



Speed Distribution on Two-Lane Rural Highways with Mixed Traffic: A Case Study in North East India

Pritam Saha¹ · Nabanita Roy¹ · Ashoke Kumar Sarkar² · Manish Pal³

Received: 11 August 2016 / Accepted: 14 June 2017 / Published online: 20 June 2017
© The Institution of Engineers (India) 2017

Abstract This work focuses on minimising inaccuracies in distributional assumptions of speed data on two-lane roads with heterogeneous traffic to improve accuracy in capacity and level of service analysis. Accordingly, field study was conducted on a two-lane highway in India that exhibits heterogeneity in its traffic composition. Two distribution functions, namely, normal and logistic were examined for the observed speed data. The appropriate function was chosen using a methodology based on K–S test and field validation. Logistic distribution function was found to exhibit its aptness in describing speed under such traffic and was, thereby, considered in estimating the limiting speed of slower vehicles that tends to obstruct the flow of traffic. Vehicles that move at or below this speed cause delay to the faster ones resulting in formation of platoons at moderate and heavy flow when passing is restricted. Since the percent time-spent-following, a performance measure to assess level-of-service of two-lane highways, considers proportion of vehicles that are trapped inside platoons, it is imperative to estimate the limiting speed of slower vehicles correctly to approximate the delayed vehicles.

Keywords Two-lane highways · Heterogeneous traffic · Speed distribution · Slower vehicles

Introduction

In most of the developing countries like India the prevailing traffic is heterogeneous in character and comprises of a wide range of vehicle category in terms of static and dynamic characteristics. They share the same road space resulting in an increased interaction between them which eventually reduce mobility considerably. This is more acute on two-lane roads where such interaction takes place even with the opposing traffic. Vehicles, which are ‘less mobile’, like non-motorized and also, low performance motorized ones cause significant level of friction to the movement of other vehicles in the traffic stream [1]. Accordingly, they create impedance to the movements of faster vehicles particularly at moderate and heavy flow when they have limited passing opportunities. As a result, they are forced to follow impeding slower vehicles and start moving in a platoon. Formation of such platoon is frequent on such roads in the event of significant presence of slower vehicles and absence of traffic segregation.

Vehicles that are entrapped inside platoons get delayed and sometimes drivers become impatient [2]. A few drivers take considerable risk to complete passing manoeuvre if delay is long enough. At the same time, however, a few vehicles in the traffic stream travel inside a platoon at their desired speed i.e. their speed is not impeded by the slower ones [1]. This attributes to their movement in a platoon by choice as their desired speed is close to the speed of platoon leaders. This reveals the fact that different types of vehicles moving on the same road may enjoy different

✉ Pritam Saha
saha.pritam@gmail.com

¹ Civil Engineering Department, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal 711103, India

² Civil Engineering Department, Birla Institute of Technology and Science Pilani, Vidya Vihar Campus, Pilani, Rajasthan 333 031, India

³ Civil Engineering Department, National Institute of Technology Agartala, Agartala, Tripura 799046, India

levels of service (LOS) even under the same operating condition.

The *highway capacity manual* suggests the use of Percent Time-Spent-Following' (PTSF) as a major determinant of LOS on two-lane roads [3]. Assessment of PTSF considers the average percentage of total travel time that vehicles must travel in platoon behind slower vehicles due to inability to pass. However, both, risk taking behaviour of drivers and their movement in platoons by choice under heterogeneous traffic, have significant implications on its assessment. Accordingly, such traffic poses a serious challenge to traffic planners and engineers who are on the lookout for establishing an indigenous method for assessing LOS.

