Fusing loop detector and probe vehicle data to estimate travel time statistics on signalized urban networks

Ashish Bhaskar*
Laboratory of Traffic Facility
Swiss Federal Institute of Technology
Lausanne, Switzerland
Ph: +41 21 693 2341;
Fax: +41 21 693 63 49
ashish.bhaskar@epfl.ch

Prof. Edward Chung
School of Urban Development
Queensland University of Technology
Brisbane, Australia
Ph: +61 731381143
Fax: +61 3138 1827
edward.chung@qut.edu.au

Prof. André-Gilles Dumont
Laboratory of Traffic Facility
Swiss Federal Institute of Technology
Lausanne, Switzerland
Ph: +41 21 693 2345
Fax: +41 21 693 63 49
andre-gilles.dumont@epfl.ch

First submission: 22nd June 2009
Final submission: 29th April 2010

*Corresponding author

Abstract
This paper presents a methodology that integrates cumulative plots with probe vehicle data for estimation of travel time statistics (average, quartile) on urban networks. The integration reduces relative deviation amongst the cumulative plots so that the classical analytical procedure of defining the area between the plots as the total travel time can be applied. For quartile estimation, a slicing technique is proposed. The methodology is validated with real data from Lucerne, Switzerland and it is concluded that the travel time estimates from the proposed methodology are statistically equivalent to the observed values.
1. Introduction

Travel time is defined as the time needed to travel from a point upstream (u/s) to a point downstream (d/s) on the network. It quantifies congestion and is easily perceived by all road users and operators. It is an important network performance measure and a decision making variable. For instance, travel time information is utilized by the operators to develop traffic control strategies for reducing congestion on both spatial and temporal scales.

Literature is abundant with models for obtaining travel time values. Majority of the literature is limited to freeways (Nam and Drew, 1996; Dharia and Adeli, 2003; Jintanakul et al., 2009) and cannot be applied on urban networks. The development of a travel time model on urban networks is more challenging than freeways due to number of reasons. For instance, interruption in traffic flow due to traffic signals; non conservation of traffic flow on urban link due to mid-link sources and sinks (e.g. parking, side street) etc. Travel time models can be differentiated into estimation models and prediction models. Estimation models provide experienced travel time whereas, prediction models (Park et al., 1999; You and Kim, 2000; Zhang and Rice, 2003; Vlahogianni et al., 2004; Vlahogianni et al., 2008; Hamad et al., 2009) provide expected (forecasted) travel time in future. The future can be immediate (i.e., for trips just starting) to several minutes (say 30 minutes) ahead.

Traffic flow on an urban network is in stop-and-go running condition i.e., vehicles have to stop at the intersection during signal red phase and queue of vehicles is formed. The individual vehicle travel time on urban link depends on number of factors. For instance, traffic demand; signal parameters; vehicles’ entry time on the link relative to downstream intersection signal red phase; and number of vehicles queued in front of it when it reaches the downstream signal etc. The distribution of travel time from different vehicles departing within an estimation interval (Here, referred as within interval distribution.) is bi-modal (or even multi-modal) with modes corresponding to the vehicles that can pass the link without stopping at intersection and for vehicles that experience delay at intersection.

Average travel time is an important indicator for network performance measure and generally most of the in-practice models are applicable for average travel time estimation (Review provided in Section 2). Due to the above mentioned bi-modal distribution; none of the vehicle can encounter average travel time. Hence, for better understanding of the network performance it is important to estimate other statistics, such as upper quartile of travel time in addition to the average travel time. Little research is performed on the estimation of within interval distribution mainly due to unavailability of individual vehicle travel time data. Robinson (2005) has applied k-NN approach, on the Automatic Vehicle Identification (AVI) data from central London, to estimate travel time variance for within interval distribution. He has observed around 23% variance in travel time between vehicles traversing the link during 15 minutes estimation interval. This is more than three times of the variance observed by Part et al., (1999) on Houston freeways.

The objective of this paper is to develop and validate a methodology for estimation of travel time statistics (average and quartiles) on signalized urban networks. The estimates are for certain time period that is integral multiple of signal control cycle. For instance, average travel time during five
signal control cycles. The develop methodology should be robust with respect to urban complexities, such as mid-link sources and sinks and detector counting error. The proposed methodology is 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 through integrating or fusing cumulative plots with probe vehicle data for accurate estimation of travel time statistics (average travel time and quartile of travel time). The proposed methodology can be applied for real time application, where the estimated travel time is the experienced travel in the last estimation interval.

The paper is organized as follows: section 2 provides the literature review for the average travel time estimation on urban networks. The classical analytical procedure for travel time estimation and its vulnerability for application on urban environment are introduced in section 3. Thereafter, the proposed methodology is developed in section 4, followed by its validation with real data in section 5. Finally, the conclusions are presented in section 6.

2. Literature review

Advancement of technology has resulted in different traffic data retrieval systems from traditional loop detectors to advanced electronic systems onboard a vehicle, such as Vehicle Information and Communications Systems (VICS). Fixed sensors, such as loop detectors provide traffic flow and occupancy at the specific location on the network whereas; mobile sensors, such as probe vehicle provide data for the entire journey of the vehicle. Based on the type of data available, different models are proposed to estimate average travel time for all the vehicles traversing the road. Moreover, the availability of different data systems provide avenue for application of data fusion techniques for more reliable and robust travel time estimation. Thereafter, here we classify the literature into: i) fixed sensor; ii) mobile sensor; and iii) data fusion based models for average travel time estimation.

**Fixed sensor based:**

The initial motivation for the development of travel time estimation models was to consider effect of congestion in conventional traffic assignment step used in four-step transportation modeling. Several travel time functions (or volume delay functions) are proposed that define relationship between link travel time and traffic intensity (flow/capacity ratio). These include: Bureau of Public Roads (BPR, 1964); Davidson’s function (Akçelik, 1978; Tisato, 1991); conical-volume delay function (Spiess, 1990) etc. Webster delay model (Webster and Cobbe, 1966) is a pioneer model for estimating average deterministic delay at undersaturated signalized intersection. Researchers have followed Webster’s work to suit different field conditions and modified models are proposed, such as Akçelik (Akçelik, 1988; Akçelik, 1991; Akcelik and Routhail, 1993; Akcelik and Routhail, 1994) and Highway Capacity Manual’s procedure for delay estimation (TRB, 1998; TRB, 2000).

