe vehicles in an estimation interval are required. If number of survey vehicles are 10, then there can be 45 different combinations of two probe vehicles. In this case, CUPRITE is applied 20 times by randomly selecting (without repetition) a combination each time. However, if there are 5 survey vehicles then only 10 combinations of two probe vehicles is possible. In this case, CUPRITE is applied 10 times and all the combinations are considered.

3.3 Ground truth travel time

The number plate survey captures the sample of vehicles traversing the link. We are interested in actual average travel time for all the vehicles departing the link during travel time estimation interval. Say the mean and standard deviation of the travel time obtained from the survey be $\overline{X}_s$ and S_s, respectively. We estimate the confidence bounds in the actual average travel time (μ_s) of the vehicles as:

$$\overline{X}_s - t_{\alpha/2, n_s - 1} \frac{S_s}{\sqrt{n_s}} \leq \mu_s \leq \overline{X}_s + t_{\alpha/2, n_s - 1} \frac{S_s}{\sqrt{n_s}} \tag{11}$$

Where: $t_{\alpha/2, n_s - 1}$ is the t-statistic with α level of significance and n_s-1 degrees of freedom; n_s is number of survey vehicles in an estimation interval.

3.4 Validation indicator

We present the results: graphically by overlapping the time series of travel time from survey and CUPRITE application; and qualitatively as statistical test of hypothesis and significance.

3.4.1 Graphical presentation of results

Figure 7 illustrates an example for the presentation of results. For each estimation interval, the black box represents the confidence bounds for the ground truth average travel time (*see* Figure 7*a*) and the orange box represents the confidence bounds for the travel time estimates from the CUPRITE (*see* Figure 7*b*).

Accuracy of the estimates from CUPRITE is defined as following:

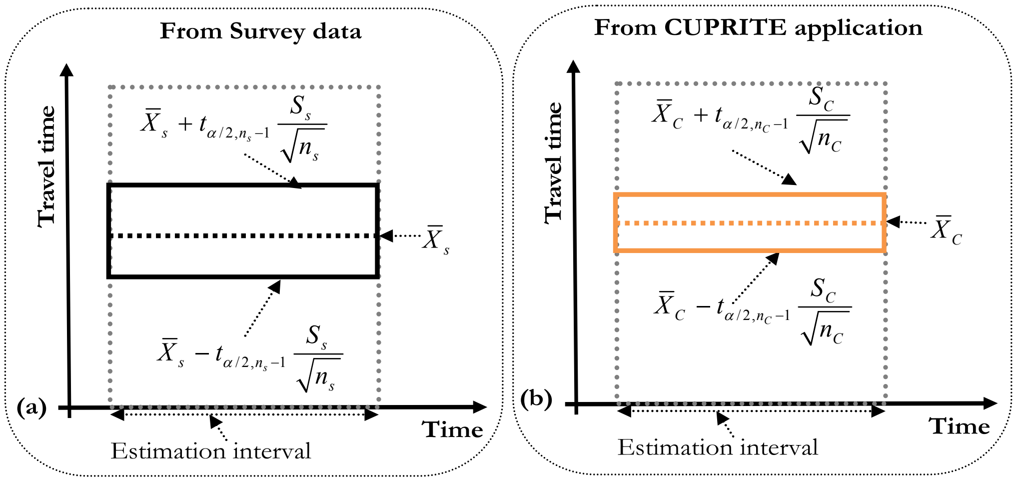
$$Error_i = \left(\frac{\left|\overline{X}_{s_i} - \overline{X}_{C_i}\right|}{\overline{X}_{s_i}}\right) \tag{12}$$

$$MAPE = \sum_{i=1 ton} \frac{Error_i}{n} \tag{13}$$

$$Accuracy(\%) = 1 - MAPE \tag{14}$$

Where: $Error_i$ is the absolute percentage error for i^{th} estimation interval; $\overline{X}_{s_i}$ and $\overline{X}_{C_i}$ are the mean of survey travel time and mean of travel time estimates from CUPRITE application during i^{th} estimation interval, respectively; n is the number of estimation intervals; and $MAPE$ is the Mean Absolute Percentage Error obtained from the CUPRITE application for different estimation intervals during survey period.

