

Traffic Congestion Quantification for Urban Heterogeneous Traffic Using Public Transit Buses as Probes

47(4), pp. 257-267, 2019

<https://doi.org/10.3311/PPtr.9218>

Creative Commons Attribution 

Selvaraj Vasantha Kumar^{1*}, Ramaswamy Sivanandan²

RESEARCH ARTICLE

Received 21 March 2016; accepted 20 September 2017

Abstract

Understanding congestion in space-time domain using performance measures is essential prior to suggesting improvement schemes to reduce congestion. With technological advances like Global Positioning System (GPS), many metropolitan planning organizations give more emphasis on travel time based performance measures to quantify congestion, than on traditional way of using volume-to-capacity (V/C) ratios. In India, often it may not be possible to use personal vehicles as probes for travel time data collection. However, the public transit buses fitted with GPS devices could be used as cheap and effective probes to estimate the congestion status of other types of vehicles in the stream. The present study is an attempt in this direction. Two bus transit routes in Chennai, India were considered as case studies in order to cover the wide range of geometric and traffic conditions on urban arterials. GPS-fitted buses on these routes were used as probes in congestion quantification. As the dwell time at bus stops is a unique characteristic of transit buses when compared to other vehicles in the stream, a methodology has been proposed to find the dwell times including acceleration and deceleration times based on the approaching and departing speeds at bus stops. Regression models were then developed to predict the Congestion Index (CI) for various types of vehicles using bus CI, weighted carriageway width and the presence or absence of signalized intersection as independent variables. The results are promising and could be considered for real-time display of congestion levels for Advanced Traveler Information System (ATIS) applications.

Keywords

Urban traffic, Congestion index, Public transit probes, Personal vehicles, Global Positioning System, Heterogeneous traffic

1 Introduction

The problem of traffic congestion and ways to tackle it is a major concern in most of the metropolitan cities around the world and India is no exception to this. The exponential growth of personal vehicles, combined with increase in trips and trip lengths are the major reasons for traffic congestion in India. In Chennai, which is the fourth largest Metropolitan City in India, the total personal vehicle population has increased from 1 million in 1999 to almost 3.7 million in 2012. This accounts for about 270 percent rise in the last 13 years. The modal share of public transport in Chennai is currently 27%, which should be eventually increased to 46% by 2026 (CMRL, 2011). The decreasing use of public transport further exacerbates the congestion situation.

The solution options for reducing congestion are infrastructure expansion, Transportation System Management (TSM) measures, congestion pricing and technology applications like Intelligent Transportation System (ITS). Before suggesting any of the above measures to reduce congestion, it is essential to first study the current system performance. The process of measuring or estimating congestion by one or many of the performance measures is called congestion quantification. The quantification of traffic congestion is useful in many areas of transportation engineering, such as the traditional capacity improvement, alternatives analysis, wide range of planning and policy evaluations, optimization of traffic control strategies, providing alternate route choices in ITS applications, and for air quality and energy models.

The methods to quantify congestion can be grouped into Highway Capacity Manual (HCM) measures, queuing-related measures, and travel time-based measures. Since traffic congestion is a dynamic phenomenon with elements of both space and time, travel time based measures are more appropriate as they are good for dynamic conditions (d'Abadie and Ehrlich, 2002). The only disadvantage associated with the travel time based approach is the budgetary limitation which can impose restrictions on the number and coverage of travel time studies using test vehicles. In such cases, the personal or commercial vehicles fitted with GPS could be used as probes. In India, equipping private vehicles with GPS may be a difficult task due to privacy

¹ School of Civil and Chemical Engineering, VIT University, Vellore, India

² Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, India

* Corresponding author, e-mail: svasanthakumar1982@gmail.com

and other issues. However, the public transit vehicles in major metropolitan cities of India are gradually being equipped with GPS devices. The Metropolitan Transport Corporation (MTC) of Chennai has GPS in 600 buses out of the total fleet size of about 3,400. Process is underway to install the GPS in another 1,000 buses. Since there are no exclusive bus lanes for buses in Chennai, the public transit buses have to travel alongside other vehicles and quite often both experience similar traffic conditions, intersection control and incidents or special events. The use of public transit buses as probes also offers other advantages like frequent trips during peak hours, wide range network coverage, easy accessibility to the data, low initial and maintenance cost when compared to that of location based smart sensors, trouble-free installation and maintenance when compared to the fixing of loop detectors and other video based sensors. While these are the advantages of using public transit buses as probes, there are some associated challenges too. The dwell time or stopping time at bus stops, which is a unique characteristic of transit buses when compared to other vehicles in the stream, is one such issue. A careful consideration of the above aspect is very important while developing models to predict the congestion level for other vehicle types using only buses as probes. As the carriageway width and presence or absence of signalized intersection also influences the duration and extent of congestion, they have been considered in the present study.

Unlike in developed countries, in India and in other developing countries, the use of automatic traffic data collection techniques is in nascent stage. Also, the traffic comprises wide variety of vehicle classes of varying static and dynamic characteristics and lane discipline is poor on urban roads. This restricts the applicability of certain performance measures which are lane based or vehicle based units for congestion quantification (eg., congested roadway expressed in lane-miles, congested travel expressed in vehicle-miles). Hence, in the present study, one of the simple and most widely used performance measures, namely CI has been used. Most of the studies using CI adopt a constant free flow speed/travel time (FFS/FF(TT)) across all the modes and sections assuming that the driver obeys speed limit restrictions. In India, due to various kinds of vehicles with different speed characteristics, the assumption of constant FFS/FF(TT) for all the vehicle modes may not be realistic. Hence the present study takes into account this variable nature of FF(TT) across modes and sections, by conducting GPS probe runs during early hours when the free flow conditions prevail on the roads.

The above issues highlight the need and challenges in developing a model which can predict the congestion level for various kinds of vehicles in the stream using only bus transit as probes.

2 Literature Review

The various methods to quantify congestion along with some notable studies under each method are briefly reviewed in this section.

2.1 Highway Capacity Manual (HCM) based method

The HCM based method mainly uses the volume-to-capacity (V/C) ratio and level of service (LOS) as performance measures to measure/quantify congestion (Highway Capacity Manual, 2010). The use of V/C ratio is one of the traditional ways to quantify congestion because of the relative ease of traffic volume data collection. They usually require detailed, location-specific input data, which makes them more appropriate for individual highway segments or intersections, rather than for corridors or region-wide analysis (Quiroga, 2000; Grant and Fung, 2005). The HCM measures are difficult to use for long-range comparisons because concepts such as capacity and speed-flow relationships tend to change over time. Sometimes, estimating capacity even within 10 percent of its actual value can be a difficult task because of many variables which can influence capacity. Also, HCM based measures break down in oversaturated conditions (Beverly, 2004). Many HCM based studies adopt different classification of V/C ratios for depicting various congestion levels (CMP, 2011). In India, the notable studies on HCM based approach are by Patel and Varia (2010), Maitra et al. (1999) and Anjaneyulu and Nagaraj (2009).

