Urban traffic state estimation: Fusing point and zone based data

Article *in* Transportation Research Part C Emerging Technologies · November 2014

DOI: 10.1016/j.trc.2014.08.015

CITATIONS
21

READS
43

4 authors:

Ashish Bhaskar
Queensland University of Technology
110 PUBLICATIONS **884** CITATIONS
SEE PROFILE

Takahiro Tsubota
Queensland University of Technology
18 PUBLICATIONS **103** CITATIONS
SEE PROFILE

Minh Le Kieu
University of Auckland
28 PUBLICATIONS **268** CITATIONS
SEE PROFILE

Edward Chung
The Hong Kong Polytechnic University
196 PUBLICATIONS **1,648** CITATIONS
SEE PROFILE

Urban traffic state estimation: Fusing point and zone based data

By

Ashish Bhaskar
Smart Transport Research Centre
Science and Engineering Faculty
Civil Engineering and Build Environment School
Queensland University of Technology
Gardens Point Campus – S723
2 George St. GPO Box 2434, Brisbane QLD 4001, Australia
Phone: +61 7 3138 9985
e-mail: ashish.bhaskar@qut.edu.au

Takahiro Tsubota
Smart Transport Research Centre
Science and Engineering Faculty
Civil Engineering and Build Environment School
Queensland University of Technology
Gardens Point Campus – P872
2 George St. GPO Box 2434, Brisbane QLD 4001, Australia
Phone: +61 7 3138 9994
e-mail: t.tsubota@qut.edu.au

Le Minh Kieu
Smart Transport Research Centre
Science and Engineering Faculty
Civil Engineering and Build Environment School
Queensland University of Technology
Gardens Point Campus – S704
2 George St. GPO Box 2434, Brisbane QLD 4001, Australia
Phone: +61 423622443
e-mail: leminh.kieu@student.qut.edu.au

and
Edward Chung
Smart Transport Research Centre
Science and Engineering Faculty
Civil Engineering and Build Environment School
Queensland University of Technology
Gardens Point Campus – P850
2 George St. GPO Box 2434, Brisbane QLD 4001, Australia
Phone: +61 7 3138 1143
e-mail: edward.chung@qut.edu.au

ABSTRACT
Loop detectors are the oldest and widely used traffic data source. On urban arterials, they are mainly installed for signal control. Recently state-of-the-art Bluetooth MAC Scanners (BMS) has significantly captured the interest of stakeholders for exploiting it for area-wide traffic monitoring. Loop detectors provide flow- a fundamental traffic parameter; whereas BMS provides individual vehicle travel time between BMS stations. Hence, these two data sources complement each other, and if integrated should increase the accuracy and reliability of the traffic state estimation.

This paper proposed a model that integrates loops and BMS data for seamless travel time and density estimation for urban signalised network. The proposed model is validated using both real and simulated data and the results indicate that the accuracy of the proposed model is over 90%.

Keywords: *Bluetooth*, *Loop*, *Data fusion*, *Travel time*, *Density*, *Signals*

1 Introduction

Road traffic is a dynamic system where its state changes over both space and time. For instance, at any given time the level of congestion at different locations on the network is different. Similarly, at any given location, the level of congestion at different time is different. The system also has a self degrading behaviour, i.e., once the congestion occurs the system moves from bad to worse at both local and global scale. For instance, the capacity of the congested section drops and the congestion spreads over the network blocking further traffic. Traffic operators dynamically control traffic with the aim to reduce congestion and limit its degrading impacts. The control algorithms need accurate spatial temporal estimation of traffic state parameters (characterised by speed or travel time, flow and density), network parameters and demand. This paper contributes to the accurate and reliable estimation of traffic state parameters over the signalised urban networks.

The state of the traffic is estimated using data from various sensors that range from traditional inductive loops to advanced Bluetooth MAC scanners. Loops are the oldest and widely available data sources. Models have been proposed to estimate travel time (or speeds) [1-7] and density [8-11] from loop detector data. Most of these models are limited to motorway traffic only, where traffic is free to flow with no intersections (conflicting areas). On urban networks, the traffic state estimation is more challenging due to presence of external factors such as signals, side friction due to bus stops and parking, and mid-link sources and sinks (e.g. side streets).

On signalised urban network, loops are primarily installed for signal control. The location of the loop on the network is defined by the controller. For instance, SCATS [12] requires loops at stop-line whereas, SCOOT [13] requires loops to be at the upstream of the signalised link. The nature of traffic on the urban network is stop-and-go running condition where vehicles queue at the stop-line during signal red phase whereas, upstream of the link can be free flow. Thus, the speed obtained from the loop cannot be generalised over the signalised link. Similarly, unlike the motorway traffic the occupancy of the loops cannot easily provide the density of the entire link.

With the advancement in technology, advanced data sources such as Bluetooth MAC Scanners (BMS) [14, 15] and Wifi MAC Scanners [16-18] are also available. These scanners are extensively utilised for travel time estimation on both motorways [19] and arterials [20, 21]. However, they cannot directly measure density and flow.

The objective of this paper is to fuse loops and BMS for seamless, accurate and reliable traffic speed (travel time) and density estimation. The paper is structured as follows: First, in Section 2 the classical cumulative plot based traffic state estimation model and its issues are discussed. Section 3 provides detailed insight on BMS data. Section 4 presents the proposed traffic state estimation model with BMS and loop data. Thereafter, Section 5 and Section 6 discuss the testing and validation of the proposed model on the real and simulated network, respectively. Finally, the paper is concluded in Section 7.

2 Classical theoretical model for traffic state estimation

Cumulative plots are time series of the cumulative number of vehicles over time at a specific location on the network. Say, we have loop detectors at upstream and downstream of the road section. The counts from the loop detector at upstream and downstream can provide cumulative plots U(t) and D(t), respectively. Under ideal conditions of no counting errors from detectors and vehicles are conserved between the detectors i.e., vehicles observed at upstream and downstream are same and there is no loss or gain of vehicles within the section:

a) The vertical distance between the two plots (see Figure 1*a*) represents the number of vehicles between the two locations *veh(t)* (equation 1). The density of the road section is the number of vehicles divided by the length (L) of the section (distance between the points where U(t) and D(t) are defined) (equation 2)

$$veh(t) = U(t) - D(t) \quad (1)$$

$$Density(t) = \frac{U(t) - D(t)}{L} \quad (2)$$

b) The area between the plots as defined by the vertical cuts in Figure 1*c* represents the total density during time period of [t-λ/2, t+λ/2,]. Average density during this time period is the area divided by length of time period (λ) and section length (L) (see equation 3 where dt is the differentiation).

$$\bar{d}(t) = \frac{\int_{t-\lambda/2}^{t+\lambda/2} U(t).dt - \int_{t-\lambda/2}^{t+\lambda/2} D(t).dt}{\lambda.L} = \frac{\int_{t-\lambda/2}^{t+\lambda/2} Density(t).dt}{\lambda} \quad (3)$$

c) The horizontal distance between the plots (see Figure 1*a*) represents individual vehicle travel time (equation 4) under First–In-First-Out (FIFO) assumption.

$$tt_i = D^{-1}(i) - U^{-1}(i) \quad (4)$$

d) The area between the plots (see Figure 1*b*), defined by horizontal cuts, represents the total travel time, and average travel time is the area divided by the number of vehicles. Equation 5 represents average travel time at time *t* for all vehicles that arrive upstream during time period of [t-λ/2, t+λ/2] where λ is the time period over which the average is estimated.

$$\bar{tt} = \frac{\sum_{D^{-1}(U(t-\lambda/2))}^{D^{-1}(U(t+\lambda/2))} D(t) - \sum_{t-\lambda/2}^{t+\lambda/2} U(t)}{U(t-\lambda/2) - U(t+\lambda/2)} \quad (5)$$

e) The average speed $\bar{v}(t)$ (equation 6) is obtained from the average travel time (equation 5) and link length (L)

$$\bar{v}(t) = L\left(\frac{U(t+\lambda/2) - U(t-\lambda/2)}{\sum_{D^{-1}(U(t-\lambda/2))}^{D^{-1}(U(t+\lambda/2))} D(t) - \sum_{t-\lambda/2}^{t+\lambda/2} U(t)}\right) \quad (6)$$

Figure 1: Systematic illustration of classical cumulative plots theory: a) individual vehicle travel time and density b) total travel time; c) total density [INSERT]

As explained above, under ideal conditions, traffic state defined by time series of average speed (equation 6) (or travel time (equation 5)) and density (equation 2) can be easily estimated using cumulative plots. However, ideal conditions are difficult to achieve in practice. If cumulative plots are defined using loop detectors, then even under normal running conditions one can easily expect random or systematic errors in counting due to reasons such as, cross-talk, etc. [22]. Moreover, urban networks have mid-link sources or sinks (endogenous demand) such as parking, or side-street, which violates the conservation of vehicle assumption. Due to the detector counting error; non-conservation of vehicles between plots location; and any such combination over time, there is drifting or Relative Deviation (RD) among the plots. For instance, if there is a mid-link sink (vehicles have destination to mid-link) then we observe more vehicles at upstream and a proportion of these will be lost in the mid-link sink, resulting in lower counts at downstream. With time the difference between U(t) and D(t) will increase and the error in the traffic state estimation will exponentially grow. On the contrary, if there is a mid-link source (vehicles have source from mid-link), then number of vehicles observed at downstream will be higher than the one observed at upstream. This can result in $U(t) < D(t)$ and hence negative values of travel time and density.

Bhaskar et al., [23] has developed a model named CUmulative plots and PRobe Integration for Travel time Estimation (CUPRITE) where travel time from probe vehicle is utilised to resolve the RD issue. The objective for CUPRITE is to correct the RD issue and estimate average travel time with the following assumptions:

a) The time stamps of probe vehicles are available at the points where the cumulative plots are defined.
b) RD issue can be resolved by correcting only one of the plots. Under such assumption, one of the cumulative plots is considered as accurate and other is redefined using probe vehicle data. Here, if U(t) is redefined then:
    a. Mid-link sink vehicles which are observed only at U(t) are assumed to be removed in the redefine U(t), and
    a. Mid-link source vehicles which are observed only at D(t) are assumed to be inserted in the redefined U(t).

    On the contrary, if D(t) is redefined then:
    a. Mid-link sink vehicles which are observed only at U(t) (and not at D(t)) are assumed to be inserted in the redefine D(t), and
    b. Mid-link source vehicles which are observed only at D(t) (and not at U(t)) are assumed to be removed in the redefined D(t).

Probe vehicles are generally vehicles equipped with GPS (e.g. taxi fleets). The data from GPS includes time stamp and position coordinates. The frequency of data varies from applications to application. To know the precise time when the probe is observed at the detector location (where cumulative plots are defined) additional work related to the map-matching and interpolation between the data points is needed. Taxi data is managed

and owned by the private operators and might not be easily available. In this paper, we explore the use of BMS data (generally managed and owned by traffic operators) as a proxy for the probe in CUPTIRE application, the details for which are discussed in the next section (Section 3).

The aforementioned second assumption is not an issue when CUPRITE is applied for average travel time estimation. However, for density estimation it can result in error. For instance, if U(t) is redefined then a) for mid-link sink vehicles case, density should be underestimated (because vehicles are removed from U(t)), and b) for mid-link source case density should be overestimated (because vehicles are added to U(t)).

The originality of this paper is exploiting the BMS data as proxy for probes in CUPRITE, and utilising CUPRITE for both travel time and density estimation with BMS and loop data.

3 Bluetooth Media Access Control Scanner (BMS) Data

Recently there has been a significant interest of traffic stakeholders and researchers to exploit Bluetooth technology for estimating experienced travel time on the road network. The concept behind this is rather simple. Here, a Bluetooth MAC Scanner (BMS) targets to inquire the Media Access Identification (MAC-ID) codes of the discoverable Bluetooth devices within its communication range, termed as *zone.* Most of the portable electronic devices (such as mobile phones, car navigation systems etc.) are equipped with Bluetooth. A device in discoverable mode can respond back to the inquiry of the BMS with an inquiry package that contains the unique ID of the device (MAC-ID). BMS scans these packages and stores the respective MAC-ID and the time-stamp linked to the inquiry process. Time-synchronised BMSs on the road network has the potential to provide travel time from one BMS *zone* to another. For instance, travel time ($tt_{m,i,j}$) (equation 7) for MAC-ID *m*, from *i*th BMS to *j*th BMS is the difference of the time when the device is observed at i^{th} BMS($t_{m,i}$) and j^{th} BMS ($t_{m,j}$)

$$tt_{m,i,j} = t_{m,j} - t_{m,i} \qquad 7$$

A device travelling through the communication *zone* of the BMS can be detected multiple times, that depends on the time spent by the device in the zone. For instance, Figure 2 illustrates a trajectory of a vehicle, equipped with Bluetooth, through the BMS *zone*. The device has arrived at and departed from the zone at time τ_a and τ_d, respectively and during its travel it has been observed seven times by the BMS. The time-stamp corresponding to the first and last observation by the BMS are τ_{fr} and τ_{lr}, respectively. In order to reduce the data storage, BMS can be configured to only store the MAC-ID, τ_{fr} and duration (d_R). For instance, Table 1 provides sample data from a BMS.

Here, duration d_R is the time difference between τ_{lr} and τ_{fr}, which represents the reported travel time of the device through the communication zone. However, the actual duration of the device is d_A represented as time difference between τ_d and τ_a, which may not be equal to d_R due to the technical factors such as time needed for communication and BMS scanning cycle. Interested reader should refer to Bhaskar and Chung [14] for the detailed discussion

on the Bluetooth communication, BMS data acquisition and accuracy and reliability of the travel time estimation from BMS.

Figure 2: Conceptual illustration of a device travelling through the BMS communication zone and detected multiple times by the BMS. [INSERT]

Table 1: Sample data fields from a BMS [INSERT]

The shape and size of the communication zone of the BMS depends on the type and the strength of the antenna used, respectively. As a general rule of thumb, to capture the Bluetooth devices the communication zone should be large enough so that the targeted Bluetooth vehicles spend at least 5 seconds within the zone. For instance, the BMS scanners on Brisbane network have communication zone that is circular (omni-directional) with radius of 100 m and 150 m on signalised arterials and motorways, respectively. Travel time is defined between two points on the road network; hence, the travel time reported from the BMS, especially on the urban arterials, should clearly specify the type of matching used. For instance: (refer to Figure 3) If the time $t_{m,i}$ and $t_{m,j}$ in equation 7 are the time when the device is first detected (τ_{fr}) in respective zone. Then the travel time is for the section that corresponds to the entrance of the u/s BMS zone to the entrance of the d/s BMS zone (*En2En* in Figure 3). Whereas, if exit time ($\tau_{fr}+d_R$) is used then the travel time is for the section that corresponds to the exit of the u/s BMS zone to the exit of the d/s BMS zone (*Ex2Ex* in Figure 3). If traffic is free-flow in each zones (u/s and d/s) then travel time for *En2En* section and *Ex2Ex* section is expected to be the same. For urban networks, if BMSs are at the intersections then *En2En* travel time consists of partial delays from both u/s and d/s intersections. Whereas, *Ex2Ex* travel time only contains delay at d/s intersection and no delay from u/s intersection. Depending on the application, one can be interested to estimate point-to-point (P2P) travel time where the points P_u and P_d can be any point within the *u/s* and *d/s* BMS zones, respectively. For instance, P_u and P_d can be stop-line location at u/s and d/s intersections, respectively.

