Research Article

Data Fusion Based Hybrid Approach for the Estimation of Urban Arterial Travel Time

S. P. Anusha, R. A. Anand, and L. Vanajakshi

Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600036, India

Correspondence should be addressed to L. Vanajakshi, lelitha@iitm.ac.in

Travel time estimation in urban arterials is challenging compared to freeways and multilane highways. This becomes more complex under Indian conditions due to the additional issues related to heterogeneity, lack of lane discipline, and difficulties in data availability. The fact that most of the urban arterials in India do not employ automatic detectors demands the need for an effective, yet less data intensive way of estimating travel time. An attempt has been made in this direction to estimate total travel time in an urban road stretch using the location based flow data and sparse travel time data obtained using GPS equipped probe vehicles. Three approaches are presented and compared in this study: (1) a combination of input-output analysis for mid-blocks and Highway Capacity Manual (HCM) based delay calculation at signals named as base method, (2) data fusion approach which employs Kalman filtering technique (nonhybrid method), and (3) a hybrid data fusion HCM (hybrid DF-HCM) method. Data collected from a stretch of roadway in Chennai, India was used for the corroboration. Simulated data were also used for further validation. The results showed that when data quality is assured (simulated data) the base method performs better. However, in real field situations, hybrid DF-HCM method outperformed the other methods.

1. Introduction

Characterization of traffic systems is complex in nature due to the dynamic interaction between the system components, namely, the vehicles, road, and the road users. The uncertainties associated with human behavior makes the system more complex making modeling of the system a challenging task. Estimation and prediction of various parameters associated with this system is also difficult due to the associated uncertainties. The usual parameters used for characterizing the system include flow, speed, density, and travel time. The present study is dealing with the estimation of one of these parameters, namely, travel time. To obtain travel time information of all vehicles in a stream by direct measurement is both time consuming and costly, and it is impractical to collect this information from

all the road stretches in a network. Travel time in urban roads experience high short-term variability and hence cannot be measured using point detection. Being a spatial parameter, direct measurement of it needs either vehicle tracking devices or vehicle reidentification feature. However, majority of the vehicle tracking or reidentification techniques available such as automatic vehicle locators (AVL) and automatic vehicle identifiers (AVI) require participation, which limits the sample size. This underlines the need to estimate travel time from other easily measurable location based parameters such as flow and speed and has been an important research topic for many years. However, majority of researches on travel time estimation and prediction were reported for freeways, where traffic flow is not much affected by external factors such as traffic signals and conflicting movements. Travel time estimation and prediction is more complex and challenging on an urban network due to the influence of signals, presence of opposing movements, mid-link sources and sinks, and random fluctuations in travel demand. The situation is grave for Indian conditions because of the additional complexities related to heterogeneity, lack of lane discipline, and nonavailability of a reliable historic data base. Hence, methods which are cost effective and less demanding in terms of data base need to be explored.

Travel time, being a spatial parameter, is difficult to be measured directly from field. Most of the direct travel time measurement techniques such as test car methods or vehicle re-identification are expensive, immature, or involve privacy concerns and hence majority of the studies depend on indirect methods for travel time estimation. Most of the indirect travel time estimation and forecasting methods can be grouped under extrapolation techniques [1], regression models [2, 3], pattern recognition techniques [4], time series analysis [5], use of filtering techniques [6, 7], neural networks [8], methods based on traffic flow theory [9–14], data fusion techniques [15], and combination of above methods [16–19].

Many of the above methods require a good data base and may not be feasible for locations where automated data collection is not yet functional. Under such conditions, methods which demand less amount of data are required. Indian traffic characterised by its heterogeneity and lack of lane discipline poses additional challenges in terms of automated data collection. Most of the existing location based sensors are lane based and will fail under less or no lane disciplined traffic. Thus, accurate measurement of all the traffic parameters automatically is still a difficult task under Indian traffic conditions. Preliminary developments in this area are showing some promise in terms of traffic counts and hence the present study assumes that traffic count is the only location based data available. On the other hand, spatial data collection using GPS is a proven technology and is applicable under Indian conditions too. However, due to less participation, data from only a sample of the entire population, mainly from public transit, can be obtained using this technology. Thus, there is a need to have multiple sensors to characterise the entire traffic stream.

The present study develops a methodology for estimating stream travel time for an urban arterial using flow data obtained from location based sensors and GPS data obtained from limited number of probe vehicles. This approach known as data fusion (DF) is not explored under Indian traffic conditions. To improve the estimation accuracy, a hybrid DF-HCM method using data fusion for mid-block sections and HCM approach for the delay calculation at intersection is attempted. To compare the performance, a base method which employs input-output analysis for mid-blocks and HCM for intersection is also carried out. The usefulness of analysing separately the delay at signals is tested by comparing with the total travel time till intersection being estimated by using the nonhybrid method which employs data fusion alone for the whole stretch. A brief literature review on these approaches is given below.

Data fusion is a broad area of research in which data from several sensors are combined to provide comprehensive and accurate information [20]. The advantages of using data fusion include increased confidence, reduced ambiguity, improved detection, increased robustness, enhanced spatial and temporal coverage, and decreased costs [20–22]. The basic idea of data fusion is to estimate parameters by using more than one measurement from different sources or sensors. This may be due to lack of availability of enough data from a single source or to capture the advantages of different data sources. Some specific applications of data fusion in the field of transportation engineering are discussed below.

Kwon et al. [23] proposed a linear regression model for travel time prediction by combining both loop detector and probe vehicle data. They showed that linear regression on current flow, occupancy measurements, departure time, and day of week is beneficial for short-term travel time prediction while historical method is better for long-term travel time prediction. Zhang and Rice [24] used a linear model with varying coefficients to predict the travel time on freeways using loop detector and probe vehicle data. The coefficients vary as smooth functions of departure time. The coefficients have to be estimated offline and stored and after that the model can be used real-time. El-Faouzi et al. [25] put forward a model based on the Dempster-Shafer theory. They used travel time from loop detector and toll collection data to estimate travel time. The model required the likelihood that the data sources are giving the correct data. El-Faouzi [22] carried out a similar work using Bayesian method using travel time data from loop detector and probe vehicle to estimate travel time. The results showed that the travel time estimate using data fusion approach was better than the estimate obtained if the data sources were used individually. Chu et al. [21] used simulated loop detector and probe vehicle data to estimate travel time using a model based approach with Kalman filtering technique. Ivan [26] used the ANN technique to detect traffic incidents on signalized arterials using simulated travel time data from loop detector and probe vehicle data.