2.2 Method based on queuing measures

The queuing-related measures concentrate on measurement of queue length and lane occupancy as performance measures to quantify congestion. Since it is usually impractical to measure queues on a broader spatial scale, queuing-related measures tend to be inappropriate for planning and policy-related analyses (Quiroga, 2000). Installation of multiple loop detectors for queue detection in city streets has almost been precluded in view of the environmental cost and the need for systems maintenance (Yi et al., 2001). The study by Geroliminis and Skabardonis (2011) concluded that the queue length estimation helps to preempt congestion by predicting congestion locations/times. It can also be integrated in real-time traffic management schemes, either at intersection scale or at larger complex urban systems.

2.3 Travel time based method

The travel time-based measures to quantify congestion are primarily based on travel time, travel speed, and delay. Since traffic congestion is a dynamic phenomenon with elements of both space and time, travel time based measures are more appropriate as they are flexible enough to describe traffic conditions at various levels of resolution in both space and time. This makes travel time based measures appropriate for handling specific locations as well as entire corridors (Quiroga, 2000). Since most of the travel time based measures are dimensionless, it helps to compare mobility levels on different roadways or among different modes of transportation. It also allows analysts to perform comparisons over long periods of time, e.g., years or decades. The measures associated with the

time or speeds are easy to understand and interpret by both the transportation professionals and the travelling public. Travel time-based measures translate easily into other measures like user costs, and can be used directly to validate planning models such as travel demand forecasting models (Grant, 2011). Another advantage is that the travel time-based measures are applicable across modes (Laird, 1996) and reflect the combined effects of geometric and operational features of the road (Beverly, 2004). Travel time measures can do a better job at pinpointing locations of congestion and can help in identifying congestion causes (Beverly, 2004). All these reasons make travel time based measures extremely powerful, versatile, and desirable for congestion quantification. An increasing number of transportation agencies are switching to travel time measures to monitor and manage congestion (Quiroga, 2000; Grant and Fung, 2005). The National Cooperative Highway Research Program project on "Quantifying Congestion" (Lomax, 1997), has recommended that travel time-based measures be used to estimate congestion. The latest 2011 Urban Mobility report (Schrank, 2011) gives utmost priority for travel time based procedures to quantify congestion.

Many studies have been reported worldwide which use travel time estimates to quantify congestion. The important ones are by Quiroga (2000), Taylor et al. (2000), Tong (2000) and D'Este et al. (1999). In the above studies, the test vehicle (passenger car) is adopted and a constant free flow travel time was assumed while calculating the performance measures such as delay, time moving, congestion index, proportion stopped time, etc. The studies on travel time based congestion quantification are still in a rudimentary stage in India. One of the earlier attempts in India was by Srinivasan and Shetty (1971) and a similar recent study was by Lovely and Madhu (2006).

In the studies on travel time based congestion quantification, the posted speed limit was considered as the FFS assuming that the drivers follow speed regulations and FF(TT) was derived using this constant FFS. The variable or actual FF(TT), which is more realistic in urban context, has not been considered while calculating congestion measures. Most of the studies on travel time based congestion quantification use only test vehicles (mostly passenger cars) which are specifically dispatched for travel time data collection. The objective of most of the travel time based congestion studies as reviewed above was to assess the performance of the current system using measures such as delay, congestion index, etc., and so they employed test vehicles for a fixed duration. But if one wants to display the real time congestion status through web for ATIS applications, running test vehicles at frequent intervals will not be feasible all the time due to financial constraints. In such cases, an attractive option is to use probes which are already in the stream (and not specifically for the purpose of data collection), such as personal and commercial vehicles, public transit buses, etc. In countries like India, fixing GPS on personal vehicles is not

practically feasible due to privacy and other issues. For commercial vehicles such as call-taxi, though they have GPS systems for tracking their own vehicles, they may not be willing to share the GPS data due to business reasons. Hence, a viable option is to use public transit buses fixed with GPS instruments for quantifying the congestion level in the stream. The studies on the use of bus as probes for congestion quantification are very limited (Chen, 2010; Berkow et al., 2008). The driving maneuvers for heavy vehicles especially in urban roads are more difficult than passenger cars because of its passenger car equivalent factor. This is one of the main reasons why less number of studies were attempted on the use of bus as probes for congestion quantification. The above review of past studies indicate the need and challenges in developing a model which can predict the congestion level for various kinds of vehicles in the stream using only bus transit as probes.

3 Data Collection and Extraction

The study stretch selected for the present study was two bus transit routes, namely 5C and 23C in Chennai, India. These routes cover the wide range of geometric and traffic conditions on urban arterials. The selected stretches are a typical representation of urban routes in India comprising road links of different categories like major arterials, and collector streets with varying geometrics and volume levels. A photograph of the 5C route at one of the locations is shown in Fig. 1. As can be seen, the traffic is highly heterogeneous in nature with mix of vehicles of different static and dynamic characteristics such as two-wheeler, three-wheeler, light motor vehicle and heavy motor vehicles. The lane discipline is also poor with no exclusive bus lanes and the buses have to share the road with other vehicles. The route number 5C connects the Parrys bus terminus in the northern part of Chennai, and the Taramani bus terminus in the southern part. The time headway between the buses varies between 15-30 minutes. The total route length of 5C is 15 km and has 23 bus stops and 14 signalized intersections. The route number 23C connects the Ayanavaram bus terminus in the north western part of the city and the Tiruvanmiyur bus terminus in the southern part. The time headway between the buses varies between 15-30 minutes. The total route length of 23C is 19 km in outbound direction (Tiruvanmiyur to Ayanavaram) and 21.5 km in inbound direction (Ayanavaram to Tiruvanmiyur) and has 18 and 21 bus stops in outbound and inbound directions, respectively. The number of signalized intersections in 23C route is 21 and 22 in outbound and inbound directions, respectively.

Data collection was carried out at different times of the day during April - June 2012 using GPS units, which were fixed in the public transit buses and by manually carrying GPS devices in three types of personal vehicles - two-wheeler, three-wheeler and car. Data collection was carried out using the funds available through Centre of Excellence in Urban Transport at IIT Madras, sponsored by Ministry of Urban Transport (MoUD),



Fig. 1 Snapshot of one of the locations on bus route 5C

Govt. of India. Fourteen trips were made for each mode in each bus route considered. The personal vehicle trip was started at the same time and location as that of the public transit bus and followed the same route with a speed representative of other vehicles. According to the travel time data collection handbook (Turner, 1998), the suggested test vehicle runs required for travel time data collection on urban arterial streets is 5, 6 and 15 at 90% confidence level with permitted relative error of $\pm 10\%$, 95% confidence level with $\pm 10\%$ error and 95% confidence level with $\pm 5\%$ error respectively. This states that, the minimum number of test vehicle runs should be in the range of 5 to 15. The 14 trips for each mode in each bus route as adopted in the present study was close to the suggested number of test vehicle runs at 95% confidence level with $\pm 5\%$ error. Also, it states that the test vehicle runs should be evenly distributed over various times in order to capture the off-peak and peak traffic characteristics. The 14 trips were well distributed across varying time over several days in order to capture the off-peak and peak traffic characteristics as shown in Table 1 and 2 for 23C-outbound and inbound directions, respectively.