In this paper, we aim to utilise BMS travel time points as a proxy for the probe requirement for CUPRITE. Assuming loop detectors are at the stop-line. Therefore, we define P_u and P_d at the stop-line locations. In other words, we are interested to estimate travel time from the stop-line at the upstream intersection to the stop-line at the downstream intersection.

Figure 3: Illustration of travel time models between two BMS locations [INSERT]

3.1 Transforming zone to point

BMS data lacks the position of the vehicle within the zone. Therefore, the time when the vehicle is at the point of interest within the zone is to be estimated from the available BMS data. For this, the traffic behaviour within the zone should be known.

Here, we are interested to estimate the time when the vehicle is at the stop-line (equation 8).

$$t'_{m,i} = t_{m,i} - \Delta_{m,i} \quad (8)$$

Where :

$t'_{m,i}$ and $t_{m,i}$ is the time when the vehicle (with MAC-ID=m) is observed at the stop-line of intersection *i* and the exit of the BMS zone at intersection *i*, respectively.

$\Delta_{m,i}$ is the travel time of the vehicle (with MAC-ID=m) from the stop-line of intersection *i* to the exit of the respective BMS zone.

Stop-line to stop-line travel time can be defined as (equation 9):

$$tt_m = t'_{m,d/s} - t'_{m,u/s} = (t_{m,u/s} - \Delta_{m,u/s}) - (t_{m,d/s} - \Delta_{m,d/s}) \quad (9)$$

Where: *u/s* and *d/s* represents upstream BMS zone and downstream BMS zone, respectively.

The behaviour of the traffic at the upstream and at the downstream of the stop-line is different. Assuming: a) there is no spill-over from the downstream signal and b) BMS is located at the signal controller. Within the BMS zone, vehicles queue upstream of the stop-line and accelerate or cruse downstream of the stop-line. The dynamics of the vehicle at the downstream of the stop-line depends on:

a) The speed at which the vehicle will cruise (v_c): This speed is defined by the speed-limit of the road, driver's acceptance to the speed limit, and cruising speed of the leading vehicle in the general car-following condition.
b) The initial speed at the stop-line (u): This speed should not be greater than v_c. (i.e., $0 \le u \le v_c$). The value of which depends on the position of the vehicle within the queue. The first vehicle from the queue, which is waiting over the stop-line, will start to accelerate with initial speed of zero at the stopline whereas, other vehicle have non zero speed at the stop-line.
c) Acceleration rate (*a*) of the vehicle: Different vehicles have different acceleration rate, that depends on the vehicle mechanical characteristics and drivers' driving behaviour (harsh, smooth, eco-friendly).

Say: a) *r* is the distance from the stop-line to the exit (at further downstream) of the zone (*see* Figure 4); and b) *s* (equation 10) is the distance travelled by the vehicle while accelerating in the downstream region of the zone.

$$s = \frac{v_c^2 - u^2}{2.a} \quad (10)$$

Figure 4 illustrate examples of a vehicle trajectory through the BMS zone, where its deceleration, stopped and acceleration profile is marked as MN, NO, and OP, respectively. LM and PQ represent the cruising profile:

a) In Figure 4*a* vehicle does not observe any delay in the BMS zone (i.e., s = 0) and its duration is among the lowest values. In Figure 4*b*, vehicle observes delay only in the upstream of the stop-line. The value of Δ in both these cases should be simply r/v_c. Duration in Figure 4*b* is higher than that of Figure 4*a*.

b) In Figure 4*c* and Figure 4*d* vehicle observes delay at the upstream of the stop-line and has an accelerating delay at the downstream of the stop-line (s > 0). In this case, both Δ and duration are high.
    a. For Figure 4*c* vehicle reach its desired speed v_c before exiting the zone (r>s).
    b. In Figure 4*d* vehicle is still accelerating at the time of exit from zone (s ≥ r). This case is possible only when if $r \leq \frac{v_c^2}{2.a}$. Assuming v_c = 60 km/hr (16.7 m/s) ; a = 2.5 m/s², leads to $\frac{v_c^2}{2.a}$ = 55.6 m. Generally, the BMS zone is greater than 100 m in radius (r > 100m), hence this case is not generally applicable. However, for the completeness of discussion we present it here.

Figure 4: Examples of vehicle trajectory through the BMS zone: a) vehicle observing no delay through the BMS zone; b) vehicle observing delay at upstream region of the BMS zone, but no delay at downstream region of the BMS zone; c) and d) vehicle observing delay at both upstream and downstream regions of the BMS zone [INSERT]

As can be seen from above, Δ (equaton 11) depends on the parameters: r, v_c , u and a and hence it should be different for different vehicles.

$$\Delta = \begin{cases} \frac{r}{v_c} & if \quad s = 0 \\ \frac{v_c - u}{a} + \frac{r - s}{v_c} & if \quad r > s \\ \frac{-u + \sqrt{u^2 + 2.a.r}}{a} & if \quad r \leq s \end{cases} \qquad 11$$

If r, v_c , u and a are known then Δ can be easily estimated. However, BMS data only provides duration d for each vehicle. Duration depends on the delay of the vehicle within the BMS zone. The higher delay results in longer duration. Higher delay means more vehicles are in the queue. Position of the vehicle within the queue at the stopline defines its initial speed (u) at the stop-line which further defines Δ for a given r and a and v_c. Hence, there should be a relationship between Δ and d. We hypothesise that such relationship can be empirically estimated using simulation.

In our previous work [14], we have developed a multi-layered Traffic and Communication Simulation (TCS) model, where microscopic traffic simulation (AIMSUN) is integrated with Bluetooth Communication simulation. Microscopic traffic simulation simulates the behaviour of individual vehicle and provides details of the vehicle dynamics (position, instantaneous speed, etc.). TCS randomly assigns vehicles with an active Bluetooth considering the user defined penetration rate of the Bluetooth equipped vehicles within the traffic stream. The dynamics of the vehicles equipped with Bluetooth are input to the Bluetooth communication simulation. The communication simulation replicates the communication behaviour of an active Bluetooth device within the vehicle with that of BMS. TCS model stochastically considers the discovery time needed by BMS to read the MAC-IDs

of the vehicle. It simulates the data acquisition process of BMS and provides the data (MAC-ID, time) from the vehicles within the communication zone of the BMS.

As illustrated in Figure 2, the reported timestamps of the BMS are different from the actual time of entering at or exiting from the communication zone. The actual entering and exiting time, τ_a and τ_d, are directly available from the simulation by installing detectors at R metres upstream and downstream of the intersection (assuming BMS communication zone is R metres radius). The TCS emulates the reported BMS timestamps of the first and last detections, τ_{fr} and τ_{lr}, by introducing the temporal errors, ε_a and ε_d, to the actual arrival/departure time, τ_a and τ_d. As proven by Bhaskar and Chung [14], the temporal errors, ε_a and ε_d, follow a Generalised Gaussian Distribution (GGD) with the constraints:

$$\varepsilon_a \geq 0 \; ; \quad \varepsilon_d \geq -C_1 \tag{12}$$

where C_1 denotes the BMS inquiry cycle. Reported duration d_R is then defined as:

$$d_R = \tau_{lr} - \tau_{fr} = (\tau_d + \varepsilon_d) - (\tau_a + \varepsilon_a) \tag{13}$$

Interested, readers should refer to Bhaskar and Chung [14] for the details of the TCS model.

TCS is used to develop an empirical relationship between Δ and d (duration d is considered as reported duration d_R) (equation 14). Here, traffic is simulated for a network with two signalised intersection. The BMS is considered at the signal controller placed at the corner of the intersections. The BMS scan zone is 100 m. Following parameters are considered for the simulation:

a) The speed limit of the road as 60 km/hr.
b) The stochastic vehicle (drivers) speed limit acceptance factor (f) is assumed to be normally distributed with mean of 1.1 and standard deviation of 0.1. The speed at which the vehicle will cruise is f times the speed limit of the road.
c) The vehicle acceleration is also assumed to be normally distributed with mean of 3 m/s^2 and standard deviation of 0.25 m/s^2.

The simulation is performed over degree of saturation of 0.5 to 1.2 so as to replicate the duration over the different ranges. The parameters (α and β) for the equation 14 are calibrated with the data obtained from the simulation. Here, Δ is obtained from the trajectory of the Bluetooth vehicle and d is obtained from the BMS data acquisition replication. The parameters calibrated parameters using the aforementioned simulation are: α = 8.2624 and β = 0.978.

$$\Delta = \alpha * d^{(1-\beta)} \tag{14}$$

In this paper, we adopt the aforementioned relationship to estimate Δ from d. It is recommended that α and β should be calibration for the given site characteristics.

Substituting the equation 14 in equation 9 we define the stop-line to stop-line travel time for a vehicle (with MAC-ID = m) as follows (equation 15).

$$tt_m = t_{m,u/s} - t_{m,d/s} - \alpha * d_{m,u/s}^{(1-\beta)} + \alpha * d_{m,d/s}^{(1-\beta)} \quad (15)$$

3.2 An overview on the sample size of filtered BMS on signalised urban arterial

There can be significant noise in the estimated travel time (equation 15) from the BMS due to various reasons such as:

a) *Unknown mode:* The data from BMS corresponds to the device being transported by the traveller. The actual mode of transport is unknown. Pedestrian or cyclist can have higher travel time than that of motorised vehicles.
b) *Only Zone Information*: Travel time is only based on the information available within the BMS zone, the actual trajectory of the vehicle is unknown. If the vehicle stops along the route, due to any reason such as parking or rest area, then it will have higher travel time. Similarly, if there are multiple route choices between the two BMS zones, then the route taken is unknown.

The noise in the travel time estimates from BMS data can be filter out by applying standard statistical techniques such as Box and Whiskers [24], MAD [25] or adaptive [26]. More details in section 4.2.

On urban arterials, the sample size of the travel time estimates from BMS (after filtering) can range from one to three travel time points per minute (on average) and there can be periods when there is no data available. For instance Bhaskar et al., [15] has reported the average number of travel time points per minute from the morning peak periods (7 am to 9 am) varying from 1.2 to 3.6 points per minute during the working days. They have also highlighted the peak periods when no travel time point is available. Algorithms need to develop to fill this gap.

Due to the presence of signals, the variability of individual vehicle travel time on signalised arterials is quite large. The observed coefficient of variation (CV) of vehicle to vehicle travel time during 5 minute time periods on signalised routes in Switzerland has been reported to vary from 26% to 55% [22]. The day-to-day travel time variability on Brisbane is reported to vary from 20% to 50% (see Figure 5). Figure 5*a* and Figure 5*b* provides CV's for Coronation drive and Wynnum Road from 20 working Tuesdays. In the figure, the dotted points are the individual vehicle filtered travel time points (primary Y axis) and the solid black line is CV (secondary Y axis). It is observed that the CV during congestion shoulder (build-up and dissipation) periods is higher, which indicates that the uncertainty of travel time prediction during shoulder conditions is high.

Statistically, the number of samples needed such that the probability (equation16) of absolute relative deviation of error being less than ε_{max} is more than α (level of significance) is given by equation 17

$$P(|relative\,error| < \varepsilon_{max}) > \alpha \quad (16)$$

$$n = \frac{Z_{\alpha/2}^2 . CV^2}{\varepsilon_{max}^2} \quad (17)$$

Where:

n: sample size

$Z_{\alpha/2}$: Standard normal variates for α level of significance

($Z_{0.05/2}$= 1.96; $Z_{0.01/2}$=1.645);

CV: Coefficient of variation of the population

Rearranging equation 17 one can define the expected maximum ε_{max} for given sample size, CV and α as follows:

$$\varepsilon_{\max} = \frac{Z_{\alpha/2}.CV}{\sqrt{n}} \qquad 18$$

Bluetooth travel time data is the sample from the population of vehicles and, on average (on signalised arterial) we do not observe many Bluetooth travel time points. Figure 6 we define the time series for ε_{max} with 10% level of significance by applying equation 18 on the sites for which CV had been determined (in Figure 5). I.E., for the observed number of Bluetooth travel time points (sample size) and CV we are 90% confident that the absolute relative error in the average travel time estimation from the sample will be less than ε_{max} (illustrated in Figure 6). Figure 6*a* to Figure 6*h* illustrates plots for 8 days. The solid line represents the time series of the ε_{max} defined for α =0.01. The dotted line represents 10% error mark. If traffic operators are interested in absolute relative errors to be less than 10% for 90% of times, then one can easily see the time periods' when the errors (solid line) exceed the 10% mark (dotted line).

Figure 5: Travel time samples (green points, primary Y-axis) and CV (black line, secondary Y-axis) along a) Coronation drive, Brisbane and b) Wynnum Road, Brisbane obtained from 20 working Tuesdays. [INSERT]

Figure 6 Time series of ε_{max} (percentage of error) defined using sample size equation for the CV for different days along Coronation drive, Brisbane. [INSERT]

Although Bluetooth provides far better sample size than that compared to what is reported for floating car or probe vehicles. However, there are times, even during peak periods, when the sample size is statistically not strong enough for desired confidence in the average travel time estimation. This encourages researchers to explore avenues to have seamless and reliable travel time estimation on urban networks. In this paper, we propose one such methodology.

4 Traffic state estimation

If loops provide pulse information, then individual vehicle detection time at the location of the detector can be obtained, though the vehicle can't be identified. But loops generally provide aggregated counts during a detection interval (say 1 minute), where the individual vehicle detection time is unknown. Hence, exact matching of the BMS record with the detection of a vehicle by loop detector is not possible. The proposed model fuses loop data with Bluetooth by first defining cumulative plots using loops and thereafter, correcting the drifting of plots using Bluetooth. The architecture for the model is presented in Figure 7. The steps include:

a) Prepare cumulative plots at upstream and downstream intersections (*see* Section 4.1)
b) Prepare Bluetooth travel time from location of upstream detectors to downstream detectors (*see* Section 4.2)
c) Reduce drifting between the plots, by integrating Bluetooth with cumulative plots (*see* Section 4.3)
d) Estimate time series of traffic states (*see* Section 4.4).

Figure 7: Architecture for the traffic state estimation model [INSERT]

4.1 Prepare cumulative plots at upstream and downstream intersections

Say U(t) and D(t) represent cumulative plots at the upstream and the downstream intersections, respectively. These plots are defined by the stop-line loops at the respective intersection:

a) If the loop data is in raw format, i.e., it contains pulses (value 0 or 1) corresponding to the presence of a vehicles then the cumulative plot can be easily defined by cumulating the pulses. Here, depending on the quality of the data available, filtering can be applied on the pulses (interested reader should refer to Chapter 6 of Bhaskar [22] for details on the filtering the pulse data).
b) Generally, loop data available for the application is aggregated over a time period termed as detection interval, that can vary from 1 minute to 5 minutes. If data is aggregated then loop data should be integrated with signal timings to capture the fluctuations in flow profile. Here, during signal red phase there should not be any flow (i.e., slope of the cumulative plot should be zero) whereas, during signal green phase the flow should reflect saturation flow rate and demand pattern. Refer to Bhaskar et al., [27] for detailed modelling of accurate cumulative plots from aggregated loop data.

For the ease of understanding, conceptual representation of the cumulative plot estimation using loop data and signal timings is illustrated in Figure 8: Figure 8*a* illustrates time series of cumulative plot defined by cumulating the vehicle counts obtained from a loop at an interval of DI (detection interval) seconds. The plots here are granulated to the level of detection interval. The flow during a detection interval is assumed as uniform and is defined as the ratio of the counts (N_d) obtained during a detection interval and corresponding detection interval time. Figure 8*b* illustrated a portion of the cumulative plot during a detection interval (from Figure 8*a*). Figure 8*c* conceptually illustrates the transformation of Figure 8*b* to more accurate representation, where the flow during the signal red phase is modified to zero, and flow during the signal green phase is redistributed considering the signal green time. The cumulative plot defined by the integration of signal timings and aggregated loop data is more accurate as it captures the stop and running traffic behaviour at the signalised arterials. Interested readers should refer to the above-mentioned references for details.

