An investigation of traffic speed distributions for uninterrupted flow at blackspot locations in a mixed traffic environment

Debashis Ray Sarkar[1], Parveen Kumar[1], K. Ramachandra Rao[2], Niladri Chatterjee[3] and Sourabh B Paul[4]

[1] TRIP Centre, IIT Delhi, New Delhi, 110016, India
[2] Dept. of Civil Engineering &TRIP Centre, IIT Delhi, New Delhi, 110016, India
[3] Dept. of Mathematics &TRIP Centre, IIT Delhi, New Delhi, 110016, India
[4] Dept. of Humanities and Social Sciences, IIT Delhi, New Delhi, 110016, India

Abstract. Modelling traffic characteristics is the foundation for resolving various traffic and transportation issues. Among them, traffic speed has a significant impact on roadway crashes at blackspot (BS) locations. Speed is a random variable; several studies have recommended normal distribution to characterize the distribution of traffic speed for uninterrupted flow. However, a mixed-traffic situation causes heterogeneity, and the distribution of speeds deviates from the normal distribution. The present study investigates the distributions of traffic speeds for uninterrupted flow at 18 blackspot locations and individual vehicle types in mixed-traffic environments. Seven distribution models, namely Normal, Lognormal, Gamma, Logistic, Weibull, Burr, and Generalized Extreme Value (GEV), are considered to determine the speed characteristics. Different parametric distribution models are fitted to the vehicular speeds using maximum likelihood estimation (MLE) methods. Kolmogorov-Smirnov (KS), Anderson-Darling (AD), and two penalized criteria, i.e., Akaike and Bayesian Information Criteria (AIC and BIC), are used as goodness-of-fit (GoF) measures to find the best-fitting distribution. The overall suitability of each predicted distribution is also determined using a novel ranking method. The test findings suggest that GEV and Burr are the most suitable empirical speed distributions, with GEV fitting best above 96%. When the heavy vehicle composition (truck, bus, and tractor) is below 10%, 10–14%, 15–20%, and above 20%, it follows the Weibull, Gamma, GEV, and Burr distributions, respectively, in a mixed traffic environment.

Keywords: Speed distribution, Statistical test, Mixed traffic, Unsignalized intersection.

1 Introduction

Modelling traffic characteristics could help in identifying issues related to traffic and transportation operations [14]. An appropriate interpretation of traffic speed distribution could be important in a variety of applications such as speed limit evalua-

*Corresponding Author: Debashis Ray Sarkar
drsredefine@gmail.com

K. R. Rao et al. (eds.), *Traffic and Granular Flow '22* , Lecture Notes
in Civil Engineering 443, https://doi.org/10.1007/978-981-99-7976-9_51

tion, roadway design, safety analysis, capacity estimation, traffic noise prediction, Level of Service (LOS) analysis, bicycle performance, pedestrian walking, bus operation analysis, kinematical traffic simulation model [11–13, 22]. They may help in understanding traffic flow behavior to provide appropriate design features and regulate the traffic stream for various facility types [7, 10, 20]. A few key parameters are frequently used to analyze and build models for traffic stream characteristics [15]. Traffic speed contributes significantly to roadway crashes at blackspot locations [21]. In addition, speed distribution considerably impacts traffic generation operations in microsimulation approaches[4, 17]. Speed is a constantly varying random variable; several studies have recommended normal distribution to characterize the distribution of traffic speed for uninterrupted flow [6, 18, 19, 22]. Few studies have looked at how other distribution models could be used to describe speed patterns in mixed-traffic environments [8, 12, 13, 24]. It was suggested that a lognormal or gamma distribution could better show speed distributions than a normal distribution [5]. Zou Y emphasized skew-normal and skew-t distributions to account for the excess kurtosis in the vehicle speed distribution for the freeway traffic stream [23]. The speed variation causes the traffic stream to be heterogeneous due to slow and fast-moving vehicles. Additionally, smaller vehicles can maneuver and travel through the lateral space between larger vehicles, making the intersection more congested. However, a mixed-traffic situation causes heterogeneity, and the distribution of speeds deviates from the normal distribution [3]. At uninterrupted intersections, an appropriate speed distribution model is intended for both theoretical and simulation-based traffic modelling [9]. Statistics help in theoretical and analytical evaluation of various traffic flow applications. On the other hand, it is an integral part of the simulation study that results in the traffic simulator vehicle generation process. Thus, it is essential to recognize the correct speed distribution pattern with reasonable accuracy for different traffic compositions and vehicle types in a mixed-traffic environment. The remainder of this paper is organized as follows. First, field data collection is considered in Section 2. Statistical modelling and investigation of traffic speed is followed by Section 3. Distribution trends, GoF test results and model ranking are described in Section 4. Finally, the findings of this study are concluded in the last section.