The simplicity of these travel time function make them favorable candidate for transport planning and policy analysis. They are not suited for ITS applications where more accurate and reliable analysis especially for variable traffic conditions in real time is required.
Regression analysis based models to estimate link travel time, as a function of site characteristics and detector data are also proposed. Wardrop (1968) has defined regression relationship between average journey speed in central urban area as a function of average traffic flow, width of the road, the number of controlled intersections per miles and average proportion of green time. Researchers (Gault, 1981; Young, 1988; Sisiopiku and Rouphail, 1994; Sisiopiku et al., 1994) have observed a regression relationship between average travel time and certain ranges of detector occupancy, for mid-link detectors, with queue that does not persist over detector location. Zhang (1999) has proposed a regression equation for journey speed as a function of volume to capacity ratio and mid-link detector flow and occupancy.

Generally, the regression models are site specific and require calibration to suite different environment. These models should not be generalized and the effect of parameters, such as detector location, effective green time, progression quality, link length, opposing flow from permissive phasing, traffic composition etc. should not be overlooked.

Researchers have also applied machine learning algorithm, such as k-Nearest Neighbor (Robinson and Polak, 2005) and Artificial Neural Networks (Palacharla and Nelson, 1999; Liu et al., 2005), for travel time estimation and prediction (Hamed et al., 1995). Such models require measured link travel time values and the corresponding detector data for a training period. The training data set should properly represent the required extend of the solution space and the model should be applied well within the limits for which it is trained.

Skabardonis and Geroliminis (2005) have proposed a model for travel time estimation based on upstream detector data and signal parameters. The queueing at the signal is considered by applying kinematic wave theory. The required detector should be sufficient upstream from the intersection stopline, so that the flow and occupancy measures from the detector are not affected by the presence of queue at the signals. The model involves calibrating the fundamental diagram (flow-density relationship) for links using detector data.

Recent advancement in sensor technology has produced Advance Inductive Loop Detectors (AILD) that can provide magnetic vehicle signatures. The signatures from upstream and downstream detectors can be correlated to reidentify the vehicles (or platoon) at downstream location for travel time estimation. Ritchie et al., (2002; 2005) have demonstrated the potential application of vehicle signature for travel time estimation on urban arterial. However, this approach is still in initial research stage, and further study is needed to increase the accuracy, reliability and reidentification rate. Moreover, for the implementation of reidentification technique, existing infrastructure should be upgraded with AILD and a high bandwidth in the data communication channel.

**Mobile sensor based:**

Mobile sensors, such as probe vehicle 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 a small fraction of all the vehicles traversing the link are probe vehicles. Average travel time for all the vehicles traversing the link can be estimated by applying
statistical sampling techniques on the travel time obtained from the probe vehicles (Hellinga and Fu, 2002; Long Cheu et al., 2002). Researchers (Srinivasan and Jovanis, 1996; Long Cheu, Xie and Lee, 2002) have shown interest to determine minimum number of probes required for statistically significant travel time estimation.

**Data fusion based:**

Researchers have also applied data fusion techniques to fuse data from different sources, specifically detector and probe vehicles data for travel time estimation. El Faouzi (2004) provides an overview of the application of data fusion techniques in road traffic engineering. Dailey et al., (1996) summaries ITS data fusion projects.

Berka et al. (1995) has applied weighted average based fusion technique where the fused average travel time is the weighted average of the estimated average travel time from detectors and estimated average travel time from probe vehicles. The weights are defined by considering variables, such as the standard deviation of the travel time from detector data and from probe data, respectively; weights assigned to detector travel time in data screening; the sum of weights of reasonable probe reports etc. A similar weighted average based data fusion approach for travel time estimation is proposed by El Faouzi (2004).

Choi and Chung (2002) have applied the data fusion technique for 5 minutes average travel time estimates using detector and probe vehicle data. The algorithm first estimates space-mean speed from detector counts and occupancy using Dailey (1999) equation, which provides travel time estimates for each minute. Each minute travel time estimates are aggregated using Voting Technique for 5 minutes average travel time ($TT_d$). Average 5 minutes travel time ($TT_g$) from GPS probes is obtained using Fuzzy regression. Finally, fused link travel time is obtained by applying Bayesian Pooling Method on $TT_d$ and $TT_g$. The algorithm is tested for undersaturated traffic condition and should be tested for oversaturated traffic condition too. They quote that “a different level of service might produce totally different weights of each data collection mechanism. In such cases, a different data fusion method and/or a revision of the proposed algorithm may be needed”.

Xie et al., (2004) have applied two independent neural network methodologies: Multi-Layer Perception (MLP) and Multi-Layer regression (MLR) models to fuse average travel time estimates from detector data and probe vehicles. The average travel time from detector is the sum of the free flow travel time and signal delay. The signal delay is estimated using Singapore model (Xie et al., 2001). Average travel time from probe samples are considered only if the sample size during estimation interval is more than 10 vehicles or is more than the minimum required sample size determined by central limit theorem.
3. The classical analytical procedure for travel time estimation

3.1 The procedure

Cumulative plot is a plot of cumulative count of vehicles versus time at a specific location on the network. The classical analytical procedure for travel time estimation considers cumulative plots $U(t)$ and $D(t)$ at upstream (u/s) and downstream (d/s) locations, respectively (Daganzo, 1997). It defines the total travel time from u/s to d/s as the area (Refer to Figure 1a) between the plots. Say, time $t_1$ and time $t_2$ correspond to the start and end of the $U(t)$ represented in the area, respectively. Similarly $t_3$ and $t_4$ are time corresponding to the start and end of $D(t)$ represented in the area, respectively. Then, mathematically, the average travel time $TT$ is represented as follows:

$$TT = \frac{\sum_{i=1}^{N} [D^{-1}(i) - U^{-1}(i)]}{N} = \frac{\sum_{i=1}^{N} D^{-1}(i) - \sum_{i=1}^{N} U^{-1}(i)}{N}$$

(1)

$$N = U(t_2) - U(t_1) = D(t_4) - D(t_3)$$

(2)

Here $N$ is the number of vehicles that depart downstream (arrives upstream) during the time interval from $t_3$ to $t_4$ ($t_1$ to $t_2$).