Figure 8: Conceptual representation of the estimation of cumulative plot from loops and signal timings: a) illustrates a time series of cumulative counts, with each interval of DI; b) Illustrates integration of signal timings with detector data to capture the fluctuations in the cumulative plots due to signals. [INSERT]

4.2 Prepare Bluetooth based individual vehicle stop-line to stop-line travel time points

Here, data from the BMSs at upstream and downstream are matched using equation 15. This provides a list of [$t'_{m,u/s}$, tt_m] where $t'_{m,u/s}$ is the time when the m-th Bluetooth device is observed at the stop-line of upstream intersection ($t'_{m,u/s} = t_{m,u/s} - \alpha * d_{m,u/s}^{(1-\beta)}$) and tt_m is the corresponding stop-line to stop-line travel time (equation 15)

Thereafter, filtering is applied on the matched travel time points to differentiate between the valid travel time points and noise. For this a moving window of Δt (= 6 min) is defined. For each time window, confidence bounds of travel time are defined in terms of Upper Bound Value (UBV) and Lower Bound Value (LBV). Valid travel time points are the one which lie within the UBV and LBV, and points outside this range are considered as noise.

UBV and LBV can be defined using different statistical filters such as Box and Whiskers, Median Absolute Deviation (MAD filter) etc. For this, paper we consider MAD filter. MAD is the median of the absolute deviations from the data median. For time *t,* and moving window of Δt the median (M_t) and MAD_t is defined as below (equation 19):

$$\forall m \ \ such that \ t'_{m,u/s} \in [t - \frac{\Delta t}{2}, t + \frac{\Delta t}{2}]$$

$$M_t = median(tt_m) \qquad 19$$

$$MAD_t = median(|tt_m - M_t|)$$

Where *median* is an operator that provided the median of the data values.

The UBV_t(equation 20) and LBV_t (equation 21) at time *t* are defined as follows:

$$UBV_t = M_t + f.\sigma_t \qquad 20$$

$$LBV_t = M_t - f.\sigma_t \qquad 21$$

Where: σ_t is the standard deviation from the MAD and for a normally distributed data, it is approximated as (equation 22):

$$\sigma_t = 1.4826 * MAD_t \qquad 22$$

f represents the scale factor for the confidence bounds. Lower *f* results in higher confidence but with lower sample size, and *vice versa*. The value of *f* has been suggested by some authors to be from 2 to 5 [28, 29]. We recommend *f=2* for higher confidence in the travel time estimates from the BMS data.

By applying the MAD filter, valid travel time points are obtained. Valid travel time points provides a database of the time when the bluetooth vehicle is observed at upstream intersection stop-line [t_u] (equation 23) and time when it is observed at downstream intersection stop-line [t_d] (equation 24).

$$[t_u] = [t'_{m,u/s}] \quad \forall m \tag{23}$$

$$[t_d] = [t'_{m,u/s} + tt_m] \quad \forall m \tag{24}$$

Thereafter, list $[t_d]$ and $[t_d]$ are integrated with the U(t) and D(t) to estimate the time series of the average travel time and density and discussed in the following sections.

4.3 Reduce drifting: Integrating cumulative plots with BMS data

BMS provides the lists $[t_u]$ and $[t_d]$ that should be integrated with U(t) and D(t). There are two possible ways to integrate:

a) *Fix BMS with D(t)*: Here the rank (equation 25) of the Bluetooth vehicle in the cumulative plots is defined using D(t). The error in integration along y-axis (cumulative counts) and x-axis (time) is represented as e_y (equation 26) and e_x (equation 27), respectively (*see* Figure 9*a*).

$$rank = D(t_d) \tag{25}$$

$$e_y = D(t_d) - U(t_u) \tag{26}$$

$$e_x = t_u - U^{-1}(D(t_d)) \tag{27}$$

b) *Fix BMS with U(t)*: Here the rank (equation 28) of the Bluetooth vehicle in the cumulative plots is defined using U(t). The error in the integration along y-axis (cumulative counts) and x-axis (time) is represented by e_y (equation 29) and e_x (equation 30), respectively (*see* Figure 9*b*).

$$rank = U(t_u) \tag{28}$$

$$e_y = U(t_u) - D(t_d) \tag{29}$$

$$e_x = t_d - D^{-1}(U(t_u)) \tag{30}$$

Figure 9: Illustrative example of fixing a BMS data with cumulative plots: a) BMS fixed with D(t); b) BMS fixed with U(t). [INSERT]

In the absence of drifting: a) both e_x and e_y for each vehicle should be zero for FIFO situations; and b) the summation of errors for all the vehicles should be zero (equation 31) for non-FIFO situations

$$\sum_{\forall m} e_{x,m} = 0 \ \ \& \sum_{\forall m} e_{y,m} = 0 \tag{31}$$

Note: If e_x is zero then e_y will also be zero and *vice versa*.

Due to drifting, equation 31 will not be satisfied. We can reduce drifting by assuming one of the cumulative plots as correct and re-defining the other plots so that equation 31 is satisfied.

Whether BMS should be fixed to U(t) or D(t) is identified considering the confidence we have in the accuracy of the respective cumulative plots. If loops are at the stop-line (SCATS) then we have more confidence in D(t) than U(t) whereas, if loops are at the upstream entrance (SCOOTS) of the link then we have more confidence in U(t) than that of D(t). For instance:

a) In Figure 10*a*, loops are at the stop-line. Here D(t) is directly obtained by summing the cumulative plots from detector d_{d1}, d_{d2}, and d_{d3}. Whereas, U(t) should be defined as :

$$U(t) = \eta_{u2} CP_{u2} + \eta_{u5} CP_{u5} + \eta_{u6} CP_{u6} \quad (32)$$

Where CP_{u2}, CP_{u5} and CP_{u6} are the cumulative plots defined at detectors d_{u2}, d_{u5} and d_{u6}, respectively. And η_{u2}, η_{u5} and η_{u6} are scaling factors to the cumulative plots CP_{u2}, CP_{u5} and CP_{u6}, respectively. In other words, η_{u2}, η_{u5} and η_{u6} represent the proportion of the counts observed at upstream detectors d_{u2}, d_{u5} and d_{u6}, respectively towards the study link. In the example presented, η_{u2} is unity as detector d_{u2} is not on a shared-use lane. The values of η_{u2}, η_{u5} and η_{u6} depend on the turning proportions and need to be estimated in real time. The best estimate of scaling factor during different time periods can be obtained using the historical data clustered during time of the day and day of the week (Refer to Bhaskar et al. [30] for details). However, the process can be complicated and hence it is easier to assume that D(t) is accurate and we perform corrections on U(t).

b) In Figure 10*b*, loops are at the upstream entrance of the link. Here U(t) is directly obtained by summing the cumulative plots from detector d_{u1}, d_{u2}, and d_{u3}. Whereas, D(t) is defined as sum of the cumulative plots from detectors d_{d1}, d_{d2}, and d_{d3}. Here D(t) for the study link is not accurate due to additional counts from link L1 and L2 and loss in counts from L3 and L5. An estimate for D(t) can be obtained by considering the signal parameters. However, the process can be complicated. In this case it is easier to assume that U(t) is accurate and we perform corrections on D(t).

Figure 10: Examples of detector configurations a) at stop-line and b) at upstream entrance of the link [INSERT]

Drifting is due to the errors in cumulating the number of vehicles (y-axis), hence the corrections applied on the cumulative plots should aim to set e_y to zero. For this:

a) *Points2Pass ($P2P_i$) are defined*: These are points through which the cumulative plot should pass.
    a. Refer to Section 4.3.1 if BMS is fixed to D(t) i.e., we assume D(t) as correct and re-define U(t)
    b. Refer to Section 4.3.2, if BMS is fixed to U(t) i.e., we assume U(t) as correct and re-define D(t)

b) *Cumulative plots are scaled vertically so that they pass through the $P2P_i$.*: Say, CP(t) represents cumulative plots, which can be either D(t) or U(t) depending on how $P2P_i$ are defined. The details for the scaling are defined in Section 4.3.3

Say we have *m* BMS travel time points that are used to correct drifting in cumulative plots. I.E., size of the list [t_u] and the list [t_d] is *m* each. We append these lists with initial value of 0, that corresponds to the zero counts at time zero. The size of the lists is now m+1, each.

4.3.1 Define Points2Pass: Assuming D(t) as correct

When we fix BMS data to D(t) it provides two lists: [t_u] and [$D(t_d)$]. The lists are sorted independently in ascending order of its value and $P2P_i$ 33 are defined as coordinates in the cumulative plot diagram:

$$\begin{aligned} &\text{Points 2 Pass}\,(P2P_i) = (x_i, y_i) \quad \forall i = 0,1,2,\ldots,m \\ &where: x_0 = 0; and\ y_0 = 0 \\ &\qquad x_i = ith\ element\ of\ the\ sorted\ list\left[t_u\right] \qquad \forall i = 0,1,2,\ldots,m \\ &\qquad y_i = ith\ element\ of\ the\ sorted\ list\left[D(t_d)\right] \ \forall i = 0,1,2,\ldots,m \end{aligned} \tag{33}$$

4.3.2 Define Points2Pass: Assuming U(t) as correct

When we fix BMS data to U(t) then it provides two lists: [t_d] and [$U(t_u)$]. The lists are sorted independently in ascending order of its value and $P2P_i$ (equation 33) are defined as coordinates in the cumulative plot diagram:

$$\begin{aligned} &\text{Points 2 Pass}\,(P2P_i) = (x_i, y_i) \quad \forall i = 0,1,2,\ldots,m \\ &where: x_0 = 0; and\ y_0 = 0 \\ &\qquad x_i = ith\ element\ of\ the\ sorted\ list\left[t_d\right] \qquad \forall i = 0,1,2,\ldots,m \\ &\qquad y_i = ith\ element\ of\ the\ sorted\ list\left[U(t_u)\right] \ \forall i = 0,1,2,\ldots,m \end{aligned} \tag{34}$$

4.3.3 Vertical scaling on CP(t) (U(t) or D(t))

Here, CP(t) is corrected so that it passes through the defined *Points2Pass*. For each $P2P_i$ (i>0) corrections are applied along y-axis on CP(t) such that it passes through *Point2Pass.* Refer to Figure 11 that illustrates the variables used here. We first define scale factor (*s*), which is the ratio of the errors along y-axis at $P2P_i$ and the increase in cumulative counts from time x_{i-1} to time x_i. If the increase in cumulative counts is zero, then scale is considered as 1.

For all time before x_{i-1}, there is no need for correction, as CP(t) has already been corrected until x_{i-1}. For time between x_{i-1} and x_i, scaling is applied whereas, for time beyond x_i CP(t) correction has the same magnitude as that at x_i. The pseudo code for this is provided in equation 35.

$$\forall i > 0$$

$$\forall t$$

$$CP(t) = CP(t) + correction$$

$$s = \begin{cases} \dfrac{CP(x_i) - y_i}{CP(x_i) - y_{i-1}} & if \quad CP(x_i) \neq y_{i-1} \\ 1 & if \quad CP(x_i) = y_{i-1} \end{cases}$$

$$correction = \begin{cases} 0 & \forall t \leq x_{i-1} \\ (s-1)*(CP(t) - y_{i-1}) & \forall t \in (x_{i-1}, x_i) \\ y_i - CP(x_i) & \forall t \geq x_i \end{cases}$$

Figure 11: Illustration of variables used to model the vertical scaling [INSERT]

4.4 Traffic state estimation

Once the drifting between U(t) and D(t) are corrected by applying the above procedure- where for stop-line detector U(t) is redefined (*see* Figure 12) and for upstream detector D(t) is redefined (*see* Figure 13). Traffic state is estimated by considering the revised cumulative plots. Here, average travel times for different estimation period are estimated to define the time series of travel time using equation 5. The time series of travel time can be transformed to the time series of average speed considering the link length using equation 6. Similarly, time series of average density is estimated using equation 3.

Figure 12: Flowchart for traffic state estimation from stop-line detectors [INSERT]

Figure 13: Flowchart for traffic state estimation from detectors that are at upstream entrance of the links [INSERT]

5 Model application on real data

The proposed model is applied on the real data from Brisbane along Coronation drive (traffic inbound towards CBD) as illustrated in Figure 14. The real data includes:

a) Signal timings from SCATS
b) SCATS stop-line loop counts aggregated over a period of 5 minutes
c) Bluetooth travel time from BMS stations

Length of section is 2.5 km, with 3 lanes in one direction. BMS scanners are located at the signalised intersection the locations for which are highlighted as yellow markers in Figure 14.

Figure 14: Study site, Coronation Drive (inbound traffic), Brisbane. [INSERT]

Figure 15 represents the travel time from Bluetooth on the study day. Bluetooth outliers are identified using MAD filter and are illustrated as grey points in the figure. Time series of 5 minute average travel time from filtered Bluetooth data is considered as ground truth

(see Figure 16*a* for the morning peak period). To demonstrate the effectiveness of CUPRITE we remove 30 minutes BMS data as if BMS was not available and apply CUPRITE the fill the gap. For instance Figure 16*b* to Figure 16*h* illustrates the periods (marked rectangle) when BMS data is purposely removed and the gap is filled by applying CUPRITE during the highlighted period. Filled travel time profile by CUPRITE is represented as red rectangles with red line. As can be seen from the figure, CUPRITE can very well fill the missing gap.

Figure 17 compares CUPRITE travel time with that of actual travel time (from Bluetooth) for different days with 30 minutes of missing Bluetooth data, R^2 (Coefficient of determination) for which is 0.902. The accuracy of CUPRITE is evaluated using the following performance indicators:

a) A_m (Equation37): This is the average of the accuracies (A(d,p), equation 36) obtained from all the estimation periods. It indicates the average performance, and is mathematically equivalent to 100(%) minus MAPE (Mean Absolute Percentage Error).
b) A_5 (Equation 38): This is the 5th percentile of the individual accuracies obtained (A(d,p) Equation 36) which means that 95% of the times the accuracy is more than A_5.

$$A(d,p)(\%) = 1 - \frac{\left|TT_{Act}(d,p) - TT_{Est}(d,p)\right|}{TT_{Act}(d,p)} \quad 36$$

$$A_m(\%) = \frac{\sum_{d=1}^{d=D}\sum_{p=1}^{p=P} A(d,p)}{D*P} \quad 37$$

$$A_5(\%) = 5th\ percentile\ of\ A(d,p) \quad 38$$

Where $TT_{Act}(d,p)$ and $TT_{Est}(d,p)$ are the actual average travel time (from BMS) and estimated (from CUPRITE) travel time, respectively during p^{th} estimation period (5 minute each) from d^{th} day.

For the current analysis we have 42 (=P) estimation periods (6 points per 30 minutes for seven 30 minutes intervals per day) for 5 (=D) days. The analysis results have 93.3% of A_m and 80.8% of A_5, which validates the capacity of CUPRITE for seamless travel time estimation.