2 Field data collection

An accident blackspot, sometimes known as a "black spot," is where crash activity has historically been concentrated. A mixed traffic condition is a designated traffic condition when you have one single road carriageway with no physical or lane markings utilized by different vehicles. If there are road lane markings, people do not follow them. Therefore, it causes heterogeneity, and the distribution of speeds deviates from the normal distribution [3]. Consequently, relying on the bell-shaped distribution for modelling vehicular speed can yield unreliable outcomes. A total of 18 different blackspots unsignalized T- intersections identified by Uttar Pradesh Public Works Department (UP-PWD) were chosen from the Indian state of Uttar Pradesh to obtain extensive vehicular speed characteristics. All intersections on two-lane rural roads (pavement widths of 7–10 m) are of particular concern because the traffic stream follows a lack of lane discipline in a mixed-traffic environment. Therefore, field data collection becomes more difficult for this study. The speed of each vehicle is measured 80–120 m upstream and downstream of the conflict zone on both approaches with a handheld LIDAR speed gun. A 30-minute video survey was used at each site to collect traffic volume data. Vehicles are classified into seven distinct categories, such as motorized two-wheelers (2W), motorized three-wheelers (3W), standard cars (Car), trucks, buses, tractors, and light commercial vehicles (LCV) as per Indo-HCM [1]. Whereas a truck, a bus, and a tractor are considered heavy vehicle composition.

3 Statistical modelling and investigation of traffic speed

This study considers seven hypothesized distributions, such as normal, lognormal, logistic, gamma, Weibull, burr and GEV distribution. Different parametric distribution models are fit to the vehicular speeds using MLE (maximum likelihood estimation) methods [2]. The investigated empirical distributions are as follows:

Normal: $$f_n(x;\sigma,\mu) = \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}} \quad (1)$$

Lognormal: $$f_{ln}(x;\sigma,\mu) = \frac{1}{x\sigma\sqrt{2\pi}}\left(exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)\right) \quad (2)$$

Logistic: $$f_{lo}(x;\sigma,\mu) = \frac{exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)}{\sigma\left(1+exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right)^2} \quad (3)$$

Gamma: $$f_g(x;\sigma,k) = \frac{\sigma^k}{\Gamma(k)}x^{k-1}\exp(-\sigma x) \quad (4)$$

Weibull: $$f_w(x;\sigma,k) = \frac{k}{\sigma}\left(\frac{x}{\sigma}\right)^{k-1} exp\left(-\left(\frac{x}{\sigma}\right)^k\right) \quad (5)$$

Burr: $$f_{br}(x;\alpha,\beta,k) = \frac{\alpha k\left(\frac{x}{\beta}\right)^{\alpha-1}}{\beta\left(1+\left(\frac{x}{\beta}\right)\right)^{k+1}} \quad (6)$$

GEV: $$f_{gev}(x;\mu,\sigma,\xi) = \frac{1}{\sigma}exp\left(-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right)\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{\xi}} \quad (7)$$

Where μ is the location; σ, β are scale, and k, ξ are shape parameters.

4 Results and discussions

4.1 Descriptive statistics for speed data

Descriptive statistics for speed data at each location are summarized in Table 1. The speed data collected is skewed; thus, it may not follow the normal distribution.

Individual vehicle speed data is also examined while considering the total sample size. As shown in Fig. 1, the box and whisker plot show that different vehicle types and driving behavior result in different speed ranges. Therefore, the normal distribution is not the best way to describe vehicular speed characteristics at unsignalized intersections. Thus, finding a suitable speed distribution for different traffic compositions and vehicle types could be useful for modelling the traffic behavior.

Table 1. Descriptive statistics for speed data at each location.

