Leader-follower identification methodology for non-lane disciplined heterogeneous traffic using steady state features

Susan Eldhose*[1], Bhargava Rama Chilukuri[2], Chandrasekharan Rajendran[3]

[1]Research Scholar, Department of Civil Engineering, Indian Institute of Technology Madras, India
[2]Associate professor, Department of Civil Engineering, Indian Institute of Technology Madras, India
[2]Professor, Department of Management Studies, Indian Institute of Technology Madras, India

SHORT SUMMARY

Road traffic in developing countries is characterized by heterogeneous vehicle types and weak lane discipline, with motorized and non-motorized vehicles of varying sizes and maneuverability leading to diverse driver behaviors. Identifying leader-follower (LF) pairs and analyzing vehicle-following (VF) behavior under such conditions is challenging, as proximity alone may not capture a leader vehicle's (LV) influence on a subject vehicle's (SV) behavior. Non-following episodes, even with similar gaps or time headways, highlight the limitations of fixed longitudinal clearance thresholds. This study addresses these challenges by combining the k-v fundamental diagram to estimate desirable longitudinal gaps and wavelet transforms (WT) to match LV and SV speed profiles. The proposed methodology improves accuracy over heuristic methods, increasing the R-squared value from 0.268 to 0.349 and reducing RMSE from 0.764 to 0.652, offering a robust framework for LF pair identification in heterogeneous traffic conditions.

Keywords: Heterogeneous and non-lane-based traffic, leader-follower identification, wavelet transform.

1. INTRODUCTION

Developed countries typically exhibit homogeneous, lane-based traffic, distinguishing VF and lane-changing behaviors. In contrast, developing nations like India experience heterogeneous weak lane-based (HWLD) traffic, where motorized and non-motorized vehicles with varying characteristics share road space. Drivers navigate longitudinally by utilizing gaps, favoring an area-based arrangement to optimize space (Madhu et al., 2020).

Analyzing VF behavior in HWLD traffic requires identifying LF pairs, where an SV's motion is influenced by an LV. However, LF identification in HWLD traffic remains largely underexplored. Studies often rely on synthetic or simulated data (Das et al., 2019; Mathew & Ravishankar, 2011) or aggregated macroscopic measures (Asaithambi et al., 2018; Mathew & Radhakrishnan, 2010).

The literature presents various criteria for identifying LF pairs. Longitudinal clearance between the assumed leader and follower is a common method used to define the LV and its influence zone on the SV (Anand et al., 2019; Madhu et al., 2020, 2022; Nirmale et al., 2021). Lateral clearance or overlap width is another important factor for identifying the LV (Anand et al., 2019; Madhu et al., 2020, 2022; Nirmale et al., 2021; Papathanasopoulou & Antoniou, 2018; Raju et al., 2021).

Once the LV is identified, the pair is classified as LF or non-LF. Anand et al. (2019) used a minimum threshold for the continuous following duration. Madhu et al. (2020, 2022) identified the most influencing

LV based on the closest gap between SV and LV when more than one LV is present.

In HWLD traffic, the influence of other surrounding vehicles between the leader and follower must be considered. Smaller vehicles, such as two-wheelers (TW), often partially occupy the influence region, causing intermittent VF behavior. Such dynamics, uncommon in lane-based traffic, frequently occur in non-lane-based conditions due to vehicle size variations. Raju et al. (2021) used space-time plots to identify other influencing vehicles but noted that visual analysis becomes impractical for large datasets. Recently Kulkarni et al. (2025) proposed a robust methodology for LF identification, focusing on strong LV-SV interaction, significant lateral overlap, and the absence of intervening vehicles. They observed that the presence of an intermediate vehicle between SV and LV weakens their interaction, leading to classification as non-LF pairs.