Performance of the CUPRITE is evaluated against the naïve travel time estimation, where the missing travel time records are filled with historical average, i.e., aforementioned for each 30 minutes of missing Bluetooth records, instead of using CUPRITE we use historical average to evaluate 5 minute average travel time records. Here the historical average is from 15 working days and the validation days are not included in estimating the average. Figure 18 illustrates graph for estimated historical travel time versus actual travel time during the validation periods. One can see that the historical average highly underestimates the travel time during the validation periods. The performance using historical average is 84% of A_m and 65% of A_5. This indicates that CUPRITE significantly outperforms historical average for both average and reliability of estimation, with around 10% improvement in mean accuracy and 15% improvement in reliability.

Figure 15: Travel time profiles along BMS site [INSERT]

Figure 16: a) Bluetooth travel time points and average for 5 minutes; b to h: Results from the application of CUPRITE on 22nd October 2012 with 30 minutes of missing Bluetooth data. [INSERT]

Figure 17: CUPRITE travel time verus actual travel time (from BMS) for different validation days [INSERT]

Figure 18: Historical average travel time verus actual travel time (from BMS) for different validation days [INSERT]

Figure 19 illustrates the results for the application of CUPRITE for density estimation for the entire day. As expected, the time series of density follows similar patters as that of the time series of travel time, with peaks during the morning and evening. Density cannot be measured directly, and hence no ground truth for density is know. Due to which the estimated density from CURPRITE cannot be validated with real data. To overcome this problem we move to synthetic data as explained in the next section.

Figure 19: Density estimation using CUPRITE using real data from Coronation drive, Brisbane. [INSERT]

6 Validating the density estimation from CUPRITE using simulation

The accuracy of CUPRITE for density estimation is evaluated in a simulation environment under Traffic and Communication Simulation (TCS) framework[14]. TCS integrates the BMS communication process with a microscopic traffic simulation to reproduce the temporal errors associated with the Bluetooth measurement (refer to Section 3.1).

Figure 20: Illustration of the study site and the benchmark density estimation [INSERT]

6.1 Simulation setting

The TCS is embedded in AIMSUN microscopic simulator, and the performance of the CUPRITE for density estimation is evaluated under the scenarios where cumulative plots have significant drifting. We consider both mid-link source and sink (side street along the route as shown in Figure 20). The test network consists of a two-lane section between two consecutive signalised intersections with 120 seconds cycle time and 0.25 green split. The section has one mid-link sink (or source) point at 840 metres from the downstream intersection (Figure 20). The vehicles for the mid-link sink (or source) are randomly selected from the vehicles traversing the section.

Demand is generated for 2 hours, which increases during the first hour, then becomes constant for the next hour. In order to test under different congestion levels, various scenarios are simulated with the degree of saturation at the downstream intersection during the peak period ranging from 0.9 to 1.2; each scenario is replicated 10 times with different random seeds. With regard to the significance of drifting, 10% mid-link sink and

source cases are tested. The percentage of sink vehicles is defined as the ratio of vehicles lost into the sink to the vehicles observed at upstream. The percentage of source vehicles is defined as the ratio of vehicles gained from the source to the vehicles departing from downstream.

Bluetooth equipped vehicles are randomly selected from the individual simulated vehicles. Regarding the penetration rate, 5 cases are assumed; 1%, 5%, 10%, 15% and 20%, being consistent with the observed penetration of Bluetooth data from Brisbane arterial network [31]. Simulation is performed with 20 replications where each replication has a different random seed, resulting in different selection of the Bluetooth vehicle data.

6.2 Benchmark density

In order to keep the consistency, the benchmark density is also measured with cumulative plots. For this, additional detectors are installed at the mid-link (M1 and M2 in Figure 20), that is, the immediate upstream and downstream of the mid-link source and/or sink, which provides cumulative curves, M1(t) and M2(t), respectively. The set of plots directly measure the number of vehicles between upstream and mid-link (from U to M1) and between mid-link and downstream (from M2 to D), which is, in turn, converted to the benchmark density. The benchmark value is calculated as:

$$Benchmark(t) = \frac{\left(U(t) - M1(t)\right) + \left(M2(t) - D(t)\right)}{L} \quad (39)$$

$$\bar{d}(t) = \frac{\int_{t-\lambda/2}^{t+\lambda/2} Benchmark(t).dt}{\lambda} \quad (40)$$

where L (270 m +840 m for Figure 20) denotes the section length.

6.3 Results

Figure 21 and Figure 22 summarise the CUPRITE density estimation results from the aforementioned simulation settings. Given the section is equipped with stop-line detectors, the results are based on the case where the downstream curve is assumed correct and the upstream curve is redefined accordingly. Density is estimated over the time period of three signal cycles (i.e., λ= 6 minutes in equation 40). Figure 21 and Figure 22 are from mid-link sink and mid-link source, respectively. The left graphs compare the benchmark density (X-axis) and the estimated density from CUPRITE (Y-axis), only for 1% penetration case. The graph and the R^2 value confirm the reasonable performance of CUPRITE for density estimation.

The right graphs compare the accuracy indicators, A_m and A_5, as presented in equation 37 and equation 38, respectively. Overall, the better accuracy is obtained as the penetration rate increases. When 20% of vehicles are equipped with the Bluetooth communication, the results show 96.96% of A_m and 90.53% of A_5 for 10% sink case, and 95.76% of A_m and 87.53% of A_5 for 10% source case, which validates the effectiveness of CUPRITE for the density estimation of signalised arterial sections.

Figure 21: CUPRITE density estimation vs Benchmark for mid-link sink cases with Bluetooth measurement errors [INSERT]

Figure 22: CUPRITE density estimation vs Benchmark for mid-link source cases with Bluetooth measurement errors [INSERT]

7 Conclusions

Road network is equipped with variety of traffic data retrieval systems ranging from traditional loops to advanced BMS. In this paper, loops and BMS are integrated for seamless traffic state estimation. Following are the conclusions from the analysis performed in the paper:

a) BMS provides a good estimate of experienced individual vehicle travel time along the road network. The sample size for which depends on various factors and there are periods when the observed sample size is statistically low to have sufficient confidence in the average travel time estimation. There are also periods when the data is not available due to malfunctioning of the BMS. Hence, for seamless and reliable travel time estimation BMS should be complemented with other data sources.

b) Loops provide counts that are cumulated to define cumulative plots at the location of the loop location. The classical theoretical model of using cumulative plots can provide accurate travel time and density. However, this theoretical model can't be applied in real world due to violation of conditions for accurate counting and conversation of traffic flow. The violation results in drifting of the plots. The proposed model reduces drifting by fusing cumulative plots with BMS data. For this BMS data is transformed from zone data to point data and empirical model for the transformation is proposed.

c) The strength of the proposed model is that it not only provides seamless travel time but also the density. Density is very difficult to measure and estimate in real network. For the measurement of density, very sophisticated measurements such as helicopter data can be utilised, which is practically not achievable. A simple method like the one proposed here provides an accurate estimation of density, which is significant contribution to the state-of-the-art.

d) For travel time estimation the model is validated using real data. The results indicate that the mean accuracy (A_m) and reliability (A_5) for travel time estimation is over 90% and 80%, respectively.

e) For density estimation, the performance is evaluated using simulation. Here, Bluetooth penetration rate of 1% to 20% is considered. The results indicates A_m varies from 93% to 96% and A_5 varies from 80% to 90% with increasing penetration rate.

f) The travel time estimated from the proposed model is compared with that of naïve travel time estimation (historical average). The proposed model has 10% improvement in mean accuracy and 15% improvement in reliability of the estimation than that of the naïve model.

REFERENCES

[1] B. Coifman and S. Kim, "Speed estimation and length based vehicle classification from freeway single-loop detectors," *Transportation Research Part C: Emerging Technologies,* vol. 17, pp. 349-364, 2009.

[2] B. Coifman and S. Krishnamurthy, "Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure," *Transportation Research Part C: Emerging Technologies,* vol. 15, pp. 135-153, 2007.

[3] C. Cortés, R. Lavanya, J.-S. Oh, and R. Jayakrishnan, "General-Purpose Methodology for Estimating Link Travel Time with Multiple-Point Detection of Traffic," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1802, pp. 181-189, 2002.

[4] B. Coifman and S. Kim, "Measuring freeway traffic conditions with transit vehicles," *Transportation Research Record,* vol. 2121, pp. 90-101, 2009.

[5] D. J. Dailey, "A statistical algorithm for estimating speed from single loop volume and occupancy measurements," *Transportation Research Part B: Methodological,* vol. 33, pp. 313-322, 1999.

[6] Y. Wang and N. L. Nihan, "Can single-loop detectors do the work of dual-loop detectors?," *Journal of Transportation Engineering,* vol. 129, pp. 169-176, 2003.

[7] A. Bhaskar, M. QU, and E. Chung, "A Hybrid Model for Motorway Travel Time Estimation- Considering Increased Detector Spacing," in *Transportation Research Board 93nd Annual Meeting*, Washington, D.C., 2014.

[8] M. Papageorgiou and G. Vigos, "Relating time-occupancy measurements to space-occupancy and link vehicle-count," *Transportation Research Part C: Emerging Technologies,* vol. 16, pp. 1-17, 2008.

[9] G. Qian, J. Lee, and E. Chung, "Algorithm for Queue Estimation with Loop Detector of Time Occupancy in Off-Ramps on Signalized Motorways," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2278, pp. 50-56, 2012.

[10] A. Sharma, D. Bullock, and J. Bonneson, "Input-Output and Hybrid Techniques for Real-Time Prediction of Delay and Maximum Queue Length at Signalized Intersections," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2035, pp. 69-80, 2007.

[11] G. Vigos, M. Papageorgiou, and Y. Wang, "Real-time estimation of vehicle-count within signalized links," *Transportation Research Part C: Emerging Technologies,* vol. 16, pp. 18-35, 2008.

[12] P. R. Lowrie, "SCATS, Sydney Co-Ordinated Adaptive Traffic System : A Traffic Responsive Method of Controlling Urban Traffic," Roads and Traffic Authority, Darlinghurst, NSW Australia1990.

[13] P. B. Hunt, D. I. Robertson, R. D. Bretherton, and R. I. Winton, "SCOOT - a traffic responsive method of co-ordinating signals," *TRL Laboratory Report 1014,* 1981.

[14] A. Bhaskar and E. Chung, "Fundamental understanding on the use of Bluetooth scanner as a complementary transport data," *Transportation Research Part C: Emerging Technologies,* vol. 37, pp. 42-72, 2013.

[15] A. Bhaskar, L. M. Kieu, M. Qu, A. Nantes, M. Miska, and E. Chung, "Is bus overrepresented in Bluetooth MAC Scanner data? Is MAC-ID really unique?," *International Journal of Intelligent Transportation Systems Research (http://dx.doi.org/10.1007/s13177-014-0089-9),* 2014.

[16] N. Abedi, A. Bhaskar, and E. Chung, "Bluetooth and Wi-Fi MAC Address Based Crowd Data Collection and Monitoring: Benefits, Challenges and Enhancement," in *36th Australasian Transport Research Forum (ATRF)*, Brisbane, Australia, 2013.

[17] M. Abbott-Jard, H. Shah, and A. Bhaskar, "Empirical evaluation of Bluetooth and Wifi scanning for road transport " in *36th Australasian Transport Research Forum (ATRF)*, Brisbane, Australia, 2013.

[18] N. Abedi, A. Bhaskar, and E. Chung, "Tracking spatio-temporal movement of human in terms of space utilization using Media-Access-Control address data," *Applied Geography,* vol. 51, pp. 72-81, 2014.

[19] Y. Wang, Y. Malinovskiy, Y.-J. Wu, and U. K. Lee, "Error Modeling and Analysis for Travel Time Data Obtained from Bluetooth MAC Address Matching," Department of Civil and Environmental Engineering University of Washington2011.

[20] Z. Mei, D. Wang, and J. Chen, "Investigation with Bluetooth Sensors of Bicycle Travel Time Estimation on a Short Corridor," *International Journal of Distributed Sensor Networks,* vol. 2012, 2012.

[21] S. Quayle, P. Koonce, D. DePencier, and D. Bullock, "Arterial Performance Measures with Media Access Control Readers," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2192, pp. 185-193, 2010.

[22] A. Bhaskar, "A methodology (CUPRITE) for urban network travel time estimation by integrating multisource data," Doctor of Sciences (Ph.D.), School of Architecture, Civil and Environmental Engineering, Ph.D. Thesis ,The Ecole Polytechnique Fédérale de Lausanne (EPFL) Lausanne, 2009.

[23] A. Bhaskar, E. Chung, and A.-G. Dumont, "Fusing Loop Detector and Probe Vehicle Data to Estimate Travel Time Statistics on Signalized Urban Networks," *Computer-Aided Civil and Infrastructure Engineering,* vol. 26, pp. 433-450, 2011.

[24] T. Tsubota, A. Bhaskar, E. Chung, and R. Billot, "Arterial traffic congestion analysis using Bluetooth duration data," presented at the 34th Australasian Transport Research Forum (ATRF), Adelaide, South Australia, Australia 2011.

[25] L. M. Kieu, A. Bhaskar, and E. Chung, "Bus and car travel time on urban networks: Integrating Bluetooth and Bus Vehicle Identification Data," presented at the 25th Australian Road Research Board Conference, Perth, Australia, 2012.

[26] S. Salek Moghaddam and H. Bruce, "Algorithm for Detecting Outliers in Bluetooth Data in Real Time," presented at the 93rd Transport Research Board Annual Meeting, Washington D.C., USA, 2014.

[27] A. Bhaskar, E. Chung, and A.-G. Dumont, "Analysis for the Use of Cumulative Plots for Travel Time Estimation on Signalized Network," *International Journal of Intelligent Transportation Systems Research,* vol. 8, pp. 151-163, 2010.

[28] L. Davies and U. Gather, "The identification of multiple outliers," *Journal of the American Statistical Association,* pp. 782-792, 1993.

[29] R. K. Pearson, "Outliers in process modeling and identification," *Control Systems Technology, IEEE Transactions on,* vol. 10, pp. 55-63, 2002.

[30] A. Bhaskar, E. Chung, and A.-G. Dumont, "Average Travel Time Estimations for Urban Routes That Consider Exit Turning Movements," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2308, pp. 47-60, 2012.

[31] A. Bhaskar, L. M. Kieu, M. Qu, A. Nantes, M. Miska, and E. Chung, "On the use of Bluetooth MAC Scanners for live reporting of the transport network," in *The 10th International Conference of Eastern Asia Society for Transportation Studies*, Taipei, Taiwan, 2013.