Figure 7 Systematic representation of the results for CUPRITE validation.



3.4.2 Statistical test

We perform statistical test so as to make qualitative decisions about the CUPRITE validation. The intension is to determine whether there is enough evidence to "reject" a (null) hypothesis about the CUPRITE validation. Here, two different processes: a) number plate survey and b) CUPRITE application; provide dataset for average travel time. We are interested to know if these two processes provide statistically similar results, i.e., the mean of the two processes are the same.

We make a null hypothesis H_0 (15): that the true mean of the first process (μ_s) is equal to the true mean of the second process (μ_C). Or in other words the two sets of data (number plate and CUPRITE) with sample means $\overline{X}_s$ and $\overline{X}_C$, respectively are both part of the same population so that their population means are equal. Null hypothesis is tested against the alternate hypothesis (H_a) that the two means are not equal (16).

$$\textit{Null Hypothesis } (H_0): \mu_s = \mu_C \tag{15}$$

$$\textit{Alternative Hypothesis } (H_a): \mu_s \neq \mu_C \tag{16}$$

If we "do not reject" the null hypothesis (H_0), then we are saying that despite the fact that the travel time estimates come from two different processes there is not enough evidence to say that they are not part of the same overall population.

The statistical test to make the above decision is *t-test to compare two sample means (two-tailed t-test)*. We form the test statistics assuming that the true standard deviations for the two processes are not equivalent (NIST).

The degree of freedom (*df*) is estimated using the Welch-Satterthwaite approximation (17).

$$df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{(\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)})} \tag{17}$$

$$t_{test\ statistics} = \frac{X_1 - X_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{18}$$

Where: X_i, s_i and n_i is the mean, standard deviation and number of observations, respectively for the two processes. $X_1 = \overline{X}_s$; $X_2 = \overline{X}_C$; $s_1 = S_s$ and $s_2 = S_C$; $n_1 = n_s$ (number of survey vehicles during the estimation interval); $n_2 = n_C$ (10).

For α level of significance we reject the null hypothesis Ho, if:

$$\left| t_{test\ statistics} \right| \geq t_{\alpha/2, df} \tag{19}$$

Else we do not reject the null hypothesis and reject the alternate hypothesis.

Where: $t_{\alpha/2,df}$ is the upper critical value of the Student's-t distribution at α level of significance with *df* degree of freedom.

"Do not reject H_0" indicates there is not enough evidence to reject the assumption that: *CUPRITE estimates are statistically equivalent to the real travel time from the number plate survey*.

Note: Statistically, both the indicators defined in the previous subsections are connected. If the confidence bounds of the CUPRITE application (defined in Section 3.4.1) contain the mean of the survey travel time then we do not reject the null hypothesis (defined in Section 3.4.2).

