Effect of Rain on Probability Distributions Fitted to Vehicle Time Headways

Hashim Mohammed Alhassan*, 1, Johnnie Ben-Edigbe2.

* Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Malaysia
Department of Civil Engineering, Bayero University. 700243, Kano, Nigeria.
E-mail: 1 mahashim2@live.utm.my; 2 edigbea@utm.my

Abstract—Time headway data generated from different rain conditions were fitted to probability distributions to see which ones best described the trends in headway behaviour in wet weather. Data was generated from the J5, a principal road in Johor Bahru for two months and the headways in no-rain condition were analysed and compared to the rain generated headway data. The results showed a decrease in headways between no-rain and the rain conditions. Further decreases were observed with increase in rainfall intensity. Thus between no-rain to light rain condition there was 15.66% reduction in the mean headways. Also the mean headway reduction between no-rain and medium rain condition is 19.97% while the reduction between no-rain and heavy rain condition is 25.65%. This trend is already acknowledged in the literature. The Burr probability distribution ranked first amongst five others in describing the trends in headway behaviour during rainfall. It passed the goodness of fit tests for the K-S, A2 and C-S at 95% and 99 % respectively. The scale parameter of the Burr model and the P-value increased as the rain intensity increased. This suggests more vehicular cluster during rainfall with the probability of this occurring increasing with more rain intensity. The coefficient of variation and Skewness also pointed towards increase in vehicle cluster. The Burr Probability Distribution therefore can be applied to model headways in rain and no-rain weather conditions among others.

Keywords—No-rain Condition; Rain Condition; Probability Density Function; Cumulative Distribution Function,
collection is presented in section III, while the probability distributions fitted to the data are presented in section IV. The results comes in section V and the conclusions in section VI.

II. LITERATURE REVIEW

Much research has been devoted to the concept that traffic stream behaviour can be analysed at the microscopic level. At this level the behaviour of individual vehicles drivers must be examined and modelled. Microscopic models use car following laws to describe the behaviour of each driver-vehicle system in the traffic stream as well as their interaction. Microscopic parameters of headway and speed have been moulded by numerous authors, all of which have been based on probability theory. Adams,[2] used the Poisson process to model headways for free flowing traffic. Miller [3] used a queuing model to explain the behaviour of slow moving vehicles in the traffic stream. The shifted exponential model and the modified semi-Poisson process were both employed by Ashton [4] and Schuhl [5] to model headways of vehicles. Modifications of the Poisson model resulted in the use of the Erlang [6] and the gamma distributions [7] in the modelling of headways. Another popular model which has stood the test of time is the Log-normal model proposed by Greenberg [8]. These models are called single models because they employ a single model at a time.

Combined probability distributions have also been used to model headways of traffic with high flow rates. Traffic is then divided into two. One group relating to traffic moving in a free flow fashion and another group in a constrained flow. A threshold is then established between the two conditions resulting in a distribution which is a linear combination of two components. This is stated as:

\[ f(t) = \theta g(t) + (1-\theta)k(t) \]  

\[ F(t) = \theta G(t) + (1-\theta)K(t) \]  

Where f, g, k and F, G, K are the PDF and CDF of the variables H, U, and V respectively. The probability that a vehicle is a follower is given by the parameter \( \theta \). The function k representing the free flow regime is widely taken to be the exponential distribution but a variety of distributions could be used for the constrained portion. These include, the Cowan M3, DDNED and the Hyperlang distributions HA et al. [9].

Also used to describe headways of vehicles are the so-called mixed models. These emanated from [10] and [11]. Buckley proposed the Semi-Poisson Model which in its essence identifies a headway which is greater than a threshold value U and follows the exponential distribution. Branston [11] and [12] working separately used a queuing model to explain the headways of following vehicles. The model proposed by them also has a component of following headways and an exponential component. The pdf of the two models are:

\[ f(t) = \theta g(t) + (1-\theta)\lambda e^{-\lambda t} \int_0^t (u)e^{\lambda u} du \]  

Since both the combined and the mixed models contain portions of free flow and constrained flow, it is essential to distinguish between the two. This is done by setting the parameter \( \theta \) to zero. The combined case degenerates to an exponential distribution but the mixed model does not.

A. Goodness of Fit and Parameter Estimation of Distributions

Goodness of fit is the method used to verify and ascertain the appropriateness of a probability distribution to modeling a particular phenomenon. The methods employed include Kolmogorov-Smirnov test (K-S test), Anderson Darling Test (\( \chi^2 \)) and the Chi-Square test (C-S). Parameter estimation methods commonly employed include moment estimator (ME), maximum likelihood Estimator (MLE) and the Bayesian method (BM). Variants of the maximum likelihood methods such as local maximum likelihood estimators (LMLE) and the modified maximum likelihood method (MMLM) have been used by Cohen and Whitten [12].

Parameter estimation for the combined and mixed models calls on a combination of the common estimation methods or variants of them to achieve the desired results. Branston [11] combined the ME and the method of maximizing the chi-squared statistics to estimate the parameters of the general queuing model (GQM). Hoogendoorn and Botma, [13] have used the mean integrated squared error (MISE) distance in the frequency domain to estimate parameters of the GQM.