Another simple analytical model that uses readily available count data from upstream and downstream ends of a link for the estimation of travel time is the cumulative counts (input-output) method [9, 10, 12–14, 27]. However, a major drawback of the input-output method is its dependency on the accuracy of flow counts for travel time estimation [9–13, 27, 28]. Some of the other reported approaches include traffic flow theory based [11, 29, 30]. It can be seen from the above literature review that majority of the models discussed above are limited to freeways, and it may not be feasible to apply them directly on urban networks without further calibration due to differences in behaviour of traffic on the freeway and urban facilities. Moreover, the models developed for freeways generally provide average travel time for the link as a whole, which may not be a true representation in case of links with intersections, turning movements, and so forth. Thus, for better performance, intersection delays may have to be dealt separately. The present study analyse the validity of this assumption by comparing the accuracy of the estimated travel time with and without considering the intersection separately.

The first and one of the most popular methods for intersection delay estimation was developed by Webster [31] from a combination of theoretical and numerical simulation approaches that became the basis for all subsequent delay models. Modifications to the above model under varying traffic conditions were reported by Miller [32] and Newell [33]. The delay model suggested in Highway Capacity Manual usually known as HCM model [34] is a modified Webster's model incorporating the effect of progression and platooning. Attempts to overcome the assumption of steady-state condition by using time-dependent functions are reported in [35]. Other reported studies include deterministic queueing method [28, 36],

modified input-output technique [18], shock-wave theory based models [37, 38], and the use of Markov Chain processes [39, 40].

Overall, it can be seen that most of the reported studies on travel time and delay estimation used data collected from homogeneous and lane-disciplined traffic, either directly from the field or indirectly through simulation models. The traffic conditions existing in India is complex and different with its heterogeneity and lack of lane discipline. There are only limited studies [7, 41, 42] which addressed heterogeneous traffic characteristics. None of those studies estimated the stream travel time in an urban arterial taking signals into account. Also, lack of automated data collection methods in India makes it difficult to explore many of the statistical, time series, and machine learning techniques which are data driven since a good data base is required for applying such techniques. Thus, the application of a new approach for urban arterial travel time estimation with less data requirement is an area that will be of interest for countries like India and requires additional research and is discussed in this study.

The present study compares the performance of three different travel time estimation methodologies, which uses flow data as the main input. The estimation methodologies include two hybrid methods namely base method and data fusion-HCM and a non-hybrid method. The study stretch consists of a midblock and an intersection. The total travel time of the stretch is considered as the sum of travel time in the midblock section, without being influenced by the intersections, and the travel time at intersection, taking into account the delays at signals. Mid-block travel time is estimated using two approaches, namely, input-output analysis and data fusion approach. Input-output analysis is a popular approach and utilizes the cumulative count at entry and exit to find the travel time of vehicles within the section. HCM method is the most popular approach for estimating delay at signalized intersections. The method of applying input-output analysis for the midblock and HCM method for the intersection to obtain the total travel time of vehicles in the study stretch can be considered as a base approach and is entitled as base method. The other approach presents a data fusion method for mid-block travel time estimation and HCM analysis for the intersection area and will be called as hybrid DF-HCM method. The data fusion approach utilizes the location based flow data and the sparse travel time data obtained from probe vehicles for estimating the travel time of the stream. The total travel time of the stretch is then obtained by summing up the mid-block travel time and delay incurred at the bounding intersection. The necessity for analyzing the intersections separately was validated using a non-hybrid approach, where the total travel time is estimated using data fusion alone for the whole stretch without separating into mid-block and intersection area. In this approach, the total travel time is directly estimated using the data fusion approach without separately analysing the delay at intersection. The data till intersection stop line is used in this case for the data fusion approach assuming that the delay is implicitly captured by the flow and travel time data till stop line. A comparison of these three methodologies is carried out to understand the best method for travel time estimation under heterogeneous traffic conditions. This is one of the first study under Indian conditions that have applied data fusion techniques as well as hybrid technique for the estimation of travel time. The study has illustrated an efficient method to estimate the stream travel time in urban arterials with limited GPS data and location based flow data. The results of the study stressed the necessity of analyzing the intersections separately for more reliable estimates of travel time in urban roads.

2. Data Collection

Under Indian traffic conditions, a ready-to-use data archive is not available and hence the study relied upon field data collected manually and simulated data using VISSIM simulation package for corroborating the estimation methods.

2.1. Field Data

The test bed selected for the present study is a six-lane busy arterial road, namely, Rajiv Gandhi Salai in Chennai, India. Traffic in one direction only was considered for the present study. Data requirements based on the selected methodology included flow data from three locations as shown in Figure 1. The distance between location 1 and location 2, which is before the influence of intersection, is 1.72 km and between location 2 and intersection (location 3) is 0.1 km, making the total section under consideration to be of 1.82 km.

Two pedestrian over bridges were identified to mount the cameras for covering location 1 and 2. The intersection area covering the stop line at location 3 was recorded using a camera mounted on a convenient luminaire support. To ensure that the three-video recordings could be synchronized during replay in the lab, the time clocks in all the three cameras were set to a common time at the start of the data collection. Data were collected over three days for a total of six hours for the complete analysis and another two days for a total of three and a half hours from the mid-block alone for comparing input-output method with data fusion approach.

The required location based data, namely, the flow was collected using videographic technique. Initial snap shots of the traffic inside the study stretch were taken from elevated points to get the initial count of vehicles in the section required for input-output analysis. Photographs taken from entry and exit points along with additional photographs taken from in-between elevated points were required to capture the whole length of the section. Classified flow data at the three-data collection points were extracted manually for two-wheeler, three-wheeler, and four-wheeler categories from the videos in a temporal resolution of one minute. The required flow data were extracted manually due to lack of automated procedures. Travel time data required for validation was also extracted manually from videos by reidentifying vehicles at various locations.

