JOURNAL OF TRANSPORTATION
SYSTEMS ENGINEERING AND INFORMATION TECHNOLOGY
Volume 12, Issue 6, December 2012
Online English edition of the Chinese language journal

Cite this article as: J Transpn Sys Eng & IT, 2012, 12(6), 72−84.

RESEARCH PAPER

Time Gap Modeling under Mixed Traffic Condition: A Statistical Analysis

DUBEY Subodh Kant, PONNU Balaji, ARKATKAR Shriniwas S*

Civil Engineering Department, Birla Institute of Technology & Science Pilani-333031, Rajasthan, India

Abstract: This paper attempts to model vehicular time gap, which is defined as the time interval between any two successive arrivals of vehicles at a reference point of measurement on a road segment. Such an approach is justified under the non-lane-based heterogeneous traffic conditions prevailing in developing countries such as India, characterized by many "zero" time gaps due to simultaneous arrivals within a given road width. In addition, time gap data are characterized by a significant amount of data in the tail region due to long headways. Nevertheless, many researchers of time gap modeling have used light-tailed distributions that modeled time gaps satisfactorily due to two reasons: (a) The tail data was merged into a single bin; and (b) goodness-of-fit tests such as the Chi-square test, which has many limitations, were used. Further, some researchers have suggested different distributions for the same range of traffic flows, leading to ambiguity in distribution selection. In addition, bin size, which dictates the degree of fit of any distribution, has been ascribed very less importance in time gap modeling. Hence, this paper tries to consolidate and standardize the existing research in time gap modeling research by addressing all these issues. Two new distributions, namely Generalized Pareto (GP) and Generalized Extreme Value (GEV) with better tail modeling properties, have been proposed along with other conventional distribution to model vehicular time gaps over a wide range of flow from 550 vph to 4,100 vph. Two types of goodness-of-fit tests, namely Area-based and Distance-based tests, have been used. It has been found from the study that GP distribution fits the time gap data well (overall and tails) up to a flow range of 1,500 vph based on both kinds of tests, and GEV fits the data well for the flow levels above 1,500 vph based on the area test only.

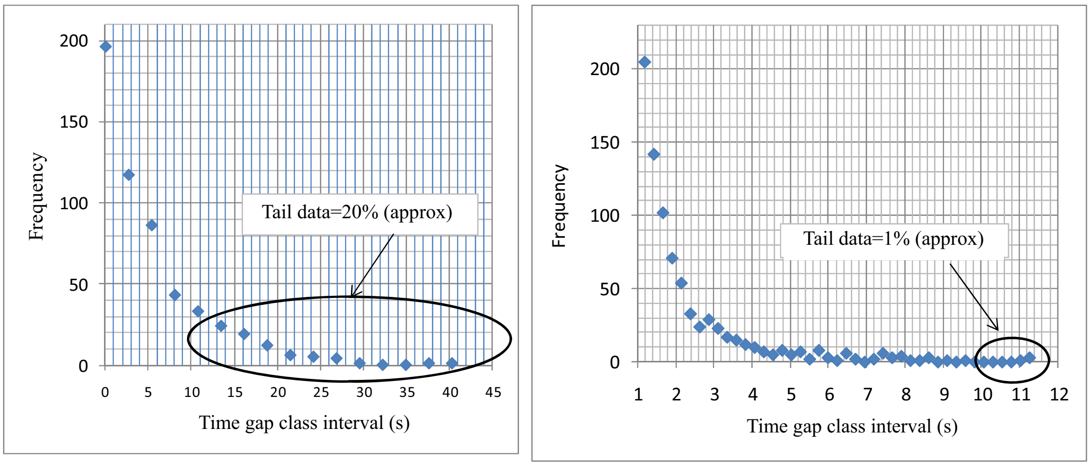
Key Words: highway transportation; vehicular time gap; heterogeneous traffic flow; bin size; K-S tests; A-D tests

