

Digital Object Identifier

# Empirical investigation of fundamental diagrams in mixed traffic

NANDAN MAITI<sup>1</sup>, DR. BHARGAVA RAMA CHILUKURI<sup>2</sup>

<sup>1</sup>Department of Civil Engineering, Indian Institute of Technology, Madras, Chennai-600036, India (e-mail: ce18d706@smail.iitm.ac.in)

<sup>2</sup>Department of Civil Engineering, Indian Institute of Technology, Madras, Chennai-600036, India (e-mail: bhargava@civil.iitm.ac.in)

Corresponding author: Nandan Maiti (e-mail: ce18d706@smail.iitm.ac.in).

“The authors would like to thank the IIT Madras, and MHRD, Government of India for supporting this research.”

**ABSTRACT** A thorough understanding of the fundamental relation of traffic flow variables is critical for the efficient operation of traffic systems. However, their relationships in mixed traffic are challenging to model due to the continuously changing vehicle composition. This paper proposes a composition-based approach for estimating the fundamental relationships between traffic flow variables using empirical data. The methodology seeks to eliminate the difficulties in class-specific *ss* identification by introducing a continuous wavelet transformation with oblique cumulative arrival and oblique occupancy time plots. We used machine learning (ML) algorithms to delineate regimes and showed the fundamental diagrams for a given location that has a composition-invariant free-flow branch but has distinct composition-specific branches in congestion. Also, it was observed that the congested regime (CR) has a wide scatter indicating possible stochastic inter-class interactions for varying vehicular composition. We proposed a distance optimization method to re-cluster the CR data and found that the proposed method improves the fit with the empirical observations. The inter-class interactions result illustrates that the heavy vehicles will dominate the high-speed vehicles with the increase of AO. It is found that beyond a critical level of AO in congestion, all vehicle class travel at the same speed. Finally, it is found that validation with different datasets shows that the proposed methodology is robust in estimating fundamental diagrams under mixed traffic conditions.

**INDEX TERMS** Mixed traffic, Non-lane-based traffic, Machine learning, Fundamental diagram, AO.

## I. INTRODUCTION

TRAFFIC flow models have been developed over the years to understand and mathematically represent the nature of traffic flow and optimize traffic conditions. Fundamental diagrams (FD) that refer to the steady-state (*ss*) relationships between any two of the traffic flow variables of flow, occupancy, and mean speed of traffic stream, are one of the fundamental models used by transportation researchers and engineers. They are helpful in describing both static and dynamic properties of traffic flow such as road capacity, optimal speed, congestion propagation, etc. For example, at signalized intersections, the rate of queue buildup and dissipation due to the cyclic changes in the signal indications can be explained using the ratio of flow change and density change. Similarly, the *ss* flow and densities during the green indication are given by the saturation flow and critical density and during the red indication by the jam density and zero flow.

The FD models are classified into single and multi-regime models. Greenshields (1935) [1] proposed a linear speed-

density single regime model for the traffic in both congestion and free-flow states. Later, Grinberg (1959) [2] proposed an FD, assuming density as a logarithmic speed function. Since then, several researchers have proposed several variations such as exponential relation [3]–[5], and complex functional relation between speed-density relationship [6]–[9]. However, the drawback of these single regime models is the discrepancy between mathematical relations and empirical evidence [10]. To deal with the challenges of the single regime models, researchers developed multi-regime models where the whole density range is classified with multi-regions for improved fit with the empirical data. Multi-regime models include two and three regimes to describe traffic conditions. However, one of the major challenges of the model is the selection of boundaries for the regimes [11]–[13].

The traditional FD models were developed under homogeneous vehicle classes and lane-based movement that mimic the traffic behavior in the developed countries. However, the traffic in south Asian and developing countries includes a



**FIGURE 1.** Sample aerial view of the study area representing mixed traffic conditions from SP road, Chennai, India

mix of vehicle classes ranging from two-wheelers, three-wheelers, and four-wheelers, to public transport like buses and non-lane-based movements [14], representing mixed traffic conditions (see Figure 1). As the vehicle dynamics and driving characteristics of mixed traffic are different from that of homogeneous traffic, the FDs' functional form and parameter values can be significantly different. Moreover, *ss* identification to estimate the speed-density relationship for mixed traffic is also a challenging task, as it deals with *ss* across multiple classes of vehicles.

Another challenge in modeling mixed traffic is the proper characterization of the traffic state. The traditional flow and density need extra information like composition that defines the properties of vehicle classes in the traffic stream. To overcome this, passenger car equivalent (PCE) [15]–[19], area defined density [20], [21], and AO [22] concepts were developed. Multi-class models have been proposed as an alternative to traditional density measurements to address the challenges of PCE [23]–[25].

In multi-class models, the vehicles are generally subdivided into classes based on their kinematic and physical features. Jin (2002) [26] developed a multi-class model in which vehicle classes are grouped based on their free-flow speed. Vehicles with identical free-flow speeds are grouped into one vehicle class. An example of such a multi-class model is the Logghe & Immers (2003) [27] model. In this model, they differentiated different vehicle classes using scale factors. An assumption made is that the fundamental diagram for any class can be a scaled version of a reference fundamental diagram. It is assumed that the mixed traffic flow also conforms to the same fundamental diagram. Heterogeneous multi-class models are developed based on the distinguishing class-specific speed  $V_j(\cdot)$ . Wong et al. (2002) [28] developed a model where the velocity function of each class is a function of total density ( $v_j = V_j(\sum_i \rho_i)$ ). This model differentiated vehicles based on vehicle length. The velocity of each class is distinct, and the model was able to describe various traffic phenomena such as two capacities, hysteresis, and platoon dispersion. Benzoni-Gavage & Colombo (2002) [29] extended it and developed a *n-population* model where vehicle size was taken into consideration. In this model, a system of conservation equations was applied for space occupied by each class of vehicles rather than conserving a number of vehicles. In these heterogeneous multi-class models, speed is defined by the function of effective density and assumes that vehicles are only in static condition when their density reaches jam density ( $\rho_{jam}$ ) or at a common maximum occupied space ( $V_j(r^m) = 0$ ), where  $r^m$  is

maximum occupied space.

In car following, follower and leader vehicle characteristics (size, speed, etc.) affect the following vehicle's decisions, which in turn influences the traffic stream behavior [22]. Congestion propagation wave speed depends on the vehicle size; usually, longer vehicles have larger wave speeds than shorter ones [30]. Vehicle size and its free-flow speed also significantly affect the saturation flow and critical density [31]. Coifman (2015) [32] proposed vehicle length and speed-based bins to calculate the fundamental relations for each bin. However, in the case of a mixed environment, headway calculation will be more difficult due to possibly multiple leaders and their complex influences. Also, in the study, vehicle dynamics of different sizes of vehicles were estimated, but the effect of composition changes was not addressed. Hence, when modeling mixed traffic, it is necessary to incorporate the kinematic and physical characteristics in the models. Traffic composition is one of the essential factors that influence the lateral distribution of vehicles under mixed traffic conditions and the road capacity [22]. The effects of lane width on FD were examined for different compositions of heterogeneous traffic and found that composition influence the capacity of road [33].

Researchers have been exploring composite fundamental diagrams of human-driven vehicles (HVs) with different penetrations of connected autonomous vehicles (CAVs). Few studies follow average stable headway to estimate the composite fundamental diagrams where the combinations of CAVs and HVs were modeled probabilistically [34], [35]. Also, microsimulation-based IDM car-following models [36] were used to model the CAVs and HVs stream with different perceptions and reaction times for analyzing the properties of FDs in different composition levels [37]. However, the complexity of probabilistic methods increases for traffic streams with more traffic classes. Also, these studies deal with vehicles of similar physical characteristics (dimensions) and are primarily based on microsimulation approaches and not empirical data.

It is important to estimate *ss* periods since traffic composition varies during the day, but there is a lack of literature on the subject. Thomas et al. (2012) [38] developed a multi-regime class-wise fundamental relationship for the mixed traffic and investigated the model performance for the impact of vehicle composition and the volume of the traffic stream but did not consider the class-wise *ss* and spatial traffic composition. So, the inter-class interactions were ignored in the fundamental relationship. The study fails to explain how to group the different compositions of vehicles into FDs. The reported study on traffic stream modeling is a very limiting resource for understanding the *ss* phenomena and the interactions of the different vehicle classes in mixed traffic conditions. Also, unlike homogeneous traffic, there are no prior studies to delineate the shape and nature of the fundamental relations in varying compositions of the traffic stream. Most of the existing studies on mixed traffic homogenize traffic conditions using passenger car units. However, it is

well known that such homogenization masks the important intra-vehicular interactions. Also, these studies aggregate the variables across multiple traffic states leading to significant noise in the FDs.

Studies in the literature attempted to estimate composite FDs in the context of connected autonomous vehicles for different penetration rates, they typically considered only two vehicle classes. Moreover, these studies typically used simulation for estimating composition-specific FDs. To overcome these limitations, the study proposes a methodology to systematically identify and estimate the compositions during the *ss* to estimate composition-specific FDs using the class-specific cumulative count curve and occupancy time curve. The proposed model composition-specific fundamental diagrams unveil interesting insights regarding the properties of the free-flow and congested regimes and class-specific interactions in mixed traffic conditions. The major contributions of this paper are listed in more detail below:

- Developed a methodology to identify *ss* duration in the mixed traffic stream.
- Proposed an ML-based and traffic state-based optimization technique for estimating composition-specific congested branches of the FD.
- Formulated composition-specific FDs and analyzed their properties.
- Investigated inter-class interactions in lane-free traffic.

The rest of the paper is organized as follows: Section II presents the data collection techniques using image processing, the data attributes, and the study location. Section III discusses the methodology involving *ss* identifications and AO derivations from class-specific traffic data. Section IV proposes the methodology for FD estimation using unsupervised ML. Section V shows the results of speed-AO-based FD and their property. Finally, Section VI presents the discussion and conclusions. The overview of this paper is shown in Figure 2.

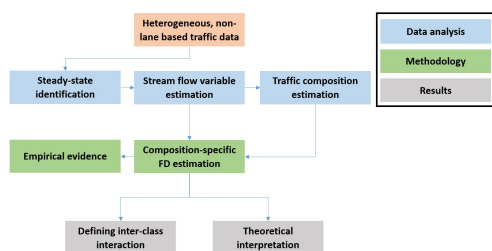


FIGURE 2. Overview of this paper

## II. DATA

To study the FD characteristics in mixed traffic, trajectory data was collected from Sardar Patel Road (SRP), Chennai, India. Data was collected during the evening peak period ranges from 4:30 pm to 6:30 pm on June 6, 2019. Videos were recorded from a vantage point using three cameras, each covering around 70m. From the video, vehicle trajectories

were extracted using the image processing techniques as discussed in [39].