Table 1 Details of data collection for 23C route – outbound direction

Two-wheeler	Three-wheeler	Car
Date (Trip starting time)	Date (Trip starting time)	Date (Trip starting time)
June 05, 2012 (8.36 am)	June 05, 2012 (8.36 am)	May 18, 2012 (8.24 am)
May 07, 2012 (8.38 am)	June 04, 2012 (9.03 am)	May 16, 2012 (8.48 am)
May 08, 2012 (9.38 am)	May 25, 2012 (9.19 am)	May 14, 2012 (9.27 am)
May 09, 2012 (10.00 am)	June 01, 2012 (9.49 am)	May 23, 2012 (9.30 am)
April 27, 2012 (10.44 am)	May 24, 2012 (9.58 am)	May 11, 2012 (9.46 am)
May 03, 2012 (11.39 am)	May 25, 2012 (11.58 am)	May 17, 2012 (10.15 am)
April 26, 2012 (1.44 pm)	May 29, 2012 (1.10 pm)	May 16, 2012 (11.51 am)
April 27, 2012 (1.49 pm)	May 28, 2012 (1.55 pm)	May 14, 2012 (12.34 pm)
May 02, 2012 (2.50 pm)	May 24, 2012 (2.24 pm)	May 17, 2012 (2.14 pm)
April 26, 2012 (4.19 pm)	May 29, 2012 (3.36 pm)	May 11, 2012 (4.12 pm)
April 27, 2012 (4.33 pm)	May 28, 2012 (4.42 pm)	May 18, 2012 (4.37 pm)
May 07, 2012 (5.35 pm)	May 24, 2012 (4.47 pm)	May 08, 2012 (5.21 pm)
May 02, 2012 (6.22 pm)	June 01, 2012 (5.33 pm)	May 09, 2012 (5.38 pm)
May 03, 2012 (7.36 pm)	May 25, 2012 (6.59 pm)	May 14, 2012 (6.59 pm)

Table 2 Details of data collection for 23C route – inbound direction

Two-wheeler	Three-wheeler	Car
Date (Trip starting time)	Date (Trip starting time)	Date (Trip starting time)
June 05, 2012 (9.58 am)	June 05, 2012 (9.57 am)	May 18, 2012 (9.37 am)
May 07, 2012 (10.33 am)	May 25, 2012 (10.42 am)	May 16, 2012 (10.23 am)
May 08, 2012 (11.36 am)	June 04, 2012 (10.43 am)	May 14, 2012 (10.53 am)
May 09, 2012 (11.36 am)	June 01, 2012 (11.13 am)	May 23, 2012 (11.02 am)
April 27, 2012 (12.07 pm)	May 24, 2012 (11.28 am)	May 11, 2012 (11.47 am)
May 03, 2012 (1.09 pm)	May 25, 2012 (1.30 pm)	May 17, 2012 (12.08 pm)
April 26, 2012 (2.53 pm)	May 29, 2012 (2.21 pm)	May 16, 2012 (1.17 pm)
April 27, 2012 (3.10 pm)	May 28, 2012 (3.13 pm)	May 14, 2012 (2.57 pm)
May 02, 2012 (4.48 pm)	May 24, 2012 (3.40 pm)	May 17, 2012 (3.30 pm)
April 26, 2012 (5.59 pm)	May 28, 2012 (5.56 pm)	May 11, 2012 (5.49 pm)
April 27, 2012 (6.10 pm)	May 29, 2012 (6.13 pm)	May 18, 2012 (6.19 pm)
May 07, 2012 (7.12 pm)	May 24, 2012 (6.38 pm)	May 08, 2012 (7.15 pm)
May 02, 2012 (8.06 pm)	June 01, 2012 (7.13 pm)	May 09, 2012 (7.33 pm)
May 03, 2012 (9.09 pm)	May 25, 2012 (8.40 pm)	May 14, 2012 (8.44 pm)

Sometimes the data may become unusable if the GPS data contains data losses due to high rise buildings or tree cover in which case, a proper data quality check is essential to ensure accuracy in the travel time calculation using cumulative distance values calculated through Haversine formula. The Haversine formula provides the great circle distance (i.e., shortest distance over the earth's surface) between two pairs of latitudes and longitudes. For the case of GPS data loss along a straight road, the loss in data was not a major issue because the Haversine formula for cumulative distance assumes the missing portion as a linear section (shortest distance) which gives reasonably accurate results. However, for the GPS data loss along a curved road, the assumption of straight section (shortest) for the missing portion is not correct and this type of error if unnoticed may underestimate the travel time for that particular section. Under such circumstances, a suitable GPS data point at the curve was assumed with known latitude and longitude values (can be identified by placing the cursor in the selected point on the map). For finding the time stamp of the hypothetical GPS data point, a recently passed trip was used as the reference to find the time of travel from the last point (point from where the data loss started) to the hypothetical GPS point location. This travel time was then added to the corresponding time of the last GPS point in the missing data to find the time stamp of the hypothetical GPS data point as shown in Eq. (1).

$$T_{s(h)} = T_{s(l)} + t \quad (1)$$

Where, $T_{s(h)}$ is the time stamp of the hypothetical GPS data point at the curve expressed in hh:mm:ss (eg., 08:10:50); $T_{s(l)}$ is the time stamp of the last GPS data point (from where the data loss started) expressed in hh:mm:ss (eg., 08:10:20); t is the travel time from the point where the data loss started to the hypothetical GPS point location calculated from a recently passed trip. The data quality control as described above to deal with the missing GPS data due to signal loss helped to obtain the correct travel time observations.

The total study stretch was divided into 500 m subsections in order to facilitate the comparison of section travel times of the bus and other vehicles. Uniform section length based discretization was adopted rather than link based discretization, because the number of intersections was less in the present case and so the number of links will be less for analysis. The uniform length based discretization inherently takes into account the delays at signalized intersections. For each trip, the travel times of bus probes and the personal vehicles in each of these 500 m subsections were calculated. This process is repeated for all the 168 trips made (3 directional routes x 4 modes x 14 trips). In order to find the free flow travel times (FF(TT)) to be used in the CI calculations, three sample runs were carried out for each mode in each route considered during early hours of 4 to 6 am over a two-week period and FF(TT) for each 500 m section was extracted. As carriageway width is one of the

independent variables in the proposed model, it was measured manually using tapes. For sections with varying carriageway widths, a weighted carriageway width has been arrived at based on the section lengths as shown below.

$$\text{Weighted carriageway width} = \frac{\sum_{i=1}^n (C_i \times l_i)}{\sum_{i=1}^n (l_i)} \quad (2)$$

where, C_i is the carriageway width for section of length l . The presence or absence of signalized intersection(s) in each 500 m section of both the routes were also noted down.