3.2 The Relative Deviation (RD) issue with the procedure

The area between the plots is the total travel time from upstream to downstream as long as all the vehicles represented in $U(t)$, from time $t_1$ to $t_2$, and in $D(t)$, from time $t_3$ to $t_4$, are same. Therefore, the plots should be based on only those vehicles that traverse from upstream to downstream.

Cumulative plots are defined based on the detector counts at a specific location. Practically, detectors are not perfect and one can easily observe 5% error in detector counting. Moreover, urban network has mid-link sources and sinks, such as parking or side-street. This results in non conservation of vehicles (loss or gain of vehicles) between upstream and downstream locations. Due to detector counting error; non conservation of vehicles between plots location; and any such combinations over time, there is relative deviation (RD) amongst the plots (also termed as “drift”).

Let us explain RD with the help of an example. Consider a scenario where upstream detector is overcounting. In Figure 1b: $U(t)$ is the cumulative plot observed from the overcounting upstream detector; $U'(t)$ is from a perfect detector. $U(t)$ deviates from $U'(t)$, or there is a relative deviation between $U(t)$ and $D(t)$. The observed cumulative plots are $U(t)$ and $D(t)$ and if the classical procedure is applied then the error in the estimation of travel time, during $T_{EI}$ travel time estimation interval, is represented as the shaded region in the figure. If RD is left unchecked then the error can exponentially grow with time. Hence, the RD issue is critical in the application of the classical procedure.

Note: $U(t)$ and $D(t)$ will eventually “diverge” from each other if: upstream detector is overcounting; or downstream detector is undercounting; or there is mid-link sink. $U(t)$ and $D(t)$ will eventually “cut” each
other if: upstream detector is undercounting; or downstream detector is overcounting; or there is midlink source. If the plots “diverge” then the travel time is highly overestimated and if the plots “cut” then travel time estimates are negative. In practice, there is complex combination of detector errors, midlink sources and midlink sinks over time, which defines the relative deviation for each estimation interval.

Figure 1: Classical analytical procedure and its vulnerability to relative deviation (RD) amongst the plots.
3.3 How RD issue can be addressed?

Refer to Figure 1a: The vertical distance (counts separation) between the plots (at time $t$) is the number of vehicles ($n$) between upstream and downstream locations. The horizontal distance (temporal separation) between the plots (for rank $i$) is an estimate of travel time, $t_{ti}$ for the $i^{th}$ vehicle under FIFO assumption. Therefore, the knowledge about the number of vehicles between upstream and downstream locations; or the travel time of individual vehicle can be applied to address the RD issue.

The number of vehicles between upstream and downstream locations is difficult to obtain. Alternatively, we can consider the knowledge about the queue length. Theoretically, the queue length can be defined as follows:

$$\text{Queue at time } t = U(t - t_{ff}) - D(t)$$  \hspace{1cm} (3)

Here, $t_{ff}$ is the free flow travel time of the link. This is further discussed in section 4.4.

There is an increasing use of probe vehicle, which can provide its travel time. Hence, in this paper we propose a methodology that integrates probe vehicle with cumulative plots to resolve the RD issue; and applies slicing technique for estimation of travel time quartiles.

4. The proposed methodology

4.1 Probe vehicle data

Here, probe vehicle is a vehicle equipped with vehicle tracking equipment. There are issues related to the probe vehicle data, such as map matching, frequency of probe data etc. To address these issues is beyond the scope of this paper. We assume that the time when the probe vehicle is at upstream ($t_u$) and downstream ($t_d$) locations is accurately obtained and its travel time is $t_d - t_u$.

4.2 Architecture of CUPRITE

The proposed methodology integrates cumulative plots with probes. The basic concept for the integration is introduced in Section 4.3. The architecture (see Figure 2) of the proposed algorithm is as follows (the details for which are provided in the following sections):

Step 1 Cumulative plots $U(t)$ and $D(t)$ are defined. Here, if the detector data is individual vehicle data (pulse data), then the cumulative counts can be obtained by cumulative the vehicles. However, if detector data is not a pulse data but an aggregated traffic counts during certain detection interval (for instance counts per 60 seconds), then cumulating the counts for each detection interval will not reflect the actual traffic fluctuations within the detection interval. These fluctuations can be captured by integrating the detector counts with signal timings, where the counts during the signal red phase is assigned to be zero, and counts during the signal green phase is segregated into counts from the saturation flow and counts.
from non saturation flow. Refer to Bhaskar et al., (2008) for the methodology for integration of signal timings with aggregated traffic counts from detector data for accurate representation of cumulative plots.

Step 2 Probe vehicle data, list of $[t_u]$ and $[t_d]$, is defined (Refer to Sections 4.1 and 4.3). Moreover, if the conditions for virtual probe (Section 4.4.1) 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 $t_{GE}$ and $t_{ff}$ are time corresponding to the end of signal green phase and link free flow travel time, respectively). Else, only real probes are considered.

Step 3 Points through which $U(t)$ should pass are defined (Section 4.5).

Step 4 $U(t)$ is redefined by first vertical scaling and shifting the plots (Section 4.6) so that it passes through the above defined points (Step 3); and

Step 5 Finally, for each estimation interval: a) average travel time (Section 4.7) is estimated using classical analytical procedure; and b) quartile for travel time (Section 4.8) is estimated using slicing technique.
Figure 2: CUPRITE architecture for estimation of travel time statistics.