We excluded the trajectory of stopped vehicles, as well as lost-tracking trajectories in the video. After pre-processing, we get 10800 unique vehicles and 432,000 seconds of vehicle trajectories for the study. Unique vehicle trajectories consist of longitudinal and lateral local coordinates with instant speed and acceleration every 1/24 seconds. Among the observed vehicles, the proportion of different vehicles can be estimated from the average classes-wise proportion (see Figure 4b). Figure 1 shows a screenshot of the video data, and Figure 3 is a schematic diagram of the location showing the selected study stretch, where "A" is the data collection loop. The selected study stretch is a six-lane roadway with three lanes in each direction. For the study, westbound traffic flow was considered. The prevailing traffic on this road is highly heterogeneous in vehicle classes with an approximate composition of 52% motorized two-wheelers (TW), 37% passenger cars (car), 6% motorized three-wheelers (auto), and 5% combined proportion of heavy vehicles with bus and truck (BLT<sup>1</sup>) (see Figure 4(b)). Figure 4(a) shows the distribution of vehicle proportions by TW and car class during the *ss* period of data collection time. It represents that proportion of TWs, and cars are more frequent at 0.45 in a *ss* period and follow the symmetric kind distribution. Further, we tested the observed data set using a  $\chi^2$  test with a null hypothesis that the vehicle proportions are normally distributed and found  $\chi^2_{observed} < \chi^2_{critical}$  (see Table 1). Hence, a significant proportion of traffic (%TW+%car  $\geq$  90%) follows normal Gaussian distribution. Traffic volume was obtained by counting the number of vehicles traveling in all lanes at the entry of virtual loop detector A (see Figure 3). The classified space mean speed of all vehicles was measured using the travel time to the detector length. The classified count was carried out for every one-second interval. The extracted data is further analyzed to estimate other traffic flow variables of interest, i.e., density, AO, flow, etc.

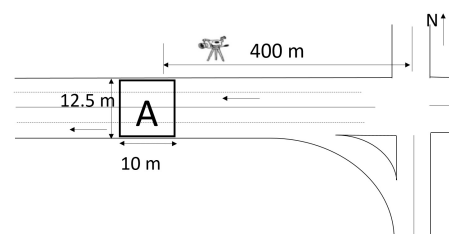


FIGURE 3. Schematic diagram of the selected study stretch

### A. ESTIMATION OF SPACE MEAN SPEED, DENSITY, AND FLOW IN MIXED TRAFFIC CONDITIONS

Longitudinal and lateral position of each vehicle for every 1/24 second is collected, along with the vehicle class information. Suppose 'p' represents the vehicle class, where

<sup>1</sup>bus-lev-truck together:BLT, as their flow is less than 10% in traffic stream

TABLE 1. Chi. sq test for TW and car proportion in the stream

H0: vehicle proportions follow normal Gaussian distribution							
Proportion		Observed		Estimated		Chi.sq	
L. limit	U. limit	TW	Car	TW	Car	TW	Car
0.15	0.2	0	4	1	2		
0.2	0.25	3	1	3	4		
0.25	0.3	10	7	5	7	5	0
0.3	0.35	8	14	8	10	0	1.6
0.35	0.4	8	9	11	12	0.81	0.75
0.4	0.45	14	20	12	13	0.33	3.76
0.45	0.5	8	6	12	11	1.33	2.27
0.5	0.55	5	6	10	8	2.5	0.5
0.55	0.6	12	2	7	5	3.57	1.8
0.6	0.65	5	5	4	3	0.25	1.33
0.65	0.7	2	2	2	1	0.14	0.12
0.7	0.75	1	1	1	1		
df=11				Total chi.sq		13.94	12.15
p<0.05				Critical chi.sq		19.67	19.67

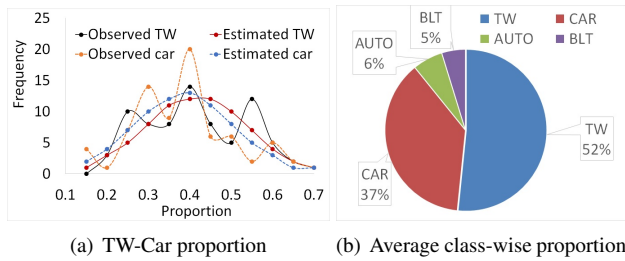


FIGURE 4. (a) Histogram plot of TW, and car proportion during ss (b) Average proportion of TW, car, auto, BLT in several ss periods

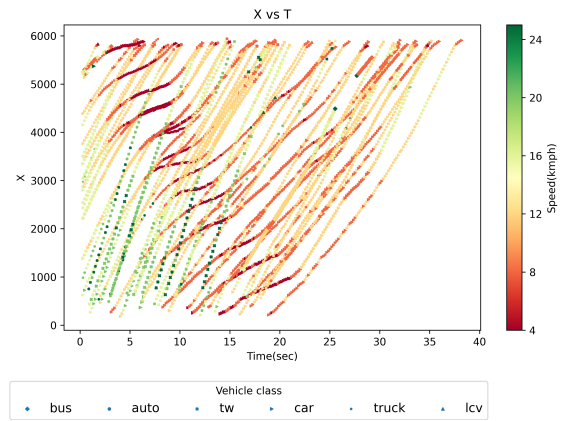


FIGURE 5. Sample trajectory of the mixed vehicles along with the speed label at every instance.

$p \in (TW, car, auto, BLT)$  and  $j, k$  is the starting time and end time of the  $ss$  period and the speed of the vehicle class  $p$  during the  $ss$  period  $j, k$  is  $(v_p^{jk})$  which is the harmonic mean of all the individual vehicle's speed on that particular  $ss$  period, so  $v_p^{jk} = \frac{1}{\sum_{t=j}^k \frac{1}{v_p}}$ . Class-specific density  $(\rho_p^{jk}) = \frac{q_p^{jk}}{v_p^{jk}}$ .

Stream speed during a particular  $ss$ ,  $v^{jk}$ , is the weighted average speed of the density proportion of all classes of vehicles, and the other stream fundamental parameters are estimated as follows.

$$v^{jk} = \sum_p \frac{\rho_p^{jk} v_p^{jk}}{\rho^{jk}}, q^{jk} = \sum_p q_p^{jk}, \rho^{jk} = \sum_p \rho_p^{jk} \quad (1)$$

Sample trajectories of vehicles along with the instantaneous speed and the class of vehicle are shown in Figure 5, which shows urban, multi-class, free-flow, and congested traffic conditions in mixed traffic conditions.

### III. METHODOLOGY

This study focuses on inter-class interactions and their implication for composition-specific FDs. First, we analyze the traffic data to estimate the  $ss$  periods using continuous wavelet transformation (CWT), cumulative count, occupancy plots, and oblique plots of arrival and occupancy. Next, we separate the  $ss$  data using machine learning and traffic state

optimization techniques into different clusters based on their compositions and traffic states. Then, we aggregated all the data points into 0.5-sec intervals to capture temporal variation. Finally, flow-density and flow-occupancy relationships for  $ss$  points were plotted to estimate the class-specific FDs. More details of these steps are given in the subsections below.

#### A. STEADY STATE IDENTIFICATION

Cassidy (1998) [40] proposed oblique cumulative count (cc) and occupancy (co) plots to identify the nearly stationary traffic in homogeneous traffic. However, it may not suit mixed traffic as the inter-class equilibrium is a critical factor. Researchers have used mathematical transformations to study the wave propagation and  $ss$  period in traffic flow using Fast Fourier Transforms [41] and Wavelet transformation (WT) [42]. These methods are very efficient for identifying traffic speed variation over time-series data but heterogeneous, class-specific traffic needs a methodology to apply those transformations in  $ss$  identification. WT assists in identifying

whether all classes of vehicles follow *ss* in mixed traffic conditions or if only some classes do. WT findings will be used in developing a methodology to identify *ss* periods in mixed traffic environments. Therefore, we focus more closely on WT of the trajectories in the next sub-section.

### 1) Wavelet Transformation:

To understand wave propagation behavior in the time-space (X-T) domain of intra-class traffic and interclass traffic in mixed traffic conditions, we take the help of the CWT based energy method. A Wavelet is a wave-like oscillation that is localized in time. The WT is similar to the short-time Fourier transformation (STFT) based on the operations of time and frequency shifts; the basis functions for the decomposition are time and frequency-shifted versions. The CWT is an alternative, but related, decomposition based upon the operations of time shifts and scalings [43]. STFT-produced time-frequency plots do not always clearly locate the stationary time intervals. Whereas a wavelet is a real or complex mathematical function,  $\psi(t, a)$ , that can transform a time series of traffic speed profiles into various scale components to help identify stationary intervals. The properties of Wavelet are shown in (2).

$$E = \int_{-\infty}^{\infty} |\psi(t, a)|^2 dt < \infty, \quad \int_{-\infty}^{\infty} \psi(t, a) dt = 0 \quad (2)$$

Here E is the wavelet-based energy with a finite value, and the mean of the wavelet is zero. As we want to process a long-duration continuous speed profile, the coefficient of CWT is the best method to achieve the goal. From many mother wavelets available (e.g., Haar, Daubechies, Mexican hat, Morlet, Coiflet) for a continuous signal, we choose the Mexican hat wavelet for the time-varying speed data, as others wavelets are near-optimal and provide similar results for a wide variety of signals [44]. Adeli et al. [45], [46] describe more details of WT and its applications. For our study, speed,  $v(t)$  time-series is the continuous signal function. Thus, the WT coefficient of  $v(t)$  can be obtained as follows (3):

$$T(t, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} v(t) \left[ 1 - \left( \frac{t - \tau}{a} \right)^2 \right] \exp \left( -\frac{t - \tau}{2a} \right) dt \quad (3)$$

Then, the average wavelet-based energy at  $\tau$  is computed based on the WT coefficients for different scales, i.e.,

$$E = \frac{1}{\max(\tau)} \int_0^{\infty} |T(t, a)|^2 dt \quad (4)$$

Here  $\tau$  is the time shift parameter i.e.,  $\psi(t, a) \leftarrow \psi(t - \tau, a)$ , is given by:

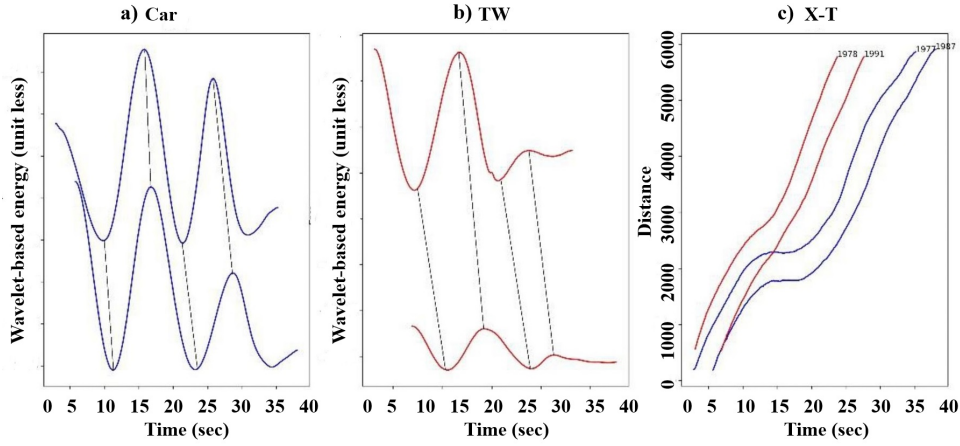
$$\psi(t, a) \leftarrow a_0 \psi(t/a_0, a_0 a), a_0 > 0$$

In this study, we used the detector data to identify the wave propagation of class-specific vehicles. Here,  $v(t)$  is the individual vehicle speed for a sample study period. T(t,a) is computed for the speed value of vehicles with the shape value