4 Congestion Analysis and Modelling

The methodology proposed here aims to find the relationship between the personal vehicle(s) CI which is the variable of interest (dependent variable) and the public transit CI, which is taken as an independent variable. As road width and presence/absence of an intersection influences traffic congestion, they have also been considered as additional independent variables to quantify congestion levels for vehicles in the stream, using only buses as probes. The methodology proposed here uses regression technique for determining the model parameters. This section starts with the procedure to remove the dwell times at bus stops with associated acceleration and deceleration times. The correlation analysis between bus and other vehicle travel times is discussed next. The details of the regression technique and the model formulation are presented after that.

4.1 Procedure to find the dwell times with associated acceleration and deceleration

Since all modes of vehicles are sharing the same roadway without any exclusive bus lanes, the only characteristic that differentiates the bus probes from the remaining vehicles is the dwell time at bus stops. The methodology to determine the dwell time or stopping time at bus stops including the time of deceleration and acceleration based on the speed of movement of the transit buses is explained below.

After fixing the latitude and longitude range of each bus stop, a program in MATLAB has been written which will check for the lowest speed value (denoted as K_t , where t is the GPS time) corresponding to the selected bus stop range. For finding the time of start of deceleration, each pair of speed values (K_t and K_{t-1} , K_{t-1} and K_{t-2} , K_{t-2} and K_{t-3} , etc.) will be checked, until the speed $K_{t-n} > K_{t-(n+1)}$, where $n=0, 1, 2, 3, \dots$. The time corresponding to K_{t-n} is considered as the time of start of deceleration. In other words, for finding the time of start of deceleration, each pair of successive speed values (prior to the lowest speed) will be checked, until the prior speed ($K_{t-(n+1)}$) in the pair is lower. A similar procedure is adopted for finding the time of end of acceleration. For finding the time of end of acceleration, each pair of speed values (K_t and K_{t+1} , K_{t+1} and K_{t+2} , K_{t+2} and K_{t+3} , etc.) will be checked, until the

speed $K_{t+n} > K_{t+(n+1)}$, where $n=0, 1, 2, 3 \dots$. The time corresponding to K_{t+n} is considered as the time of end of acceleration. The difference between the two times is taken as the dwell time, which includes the acceleration and deceleration times.

Once the dwell time with associated acceleration and deceleration times were found for all the bus stops in each bus trip, they were removed from the actual section travel times of the bus. After removing the bus stop dwell times, the bus travel time was correlated with the corresponding personal vehicle travel time in each section as explained in the following section to check whether buses could be considered as probes.

4.2 Correlation analysis between bus travel times and other vehicle

The two variables considered for the correlation analysis are the section-wise bus travel times (dwell time removed) and the personal vehicle travel times for each trip. The results of correlation analysis show that the correlation coefficient is positive for all the 126 trips. It shows that if the traffic is higher in a particular section, both transit probe and other vehicles will get slowed down. Similarly, if free flow condition exists, both the modes travel fast. The hypothesis testing to check whether the correlation coefficient is zero or not shows that at 5% level of significance, 109 out of 126 trips (here, each trip is a pair of public transit and personal vehicle) have correlation between the bus travel times and other modes travel times. Thus, the bus travel time after removing the bus stops dwell times could be considered as a probe input for predicting the congestion status for other vehicles via travel time based measures, such as CI, as explained in the following section.

4.3 Regression model to determine the personal vehicle CI

In the present study, regression analysis was used to develop models to predict personal vehicle congestion index (CI). The reason for selecting regression analysis is that it offers many advantages when compared to other methods such as simplicity, easy interpretation of the coefficients and its signs, availability in many statistical software packages and acceptance by scientific community. Other statistical methods such as artificial neural network (ANN) or support vector machines (SVM) haven't been used as they required a huge database for model building. The regression analysis used in the present study required only limited input data, i.e., 14 trips for each mode in each bus route considered. One of the disadvantages of regression analysis is that the value of the dependent variable cannot be correctly estimated if the value of the independent variable falls outside the range of values used for determining the linear regression equation. However this may not be an issue in the present case as the input dataset used for model building was well distributed across varying time in order to capture the off-peak and peak traffic characteristics.

One of the simple and most widely used congestion measures called CI is used in the present study to depict the congestion level of personal vehicles using public transit buses as probes. The equation for finding the CI is given by (Richardson and Taylor, 1978)

$$\text{Congestion Index}(CI) = \frac{\text{Actual travel time} - \text{free flow travel time}}{\text{Free flow travel time}} \quad (3)$$

A CI value of zero means that the actual travel time and free flow travel time are equal. A value of one means that the actual travel time is twice the free flow travel time. An index greater than 2 indicates congested condition (Taylor et al., 2000; D'Este et al., 1999). In the present study, three traffic periods were considered, namely, the off-peak, peak-morning or peak-evening, to study the congestion level of vehicles during different time periods. The time period from 8 am to 11 am and 5 pm to 8 pm were considered as the morning and evening peak periods, respectively, and the remaining time periods were considered as off-peak. Based on this, a total of 126 trips (here each trip is a pair of public transit and personal vehicle, both of which started at same time and location) in all three directional routes considered in this study were classified into off-peak, morning peak hour or evening peak hour trips based on their trip departure times. Section-wise average travel times were calculated for the off-peak, morning peak and evening peak hour trips. The average section-wise travel times were then used as actual travel times in Eq. (3) to calculate the section-wise CI.

For public transit buses, the dwell time removed section travel times were used for bus CI calculation. Both fixed and variable FF(TT) were considered while calculating CI using Eq. (3). For fixed FF(TT), a constant free flow speed (FFS) of 40 km/h (the corresponding FF(TT) is 45 sec. for a 500 m section) was assumed for all the modes. The reason for selecting 40 km/h as FFS is that, in most parts of Chennai city, the speed limit is 40 km/h. The section-wise variable FF(TT) is based on three sample probe runs for each mode in each of the three routes during the free-flow hours of 4:00 to 6:00 am. The extracted section-wise FF(TT) was averaged across the three trips for each mode and are shown in Figs. 2, 3 and 4 for 5C, 23C- outbound and 23C-inbound directions, respectively.

