Kieu, Le Minh, Bhaskar, Ashish, & Chung, Edward
(2015)
Empirical modelling of the relationship between bus and car speeds on signalised urban networks.
*Transportation Planning and Technology*, *38*(4), pp. 465-482.



*https://doi.org/10.1080/03081060.2015.1026104*

# Empirical modelling of the relationship between bus and car speeds on signalised urban networks


- **Journal name**: Journal of Transport Planning and Technology


- **Corresponding author**:

Le Minh Kieu

*Smart Transport Research Centre, Science and Engineering Faculty, School of Civil Engineering and Built Environment, Queensland University of Technology, Australia*
Email: le.kieu@qut.edu.au

- **Co-authors**

Ashish Bhaskar

*Smart Transport Research Centre, Science and Engineering Faculty, School of Civil Engineering and Built Environment, Queensland University of Technology, Australia*
Email: ashish.bhaskar@qut.edu.au

Edward Chung
*Smart Transport Research Centre, Science and Engineering Faculty, School of Civil Engineering and Built Environment, Queensland University of Technology, Australia*
Email: edward.chung@qut.edu.au



- **Acknowledgements**

The data used for this research was provided by the Brisbane City Council. This research contributes to the Integrated Traveller Information Research Domain at the Queensland University of Technology- Smart Transport Research Centre. The conclusions of this paper reflect the understanding of the authors, who are responsible for the accuracy of the data.

# Empirical modelling of the relationship between bus and car speeds on signalised urban networks

Vehicle speed is an important attribute for the utility of a transport mode. The speed relationship between multiple modes of transport is of interest to the traffic planners and operators. This paper quantifies the relationship between bus speed and average car speed by integrating Bluetooth data and Transit Signal Priority data from the urban network in Brisbane, Australia. The method proposed in this paper is the first of its kind to relate bus speed and average car speed by integrating multi-source traffic data in a corridor-based method. Three transferable regression models relating not-in-service bus; in-service bus during peak; and in-service bus during off peak periods with average car are proposed. The models are cross-validated and the interrelationships are significant.

Keywords: Bus and Car relationship, Bluetooth, Transit Signal Priority, Multimodal modelling, signalised urban networks

## 1 Introduction

Intelligent Transport Systems collect various data for real-time monitoring management and the control of road networks. For instance, Bluetooth is used in live reporting of traffic, and Transit Signal Priority (TSP) is used to give priority to buses, etc. The availability of these data sources provides an excellent opportunity to develop models for other applications, such as transport planning.

Traffic demand modelling plays a vital part in strategic transport planning and policy making. Most of the approaches are predicated upon the idea of the classical four-step demand modelling. Modal split is the third stage in the four steps, which models the mode choice of passengers. Passengers are visualized as choosing a mode that maximizes their utility. The utility is defined as a function of various attributes such as access time, egress time, fare cost, waiting time, travel time, etc. Among them, travel time/speed is one of the most important attributes. Unlike most of the other ones, it needs on-going monitoring for collecting the data.

However, both the initial investment and maintenance of equipment for monitoring of travel time/speed are very costly. For this reason, the traffic data for all modes of transport are not always available. Conversely, travel time or space mean that the speed of different modes of transport is needed for multimodal modelling, strategic planning, management and control of transport networks.

Understanding the relationship between multimodal forms of transport could provide an alternative way to overcome the issue of the lack of traffic data. A quantified relationship between bus and car travel time/speed could be useful for estimating the generalised cost of each mode using the knowledge of the other mode. The generalised cost could be used in the traditional four-step demand modelling. However, it is challenging due to the differences in their operational and mechanical behaviours. Buses have to stop for boarding/alighting of passengers. In-service (servicing) buses also tend to use the left most lanes (for left hand driving) on the roads and their mechanical characteristics are different from cars. Those differences between buses and cars increase the complexities of multimodal speed estimation and complicate the relationship between bus and car speed (or travel time).

Travel time has long been topic of research but most of the research is limited to average car travel time estimation only (Bhaskar, Chung, and Dumont 2011, 2010, 2012; Liu et al. 2010). The literature on a transferable multimodal travel time relationship for large scale application is limited. Most of the literature focuses on finding the impact of different factors on the transit performance (Levinson 1983; McKnight et al. 2004) or using buses as traffic probes in urban areas (Bertini and Tantiyanugulchai 2004; Cathey and Dailey 2002; Chakroborty and Kikuchi 2004; El Esawey and Sayed 2012) to estimate the car travel time or speed. Levinson (1983) analysed the traffic data in a few U.S cities and revealed that average car speeds are 1.4 to 1.6 times faster than average bus speed. McKnight (2004) proposed an equation between bus travel time per mile and car travel time per mile. Some other studies from the US aimed to utilize buses as traffic probes for identifying traffic conditions and estimate car travel time/speed. Examples of them are the projects in King County in WA (Cathey and Dailey 2002), TriMet in Oregon (Bertini and Tantiyanugulchai 2004) and Delaware Department of Transportation (Chakroborty and Kikuchi 2004). These studies concluded that buses could be used as traffic probes effectively in urban arterials.

In an effort to improve the understanding of the relationship, and to utilise it in large scale traffic demand modelling, this paper proposes the following:

- Most of the past studies on bus-car travel relationship use floating cars and Automatic Vehicle Location (AVL) data for car and bus, respectively. They followed a bus-route based study, in which the floating cars travelled along the same route with the studied bus and compare their travel times or speeds (Chakroborty and Kikuchi 2004; Bertini and Tantiyanugulchai 2004; Cathey and Dailey 2002). However, the AVL and floating car data sources are expensive, and the bus-route based approach limits the bus sample size to one bus route and selected cars (floating cars) only. Furthermore, the studied dataset contained only selected samples but not all the vehicles over the studied area. Therefore, their conclusions are only valid for the studied routes, with specific buses and cars and are not transferable over large scale application. A corridor based method which considers all buses and cars along a study corridor is required for traffic demand modelling. Some more accessible data sources also need to be explored. This paper proposes a method for large scale analysis of bus and car space mean speed relationships in urban areas.

- To the best of the authors' knowledge, there is no study in the literature utilising the data of buses which are not-in-service. The not-in-service buses are the ones which do not stop along the corridor. They are travelling to the depots to start new services after finishing other journeys or travelling from the bus garages. AVL datasets used in many past studies do not contain not-in-service buses' data since they are not on duty. Due to the unavailability of data, this kind of bus does not receive much attention in the existing literature. This paper proposes a method for relating the not-in-service buses with cars in urban areas.

The research aims to model the relationship between bus and car speed by integration of multi-source historical ITS data. The study does not aim to estimate real-time bus speed using car speed and *vice versa*, which should take into account the stochasticity of various factors such as signal delay, weather etc. However, different factors that affect the speeds of both bus and car on urban arterials are modelled and their impacts are explored.

The paper is structured as follows. First, the research methodology is discussed, followed by the description of the study site and available data. Thereafter, the procedure for cleaning the data is discussed and the proposed models are developed. Finally, the model is cross-validated and the paper is concluded.

## 2 Methodology

Modelling speed is equivalent to modelling travel time normalised with distance. As data from different sites are used, speed is chosen as the direct parameter for modelling in this paper. A generic equation between bus and car speed is developed for an urban corridor where buses and cars share the same road.