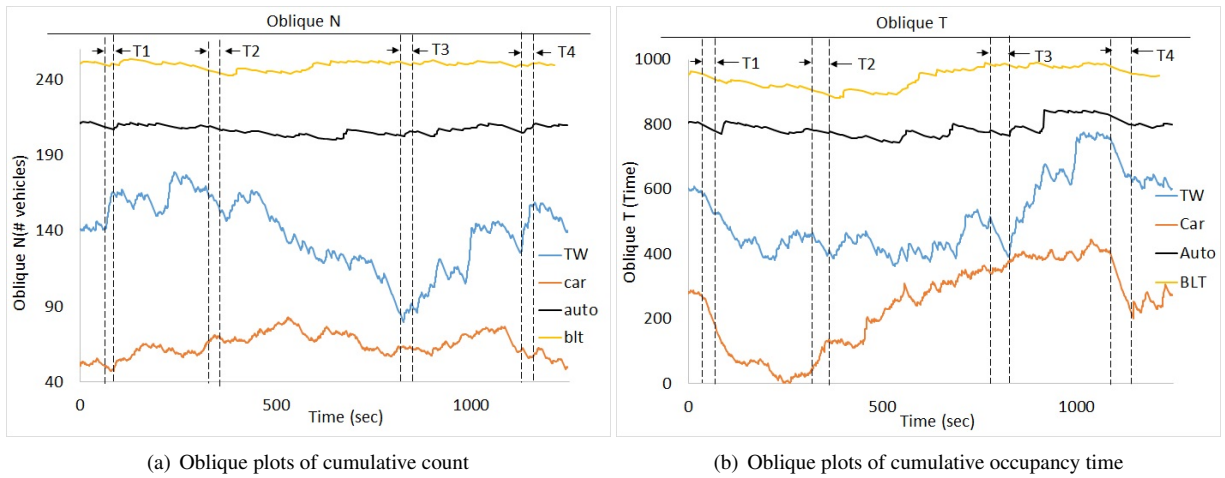
of  $a = 64$ . The temporal wavelet-based energy plots of the TW and the car is shown in Figure 6(a,b). The peaks and dips in the plot imply the abrupt speed change during the travel. In Figure 6, black dotted lines joining through the dips and peaks of energy plots have different slopes for car and TW throughout the travel period. The analysis showed that class-specific wave speeds exist during congestion periods, and they propagate backward, showing that all vehicles have different wave speeds in congestion. Three different detectors were placed in the road section at entry, mid-block, and exit locations where the space mean speed of individual vehicles crossing from the detector was collected with respect to time. Similarly, the time-varying speed profile of three detectors was transformed into a CWT energy plot. There has been a constant change in the slope of peaks and dips connecting lines throughout time and space. Hence, it is concluded that traffic composition greatly affects wave speed propagation and fundamental parameter estimation of the traffic stream.

### 2) Cumulative count and cumulative occupancy curve

To estimate fundamental relationships among the traffic flow variables, i.e., flow ( $q$ ), density ( $\rho$ ), and speed ( $v$ ) in mixed traffic conditions, one needs to identify them during *ss*. Here, we extended Cassidy's [40] methodology to identify the class-specific common *ss* periods in the mixed traffic stream. The first step involves plotting the vehicle class-specific cumulative count curve (N-curve), curves of cumulative vehicle arrival number to  $x$  by time  $t$  for each class,  $N_i(x, t)$ . To visually identify the linear arrival rates of vehicles, oblique plots of arrival ( $N_i(x, t) - q_{i0}t$ ) were constructed. Next, to check the constant rate of speed during the *ss* periods of each vehicle class, cumulative occupancy time (T-curve) and the oblique plots of occupancy time ( $T_i(x, t) - T_{i0}t$ ) were plotted over the detector length. The occupancy time (T-curve) was calculated in the middle of the study section by considering a virtual loop detector of 10 meters in length over the whole cross-section (see Figure 3a). This data is used to develop the oblique plots of the cumulative count and travel time to magnify the fluctuation of the vehicle arrival and travel time during the observation period. Note that  $q_{i0}$  and  $T_{i0}$  represent the average rate of vehicle arrivals and average rate of travel time of vehicle during the period and  $N_i(x, t)$  and  $(T_i(x, t))$ , the cumulative vehicles arrival and the cumulative occupancy time curves for  $i^{th}$  class vehicles. The *ss* periods for each vehicle class are identified when both the oblique N-curve and T-curve have a linear slope, indicating a *ss* period. Thresholds on the minimum duration of 10 seconds and maximum coefficient of variation of the speed of 20 were imposed during the *ss* identification. Finally, the common *ss* periods for all vehicle classes were identified. In Figure 7, the *ss* periods of the stream are illustrated for the time window of 20min for peak hour traffic. It can be seen in Figure 7(a,b) for a few sample time periods  $T_1, T_2, T_3$ , and  $T_4$  where class-specific oblique plot for N-curve and T-curve have a linear slope in the mentioned period. The data from these *ss* periods are collected for further analysis.



**FIGURE 6.** (a) CWT-based energy plots of the class-specific vehicles, black lines joining through peaks, and dips of the energy plots give the wave propagation idea for each class a) car b) two-wheeler, c) x-t Plot shows the trajectories of those vehicles (tw,car) in mixed traffic conditions (blue color: car, red color: two-wheeler).



**FIGURE 7.** Class-specific oblique plots of (a) cumulative count curves and (b) cumulative occupancy time curves.

## B. FUNDAMENTAL DIAGRAM ESTIMATION

The FD represents the  $ss$  relationship between  $q$ ,  $\rho$ , and  $v$ . The critical points of the FD are jam density ( $\rho_{jam}$ ), critical density ( $\rho_{cr}$ ), free-flow speed ( $v_f$ ), capacity ( $q_{max}$ ), and wave speed ( $w$ ). Existing methods for FD estimation are developed for homogeneous traffic conditions. However, vehicle composition is an important factor in characterizing traffic conditions. Density is essentially defined for spatial measurements suitable for lane-based homogeneous traffic. In the recent past, researchers proposed that occupancy is a better variable for highly mixed traffic to incorporate widely varying vehicular dimensions and speed [22]–[24]. However, occupancy may not precisely represent the combined mixed traffic behavior moving in both lateral and longitudinal directions [47]. To deal with these difficulties, the width of the road is considered in occupancy calculation, known as AO.

### 1) Area occupancy estimation

Macroscopic traffic flow characteristics can be described using density, occupancy, and AO. While density and oc-

cupancy are more relevant for homogeneous and lane-based traffic, AO is suitable for heterogeneous and non-lane-based traffic. The AO expresses how long a particular size of the vehicle is over the detector on the road. AO is calculated based on the width of the road irrespective of the number of lanes. Hence, multiple vehicle presence over the road section can be considered. The AO of the traffic stream is the summation of all vehicles available during a particular time. Derivation of AO can be found in [24].

$$\begin{aligned}
 AO &= \sum_{p=1}^m \frac{\sum_{q=1}^n \Delta x w_{pq} \tau_{pq}}{\Delta x WT} \\
 &= \sum_{p=1}^m \frac{\sum_{q=1}^n w_{pq} \frac{(l_{pq} + \Delta x)}{v_{pq}}}{WT} \\
 &= \sum_{p=1}^m \bar{w}_p (l_p + \Delta x) \rho_p, \quad \bar{w}_p = \frac{w_p}{W}
 \end{aligned} \tag{5}$$

So, the AO for a particular period for  $p^{th}$  vehicle class of

$q$  number of the vehicle can be estimated using (5). Density  $\rho_p$ , vehicle length  $l_p$ , and vehicle width  $w_p$  for a particular class is the only independent variables to estimate AO for that class at any instant.  $\tau_{pq}$  is the occupancy time of  $p^{th}$  class  $q^{th}$  vehicle. In this study, the width,  $W$ , of the road section is uniform throughout the detector length, is 12.5 meters. There were only six classes of vehicles available based on the size and vehicle dynamics in the study location, their length ( $l_p$ ) and width ( $w_p$ ) can be found in Table-2. As the frequency of arrivals of bus, truck, and lcv during any time of day are very less compared to the car and TW, we grouped them into one class named BLT class. The average size of BLT class was taken as 2.5m in width and 8.5m in length. The longitudinal length of the detector  $\Delta x$  is taken as 10m. The class-specific AO is calculated from the estimated  $\rho_q$  and using other variables. The total AO for the stream of traffic is the summation of class-specific AO in a  $ss$  period.

TABLE 2. Physical characteristics of Vehicles observed in Indian road ([24])

Veh. Class ( $p$ )	length( $l_p$ )	Width( $w_p$ )
Car	4.7	1.7
lcv	5.0	1.9
TW	1.8	0.6
truck	8.4	2.5
auto	2.6	1.4
bus	10.5	2.5

This section proposes a method for FD estimation using different unsupervised learning algorithms to segregate composition-specific FDs. Large vehicles significantly reduce the speed of smaller vehicles compared to the same AO road section with more small vehicles [48]. Composition change leads to changes in traffic characteristics even in the same density of the same road section. In this paper, we developed a methodology for estimating composition-based FD on the collected traffic data. The collected data set contains a large percentage of TW-class of vehicles. In the highly congested road, when all other classes of vehicles are stopped, TW can move in between the available safe gap of other vehicles, making traffic more volatile during congestion. The congested regime of fundamental diagrams is more scattered (see Figure 8) than the free-flow regime in the mixed traffic stream. It represents that the extensive flow range can be experienced at a particular AO or density in congestion. One possible reason for these widespread data points could be the complex interactions between the available vehicle classes across the different traffic compositions. Segregation of the traffic compositions from the stream will greatly help understand the vehicle interactions in the traffic stream.

## 2) Composition-based clustering

Every  $ss$  traffic stream has a unique composition of classes of vehicles. The composition of vehicles in a traffic stream is described by the proportion of all the available vehicle classes during a particular  $ss$  period. Density proportions give

a better understanding of spatial traffic variability. To investigate the effect of vehicle interactions in the road space, We used density proportion to estimate traffic composition rather than volume proportion. As there are predominantly four classes of vehicles appearing in the stream and the proportion of blt class are always lesser than other classes, the stream composition is expressed as  $(\rho_{TW}/\rho_{BLT} : \rho_{car}/\rho_{BLT} : \rho_{auto}/\rho_{BLT} : 1)$ . This proportion value changes every period. To classify these proportions into several groups, we use an unsupervised ML called Gaussian Mixing Model (*GMM*), which is an expectation-maximization algorithm [49]. The core idea of this model is traffic composition can be modeled by Gaussian Distribution. The proportion distribution of the predominant vehicle class, i.e., TW, and car follows Gaussian normal distribution as shown in Figure 4 (a). So it is quite natural and intuitive to assume that the composition clusters come from different Gaussian Distributions. Or in other words, it tries to model the dataset as a mixture of several Gaussian Distributions. For multivariate Gaussian Distribution, the probability density function is given by (6).

$$G(X|\mu, \sigma) = \frac{1}{\sqrt{(2\pi)^d |\sigma|}} \exp\left(-\frac{1}{2}(X - \mu)^T \sigma^{-1}(X - \mu)\right) \quad (6)$$

Here  $\mu$  is the  $d$ -dimensional vector denoting the mean of the distributions, and  $\sigma$  is the  $(d \times d)$  covariance matrix.  $X$  is the data frames of compositions  $X \in [(\rho_{TW}/\rho_{BLT}, \rho_{car}/\rho_{BLT}, \rho_{auto}/\rho_{BLT})]$ . Suppose in our case for traffic composition, and there are  $K$  clusters based on composition. So,  $\mu$  and  $\sigma$  are also estimated for each  $K$ . These parameters can be estimated by the maximum-likelihood method. Since there are  $K$  such clusters, pdf is defined as the linear function of densities of all these  $K$  distributions, i.e.

$$p(X) = \sum_{k=1}^K \pi_k G(X|\mu_k, \sigma_k)$$

where  $\pi_k$  is the mixing coefficient for the  $k$ -th distribution. In this study, we used a python package of *GaussianMixture* ( $n_{components} = K$ ) [50] and the initial and essential parameters ( $n_{components}, n_{init}, random_{state}$ ) are customized for the package. *GMM* requires the user to specify the number of components or number of clusters ( $n_{components}$ ) before training the model. Here, we used the famous Aikaki Information Criterion (AIC) and the Bayesian Information Criterion (BIC) to aid us in this decision. In our study, both AIC and BIC [51] converge approximately at  $K = 2$  (see Figure 9 (a)). So,  $n_{components}$  or the number of feasible clusters based on the traffic composition is taken as 2. We named this classification Class-A, so the two clusters are A1 and A2. Based on the A1 and A2 clustering FD of flow vs. AO are plotted (see Figure 9(b)).