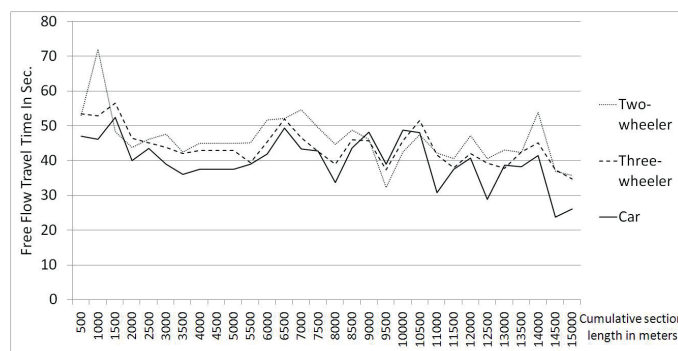


Fig. 2 Free Flow Travel Times of Two-Wheeler, Three-Wheeler and Car for 5C Route

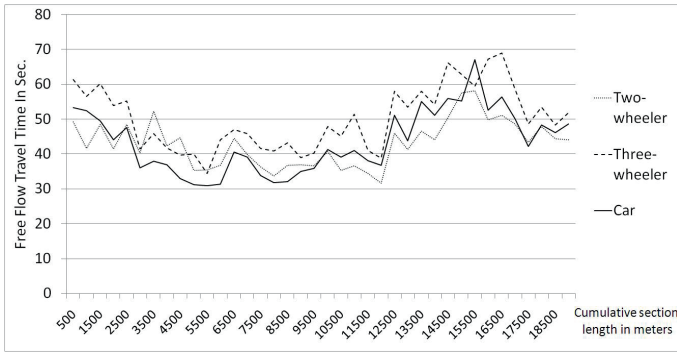


Fig. 3 Free Flow Travel Times of Two-Wheeler, Three-Wheeler and Car for 23C (Outbound) Route

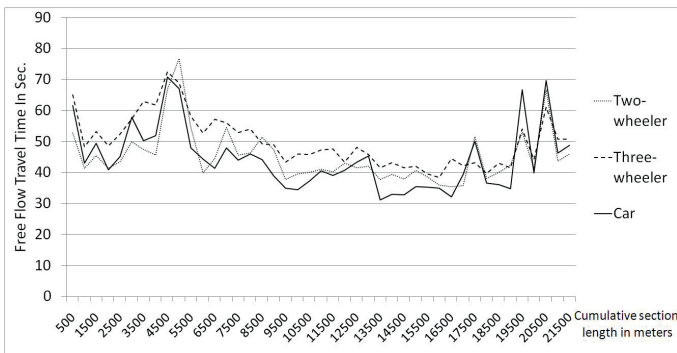


Fig. 4 Free Flow Travel Times of Two-Wheeler, Three-Wheeler and Car for 23C (Inbound) Route

It can be observed that car travels much faster than two-wheeler and three-wheeler and shows comparatively less travel times in most of the sections during the free flow hours. It is also observed that, all the three modes exhibit a similar trend (increasing or decreasing pattern) in travel time over the sections. This indicates that road geometry plays a significant role in FF (TT) of various modes. For example, if the carriageway width is less in a certain section, the travel speed of all three modes gets reduced and shows a similar trend of travel time values. Hence, instead of directly using the averaged FF (TT) for each section, which will be more site-specific, the FF (TT) classified based on carriageway width was considered. For this, the weighted carriageway width was divided into four categories, namely, 4 to 8 m, 8 to 12 m, 12 to 16 m and 16 to 20 m, and the section-wise FF (TT) s were averaged. This process is repeated for all the modes. Similar to personal vehicles, three early morning trips of public transit buses were considered to derive the section-wise variable FF (TT) after removing the dwell time. The results of FF (TT) classified based on carriageway width is shown in Fig. 5. It can be seen that the FF (TT) gradually decreases when the width of the carriageway increases. The corresponding FFS is shown in Fig. 6 for various categories of carriageway widths. This way, one can get the FF (TT)/FFS for a specific mode and for a specific carriageway width, which could be used in CI calculations.

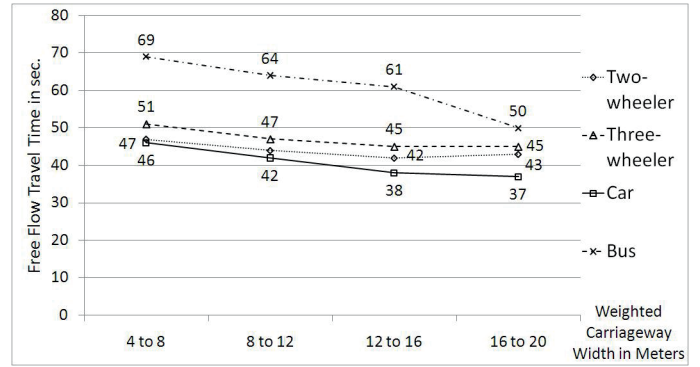


Fig. 5 Free Flow Travel Times for Various Modes Classified Based on Weighted Carriageway Widths

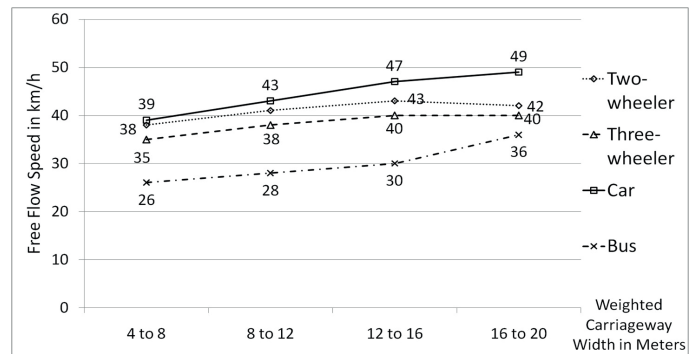


Fig. 6 Free Flow Speeds for Various Modes Classified Based on Weighted Carriageway Widths

Seven regression models based on categorical regression analysis were developed to study the relationship between personal vehicle(s) CI and public transit CI for each period (off-peak, peak-morning, peak-evening) and each mode (two-wheeler, three-wheeler and car). A final combined model considering all the modes and periods together was also developed to predict the congestion level for other vehicles in the stream using bus probe data. The general forms of the estimated regression models are shown in Eqs. (4), (5) and (6).

A). Off-peak period for all the three modes

$$y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 \quad (4)$$

Where, y = Personal vehicle CI for off-peak period

X_1 = Public transit CI

X_2 = Carriageway width

X_3 = Presence (1) or absence (0) of Signalized intersection

X_4 = 1 if three - wheeler, X_4 = 0 otherwise,

X_5 = 1 if car, X_5 = 0 otherwise,

where, a_0 is the regression constant; a_1, a_2, a_3, a_4 and a_5 are the estimated regression coefficients for the variables X_1, X_2, X_3, X_4 and X_5 , respectively. X_4 and X_5 are the categorical regression variables to account for the three modes considered. A similar equation will apply for the peak-morning and peak-evening periods as well.

B). Two-wheeler for all the three periods

$$\hat{y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 \quad (5)$$

Where, \hat{y} = Two-wheeler CI

X_1 = Public transit CI

X_2 = Carriageway width

X_3 = Presence (1) or absence (0) of Signalized intersection

$X_6 = 1$ if peak morning period, $X_6 = 0$ otherwise

$X_7 = 1$ if peak evening period, $X_7 = 0$ otherwise

where, b_0 is the regression constant; b_1, b_2, b_3, b_6 and b_7 are the estimated regression coefficients for the variables X_1, X_2, X_3, X_6 and X_5 , respectively. X_6 and X_7 are the categorical regression variables to account for the three periods considered. A similar equation will apply for the three-wheeler and car as well.