The limited travel time data required for the data fusion model were collected using test vehicles equipped with GPS units. The test vehicles comprised of two cars, two three-wheelers, and two two-wheelers which provided travel time data. These vehicles were travelling back and forth between entry and exit points continuously during the data collection period. Moreover, GPS data were available from route number 5C of the public transport bus passing through the stretch. The GPS raw data included time, latitude, and longitude at every five seconds interval. From this, travel time data were extracted using the software package ArcGIS [43].

Due to the lack of automated video data extraction, all the above field data collection and extraction were carried out manually, which was laborious and time consuming. The data collection procedure required lot of coordination for collecting the video simultaneously from three different locations, along with initial snapshots, and GPS data from all different types of vehicles. An error in video recording even in one location makes the entire data not useful for analysis. Due to these difficulties, it was decided to simulate the field traffic conditions using collected field data, and carry out further analysis using the data generated using the calibrated simulated network. The details of the simulation are detailed below.
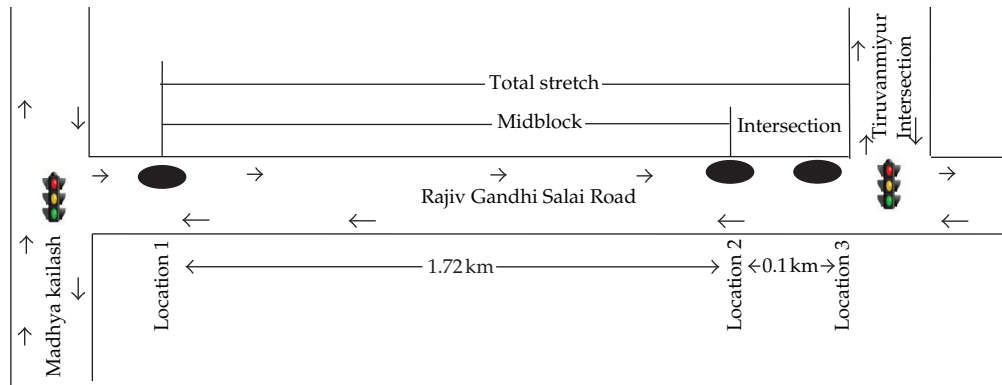
Figure 1: Schematic representation of the study corridor.

2.2. Simulated Data

VISSIM 5.3 from the PTV vision [44] was used in the present study to simulate the traffic conditions for testing the accuracy of the travel time estimation model under varying traffic conditions. A road stretch similar to the field test bed was created in VISSIM using a satellite image. For realistic representation of field conditions, data on intersection geometry, signal timing and phasing, vehicle types, traffic composition, vehicle input, proportion of turning traffic, and speed distribution were entered from field. Less lane disciplined, traffic movement was achieved by placing the vehicles anywhere on the lane, by setting the option for over-taking through left and right side of vehicle and allowing a diamond-shaped queuing at the intersection. To account for the nonstandard vehicles types, static, and dynamic characteristics of most of the regular vehicle types in terms of length, width, acceleration, and deceleration, and speed ranges were defined based on field values.

Signal timing and phase change data from the field were used in the simulation with a total cycle time of 145 s, red time of 98 s, green time 45 s, and amber time 2 s. Classified flow data for two-wheeler, three-wheeler, and four-wheeler categories extracted manually for the three locations is used for calibration and validation of the simulation. Five-minute aggregated flow and composition at location 1 were used as dynamic inputs for calibration. Data generated at the other two locations were compared with the field values for validation.

During calibration, several parameters in VISSIM were adjusted to match the field scenario. Simulations were performed with different random seeds with an average of five to ten values for each influencing parameter. Parameters were calibrated such that the error in flow, density, and travel time/speed was reduced. The errors were quantified in terms of mean absolute percentage error and were comparable with other similar studies in VISSIM [45, 46].

3. Methodology

3.1. Estimation Schemes

As mentioned already, in this study, the total travel time is estimated using a hybrid DF-HCM method making use of data fusion approach for the mid-block and HCM approach for intersection. A comparison is carried out using the base method employing input-output analysis which uses a simple deductive principle of cumulative counts (input-output method) for estimating the link travel time and HCM analysis for the intersection

area. The total travel time of the stretch is then estimated as the sum of mid-block and intersection travel time. Also, the need for analysing intersection delay separately is verified by comparing with a direct estimation of travel time till intersection using non-hybrid method which employs data fusion approach alone for the whole section. The basic approach of the above methods, namely, model based approach using data fusion, input-output method, and HCM approach are discussed below.

3.2. Data Fusion Method

This section details the methodology adopted for fusing both Eulerian video data and Lagrangian GPS data for estimating the travel time. The methodology was motivated from the study of Chu et al. [21]. The estimation scheme is based on the conservation equation and the fundamental traffic flow equation given in (3.1), and (3.2), respectively:

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0, \tag{3.1}$$

$$q = kV, \tag{3.2}$$

where q is the flow in PCU/hour, k is the density in PCU/km, and V is the space mean speed in km/hour, with x being the distance and t being the time.

By discretising (3.1), the density at time t can be represented as

$$k(t) = k(t-1) + \frac{\Delta t \times \{q_{\text{entry}}(t-1,t) - q_{\text{exit}}(t-1,t)\}}{\Delta x}, \tag{3.3}$$

where $q_{\text{entry}}(t-1,t)$ and $q_{\text{exit}}(t-1,t)$ are, respectively, the flow in PCU/h at the entry and exit points during the time interval $(t-1)$ to t. Δt is the data aggregation interval (1 minute in this study).