Theoretically, there would be no platooning and consequent delays, if all the vehicles start travelling at the mean speed. However, diverse vehicular characteristics virtually lead to a possibility of wide range of operating speeds in the traffic stream. These calls for an initiative in identifying the limiting speed of slower vehicles that tends to obstruct the flow of traffic and causes delay to the faster ones. Proportion of vehicles that move at or below this limiting speed is considered while assessing PTSF. Typically, such speed refers to a threshold above which 85% motorists' travel [4] and thereby, could be determined on the basis of appropriate distribution function. Conventionally, normal distribution exhibits its aptness in describing speed data; however, it deviates significantly in the event of heterogeneity in traffic mix [5]. Thus, it becomes imperative to develop a compatible model for such traffic as well in order to identify the threshold.

Literature Review

There have been a number of studies that investigated appropriate mathematical model to describe observed speed data. The majority of the previous studies confirmed that the spot speed data usually follow a normal distribution [6–10]. A study, however, indicates that car speeds are approximately normally distributed if coefficient of variation is in the range of about 0.11–0.18 on low volume two-lane roads [11, 12]. At the same time, some studies reported the application of several other distributions to describe the speed data. Gamma or log-normal was found to have adequate compatibility with time speeds [13]. The advantages that they offer are the same functional form is retained even when the time speed distribution is transformed into a space-speed distribution and avoid the theoretical difficulty of the negative speeds given by the infinite tails of the normal distribution. Experience on Indian traffic makes it clear that a normal distribution

describes the spot speeds of cars, heavy commercial vehicles, light commercial vehicles, scooters, and cycle rickshaws (pedal tricycles) well, whereas a log-normal distribution holds good for bicycles [14–17]. A fairly recent study in India clearly indicates speed deviates from the normal curve in the event of heterogeneity in traffic mix due to the large variation in speeds of faster and slower particularly the non-motorized vehicles [5]. The study introduced the concept of 'spread ratio' and suggested a threshold of it for examining the normality of a distribution curve.

Accurate measurement of speeds and their evaluation based on in-depth statistical analysis, however, invite further attention. Couple of international studies suggest use of either moving car observer [18] or free gap evaluation [19] method while determining free-flow speed. An effort aimed at analysing such data reveals that obtaining the 85th percentile speed from regression modelling gives much better estimates than those from the normal approximation model [20]. A study suggests developing neural-network-based speed models to establish a relationship between the roadway characteristics and the 85th percentile speed [21] and ensures reasonable degree of accuracy of model predictions. By the same token, a few studies suggest the use of several other models: back propagation artificial neural networks particularly while predicting speed of passenger cars on two-lane rural highways [22], speed-profile model to evaluate the design consistency for prediction of the speed along an alignment [23], operating speed prediction model for trucks on two-lane rural highways for use in design consistency [24].

Most of the above studies have made it evident that wide variation in speed under heterogeneous traffic would cause the speed data to deviate from the normal distribution. This is more acute on most of the two-lane roads in India where the same road space is shared by a variety of motorized and non-motorized modes of transportation [25, 26] resulting in evaluation of operational conditions challenging. Therefore, it is imperative to develop an in-depth understanding of the distribution pattern of the speed data and also to predict the percentile speed with reasonable amount of accuracy. This would help in arriving at the threshold of limiting speed, a major platooning variable of two-lane roads, which has significant implications on PTSF.

Study Sites and Field Data

The limiting speed of slower vehicles that frequently creates impedance to the flow of traffic and causes formation of platoons as a consequence could be assessed based on

speed percentiles or quantiles. An analyst, therefore, has to resort to field measurement in order to estimate the values. Further, it is critical to keep out the faster vehicles that are entrapped inside a platoon to explicitly explain such percentiles; otherwise, increased proportion of impeded vehicles would eventually mislead the results. Thus, a field study at low flow level when platoon formation is infrequent is imperative for the collection of speed data.

Field study was conducted on a two-lane national highway (popularly known as Assam–Agartala road) in north-east India. The study segment was selected close to the capital city, Agartala of the state of Tripura where platoon movements are frequent due to large volume of city bound traffic. Also, it was free from the effect of intersection, curvature, ribbon development and the pavement condition was good and uniform (Fig. 1). The free-flow situation was approximated when the traffic volumes do not exceed 200 veh/h [3] and the hourly traffic variation was, accordingly contemplated to identify it.