4.3 Integration of cumulative plots and probes

Suppose, there is no RD and both \( U(t) \) and \( D(t) \) are perfect. Due to non-FIFO traffic behavior, even in the absence of RD, the rank of a vehicle in upstream and downstream cumulative plots may not be same i.e, \( U(t_u) \neq D(t_d) \) (Figure 3a). We fix the vehicle to downstream cumulative plot, i.e., we fix the rank of the vehicle in the cumulative plots as \( D(t_d) \) (Refer to Figure 3b) and define a parameter \( \Delta t \):

\[
\Delta t = U^{-1}(D(t_d)) - t_u
\]  

(4)

Where \( U^{-1}(D(t_d)) \) is the time when the vehicle is represented at \( U(t) \) given that we fix its rank to \( D(t_d) \).
Figure 3: a) Illustration for the relationship between probe data and cumulative plots; b) Fixing of probe data to D(t).

If all the vehicles in U(t) and D(t) are same then $\sum \Delta t$ from all the vehicles should be zero (5). This is an important property and is the explanation for the area between the plots to be the total travel time.

$$\sum_{\forall i} \Delta t_i = 0$$  \hspace{1cm} (5)

If there is presence of RD then the equation (5) is not satisfied. Therefore, RD issue can be addressed by correcting the cumulative plots such that equation (5) is satisfied.

Equation (5) is satisfied when we are considering all the vehicles. Each vehicle has an equal probability of being a probe vehicle and only a fraction of vehicles are randomly selected as probes. We make a hypothesis that we can remove or at least reduce the RD by redefining U(t) such that $\sum \Delta t$ from all the probes is zero.
In practice, we do not know which plot is responsible for RD issue. It can be U(t), D(t) or both. It is complicated to correct both U(t) and D(t) simultaneously. As the deviation amongst the plots is relative therefore, we can correct either U(t) or D(t). Here, we redefine U(t) because we fix the rank of the probe considering D(t). Alternatively, we can redefine D(t), if we fix the rank of the probe considering U(t).

4.4 Virtual probe

Virtual probe (Figure 4) is defined as a virtual vehicle that, during undersaturated traffic flow, departs from the downstream at the end of signal green phase (at time \( t_{GE} \)) and its travel time is free-flow travel time (\( t_{ff} \)) of the link. The probe is not real and is defined with the aim to reduce RD.

For undersaturated traffic condition, the vehicle queue formed during the signal red phase should be completely served during the signal green phase i.e., the queue length at time \( t_{GE} \) should be zero. Considering equation (3): \( U(t_{GE} - t_u) = D(t_{GE}) \) i.e., the travel time of the vehicle that enters the intersection at time \( t_{GE} \) should be close to \( t_u \). Therefore, under such conditions we can define virtual probe (see Figure 4) such that it is observed at upstream and downstream locations at time \( t_{GE} - t_u \) and \( t_{GE} \), respectively (i.e. for virtual probe \( t_u = t_{GE} - t_u \) and \( t_d = t_{GE} \)).

Note: Virtual probe is only defined if the following conditions for virtual probe are satisfied.

4.4.1 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 link the sources for significant mid-link delay, such as mid-link intersections and on-street bus stop should be absent.

ii. Virtual probes are defined only for undersaturated condition with logic of zero queue length at the end of signal green phase. Traffic condition is defined as undersaturated if counts during the signal cycle (or more specifically during signal green time) are less than the corresponding capacity (Figure 4) i.e.,

\[
D(t_{GE}) - D(t_{GE} - c) + \Delta < s^*g
\]  

Where: \( s \) and \( g \) are saturation flow rate and effective signal green time, respectively; \( s^*g \) is the capacity and \( \Delta \) is a calibration parameter to take into account the error in the estimation of capacity.

To define equation (6) it is assumed that there is no spill-over from downstream link. If there is spill-over, then vehicles are restricted to flow resulting in low counts at stop-line detector. Capacity is generally not corrected to account for the spill-over from downstream link. Due to which equation (6) is satisfied and system can falsely indicate undersaturated situation for spill-over cases. Though under such situation the queue may not vanish and hence virtual probe should not be defined.
iii. Virtual probe is defined with the aim to reduce RD. Hence, it should only be defined if there is presence of RD i.e., the following equation should be satisfied:

\[
U^{-1}(D(t_{GE}))-t_{GE} \not\in [t_g - \delta, t_g + \delta]
\]  

(7)

Where: \(\delta\) is a calibration parameter taking into account the variation in the estimation of \(t_f\). It can be considered equal to the standard deviation of the estimate of \(t_f\).

![Figure 4: Illustration of a virtual probe fixed to D(t).]

4.5 How to define the points through which \(U(t)\) should pass?

Say, we have \(n\) probe vehicles and the database for the probe is defined as list of \([t_u]\) and list of \([t_d]\), where the size of each list is \(n\). The value of \(f^p\) element in the list represents the data from the \(f^n\) probe. These lists are appended with additional elements satisfying the conditions for virtual probe (Section 4.4.1). If the conditions are satisfied, then time \(t_{GE}\) is appended to the list \([t_d]\); and time \((t_{GE} - t_f)\) is appended to the list \([t_u]\).

4.5.1 Grid technique

Consider an example, in Figure 5a, where we have four probes fixed to D(t). The \(U(t)\) should pass within the region satisfying the following constrain (Refer to the rectangular region in Figure 5a)

\[
\min \{t_u\} \leq t \leq \max \{t_u\} \quad \forall \text{probes}
\]  

(8)

\[
\min \{D(t_d)\} \leq \text{counts} \leq \max \{D(t_d)\} \quad \forall \text{probes}
\]  

(9)
We can define a grid with rows corresponding to \( D(t_d) \) and columns corresponding to \( t_u \) within the above region (Refer to Figure 5b). If \( U(t) \) passes through the diagonal nodes of the above grid then \( \sum \Delta t = 0 \) is satisfied. Therefore, the required points to pass are the diagonal nodes of the grid and can be obtained from the following algorithm:

**Step 1**  
Sort list \([t_d]\) in ascending order of its values. This is required as the rank of the probe is defined considering \( D(t) \).