III. DATA COLLECTION

Data for this study was collected on the J5 Highway in the Southern Malaysian State of Johor Bahru. A basic section devoid of the influence of intersections and other disturbances was selected and has an average traffic volume of 12000 vehicles per day with 79% of the traffic being cars. The section is a two-way two-lane facility and has a posted limit of 60km/hr. The pavement has uniform section and is well marked.

Data was collected for two months starting from November 2010 to December 2010 and was filtered to remove overtaking maneuvers. The data was collected using pneumatic tube detectors which recorded each individual vehicle detail such as arrival times, instantaneous speed, headway, gap, wheel base etc. Also the traffic data were separated into daylight data and night data. The periods of the daylight traffic data that coincided with rainfall events were identified and filtered. Other traffic data which did not relate with rainfall were also filtered. The rainfall traffic data were further classified according to rainfall intensity. The rainfall event times were used to select the corresponding dry weather traffic data and the headways for both rain and non-rain periods were extracted for analysis. For instance, if rain event occurred during the morning peak hour, the corresponding morning peak no-rain data was used for analysis. Rainfall data was obtained from a nearby rain...
gauge station located 750m away from the data collection site.

IV. METHOD

Computer software was used to explore available probability distributions and candidate distributions were identified for goodness of fit tests. Six distributions that fitted the frequency distribution were ranked in order of best fit. For the headways data, the following distributions provided the best fit.

A. Burr Distribution

The Burr Type XII Distribution is a continuous probability distribution for a non-negative random variable sometimes also called the generalized log-logistic distribution. It is commonly used to model household income. It has continuous shape parameters \( k > 0, \alpha > 0 \) continuous scale parameter \( \beta > 0 \) and continuous location parameter \( \gamma > 0 \) with a pdf and CDF given by:

Probability Density Function (PDF):

\[
f(x) = \frac{\alpha k}{\beta} \left( \frac{x}{\beta} \right)^{a-1} \left[ 1 + \left( \frac{x}{\beta} \right)^{a} \right]^{-k-1}
\]

Cumulative Distribution Function (CDF):

\[
F(x) = 1 - \left[ 1 + \left( \frac{x}{\beta} \right)^{a} \right]^{-k}
\]

B. Frechet or Maximum Extreme Value Type II Distribution

This distribution is used in hydrology to model annual maximum one-day rainfalls and river discharges. It has continuous shape parameter \( \alpha > 0 \), continuous scale parameter \( \beta > 0 \) and continuous location parameter \( \gamma > 0 \) but reduces to the 2-parameter function when \( \gamma = 0 \). It has PDF and CDF as follows:

Probability Density Function:

\[
f(x) = \frac{\alpha}{\sigma} \left( \frac{x-\mu}{\sigma} \right)^{\alpha-1} \exp \left( -\frac{x-\mu}{\sigma} \right)
\]

Cumulative Distribution Function:

\[
F(x) = 1 - \exp \left( -\frac{x-\mu}{\sigma} \right)
\]

C. Generalized Extreme Value Distribution.

This is a family of continuous probability distributions developed from extreme value theory. Thus, it is used as a limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. It is therefore used as an approximation to model the maxima of long (finite) sequences of random variables. The distribution has continuous shape parameter \( k > 0 \), continuous scale parameter \( \alpha > 0 \). The pdf and CDF of this distribution is stated as follows:

\[
f(x) = \begin{cases} 
\frac{1}{\alpha} \exp\left( -\left(1+kz\right)^{-1/k} \right) \left(1+kz\right)^{-1/k} & k \neq 0 \\
\frac{1}{\alpha} \exp \left( -z - \exp \left( -z \right) \right) & k = 0
\end{cases}
\]

Cumulative Distribution Function:

\[
F(x) = \begin{cases} 
1 - \left(1 + k \left( \frac{x-\mu}{\sigma} \right)^{-1/k} \right) & k \neq 0 \\
1 - \exp \left( -\frac{x-\mu}{\sigma} \right) & k = 0
\end{cases}
\]

D. Generalized Pareto Distribution

The Generalized Pareto distribution has continuous shape parameter \( k > 0 \), continuous scale parameter \( \sigma > 0 \) with probability and cumulative distribution functions as follows:

\[
f(x) = \begin{cases} 
\frac{1}{\sigma} \left( 1 + k \left( \frac{x-\mu}{\sigma} \right)^{-1/k} \right) & k \neq 0 \\
\frac{1}{\sigma} \exp \left( -\frac{x-\mu}{\sigma} \right) & k = 0
\end{cases}
\]

Cumulative Distribution function:

\[
F(x) = \begin{cases} 
1 - \left(1 + k \left( \frac{x-\mu}{\sigma} \right)^{-1/k} \right) & k \neq 0 \\
1 - \exp \left( -\frac{x-\mu}{\sigma} \right) & k = 0
\end{cases}
\]