Location	Sample size	Min speed (kmph)	Mean speed (kmph)	Max speed (km)	Variances	Skewness	Kurtosis
BS 17	164	17.00	45.06	86.00	226.35	0.39	-0.16
BS 94	158	23.00	46.95	72.00	132.13	0.14	-0.57
BS 95	204	10.00	49.24	94.00	261.39	0.26	0.21
BS 96	200	23.00	55.92	96.00	252.76	0.23	-0.28
BS 114	122	18.00	44.02	66.00	125.42	-0.27	-0.56
BS 116	158	12.00	41.94	78.00	222.52	0.21	-0.72
BS 202	158	25.00	57.85	88.00	239.52	-0.12	-0.70
BS 214	152	11.00	33.93	58.00	100.41	0.38	-0.26
BS 253	116	12.00	44.36	73.00	218.73	-0.14	-0.54
BS 254	120	14.00	41.63	70.00	155.42	0.05	-0.26
BS 258	190	15.00	39.19	66.00	123.23	-0.08	-0.52
BS 260	164	13.00	43.77	70.00	211.54	-0.18	-0.85
BS 262	176	19.00	39.40	65.00	109.48	-0.08	-0.60
BS 264	158	18.00	51.85	78.00	161.16	-0.43	0.33
BS 268	174	18.00	50.05	87.00	245.14	0.06	-0.61
BS 273	178	18.00	51.76	94.00	247.91	0.22	-0.10
BS 276	116	12.00	45.12	72.00	209.37	-0.16	-0.48
BS 277	194	14.00	50.54	90.00	283.63	-0.09	-0.58

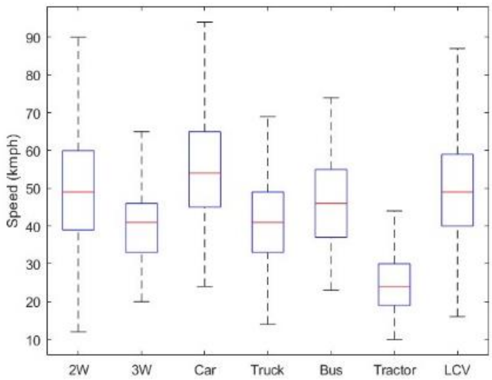


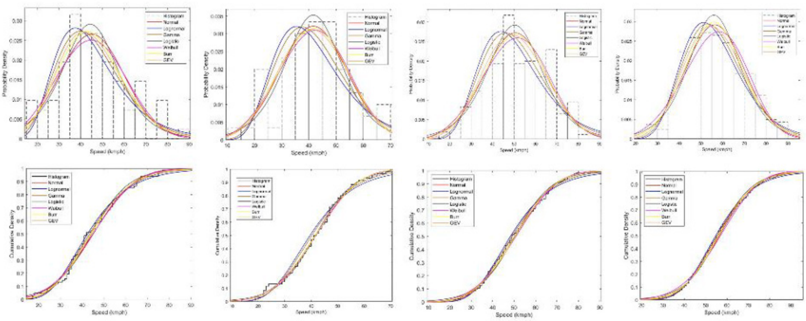
Fig. 1. Speed characteristics of each vehicle type.

4.2 Distribution trends of observed speed data

Before implementing one or more parametric candidate distributions on a dataset, choosing a predetermined set of distributions is essential. Firstly, plotting histograms and empirical distribution models for vehicular speeds implies the selection of candidate distributions. This study examines seven empirical distribution models to determine speed characteristics: normal, lognormal, Gamma, Logistic, Weibull, Burr, and GEV. MLE methods fit the location, shape, and scale parameters of different parametric distribution models to vehicular speeds at each location and for individual vehicle types, as shown in Table 2. The PDF and CDF (distribution trends) of different distribution models for two different locations and vehicle types are shown in Fig. 2. In the same way, the PDF and CDF of other selected locations and individual vehicle types are determined.

Table 2. Distribution parameters for different analytical models for each location/vehicle.