A longitudinal trap of 10 m length was made on the carriageway and a video camera was placed away from the trap in order to avoid any influence in the operating speeds of vehicles. Further, it was mounted on a stand height of which was adjusted in such a way such that it covers the entire trap length keeping some margin on either side. The recorded video files were then played in a computer and the time taken to cover the trap length by each vehicle was measured with an accuracy of 0.01 s. The spot speeds were computed on the basis of lapsed times of vehicles passing through the section [26].

A variety of vehicle categories including both motorized and non-motorised modes were observed in the traffic stream. The directional segments were studied separately and presence of car (about 30–35%), bike (motorized two-wheeler) (about 30–35%) and non-motorised modes of transportation (about 10%) was significant when compared

to other modes of transportation. An informal opinion poll of about 50 users indicates that lack of adequate public transportation facilities to the commuters who live in city outskirts, compelled them to rely more on their own vehicles or para-transit modes like bi-cycles or paddle tri-cycles.

Trends in Speed Distribution

Speed is a continuous random variable and, on highways, normally it exhibits a symmetrical distribution about a central value. Three distribution functions, namely, triangular, normal and logistic could, therefore, be applied while describing such data (Fig. 2). The basis of selecting an appropriate one is, however, the extent of outliers in the sample data; in a way that data points are further away from the sample mean. Notably, occurrence of such outliers is very likely at the time of recording speed data due to the measurement error and also variability in vehicular characteristics. It exaggerates further in the event of heterogeneous traffic composed of a wide range of vehicle categories in terms of static and dynamic characteristics. Accordingly, proportion of outliers increases considerably under such traffic resulting in a heavy-tailed distribution of the samples.

Statistically, speed data would follow a normal distribution if the proportion of outliers is insignificant; this is apparent when the traffic is more or less homogeneous in character. The distributional choice, however, deviates from the generally accepted normal distributions under heterogeneous traffic with large speed differential [5]. Suitable statistical models should, therefore, be resorted to in the event of such heterogeneity in traffic mix. The present study considers logistic distribution as an alternative of normal distribution in describing speed data as it has a



Fig. 1 A view of the study section on Assam–Agartala road (photos by the authors)

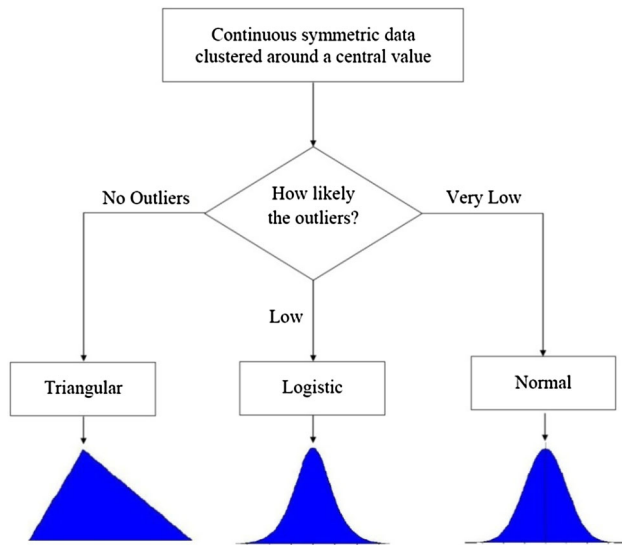


Fig. 2 Statistical basis of distributional choices for speed data

quite similar shape. Besides, it has heavier tails, which often increases the robustness of analyses based on it compared to normal distribution.

Hypothetically, the logistic distribution offers advantages over the conventional normal distribution by means of its functional simplicity and mathematical accuracy. A normal distribution needs numeric approximation, whereas the proposed one can be solved analytically, thus, could be used instead at the time of modelling speed data. Equations 1 and 2 demonstrates the probability density functions of normal and logistic distributions which are applied to the field data.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \times \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad (1)$$

$$f(x) = \exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right) \times \frac{1}{\sigma(1 + \exp(-\left(\frac{x-\mu}{\sigma}\right)))^2} \quad (2)$$

where σ , continuous scale parameter ($\sigma > 0$); μ , continuous location parameter.