## 3) Traffic state-based clustering

Composition-specific FDs are scattered plots of  $q-\rho$  or  $q$ -AO over the whole regime. It is not trivial to separate the different

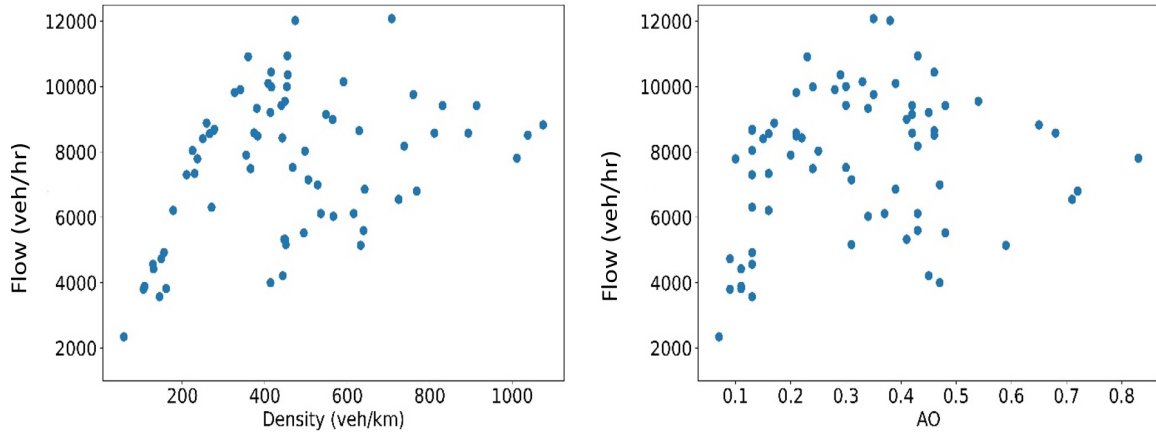


FIGURE 8. AO vs Flow diagram for the *ss* dataset

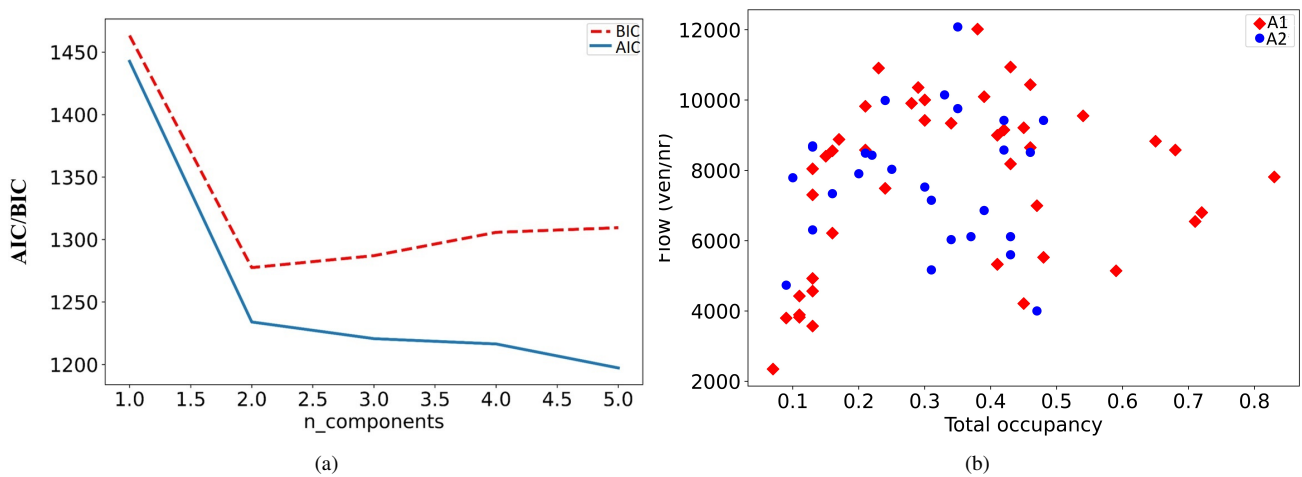


FIGURE 9. (a)  $n_{components}$  estimation for traffic composition using AIC and BIC criterion (b) Composition-based clustering of AO vs. flow plots

traffic regimes in the FDs from the composition-based methods since it is difficult to conclusively identify the transition point or the critical density visually. Therefore, there is a need to design a methodology that can systematically segregate the free-flow and congested regime data. In the traffic-state-based clustering, we used the same procedure as the *GMM* clustering technique, except that  $X$  is the data frames of flow ( $q$ ), *sms* speed ( $v_{sms}$ ), and AO ( $X \in [q, v_{sms}, AO]$ ). It is found that the number of optimal clusters based on the traffic composition is 3. It divided the fundamental diagram into three parts, two clusters in the free-flow regime and a cluster for a congested regime based on the distribution of  $X$ . The classes of this state-based FD are named class-B (B1, B2, B3). Based on the B1, B2, and B3 clustering FD of flow vs. AO is plotted in Figure 10.

#### 4) Composite fundamental diagrams

Based on the above two methods, the whole regime of FD is segregated into two broader groups, i.e., class A and class B. The overall methodology of the FDs estimation using unsupervised ML can be found in a flowchart shown in Figure-13.

Composition-specific FD refers to the line joining through mean points of the class-B group for a unique composition of class-A ( $B \in A$ ). The main idea behind connecting the mean points to construct the composite FD is that a conventional FD always represents the average traffic behavior in the equilibrium conditions. Spreading of the equilibrium points around the mean line infers the acceleration and deceleration of vehicles for a unique composition. We applied k-mean clustering to each of the clusters to identify the centroid or mean point for each group of clusters. k-means algorithm [52] is an iterative algorithm that clusters the dataset into  $K$  pre-defined distinct non-overlapping subgroups (clusters). It is a centroid-based algorithm where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. In this study, the k-mean algorithm takes unlabeled data of each group as input and generates the mean point of the cluster as output ( $k = 1$ ). The proposed best-fit FD is developed as follows:

- Line joining mean points of free-flow regime and origin gives the free-flow speed of the traffic stream for differ-



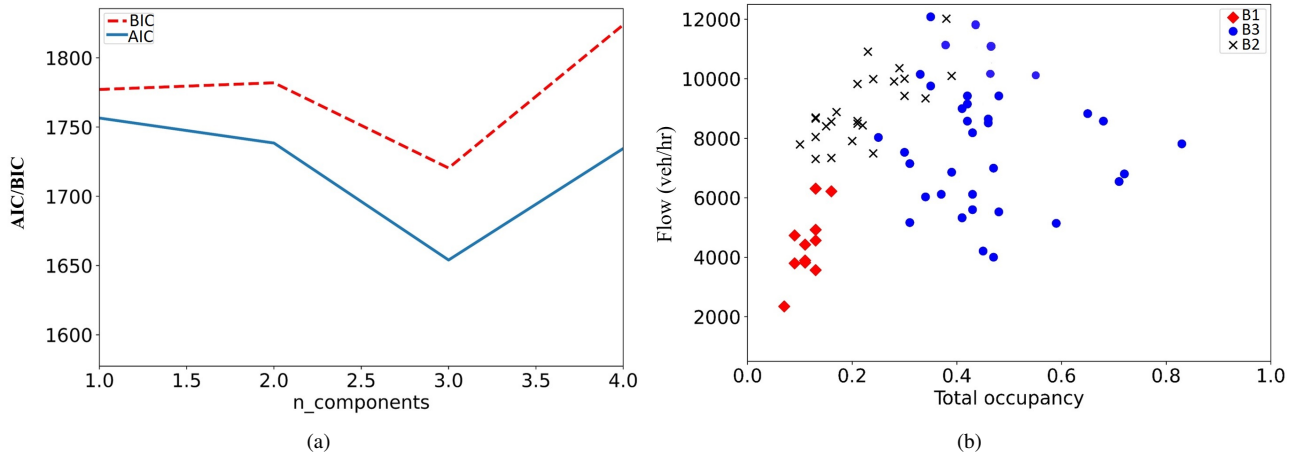


FIGURE 10. (a)  $n_{components}$  number of the optimum cluster for the input (q, AO, v) dataset (b) Traffic state-based clustering of AO vs. flow plots

ent compositions.

- The capacity flow ( $Q_{max}$ ) for each composition-specific FDs can be estimated by intersecting the free-flow line with a line passing through maximum AO point ( $AO = 1.0$ ), and mean points of congested regime's cluster.
- Jam density ( $\rho_{jam}$ ) can be estimated by extending the line from ( $Q_{max}, \rho_{cr}$ ) through the mean points of the congested regime's cluster.
- Line starting from  $\rho = \rho_{jam}$  at zero flow through mean points of the congested regime for each composition gives the congested regime wave speed.
- The meeting point of the free flow speed line and congested wave speed line gives the saturation flow and the critical density for a particular composition of the traffic stream.

The primary purpose of q-AO FDs is to estimate the capacity flow based on the intersection point of the best-fit line in the free-flow regime and the best-fit line in the congested regime. Since the best-fit congested line in the q-AO FDs can be bounded by a maximum value of  $AO = 1$ , it is easy to estimate q-AO FDs for the observed data set. Once the capacity flow is known from Q-AO FDs, q- $\rho$  FDs and jam density can be estimated using free flow speed, capacity flow, and the congested regime mean points. Therefore, all steps are essential for estimating composition-specific FDs from observations.

#### Distance optimization

Classification of congested regime data based on the composition and traffic state showed a promising result with several clusters. However, congested regime data are scattered more than free flow regime data, excluding several points from estimated FDs.

To overcome this, we introduced another level of clustering for congested regime data points based on the distance from the w-line in the direction of speed of each ( $q, \rho$ )

point. w-line is the congested regime wave speed, connecting saturation flow and jam density through the specific composition's mean point, shown in Figure 11. An optimization algorithm based on traffic state has been developed. Points with equal travel speeds are grouped according to the distribution of  $ss$  points in  $q$ - $\rho$  space. The scattered equal travel speed points represent the vehicles moving with unique equal speeds when the flow and density of the stream change. This phenomenon indicates that the scattered points with equal speed have different compositions, and the unique composition points are placed close to each other. The steps for the proposed distance optimization algorithm are explained as follows:

- Lines emanating from the origin and passing through each ( $q, \rho$ ) points have to extend up to the w-lines derived in the previous method.
- Calculate the distance ( $d_i$ ) between the original points ( $q_i, \rho_i$ ) and the extension points at w-line ( $q_{ei}, \rho_{ei}$ ).
- Existing points are classified based on the magnitude of distance from the original to the extension points ( $d_i < d_j$  two points are in different classes)
- Set the ranges of the distance ( $d$ ) from the w-line for both side data to classify them based on their magnitude in the direction of speed to the w-line.
- Estimate revised w-line and repeat the process until the centroid position does not change.

Steps to calculate  $d_i$ :

- Equation of line originating from the origin and passing through ( $q_i, \rho_i$ )

$$q_i = v_i \rho_i$$

here  $v_i$  sms speed of each point.

- Equation of w-line:

$$q_i = -w\rho + c$$

passing through ( $0, \rho_{jam}$ ), so  $c = w\rho_j$ .