A filtering technique is used to estimate the density by assuming a value for the initial density, $k(0)$. Then the average travel time taken by the vehicles to reach the exit point from the entry is given by

$$\text{tt}(t) = \frac{\Delta x}{V(t)} = \frac{\Delta x \times k(t)}{q(t-1,t)}, \tag{3.4}$$

where tt(t) is the travel time at time t, $q(t-1,t)$ is the flow along the section during $(t-1)$ to t which is given by

$$q = \begin{cases} q_{\text{entry}}, & \text{if } (q_{\text{exit}} - q_{\text{entry}}) > q_{\text{critical}} \\ q_{\text{exit}}, & \text{if } (q_{\text{entry}} - q_{\text{exit}}) > q_{\text{critical}} \\ \dfrac{q_{\text{entry}} + q_{\text{exit}}}{2}, & \text{if } |q_{\text{entry}} - q_{\text{exit}}| < q_{\text{critical}}. \end{cases} \tag{3.5}$$

Average of the flows at entry and exit points is used under normal conditions, when the flows at both ends are comparable without any shock-wave propagation. When

the flows at the entry and exit are not comparable, minimum of the two was adopted to capture the density variation within the stretch. In the present study, $q_{critical}$ was selected as 20 PCU/minute.

In the above formulation, since the initial density in the section, $k(0)$ is unknown, there is a need for a parameter estimation scheme. The use of techniques such as Kalman Filtering or High Gain Observer (HGO) based parameter identification are reported in literature for similar applications [21, 47, 48]. In the present study, a Kalman filter based estimation scheme is adopted.

The Kalman filter (KF) is a recursive algorithm [49] and is usually applicable to system models which can be written in the state space representation. It is a model based tool for estimation and prediction and incorporates the stochastic nature of parameters. The KF can be of different types such as discrete Kalman filter, extended Kalman filter, and adaptive Kalman filter. The selection of the filter depends on the nature of the governing equations. In the present problem, as state equation (3.3) and the measurement equation (3.4) are linear, the Discrete Kalman Filter (DKF) is used. The present study uses flow data from the video and travel time from limited test vehicles to estimate average stream travel time. The state variable used is traffic density and the travel time is the measurement variable. The state (process) and measurement (observation) equations of DKF can be derived from (3.3) and (3.4) and are given below.

State equation:

$$k(t) = k(t-1) + u(t-1, t) + w(t-1), \tag{3.6}$$

Measurement equation:

$$tt(t) = H(t) \times k(t) + z(t), \tag{3.7}$$

where $u(t-1, t)$ is the input which depends on the flow and is given by

$$u(t-1, t) = \frac{q_{entry}(t-1, t) - q_{exit}(t-1, t)}{\Delta x}, \tag{3.8}$$

$H(t)$ is the transition matrix which converts the density to travel time and is given by

$$H(t) = \frac{\Delta x}{q(t-1, t)}, \tag{3.9}$$

and $w(t-1)$ and $z(t)$ are the process disturbance and the measurement noise, respectively. These are assumed to be Gaussian with zero mean and variances Q and R, respectively.

The Kalman filter algorithm is given by

$$\overline{k}(t) = \widehat{k}(t-1)^+ + u(t-1,t),$$

$$\overline{P}(t) = P(t-1)^+ + Q,$$

$$G(t) = \overline{P}(t)H(t)^T\left[H(t)\overline{P}(t)H(t)^T + R\right]^{-1}, \qquad (3.10)$$

$$\widehat{k}(t)^+ = \overline{k}(t) + G(t)\left[\text{tt}(t) - H(t)\overline{k}(t)\right],$$

$$P(t)^+ = \overline{P}(t) - G(t)H(t)\overline{P}(t),$$

where $\overline{k}(t)$ is the a priori estimate of density calculated using the measurements prior to the instant t and $\overline{P}(t)$ is the a priori error covariance associated with $\overline{k}(t)$. $\widehat{k}(t)^+$ and $P(t)^+$ are the a posteriori density estimate and its covariance, respectively, after incorporating the measurements till time t. $G(t)$ is the Kalman gain which is used in the correction process.

The above steps are repeated at every time steps, and the correction step was carried out only at the intervals when a measurement of GPS travel time is available.

3.3. Input-Output Method

The input-output (cumulative count) method as given by Nam and Drew [50] involves constructing the cumulative vehicle counts (N) on the y-axis and time on the x-axis, as shown in Figure 2.

The classical analytical procedure for travel time estimation considers cumulative flow plots $N(X_1, t)$ and $N(X_2, t)$ at upstream entrance and downstream exit of the link. The total travel time of vehicles during a given time interval, say between t_n and t_{n-1}, is then given by the area between the two curves for that time period, represented by the shaded region in Figure 2. The area can be calculated considering all vehicles that are entering, exiting, or entering and exiting. In this study, all vehicles that are exiting in the time period are considered, and the area is calculated accordingly. Corresponding analytical expression for total travel time (area of trapezoid) is given by

$$T(t_n) = \frac{1}{2}\left[(t_{n-1} - t'') + (t_n - t')\right] \times m(t_n), \qquad (3.11)$$

where, t' = time of entry of the last vehicle that exits the link during t_{n-1} to t_n, t'' = time of entry of the first vehicle that exits the link during the t_{n-1} to t_n, $m(t_n)$ = the total number of vehicles that exit the link during t_{n-1} to t_n.

Under the first-in first-out condition $m(t_n)$ can be given as

$$m(t_n) = Q(X_2, t_n) - Q(X_2, t_{n-1}), \qquad (3.12)$$

where, $Q(X_2, t_n)$ = cumulative number of vehicles exiting at t_n, $Q(X_2, t_{n-1})$ = cumulative number of vehicles exiting at t_{n-1}.
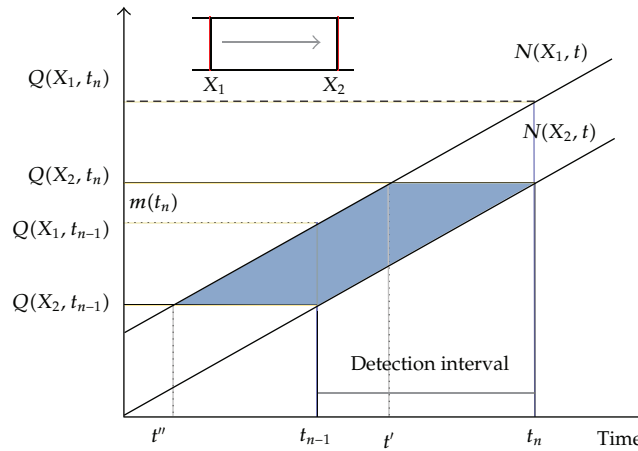
Figure 2: Illustration of input output analysis.