Appropriate distribution function for the observed field data is, however, decided based on goodness-of-fit tests. In traffic engineering problems, two such tests namely Chi square test and the Kolmogorov–Smirnov (K–S) test are commonly used. The present study applied the K–S test while identifying the best fitted models as it offers certain advantages over the Chi square test; K–S test can use data with a continuous distribution and there is no minimum frequency per test interval [27]. The test statistic is calculated by determining the difference between the cumulative percentage of the measured frequency and the cumulative percentage of the expected frequency. The largest of these differences over the entire measured population interval is the test statistic ‘D’ [28] and it is computed for the

distribution models at the desired significance level in order to find the best fitted one.

Statistical Investigation

The field data was grouped in the form of frequency distribution based on Sturges’ rule [29, 30]. According to the rule, the width of class intervals should be equal to $\text{Range}/[1 + 3.322 \log_{10}n]$, where, n is the number of observations, and Range, the difference between maximum and minimum values of observed speed data. However, if the class interval is too small or large, the resultant histogram will have either a ‘ragged’ or ‘block like’ shape. It would subsequently mask the true shape of underlying density and affect the statistical investigations at large. In such cases, different class intervals may be compared, while choosing the appropriate one that gives a smooth histogram [31].

A class interval of 9 kmph was accordingly obtained on the basis of speed data analysis and used in frequency distribution. The histogram shows that speeds of vehicles tend to cluster about the mean value and frequency drops as the speeds depart from the mean (Fig. 3). Probability density functions of normal and logistic distributions (Eqs. 1 and 2) were fitted to the field data and observed to have symmetrical shape (Fig. 3). Table 1 provides the descriptive statistics and goodness-of-fit details of the distribution functions. A close look into the table reveals that coefficient of variation decreases from 0.295, as obtained respectively from field data and normal distribution model, to 0.163 when derived from logistic distribution model; this attributes to a reduction in standard deviations. Further, Fig. 3 shows that the smaller standard deviation as obtained from logistic distribution is largely due to thicker tails and higher kurtosis than the normal.

The goodness-of-fit of the models was checked using Kolmogorov–Smirnov (K–S) tests with 5% level of significance (α). The null hypotheses for each test were as follows: ‘The compatibility hypotheses of speed distribution with fitted model is rejected (P value $< \alpha$) or not rejected (P value $> \alpha$)’. The statistic D value of K–S test was calculated for the fitted models and compared in order to determine the extent of fit to the observed data. Table 1 shows the model parameters estimated when it best fits the data based on the minimum statistic value. The logistic distribution is deemed to describe the speed data well as the test statistic that represents the largest difference between measured and expected frequency was observed to be lower. Also, the observed P values were more than 0.05 which indicates that null hypotheses are accepted for both the models.

With the aim of visual assessment of the goodness-of-fit for logistic distribution model, P–P (probability vs.

probability) and Q–Q (quantile against quantile) plots are provided. In statistics, a Q–Q plot is a probability plot, which is a graphical method of comparing the distribution of sample data and the chosen theoretical distribution by plotting their quantiles against each other. Conversely, the P–P plot is the graph of percentiles of one distribution versus the percentiles of another. If the proposed distribution is consistent with the characteristics of the sample data, the plot will lie on the 45-degree line. The illustration of the probability plots obtained for the observed speed data with logistic distribution model (Fig. 4) makes it evident that the data points (considering both quantiles and percentiles) are very close to the 45-degree line, thereby, signifying satisfactory agreement.