- $q_i = v_i \rho_i$  line intersects w-line at ( $q_{ei}, \rho_{ei}$ ) i.e.,  $(\frac{w\rho_{jam}v_i}{v_i+w}, \frac{w\rho_{jam}}{v_i+w})$

- The distance between  $(q_i, \rho_i)$  and  $(q_{ei}, \rho_{ei})$  can be calculated as:

$$d_i = \sqrt{(q_i - q_{ei})^2 + (\rho_i - \rho_{ei})^2}$$

- $w = \operatorname{argmin}_w(\sum d_i)$

Figure 11 describes the distance calculation procedure in detail.

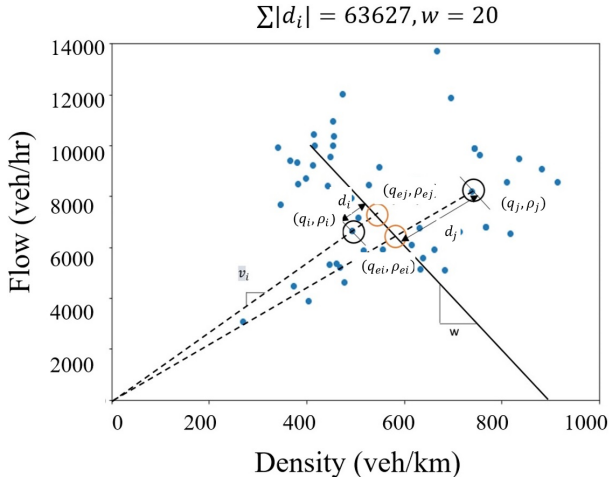


FIGURE 11. Scattered  $(q_i, \rho_i)$  points in the congested regime, solid black line is the congested wave speed which was estimated from the composition-based classification of the traffic stream

Figure 12 shows two iterations of  $w = 12.6$  and  $25$  km/hr with the total sum of the distance of steady state points from w-line. W-line ending at a jam density approximate of  $875$  veh/km (Figure 11) illustrated minimum  $\sum d_i$  than the w-line of 25, and 12.6 (see Figure 12). This method ensures that all the points in the new clusters have the same characteristics of wave speed, composition, and traffic state. The objective function is to minimize the sum of the distance of all points,  $\sum d_i$  for a cluster. This is a performance measure that doesn't have a physical meaning. Therefore, a minimum of  $\sum d_i$  helps to estimate w-line through the mean points of a cluster of unique composition. The objective function is to minimize the sum of the distance,  $\sum d_i$  of all points for a cluster. The minimum of  $\sum d_i$  ensures that the ss points for a unique composition has the same speed (w-line) and the cluster is closely packed in the  $q - \rho$  domain. Thus, the w-line through the optimum clusters is fine-tuned till the convergence of the algorithm. In this study, the optimum w-line for one composition class is shown in Figure 11.

#### IV. RESULTS

This section describes the properties of the composite FDs, developed by combining class-A and class-B data sets. The data set in this study has two composition-based (class-A) and three state-based (class-B) clusters. Hence, there are three groups for each traffic composition, resulting in six groups. It is found that the free flow regime centroids of four data groups lie on a unique slope line, whereas two groups in the congested regime had distinct means as can be seen from

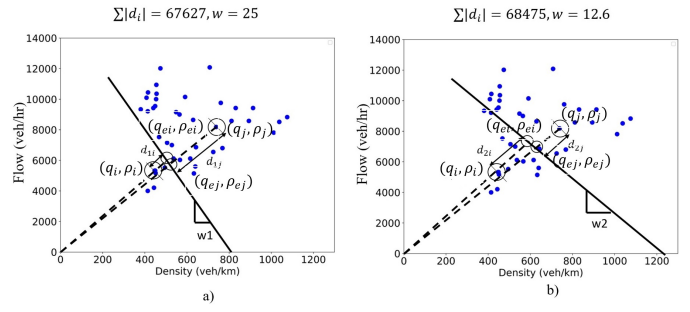


FIGURE 12. a) The first iteration of the distance optimization algorithm, the total sum of the distance from the w-line to steady state points is 67627 when  $w = 25$  b) the Second iteration of the distance optimization algorithm, the total sum of the distance from the w-line to steady state points is 68475 when  $w = 12.6$

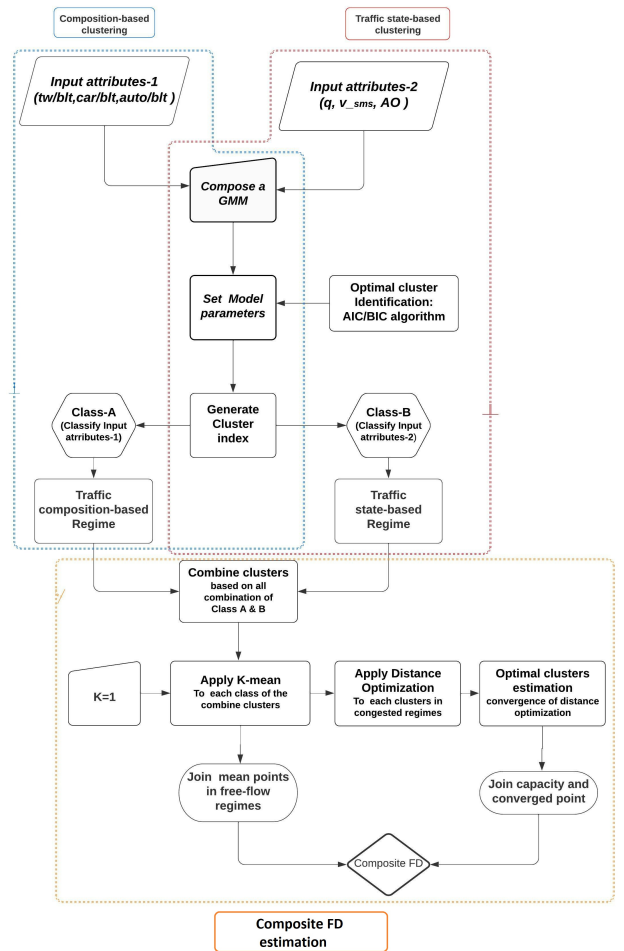
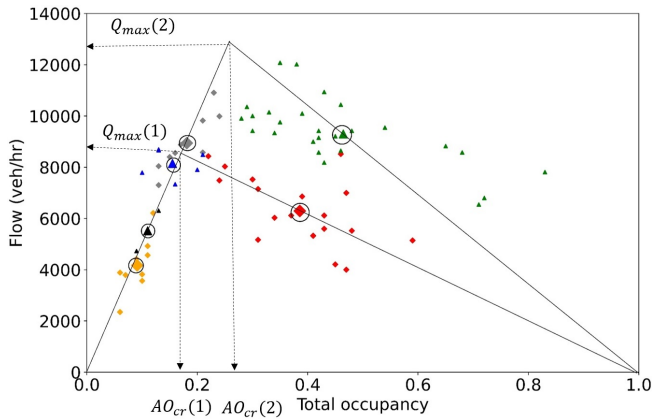


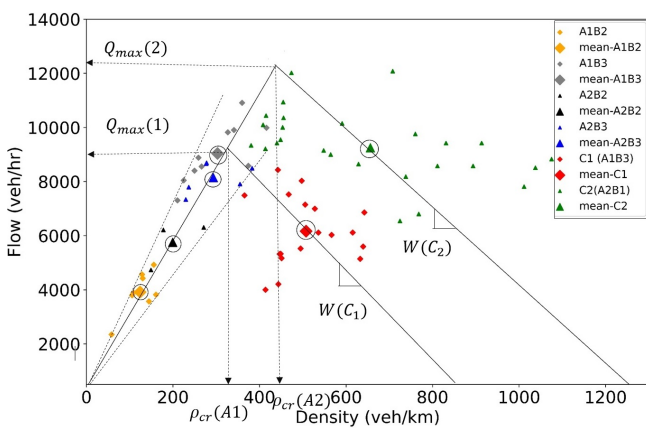
FIGURE 13. Unsupervised learning approach to fundamental diagram estimation

Figure 14. Line joins  $AO = 1$  and two mean points which meet with the free-flow regime at  $(AO_{cr}(C1), Q_{max}(C1))$ ,  $(AO_{cr}(C2), Q_{max}(C2))$ , representing the wave speed. Here,  $(AO_{cr}(Ci), Q_{max}(Ci))$  are critical AO and a maximum flow of traffic for the  $Ci$  class ( $i \in 1, 2$ ). The assumption we made

in this study is that vehicles will occupy the maximum area of the road section during jam conditions leading to  $AO = 1$ . We estimated the jam density and generated  $q - \rho$  FDs (see Figure 14b) for the clustered data sets using capacity flow ( $Q_{max}$ ) of AO- $q$  FDs and mean points of  $q - \rho$  FD clusters. In Figure 14b, the free-flow regime continuous line represents the average free-flow speed of the traffic stream. Two dotted lines above and below the best-fitted line in the free-flow regime represent the 95% confidence interval of mean free-flow speed.  $Q_{max}(C1)$ ,  $Q_{max}(C2)$  are the saturation flow for C1 and C2 cluster, respectively, which were estimated from Figure 14(b). The point of intersection of the mean free-flow speed line and the two reference saturation lines gives the critical density for class-C1 ( $\rho_{cr}(C1)$ ) and class-C2 ( $\rho_{cr}(C2)$ ). Line trough ( $\rho_{cr}$ ,  $Q_{max}(C)$ ) and mean point of congested regime meets at  $q = 0$ , gives the jam density ( $\rho_{jam}$ ) for C1 and C2. Wave speed in congestion for two different compositions  $W(C1)$ ,  $W(C2)$  can be estimated from ( $Q_{max}(C)$ ,  $\rho_{cr}$ ) and  $\rho_{jam}$ .



(a)



(b)

**FIGURE 14.** a) AO- $q$  FD after applying distance optimization algorithm, new clusters in the congested regime are C1 and C2 b)  $q - \rho$  FD after applying distance optimization algorithm.

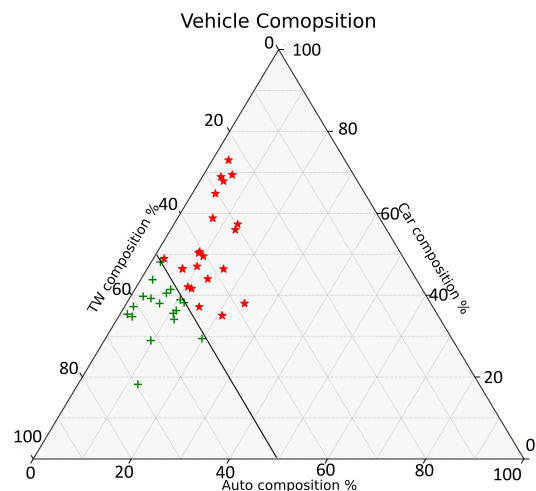
Figure 14(a) shows two classes of scattering data in a congested regime. To evaluate the distribution of compositions of the new cluster's data, we plotted a ternary phase diagram

VOLUME ,

of composition for the two estimated classes. A ternary plot is a triangular plot of three variables (TW/BLT: car/BLT: auto/BLT) which must sum to 100 %. The scatter triangular plot of the three compositions variable for the two classes are shown in Figure 15. It is evident from Figure 15 that the distribution of compositions is different from one class to another, demonstrating that all compositions with 2W percentage over 50% and 3W proportion less than 50% have a distinct wave speed compared to the other way around. Table (3) shows the summary statistics of the traffic composition and the corresponding FD parameters. From Table (3), it can be seen that the mean free-flow speed of composition lies between 95% confidence interval, i.e. [34,28] km/hr. It can also be seen that  $\rho_{jam}$  for C2 is higher than C1 as the TW and BLT proportion is higher for C2. Wave speed in congestion for two different compositions can be estimated from ( $Q_A$ ,  $\rho_{cr}$ ) and  $\rho_{jam}$ . It is comprehensible from Figure 14 (b) that the C1 class has a higher wave speed than C2. So, it can be concluded that higher TW and BLT proportion leads to lower wave speed in congestion.