The above literature on the identification of LF pairs reveals several gaps as follows

i. Fixed thresholds for longitudinal clearance may not capture realistic interactions, as vehicle influence depends on speed and type. For example, TW's generally have shorter look-ahead distances than larger vehicles like trucks.
ii. Kulkarni et al.'s (2025) approach, based on Wiedemann-99 driving regimes, is limited to cars.
iii. A potential LV influences the SV's speed leading to similar speed patterns between them. However, such relationship for LF pair identification has yet to be explored in the literature.
iv. Although WTs have been applied to traffic flow analysis, their use for LF identification under HWLD traffic remains unexplored. Maiti and Chilukuri (2023) used Mexican Hat WT to analyze speed profiles and detect abrupt changes, but its potential for LF identification has yet to be examined.

This study addresses these gaps by proposing an enhanced LF identification method, incorporating desirable longitudinal gaps and speed profile correlations. By leveraging these criteria, the methodology demonstrates improved accuracy over existing heuristic approaches, with broader applicability to diverse vehicle types in HWLD traffic.

2. LEADER FOLLOWER IDENTIFICATION

This section presents an overview of the proposed LF pair identification method.

Data

This study utilizes open-source HWLD trajectory data from Saidapet, Chennai, as provided by Kanagaraj et al. (2015). The processed data include individual trajectories of 3005 vehicles with each vehicle's trajectory including the spatial position, speed, and acceleration/deceleration values in both the longitudinal and lateral directions at a 0.5 s resolution.

Base model

This study aims to predict the longitudinal response (acceleration or deceleration) of the SV using multiple linear regression model. Three fundamental stimuli are considered for modelling: the relative velocity between the SV and the LV, the longitudinal gap between the SV and the LV, and the SV speed. Nirmale et al. (2021) reported optimal acceleration response prediction using a 0.5-second update time with the same open-source Chennai data. Consequently, this study also adopts a 0.5-second update time. The model structure is represented in Equation 1.

$$y(t+\tau) = \beta_0 + \beta_1 x1(t) + \beta_2 x2(t) + \beta_3 x3(t) + \varepsilon \qquad (1)$$

Where $y(t+\tau)$ is the acceleration or deceleration response of the SV at time $(t+\tau)$, $x1$ is the relative speed between LV and SV, $x2$ is the bumper-to-bumper gap between LV and SV in the longitudinal direction, $x3$ is the SV speed, β_x is the parameter associated with variable x and ε is the error term.

Identify leader-follower pairs based on the literature

The development of a longitudinal response model begins with identifying LF pairs, a challenging task in HWLD traffic conditions due to varying vehicle dimensions, intermittent following, and multiple potential leaders. The following four criteria, based on existing literature, were employed to identify LF pairs:

i. Longitudinal threshold: the leader is expected to be present within 30 m from the front bumper of the SV (Madhu et al., 2020, 2022; Nirmale et al., 2021)

ii. Lateral overlap: LV's lateral dimensions should be fully or partially overlapping with SV (Anand et al., 2019; Madhu et al., 2020, 2022; Nirmale et al., 2021; Papathanasopoulou & Antoniou, 2018; Raju et al., 2021).

iii. Duplicate leader: when more than one LV is present, the most influencing LV is identified based on the closest gap with the SV (Madhu et al., 2020, 2022)

iv. A minimum of 5 s continuous following duration (Anand et al., 2019).

Based on these four criteria from the literature, 2125 LF pairs were identified from the trajectory data. The class-wise distribution is presented in Table 1. This study specifically focuses on 707 potential pairs where a Car follows LVs of any type.

Table 1: LF pairs identified based on the literature

SV type	**LF pairs**
TW	942
CAR	707
HV	90
LCV	34
AUTO	352
Total	2125

Propose LF identification methodology

This study introduces an enhanced methodology for identifying LF pairs in HWLD traffic. The following modifications were incorporated to refine the LF pair selection process:

i. **Modification 1**: Exclusion of multiple SVs for a single LV

When multiple SVs follow a single LV, their overlap with the LV is minimal, often indicating tailgating behavior with overtaking intent rather than genuine following. Such pairs were excluded to ensure accurate LF identification.

ii. **Modification 2:** Removal of lateral clearance for overlap calculations

Initially, a lateral clearance of 0.2 m (Papathanasopoulou & Antoniou, 2018) was included for calculating vehicle boundaries. However, this approach overestimated lateral overlaps. Eliminating lateral clearance refined LF pair selection.