Interpolating the values of t' and t'' and substituting them in (3.11), the total travel time $T(t_n)$ can be calculated. The average travel time of vehicles exiting the link during the given interval ($\overline{TT}$) is then obtained by dividing the total travel time $T(t_n)$ by the number of vehicles exiting the link for the same period as $T(t_n)/m(t_n)$ where $m(t_n)$ is the number of vehicles that exit during the interval.

3.4. HCM Delay Method

HCM delay method [34] is for estimating delay at an intersection over a given time period. Using this method, the average delay per vehicle for a lane group can be calculated using (3.13).

$$d = (d_1 \times f_{PF}) + d_2 + d_3,$$

$$d_1 = \frac{0.5 \times C[1 - g/C]^2}{1 - [\min(1, X)(g/C)]},$$

$$d_2 = 900 \times T\left[(X - 1) + \sqrt{(X - 1)^2 + \frac{8kIX}{cT}}\right],$$

$$f_{PF} = \frac{(1 - P)f_p}{1 - g/C},$$

(3.13)

where, d = average overall delay per vehicle (seconds/vehicles); d_1 = uniform delay (s/veh); d_2 = incremental or random delay (s/veh); d_3 = residual demand delay or initial queue delay (s/veh); PF = progression adjustment factor; X = volume to capacity ratio of the lane group; C = traffic signal cycle time (seconds); c = capacity of the lane group (veh/h); g = effective green time for the through lane group (seconds); T = duration of analysis period (hours); k = incremental delay factor (0.50 for pre timed signals); I = upstream filtering/metering adjustment factor (1.0 for an isolated intersection); P = proportion of vehicles arriving during the green interval; f_{PF} = progression adjustment factor.

The above delay calculation of HCM requires flow values at location 3, free flow running time between the location 2 and location 3, cycle timings, capacity of the lane group, vehicle arrival type, and progression adjustment factor as input values. Out of these, the flow values and cycle timings were directly obtained from the field. The capacity of the lane group was calculated from the field using saturation flow rate and green cycle time ratio as given by $c = s \times g/C$ [34]. A value of 3564 vehicles per hour was obtained as capacity value which closely matched the value given in IRC: 106-1990 [51] for three-lane arterial road. Hence, the standard value of 3600 vehicles per hour as per IRC: 106-1990 for three-lane arterial road was taken for analysis. HCM Exhibits 16-11 and 16-12 [34] were used to determine the arrival type and progression adjustment factor for the known volume condition and vehicle distribution over green time. The value corresponding to arrival type 4 and green cycle time ratio of 0.3 was chosen for the progression adjustment factor.

The additional free flow running time for the delay stretch (constant value for a particular link) is added to the estimated delay to obtain the total travel time between the two locations. The free flow running time is obtained by dividing the distance between locations 2 and 3, $L_{a,s}$ by the free flow speed ff_s of the stretch. Thus, the total travel time of the delay stretch is obtained as

$$\overline{\mathrm{TT}} = d + \frac{L_{a,s}}{\mathrm{ff}_s}, \tag{3.14}$$

where, $\overline{\mathrm{TT}}$ is the total travel time of the delay stretch and d is the estimated delay value.

The total link travel time of 1.82 km stretch is then computed as the sum of mid-block travel time (1.72 km) and the travel time in the delay stretch (0.1 km). Each of the modules of base, hybrid DF-HCM, and non-hybrid method is corroborated using field data and simulated data and is detailed in the section below.

4. Corroboration of the Estimation Scheme

In order to evaluate the performance of the above methods, the estimated mean link travel times using these methods were plotted against the actual travel times using both field data and simulated data. The actual travel time data required for the validation of the results while using field data were obtained through GPS equipped test vehicles as well as by manually re-identifying vehicles from videos. GPS data were collected using three cars, three auto rickshaws, two motor bikes, and five buses as representative samples of each classification. Throughout the data collection period, these GPS test vehicles, except the buses, were made to travel along the study section repeatedly. Figures 3, 4, and 5 shows the plots of travel times predicted by the three methods compared to the actual travel time from the field data and simulated data. It can be seen that the hybrid DF-HCM model is able to capture the variations in the actual travel time better than the other two methods.

The errors in travel time estimation in the above cases were quantified using mean absolute percentage error (MAPE) and root mean squared error (RMSE). The mean absolute percentage error is obtained using

$$\mathrm{MAPE} = \left(\frac{1}{N} \sum_{k=1}^{N} \frac{|N_{\mathrm{meas}}(k) - N_{\mathrm{est}}(k)|}{N_{\mathrm{meas}}(k)} \right) \times 100\%, \tag{4.1}$$

where $N_{\mathrm{est}}(k)$ and $N_{\mathrm{meas}}(k)$ are the estimated and the measured average travel time of the study stretch during the kth interval of time with N being the total number of time intervals.
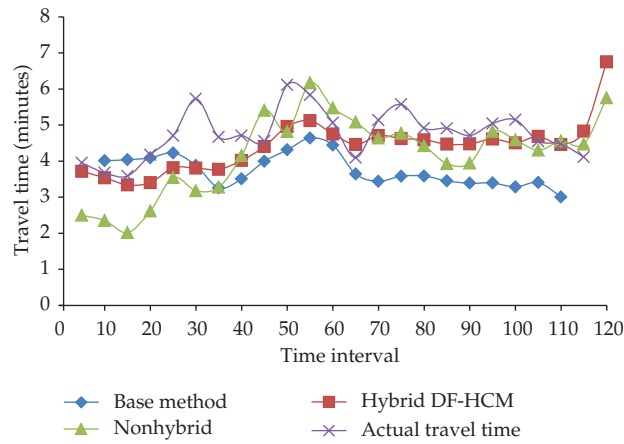
Figure 3: Sample travel time results using field data.
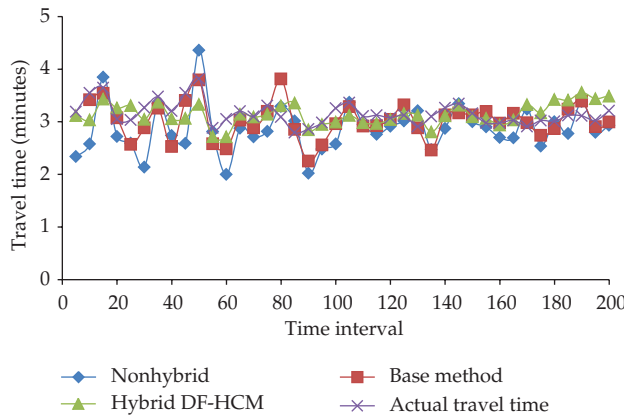