**TABLE 3.** The parameter value of composition-based FD after distance optimization algorithm

Parameters	Class-C1	Class-C2
Mean proportion [ $\frac{TW}{BLT} : \frac{car}{BLT} : \frac{auto}{BLT}$ ]	[12.3:12:2]	[17.6:13.2:3]
SD of proportion [ $\frac{TW}{BLT} : \frac{car}{BLT} : \frac{auto}{BLT}$ ]	[3.8:2.5:0.65]	[7.6:3.8:3.2]
Mean $v_f$ (km/hr)	30	30
95% CI of $v_f$ (km/hr)	[34,28]	[34,28]
$\rho_{cr}$ (veh/km)	350	410
$AO_{cr}$ (veh/km)	0.21	0.3
$\rho_{jam}$ (veh/km)	870	1300
$Q_{max}$ (veh/hr)	9800	12500
wave speed, $w$ (km/hr)	18.9	14.0



**FIGURE 15.** Ternary phase diagram of the percentage of composition (TW/BLT: car/BLT: auto/BLT) for the optimized distance clusters in congested regime

**A. INTER-CLASS INTERACTION**

We studied the interactions between different classes of vehicles using composition-specific FDs. In this study, the interactions are defined as the speed change of one class in the presence of other classes with AO. The interactions were investigated to understand the response of the one-class vehicle to the others in the congested regime. Therefore, we represented the interaction in terms of vehicle speed based on the class of the vehicles.

The *ss* points of composition-specific FDs are weighted averages of class-specific variables and can be grouped based on similar behavior. The similar behavior points on composite FDs are identified by class-specific mean speed of the TW ( $v_{TW}$ ), cars ( $v_{car}$ ), auto ( $v_{auto}$ ), BLT ( $v_{BLT}$ ) and their relative speeds. These parameters are used as a feature to classify the congested points using the *GMM* model. The characteristics of each vehicle class in a unique group of *GMM* models are analyzed to understand the inter-class interactions.

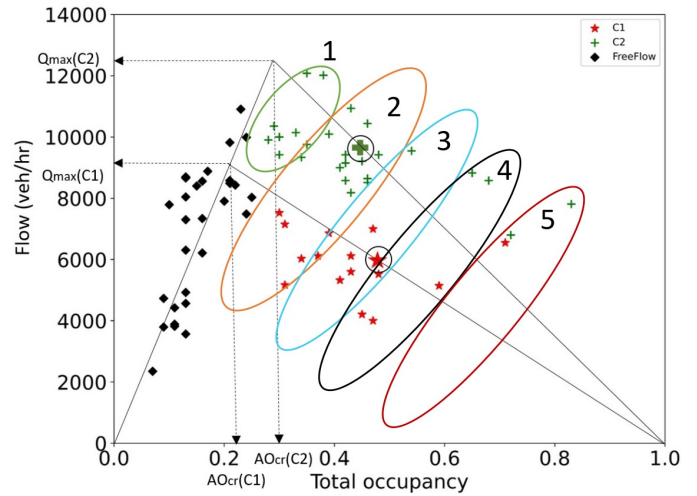
**TABLE 4.** Inter-class interaction regimes in the congested branch of composition-specific FDs

Regime	Speed description	C1		C2	
		min_AO	max_AO	min_AO	max_AO
1	TW>auto>car>BLT	0.32	0.46	-	-
2	TW>car>auto>BLT	0.48	0.56	0.4	0.51
3	TW>BLT>car=auto	0.46	0.69	0.35	0.567
4	TW>auto=BLT>car	0.56	0.84	0.48	0.59
5	TW=car=auto=BLT	0.62	0.85	0.61	0.78

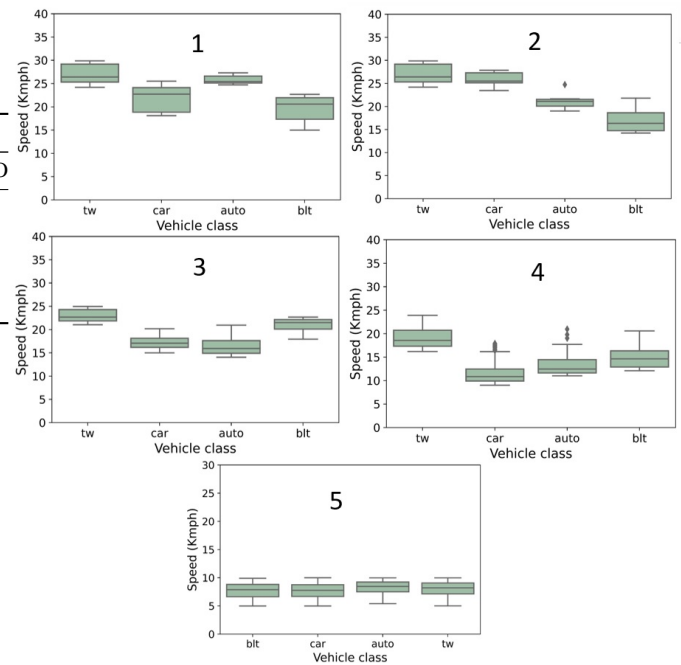
The mean speed, standard deviation, and range of observed AO for each group are shown in Table 4 and Figure 17 to understand the behavior of each class. TWs always move faster than other classes of vehicles, irrespective of the regimes. In the lower congestion levels (in regime 1,2), cars and autos move faster than heavy vehicles (BLT). However, BLT dominates the cars and auto speed with the increase of congestion (in Regime 3,4). It demonstrates that the presence of heavy vehicles at the medium congestion level affects high-speed vehicles. Further, all the vehicles move together at a very high congestion level (in Regime 5), beyond the AO of 0.6, representing single pipe flow in high congestion. Since the maximum and minimum AO values for C1 are higher than C2, the shape of the regimes can be represented by joining an oval diagonally to the minimum and maximum points of C1 and C2 in the FDs. Figure 16 shows the regimes 1 to 5 in the fundamental diagrams.

**V. DISCUSSION**

We have introduced a framework based on empirical observation, which estimates the composition-specific fundamental diagrams for heterogeneous and non-lane-based traffic. In this section, we showed the potential of the proposed methodology to produce promising results for different data sets and the theoretical insights from the empirical analysis of the composition-specific FDs.



**FIGURE 16.** composition-specific fundamental diagram with vehicle interaction regimes in congestion.



**FIGURE 17.** Box plots of speed for TW, car, auto, and BLT for different regimes.

**A. EMPIRICAL EVIDENCE**

To the authors' understanding, since there are no suitable studies in the literature for fair benchmarking, the observations are validated theoretically by proposing several lemmas. Furthermore, this methodology was applied to multiple data sets to show the generalizability of the observations. The proposed methodology is tested on four mixed traffic datasets from India and Greece to demonstrate its universality. The datasets include Stadiou corridor, Athens, Greece [53] (named pNEUMA-1; 9:00-9:30 and pNEUMA-2; 9:30-10:00), Sardar Patel Road, Chennai, India [39] (named view2-Chennai), and Surat-Dumas road, Surat, India [54]

(named Surat). Figure 18 shows the composition-specific  $q - \rho$  FDs for the four datasets. The FDs show variable free-flow and wave speeds based on the compositions. Also, the FDs show that the wave speed is higher for the composition with more heavy vehicles and decreased with the increase of fast-moving vehicles. Also, the jam density increased with increasing TW proportion. These results are in line with the observations from the literature and indicate that the proposed methodology is robust in accurately estimating the FDs for mixed traffic conditions.

## B. THEORETICAL DERIVATION

In this section, we introduce and prove some characteristics of proposed composition-specific fundamental diagrams analytically from the observed data set.

**Lemma-1:** In free flow, the mean speed of different compositions is equal. However, the range of the speed of the stream composition for relatively higher when the proportion of fast-moving vehicles (TW, car) is higher than that of the slow-moving vehicles.

**proof:** We consider the case of two traffic streams where composition-1 (C1) has more slow-moving vehicles (truck, heavy vehicles), and composition-2 (C2) has more fast-moving vehicles (TW, cars). Let's assume the proportion of fast vehicles to slow vehicles for C1 and C2 are  $(\alpha_1 : \beta_1)$  and  $(\alpha_2 : \beta_2)$ , respectively. Where  $(\alpha_2 > \alpha_1 : \forall \alpha, \beta > 0)$ , then the stream free flow speed for C2 ( $v_2$ ) will be higher than stream free flow speed for C1 ( $v_1$ ) (7).

$$\begin{aligned} v_1 &= (v_{f(\text{fastvehicle})} \times \alpha_1 + v_{f(\text{slowvehicle})} \times \beta_1) \leq v_2 \\ &= (v_{f(\text{fastvehicle})} \times \alpha_2 + v_{f(\text{slowvehicle})} \times \beta_2) \quad (7) \end{aligned}$$

$v_{f(\text{slowvehicle}/\text{fastvehicle})}$ - free flow speed of slow-moving and fast-moving vehicles respectively. The observed maximum speed ( $v_{sl}$ ) for the urban road was 40 kmph which was lesser than  $v_{f(\text{slowvehicle}/\text{fastvehicle})}$ . Hence, it is evident that any traffic-stream compositions in free-flow states will only allow to move in  $v_{sl}$  subjected to maximum variation up to  $v_{f(\text{slowvehicle}/\text{fastvehicle})}$ . The free flow speed variation of slow vehicles  $[v_{sl}, v_{f(\text{slowvehicle})}]$  and fast vehicle  $[v_{sl}, v_{f(\text{fastvehicle})}]$ .

$$v_2 \geq v_1 \geq v_{sl}, \quad \exists [v_{sl}, v_{f(\text{slowvehicle})}] \subset [v_{sl}, v_{f(\text{fastvehicle})}] \quad (8)$$

Empirical evidence for hypothesis 1 can be seen in Figure 19 for our data set. C2 composition of the stream had a more significant proportion of fast-moving vehicles than composition C1. Hence, the best-fit lines of the speed-AO relationship for C2 lay above C1 in the congested regime, but they almost overlapped in free-flow regimes. Figure 19 shows the coefficient of determination ( $R^2$ ) and logarithmic relation between speed and AO.

**Lemma-2:** Traffic composition of a higher proportion of fast-moving vehicles has a higher capacity than the lower fast-moving vehicle composition.

**proof:** The critical density of a higher proportion of fast vehicles is greater than the lower proportion of fast-moving vehicles, and their speed in the free-flow regime is unique,  $v_{sl}$ . As it appears that the FD obtained in this study, the higher proportion of fast-moving vehicle's composition impacts higher capacity than the lower proportion of the fast-moving vehicles (see Figure 14(b)).

**Lemma-3:** In congestion, the wave speeds are different for different compositions. Wave speed for a relatively higher proportion of large vehicles is higher.