The car travel time is obtained from time synchronised Bluetooth scanners located at the intersections. Bluetooth technology has been described in literature as an effective method for travel time estimation (Kieu, Bhaskar, and Chung 2012; Tsubota et al. 2011; Hamilton et al. 2012). Here, travel time, $TT(m,u,d)$, of a Bluetooth MAC-ID , $m$, observation at upstream, $u$, and downstream, $d$, intersections is defining as the difference of the time when the device is observed at downstream, $T(m,d)$, and upstream, $T(m,u)$, intersections.

$$TT(m,u,d) = T(m,d) - T(m,u) \tag{1}$$

The bus travel time is obtained in a similar way by using TSP data, where instead of MAC-ID we match the unique RFID of the bus. TSP detectors in Brisbane are RFID scanners which identify unique Bus Vehicle Identification (VID) of every bus passing the intersection. By using the same matching method as Bluetooth travel time estimation, the bus travel time from upstream to downstream intersection of the study corridor could be calculated.

All buses and Bluetooth-enabled vehicles passing the entrance and exit of the study corridor are identified and stored in the TSP database (for buses) and Bluetooth database (for cars), respectively. Upstream and downstream of the corridor are signalised intersections and a corridor is the shortest route between these two intersections. The space mean speed of individual vehicle $\bar{v}_{m,u,d}$ (bus or car) is defined as the ratio of corridor length ($d_{u,d}$) and the vehicle travel time along the corridor.

$$\bar{v}_{m,u,d} = \frac{d_{u,d}}{TT(m,u,d)} \tag{2}$$

The studied buses are classified into in-service buses and not-in-service buses. Only the buses which travel the entire corridor by the shortest route between upstream and downstream intersection are considered in the analysis. Inbound traffic during typical working days (weekdays excluding the public holidays and school holidays) from 7 AM to 9 PM are considered. An example of daily travel speed profiles on Coronation Drive is showed in Figure 1.

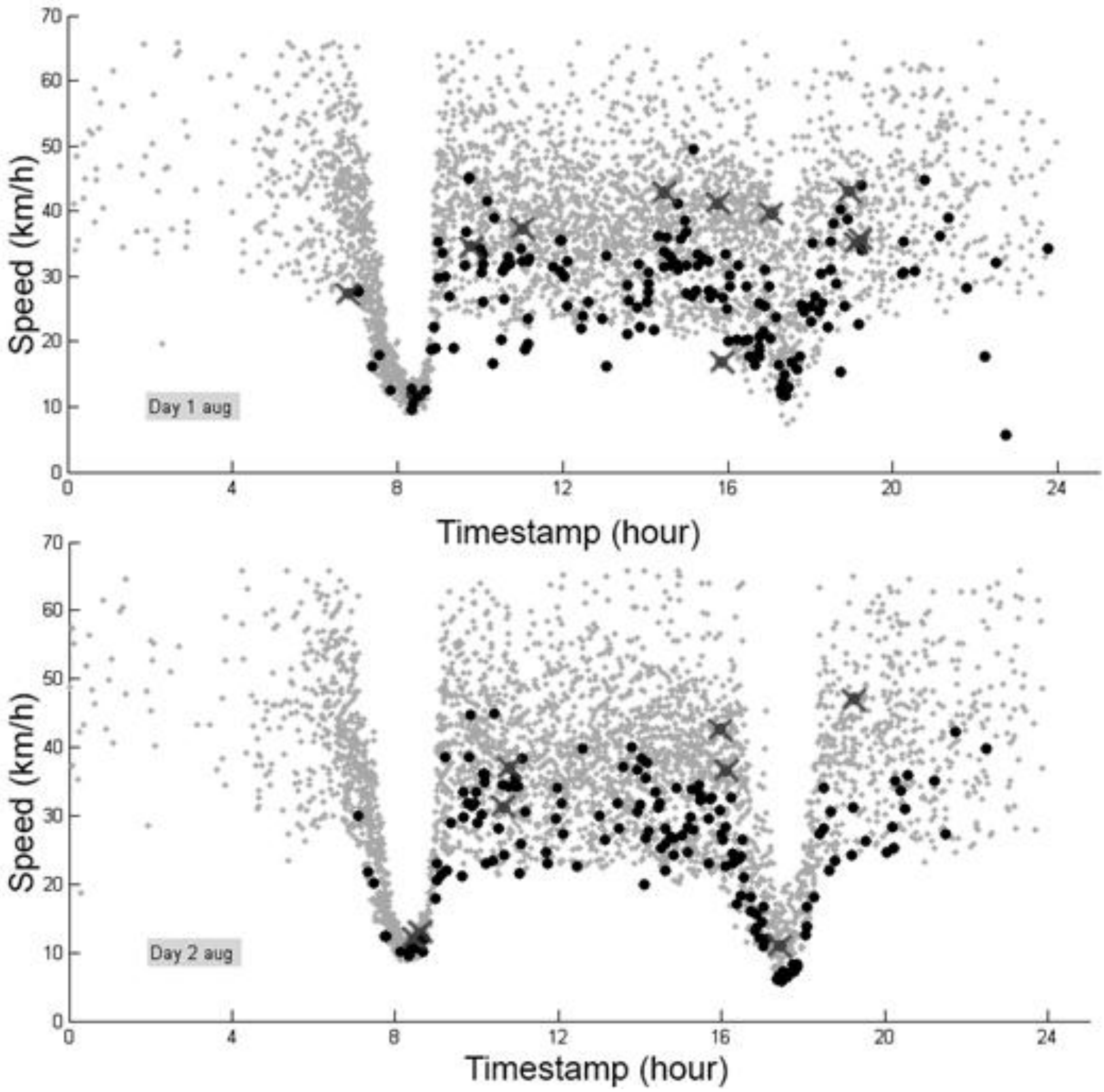


Figure 1 Example of travel time speed profiles for first week of August 2011 in Coronation Drive, Brisbane, Australia (*grey dots*: individual space mean car speeds from Bluetooth; *black dots*: individual in-service space mean bus speeds from VID; and *black cross*: individual not-in-service space mean bus speeds from VID).

The study period is classified into peak periods (7-9 AM and 4-9 PM) and mid-day off-peak periods (9AM-4PM) by examining the daily historical travel speed profiles of all the sites and considering the peak/off-peak ticketing classification in Brisbane (Translink 2012b).

The bus-car speed relationships during peak and mid-day off-peak periods are studied separately because different patterns are observed during the two periods (*see* Figure 1). The in-service bus speed samples stay within the lowest car speed samples during peak while there are larger spreads of bus samples during the mid-day off-peak period. Moreover, it also implicitly considers the difference in traffic demand between peak and midday off-peak periods.

In contrast, for inspecting not-in-service bus and car speed profiles, no significant difference in pattern is observed between peak and mid-day off peak periods. Hence, the relationship between not-in-service bus and car speeds is studied for the whole study period between 7AM to 9 PM.

The framework of this research is illustrated in the Figure 2. The individual vehicle car and bus travel times are estimated from Bluetooth and TSP data, respectively. Individual vehicle travel times are cleansed by applying filters (Section 4) and travel time are transformed into individual vehicle space mean speed by respective considering of the corridor length. Thereafter individual vehicle speeds from all the corridors are integrated (Section 5) and data is prepared for model development (Section 6) and cross-validation (Section 7).