Figure 4: Travel time comparison for a 2 hrs simulated data.
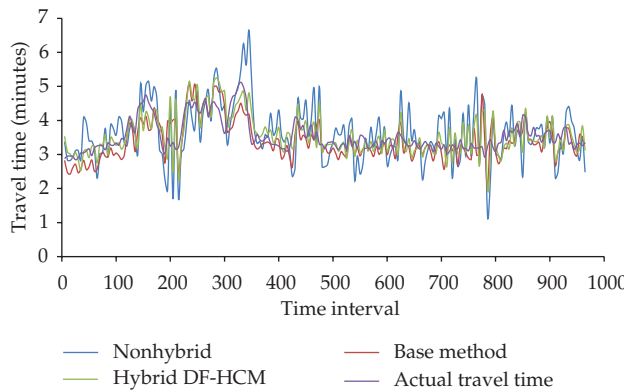

Figure 5: Sample plot for travel time comparison from 6:00 A.M–10:00 P.M simulated data.

MAPE meets most of the criteria required for a summary measure such as measurement validity, reliability, ease of interpretation, clarity of presentation, and support of statistical evaluation. However, as noted by most researchers [52], the distribution of MAPE is often asymmetrical or right skewed and undefined for zero values. Hence, a scale dependent measure called root mean square error (RMSE) is also used which is often helpful when different methods applied to the same set of data are compared. However, there is no absolute criterion for a "good" value of any of the scale dependent measures as they are on the same scale as the data [53]. The lesser the value of RMSE, the better is the forecast obtained. RMSE expresses the expected value of the error and has the same unit as the data which makes the size of a typical error visible. the root mean square error is given by

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[N_{\text{meas}}(k) - N_{\text{est}}(k)]^2}, \tag{4.2}$$

where $N_{\text{est}}(k)$ and $N_{\text{meas}}(k)$ are the estimated and the measured average travel time of the study stretch during the kth interval of time with N being the total number of time intervals.

The MAPE and RMSE values obtained are shown in Figures 6 and 7 and can be seen that the errors are within acceptable ranges [52, 53]. According to Lewis' scale of judgment of forecasting accuracy [49], any forecast with a MAPE value of less than 10% can be considered highly accurate, 11%–20% is good, 21%–50% is reasonable, and 51% or more is inaccurate.

It can be observed that in the case of field data, hybrid DF-HCM method performed better than the other two methods. The results using simulated data show both base and hybrid DF-HCM methods performing comparably and both performing better than the non-hybrid method. Thus, the results clearly show that analysing the delays at intersections separately brings in more accuracy to travel time estimation. Also, it can be observed that input-output method is too constrained by the flow data quality and can be used only when the flow data accuracy is guaranteed such as from simulation. This is further checked by comparing the performance of two more days of field data for mid-block section, and the MAPE and RMSE is as shown in Table 1. It can be seen that in these cases, the data fusion method outperformed the input-output method unlike the case of using data from simulation confirming that with uncertainty in flow values as obtained from field, the data fusion method outperforms the input-output approach.

Overall, it can be observed that the proposed data fusion method is a better candidate for travel time estimation compared to input-output method in the mid-block sections. Input-output method can be considered in cases where the accuracy of flow values is guaranteed. Also, it can be clearly observed that, analysing the intersection delay separately brings in more accuracy to the travel time estimation. This leads to the conclusion that intersection delay needs to be analysed separately to determine the total travel time of urban arterials. Thus, the proposed data fusion method can be used in mid-block sections along with HCM method for delay estimation at intersections to determine the total travel time of urban arterials.

5. Summary and Conclusions

The negative impacts of growth in vehicular population include congestion and delays, and are much debated topics currently all over the world. India, with its rapid growth in
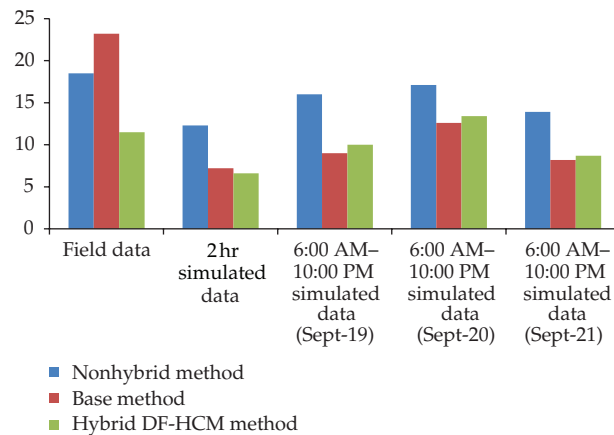
Figure 6: Performance of models in terms of MAPE for field and simulated data.
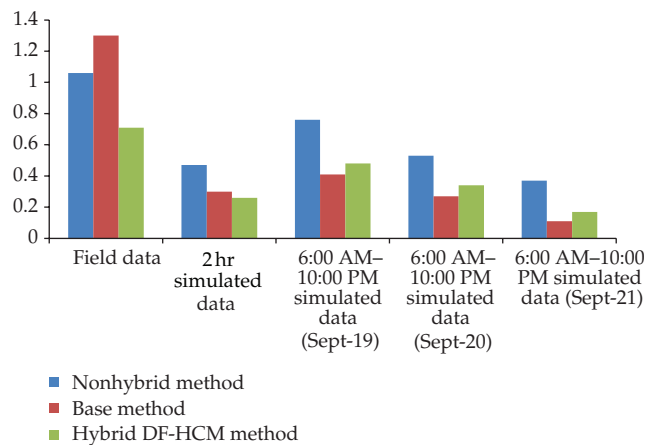

Figure 7: Performance of models in terms of RMSE for field and simulated data.

economy and corresponding growth in vehicular population, is no exception to this. The heterogeneous nature of traffic and lack of lane discipline makes these issues more complex. Characterization and analysis of this type of traffic existing in developing countries demand a different approach than what is followed in western countries with homogenous and lane disciplined traffic. While dealing with urban arterials, there are additional challenges due to the high variability in traffic characteristics. In countries such as India, another challenge is in terms of data availability and hence less data intensive methods for travel time estimation are to be employed.