However, it is imperative to test the validity of model outcomes while representing the existing traffic system; in a way that it reproduces the system behavior with good amount of accuracy. A further attempt was, therefore, made to assess the validity of the proposed model on the basis of a pilot study conducted on a different study segment of the same highway section. Theoretically calculated probabilities of speeds were plotted against those observed at field in

order to compare how rational the outcomes are with respect to field observations. Figure 5 illustrates that Logistic distribution exhibits reasonably good agreement when compared to Normal distribution. The variability of predictions was determined for both the distributions and expressed in terms of standard error of the estimate (SEE). A ‘SEE’ value of 0.055 was obtained for Logistic distribution, whereas it increases up to 0.113 in case of Normal distribution.

Since the study aims to reduce measurement errors, the logistic distribution function was considered appropriate in describing speed data as it represents the prevailing traffic better. The speed percentiles were determined and the 15th percentile speed which is regarded as the limiting speed of slower vehicles was found to be about 25 kmph. This value is considered as unreasonably slow when compared to the trend of traffic and proportion of vehicles that move at or below this limiting speed is considered while assessing PTSF. It evidently indicates two major implications on PTSF: following by choice of the low performance vehicles [1] and risk taking behaviour of drivers [2], thereby, specifying the need of identifying an alternative LOS measures for two-lane roads.

Table 1 Speed distribution parameters and goodness-of-fit details of the distribution models

Distribution function	Mean, km/h	Standard deviation, km/h	Coefficient of variation	K–S test statistic	P value	α	Null hypotheses (accept/reject)
Observed data	37.095	10.947	0.295	–	–	–	–
Logistic distribution	37.095	6.0356	0.163	0.184	0.703	0.05	Accept
Normal distribution	37.095	10.947	0.295	0.195	0.635	0.05	Accept

Bold value indicates lowest K–S test statistic

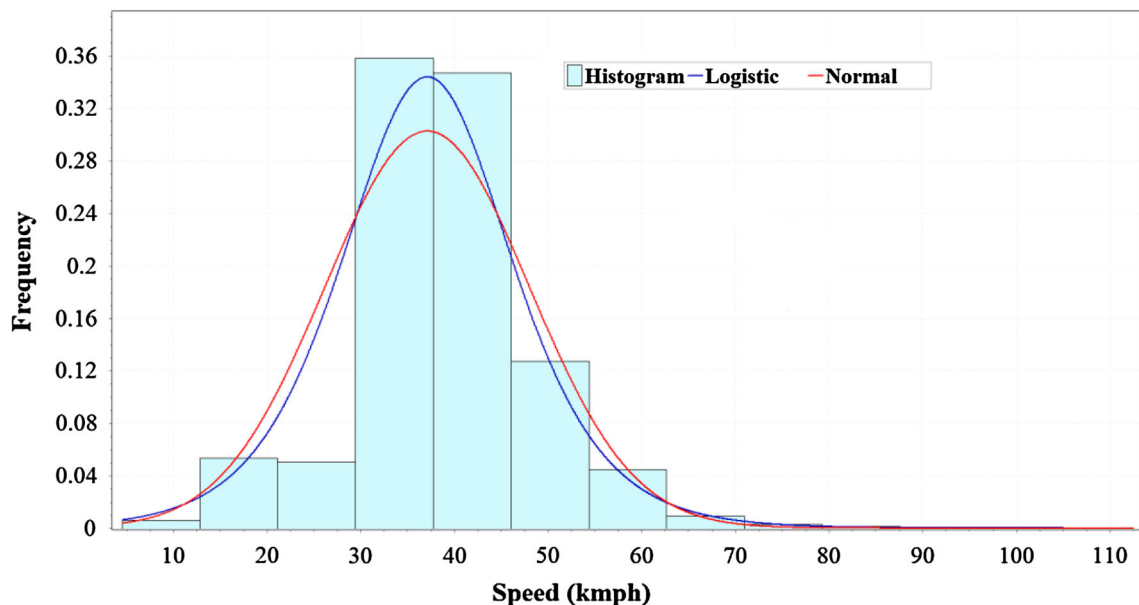


Fig. 3 Probability density functions of normal and logistic distributions fitted to the observed speed data