**Step 2**  
Sort list \([t_u]\) in ascending order of its values. This is required to make sure that the redefined \( U(t) \) is monotonically increasing and satisfies the property of \( \sum \Delta t = 0 \).

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

![Figure 5a: Region](image)

![Figure 5b: Grid](image)

- Point through which \( U(t) \) should pass

\[
\Delta t_1 = -\Delta t_2; \Delta t_3 = -\Delta t_4 \implies \sum \Delta t_i = 0
\]

**Figure 5:** Points through which the \( U(t) \) should pass.
4.6 How to redefine $U(t)$?

4.6.1 Reference point

We define reference point as the point in which we have confidence that it is a correct point on the cumulative plot. Initially, $U(t)$ and $D(t)$ are 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 $P_n$, the reference point is $P_{n-1}$ and so on.

4.6.2 Vertical scaling and shifting technique

The RD issue is the result of: a) the error in the cumulative counts - due to error in detector counting; and b) inconsistency between the cumulative count at upstream and downstream locations - due to mid-link sources and sinks. Therefore, to address the issue, the cumulative counts (vertical axis) should be corrected. For this, we apply the following vertical scaling and shifting technique.

Say, a) point $(t_{\text{ref}}, U(t_{\text{ref}}))$ is a reference point; and b) point $(t_p, Y_p)$ is a point through which $U(t)$ should pass (Section 4.5). The redefined $U(t)$ should pass through both these points.

Refer to Figure 6, we define $Y_r = U(t_p) - U(t_{\text{ref}})$ and $y_r = Y_p - U(t_{\text{ref}})$.

![Figure 6: Illustration of the abbreviations for vertical scaling.](image)

a) For time $\leq t_{\text{ref}}$
The reference point is the correct point on the cumulative plot; therefore no correction on the cumulative plot is required for time less than and equal to $t_{\text{Ref}}$.

b) For $t_{\text{Ref}} < \text{time} \leq t_p$

We perform vertical scaling on $U(t)$ such that it passes through the point $(t_p, Y_p)$. The scale is defined as follows:

$$
\text{scale} = \begin{cases} 
\frac{Y_p - U(t_{\text{Ref}})}{Y_p - U(t_{\text{Ref}})} & \text{if } U(t_p) \neq U(t_{\text{Ref}}) \\
1 & \text{if } U(t_p) = U(t_{\text{Ref}})
\end{cases}
$$

(10)

The value of the scale reflects the net effect of the RD on the cumulative plots:

i. $\text{scale} > 1$: The plots are diverging. For diverging plots, the error can exponentially grow with time. For instance, there is a mid-link sink. The vehicles from the sink are observed at upstream and not at downstream.

ii. $\text{scale} < 1$: The plots are converging. For converging plots, $U(t)$ can cut $D(t)$ resulting in negative travel time estimation. For instance, there is a mid-link source. The vehicles from the source are observed at downstream but not at upstream.

iii. $\text{scale} = 1$: RD is absent.

Here, the relative deviation in the cumulative count at time $t$, $(\varepsilon_t)$ is the result of the accumulation of the relative deviation since time $t_{\text{Ref}}$ (Refer to Figure 6). Equation (11) defines the RD $(\varepsilon_{t_p})$ at time $t_p$.

$$
\varepsilon_{t_p} = Y_p - y_{t_p} = U(t_p) - Y_p
$$

(11)

The proportion of this relative deviation to the cumulative counts $(\frac{\varepsilon_{t_p}}{Y_p})$ is assumed to be constant (12).

$$
\frac{\varepsilon_{t_p}}{Y_p} = \frac{\varepsilon_t}{Y_t} ; \quad \forall t_{\text{Ref}} \leq t \leq t_p
$$

(12)

Where: $\varepsilon_t$ is the relative deviation at time $t$; $Y_t = U(t) - U(t_{\text{Ref}})$.

The above equations can be rearranged to define $\varepsilon_t$ as follows:

$$
\varepsilon_t = (1 - \text{scale}) \cdot Y_t ; \quad \forall t_{\text{Ref}} \leq t \leq t_p
$$

(13)

c) For $\text{time} > t_p$

All the points on $U(t)$ beyond time $t_p$ are shifted vertically so that the redefined $U(t)$ is continuous. The magnitude of the shift is the relative deviation observed at time $t_p$ (Eq (11)).
The above is summarized as follows: we redefine $U(t)$ (14) by applying correction (15) on it such that all points on the plot: i) before time $t_{\text{Ref}}$ have no correction; ii) between $t_{\text{Ref}}$ to $t_p$ are scaled vertically; and iii) beyond $t_p$ are shifted vertically.

$$U(t) = U(t) - \varepsilon,$$  

$$\varepsilon = \begin{cases} 
0 & \forall \ t \leq t_{\text{Ref}} \\
(1 - \text{scale}) \cdot (U(t) - U(t_{\text{Ref}})) & \forall \ t_{\text{Ref}} < t \leq t_p \\
U(t_p) - Y_p & \forall \ t > t_p 
\end{cases}$$  

Figure 7 represents an example, where $P_0$ is the initial reference point; and points $P_1$ and $P_2$ are two points through which $U(t)$ should pass (refer to section 4.5). First, the correction for point $P_1$ is performed with $P_0$ as the reference point (Refer to Figure 7b). Thereafter, $P_1$ becomes the reference point for $P_2$ and correction for $P_2$ is performed (Refer to Figure 7c). The redefined $U(t)$ considering points $P_1$ and $P_2$ is represented in Figure 7c.
Figure 7: Example for redefining $U(t)$ based on vertical scaling and shifting technique.
4.7 Average travel time estimation

The classical procedure (see section 3) is applied between redefined \( U(t) \) and \( D(t) \) to estimate average travel time.