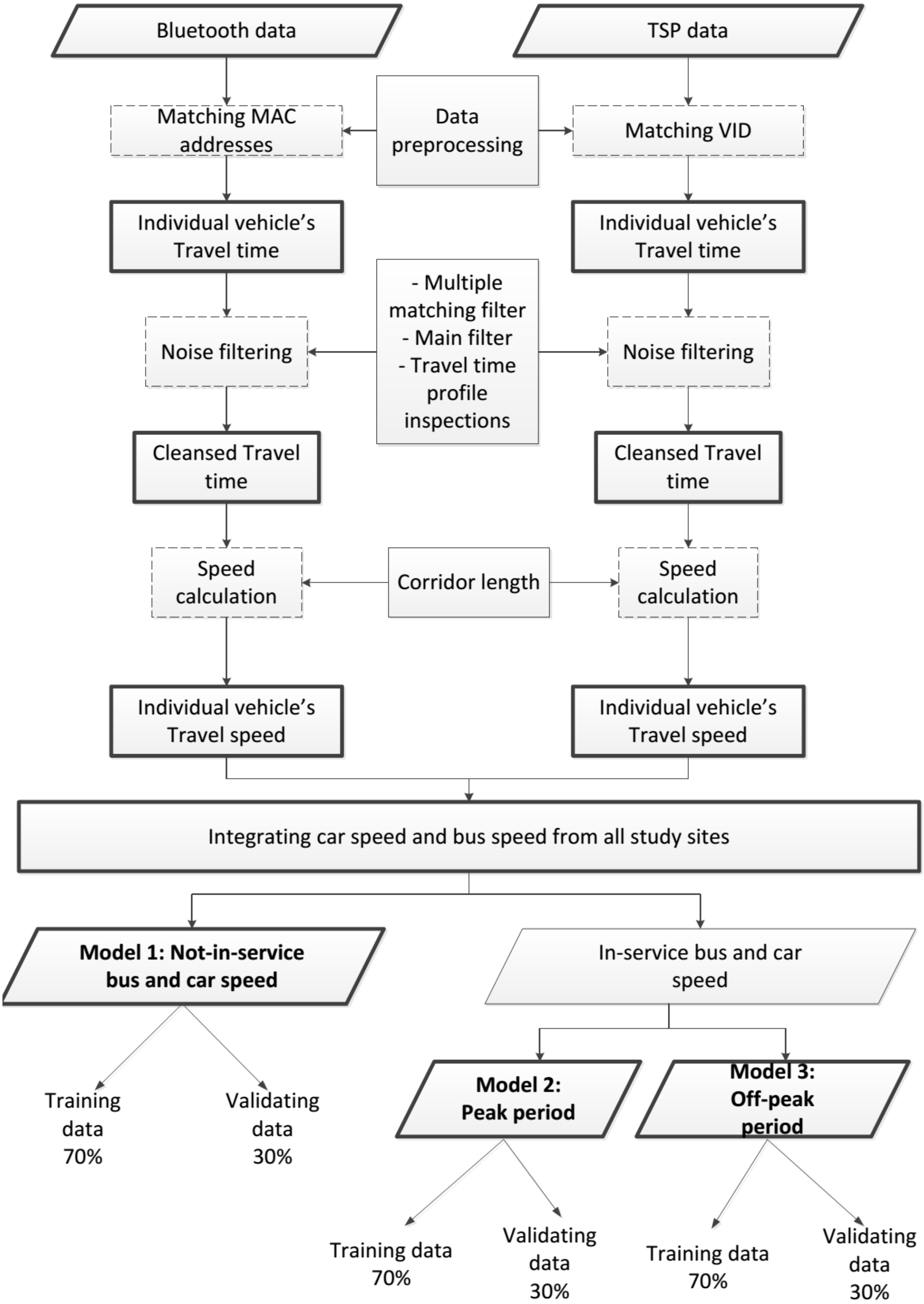


Figure 2 Research Framework

## 3 Study site

The Brisbane City Council has installed Bluetooth scanners at major intersections and TSP detectors at intersections along some busy bus routes in Brisbane, for monitoring traffic and transit operations. Four months (from August to November 2011) of Bluetooth and TSP data of Brisbane City have been used in this study.

We choose three sites for analysis: Coronation Drive, Logan Road and Wynnum Road (*see* Figure 3). The criteria for choosing the sites are: data available; data diversity; and buses and cars share the same road. Traffic volumes on these roads vary from highly congested (Coronation Drive), to only congested during morning peak (Wynnum Road) and uncongested traffic (Logan Road). The speed limit is 60 km/h on all the sites. The details for length, number of signalised intersection, number of bus stops, and corridor length etc. are provided in Table 1.