3.5 Site Description

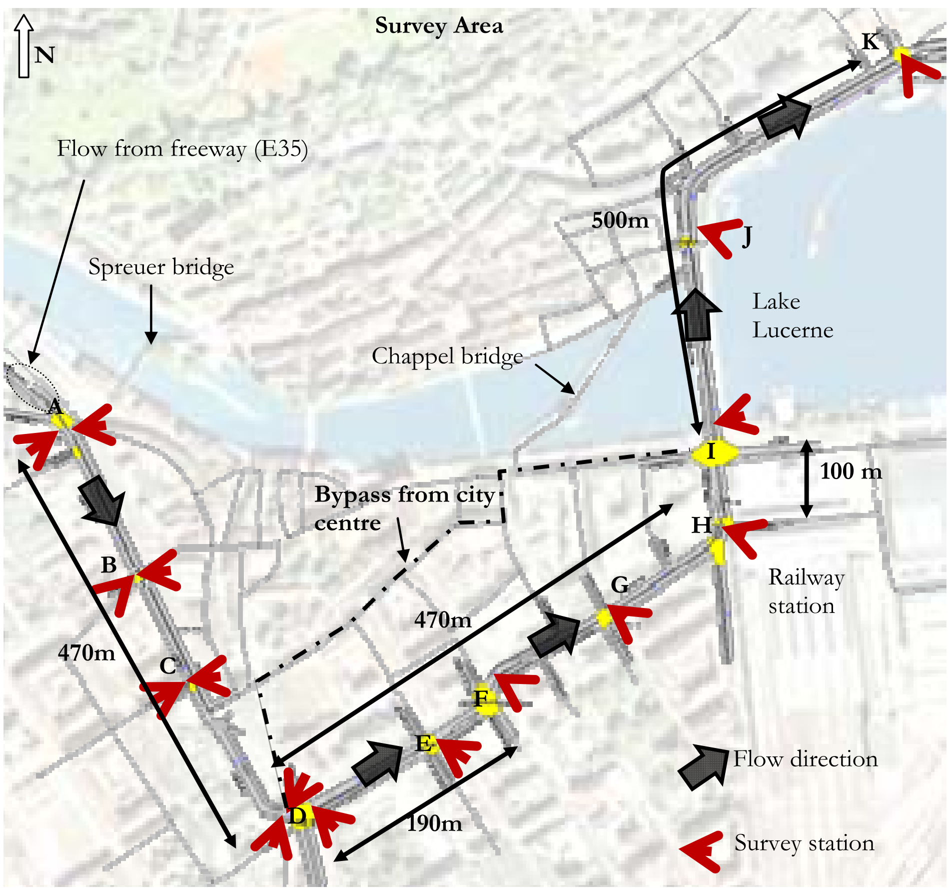
The data is collected on eleven consecutive signalized intersections (intersections *A* to *K*) as shown in the Figure 8. It consists of three legs:

i. Intersection *A* to intersection *D* in which the flow is from a freeway (Freeway number E35) with minor mid-link sinks and sources;

ii. Intersection *D* to intersection *I*, which passes through the city centre and the bottleneck mainly at intersection *F* and intersection *I*. This leg also carries traffic to the railway station; and

iii. Intersection *I* to intersection *K*, where there is no mid-link sink or source, but significant amount of mid-link delay due to pedestrians. Link from intersection *I* to intersection *K* is along the lake side with significant number of tourists.

Figure 8 Number plate survey site.



3.6 Validation results

Here, travel time estimation interval is for five signal cycles. As signals are adaptive therefore the cycle time is not fixed. Fixed number of probes per estimation interval (S_n) is considered. Two tailed t-tests were considered significant at (α=) 0.05.

Here, the results of time series of travel time and statistical decision from t-tests are presented in the same figure (*see* Figure 9). For each estimation interval: a) Orange and black boxes are as defined in Section 3.4.1; b) Green circle represents, "not enough evidence to reject H_0"; and c) Red triangle represents "Reject H_o".

For $A \rightarrow D_{Lft}$ the results obtained for one, two and three probes per estimation interval are illustrated in Figure 9*a*, Figure 9*b* and Figure 9*c*, respectively. Similary, the results for $A \rightarrow D_{Thru}$ are illustrated in Figure 9*c*, Figure 9*d* and Figure 9*e*. In most of the estimation intervals, the null hypothesis cannot be rejected. Indicating that our initial assumption (Mean estimated from CUPRITE is statistically equivalent to that of number plate survey.) is not rejected at 0.05 level of significance. The orange box overlaps with black box, indicating that the CUPRITE can estimate the true actual travel time. It can be seen that even the short term oversaturation in the system can be accurately estimated. For instance, in Figure 9: fourth, fifth, sixth and seventh estimation intervals (time from 15:30 hr to 16:00 hr) are congestion build up, and there is significant variation in average travel time between the three periods. This fluctuation is also captured accurately by CUPRITE.