Figure

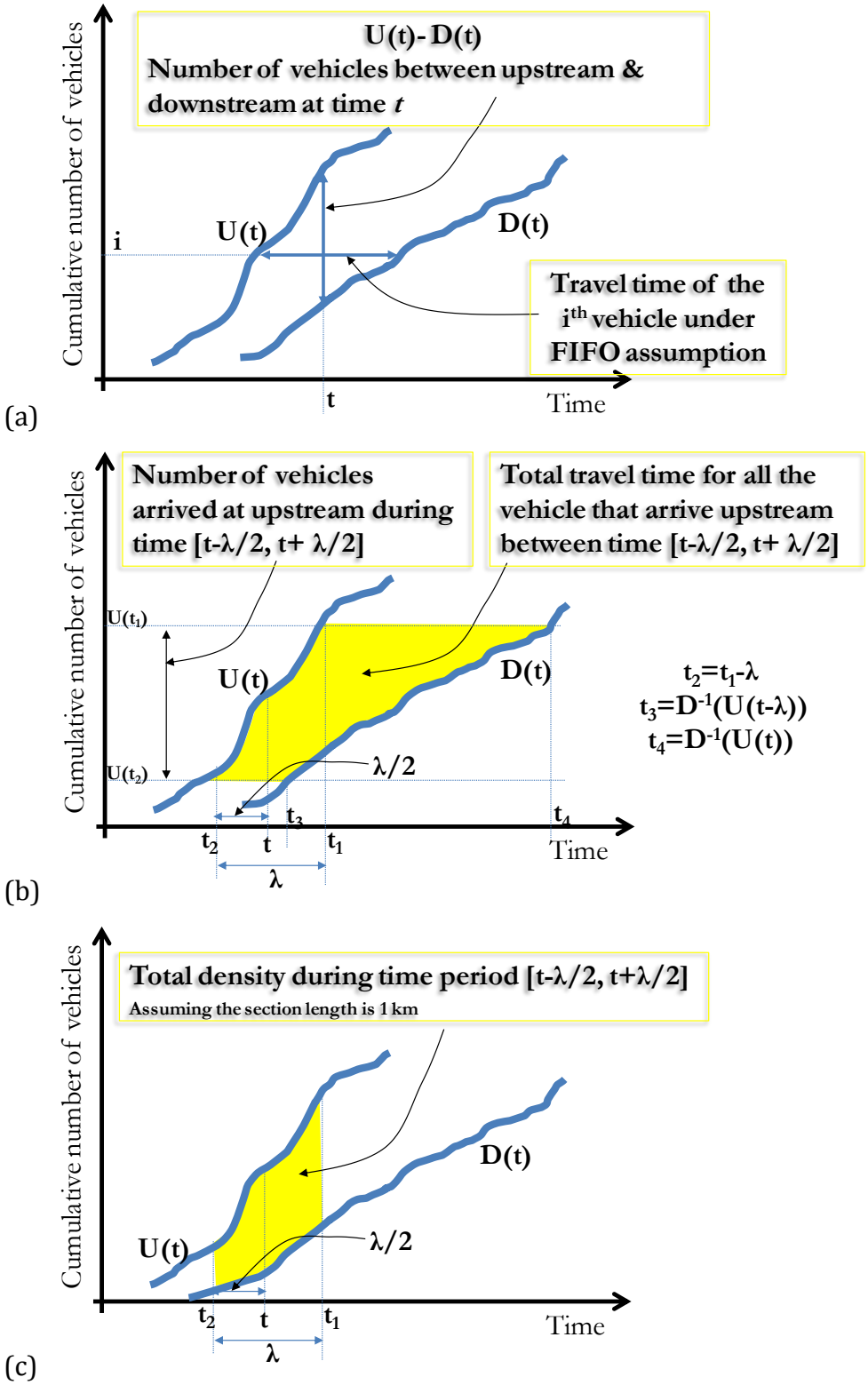

Figure 1: Systematic illustration of classical cumulative plots theory: a) individual vehicle travel time and density b) total travel time; c) total density

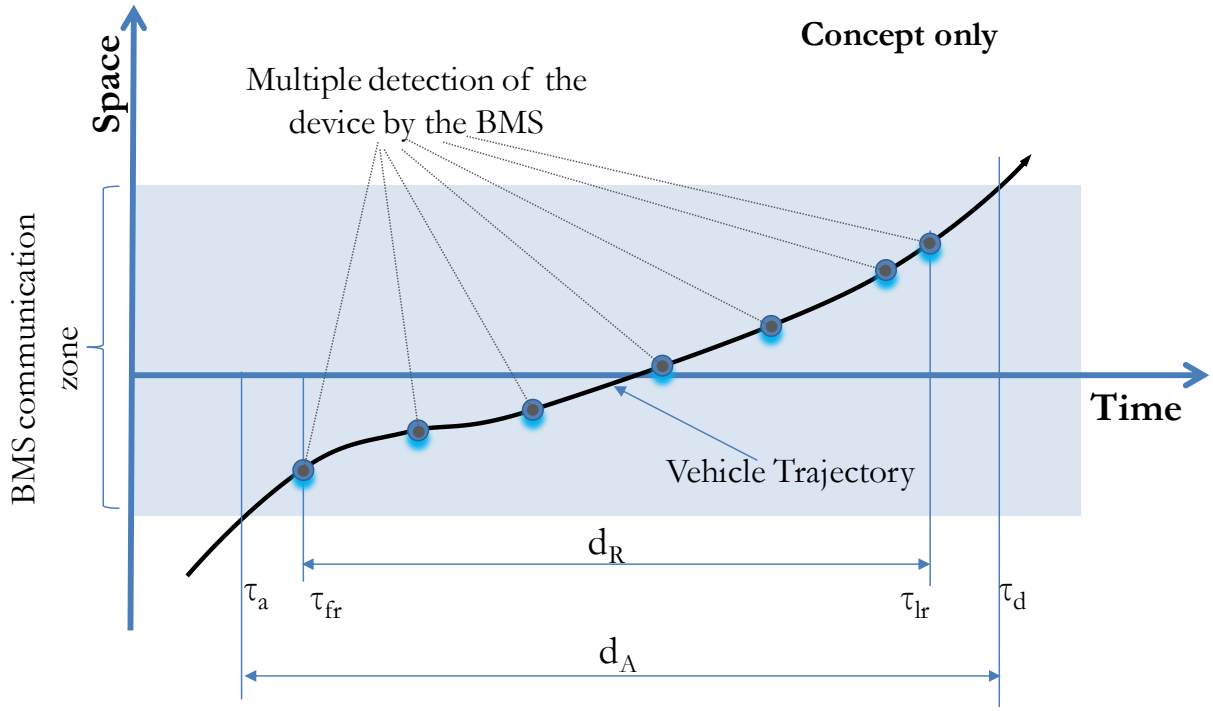

Figure 2: Conceptual illustration of a device travelling through the BMS communication zone and detected multiple times by the BMS.

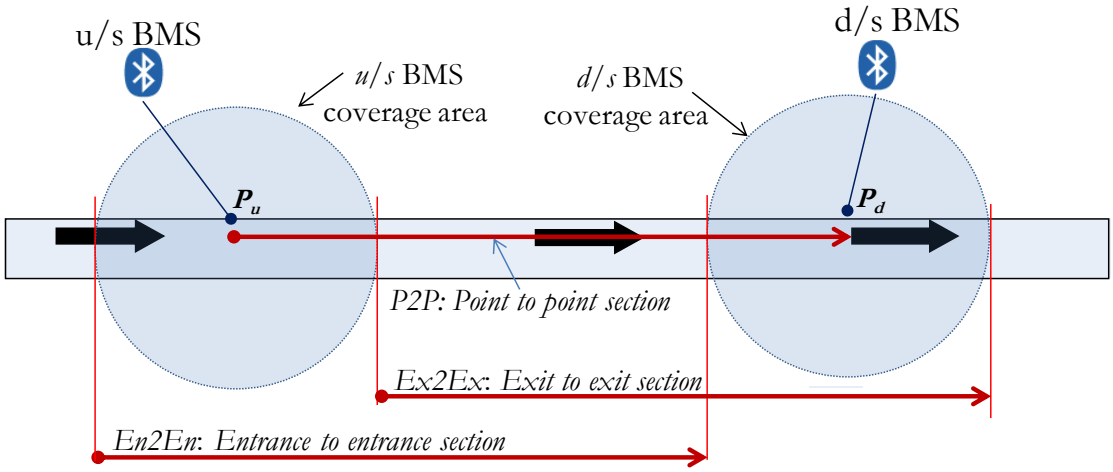

Figure 3: Illustration of travel time models between two BMS locations

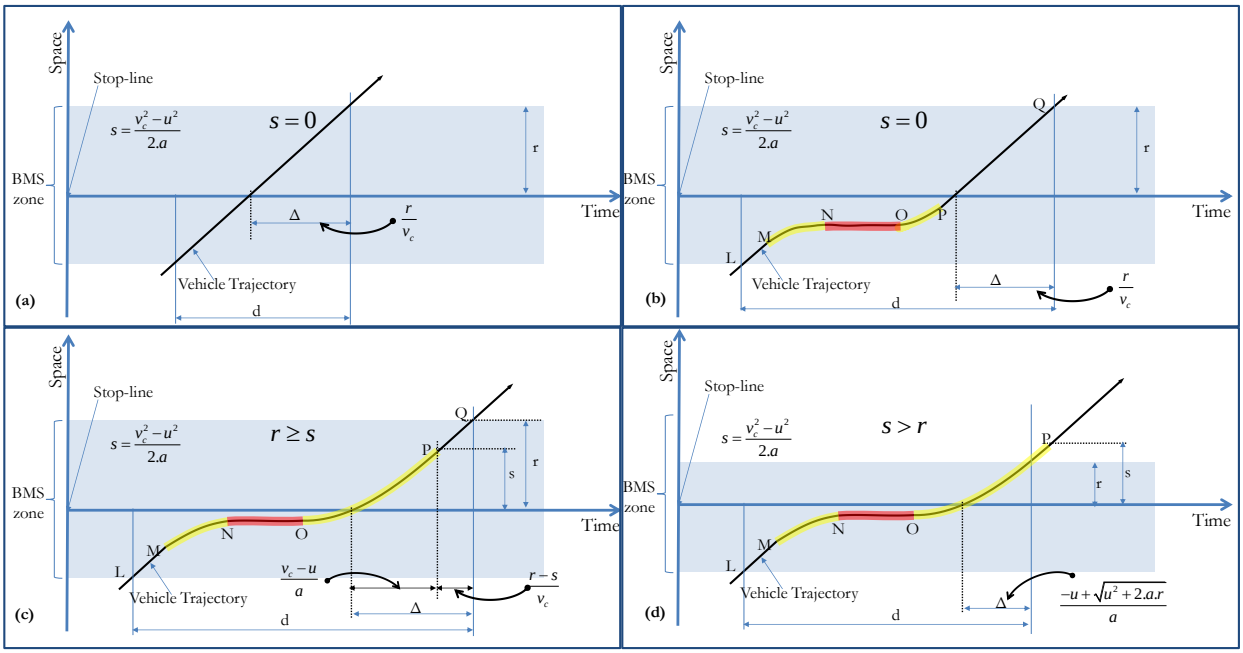

Figure 4: Examples of vehicle trajectory through the BMS zone: a) vehicle observing no delay through the BMS zone; b) vehicle observing delay at upstream region of the BMS zone, but no delay at downstream region of the BMS zone; c) and d) vehicle observing delay at both upstream and downstream regions of the BMS zone

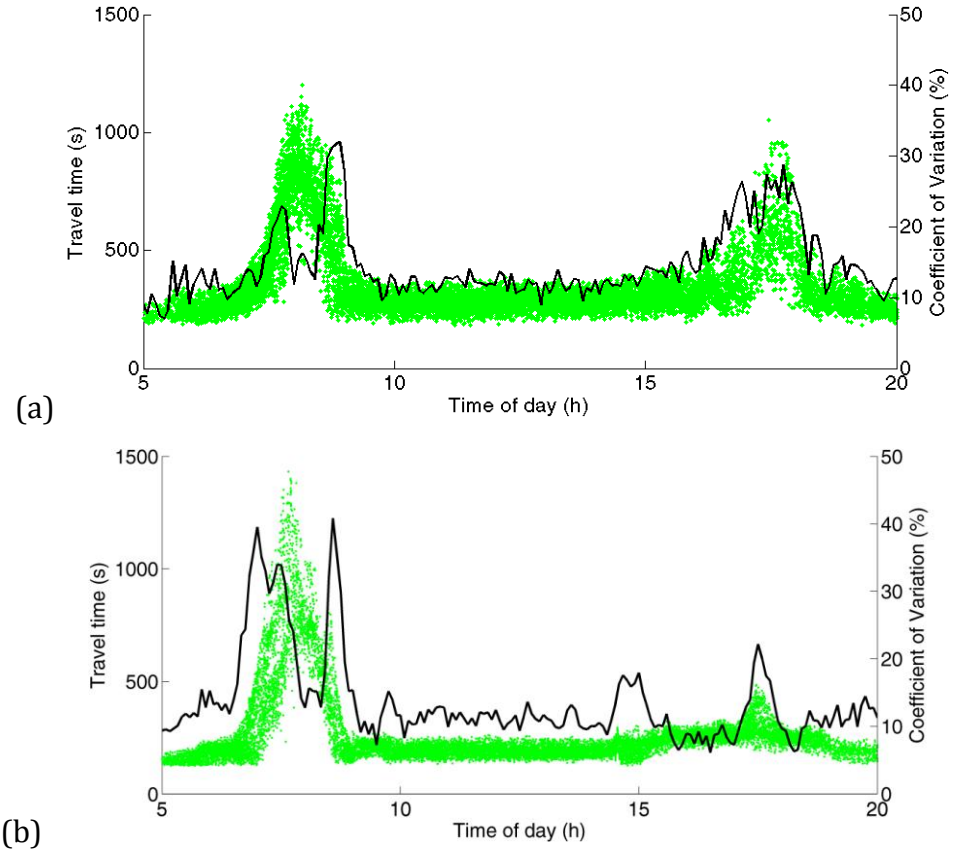

Figure 5: Travel time samples (green points, primary Y-axis) and CV (black line, secondary Y-axis) along a) Coronation drive, Brisbane and b) Wynnum Road, Brisbane obtained from 20 working Tuesdays.

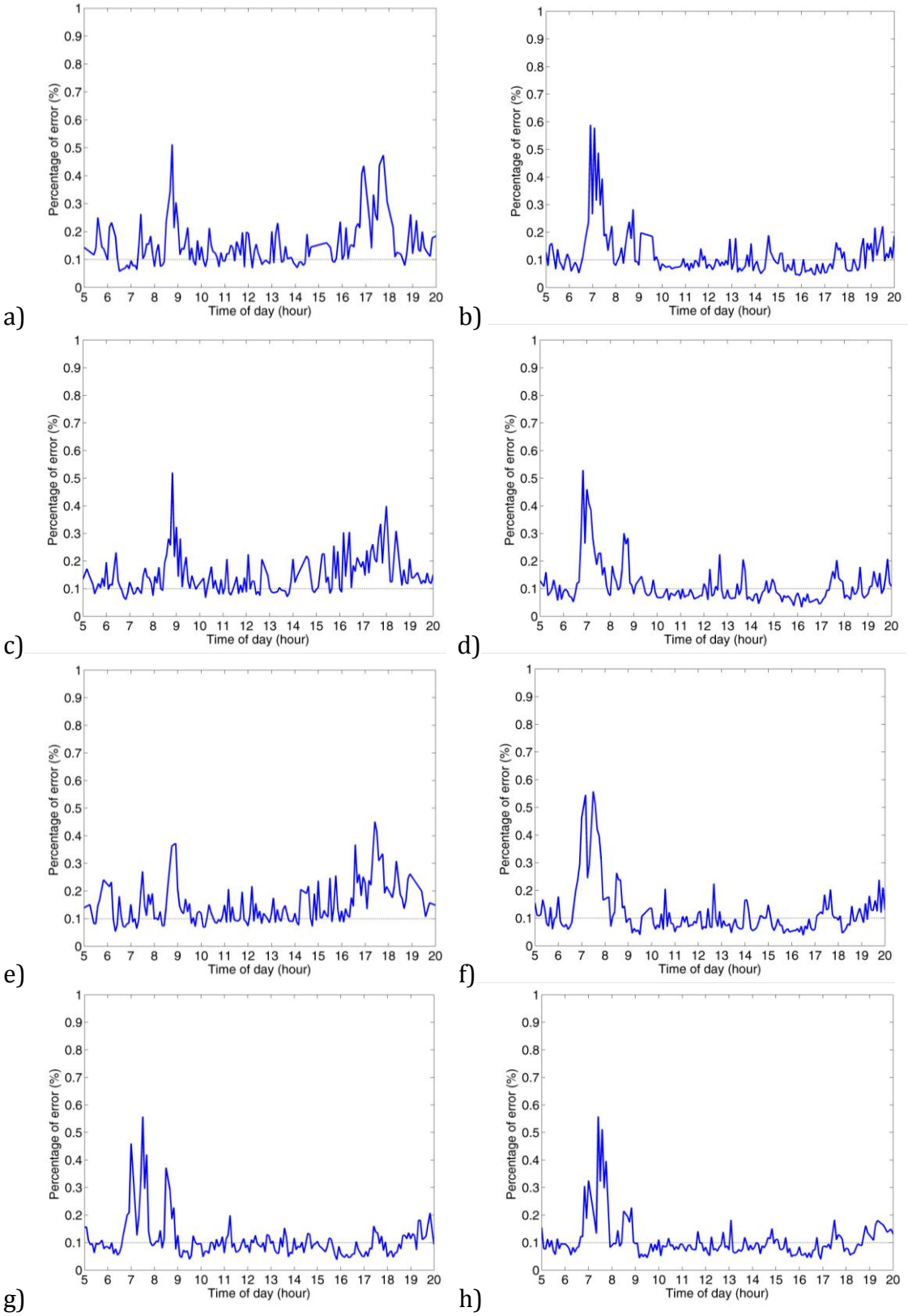

Figure 6 Time series of ε_{max} (percentage of error) defined using sample size equation for the CV for different days along Coronation drive, Brisbane.