1 Introduction

Traffic conditions prevailing in developing countries such as India, China, Bangladesh, and Sri-Lanka are heterogeneous; that is, they are mixed in nature and comprise several categories of vehicles with varying dimensions, maneuvering capabilities, and speed. There are broadly nine to ten categories of vehicles ranging from slow moving vehicles such as cycles, motorcycles, three wheelers (auto-rickshaw), and pick-ups to fast moving vehicles such as cars, vans, and 2-axle and 3-axle trucks. There is an imperfect or no-lane discipline, with vehicles rarely following the lane markings and not strictly following any leading vehicle unlike under homogeneous traffic conditions. Further, there can be more than one vehicle arriving at a point on the road at any given instant, leading to zero time gaps. The follower headway concept that measures the time interval between two successive vehicles in a traffic lane or a single file traffic stream as they pass a reference point on the roadway is not apposite to heterogeneous traffic conditions. In heterogeneous traffic, vehicles move based on the entire road width and could follow more than one vehicle. Hence, under such conditions, a better approach involves considering the time gap which is the time interval between consecutive vehicles passing a reference line on the entire road width, rather than the lane-based approach. Many researchers[1,2] have used the entire road space-based arrivals at a reference line on the road for the modeling of time gap under heterogeneous conditions. Vehicular time gap incorporates both following and non-following interactions that are typical of a heterogeneous traffic scenario in developing countries such as India. In spite of this advantage, there are a few challenges in time gap

Received date: Jan 12, 2012; Revised date: Feb 27, 2012; Accepted date: May 8, 2012
*Corresponding author. E-mail: sarkatkar@gmail.com

DOI: 10.1016/S1570-6672(11)60233-X



(a) Flow level: 1020 vph (b) Flow level: 4100 vph

Fig. 1 Rank frequency plot of time gap on log-log scale

modeling. Due to the presence of both fast and slow moving, small and large vehicles, time gaps may range from 0 to 25 s, with a significant amount (0–20%) of data in the tail regions, as shown in Fig. 1. Hence, the modeling of both zero time gaps and in the tail regions assumes paramount importance and leads to erroneous results when neglected. In a nut shell, it becomes imperative to perform statistical modeling of the entire road width-based time gaps under heterogeneous traffic conditions coupled with better tail modeling.

Vehicular time gaps under heterogeneous traffic conditions are significantly different from those of time headways under homogeneous traffic conditions, as they both measure two different quantities[2]. Nevertheless, many researchers of time gap modeling have adopted distributions such as exponential, gamma, erlang, and lognormal, which have been used by researchers of headway modeling. The researchers of heterogeneous traffic conditions have also referred to the entire road width-based inter-arrival rate as "headway," though it is not the same as the lane-based follower headway that prevails under homogeneous conditions. Under homogeneous conditions, Al-Ghamdi[3] studied exponential, shifted exponential, and erlang distribution for headways and established boundaries such as low traffic (less than 400 vph), medium traffic (400 to 1,200 vph), and high traffic (more than 1,200 vph). Similarly, in time gap research in heterogeneous traffic, Kumar and Rao[4] studied negative exponential distribution for flow ranges varying from about 100 vph to 200 vph. Chandra and Kumar[5] analyzed the headways on urban roads in India and suggested hyperlang distribution for a flow range of 900–1,600 vph. Arasan and Koshy[6] have proposed negative exponential for all flow ranges while considering the sampling approach. The reason for the[3–6] light-tailed distributions just cited to have modeled time gap data is the application of the Chi-square goodness-of-fit test, which is not only a weak powered test (Steele *et al.*[7]) but also one that ascribes no specific importance to tail data. From the literature just cited, it is also obvious that different authors on time gap modeling have used different distributions for the same flow range and vice versa, which necessitates streamlining of the existing literature.

Some authors in time gap research have also used heavy-tailed distributions. Ramanayya[8] proposed exponential distribution for flows up to 500 vph, shifted exponential distribution for 500–650 vph, and lognormal for higher flow levels. Yin *et al.*[9] reported that lognormal distribution gives the best fit regardless of traffic conditions with a maximum flow level of 617 vph. Even distributions such as lognormal that gave a better fit than their light-tailed counterparts could only do so because of two reasons: (a) merging of the data in the tails into a single bin and (b) using the goodness-of-fit test such as the Chi-square test in most of the situations. The merging of data in the tail, that is, combining two or more bins into a single bin, could result in the loss of a significant amount of information and can also lead to a non-robust modeling. One more interesting fact is that the researchers of both headway and time gap modeling have neglected the importance of bin size compared with distribution performance. Dey and Chandra[10] have used arbitrary bin sizes in their work. Arasan and Koshy[5] used Sturges' rule that uses range of the data for bin size calculation and which is applicable only to a data set with a maximum of 100 observations[11]. Sahoo *et al.*[12] chose an arbitrary bin size of three seconds for modeling time gap data and proposed negative exponential for a maximum flow of 850 vph. A detailed discussion about the consequences of using non-optimal bin size has been presented in Section 5.

Some authors used vehicle-specific headway models that considered the effect of traffic composition. Hoogendoom and Bavy[13] used Branston's General Queuing model (GQM) for headways that were aggregated according to vehicle type and