A revised LF pair list is identified based on these two modifications. Table 2 compares the number of LF pairs and the correlation values of the independent variables (longitudinal gap, relative velocity, and SV velocity) with the dependent variable (acceleration) for the base model and after applying the modifications. The correlation values highlight Modification 2 was more effective in improving LF pair selection, forming the basis for further analysis.

Table 2: LF pairs and correlations with the modification 1 and 2

Model	**Variables**	**Base model**		**Modification 1**		**Modification 2**	
		No: Pairs	**Correlation**	**No: Pairs**	**Correlation**	**No: Pairs**	**Correlation**
CAR	Long gap	702	-0.002	461	0.018	503	0.007
	Relative velocity		0.442		0.43		0.456
	SV velocity		-0.428		-0.431		-0.447

iii. **Modification 3:** Vehicle-type-specific longitudinal gap thresholds

The commonly used fixed threshold of 30 meters for identifying LVs may not accurately capture real-world interactions. For instance, slower-moving vehicles are less influenced by distant LVs, and the look-ahead distance for TWs is typically shorter than that of heavy vehicles such as trucks. To address this, the study proposes estimating a desirable longitudinal gap for different vehicle types at varying speeds.

Desirable Longitudinal Gap

The desirable gap (s) is calculated for each SV type based on speed, using density (k) derived from the fundamental k-v diagram (Figure 1) and calculated through Equations (2) and (3) for each vehicle type. LF pairs are then selected based on whether the observed longitudinal gap falls within the estimated desirable gap range.

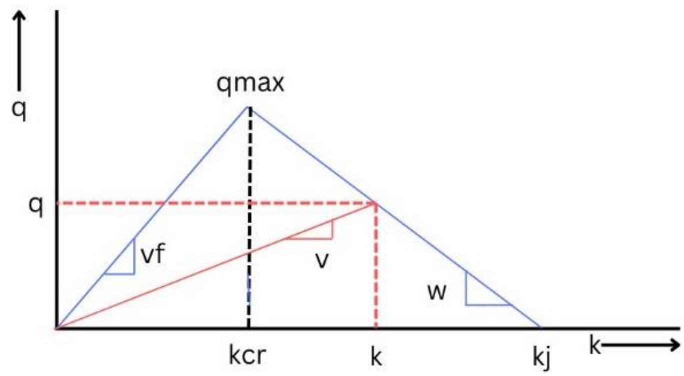

Figure 1: Fundamental diagram

$$k = \frac{w*kj}{(w+v)} \qquad (2)$$

$$gap\ (s) = \frac{1}{k} = \frac{w*kj}{(w+v)} * 1000 \qquad (3)$$

The fundamental diagram parameters such as maximum flow, jam density, critical density, and free flow speed are taken from Ashok and Chilukuri (2024). Figure 2 represents the vehicle type-specific fundamental diagram based on the values from the literature.