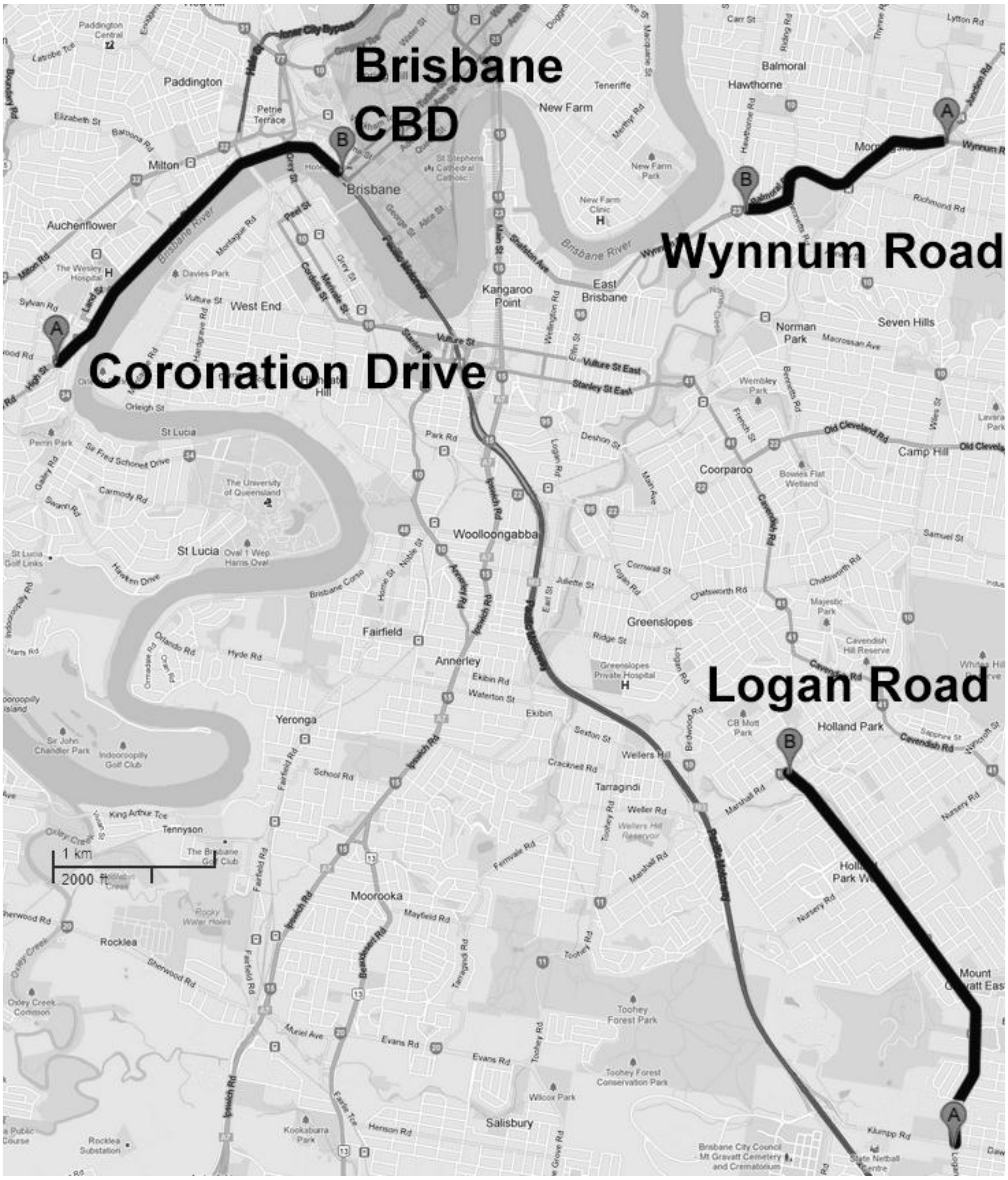


Figure 3 Case study sites map

Table 1 Case study sites description

| Site | Period (timestamp in hour) | | | # signal. intersecti-on | Max. # bus stop | Length (km) | Data Avai. | Servicing bus routes |
|---|---|---|---|---|---|---|---|---|
| | 7-9 | 9-16 | 16-21 | | | | | |
| Corronation Drive | **Peak** | Off-peak | **Peak** | 8 | 4 | 3.3 | Aug to Nov | 411,412,415,417,433,445,775 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Logan Road | Off-peak | Off-peak | Off-peak | 12 | 12 | 4.0 | Aug to Nov | 175,176,806,870 |
| Wynnum Road | **Peak** | Off-peak | Off-peak | 6 | 8 | 2.2 | Aug to Nov | 227 |

# 4 Data cleansing

Various kinds of noise could be observed in the Bluetooth and TSP dataset. These noises could lead to overly long travel time along the studied corridor, and should be removed from our analysis.

Travel time, instead of speed has been chosen as the criterion for removing the outliers from our dataset, because it is the direct product from our MAC address (from Bluetooth devices) and VID (from buses) matching algorithms.

## *4.1 Cleaning travel time from Bluetooth*

Travel time between two Bluetooth scanner stations can be directly obtained by utilising equation (1). The matched travel time data do contain noise due to reasons such as:

- *Unknown mode*: Obtained travel time is for the Bluetooth device transported by a traveller utilising any mode of transport. Different modes have different travel times depending on its operational and behavioural characteristics. If one is interested in car travel time, then the presence of pedestrian or bicycle can result in unrealistic high travel time values and vice-versa.
- *No information for the vehicle's travel along the corridor*: The travel time estimate is from the data available at upstream and downstream of the corridor. The actual travel pattern of the vehicle along the corridor is unknown. A vehicle can rest along the corridor or can take a different route with a significantly different travel time to that of the assumed route.
- *Multiple matches*: Especially on arterial network, a device can be observed at a Bluetooth scanner location (zone) and then it might take a detour, return to the same zone, and thereafter travel to the next zone. In such a situation, the device can be observed twice in the first *zone* and only once in the second *zone*,

resulting in two travel time values. Similarly, other combinations of multiple matches can occur resulting is noise.

- *Missed observation*: A Bluetooth device has a probability to be discovered at a *zone* and not all devices passing the *zone* are discovered. For instance, say a device travels twice between zones *u* and *d*. During its first trips the device was observed at *u* at time *tu1* whereas, it was missed at *d*. During its second trip, it was observed at *d* at time *td2* but missed at *u*. Such observations will result in noisy travel time from *u* to *d* as (*td2-tu1*). Similarly other combinations of observations can result in inaccurate travel time values.

Here, the following filters are applied on the time series of the individual travel time records obtained from equation (1).

#### *4.1.1 Multiple Matching filter*

Here, we first extract all the travel time points observed during a time window defined as $[t\text{-}\Delta t,\ t+\Delta t]$. Where $\Delta t$ = 5 minutes and $t$ = 0:05 am, 0:06 am, etc. to 11:54 pm, 11:55 pm. Then, we look at the MAC ID's of the travel time points observed during the given time window. The MAC ID's observed more than once are identified. The observations with similar MAC ID's are clustered. And for each cluster, the non-minimum travel time data points is identified as noise and is filtered out. By doing so, the errors related to aforementioned *multiple matches* are reduced.

#### *4.1.2 MAD outlier filter*

The previous filter logically cleans the time series of travel by removing the repetitions of a vehicle travel time reported more than once in each 10 minutes time window. The remaining noise related to unrealistic travel time values are removed by a statistical filter, termed as Median Absolute Deviation (MAD) filter or Hampel identifier (Pearson 2002) is applied. This filter removes outliers by comparing them with neighbor travel time observations within 10 minutes interval. For each minute, a window of 5 minutes before and 5 minutes after is considered, and this window is moved from the first to the last minute of the day. The outliers are identified if they are larger than the Upper Bound Value (UBV), or lower than a Lower Bound Value (LBV) of the current window as defined below.

$$UBV = median + \hat{\sigma} f \tag{3}$$

$$LBV = median - \hat{\sigma} f \tag{4}$$

Where $\hat{\sigma}$ is the standard deviation from the *MAD*, in which a normally distributed data can be approximated as $\hat{\sigma} = 1.4826 \times \mathrm{MAD}$

$$MAD = median(|X_i - median(X_j)|) \tag{5}$$

The value of $\hat{\sigma} f$ defines the scatter of data, where *f* is a scale factor which varies on a case by case basis. If *f* is small, the gap between *UBV* and *LBV* to the median value is small, and *vice versa*. The value of *f* has been suggested by some authors to be from 1 to 5 (Davies and Gather 1993; Pearson 2002).

For the current analysis, *f* = 1,2 and 3 is applied and the results are presented in Figure 4, where the green dots represents the noise identified by the application of the filters and black dots are the travel time points which are not identified as noise by the aforementioned filters. As seen in the value *f=1* gives us the most confidence in the travel time profile, but can consider valid travel time points as noise. On the other hand, with *f*=*3*, we have lower confidence in the travel time profile with few noisy points considered as valid. Considering these, *f=2* is selected for filtering.