Location/	Normal		Lognormal		Gamma		Logistic		Weibull		Burr			GEV		
Vehicle	μ	σ	μ	σ	k	σ	μ	σ	k	σ	α	β	k	μ	σ	ξ
BS 17	45.06	14.95	3.75	0.35	8.64	5.26	44.36	8.58	3.25	50.28	3.53	65.95	3.82	39.03	13.69	-0.16
BS 94	46.95	11.42	3.82	0.25	16.16	2.94	46.68	6.67	4.54	51.42	6.49	73.31	4.93	42.84	11.15	-0.27
BS 95	49.24	16.09	3.83	0.38	8.21	5.88	48.67	9.00	3.30	54.81	4.40	77.32	3.78	43.09	15.48	-0.21
BS 96	55.92	15.82	3.98	0.30	11.82	4.76	55.51	9.08	3.86	61.81	4.03	80.93	4.44	49.90	15.11	-0.22
BS 114	44.02	11.11	3.75	0.28	13.85	3.23	44.40	6.49	4.60	48.25	8.37	207.60	4.61	40.78	11.68	-0.41
BS 116	41.94	14.82	3.67	0.39	7.34	5.56	41.51	8.75	3.11	46.98	4.22	32.65	3.12	36.33	14.08	-0.23
BS 202	57.85	15.38	4.02	0.29	12.78	4.55	58.15	9.08	4.30	63.66	6.30	28.47	4.30	53.06	15.86	-0.37
BS 214	33.93	9.95	3.48	0.30	11.21	3.03	33.39	5.69	3.67	37.60	2.19	41.00	4.70	29.99	9.18	-0.18
BS 253	44.36	14.66	3.73	0.39	7.60	5.88	44.53	8.52	3.40	49.42	8.20	35.30	3.41	39.92	15.23	-0.39
BS 254	41.63	12.36	3.68	0.33	10.15	4.17	41.63	7.08	3.74	46.12	17.47	95.61	3.86	37.23	12.28	-0.27
BS 258	39.19	11.08	3.62	0.31	11.16	3.57	39.40	6.47	3.99	43.28	4.36	97.40	4.00	35.38	11.2	-0.30
BS 260	43.77	14.46	3.71	0.39	7.71	5.56	44.08	8.60	3.45	48.79	9.31	35.32	3.45	39.96	15.43	-0.46
BS 262	39.40	10.40	3.63	0.29	13.06	3.03	39.56	6.09	4.29	43.34	3.39	67.83	4.30	35.8	10.49	-0.30
BS 264	51.85	12.61	3.91	0.29	14.03	3.70	52.41	6.98	4.75	56.61	3.58	94.27	4.77	48.06	13.37	-0.39
BS 268	50.05	15.57	3.86	0.34	9.35	5.26	49.91	9.14	3.58	55.62	2.56	26.05	3.91	44.50	15.36	-0.26
BS 273	51.76	15.66	3.90	0.32	10.07	5.26	51.39	8.96	3.60	57.42	4.98	82.07	4.06	45.78	15.02	-0.22
BS 276	45.12	14.34	3.75	0.38	8.23	5.56	45.33	8.32	3.56	50.14	8.09	26.01	3.56	40.81	14.94	-0.39
BS 277	50.54	16.75	3.86	0.39	7.71	6.67	50.75	9.83	3.89	56.33	7.77	40.49	3.39	44.89	16.99	-0.33
2W	50.26	14.53	3.87	0.31	11.15	4.55	49.88	8.46	3.79	55.61	10.21	97.91	4.02	45.15	14.17	0.24
3W	40.00	9.17	3.66	0.25	17.59	2.27	40.24	5.36	4.98	43.59	8.76	25.15	4.99	39.78	11.94	-0.13
Car	55.46	14.27	3.98	0.27	14.45	3.85	55.08	8.27	4.25	60.95	2.93	71.6	5.43	51.17	13.17	-0.23
Truck	41.38	11.50	3.68	0.30	12.1	3.45	41.12	6.68	3.96	45.68	7.74	77.51	4.99	42.92	13.85	-0.19
Bus	46.46	11.69	3.80	0.26	15.37	3.03	46.09	6.92	4.38	51.00	7.62	71.90	4.25	44.07	12.82	-0.22
Tractor	24.16	7.32	3.17	0.30	11.46	2.17	24.28	4.22	3.63	27.47	1.72	27.56	4.86	34.97	16.47	-0.14



a) b) c) d)

Fig. 2. PDF and CDF of different distribution models for a) BS17; b) BS254; c) 2W; d) Car.

4.3 Goodness-of-fit (GoF) test results

Different GoF tests are used in traffic engineering to assess the effectiveness and applicability of various distribution models. The goal of the GoF statistics is to figure out how different the fitted parametric distribution is from the empirical distribution. This study uses KS, AD tests, and two penalized criteria, AIC, and BIC, to find the best-fitting parametric distribution at each location and for each vehicle type.