C). All the modes for all the periods together

$$\hat{y} = c_0 + c_1X_1 + c_2X_2 + c_3X_3 + c_4X_4 + c_5X_5 + c_6X_6 + c_7X_7 \quad (6)$$

Where, \hat{y} = Personal vehicle CI for a particular mode during a particular period

X_1 = Public transit CI

X_2 = Carriageway width

X_3 = Presence (1) or absence (0) of Signalized intersection

$X_4 = 1$ if three - wheeler, $X_4 = 0$ otherwise,

$X_5 = 1$ if car, $X_5 = 0$ otherwise,

$X_6 = 1$ if peak morning period, $X_6 = 0$ otherwise

$X_7 = 1$ if peak evening period, $X_7 = 0$ otherwise

where, c_0 is the regression constant; $c_1, c_2, c_3, c_4, c_5, c_6$ and c_7 are the estimated regression coefficients for the variables $X_1, X_2, X_3, X_4, X_5, X_6$ and X_7 , respectively. The results of regression analysis using the above models are presented in the following section.

5 Results and Discussion

Microsoft Office Excel 2007 is used to determine the regression model. The plot of personal vehicle CI versus the bus CI is shown in Fig. 7 for all the seven cases for constant FF(TT) case. Similarly, Fig. 8 shows the corresponding cases for variable FF(TT) case.

Overall, the plots in both Figs. 7 and 8 do suggest a linear relationship between the personal vehicle CI and bus CI calculated using the dwell time removed bus travel time. It is observed from both the figures that, for off-peak period, the CI variation is less when compared to that of peak-morning and peak-evening periods. The comparison of Figs. 7 and 8 suggests that the assumption of constant FFS/FF (TT) while calculating CI values may not be a realistic one as it exaggerates the congestion level, especially for public transit buses. For example, in Fig. 8, the CI values based on actual FFS/FF (TT) for bus is only in the range of 0 to 1.5 during off-peak period and 0 to 2.5 during peak periods, whereas, the CI variation for bus assuming constant FFS/FF (TT) is in the range of 0 to 3 and 0 to 4.5 for off-peak and peak periods, respectively (Fig. 7).

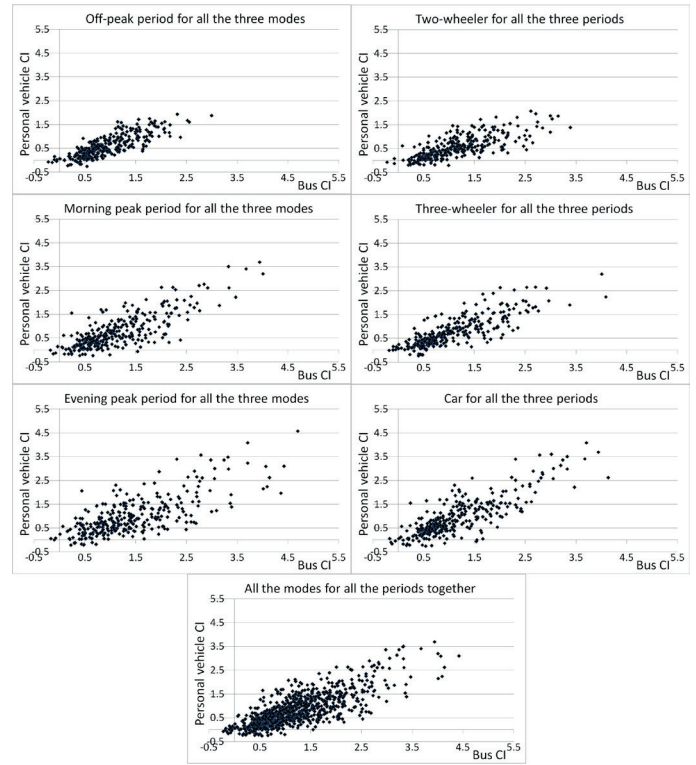


Fig. 7 Plot of Bus CI versus Personal vehicle CI for various categories of modes and traffic periods for constant FF(TT) scenario

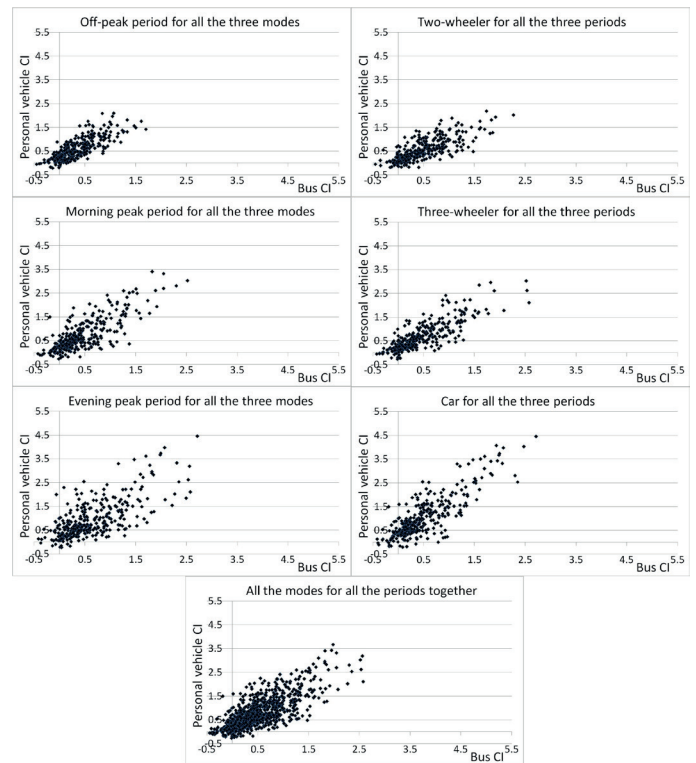


Fig. 8 Plot of Bus CI versus Personal vehicle CI for various categories of modes and traffic periods for variable FF(TT) scenario

A similar phenomenon can be observed while considering the individual modes also; CI variation for bus is exceptionally high and reaches a value of around 4.5 in Fig. 7, whereas the CI calculated using variable FF(TT) is only in the range of 0 to 2.5 as observed in Fig. 8. This clearly shows that the use

of variable or actual FFS/FF(TT) for CI calculations would be more realistic than using a constant FF(TT) across all the sections and modes. Hence, it was decided to use the variable or actual FF(TT) for CI calculations. In Fig. 8, it is interesting to see that there is a considerable difference in CI variation among the personal vehicles considered. For example, the CI variation for two-wheeler is only in the range of 0 and 2.5, whereas for three-wheeler, the range is 0 and 3.5 and for car, it is between 0 and 4.5. Since two-wheelers in heterogeneous traffic can easily maneuver through the available space between vehicles as compared to three-wheelers and cars, their travel times are lesser, thereby showing a relatively low variation in CI when compared to other modes. The wide variation in CI between classes of vehicles as observed in Fig. 8 suggests the need for defining mode-specific congestion levels under heterogeneous traffic conditions as existing in India.