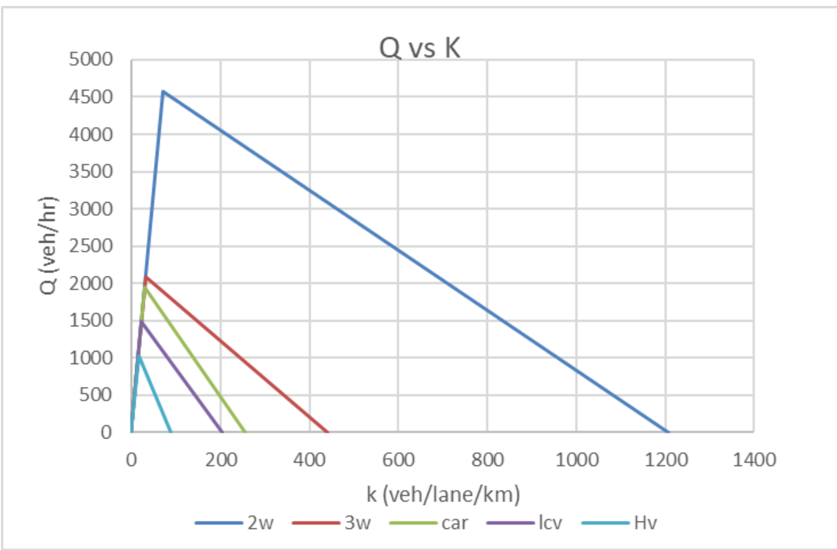

Figure 2: Fundamental diagram of heterogeneous traffic

To accommodate the variability in heterogeneous traffic, a gap allowance ranging from 20% to 100% (in 20% increments) was applied. LF pairs were selected based on thresholds where 50%, 60%, or 70% of the data points of each pair satisfied the gap allowance conditions. **Error! Reference source not found.** lists the LF pairs identified under different spacing allowance thresholds.

Table 3: LF pairs across different gap allowance thresholds

Percentage data points of each pair satisfy threshold	LF pairs with gap allowance				
	20%	**40%**	**60%**	**80%**	**100%**
50%	317	376	434	469	492
60%	291	356	414	457	484
70%	263	324	389	440	474

iv. **Modification 4:** Speed profile correlation between LV and SV

Existing methods primarily identify LVs based on relative position without confirming their influence on the SV. This study addresses this by evaluating the correlation between the speed profiles of LV and SV using the Mexican Hat continuous WT.

Speed Profile Correlation:

Wavelet energy trends were analyzed to identify LVs that influence SV behavior. LF pairs were validated if the SV's wavelet energy profile exhibited a lagged similarity to the LV's profile. For LF pairs identified through the prior modification, the LV's influence was assessed by comparing wavelet energy plots of LV and SV speeds. LF pairs with at least four matching peak points (arbitrarily selected) in their wavelet energy plots were shortlisted.

Figure 3 illustrates the wavelet energy profiles of LF pair 408-410, and Table 4 provides a breakdown of LF pairs meeting these criteria based on LV type.

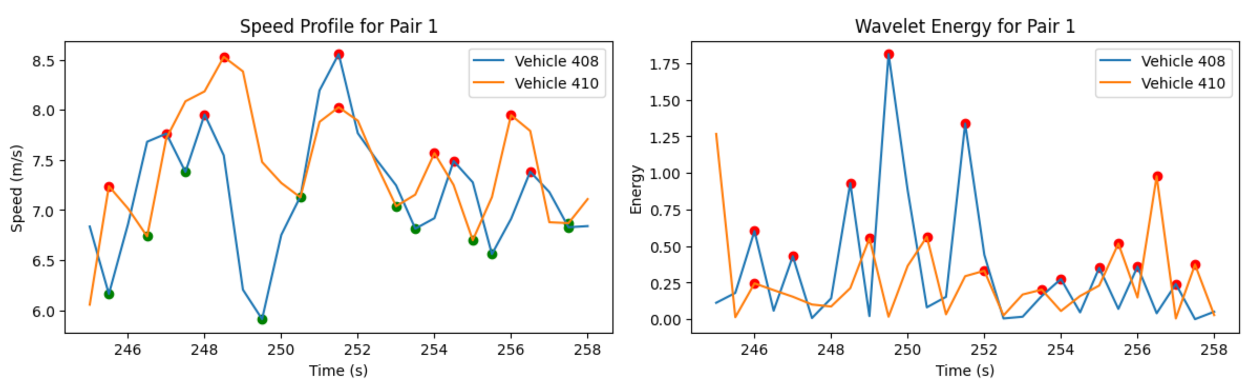

Figure 3: Sample LF pair (a) speed profile (b) wavelet energy

Table 4: LF pairs satisfying speed correlation criteria

	Original pairs	**Pairs with matching energy plots**	**Pairs with min four matching peaks**
TW-Car	114	100	35
Car-Car	217	168	63
LCV-Car	10	10	6
AUTO-Car	55	51	24