KS: $$D_{stat} = Sup|F_n(x) - F(x)| \quad (8)$$

AD: $$n\int_{-\infty}^{\infty} \frac{(F_n(x)-F(x))^2}{F(x)(1-F(x))} dx \quad (9)$$

AIC: $$AIC = 2k - 2\ln(\hat{L}) \quad (10)$$

BIC: $$BIC = k\ln(n) - 2\ln(\hat{L}) \quad (11)$$

Null hypothesis: $$H_0: F_n(x) = F(x) \quad (12)$$

Alternate hypothesis: $$H_0: F_n(x) \neq F(x) \quad (13)$$

Value of KS test: $$D_{stat} = max|F(x) - F_n(x)|, 0 < x < \infty \quad (14)$$

Critical value of KS test: $$D_{n,\alpha} = \frac{1.35810}{\sqrt{n}}, n>50, \alpha=0.05 \quad (15)$$

Where, $F_n(x)$= empirical CDF of vehicle speed, F(x)= fitted theoretical parametric CDF, k= number of estimated parameters, n= number of observations, $\hat{L}$= maximum likelihood value for the model. Furthermore, if $D_{stat} < D_{n,\alpha}$, accept the null hypothesis, which is statistically significant; otherwise, reject the null hypothesis. Fig. 3 shows the percentage of fit for each analytical distribution. The test results show that GEV and Burr are the best empirical speed distributions, with GEV fitting above 96% in a mixed-traffic environment.

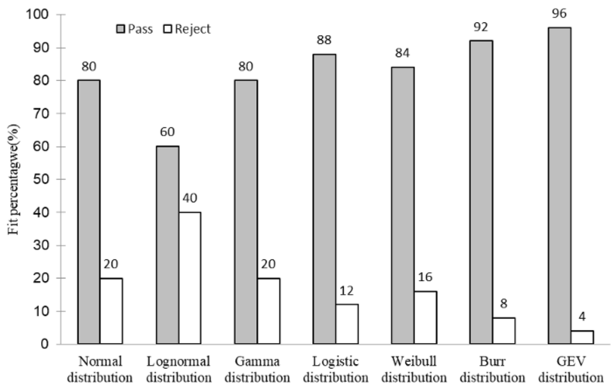


Fig. 3. Fit percentage for each hypothesis distribution.

4.4 Distribution model ranking

The overall suitability of each predicted distribution is also determined using a novel ranking method [11, 16]. The statistical test results assign a priority value (P_i) from 1 to 7 (a lower value indicates less significance) and a weightage factor (F_w) to each distribution. Therefore, a ranked value (RV_i) and total rank value (TRV) of statistically significant distributions are computed using the following equations, as shown in Table 3. Based on TRV, GEV is found the most appropriate among all other distributions.

Weightage factor: $$F_w = \frac{P_i}{\sum_{i=1}^{7} P_i} \tag{16}$$

Rank value: $$RV_i = \frac{F_w}{D_{stat}} \tag{17}$$

Total rank value: $$TRV = \sum_{i=1}^{18} RV_i \tag{18}$$

Table 3. Ranking of different analytical distributions at each location.

GoF	Normal	Lognormal	Gamma	Logistic	Weibull	Burr	GEV
KS	36.63	13.61	28.38	35.64	36.55	53.23	55.78
AD	2.98	0.00	0.87	1.97	2.76	4.00	5.60
AIC	24622.05	24947.45	24737.95	24730.49	24638.40	24609.32	24581.45
BIC	24783.87	25059.27	24849.77	24843.32	24792.22	24777.19	24749.18
Ranking	3	7	6	5	4	2	1

5 Conclusions

The results show that for most blackspot locations, Burr or GEV distributions can model speed data with superior performance under prevailing traffic conditions. In comparison, the lognormal distribution is found to fare poorly among the candidate distributions. The GEV distribution combines the Gumbel, Weibull and Fréchet families derived from extreme value theory. When the heavy vehicle composition is below 10%, 10–14%, 15–20%, and above 20%, it follows the Weibull, Gamma, GEV, and Burr distributions, respectively. This work can analyze various performance measures and flow characteristics and simulate driver behavior at unsignalized intersections. Thus, the present study suggests a new statistical distribution model that would help in choosing speed distribution characteristics for uninterrupted traffic flow at black spot locations in a mixed-traffic environment.