For $A \rightarrow D_{Lft}$ the accuracy (14) of the CUPRITE model increases from 92.3% to 94.6% with increase in number of probes from one probe per estimation, respectively. Similar results are obtained for other routes and are illustrated in Figure 10. The route from $A \rightarrow I$ and $A \rightarrow K$ is not considered. This is because there is a bypass from intersection *D* to intersection *I* (*see* Figure 8), which is used by drivers to avoid the congestion through the city centre. The vehicle observed at both *A* and *I* or *A* and *K* can be the one traversing through the bypass.

The validation of CUPRITE has demonstrated that it can be successfully applied for accurate and reliable travel time estimation on urban networks

Figure 9 Results for A→D_{Lft} and A→D_{Thru}.

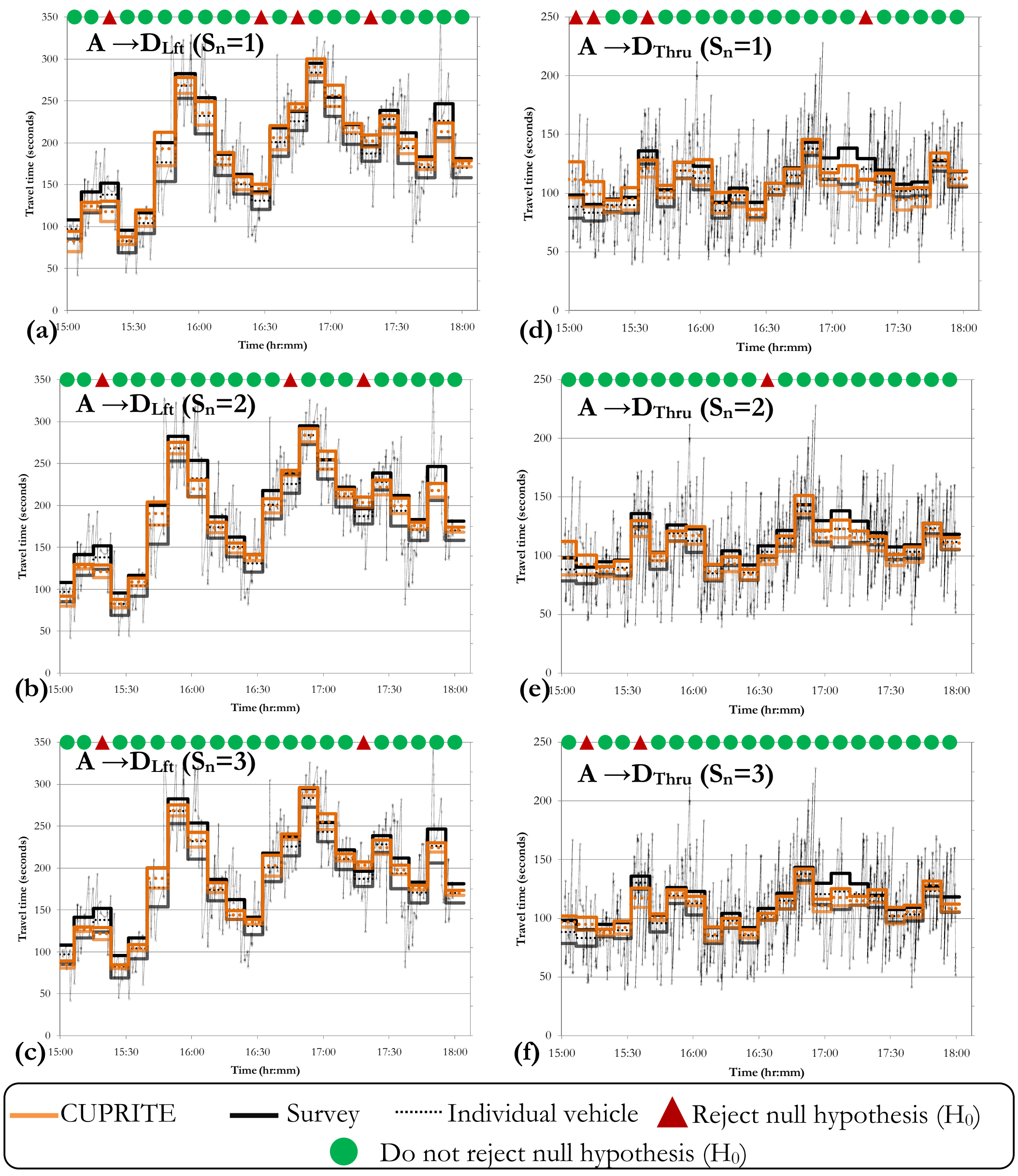
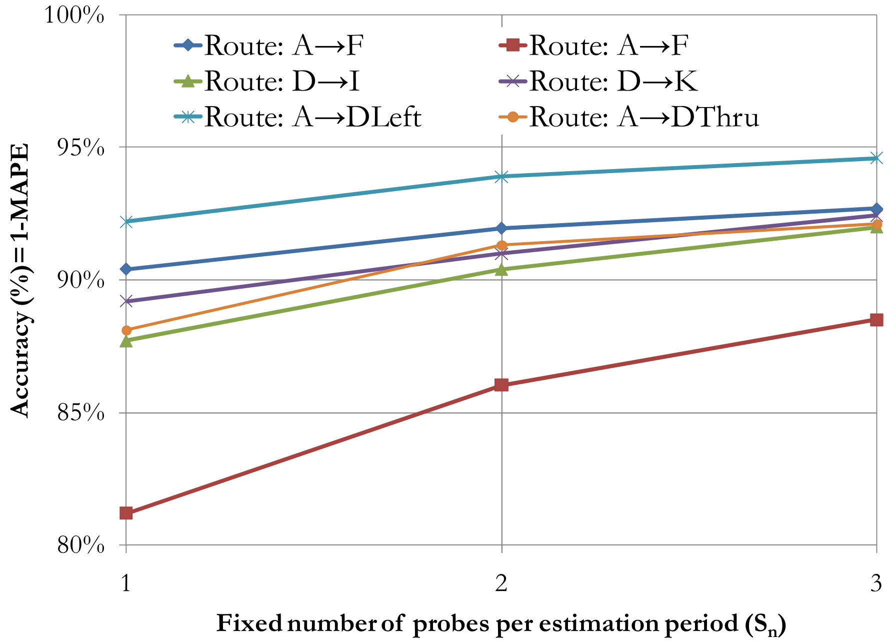