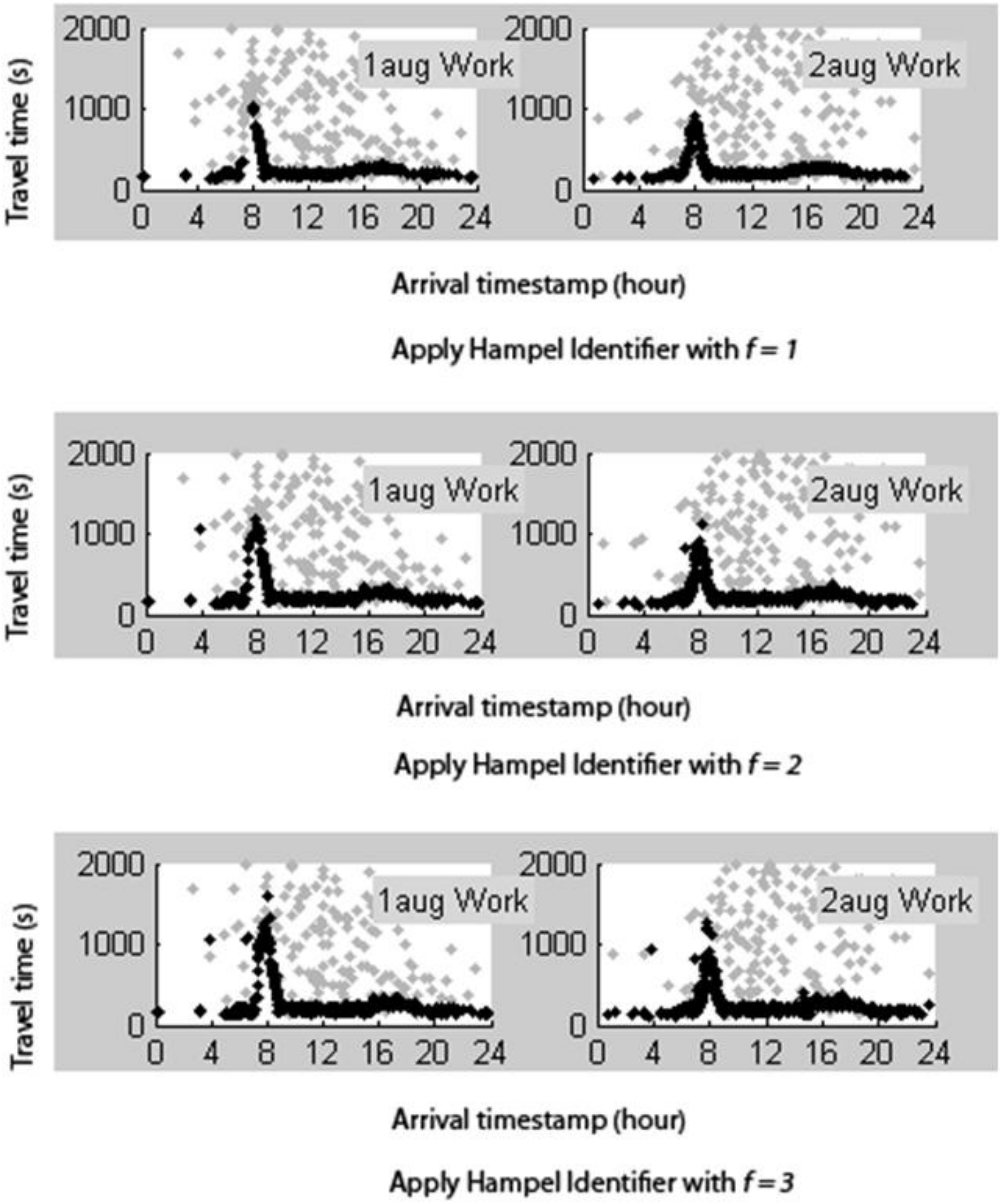


Figure 4 Examples of different values of f for the outlier filter applied on Wynnum Road in the first week of August 2011.

### *4.2 Cleaning travel time from VID*

TSP detectors are installed on every lane at the same position. This can result in noise when a bus is detected by TSP detectors on two adjacent lanes. Due to this an individual bus can have duplicate travel time values. Here the *multiple matching* filter (see sub section 4.1.1) is applied on the time series of the bus travel time and noise is reduced.

A bus has a fixed route and hence it does not have noise related to unrealistic travel time along the corridor due to a change of route. We do not observe large numbers of buses in 10 minutes of the time window, hence statistically the bus sample size is not large enough to apply MAD outlier filter. The travel time from VID is fairly accurate and represents the actual bus travel time.

## 5 Integrating bus and car data and preparing data for model development and cross-validation

The cleansed travel time is transformed into speed using the respective corridor length. This provides a time series of individual vehicle car speeds ($v_{car,i}(\zeta_u)$) and bus speeds ($v_{bus,j}(\zeta_u)$) corresponding to the time when the vehicle was observed at upstream ($\zeta_u$). Here, *i* and *j* are number of cars and buses observed at upstream at time $\zeta_u$, respectively.

For each bus sample we define a time window of 10 minutes *[t-5 minutes, t+5 minutes]*. All the car travel time points observed at upstream within the time window and for the same day are extracted. And we define the average car speed $\overline{v_{car}(t)}$ as the average of these data points.

$$\overline{v_{car}(t)} = \frac{\sum_{i=1}^{n} v_{car,i}(\zeta_u)}{n} \quad where \quad \zeta_u \in [t-5, t+5] \tag{6}$$

This provides us with an average car speed for each bus speed data point from all the corridors and study time periods. This data is saved along with the respective infrastructure characteristics into one database with following fields:

- Time at upstream
- Bus Type (not-in-service; in-service)
- Bus speed
- Car speed
- Corridor ID
- Number of bus stops along the corridor
- Number of signalised intersections along the corridor
- Corridor length

The aforementioned, database is then segregated into following three categories:

- Not-in-service bus
- In-service bus during peak periods
- In-service bus during mid-day off-peak periods

Finally, for each segregated data: 70% of the data is randomly selected for the model development and the remaining 30% is kept aside for cross-validation of the proposed models (Section 7).

## 6 Model development

Here, the objective is to find a transferable formulation between bus and car speeds along the signalised urban networks. For this we chose multiple regression for the following reasons:

- It provides flexibility to use multiple variables and facilitates exploration of how a set of independent variables are associated with the dependent variable.
- It provides simple and efficient models with relatively limited datasets.
- The developed models have easily-interpretable and transferable formulations which are possible to use in large scale demand modelling.

Here, the bus speed ($V_{bus\text{-}NIS}$ *or* $V_{bus\text{-}IS\text{-}P}$ *or* $V_{bus\text{-}IS\text{-}OP}$ representing not-in-service, in-service bus during peak periods, and in-service bus during off peak periods, respectively) is defined as a dependent variable. The independent variables are chosen to capture the similarities and differences between buses and cars on urban arterials and it includes:

- Car speed ($V_{car}$) (km/hr)
- Number of signalised intersection ($n_s$)
- Corridor length ($l$) (km) and
- Number of bus stops ($n_b$)

Since the data has been collected where buses and cars are sharing the same road, the main differences between buses and cars are caused by cross-sections and bus stops. Buses as heavy-vehicles have different mechanical characteristics to cars in running, acceleration and deceleration. Therefore, space mean speeds along the corridor and route length have been chosen to take into account the difference in cruise speed between buses and cars. Number of signalised intersections has been chosen for considering the differences at cross roads. More importantly, at each stop in-service buses have to stop for dwelling and acceleration/deceleration (Chien  and Qin 2004). The number of bus stops is collected to explore the difference caused by stopping time for boarding/alighting of passengers. The traffic demand and flow are implicitly considered in the car speed variable and the separation of peak and off-peak models. Some other regional variables such as Area characteristics (Business, Residential), Number of lane, etc. are also useful in exploring the relationship between bus and car. However, due to the limitation of number of study sites (3 sites), they are not considered in this paper.

The regression assumptions are analysed to find any violations. The assumptions include normality, linearity, homoscedasticity and independence. First, the assumption of normality is tested by looking at the distribution of the dependent variable $V_{bus}$ and the normal probability plot of residuals. The evidence of linearity assumption could be observed at the plot of observed versus predicted values, which is a part of the standard regression output. The assumption of homoscedasticity is tested by looking at the graph of Regression Standardise Predicted Values (*x-axis*) versus Regression Standardised Residual (*y-axis*). If the residuals spread around the same on both sides of the average residual, we can conclude that the assumption of homoscedasticity is not violated. Collinearity statistical analyses are also performed. Variables are identified as highly correlated with other variables if their Variance Inflation Factor (VIF) are more than 4 or Tolerance values are less than 0.25 (these limits are set based on the rules of thumb described by O'brien (2007). These tests have been performed for each model and the final results of the 3 models proposed in this paper do not violate any aforementioned assumptions.

The following sub sections define the three proposed models.