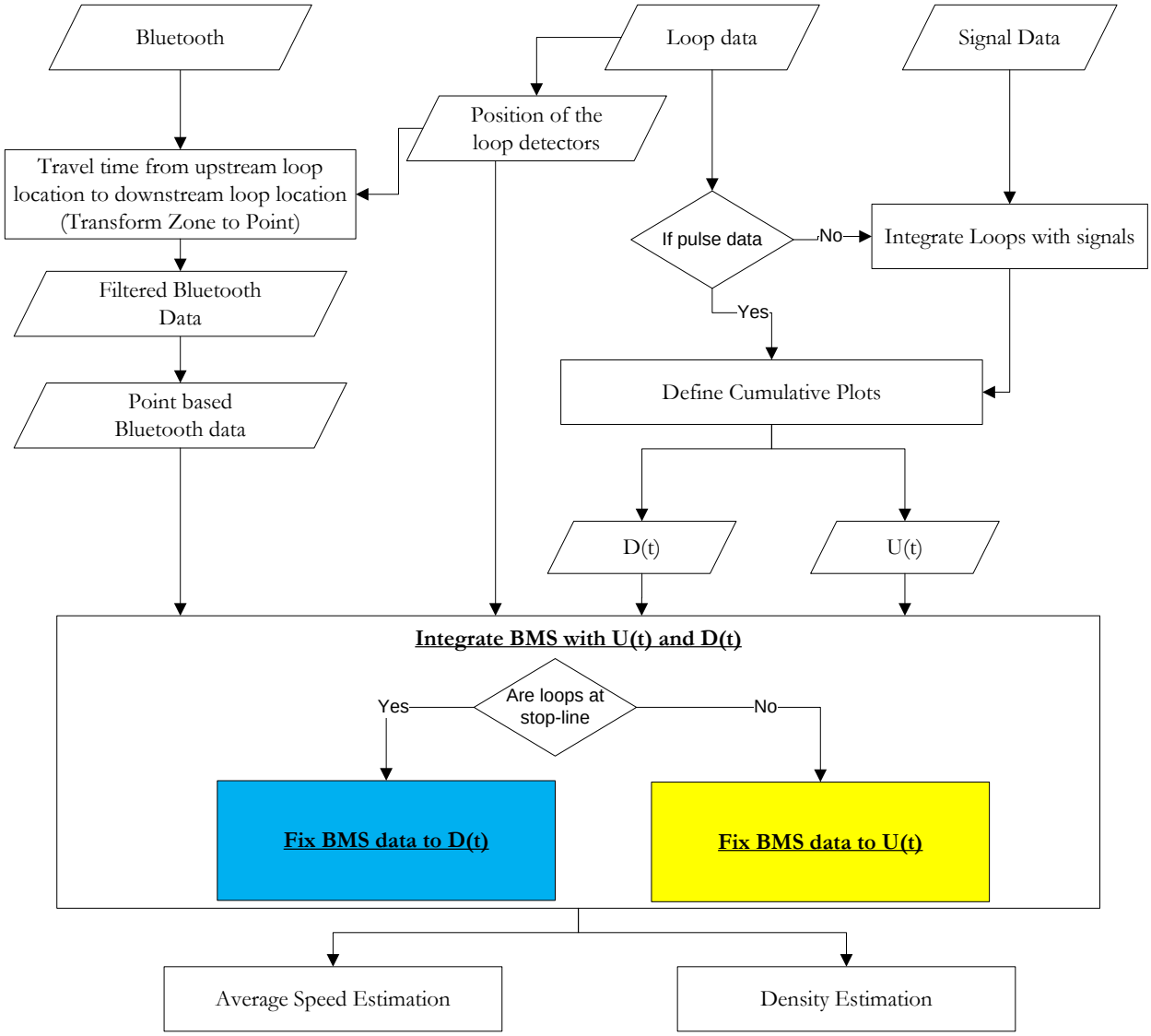

Figure 7: Architecture for the traffic state estimation model

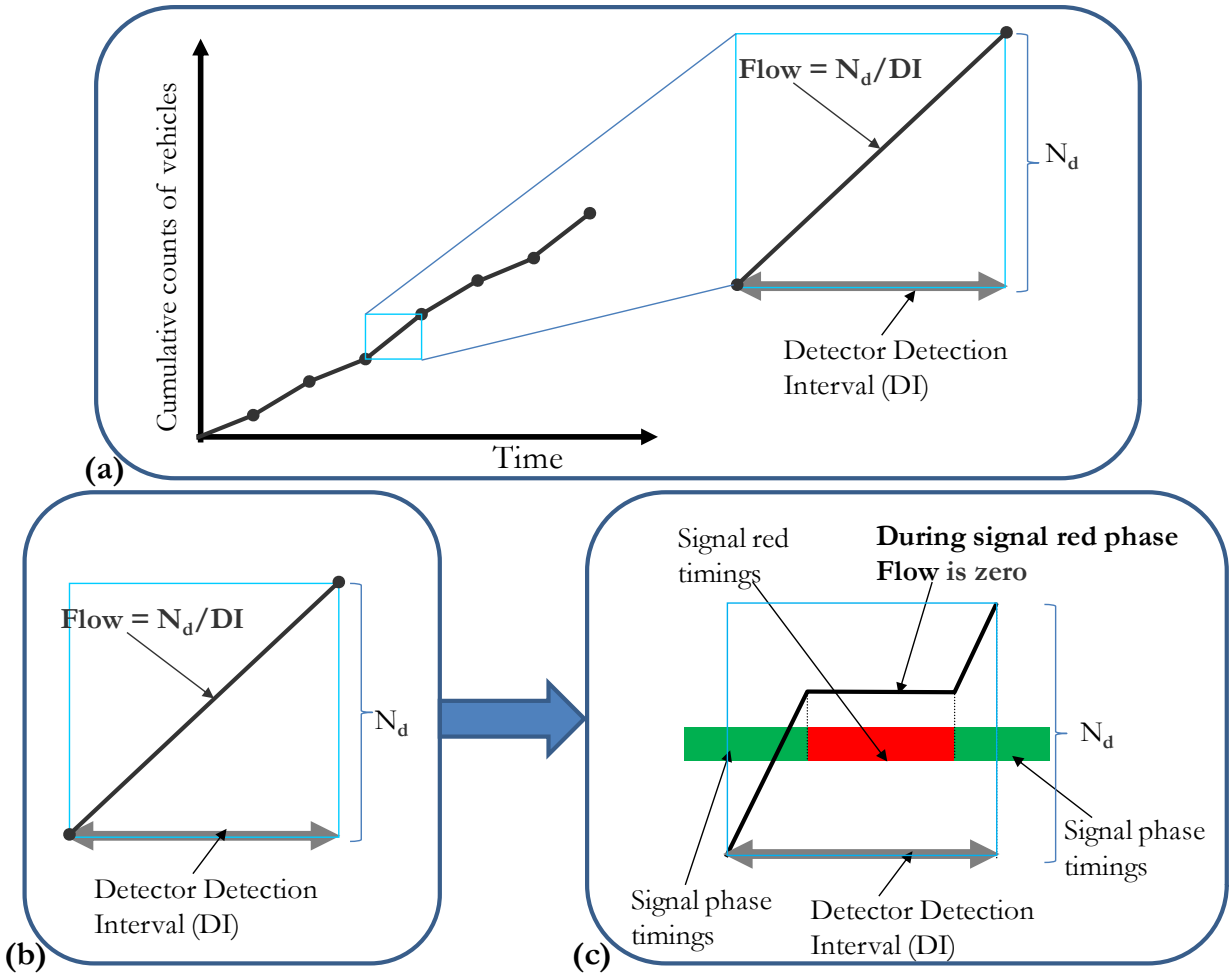

Figure 8: Conceptual representation of the estimation of cumulative plot from loops and signal timings: a) illustrates a time series of cumulative counts, with each interval of DI; b) Illustrates integration of signal timings with detector data to capture the fluctuations in the cumulative plots due to signals.

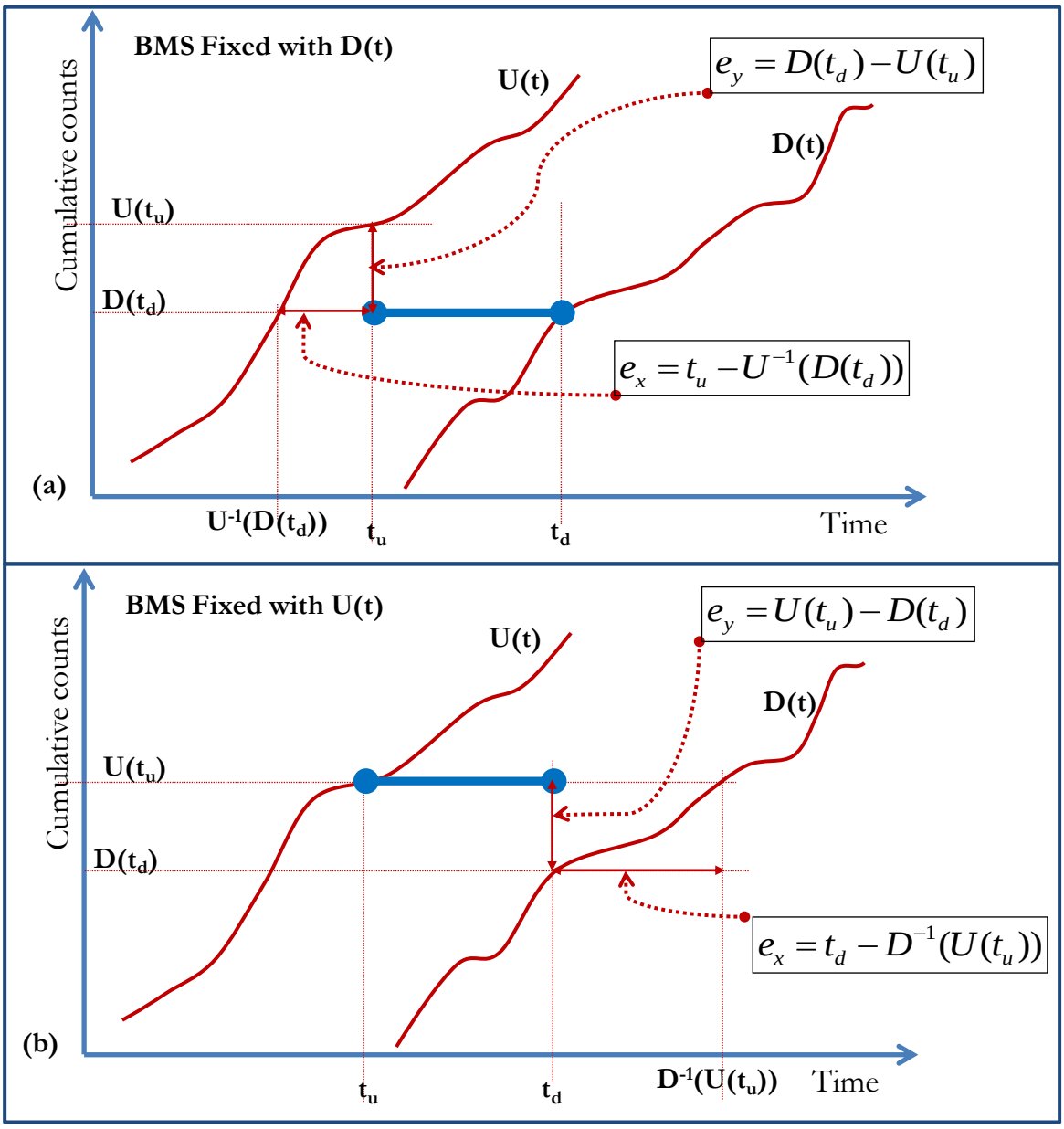

Figure 9: Illustrative example of fixing a BMS data with cumulative plots: a) BMS fixed with D(t); b) BMS fixed with U(t).