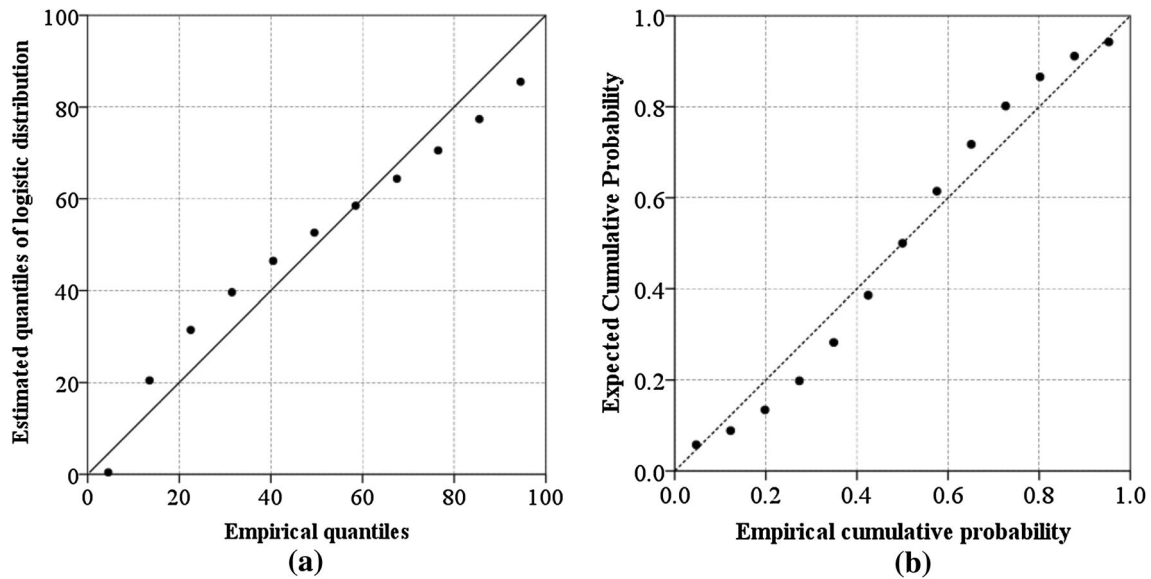


Fig. 4 Probability plots for the observed speed data with logistic distribution model: **a** Q–Q plot and **b** P–P plot

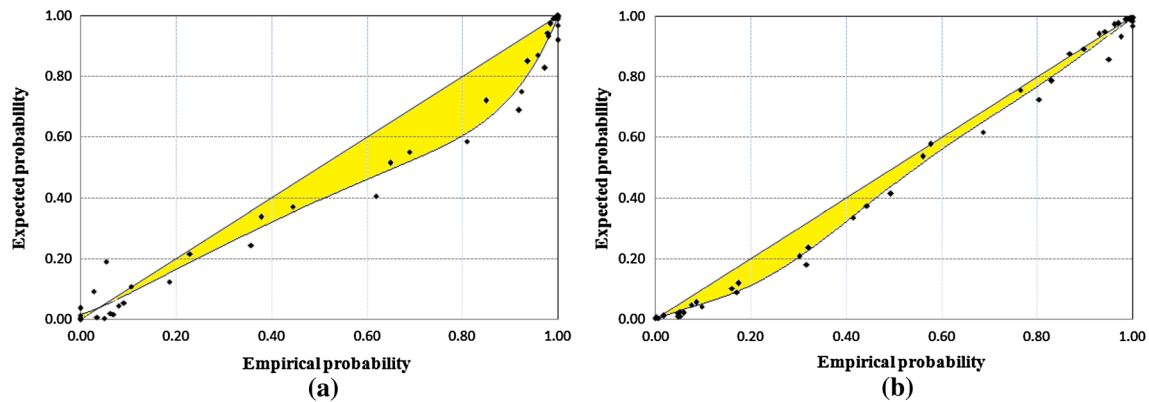


Fig. 5 Agreement between the empirical and theoretically calculated probabilities of speeds at a different study site: **a** normal distribution and **b** logistic distribution

Conclusion

The speed study was conducted on a two-lane highway during low flow when platoon formation is infrequent to collect speed data of free moving vehicles. A wide range of operating speeds due to diverse vehicular characteristics was observed in the traffic stream and their distribution was symmetrical about a central value. Two distribution functions, namely, normal and logistic were, therefore, examined for the observed speed data as they have a quite similar shape. The appropriate function was chosen based on goodness-of-fit test and field validation.