### *6.1 Model 1: Not-in-service bus speed and car speed relationship*

Here, not-in-service bus and car speeds from 7 AM to 9PM are modelled using the aforementioned database. Table 2 provides the descriptive statistic of the data used. A backward stepwise regression has been performed to choose the independent variables. The procedure started with including all the variables defined in Table 2. Each variable is excluded until a set of variables which show the highest adjusted R square is found. The final model has $V_{car}$ and $n_s$ as the predictive variables. Among the two other variables, $l$ is highly correlated with $n_s$ and $n_b$ has no relation with $V_{bus\text{-}NIS}$. The final proposed model is as follows:

$$V_{\mathrm{bus-NIS}} = 1.1 \times V_{car} - 0.5 \times n_s \tag{7}$$

The Table 3 shows that the model on equation (7) shows promising goodness-of-fit with high R square (0.80). The collinearity statistical analysis also reveals that the Variance Inflation Factors (VIF) of the variables, and the Tolerance values, are both very close to 1, which means that there is no sign of multicollinearity in the model.

Table 2 Descriptive Statistics of Model 1

| | **Sample size** | **Min** | **Max** | **Mean** | **Median** | **Std. Dev.** |
|---|---|---|---|---|---|---|
| Car speed (*CarSpeed*) | 626 | 6.85 | 49.80 | 25.66 | 25.87 | 8.59 |
| Not-in-service bus speed ($BusSpeed_{NIS}$) | 626 | 7.60 | 49.48 | 25.45 | 25.62 | 8.27 |
| Signalized intersection (*SI*) | 626 | 6.00 | 12.00 | 9.14 | 8 | 0.60 |
| Link Length (*Length*) | 626 | 2.29 | 3.33 | 2.83 | 3 | 0.44 |
| Bus Stop (*BusStop*) | 626 | 4.00 | 12.00 | 8.15 | 8 | 2.47 |

Table 3 Model 1 summary

| **Independent Variables** | **Model 1** | **R square** | **Collinearity Statistics** | |
|---|---|---|---|---|
| | | | Tolerance | VIF |
| *CarSpeed* | 1.1 | 0.80 | 0.92 | 1.09 |
| *SI* | -0.5 | | 0.92 | 1.09 |
| *Length* | # | | - | - |
| *BusStop* | # | | - | - |

The parameter of $n_s$ represents that not-in-service buses are 0.5 km/h slower than average cars per each signalised intersection. Both buses and cars experience signal delay at intersections and if additional green time is provided to a bus (from TPS) then few cars running along the same route also benefit from it. The parameter of $n_s$ can be attributed to the mechanical characteristic of the buses as heavy vehicles with lower acceleration/deceleration rate.

With average speed of the car being one of the explanatory variables, part of the bus signal delay is considered in the parameter of $V_{car}$. However, the parameter of $V_{car}$ suggests that between two signalised intersections a not-in-service bus travels around 10% faster than an average car. The reason for this can be attributed to the following: $V_{car}$ represents the average speed of the cars from all the lanes. In Brisbane, a bus by law should be given right-of-way. The left lane (left hand driving) on which there is a bus stop should require a lower speed of the car than that in the right lane. A not-in-service bus does not have to stop at the bus stop and they generally travel in the right lane, resulting in a higher speed than that of $V_{car}$.

### *6.2 Model 2: In-service bus and car during peak period*

Here, in-service bus and car speeds during peak periods (7-9 AM on Wynnum Road, and 7-9 AM, 4-7 PM on Coronation Drive) are modelled. The case study sites are Wynnum Road and Coronation Drive, because no congestion was observed at Logan Road. Table 4 provides the descriptive statistic of the data used. The in-service buses have to stop for boarding/alighting of passengers. Due to unavailability of dwell time data, it is not explicitly considered in the model, but is implicitly taken into account as $n_b$. The backward stepwise regression analysis revealed that the three variables $n_b$, $l$ and $n_s$ are highly correlated, and only one of them should be included in the model. The variable $n_b$ is selected as the independent variable in building the model.

The final proposed model (model-2) is as follows:

$$V_{bus-IS-P} = 0.87 \times V_{car} - 0.4 \times n_b \tag{8}$$

The model has high R square of 0.82. The collinearity analysis also reveals no multicollinearity among the independent variables (see Table 6 ). The parameters denote that during peak periods, the $V_{bus\text{-}IS\text{-}P}$ is slightly lower than $V_{car}$, and gets worse with the increase in the number of bus stops. As discussed in the previous sub-section, the in-service bus travels on the left lane and during congested conditions when it stops at the bus stop, it has a tendency to significantly reduce the speed of the left lane. However, the right lane is not much affected by the bus stop resulting in higher average car speed than that of the bus. The negative parameter of $n_b$ represents the penalty of bus speed due to the boarding/alighting of passengers at the bus stop.

Table 4 Descriptive Statistics of Model 2

| | **Sample size** | **Min** | **Max** | **Mean** | **Median** | **Std. Dev.** |
|---|---|---|---|---|---|---|
| Car speed (*CarSpeed*) | 580 | 6.12 | 44.97 | 23.17 | 24.87 | 8.30 |
| In-service bus speed ($BusSpeed_{IS\text{-}peak}$) | 580 | 7.06 | 37.40 | 18.57 | 19.11 | 6.11 |
| Signalized intersection (*SI*) | 580 | 6.00 | 8.00 | 7.24 | 8 | 1.90 |
| Length (*Length*) | 580 | 2.29 | 3.33 | 2.98 | 3.33 | 3.08 |
| Bus stop (*BusStop*) | 580 | 6.00 | 8.00 | 7.24 | 8 | 0.92 |

Table 5  Model 2 summary

| **Independent Variables** | **Model 2** | **R square** | **Collinearity Statistics** | |
|---|---|---|---|---|
| | | | Tolerance | VIF |
| *CarSpeed* | 0.87 | 0.82 | 0.88 | 1.14 |
| *BusStop* | -0.4 | | 0.88 | 1.14 |
| *Length* | # | | - | - |
| *SI* | # | | - | - |

## 6.3 *Model 3: In-service bus and car speed during off-peak period*

Here, in-service bus and car speeds during off-peak periods (9AM-4PM in Coronation Drive, 7AM-9PM in Logan Road and 9AM-9PM in Wynnum Road) are modelled. Table 6 provides the descriptive statistics of the data used. A stepwise regression analysis reveals that except *l* (which is highly correlated with $n_s$), all other variables could be included in the model. The two variables $n_b$ and $n_s$ are not highly correlated in this case. During mid-day off-peak periods, some school bus routes (route 775, 806, 870) start operating (Translink 2012a). These buses only dwell at some selected bus stops along the corridor. By introducing a new study site (Logan Road) and adding these school routes, the variances of variables $n_b$ and $n_s$ are increased, and the correlation between them is reduced. The Pearson's coefficients of correlation of the independent variables are shown in Table 7. All Pearson's coefficients between independent variables with the dependent variable are less than 0.8, indicates that the correlations do not cause multicollinearity in the analysis (Katz 2006). The final proposed model (Model 3) is as follows:

$$V_{bus-IS-OP} = -1 + 0.9 \times V_{car} - 0.3 \times n_b - 0.1 \times \mathrm{n_s} \tag{9}$$