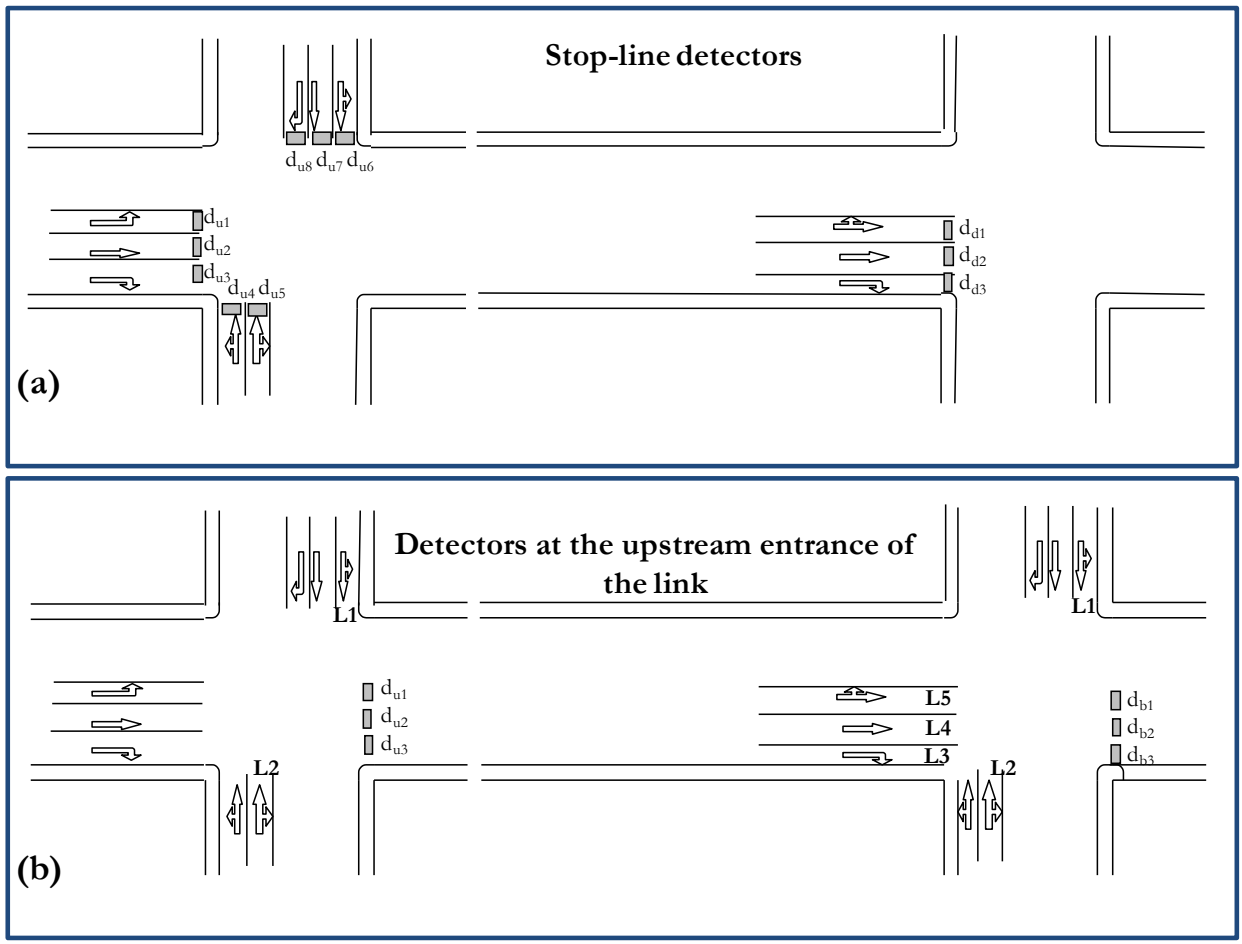

Figure 10: Examples of detector configurations a) at stop-line and b) at upstream entrance of the link

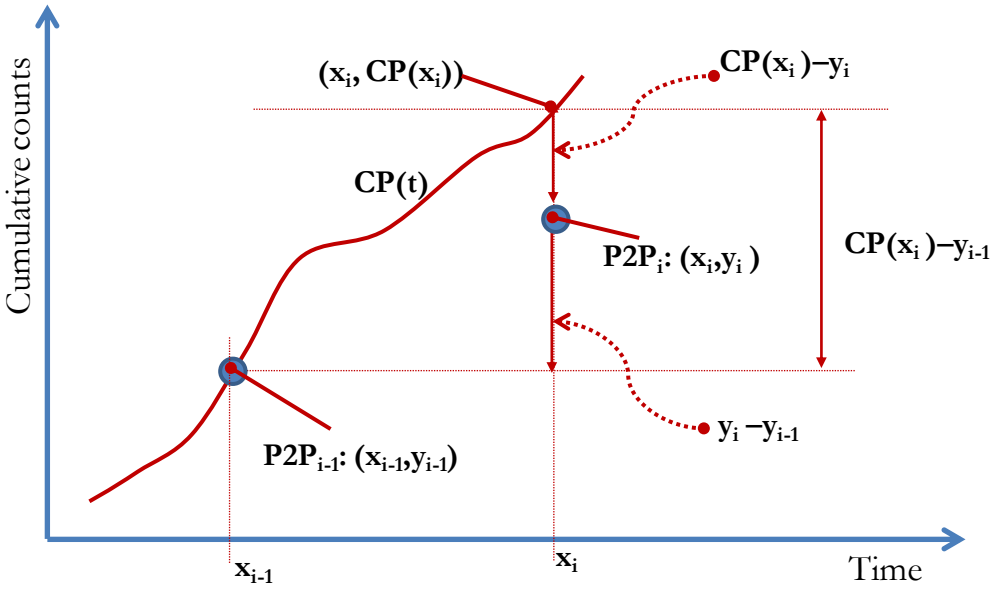

Figure 11: Illustration of variables used to model the vertical scaling

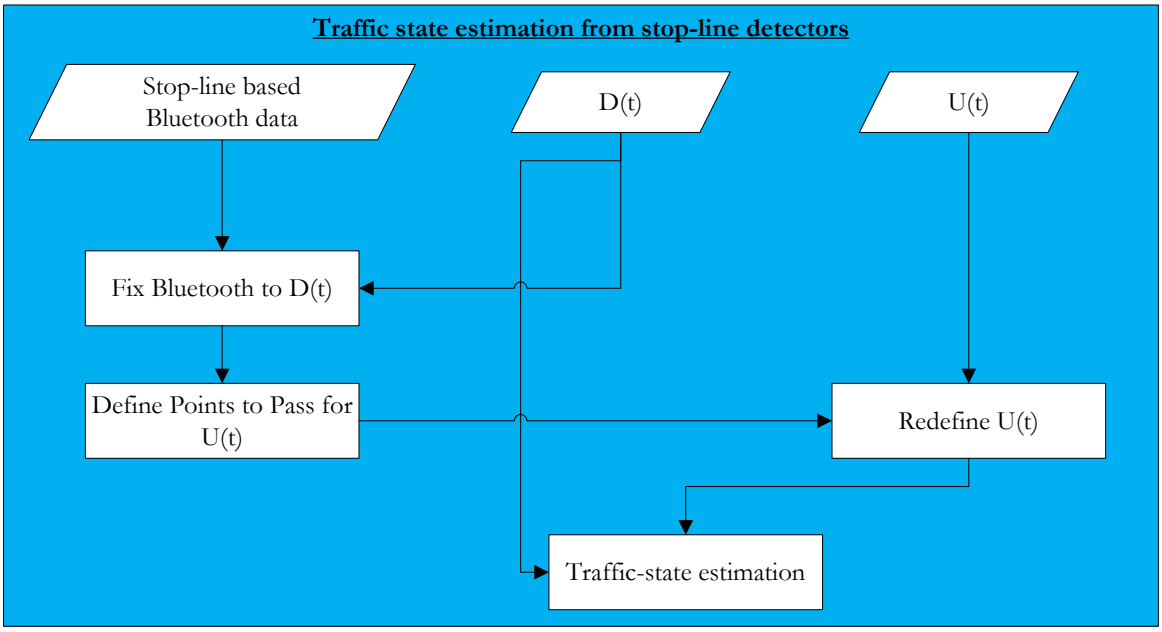

Figure 12: Flowchart for traffic state estimation from stop-line detectors

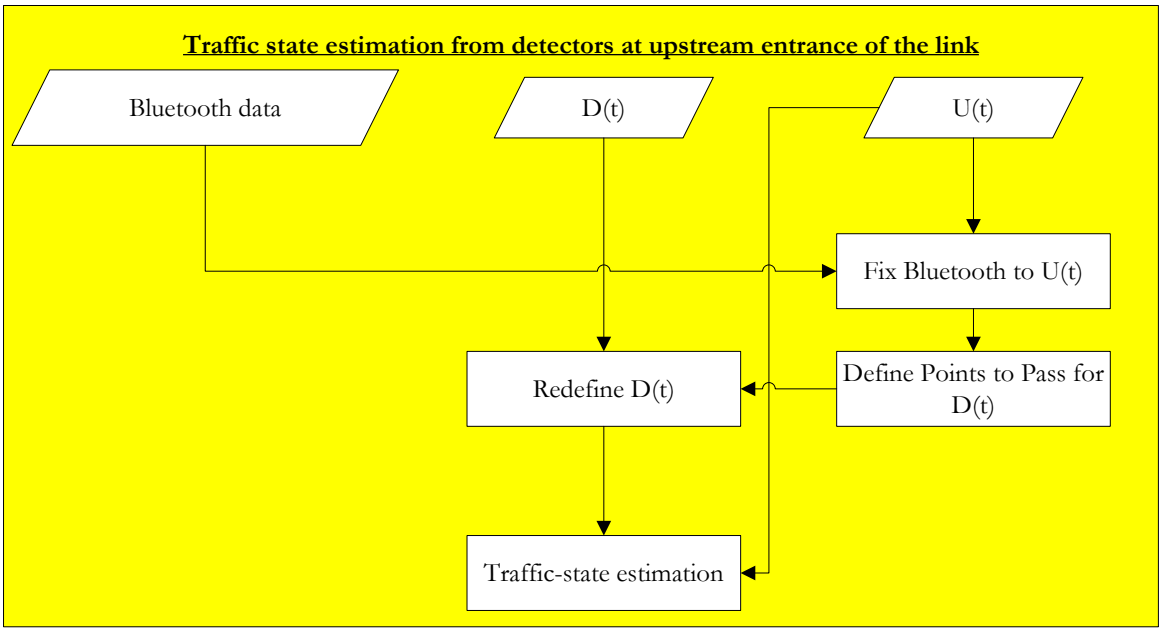

Figure 13: Flowchart for traffic state estimation from detectors that are at upstream entrance of the links

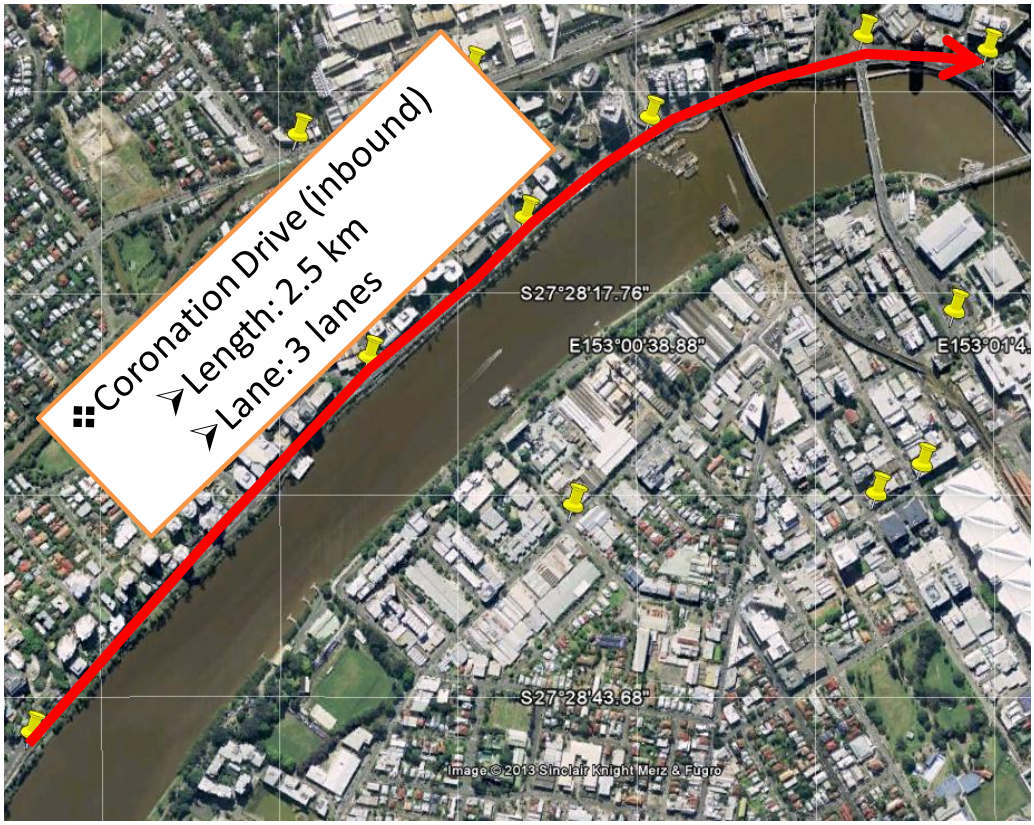

Figure 14: Study site, Coronation Drive (inbound traffic), Brisbane

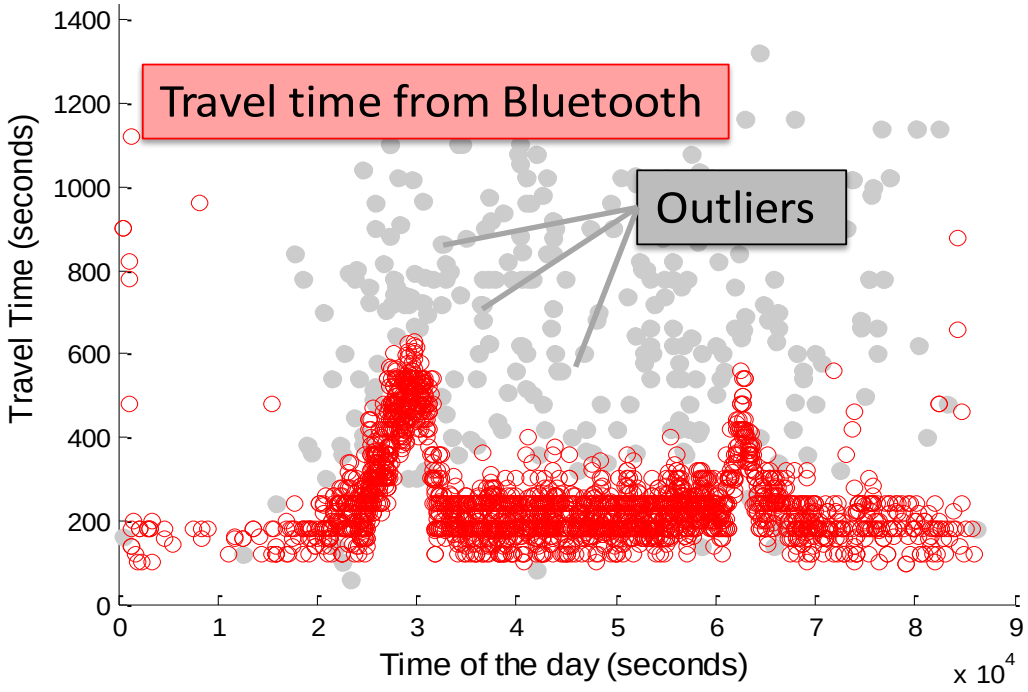

Figure 15: Travel time profiles along BMS site

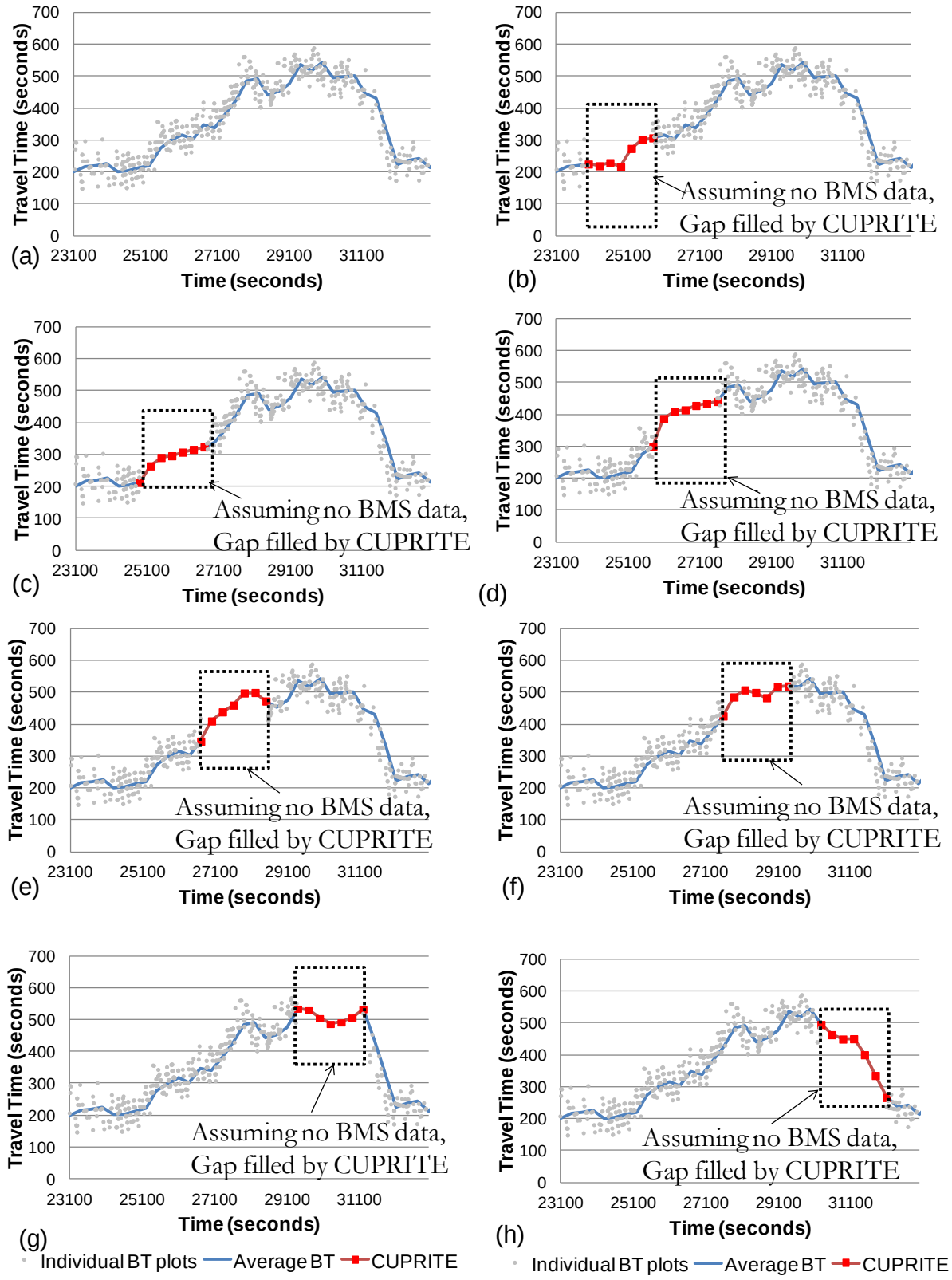

Figure 16: a) Bluetooth travel time points and average for 5 minutes; b to h: Results from the application of CUPRITE on 22nd October 2012 with 30 minutes of missing Bluetooth data.