Figure 10 Comparative results for different route.



4. Conclusion

In this paper, a methodology CUPRITE is presented and validated. The real data for the validation is obtained through number plate survey at a site in Lucerne city, Switzerland. The study site is a typical urban network with following characteristics: *a)* mixed traffic (with buses); *b)* on-street bus stops; *c)* significant loss and gain from mid-link sinks and sources, respectively; and *d)* significant mid-link delays due to pedestrian crossing. The loop detectors on the site are not perfect.

CUPRITE is applied at the above site and two tailed t-test (at 0.05 level of significance) results confirm that the travel time estimates from CUPRITE are statistically equivalent to real estimates from number plate survey data. Validation results also indicate that CUPRITE can accurately capture the time series of travel time and short-term oversaturation in the system.

It can be concluded that CUPRITE required only few probes per estimation period for accurate travel time estimation. The current market penetration of probe is low and the requirement of few probes makes the methodology directly applicable.

5. References

BAJWA, S. I., CHUNG, E. & KUWAHARA, M. (2003) A travel time prediction method based on pattern matching technique *21st ARRB and 11th REAAA Conference.* Cairns, Australia,.

BERKA, S., TARKO, A., ROUPHAIL, N. M., SISIOPIKU, V. P. & LEE, D.-H. (1995) Data fusion algorithm for ADVANCE Release 2.0. *Advance Working Paper Series, No. 48, University of Illinois at Chicago, Chicago, IL.*

BHASKAR, A. (2009) A methodology (CUPRITE) for urban network travel time estimation by integrating multisource data. *School of Architecture, Civil and Environmental Engineering.* Lausanne, The Ecole Polytechnique Fédérale de Lausanne (EPFL) http://library.epfl.ch/theses/?nr=4416.