Logistic distribution function was found appropriate in describing the observed speed data and, accordingly, the speed percentiles were derived from it. The limiting speed of slower vehicles that tends to obstruct the flow of traffic

was found to be about 25 kmph. Vehicles in the traffic stream that move at or below this speed cause delay to the faster ones and thereby, platoons are formed at moderate and heavy flow when passing is somewhat restricted. Proportion of the vehicles that are trapped inside platoons are considered while estimating PTSE, a performance measure to assess level-of-service of two-lane highways. It is, therefore, paramount to arrive at an accurate estimate of limiting speed to approximate the delayed vehicles.

Since the premise on which the present study is based aims to reduce measurement errors, it warrants efforts towards minimising inaccuracies in distributional assumptions. This is attributable to the fact that limitations in accuracy and precision would cause an uncertainty in results and mislead traffic analysts particularly while assessing the level-of-service. The present study, thus,

creates a starting point of further initiatives aimed at establishing a robust method of modelling speeds on two-lane rural highways with mixed traffic based on comprehensive field data.

Acknowledgement The authors are grateful to the anonymous referees for their suggestions to improve the paper.

References

- P. Saha, A. Bhadra, N.S. Reddy, A.K. Sarkar, Method of identifying low performance vehicles in heterogeneous traffic on two-lane highways. *Procedia Soc. Behav. Sci.* **104**, 526–532 (2013)
- P. Saha, A.K. Sarkar, M. Pal, Evaluation of performance measures of two-lane highways under heterogeneous traffic. *Pertanika J. Sci. Technol.* **23**(2), 223–239 (2015)
- National Academics, *Highway Capacity Manual*, 5th edn. (TRB, Washington, 2010)
- Y. Hou, C. Sun, P. Edara, Statistical test for 85th and 15th percentile speeds with asymptotic distribution of sample quantiles. *Transp. Res. Rec.* **2279**, 47–53 (2012)
- P.P. Dey, S. Chandra, S. Gangopadhaya, Speed distribution curves under mixed traffic conditions. *J. Transp. Eng. ASCE* **132**(6), 475–481 (2006)
- V.M. Kumar, S.K. Rao, Headway and speed studies on two-lane highways. *Indian Highw.* **26**(5), 23–36 (1998)
- M.N. Islam, P.N. Seneviratne, Evaluation of design consistency of two-lane rural highways. *ITE J.* **64**(2), 28–31 (1994)
- J.A. Kerman, M. McDonald, G.A. Mintsis, Do vehicles slow down in bends? A study into road curvature, in *Proceedings of Driver Behaviour and Design, PTRC. 10th Summer Annual Meeting, Seminar K*, pp. 57–67 (1982)
- M. Hossain, G.A. Iqbal, Vehicular headway distribution and free speed characteristics on two-lane two-way highways of Bangladesh. *J. Inst. Eng. (India) Part AG* **80**, 77–80 (1999)
- I.H. Hashim, Analysis of speed characteristics for rural two-lane roads: a field study from Minoufiya Governorate, Egypt. *Ain Shams Eng. J.* **2**(1), 43–52 (2011)
- H.J.W. Leong, The distribution and trend of free speeds on two-lane two-way rural highways in New South Wales, in *Proceedings of 4th Australian Road Research Board Conference, Part 1, Australian Road Research Board, Vermont South, Victoria, Australia*, pp. 791–808 (1968)