HV-Car	18	17	2
Total	414	346	130

The proposed framework improves LF identification under HWLD traffic conditions by incorporating vehicle-type-specific longitudinal gaps and speed profile correlations. Traditional methods are supplemented with these modifications to enhance accuracy. The methodology's performance is compared with existing approaches in the following results section.

3. RESULTS AND DISCUSSION

This section evaluates the performance of existing LF identification methods against the proposed approaches using a base model predicting the SV's longitudinal acceleration or deceleration.

Initially, LF pairs identified from the literature were used for modeling. These were subsequently refined through the proposed modifications. Among the first two modifications, Modification 2 was most effective in refining LF pair selection. Thus, the revised LF pairs identified using Modification 2 were used as the input for Modification 3.

Figure 4, illustrates the performance matrix after applying the Modification 3. Among the various threshold and allowance combinations, an allowance of 60% with 60% of data points satisfying the condition strikes a balance between model performance improvement and the number of LF pairs retained. This combination was selected as the input for Modification 4.

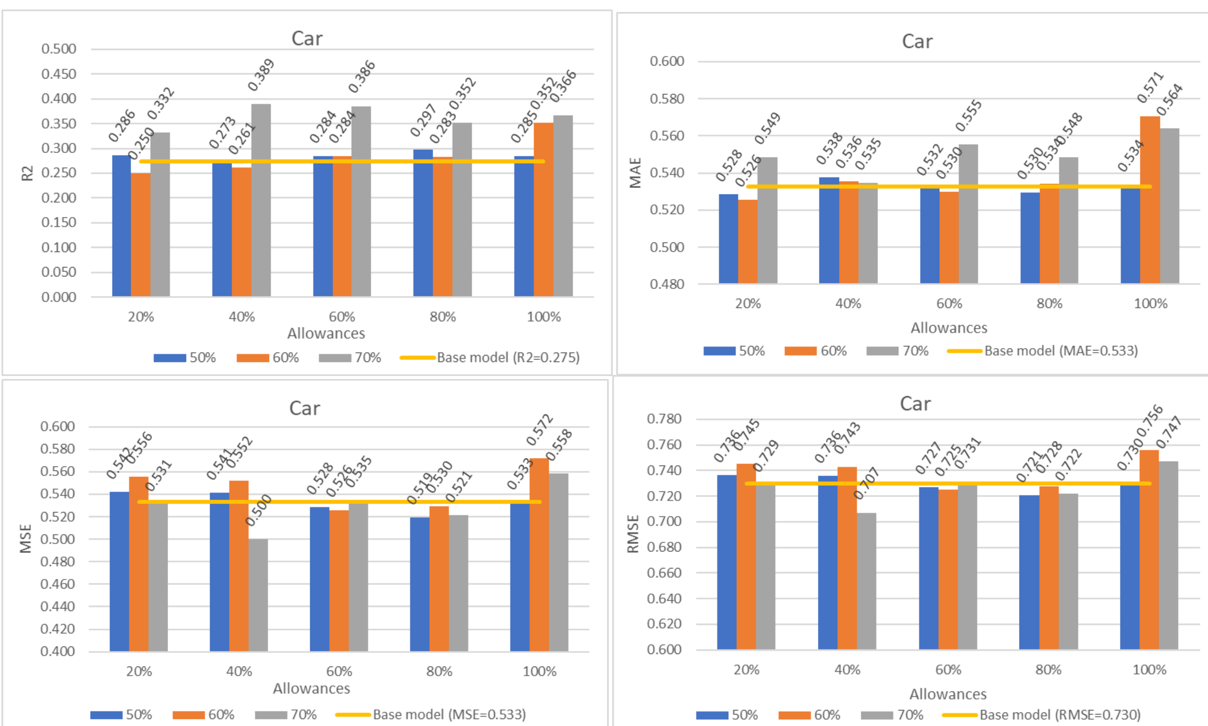

Figure 4: Performance metric with modification 3

Table 5 summarizes model performance across modifications. Incorporating vehicle-type-specific longitudinal gaps and speed profile correlations through wavelet energy analysis significantly improved LF pair identification and model accuracy.