4.8 Quartiles of travel time estimation

By definition, quartile is any value that divides the sorted data into equal parts:

i. \( Q_1 \): the first quartile is the 25\(^{th}\) percentile and 25\% of the data is lower than \( Q_1 \).

ii. \( Q_2 \): the second quartile is the 50\(^{th}\) percentile (or median) and it divides the data into two equal parts.

iii. \( Q_3 \): the third quartile is the 75\(^{th}\) percentile and 75\% of the data is lower than \( Q_3 \).

To obtain quartiles of travel time, we need either individual vehicle travel time data or grouped vehicle travel time data. The later is the data consisting of representative travel time from different group of vehicles. For FIFO systems, the horizontal distance (temporal separation) between the cumulative plots is an estimate for individual vehicle travel time. For both FIFO and non FIFO systems, the area between the plots is an estimate for total travel time for a group of vehicles represented within the plots. We propose the following slicing technique where, within an estimation interval, we slice the area between the cumulative plots to obtain the grouped vehicle travel time data.

4.8.1 Slicing technique

Cumulative plot is a two dimensional piece wise linear graph with coordinates \((t_i, i)\) as its nodes; where \( t_i \) is the time when the \( i^{th} \) vehicle is observed. In Figure 8, we illustrate a study link between u/s and d/s locations. In the figure, links \( L_1 \), \( L_2 \) and \( L_3 \) are three upstream links that contribute to the flow on the study link. Each of these links has a signal phase Phase\(_1\), Phase\(_2\) and Phase\(_3\) that permit the flow towards the study link, respectively. Time \( t_{gs1} \), \( t_{gs2} \) and \( t_{gs3} \) correspond to the start of signal green time for Phase\(_1\), Phase\(_2\) and Phase\(_3\), respectively.

Upstream cumulative plot, \( U(t) \), is defined by the flow contributions from different upstream links (\( L_1 \), \( L_2 \) and \( L_3 \)). We define a “cut node” as the node corresponding to the start of each signal green time (\( t_{gs1} \), \( t_{gs2} \) and \( t_{gs3} \)) for the upstream signal phases that permits the flow towards the study link. The “cut nodes” in the \( U(t) \) is marked in the figure. Here a node is a “cut node” if the following is satisfied:

\[
\text{if } ((t_{i-1} < t_{gs1} \leq t_i) \text{ or } (t_{i-1} < t_{gs2} \leq t_i) \text{ or } (t_{i-1} < t_{gs3} \leq t_i)) \\
\text{then node } (t_i, i) \text{ is a "cut node"}
\]  

(16)

Similarly, for downstream cumulative plot, we define a “cut node” as the node corresponding to the start of the signal green time for the downstream signal phases that permits the departure of the vehicles from the study link.
Figure 8: Simplified illustration of how a “cut node” is defined for upstream cumulative plot with flow contribution from different upstream links.

We define, $M_u$ as a matrix of the “cut nodes” for the upstream cumulative plot and similarly $M_d$ for downstream cumulative plot. Within an estimation interval, the total area $A$ between the cumulative plots, is fragmented into different areas ($A_i$) (see Figure 9), by horizontal cuts corresponding to the nodes at $M_u$ (“cut node” matrix for $U(t)$), $M_d$ (“cut node” matrix for $D(t)$) and with the following constraint:

For each fragmented area $A_i$, if the counts $N_i$ are above a certain threshold number, $N_{\text{threshold}}$, then the area ($A_i$) is further fragmented by a horizontal cut into two fragmented areas: $A_{i1}$ and $A_{i2}$ with counts $N_{\text{threshold}}$ and $N_i-N_{\text{threshold}}$, respectively. The threshold value provides an upper limit on the number of vehicles for each fragmented area.

The process is repeated until each fragmented area satisfies the constraint. Finally, each fragmented area ($A_i$) represents the total travel time for the $N_i$ number of vehicles. We assume that $N_i$ number of
vehicles experience similar travel time ($\overline{TT}_i$) equal to the $A_i/N_i$. This defines a grouped vehicle travel time data.

Figure 9: Illustration for slicing the area between cumulative plots for defining travel time for different group of vehicles within an estimation interval.

Note: if we define $N_{\text{threshold}} = 1$ vehicle, then the travel time estimates from the above procedure is equivalent to obtaining individual vehicle estimates as the horizontal distance between the plots.

An estimate for the quartiles of travel time is obtained by sorting the travel time values obtained from all the sliced areas and corresponding number of vehicles as follows:

For an estimation interval, say we have a two dimensional array with first column as list of $A_i/N_i$ (i.e., list $L_{AN}$) and second column as list of $N_i$ (i.e., list $L_N$). Following steps are followed:

- **Step 1** Sort the array with respect to the values in the list $L_{AN}$;
- **Step 2** Define a cumulative frequency list ($L_f$) by cumulating the values in the list $L_N$;
- **Step 3** Define $N_i$, as total number of vehicles in the estimation interval. This is the last element of the above cumulative frequency list;
Step 4 Define the index for the quartiles as follows:

\[ Q_{1\_index} = 0.25^*N \]
\[ Q_{2\_index} = 0.5^*N \]
\[ Q_{3\_index} = 0.75^*N \]

Step 5 Quartiles are defined as the value corresponding to the \( j^{th} \) element of the sorted list \( L_{AN} \) where \( j \) is the rank of \( L \) such that:

if \((j=0)\) and \((L[j] \geq Q_{3\_index})\),
then \( Q_3 = L_{AN}[0] \)

if \((j>0)\) and \((L[j-1] < Q_{3\_index}) \) and \((L[j] \geq Q_{3\_index})\),
then \( Q_3 = L_{AN}[j] \)

Similarly, for \( Q_2 \) and \( Q_1 \);

Note: here the elements of the list start from rank 0.

For better understanding of the above algorithm a self explaining example is presented in the Figure 10, where Table A is the original list of \( L_{AN} \) and \( L_N \); Step 1 and Step 2 are executed in Table B; Step 3 and Step 4 are performed in Table C; and finally, the quartiles are defined by performing Step 5 in Table B.
Figure 10: An example for quartile estimation using slicing technique.