Table 6 Descriptive Statistics of Model 3

| | **Sample size** | **Min** | **Max** | **Mean** | **Median** | **Std. Dev.** |
|---|---|---|---|---|---|---|
| Car speed (*CarSpeed*) | 1870 | 13.40 | 47.14 | 24.84 | 25.03 | 6.13 |
| Bus speed (*BusSpeed$_{IS\text{-}offpeak}$*) | 1870 | 17.93 | 49.50 | 33.45 | 33.81 | 6.20 |
| Signalized intersection (*SI*) | 1870 | 4.00 | 12.00 | 8.09 | 8 | 3.41 |
| Length (*Length*) | 580 | 2.29 | 3.33 | 2.88 | 3 | 2.87 |
| Bus stop (*BusStop*) | 1870 | 6.00 | 12.00 | 9.26 | 8 | 2.47 |

Table 7 Pearson's Coefficients of Correlations of independent variables

| **Variables** | ***BusSpeedIS-offpeak*** | ***CarSpeed*** | ***SI*** | ***BusStop*** |
|---|---|---|---|---|
| *BusSpeed$_{IS\text{-}offpeak}$* | 1.00 | 0.81 | -0.63 | -0.68 |
| *CarSpeed* | | 1.00 | -0.74 | -0.64 |
| *SI* | | | 1.00 | 0.71 |
| *BusStop* | | | | 1.00 |

Table 8 Model 3 summary

| **Independent Variables** | **Model 3** | **R square** | **Collinearity Statistics** | |
|---|---|---|---|---|
| | | | Tolerance | VIF |
| *CarSpeed* | 0.9 | 0.71 | 0.41 | 2.49 |
| *SI* | -0.1 | | 0.37 | 2.70 |

| | | | | | |
|---|---|---|---|---|---|
| | *BusStop* | -0.3 | | 0.44 | 2.28 |
| | *Length* | # | | - | - |

The independent variables could explain 71% of the bus speeds. The collinearity statistic shows no sign of multicollinearity between the independent variables (all the VIFs are smaller than 2.8 and Tolerances are larger than 0.36). Overall, the bus speed is 90% of car speed minus 1 km/h during mid-day off-peak hours.

Comparing the coefficient values for Model 2 & Model 3, the difference in the coefficient of $n_b$ shows that the number of bus stops has less impact on the bus speed in off-peak periods (Model 3) than in peak periods (Model 2). With the number of signalised intersection $n_s$ held fixed, the gap between bus speed and car speed increases by 0.6 km/h for each bus stop added to the network in peak periods, but only increases by 0.3 km/h in off-peak periods. This difference can be attributed to that faced during off-peak period; the dwell time at the bus stop is lower due to the low number of transit passengers.

## 7 Model cross-validation

Here the proposed models are validated with the remaining 30% of the data which has not been used for mode development. The performance is evaluated in terms of Root Mean Squared Error (*RMSE*) and Mean Absolute Percentage Error (*MAPE*) defined in terms of actual speed ($V_A$) and estimated speed ($V_E$).

$$RMSE = \sqrt{\frac{\sum_{i=1 to N} (V_{A,i} - V_{E,i})^2}{N}} \qquad (10)$$

$$MAPE = \frac{\sum_{i=1 to N} \left(\frac{|V_{A,i} - V_{E,i}|}{V_{A,i}}\right)}{N} \qquad (11)$$

Where $N$ is the number of data points used for evaluation.

Figure 5, Figure 6 and Figure 7 presents the results for Model 1, Model 2 and Model 3 respectively. Here, the x-axis is the actual speed and y-axis is the estimated speed from the respective model. The diagonal line represents the line of equality. Points below and above the line of equality represents underestimation and overestimation, respectively. The MAPE from the three models are 7.2%, 5.5%, and 5.1 % respectively.

Similarly, RMSE for the three models are 2.9 km/hr, 3.6 km/hr, and 3.4 km/hr.