The present study analysed the application of data fusion technique for travel time estimation in an Indian urban arterial. The study attempted a hybrid DF-HCM method and compared its performance with a base method and a non-hybrid data fusion model to obtain the total travel time of the whole section. The performance of the models for varying traffic flow conditions was tested using field and simulated data. The results showed hybrid DF-HCM model as the best candidate for travel time estimation in urban arterials. The non-hybrid method was found to have the highest error stressing the need for analysis

Table 1: Performance of models in terms of MAPE and RMSE for mid-block.

Method/Date	MAPE (%)		RMSE (min)	
	Input-output	Data fusion	Input-output	Data fusion
June 30	21.9	18.9	0.132	0.096
Sept 23	13.6	9.0	0.28	0.12

of the intersections separately for better performance. Among the hybrid models, data fusion method gave promising results under field conditions. When the accuracy of flow value was guaranteed, such as using simulated data, both the base and hybrid DF-HCM methods showed comparable performance with a slightly better performance from the input-output method. Hence, for real-time field implementations such as ATIS for urban arterials, the hybrid approach using data fusion for mid-block sections along with separate delay estimation at signalized intersections brings in the maximum accuracy of the predicted travel time information showing its potential for any such real time ITS implementations.

Acknowledgments

References

[1] S. M. Turner, W. L. Eisele, R. J. Benz, and D. J. Holdener, "Travel time data collection handbook," Tech. Rep. FHwA-PL-98-035, Federal Highway Administration, Washington, DC, USA, 1998.

[2] V. P. Sisiopiku, N. M. Rouphail, and A. Santiago, "Analysis of correlation between arterial travel time and detector data from simulation and field studies," *Transportation Research Record*, no. 1457, pp. 166–173, 1994.

[3] H. M. Zhang, "Link-journey-speed model for arterial traffic," *Transportation Research Record*, no. 1676, pp. 109–115, 1999.

[4] S. Robinson and J. W. Polak, "Modeling urban link travel time with inductive loop detector data by using the k-NN method," *Transportation Research Record*, no. 1935, pp. 47–56, 2005.

[5] J. Anderson and M. Bell, "Travel time estimation in urban road networks," in *Proceedings of the 1997 IEEE Conference on Intelligent Transportation Systems, ITSC*, pp. 924–929, November 1997.

[6] W. H. Lin, A. Kulkarni, and P. Mirchandani, "Short-term arterial travel time prediction for advanced traveler information systems," *Journal of Intelligent Transportation Systems*, vol. 8, no. 3, pp. 143–154, 2004.

[7] L. D. Vanajakshi, S. C. Subramanian, and R. Sivanandan, "Short-term prediction of travel time for Indian traffic conditions using buses as probe vehicles," in *Proceedings of the 87th Annual Meeting on Transportation Research Board*, Washington, DC, USA, 2008.

[8] D. Park, L. R. Rilett, and G. Han, "Spectral basis neural networks for real-time travel time forecasting," *Journal of Transportation Engineering*, vol. 125, no. 6, pp. 515–523, 1999.

[9] D. H. Nam and D. R. Drew, "Traffic dynamics: method for estimating freeway travel times in real time from flow measurements," *Journal of Transportation Engineering*, vol. 122, no. 3, pp. 185–191, 1996.

[10] D. H. Nam and D. R. Drew, "Analyzing freeway traffic under congestion: traffic dynamics approach," *Journal of Transportation Engineering*, vol. 124, no. 3, pp. 208–212, 1998.

[11] A. Skabardonis and N. Geroliminis, "Real-time estimation of travel times on signalized arterials," in *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, pp. 387–406, 2005.

[12] K. Krishnan, *Travel time estimation and forecasting on urban roads [Ph.D. thesis]*, Imperial College, London, UK, 2008.

[13] L. D. Vanajakshi, B. M. Williams, and L. R. Rilett, "Improved flow-based travel time estimation method from point detector data for freeways," *Journal of Transportation Engineering*, vol. 135, no. 1, pp. 26–36, 2009.

[14] A. Bhaskar, E. Chung, and A. G. Dumont, "Analysis for the use of cumulative plots for travel time estimation on signalized network," *International Journal of Intelligent Transportation Systems Research*, vol. 8, no. 3, pp. 151–163, 2010.

[15] K. Choi and Y. Chung, "A data fusion algorithm for estimating link travel time," *ITS Journal*, vol. 7, no. 3-4, pp. 235–260, 2002.

[16] J. You and T. J. Kim, "Development and evaluation of a hybrid travel time forecasting model," *Transportation Research C*, vol. 8, no. 1–6, pp. 231–256, 2000.

[17] C. Xie, R. L. Cheu, and D. H. Lee, "Improving arterial link travel time estimation by data fusion," in *Proceedings of the 83rd Annual Meeting on Transportation Research Board*, Washington, DC, USA, 2004.

[18] A. Sharma, D. M. Bullock, and J. A. Bonneson, "Input-output and hybrid techniques for real-time prediction of delay and maximum queue length at signalized intersections," *Transportation Research Record*, no. 2035, pp. 69–80, 2007.

[19] J. Bonneson, A. Sharma, and D. Bullock, "Measuring the performance of automobile traffic on urban streets," NCHRP Report, Transportation Research Board of the National Academies, Washington, DC, USA, 2008.

[20] D. J. Dailey, P. Harn, and P. J. Lin, "ITS data fusion," ITS Research program Final Research Report, Washington State Transportation Commission, Washington, DC, USA, 1996.

[21] L. Chu, J. Oh, and W. Recker, "Adaptive kalman filter based freeway travel time estimation," in *Proceedings of the 84th Annual Meeting on Transportation Research Board*, Washington, DC, USA, 2005.