5. Validation of the methodology with real data

The methodology is validated using real data from Lucerne, Switzerland. The traffic at the site is controlled by a fully actuated signal controller named VS-PLUS (VS-PLUS) that provides the detector counts and signal timings. Ground truth individual vehicle travel time, is obtained from manual number...
plate survey performed from 3:00 pm to 6:00 pm on 15th April, 2008 (working day). The probes are randomly selected from the survey individual vehicle data.

CUPRITE is applied on the link (see Figure 11) from Intersection A ("Kaserneplatz") to intersection D ("Pilatusplatz"). From A to B, there is minor side street acting as both source and sink; from B to C there is on-street bus stop; and from C to D, there are two different movements (left and through) associated with the link. Around 15% of the vehicles are lost in the side street in between intersection C to D. The detectors at the site are also not perfect.

The four stop-line detectors at A (a_1, a_2, a_3, a_4) provide total cumulative plot at the upstream (U_T). The downstream cumulative plots for through movement (D_{Thru}) and left movement (D_{Lft}) are obtained from stop-line detectors (d_1, d_2) and detector (d_3), respectively. U_T is scaled vertically using the average turning ratio of 55% for through movement and 30% for left movement to define the initial arrival cumulative plot for each movement.

![Figure 11](image-url)  
**Figure 11:** Illustration of the link characteristics between intersections A and D.

### 5.1 Average Travel Time estimation

#### 5.1.1 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 \( \bar{X}_s \) and \( S_v \), respectively. We define the confidence bounds in the actual average travel time (\( \mu_s \)) of the vehicles as:
\[
\bar{X}_s - t_{a/2, n_s - 1} \frac{S_s}{\sqrt{n_s}} \leq \mu_s \leq \bar{X}_s + t_{a/2, n_s - 1} \frac{S_s}{\sqrt{n_s}} \tag{17}
\]

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

5.1.2 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 with different values of the seed for random number generator to randomly select 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 \( \bar{X}_C \) and \( S_C \), respectively. Then we define the confidence bounds for the travel time estimate by CUPRITE as:

\[
\bar{X}_C - t_{a/2, n_C - 1} \frac{S_C}{\sqrt{n_C}} \leq \mu_C \leq \bar{X}_C + t_{a/2, n_C - 1} \frac{S_C}{\sqrt{n_C}} \tag{18}
\]

Where:

- \( \mu_C \) is the mean of the population of estimates from CUPRITE application;
- \( t_{a/2, n_C - 1} \) is the \( t \)-statistic at \( \alpha \) level of significance and \( n_C - 1 \) degrees of freedom.

Figure 12 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 12a) and the orange box represents the confidence bounds for the travel time estimates from the CUPRITE (Figure 12b). Note: In the results, if the mean from the CUPRITE is in within the confidence bounds from the survey then we can say that the travel time estimates from CUPRITE are statistically equivalent to that from the survey.

Accuracy of the estimates from CUPRITE is defined as following:

\[
Accuracy(\%) = (100 - MAPE) \tag{19}
\]

\[
MAPE = \sum_{i=1}^{n} \frac{Error_i}{n} \tag{20}
\]

\[
Error_i = \left( \frac{X_{s_i} - \bar{X}_C}{X_{s_i}} \right) \times 100 \tag{21}
\]
Where: Error$_i$ is the absolute percentage error for $i^{th}$ estimation interval; $\bar{X}_s$ and $\bar{X}_c$ 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.

Here, the estimation interval is five continuous signal cycles. During the analysis period, the cycle time varies between 96 s to 116 s. The level of significance for t-statistics considered is 0.05 (=α).

The results presented here are from for A→D$_{Lft}$. Figure 13, Figure 14 and Figure 15 illustrate results with one, two and three probes per estimation interval ($S_n$). 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 13: fourth, fifth, sixth and seventh estimation intervals (time from 15:30 hr to 16:00 hr) have significant variation in average travel time between the periods. This fluctuation is also accurately captured by CUPRITE.

For A→D$_{Lft}$: the accuracy (19) of the CUPRITE model increases from 92.3% to 94.6% with increase in number of probes from one probe per estimation interval (see Figure 13) to three probes per estimation interval (see Figure 15), respectively.

![Figure 12: Systematic representation of the results for CUPRITE validation.](image-url)

---

**Page 26 of 38**
Figure 13: Results for $A \rightarrow D_{Lh}$ with $S_n = 1$.

Figure 14: Results for $A \rightarrow D_{Lh}$ with $S_n = 2$. 
Figure 15: Results for A→D_Lft with S_n = 3.

5.2 Quartile for travel time estimation

Here slicing technique is applied and quartile Q_3 (75\textsuperscript{th} percentile) is estimated. The results are presented in Figure 16, Figure 17 and Figure 18 for one, two and three probes per estimation interval, respectively. Here the accuracy of the estimates from CUPRITE is defined as following:

\[ \text{Accuracy} \% = (100 - \text{MAPE}) \]

\[ \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} Error_i \]

\[ Error_i = \left| \frac{Q_{s_i} - Q_{C}}{Q_{s_i}} \right| \times 100 \]

Where: \( Error_i \) is the absolute percentage error for \( i^{th} \) estimation interval; \( Q_{s_i} \) is the Q_3 of survey travel time during \( i^{th} \) estimation interval (As CUPRITE is applied \( n_e \) number of times (Refer to section 5.1.2), therefore \( Q_{C} \) is the 75\textsuperscript{th} percentile of the different values of Q_3 of travel time obtained from CUPRITE application during \( i^{th} \) estimation interval.); \( n \) is the number of estimation intervals.
It is observed that the accuracy increases from 92.4% to 94.7% for increase in $S_n$ from one to three probes per estimation interval. The results are similar to what we have observed for application of CUPRITE for average travel time estimation.

The above analysis indicates the potential of CUPRITE for quartile travel time estimation in addition to the average travel time estimation.