Table 5: Model performance across different LF pair identification modifications

LF pairs	Data set	R-squared	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
From Literature	Training	0.245	0.539	0.578	0.76
	Test	0.268	0.552	0.584	0.764
With Modification 1	Training	0.247	0.524	0.530	0.728
	Test	0.261	0.523	0.506	0.712
With Modification 2	Training	0.275	0.533	0.533	0.730
	Test	0.286	0.536	0.535	0.731
With Modification 3	Training	0.284	0.530	0.526	0.725
	Test	0.259	0.536	0.559	0.748
With Modification 4	Training	0.352	0.496	0.436	0.660
	Test	0.349	0.483	0.425	0.652

The type of LV significantly impacts the SV's longitudinal behavior. Larger LVs, such as buses or trucks, exert greater influence on the SV's longitudinal gap and acceleration compared to smaller vehicles like TWs. Table 6 details the performance metrics for the refined model across different LV types.

Table 6: Model performance metrics with Modification 4 across different LV type

Model	Sample size	Data set	R-squared	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
Car	2128	Training	0.3518	0.4964	0.4358	0.6602
	912	Test	0.3485	0.4827	0.4248	0.6518
TW-Car	566	Training	0.3348	0.4824	0.4316	0.6569
	243	Test	0.293	0.5201	0.4799	0.6927
Car-Car	1026	Training	0.3683	0.4901	0.4242	0.6513
	440	Test	0.3829	0.5081	0.4658	0.6825
LCV-Car	1026	Training	0.397	0.4126	0.273	0.5225
	440	Test	0.5413	0.4888	0.359	0.5992
AUTO-Car	405	Training	0.3209	0.4605	0.3934	0.6273
	174	Test	0.4658	0.4552	0.3597	0.5998
HV-Car	39	Training	0.4915	0.5077	0.391	0.6253
	16	Test	0.2867	0.6641	0.5619	0.7496

Notably, Car-Car and LCV-Car pairs demonstrated higher R-squared values, reflecting improved predictive accuracy. However, the limited sample size for HV-Car pairs reduced the reliability of their results.

4. CONCLUSIONS

This study presents a novel approach to identifying LF pairs in HWLD traffic conditions. Using empirical trajectory data from Chennai, India, the proposed methodology was compared with existing LF identification techniques. By incorporating the k-v fundamental diagram, the approach estimates appropriate longitudinal gaps for various vehicle types across different speeds. Furthermore, WT was applied to identify similar speed patterns between LV and SV, facilitating the identification of influential LVs. The methodology's performance is compared by predicting the SV's longitudinal response in terms of acceleration or deceleration.

The results provide valuable insights into identifying LF pairs with influencing LVs and their role in understanding VF behaviour. This method provides a robust framework for extracting LF pairs from trajectory data in complex traffic settings, with significant implications for traffic engineering. While the study focuses on cars as SVs, future work could expand the methodology to include two-wheelers and other vehicle types, which play a major role in the traffic in the study area.

ACKNOWLEDGMENTS

The authors express gratitude to Kanagaraj et al. (2015) for providing open access to the mixed traffic trajectory data and to the Ministry of Human Resources Development, India, for supporting the first author through the Half-Time Research Assistantship (HTRA) during this study.