[22] N. E. El-Faouzi, "Bayesian and evidential approaches for traffic data fusion," in *Proceedings of the 85th Annual Meeting on Transportation Research Board*, Washington, DC, USA, 2006.

[23] J. Kwon, B. Coifman, and P. Bickel, "Day-to-day travel-time trends and travel-time prediction from loop-detector data," *Transportation Research Record*, no. 1717, pp. 120–129, 2000.

[24] X. Zhang and J. A. Rice, "Short-term travel time prediction," *Transportation Research C*, vol. 11, no. 3-4, pp. 187–210, 2003.

[25] N. E. El-Faouzi, L. A. Klein, and O. De Mouzon, "Improving travel time estimates from inductive loop and toll collection data with dempster-shafer data fusion," *Transportation Research Record*, no. 2129, pp. 73–80, 2009.

[26] J. N. Ivan, "Neural network representations for arterial street incident detection data fusion," *Transportation Research C*, vol. 5, no. 3-4, pp. 245–254, 1997.

[27] C. F. Daganzo, *Fundamentals of Transportation and Traffic Operations*, Pergamon-Elsevier, Oxford, UK, 1997.

[28] A. D. May, *Traffic Flow Fundamentals*, Prentice Hall, Englewood Cliffs, NJ, USA, 1990.

[29] R. Pueboobphan and T. Nakatsuji, "Real-time traffic state estimation on urban road network: the application of unscented Kalman filter," in *Proceedings of the Ninth International Conference on Applications of Advanced Technology in Transportation*, pp. 542–547, usa, August 2006.

[30] H. X. Liu, W. Ma, X. Wu, and H. Hu, "Real time estimation of arterial travel time under congested conditions," *Transportmetrica*, vol. 8, no. 2, pp. 87–104, 2012.

[31] F. V. Webster, "Traffic signal settings," Road Research Technical Paper 39, Department of Scientific and Industrial Research, London, UK, 1958.

[32] A. J. Miller, "Settings for fixed-cycle traffic signals," *Operational Research Quarterly*, no. 14, pp. 373–386, 1963.

[33] G. F. Newell, "Stochastic delays on signalized arterial highways," in *Transportation and Traffic Theory*, M. Koshi, Ed., pp. 598–598, Elsevier Science Publishing, 1990.

[34] HCM, *Highway Capacity Manual*, Transportation Research Board, Washington, DC, USA, 2000.

[35] R. M. Kimber and E. M. Hollis, "Traffic queues and delays at road junctions," TRRL Laboratory Report 909, Transport and Road Research Laboratory (TRRL), London, UK, 1979.

[36] D. W. Strong, R. M. Nagui, and K. Courage, "New calculation method for existing and extended HCM delay estimation procedure," in *Proceedings of the 87th Annual Meeting Transportation Research Board*, Washington, DC, USA, 2006.

[37] S. C. Wirasinghe, "Determination of traffic delays from shock-wave analysis," *Transportation Research*, vol. 12, no. 5, pp. 343–348, 1978.

[38] W. B. Cronje, "Analysis of existing formulas for delay, overflow and stops," *Transportation Research Record*, no. 905, pp. 93–95, 1983.

[39] W. Brilon and N. Wu, "Delays at fixed-time traffic signals under time-dependent traffic conditions," *Traffic Engineering and Control*, vol. 31, no. 12, 1990.

[40] F. Viti and H. V. Zuylen, "Markov mesoscopic simulation model of overflow queues at multilane signalized intersections," in *Proceedings of the Joint Conference 10th EWGT Meeting & 16th Mini-EURO Conference*, Warsaw, Poland, 2005.

[41] Y. Ramakrishna, P. Ramakrishna, V. Lakshmanan, and R. Sivanandan, "Bus travel time prediction using GPS data," http://www.gisdevelopment.net/proceedings/mapindia/2006student%20oral/mi06stu_84.htm, 2011.

[42] R. P. S. Padmanaban, L. Vanajakshi, and S. C. Subramanian, "Estimation of bus travel time incorporating dwell time for APTS applications," in *2009 IEEE Intelligent Vehicles Symposium*, pp. 955–959, chn, June 2009.

[43] ArcGIS 9.3 manual, ESRI, August 2008.

[44] PTV, *VISSIM Version 5.3 Manual, Innovative Transportation Concepts*, Planung Transport Verkehr AG, Karlsruhe, Germany, 2010.

[45] F. C. Fang and L. Elefteriadou, "Some guidelines for selecting microsimulation models for interchange traffic operational analysis," *Journal of Transportation Engineering*, vol. 131, no. 7, pp. 535–543, 2005.

[46] T. V. Mathew and P. Radhakrishnan, "Calibration of microsimulation models for nonlane-based heterogeneous traffic at signalized intersections," *Journal of Urban Planning and Development*, vol. 136, no. 1, Article ID 005001QUP, pp. 59–66, 2010.

[47] X. Dai, Z. Gao, T. Breikin, and H. Wang, "High-gain observer-based estimation of parameter variations with delay alignment," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 726–732, 2012.

[48] Z. Gao, X. Dai, T. Breikin, and H. Wang, "Novel parameter identification by using a high-gain observer with application to a gas turbine engine," *IEEE Transactions on Industrial Informatics*, vol. 4, no. 4, pp. 271–279, 2008.

[49] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, University of North Carolina, Chapel Hill, NC, USA, 2006.

[50] D. H. Nam and D. R. Drew, "Automatic measurement of traffic variables for intelligent transportation systems applications," *Transportation Research B*, vol. 33, no. 6, pp. 437–457, 1999.

[51] IRC: 106, *Guidelines for Capacity of Urban Roads in Plain Areas*, Indian Roads Congress, New Delhi, India, 1990.

[52] D. L. Kenneth and K. K. Ronald, *Advances in Business and Management Forecasting*, Emerald books, Howard House, London, UK, 1982.

[53] C. D. Coleman and D. A. Swanson, "On MAPE-R as a measure of cross-sectional estimation and forecast accuracy," *Journal of Economic and Social Measurement*, vol. 32, no. 4, pp. 219–233, 2007.