Figure 16: $Q_3$ estimation using CUPRITE for route from $A \rightarrow D_{Lr}$ ($S_n=1$).
Figure 17: $Q_3$ estimation using CUPRITE for route from $A \rightarrow D_{Lft} (S_n=2)$.

Figure 18: $Q_3$ estimation using CUPRITE for route from $A \rightarrow D_{Lft} (S_n=3)$.
5.3 Discussion on probe as percentage of vehicles traversing the link ($S_p$)

The results presented in the previous section are with fixed number ($S_n$) of probes per estimation interval. In order, to capture the effect of probe market penetration we apply the model with probes as percentage ($S_p$) of vehicles traversing the route during three hour of survey period. Here during an estimation period there can be no probe ($S_n=0$) or at least one probe ($S_n>0$) (Refer to Figure 19). In the present analysis around 60% of the estimation intervals have no probe ($S_n=0$) for $S_p=1\%$ and the percentage of estimation intervals with $S_n=0$ decreases with increase in $S_p$.

Figure 20, Figure 21 and Figure 22, illustrates results for average travel time estimation for $S_p$ equals 1%, 2% and 3%, respectively. The results indicate that even with 1% of the probes CUPRITE can capture the fluctuations in time series of travel time. The accuracy of the estimation increased from 83.5% to 92.3% with increase in $S_p$ from 1% to 3%, respectively. Similar, Figure 23, Figure 24 and Figure 25 illustrate results for $Q_3$ estimates for $S_p$ equals 1%, 2% and 3%, respectively. The accuracy of the $Q_3$ estimation is close to 90%.

![Figure 19: Percentage of estimation intervals versus $S_n$ for route A→DLft for different $S_p$.](image-url)
Figure 20: Results for $A \rightarrow D_{Lft}$ with $S_p = 1\%$.

\[
\text{Accuray (\%) = 1-MAPE = 86.7\%}
\]

Figure 21: Results for $A \rightarrow D_{Lft}$ with $S_p = 2\%$.

\[
\text{Accuray (\%) = 1-MAPE = 89.24\%}
\]
Figure 22: Results for $A \rightarrow D_{Lt}$ with $S_p = 3\%$.

Figure 23: $Q_3$ estimation using CUPRITE for route from $A \rightarrow D_{Lt}$ ($S_p = 1\%$).
Figure 24: \( Q_3 \) estimation using CUPRITE for route from \( A \rightarrow D_{Lt} (S_p=2\%) \).

Figure 25: \( Q_3 \) estimation using CUPRITE for route from \( A \rightarrow D_{Lt} (S_p=3\%) \).
6. Conclusions

A majority of research on travel time estimation provide average travel time. The distribution of travel time from different vehicles departing within an estimation interval is generally bi-modal and hence it can happen that none of the vehicle can experience average travel time. For better understanding of the network performance statistics, such as quartiles should be explored.

The methodology proposed and validated in this paper provides travel time statistics (average and quartiles). The methodology is based on the classical analytical procedure for travel time estimation. The procedure is vulnerable to the relative deviation issue. This issue is addressed by integrating cumulative plots with probe vehicle data. For this, the probe data is fixed to the downstream cumulative plot \(D(t)\) and upstream cumulative plot \(U(t)\) is redefined: First, the points through which \(U(t)\) should pass are defined by applying grid technique thereafter, the \(U(t)\) is redefined by applying vertical scaling and shifting technique. The average travel time is estimated by applying classical procedure between redefined \(U(t)\) and \(D(t)\). For estimation of quartiles, slicing technique is proposed.

The methodology is validated using real data from Luzern city, Switzerland. The application site represents a typical urban network with: a) mixed traffic (with buses); b) on-street bus stops; c) mid-link sinks and sources; and d) detector counting error. The validation of the methodology on real network demonstrates its potential for practical application. The methodology requires few probes per estimation interval for accurate estimation. The current market penetration of probe is low, and with limited number of probes per estimation interval, it can considerably enhance the accuracy of travel time estimation on urban networks.

Though, the development of methodology is based on urban networks, but it can be equally applied to freeway facilities. It can be easily integrated with traffic monitoring system to simultaneously monitor both urban and freeway networks.

The probe vehicle data for the methodology is the time when the probe is at upstream and downstream locations. Advanced loop detectors with the capacity to provide vehicle signatures can be explored for vehicle re-identification. The re-identified vehicle can be a proxy for probe vehicle data. For this, even with low re-identification rate, the methodology can accurately estimate travel time.

The proposed methodology accurately estimates travel time, which is the experienced travel time. It should be extended further, for short-term travel time prediction, by exploring forecasting techniques, such as time series analysis, Artificial Neural Network applications etc.

The methodology integrates cumulative plots with probe vehicles. The cumulative plot considered is two dimensional i.e., both \(U(t)\) and \(D(t)\) are represented in the same figure. An avenue for extension of this research is to consider three three-dimensional (3D) representations of cumulative plots i.e., to consider cumulative counts; time; and location from upstream to downstream as three different axis, respectively. The integration of 3D representation of cumulative plots with the trajectory of a probe vehicle should be explored for detailed modeling of individual vehicle trajectories.
References


Dharia, A. & Adeli, H. (2003), Neural Network Model for Rapid Forecasting of Freeway Link Travel Time, *Engineering Applications of Artificial Intelligence*, 16(7-8), 607-613.


Jintanakul, K., Chu, L. & Jayakrishnan, R. (2009), Bayesian Mixture Model for Estimating Freeway Travel Time Distributions from Small Probe Samples from Multiple Days, Transportation Research Record, 2136, 37-44.


Park, D., Rilett, L. R. & Han, G. (1999), Spectral Basis Neural Networks for Real-Time Travel Time Forecasting, Journal of Transportation Engineering, 125(6), 515-523.


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(1457), 158-165.

Skabardonis, A. & Geroliminis, N. (2005), Real-Time Estimation of Travel Times Along Signalized Arterials, 16th International Symposium on Transportation and Traffic Theory, Maryland, USA.