- J.R. McLean, Observed speed distributions and rural road traffic operations, in *Proceedings of 9th Australian Road Research Board Conference, Part 5, Australian Road Research Board, Vermont South, Victoria, Australia*, pp. 235–244 (1978)
- F.A. Haight, W.W. Mosher, A practical method for improving the accuracy of vehicular speed distribution measurements, in *HRR 341, Highway Research Board, Washington, D.C.* pp. 92–116 (1962)
- B.K. Katti, S. Raghavachari, Modelling of mixed traffic speed data as inputs for the traffic simulation models, in *Highway Research Bulletin 28, Indian Roads Congress, New Delhi, India*, pp. 35–48 (1986)
- K.S. Pillai, T.V. Ramanayya, Traffic inputs for simulation of mixed traffic on a digital computer, in *Proceedings of National System Conference, Indore, India*, pp. 66–71 (1977)
- J.R. McLean, *Two-Lane Highway Traffic Operations: Theory and Practice*, 6th edn. (Gordon and Breach Science, New York, 1989)
- M.L. Rai Chowdhury, A study of some traffic characteristics in an Indian metropolis and their application in simulation of traffic flow through a signalized intersection. Ph.D. thesis, Indian Institute of Technology, Kharagpur, India (1993)
- U.T. Abdurrahman, O.C. Puan, M.N. Ibrahim, comparison of free-flow speed estimation models, in *National Seminar on Civil Engineering Research (SEPKA), Malaysia*, vol. 1, p. 69 (2014)
- A. Lobo, M.A.P. Jacques, C.M. Rodrigues, A. Couto, Free-gap evaluation for two-lane rural highways. *Trans. Res. Rec.* **2223**, 9–17 (2011)
- A.S. Al-Ghamdi, Spot speed analysis on urban roads in Riyadh. *Transp. Res. Rec.* **1635**, 162–170 (1998)
- Y. Najjar, R. Stokes, E. Russell, Setting speed limits on Kansas two-lane highway: neuronet approach. *Transp. Res. Rec.* **1708**, 20–27 (2000)
- J. McFadden, W.T. Yang, S.R. Durrans, Application of artificial neural networks to predict speeds on two-lane rural highways. *Transp. Res. Rec.* **1751**, 9–17 (2001)
- K. Fitzpatrick, J.M. Collins, Speed-profile model for two-lane rural highways. *Transp. Res. Rec.* **1737**, 42–49 (2000)
- E.T. Donnell, Y. Ni, M. Adolini, L. Elefteriadou, Speed prediction models for trucks on two-lane rural highways. *Transp. Res. Rec.* **1751**, 44–55 (2001)
- P. Penmetsa, I. Ghosh, S. Chandra, Evaluation of Performance measures for two-lane intercity highways under mixed traffic conditions. *J. Transp. Eng. ASCE* (2015). doi: [10.1061/%28ASCE%29TE.1943-5436.0000787](https://doi.org/10.1061/%28ASCE%29TE.1943-5436.0000787)
- P. Saha, A.K. Sarkar, M. Pal, Evaluation of speed–flow characteristics on two-lane highways with mixed traffic. *Transport* (2015). doi: [10.3846/16484142.2015.1004369](https://doi.org/10.3846/16484142.2015.1004369)
- F. Ye, Y. Zhang, Vehicle type-Specific headway analysis using freeway traffic data. *Transp. Res. Rec.* **2124**, 222–230 (2009)
- H.W. Lilliefors, On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **62**(318), 399–402 (1967)
- H.A. Sturges, The choice of a class interval. *J. Am. Stat. Assoc.* **21**(153), 65–66 (1926)
- D.W. Scott, Sturges’ rule. *Wiley Interdiscip. Rev. Comput. Stat.* **1**(3), 303–306 (2009)
- A.M. Law, D.W. Kelton, *Simulation Modelling and Analysis*, 3rd edn. (McGraw-Hill Higher Education, Singapore, 2000)