Overall, the R^2 values for all the seven regression models lie in the range of 0.6 to 0.8, which suggests a reasonably good model for the present scenario of using bus as probes for predicting the other vehicle CI values, when considering the unique nature of traffic and data collection constraints in India. The results of regression analysis showing the coefficients of regression variables are shown in Table 3.

Table 3 Regression coefficients and p-values of independent variables

	Regression coefficients	p-value
Intercept	0.098	0.055
Public Transit CI (X_1)	0.937	4.2E-172
Carriageway width (X_2)	-0.010	0.019
Presence/ absence of signalized intersection (X_3)	0.188	1.74E-12
Categorical variable for mode (X_4)	0.143	1.72E-06
Categorical variable for mode (X_5)	0.418	8.05E-40
Categorical variable for period (X_6)	0.040	0.168
Categorical variable for period (X_7)	0.062	0.043

The carriageway width generally exhibits a negative coefficient. It is logical that when carriageway width reduces, the capacity gets reduced and will result in an increased travel time and congestion. Intersection presence/absence (a binary variable which takes the value of 1 or 0) exhibits a positive regression coefficient. It is consistent with the fact that when there is a signalized intersection in a 500 m section, the probability of congestion is high due to signal delays. In regression analysis, it is essential to check whether the coefficients are significant using the p-values. The p-values of independent variables are less than 0.05, except for X_6 ; the results of step-wise regression analysis also show similar results. Hence, using the five independent variables, namely, X_1 , X_2 , X_3 , X_4 and X_5 , the next step of model validation was performed. Out of the total

924 data points considering all the modes and all the periods together, 75% of the data points were selected randomly using ‘randsample’ command in MATLAB for regression model development and the remaining 25% of the data points were used for validation (Kotagi et al., 2016). The final model developed using 75% of the data points is given below in Eq. (7).

$$\hat{y} = 0.108 + 0.914X_1 - 0.008X_2 + 0.198X_3 + 0.140X_4 + 0.424X_5 \quad (7)$$

Where, \hat{y} = Personal vehicle CI for a particular mode

X_1 = Public transit CI

X_2 = Carriageway width

X_3 = Presence (1) or absence (0) of Signalized intersection

X_4 = 1 if three-wheeler, X_4 = 0 otherwise

X_5 = 1 if car, X_5 = 0 otherwise

The final model shown in Eq. (7) was used to estimate the personal vehicle CI for the remaining 25% of the data points. The estimated personal vehicle CI was then compared with the observed personal vehicle CI using the measure mean absolute error (MAE) which is given by,

$$MAE = \frac{1}{n} \sum |\hat{y} - y| \quad (8)$$

where, \hat{y} is the estimated personal vehicle CI and y is the observed personal vehicle CI. For comparison of the estimated CI values with the observed CI values, mean absolute error (MAE) was used rather than mean absolute percentage error (MAPE). This is due to the fact that the error measures based on percentage such as MAPE have the disadvantage of being infinite or having extreme values when the observed values are close to zero (Hyndman, 2006). In the present study, the CI values were in the range of 0 to 3.5 with most of the values close to zero and hence the MAPE was not considered. The MAE between the observed and estimated personal vehicle CI was found to be 0.248. The minimum absolute error is 0.0004 and the maximum error is 1.063. The value of MAE is indicative of the performance of the developed model for predicting the personal vehicle CI using only bus probe data. The plot of estimated versus observed personal vehicle CI is shown in Fig. 9. It can be observed that the estimated CI is close to the observed CI, showing good performance of the developed model for estimating the class-wise CI for personal vehicles using only bus probe data.

6 Conclusion

In recent years, more emphasis is being placed on travel time based measures such as congestion index, to quantify traffic congestion. Probe vehicles using GPS are attractive for estimating congestion measures. Due to privacy and other issues, it may not be possible to employ the personal or commercial vehicles as probe vehicles in India, leading to the research

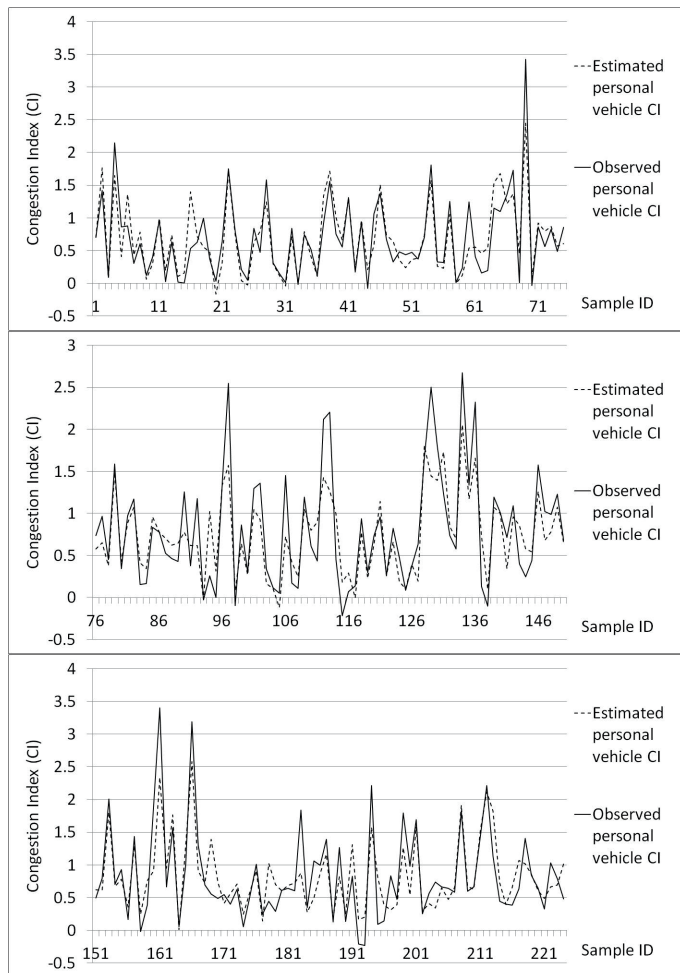


Fig. 9 Plot of Observed versus Estimated Personal Vehicle CI

problem of quantifying stream congestion from bus (serving as probe) data alone using travel time based measures such as congestion index. In order to use bus data for such an application, there is a need to remove the dwell times at bus stops. This has been given a careful attention in the present study by proposing a methodology which can calculate the dwell times with associated acceleration and deceleration times. The bus CI (calculated using the dwell time removed travel time), weighted carriageway width and presence/absence of signalized intersection were considered as independent variables in the model to predict the mode-wise congestion levels under heterogeneous traffic conditions. Three directional bus routes were considered in order to capture the wide range of geometric and traffic characteristics of urban heterogeneous traffic. Both fixed and variable FF(TT) scenarios were taken into account and their influence on CI variation of various modes during different traffic periods were studied. It was found that the assumption of constant FFS/FF(TT) for calculating CI values may not be a realistic one as it exaggerates the congestion level, especially for public transit buses when compared to actual/variable FF(TT) case. Seven regression models were built to take into account the various traffic periods and vehicle types. Wide variations in CI for each class of vehicles was observed,

which indicates the need for defining congestion levels based on modes under heterogeneous traffic conditions as existing in India. The regression coefficients of the developed models are statistically significant and the signs of the coefficients are also logical in terms of their effect on congestion. The validation of the regression model using data splitting approach shows that the average absolute error between the predicted and observed personal vehicle congestion index is 0.248, which is indicative of the performance of the developed model. The proposed methodology for quantifying the congestion levels for vehicles in the stream using only GPS-fitted buses as probes could potentially be used for real-time display of congestion status in ATIS applications in a cheap and effective manner. With wider use of GPS in public transport buses, the availability of bus probe data in a continual and uninterrupted manner renders this approach as an attractive option.