BHASKAR, A., CHUNG, E., DE OLIVIER, M. & DUMONT, A. G. (2008) Analytical Modelling and Sensitivity Analysis for Travel Time Estimation on Signalised Urban Network. *Transportation Research Board 87th Annual Meeting.* Washington, D.C.

CHEN, H., GRANT-MULLER, S., MUSSONE, L. & MONTGOMERY, F. (2001) A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. *Neural Computing and Applications,* 10**,** 277-286.

CHOI, K. & CHUNG, Y. (2002) A data fusion algorithm for estimating link travel time. *ITS Journal: Intelligent Transportation Systems Journal,* 7**,** 235-260.

COIFMAN, B. & KRISHNAMURTHY, S. (2007) Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure. *Transportation Research Part C: Emerging Technologies,* 15**,** 135-153.

DAGANZO, C. F. (1997) *Fundamentals of Transportation and Traffic Operations*, Pergamon, Oxford.

DAILEY, D. J. (1993) Travel-time estimation using cross-correlation techniques. *Transportation Research,* 27 B**,** 97-107.

EL FAOUZI, N. E. (2004) Data-driven aggregative schemes for multisource estimation Fusion: A road travel time application. IN DASARATHY, B. V. (Ed.) *Proceedings of SPIE - The International Society for Optical Engineering.* Orlando, FL.

HELLINGA, B. R. & FU, L. (2002) Reducing bias in probe-based arterial link travel time estimates. *Transportation Research Part C: Emerging Technologies,* 10**,** 257-273.

LIU, H., VAN ZUYLEN, H., VAN LINT, H. & SALOMONS, M. (2006) Predicting urban arterial travel time with state-space neural networks and Kalman filters. *Transportation Research Record.*

NAM, D. H. & DREW, D. R. (1999) Automatic measurement of traffic variables for intelligent transportation systems applications. *Transportation Research Part B: Methodological,* 33**,** 437-457.

NIST NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 13th February 2009.

OH, J. S., JAYAKRISHNAN, R. & RECKER, W. (2003) Section travel time estimation from point detection data. *In proceedings of the 82th Annual Meeting of Transportation Research Board.* Washington, D.C., U.S.A.

PARK, D. & RILETT, L. R. (1998) Forecasting multiple-period freeway link travel times using modular neural networks. *Transportation Research Record.*

ROBINSON, S. & POLAK, J. W. (2005) Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method. *Transportation Research Record,* 1935**,** 47-56.

SISIOPIKU, V. P. & ROUPHAIL, N. M. (1994) Toward the use of detector output for arterial link travel time estimation: a literature review. *Transportation Research Record***,** 158-165.

SISIOPIKU, V. P., ROUPHAIL, N. M. & SANTIAGO, A. (1994) Analysis of correlation between arterial travel time and detector data from simulation and field studies. *Transportation Research Record***,** 166-173.

SRINIVASAN, K. K. & JOVANIS, P. P. (1996) Determination of number of probe vehicles required for reliable travel time measurement in urban network. *Transportation Research Record***,** 15-22.

STADTLUZERN (2008) http://www.stadtluzern.ch/.

VS-PLUS (2008) http://www.vs-plus.com/e/vsintro.htm.

WARDROP, J. G. (1968) Journey speed and flow in central urban areas. *Traffic Engineering and Control,* 9**,** 528-532.

WESTERMAN, M., LITJENS, R. & LINNARTZ, J.-P. (1996) Integration Of Probe Vehicle And Induction Loop Data: Estimation Of Travel Times And Automatic Incident Detection. *California Partners for Advanced Transit and Highways (PATH). Research Reports: Paper UCB-ITS-PRR-96-13. .*

YOU, J. & KIM, T. J. (2000) Development and evaluation of a hybrid travel time forecasting model. *Transportation Research Part C: Emerging Technologies,* 8**,** 231-256.

YOUNG, C. P. (1988) A relationship between vehicle detector occupancy and delay at signal-controlled junctions. *Traffic Engineering and Control,* 29**,** 131-134.