**proof:** Traffic density of a stream depends on the stream composition, the fundamental relation for any composition (c) of traffic can be expressed as:

$$q(\rho(c)) = \rho(c)v(c)$$

Here  $\rho(c), q(c)$ , and  $v(c)$  are the fundamental parameters of composition (c) specific traffic stream. So, the wave speed for the composition stream in congestion

$$\begin{aligned} \frac{\partial q(\rho(c))}{\partial \rho(c)} &= v(c) + \rho(c) \frac{dv(c)}{d\rho(c)} \\ &= v(c) + \rho(c) \frac{\partial v(c)}{\partial AO} \frac{dAO}{d\rho(c)} \\ &= v(c) + \rho(c) \frac{\partial v(c)}{\partial AO(c)} \times \bar{w}(c)(l(c) + \Delta x) \quad \text{from} \quad (5) \end{aligned} \quad (9)$$

We consider two traffic streams based on the fast vehicles (TW,car) to slow vehicles (heavy vehicles) proportion i.e. C1( $\alpha_1 : \beta_1$ ), and C2( $\alpha_2 : \beta_2$ ) respectively. Where  $(\beta_1 > \beta_2) : \forall \alpha, \beta > 0$ . The heavy vehicle proportion for C1 is greater than C2, and the fast vehicle proportions for C1 and C2 are the same. So, the density relation of C1 and C2 can be expressed as:

$$\rho(C2) < \rho(C1) \quad (10)$$

The proportion of heavy vehicles or slow-moving vehicles increases for C1, and the average speed for the stream of C1 composition will be lesser than C2.

$$v(C2) > v(C1) \quad \text{as} \quad v_2 > v_1 \quad (\text{congestion} : \rho_{cr} < \rho < \rho_{jam}) \quad (11)$$

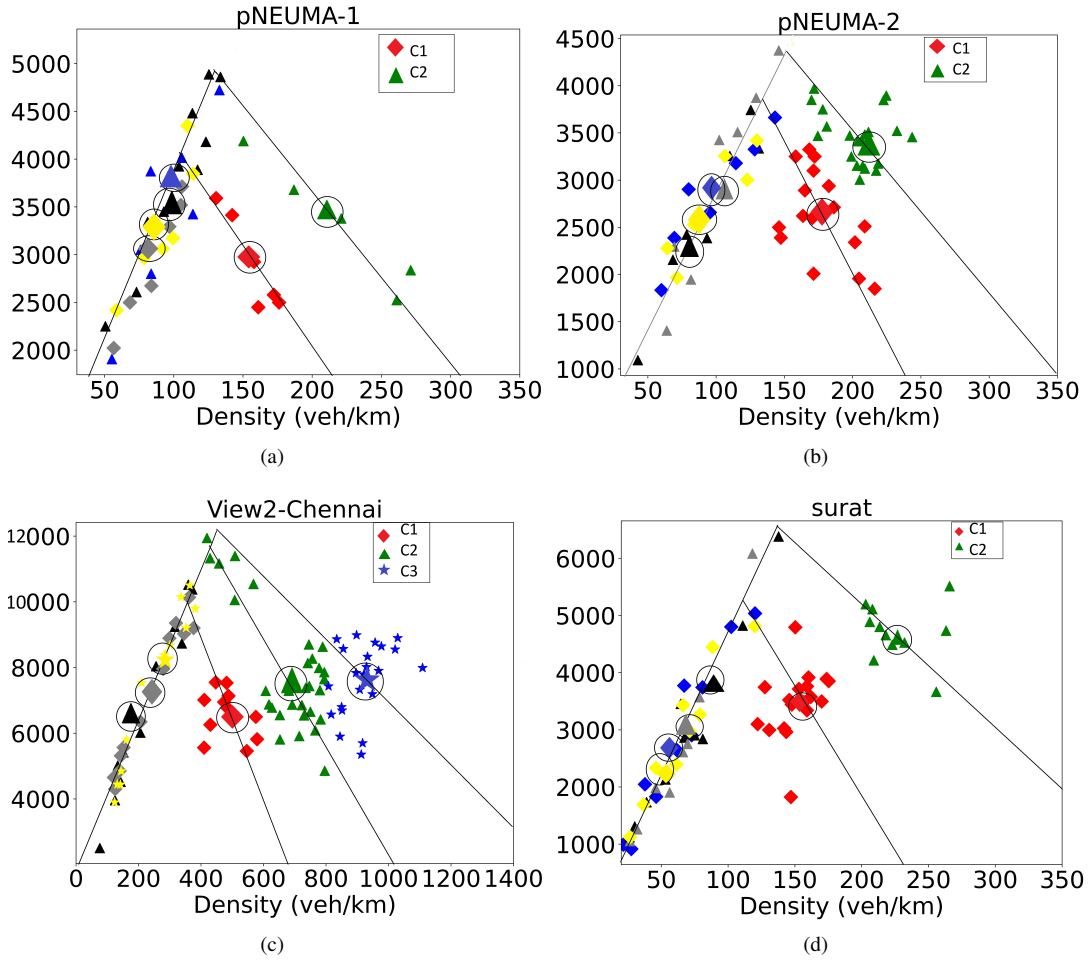
As the average width and length of C1 composition is larger than C2:

$$\bar{w}(C2)(l(C2) + \Delta x) < \bar{w}(C1)(l(C1) + \Delta x) \quad (12)$$

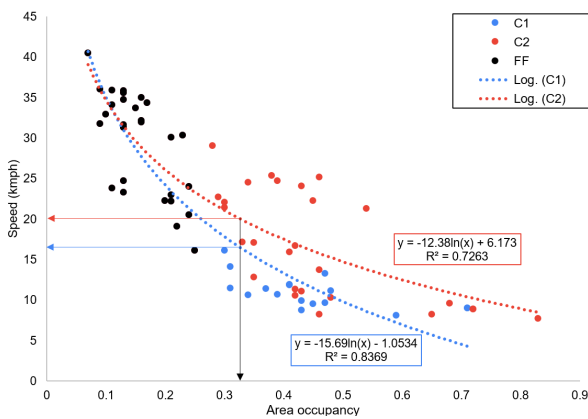
Assume the rate of change of speed with respect to AO in a congested regime is irrespective of composition and speed is a decreasing function of AO in congestion, i.e.,

$$\frac{\partial v(c)}{\partial AO} < 0, \quad AO_{cr} < AO < 1 \quad (13)$$

Now we can estimate the wave speed in congestion  $|\frac{\partial q(\rho(c))}{\partial \rho(c)}|$  using (10-13) in (9). As the 2nd term of (9) is significantly higher for C1 than C2, and from (10), it is evident that the magnitude of wave speed for C1 is higher than C2.



**FIGURE 18.** (a) composition-specific FDs were estimated from the data collected using camera-1,2,3,4 in morning 9:00-9:30,pNEUMA, Athens, Greece [53] (b) composition-specific FDs were estimated from the data collected using camera-1,2,3,4 in morning 9:30-10:00,pNEUMA, Athens, Greece [53] (c) same data location from Chennai, India which is used to develop the methodology in this paper for a different time period in evening peak(d) composition-specific FDs were estimated from surat-dumas roadway data [54]

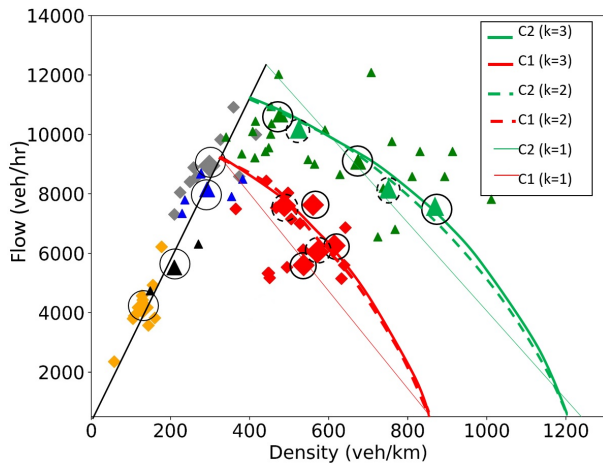


**FIGURE 19.** Speed-AO relationship for two compositions C1 and C2

$$\left| \frac{\partial q(\rho(C2))}{\partial \rho(C2)} \right| < \left| \frac{\partial q(\rho(C1))}{\partial \rho(C1)} \right| \quad (14)$$

### C. FDS WITH $K > 1$ IN THE K-MEAN CLUSTERING

The earlier sections demonstrated the methodology to estimate FDs of different compositions using the k-mean clustering. The steps assumed linear FDs to better understand the composition-specific FDs' average behavior (i.e.,  $k = 1$ ). However, one could assume a non-linear congested branch for the FDs using  $k > 1$ . Figure 20 shows the composition-specific fundamental diagrams for  $k = 2, 3$  which indicate changes in the saturation flow, critical density, and wave speed. Also, the FDs parameters for  $k=1, 2$ , and 3 show that the difference between the former two is more than the latter for the compositions observed. Thus, it may be conjectured that analysis beyond  $k = 2$  may not be required.



**FIGURE 20.** Thick continuous lines in congested regime show the composition-specific fundamental diagrams after applying k-mean clustering with  $k = 3$  and dotted lines represent the composition-specific fundamental diagrams after applying k-mean clustering with  $k = 2$ . A comparison could be made between these FDs and the proposed FDs  $k = 1$  (thin continuous line).

## VI. CONCLUSION

This paper deals with the estimation of multi-regime FDs based on traffic composition for mixed traffic conditions. Towards this, the paper first presented a methodology to identify the  $ss$  periods using the oblique plot (Obl N) of the cumulative vehicle arrivals curve with the help of the oblique occupancy time plot (Obl T). The CWT energy plots of  $sms$  speed show that the class-specific  $ss$  exists in the traffic stream. Next, the common class-specific  $ss$  parameters are extracted to estimate the stream  $ss$ . In mixed traffic conditions, the availability and accessibility of free space are the major elements that govern the speed of vehicles. In this paper, AO was used for the measure of FD parameters which would be more accurate for modeling the vehicular flow.

Unsupervised ML, *GMM* is used for clustering the traffic compositions and traffic states and finding their centroids. We show that different classes of fundamental diagrams with small variations in saturation flow and wave speed are possible for  $k = 1$  and  $k = 2$ . However, there is no significant difference in FDs was found beyond  $k > 3$ . In addition, we introduced the distance optimization algorithm to address the scatteredness of points in the congested regime. It is found that the distance optimization algorithm effectively classifies the points for each composition to obtain a tight fit for the congested branch.

Traffic flow parameters like mean free-flow speed, critical density, critical AO, saturation flow, wave speed, and Jam density are estimated from the proposed FD. Based on the observation, we proposed some hypotheses of the composite fundamental diagrams and proved those using traffic flow theory.

Some of the major insights/highlights of the paper are:

- A systematic separation of traffic states facilitates easily capturing the effects and interactions across different traffic compositions.

- Estimation of stream  $ss$  from the class-specific  $ss$  using obl-N and Obl-T is an efficient way to effectively capture the mixed traffic characteristics.
- The FDs are highly sensitive to traffic compositions, with frequent changes in compositions creating complex traffic scenarios.
- The free-flow speed only varied marginally and showed a strong linear fit for the free-flow branch of the FDs for different compositions.
- From the  $q-\rho$  FD plots, it was found that higher TW (or lower BLT) proportions lead to higher saturation flow, jam density, and lower wave speed in congestion.
- An analysis of inter-class interactions shows that heavy vehicles cause faster vehicles to slow down at medium to high congestion levels. However, this effect is negligible in low congestion and diminishes with increasing congestion levels.
- The validation results from various data sets indicate that the proposed methodology is robust in systematically developing FDs for mixed traffic conditions and the generalization of the insights across geographical locations shows the possible universality of the results.

The methods presented in this paper can be easily adapted for more compositions for better estimation of traffic flow variables and FDs. Some of the future research directions include calibrating the thresholds, improving and fine-tuning the clustering methods, and exploring alternative approaches to automatically estimate FDs from the trajectory data to reduce the computational effort for FD extraction and research continues in these directions.