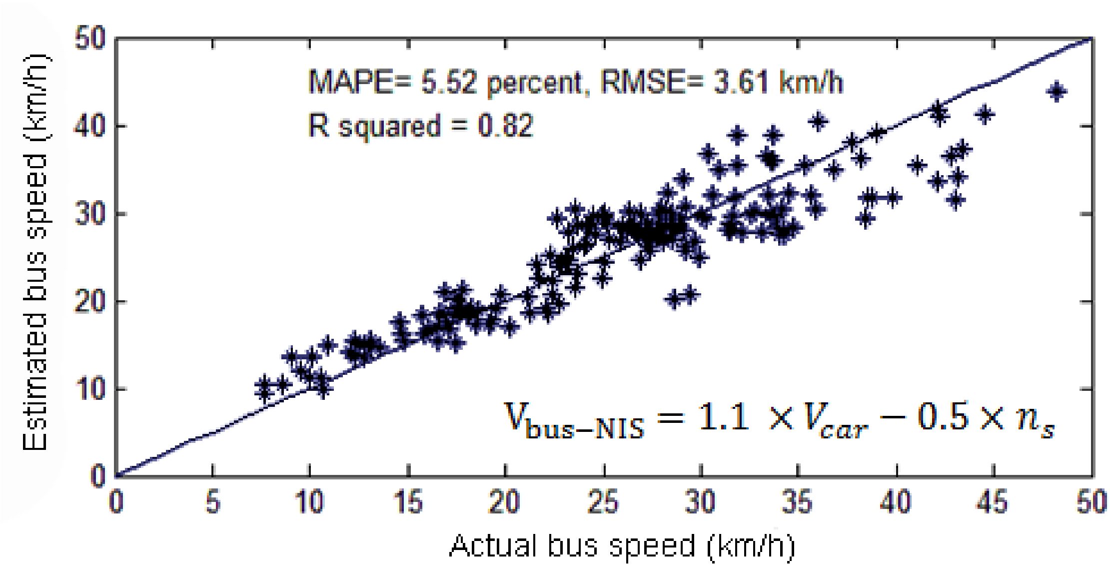


Figure 5 Model 1 validation results

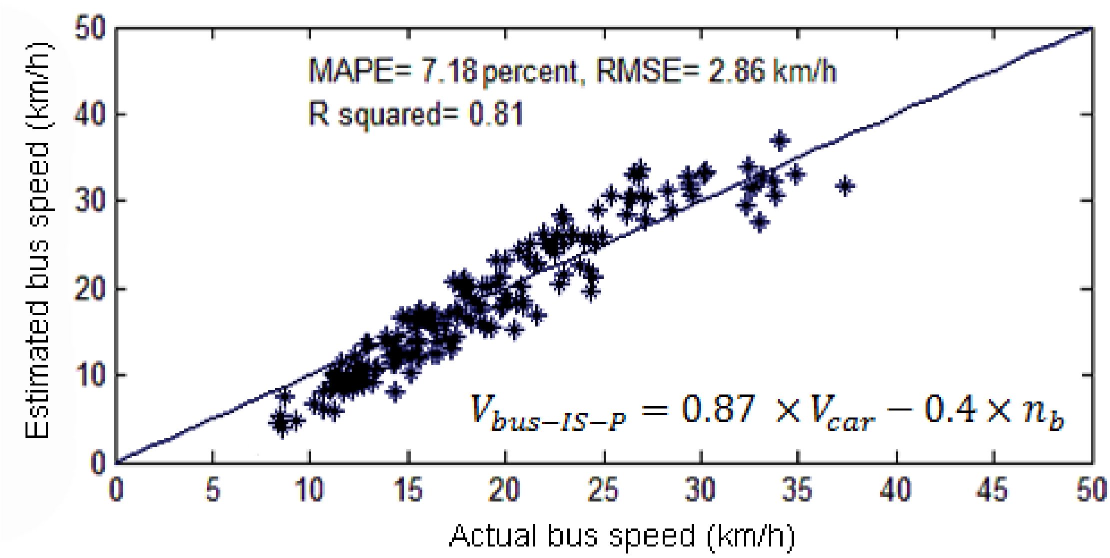


Figure 6 Model 2 validation results

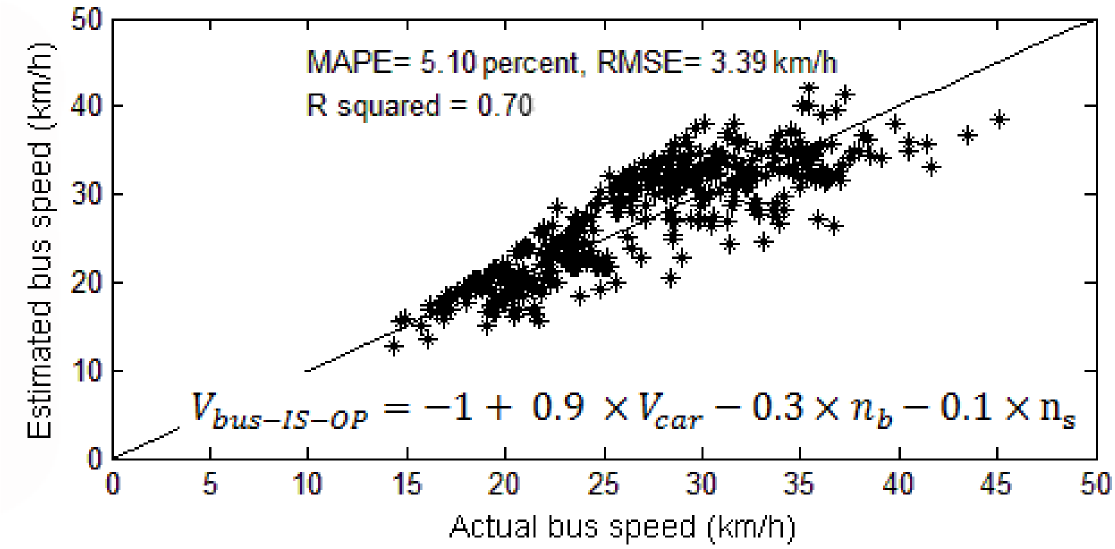


Figure 7 Model 3 validation results

The correlation R-squared values of the estimated against the actual bus speeds are less than 2% difference to the original R-squared of the models. It means the prediction models generalize well to the validation samples (Osborne 2000).

From these results, we can say that the model is cross-validated and can be applied with reasonable accuracy.

## 8 Conclusion

Understanding the relationship between multi-modals of transport could benefit strategic traffic planners and road operators. In this paper, we have exploited the real data used for traffic monitoring (Bluetooth and TSP) to model the relationship between bus and average car speed along the signalised urban corridor. The modelling results show that the relationship between bus and car speeds could be empirically established.

As expected, the coefficients of the proposed models are consistent with the expected observation that: a not-in-service bus can travel faster than the average car; an in-service bus has a slower speed than that of an average car; and number of bus stops and signalised intersection has significant impact on the in-service bus speed.

Sample size is not a problem in this study, since all buses and Bluetooth-enabled cars along the study corridors are considered. However, due to the limitation of data access, only 3 corridors in Brisbane are explored. Other variables which may contribute to the difference between bus and car speeds such as the number of boarding/alighting passengers, type of bus, etc. will be explored in future studies.

## 9 References


Bertini, R.L., and S. Tantiyanugulchai. 2004. "Transit buses as traffic probes: Use of geolocation data for empirical evaluation." *Transportation Research Record: Journal of the Transportation Research Board* no. 1870 (-1):35-45.

Bhaskar, A., E. Chung, and A. G. Dumont. 2010. "Analysis for the Use of Cumulative Plots for Travel Time Estimation on Signalized Network." *International Journal of Intelligent Transportation Systems Research* no. 8 (3):151-163. doi: 10.1007/s13177-010-0012-y.

———. 2011. "Fusing Loop Detector and Probe Vehicle Data to Estimate Travel Time Statistics on Signalized Urban Networks." *Computer-Aided Civil and Infrastructure Engineering* no. 26 (6):433-450. doi: 10.1111/j.1467-8667.2010.00697.x.

———. 2012. Urban Route Average Travel Time Estimation Considering Exit Turning Movements. Paper read at Transportation Research Board 91st Annual Meeting, at Washinton, D.C.

Cathey, F.W., and D.J. Dailey. 2002. "Transit vehicles as traffic probe sensors." *Transportation Research Record: Journal of the Transportation Research Board* no. 1804 (-1):23-30.

Chakroborty, P., and S. Kikuchi. 2004. "Using bus travel time data to estimate travel times on urban corridors." *Transportation Research Record: Journal of the Transportation Research Board* no. 1870 (-1):18-25.

Chien , Steven I., and Zhaoqiong Qin. 2004. "Optimization of bus stop locations for improving transit accessibility." *Transportation Planning and Technology* no. 27 (3):211-227. doi: 10.1080/0308106042000226899.

Davies, Laurie, and Ursula Gather. 1993. "The Identification of Multiple Outliers." *Journal of the American Statistical Association* no. 88 (423):782-792.

El Esawey, Mohamed, and Tarek Sayed. 2012. "A framework for neighbour links travel time estimation in an urban network." *Transportation Planning and Technology* no. 35 (3):281-301. doi: 10.1080/03081060.2012.671028.

Hamilton, A., B. Waterson, T. Cherrett, A. Robinson, and I. Snell. 2012. "The evolution of urban traffic control: changing policy and technology." *Transportation Planning and Technology* no. 36 (1):24-43. doi: 10.1080/03081060.2012.745318.

Katz, M.H. 2006. *Multivariable analysis: a practical guide for clinicians*: Cambridge Univ Pr.

Kieu, L.M., A. Bhaskar, and E. Chung. 2012. Bus and car travel time on urban networks: integrating bluetooth and bus vehicle identification data. Paper read at 25th ARRB Conference 2012, at Perth, Australia.

Levinson, H. 1983. "Analyzing transit travel time performance." *Transportation Research Record* (915).

Liu, Henry X., Wenteng Ma, Xinkai Wu, and Heng Hu. 2010. "Real-time estimation of arterial travel time under congested conditions." *Transportmetrica* no. 8 (2):87-104. doi: 10.1080/18128600903502298.

McKnight, C.E., H.S. Levinson, K. Ozbay, C. Kamga, and R.E. Paaswell. 2004. "Impact of traffic congestion on bus travel time in northern New Jersey." *Transportation Research Record: Journal of the Transportation Research Board* no. 1884 (-1):27-35.

O'brien, R.M. 2007. "A caution regarding rules of thumb for variance inflation factors." *Quality & Quantity* no. 41 (5):673-690.

Osborne, J.W. 2000. "Prediction in Multiple Regression." *Practical Assessment, Research & Evaluation* no. 7:n2.

Pearson, R. K. 2002. "Outliers in process modeling and identification." *Control Systems Technology, IEEE Transactions on* no. 10 (1):55-63. doi: 10.1109/87.974338.

Translink. 2012a. Brisbane Transport School Bus Services. http://translink.com.au/resources/travel-information/network-information/school/120326-services-by-route.pdf.

———. 2012b. Off-peak times. (23/07/2012), http://translink.com.au/tickets-and-fares/fares/off-peak-times.

Tsubota, T., A. Bhaskar, E. Chung, and R. Billot. 2011. Arterial traffic congestion analysis using Bluetooth Duration data. In *Australasian Transport Research Forum 2011*.