References

1. CSIR-CRRI (2017) Indian Highway Capacity Manual. Indian Highw Capacit Man
2. Delignette-Muller ML, Dutang C (2015) fitdistrplus: An R package for fitting distributions. J Stat Softw 64:1–34. doi: 10.18637/jss.v064.i04
3. Dey PP, Chandra S, Gangopadhaya S (2006) Speed distribution curves under mixed traffic conditions. J Transp Eng 132:475–481. doi: 10.1061/(ASCE)0733-947X(2006)132:6(475)
4. Grumert EF, Tapani A (2020) Bottleneck mitigation through a variable speed limit system using connected vehicles. Transp A Transp Sci 16:213–233. doi: 10.1080/23249935.2018.1547332
5. Haight FA, Mosher WH (1962) A practical method for improving the accuracy of vehicular speed distribution measurements. Highw Res Rec 341:92–116
6. Hashim IH (2011) Analysis of speed characteristics for rural two-lane roads: A field

study from Minoufiya Governorate, Egypt. Ain Shams Eng J 2:43–52. doi: 10.1016/j.asej.2011.05.005
7. Jamshidnejad A, De Schutter B (2015) Estimation of the generalised average traffic speed based on microscopic measurements. Transp A Transp Sci 11:525–546. doi: 10.1080/23249935.2015.1026957
8. Jun J (2010) Understanding the variability of speed distributions under mixed traffic conditions caused by holiday traffic. Transp Res Part C Emerg Technol 18:599–610. doi: 10.1016/j.trc.2009.12.005
9. Kosun C, Ozdemir S (2016) A superstatistical model of vehicular traffic flow. Phys A Stat Mech its Appl 444:466–475. doi: 10.1016/j.physa.2015.10.042
10. Liu H, Rodgers MO, Liu F "Cherry", Guensler R (2020) Bayesian approach in estimating the road grade impact on vehicle speed and acceleration on freeways. Transp A Transp Sci 16:602–625. doi: 10.1080/23249935.2020.1722280
11. Maghrour Zefreh M, Török Á (2020) Distribution of traffic speed in different traffic conditions: An empirical study in budapest. Transport 35:68–86. doi: 10.3846/transport.2019.11725
12. Maurya AK, Das S, Dey S, Nama S (2016) Study on Speed and Time-headway Distributions on Two-lane Bidirectional Road in Heterogeneous Traffic Condition. In: Transportation Research Procedia. Elsevier B.V., pp 428–437
13. Maurya AK, Dey S, Das S (2015) Speed and Time Headway Distribution under Mixed Traffic Condition
14. Mondal S, Arya VK, Gupta A (2022) An optimised approach for saturation flow estimation of signalised intersections. Proc Inst Civ Eng Transp 175:137–149. doi: 10.1680/jtran.18.00206
15. Mondal S, Gupta A (2019) Assessment of vehicles headway during queue dissipation at signal-controlled intersection under mixed traffic. Curr Sci 116:437–444. doi: 10.18520/cs/v116/i3/437-444
16. Mondal S, Gupta A (2021) Speed distribution for interrupted flow facility under mixed traffic. Phys A Stat Mech its Appl 570. doi: 10.1016/j.physa.2021.125798
17. Ou H, Tang TQ, Zhang J, Zhou JM (2018) A car-following model accounting for probability distribution. Phys A Stat Mech its Appl 505:105–113. doi: 10.1016/j.physa.2018.03.072
18. Roess R. P. Prassas E. S. & McShane W. R. (2019) Traffic engineering (Fifth). Pearson
19. Roy R, Saha P (2018) Headway distribution models of two-lane roads under mixed traffic conditions: a case study from India. Eur Transp Res Rev 10. doi: 10.1007/s12544-017-0276-2
20. Tang TQ, Zhang J, Liu K (2017) A speed guidance model accounting for the driver's bounded rationality at a signalized intersection. Phys A Stat Mech its Appl 473:45–52. doi: 10.1016/j.physa.2017.01.025
21. Transportation Research Wing (2021) ROAD ACCIDENTS IN INDIA
22. Wang H, Ni D, Chen QY, Li J (2013) Stochastic modeling of the equilibrium speed-density relationship. J Adv Transp 47:126–150. doi: 10.1002/atr.172
23. Zou Y, Yang H, Zhang Y, Tang J, Zhang W (2017) Mixture modeling of freeway speed and headway data using multivariate skew-t distributions. Transp A Transp Sci 13:657–678. doi: 10.1080/23249935.2017.1318973
24. Zou Y, Zhang Y (2011) Use of skew-normal and skew-t distributions for mixture modeling of freeway speed data. Transp Res Rec 67–75. doi: 10.3141/2260-08