REFERENCES

Anand, P. A., Atmakuri, P., Anne, V. S. R., Asaithambi, G., Srinivasan, K. K., Sivanandan, R., & Chilukuri, B. R. (2019). Calibration of Vehicle-Following Model Parameters Using Mixed Traffic Trajectory Data. *Transportation in Developing Economies*, *5*(2), 1–11. https://doi.org/10.1007/s40890-019-0086-4

Asaithambi, G., Kanagaraj, V., Srinivasan, K. K., & Sivanandan, R. (2018). Study of traffic flow characteristics using different vehicle-following models under mixed traffic conditions. *Transportation Letters*, *10*(2), 92–103. https://doi.org/10.1080/19427867.2016.1190887

Ashok, A., & Chilukuri, B. R. (2024). A Framework to Characterise, Estimate, and Predict Vehicle Class-Agnostic Traffic States and Class-Wise Speeds for Mixed Traffic Conditions. *IEEE Access*, *12*(August), 106211–106235. https://doi.org/10.1109/ACCESS.2024.3437172

Das, S., Maurya, A. K., & Budhkar, A. K. (2019). Determinants of time headway in staggered car-following conditions. *Transportation Letters*, *11*(8), 447–457. https://doi.org/10.1080/19427867.2017.1386872

Kanagaraj, V., Asaithambi, G., Toledo, T., & Lee, T. C. (2015). Trajectory data and flow characteristics of mixed traffic. *Transportation Research Record*, *2491*, 1–11. https://doi.org/10.3141/2491-01

Kulkarni, M. M., Chaudhari, A. A., Srinivasan, K. K., Chilukuri, B. R., Treiber, M., & Okhrin, O. (2025). Leader–follower identification with vehicle-following calibration for non-lane-based traffic. *Transportation Research Part C: Emerging Technologies*, *171*(December 2024), 104940. https://doi.org/10.1016/j.trc.2024.104940

Madhu, K., Srinivasan, K. K., & Sivanandan, R. (2020). Following Behaviour in Mixed Traffic: Effects of Vehicular Interactions, Local Area Concentration and Driving Regimes. *International Journal of Engineering Research and Technology*, *13*(6), 1353–1368. https://doi.org/10.37624/ijert/13.6.2020.1353-1368

Madhu, K., Srinivasan, K. K., & Sivanandan, R. (2022). Acceleration models for two-wheelers and cars in mixed traffic: effect of unique vehicle-following interactions and driving regimes. *Current Science*, *122*(12), 1441–1450. https://doi.org/10.18520/cs/v122/i12/1441-1450

Maiti, N., & Chilukuri, B. R. (2023). Empirical Investigation of Fundamental Diagrams in Mixed Traffic. *IEEE Access*, *11*(January), 13293–13308. https://doi.org/10.1109/ACCESS.2023.3242971

Mathew, T. V., & Radhakrishnan, P. (2010). Calibration of Microsimulation Models for Nonlane-Based Heterogeneous Traffic at Signalized Intersections. *Journal of Urban Planning and Development*, *136*(1), 59–66. https://doi.org/10.1061/(asce)0733-9488(2010)136:1(59)

Mathew, T. V., & Ravishankar, K. V. R. (2011). Car-following behavior in traffic having mixed vehicle-types. *Transportation Letters*, *3*(2), 109–122. https://doi.org/10.3328/TL.2011.03.02.109-122

Nirmale, S. K., Pinjari, A. R., & Sharma, A. (2021). A discrete-continuous multi-vehicle anticipation model of driving behaviour in heterogeneous disordered traffic conditions. *Transportation Research Part C: Emerging Technologies*, *128*(November 2020), 103144. https://doi.org/10.1016/j.trc.2021.103144

Papathanasopoulou, V., & Antoniou, C. (2018). Flexible car – following models for mixed traffic and weak lane – discipline conditions. *European Transport Research Review*, *10*, 1–14.

Raju, N., Arkatkar, S., & Joshi, G. (2021). Modeling following behavior of vehicles using trajectory data under mixed traffic conditions: an Indian viewpoint. *Transportation Letters*, *13*(9), 649–663. https://doi.org/10.1080/19427867.2020.1751440