Acknowledgement

This study was made possible partly due to funds made available through Centre of Excellence in Urban Transport at IIT Madras, sponsored by Ministry of Urban Transport (MoUD), Govt. of India.

References

- Anjaneyulu, M. V. L. R., Nagaraj, B. N. (2009). Modelling congestion on urban roads using speed profile data. *Journal of Indian Roads Congress*. 549, pp. 65-74.
- Berkow, M., Wolfe, M., Monsere, C. M. (2008). Using signal system data and buses as probe vehicles to define the congested regime on arterials. In: *CD-ROM Proceedings, Transportation Research Board of the National Academies*. Washington, D.C.
- Beverly, N. (2004). Transportation 2025 - Appendix: Congestion Management Plan. In: *State guide plan element 611*, State of Rhode Island and Providence, USA.
- Chen, C. (2010). Study of indicators of recurrent congestion on urban roadway network based on bus probes. Master's Thesis, The Ohio State University, Columbus, Ohio, USA.
- CMP (2011). Congestion Management Process - State of the System Report - Fiscal Year 2011. Austin, Texas, USA, 2011.
- CMRL (2011). *Chennai Metro Rail News Letter*. Chennai Metro Rail Limited.
- d'Abadie, R. J., Ehrlich, T. (2002). Contrasting time-based and distance-based measures for quantifying traffic congestion levels: Analysis of New Jersey Counties. *Transportation Research Record: Journal of the Transportation Research Board*. 1817, pp. 143-148. <https://doi.org/10.3141/1817-18>
- D'Este, G. M. D., Zito, R., Taylor, M. A. P. (1999). Using GPS to measure traffic system performance. *Computer-Aided Civil and Infrastructure Engineering*. 14(4), pp. 255-265. <https://doi.org/10.1111/0885-9507.00146>
- Geroliminis, N., Skabardonis, A. (2011). Identification and analysis of queue spillovers in city street networks. *IEEE Transactions on Intelligent Transportation Systems*. 12(4), pp.1107-1115. <https://doi.org/10.1109/TITS.2011.2141991>
- Grant, M., Fung, C. (2005). Congestion Management Systems: Innovative Practices. Task 1 report, ICF Consulting Inc., Fairfax, VA, USA.

- Grant, M. (2011). Congestion Management Process: A Guidebook, FHWA-HEP-11-011, ICF Consulting Inc., Fairfax, VA, USA.
- Highway Capacity Manual (2010). Transportation Research Board, Washington, D.C.
- Hyndman, R. J. (2006). Another look at forecast-accuracy metrics for intermittent demand foresight. *International Journal of Applied Forecasting*, 4, pp. 43-46.
- Kotagi, P. B., Raj, P., Gowri, A. (2016). Analysis of lateral placement and movement of vehicles on urban undivided roads in mixed traffic. In: 12th International Conference on Transportation Planning and Implementation Methodologies for Developing Countries (TPMDC) Proceedings, Transportation Systems Engineering, Department of Civil Engineering, IIT Bombay, p. 23.
- Laird, D. (1996). Emerging Issues in the Use of GPS for Travel Time Data Collection. In: National Traffic Data Acquisition Conference (NATDAC '96) Proceedings, 1(96), Alliance for Transportation Research, Albuquerque, New Mexico, pp. 117-123
- Lomax, T. (1997). Quantifying congestion - Volume 1, NCHRP report 398, *Transportation Research Board*, Washington D.C.
- Lovely, L., Madhu, N. (2006). Study on congestion measurement on urban road stretch. *Proceedings of the NCTT 2006 Conference*, Trivandrum, India.
- Maitra, B., Sikdar, P. K., Dhingra, S. L. (1999). Modelling congestion on urban roads and assessing level of service. *Journal of Transportation Engineering*, 125(6), pp. 508-514.
[https://doi.org/10.1061/\(ASCE\)0733-947X\(1999\)125:6\(508\)](https://doi.org/10.1061/(ASCE)0733-947X(1999)125:6(508))
- Patel, S. G., Varia, H. R. (2010). Evaluation of level of service through congestion on urban road links. *Proceedings of the 12th World Conference on Transport Research (WCTR)*, Lisbon, Portugal.
- Quiroga, C. A. (2000). Performance measures and data requirements for congestion management systems. *Transportation Research Part C*, 8(1-6), pp. 287-306.
[https://doi.org/10.1016/S0968-090X\(00\)00008-5](https://doi.org/10.1016/S0968-090X(00)00008-5)
- Richardson, A. J., Taylor, M. A. P. (1978). Travel time variability on commuter journeys. *Journal of Advanced Transportation*, 12(1), pp. 77-99.
- Schrank, D., Lomax, T., Eisele, B. (2011). Urban mobility report. Texas Transportation Institute. The Texas A&M University System, Texas, USA.
- Srinivasan, N. S., Shetty, P. S. (1971). Travel time study in a metropolitan city. *Journal of the Indian Roads Congress*, 34(2), pp. 233-284.
- Taylor, M. A. P., Woolley, J. E., Zito, R. (2000). Integration of the global positioning system and geographical information systems for traffic congestion studies. *Transportation Research Part C*, 8, pp. 257-285.
[https://doi.org/10.1016/S0968-090X\(00\)00015-2](https://doi.org/10.1016/S0968-090X(00)00015-2)
- Tong, D., (2015). Traffic information deriving using GPS probe vehicle data integrated with GIS. [Online]. Available from: <http://www.gis-t.org/files/13auY.pdf> [Accessed: 10th January 2018]
- Turner, S. M., Eisele, W. L., Benz, R. J., Holdener, D. J. (1998). Travel time data collection handbook, Technical Report FHWA-PL-98-035, Federal Highway Administration, USA.
- Yi, P., Xin, C., Zhao, Q. (2001). Implementation and field testing of characteristics-based intersection queue estimation model. *Networks and Spatial Economics*, 1(1), pp. 205-222.
<https://doi.org/10.1023/A:1011589313578>