## ACKNOWLEDGMENT

The author would like to thank Prof. Jorge Laval for his helpful advice on various technical aspects in this paper. The authors would also like to acknowledge the Indian Institute of Technology, Madras which facilitated this collaborative research with the Georgia Institute of Technology, Atlanta, USA.

## CODE AVAILABILITY

We used python programming to demonstrate the methodology, and the link for the code can be found here: <https://github.com/tffnandan/composition-specific-FDs>. Traffic state optimization code will be available upon request from readers.

...

## REFERENCES

- [1] B. Greenshields, J. Bibbins, W. Channing, and H. Miller, "A study of traffic capacity," in Highway research board proceedings, vol. 1935. National Research Council (USA), Highway Research Board, 1935.
- [2] H. Grinberg, "An analysis of traffic flows," *Oper Res*, vol. 7, pp. 79–85, 1959.
- [3] R. T. Underwood, "Speed, volume and density relationships," *Quality and theory of traffic flow*, 1961.
- [4] J. S. Drake, "A statistical analysis of speed density hypothesis," *HRR*, vol. 154, pp. 53–87, 1967.

- [5] M. Papageorgiou, J.-M. Blosseville, and H. Hadj-Salem, "Macroscopic modelling of traffic flow on the boulevard périphérique in paris," *Transportation Research Part B: Methodological*, vol. 23, no. 1, pp. 29–47, 1989.
- [6] J. D. Castillo and F. Benitez, "On the functional form of the speed-density relationship." *Transportation Research Part B: Methodological*, 1995.
- [7] M. Van Aerde, "A single regime speed-flow-density relationship for freeways and arterials," Preprint, vol. 950802, 1995.
- [8] H. Wang, J. Li, Q. Chen, and D. Ni, "Representing the fundamental diagram: The pursuit of mathematical elegance and empirical accuracy," *Tech. Rep.*, 2010.
- [9] M. J. MacNicholas, "A simple and pragmatic representation of traffic flow," in *Symposium on The Fundamental Diagram*, vol. 75, 2011, pp. 161–177.
- [10] H. K. Gaddam and K. R. Rao, "Speed–density functional relationship for heterogeneous traffic data: a statistical and theoretical investigation," *Journal of modern transportation*, vol. 27, no. 1, pp. 61–74, 2019.
- [11] L. C. Edie, "Car-following and steady-state theory for noncongested traffic," *Operations research*, vol. 9, no. 1, pp. 66–76, 1961.
- [12] L. Sun and J. Zhou, "Development of multiregime speed–density relationships by cluster analysis," *Transportation Research Record*, vol. 1934, no. 1, pp. 64–71, 2005.
- [13] G. Asaithambi, V. Kanagaraj, K. K. Srinivasan, and R. Sivanandan, "Study of traffic flow characteristics using different vehicle-following models under mixed traffic conditions," *Transportation Letters*, vol. 10, no. 2, pp. 92–103, 2018. [Online]. Available: <https://doi.org/10.1080/19427867.2016.1190887>
- [14] R. Soni, V. Vasudevan, and B. Dutta, "Analysis of overtaking patterns of indian drivers with data collected using a lidar," *Transportation research part F: traffic psychology and behaviour*, vol. 74, pp. 139–150, 2020.
- [15] M. Adnan, "Passenger car equivalent factors in heterogenous traffic environment-are we using the right numbers?" *Procedia engineering*, vol. 77, pp. 106–113, 2014.
- [16] O. Giuffrè, A. Granà, R. Mauro, A. B. Silva, and S. Chiappone, "Developing passenger car equivalents for freeways by microsimulation," *Transportation Research Procedia*, vol. 10, pp. 93–102, 2015.
- [17] A. Dhamaniya and S. Chandra, "Concept of stream equivalency factor for heterogeneous traffic on urban arterial roads," *Journal of transportation engineering*, vol. 139, no. 11, pp. 1117–1123, 2013.
- [18] O. Giuffrè, A. Grana, M. L. Tumminello, and A. Sferlazza, "Estimation of passenger car equivalents for single-lane roundabouts using a microsimulation-based procedure," *Expert Systems with Applications*, vol. 79, pp. 333–347, 2017.
- [19] A. Dhamaniya and S. Chandra, "Conceptual approach for estimating dynamic passenger car units on urban arterial roads by using simultaneous equations," *Transportation Research Record*, vol. 2553, no. 1, pp. 108–116, 2016.
- [20] N. C. Sarkar, A. Bhaskar, Z. Zheng, and M. P. Miska, "Microscopic modelling of area-based heterogeneous traffic flow: Area selection and vehicle movement," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 373–396, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0968090X18313810>
- [21] S. K. Nirmale, A. R. Pinjari, and A. Sharma, "A discrete-continuous multi-vehicle anticipation model of driving behaviour in heterogeneous disordered traffic conditions," *Transportation Research Part C: Emerging Technologies*, vol. 128, p. 103144, 2021.
- [22] C. Mallikarjuna and K. R. Rao, "Heterogeneous traffic flow modelling: A complete methodology," *Transportmetrica*, vol. 7, no. 5, pp. 321–345, 2011.
- [23] V. T. Arasan and G. Dhivya, "Measuring heterogeneous traffic density," in *Proceedings of international conference on sustainable urban transport and environment*. Citeseer, 2008, pp. 342–346.
- [24] C. Mallikarjuna and K. R. Rao, "Area occupancy characteristics of heterogeneous traffic," *Transportmetrica*, vol. 2, no. 3, pp. 223–236, 2006.
- [25] V. T. Arasan and G. Dhivya, "Simulation of highly heterogeneous traffic flow characteristics," in *ECMS*, 2010, pp. 81–87.
- [26] W. Jin, "Kinematic wave traffic flow model for mixed traffic," *Transportation Research Record*, vol. 1802, 11 2002.
- [27] S. Logghe and B. Immers, "Heterogeneous traffic flow modelling with the lwr model using passenger-car equivalents," 2003.
- [28] G. Wong and S. Wong, "A multi-class traffic flow model – an extension of lwr model with heterogeneous drivers," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 9, pp. 827 – 841, 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0965856401000428>
- [29] S. Benzoni-Gavage and R. M. Colombo, "An n-populations model for traffic flow," *Europ. J. Appl.Math.*, vol. 14, pp. 587–612, 2002.
- [30] F. van Wageningen-Kessels, H. Van Lint, S. P. Hoogendoorn, and K. Vuk, "New generic multiclass kinematic wave traffic flow model: model development and analysis of its properties," *Transportation Research Record*, vol. 2422, no. 1, pp. 50–60, 2014.
- [31] F. L. Hall, "Traffic stream characteristics," *Traffic Flow Theory*. US Federal Highway Administration, vol. 36, 1996.
- [32] B. Coifman, "Empirical flow-density and speed-spacing relationships: Evidence of vehicle length dependency," *Transportation Research Part B: Methodological*, vol. 78, pp. 54–65, 2015.
- [33] A. K. Gupta and I. Dhiman, "Analyses of a continuum traffic flow model for a nonlane-based system," *International Journal of Modern Physics C*, vol. 25, no. 10, 2014.
- [34] Z. Yao, Q. Gu, Y. Jiang, and B. Ran, "Fundamental diagram and stability of mixed traffic flow considering platoon size and intensity of connected automated vehicles," *Physica A: Statistical Mechanics and its Applications*, vol. 604, p. 127857, 2022.
- [35] W. Wang and B. Wu, "The fundamental diagram of mixed-traffic flow with cacc vehicles and human-driven vehicles," *Journal of Transportation Engineering, Part A: Systems*, vol. 149, no. 1, p. 04022116, 2023.
- [36] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.
- [37] M. Shang and R. Stern, "A hybrid fundamental diagram for modeling mixed human and automated traffic flow," in *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, 2021, pp. 1–6.
- [38] J. Thomas, K. K. Srinivasan, and V. T. Arasan, "Vehicle class wise speed-volume models for heterogeneous traffic," *Transport*, vol. 27, no. 2, pp. 206–217, 2012.
- [39] R. Dhatbale and B. R. Chilukuri, "Deep learning techniques for vehicle trajectory extraction in mixed traffic," *Journal of Big Data Analytics in Transportation*, pp. 1–17, 2021.
- [40] M. J. Cassidy, "Bivariate relations in nearly stationary highway traffic," *Transportation Research Part B: Methodological*, vol. 32, no. 1, pp. 49–59, 1998.
- [41] D. S. Dendrinos, "Urban traffic flows and fourier transforms," *Geographical Analysis*, vol. 26, no. 3, pp. 261–281, 1994.
- [42] Z. Zheng, S. Ahn, D. Chen, and J. Laval, "Applications of wavelet transform for analysis of freeway traffic: Bottlenecks, transient traffic, and traffic oscillations," *Transportation Research Part B: Methodological*, vol. 45, no. 2, pp. 372–384, 2011.
- [43] P. White, "Transforms, Wavelets," *Encyclopedia of Vibration*, pp. 1419–1435, 2001.
- [44] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *Applied and Computational Harmonic Analysis*, vol. 1, no. 1, pp. 100–115, 1993. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1063520383710080>
- [45] H. Adeli and A. Karim, *Wavelets in intelligent transportation systems*. John Wiley & Sons, Inc., 2005.
- [46] H. Adeli and X. Jiang, *Intelligent infrastructure: neural networks, wavelets, and chaos theory for intelligent transportation systems and smart structures*. CRC press, 2008.
- [47] L. W. Lan and C.-C. Hsu, "Formation of spatiotemporal traffic patterns with cellular automaton simulation," *Tech. Rep.*, 2006.
- [48] C. Gao, J. Xu, X. Jia, Y. Dong, and H. Ru, "Influence of large vehicles on the speed of expressway traffic flow," *Advances in Civil Engineering*, vol. 2020, 2020.
- [49] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.
- [50] T. L. Athey and J. T. Vogelstein, "Autogmm: Automatic gaussian mixture modeling in python," *arXiv preprint arXiv:1909.02688*, 2019.
- [51] D. A. Reynolds, "Gaussian mixture models." *Encyclopedia of biometrics*, vol. 741, no. 659–663, 2009.
- [52] J. MacQueen, "Classification and analysis of multivariate observations," in *5th Berkeley Symp. Math. Statist. Probability*, 1967, pp. 281–297.
- [53] E. Barmponakis and N. Geroliminis, "On the new era of urban traffic monitoring with massive drone data: The pneuma large-scale field experiment," *Transportation research part C: emerging technologies*, vol. 111, pp. 50–71, 2020.
- [54] N. Raju, S. Arkatkar, S. Easa, and G. Joshi, "Developing extended trajectory database for heterogeneous traffic like ngsim database," *Transportation Letters*, vol. 14, no. 5, pp. 555–564, 2022.





modeling, Intelligent Transpiration Systems (ITS).

NANDAN MAITI was born in Kharagpur, India. He received a B.E. degree in Civil Engineering from the Indian Institute of Engineering Science and Technology, Shibpur (IEST Shibpur), India, in 2015. He is pursuing a direct Ph.D. (M.S.+Ph.D.) degree with the Department of Civil engineering, Indian Institute of Technology, Madras. From 2015 to 2017, he was an executive engineer with ITD Cementation India Limited, Mumbai. His research interests include traffic flow



Traffic, Traffic Operations, and Networks.

DR. BHARGAVA RAMA CHILUKURI holds a Ph.D. in Civil and Environmental Engineering from Georgia Institute of Technology, Atlanta, Georgia, USA. He has several years of professional experience as a traffic engineer of multiple companies in the USA. He is currently working as an Associate Professor in the Department of Civil Engineering, at the Indian Institute of Technology Madras. His research interests include Traffic Flow Theory of Homogenous and Heterogeneous