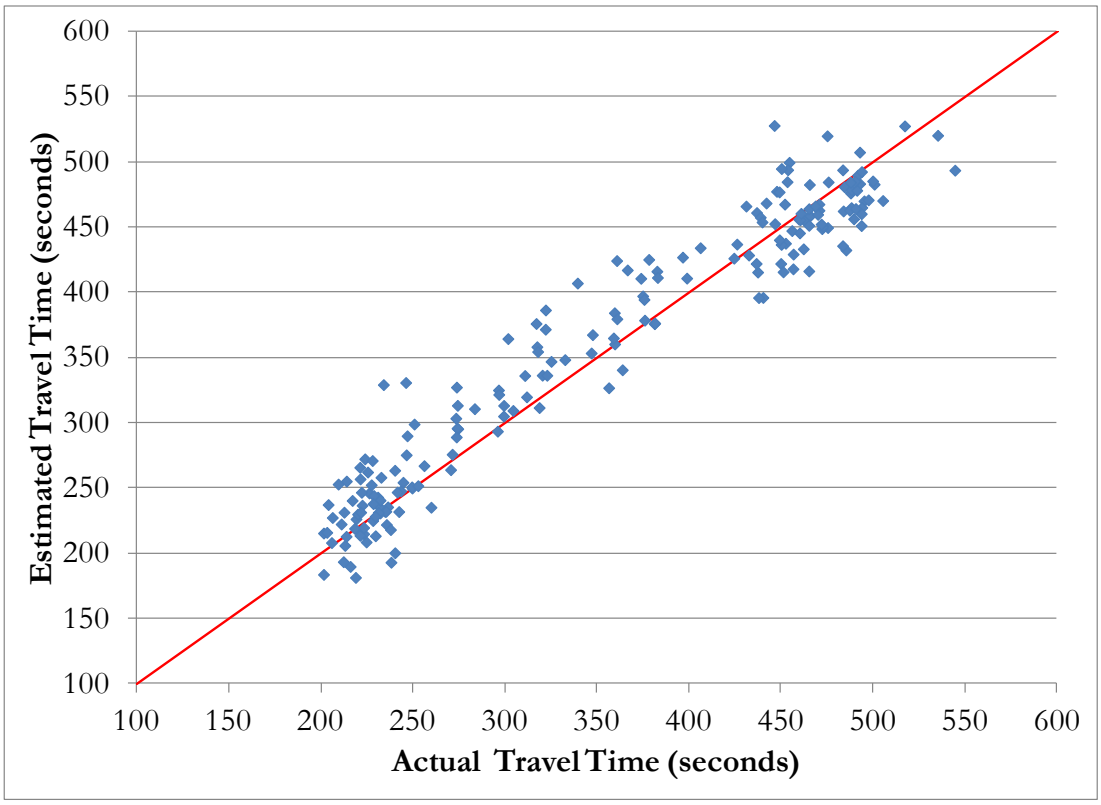

Figure 17: CUPRITE travel time verus actual travel time (from BMS) for different validation days

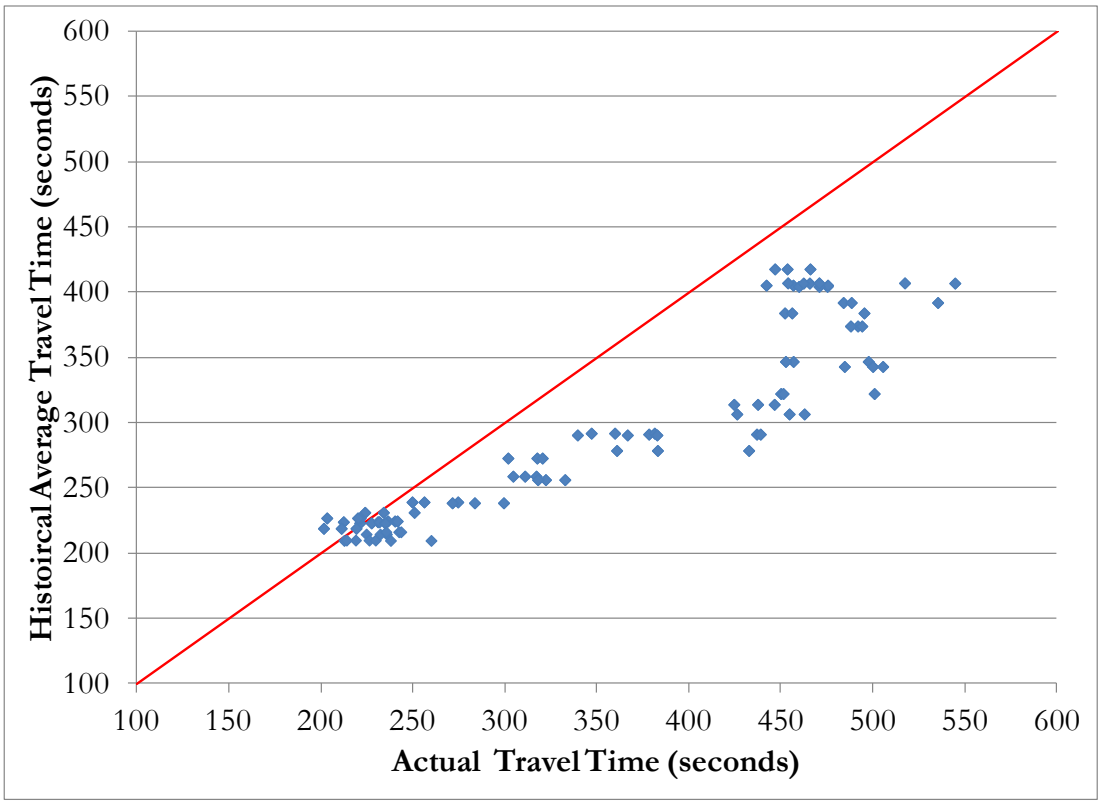

Figure 18: Historical average travel time verus actual travel time (from BMS) for different validation days

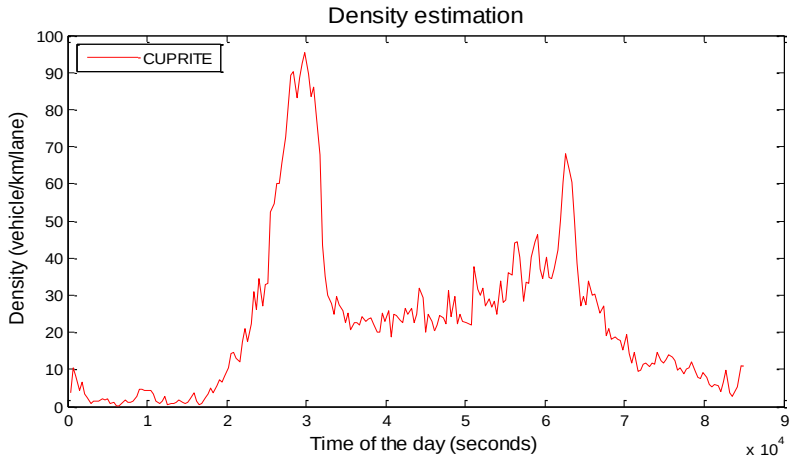

Figure 19: Density estimation using CUPRITE using real data from Coronation drive, Brisbane

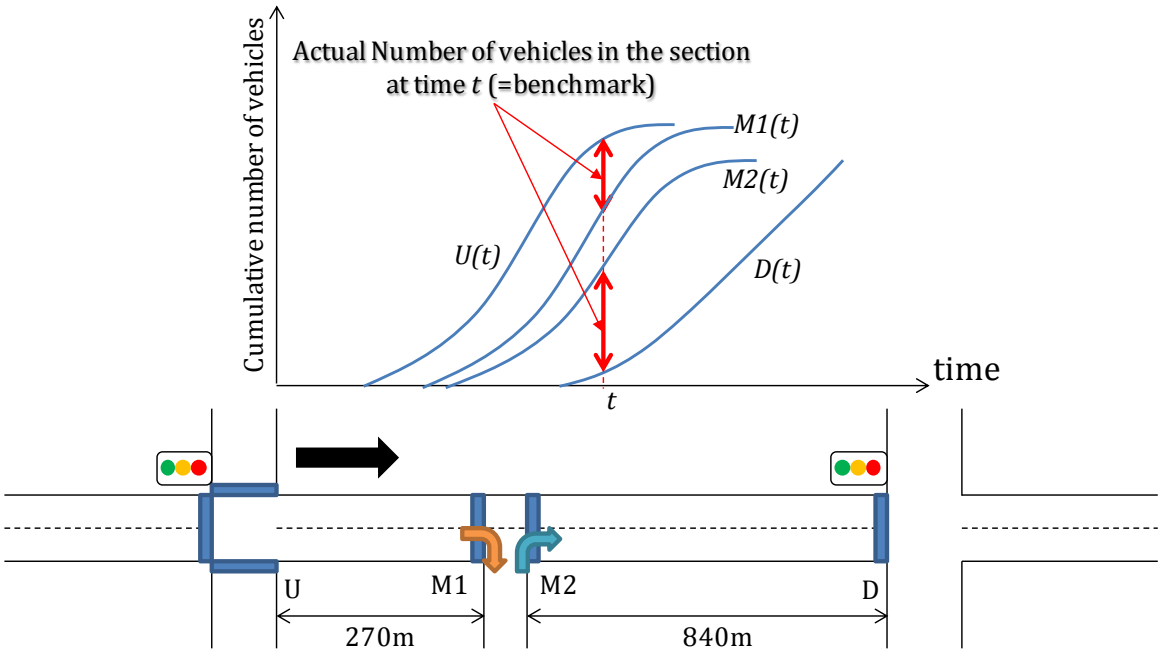

Figure 20: Illustration of the study site and the benchmark density estimation

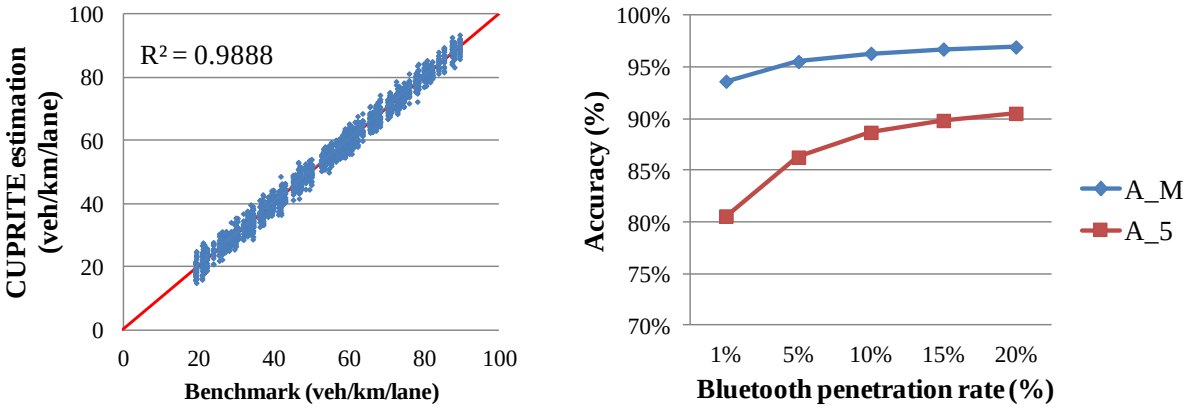

Figure 21: CUPRITE density estimation vs Benchmark for mid-link sink cases with Bluetooth measurement errors

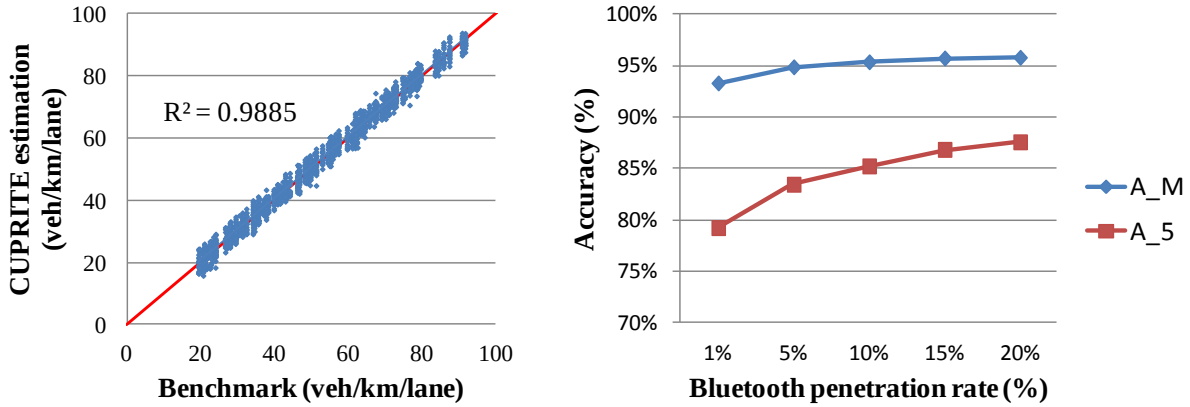

Figure 22: CUPRITE density estimation vs Benchmark for mid-link source cases with Bluetooth measurement errors

Table 1: Sample data fields from a BMS

Record number	MAC-ID (hexadecimal or encrypted)	Time first recorded (τ_{fr}) (hr:mm:ss)	Duration (d_R) (seconds)
1	F8:5F:2A:7A:8B:EA	17:30:32	10
2	F4:8E:09:40:B7:D1	17:30:12	20
3	E0:CA:94:E7:7A:38	17:30:42	10