E. Log-Normal Distribution

The Lognormal three parameter distribution has continuous shape parameter \( \sigma > 0 \), continuous scale parameter \( \mu > 0 \) and continuous location parameter \( \gamma > 0 \). This reduces to the two parameter distribution when \( \gamma = 0 \).
The probability density and cumulative distribution functions are given by:

**Probability Density Function:**

\[
 f(x) = \frac{\exp \left( -\frac{1}{2} \left( \frac{\ln(x - \gamma) - \mu}{\sigma} \right)^2 \right)}{(x - \gamma) \sigma \sqrt{2\pi}}
\]  
\[\text{(13)}\]

**Cumulative Distribution Function**

\[
 F(x) = \phi \left( \ln \left( \frac{x - \gamma}{\mu} \right) \right)
\]  
\[\text{(14)}\]

**F. Pearson Type 6 Distribution**

Pearson type 6 distribution is a 4-parameter distribution with a continuous shape parameter (\(\alpha_1 > 0\)), continuous shape parameter (\(\alpha_2 > 0\)), continuous scale parameter (\(\beta > 0\)) and continuous location parameter (\(\gamma > 0\)). This transforms to the 3-parameter distribution when \(\gamma = 0\). The two distribution functions are stated as:

**Probability Density Function:**

\[
 f(x) = \frac{\left( (x - \gamma) / \beta \right)^{\alpha_1 - 1}}{\beta B(\alpha_1, \alpha_2) \left( 1 + \left( x - \gamma / \beta \right) \right)^{\alpha_1 + \alpha_2}}
\]  
\[\text{(15)}\]

**Cumulative Distribution Function:**

\[
 F(x) = I \left( x - \gamma / \beta \right)^{\alpha_1} / \left( x - \gamma + \beta \right)^{\alpha_2}
\]  
\[\text{(16)}\]

**V. RESULTS**

Six probability distributions presented as above were used to fit the headway data. The fit of each distribution can be seen from the probability density and p-p plots for each weather condition shown from fig. 1 to 4. Visually, all the distributions provided a good fit. However, the Burr distribution provided the best fit for all weather conditions. It passed both the 95% and 99% goodness of fit test for Kolmogorov-Smirnov (K-S), but failed the Anderson Darling (AD) and Chi-squared (C-S) tests. All the other distributions failed the goodness of fit tests at 95% and 99% for all the three tests. The Log-normal model, The Gamma model, The Rayleigh model, displaced negative exponential model all performed poorly given the direction pointed by the literature about their appropriateness in modelling headway data [14].

From table 1, the Burr Distribution is the overall best fit for headway data both under rain and non-rain conditions. The Frechet Model follows closely as the second best model to describe headway data. The Pearson 6 surpassed the rest under no-rain, light rain and medium rain conditions as the third best model. However, the Generalised Pareto (GP) was a better fit under heavy rain using the K-S test criteria. There is no consistent performance by a particular model under heavy rain conditions, for Generalised Extreme Value was rated third best using the K-S test while at the same time the GEV was second best using Anderson Darling and third best again under the Chi-Squared test.

The Generalised Pareto (GP) model has a better fit to the headway data than the Lognormal and Generalised Extreme Value (GEV) models under no-rain, light rain and heavy rain conditions. The lognormal model was the least good fit under all rain conditions for all the goodness of fit tests carried out. Ironically, it has been the model most applied to time headway distribution of vehicles. Whereas HA et al. (2010) has confirmed that the LNM model provided the best fit among single distribution models; nevertheless it did not satisfy the goodness of fit for some traffic in the slow lane on R118.
A. Headway Characteristics

The headway characteristics for both rain and non-rain conditions are summarised in table 2. The mean headway decreases between no-rain and rain conditions and decreases further with increase in rainfall intensity. The variance and the standard deviation also behave in a similar fashion. The trend in the coefficient of variation suggests that there is a cluster of headways on the facility and this increase with rain and rain intensity. Even though the coefficient of variation under heavy rain is the lowest of the three categories it is higher than the no rain condition. The skewness and kurtosis of the time headway variable are all positive indicating that the bulk of the headway distribution is to the left of the mean value which presupposes smaller headways or localised clusters on the facility. Thus up to 50% of the vehicles were travelling with time headways less than 2.20 secs under no-rain condition, 2.00 secs under light and medium rain, and again 2.20 secs under heavy rain conditions.

B. Probability Distribution Parameters

The distributions fitted to the time headways have varied parameters. Of these, only the Frechet and the Lognormal have two parameters. The rest have three parameters each. These parameters are the shape, scale and the location parameters. For the Burr model, the shape parameter (k and α) and the scale parameter β increases with rain and with increase in rain intensity. The P-Value increases duly with rain and rain intensity. Thus if two rain conditions with the same headways are considered, there will be contractions in the headways skewed further to the
left of the mean. Further increase in the scale parameter will lead to cluster of vehicles with smaller headways. This aptly describes the situation under rainfall conditions on the facility. The parameters of the fitted distributions are shown in table 3. With the Frechet model, the shape parameter increases with rainfall intensity while the scale parameter decreases. The heavy rain model parameters show an increase in the two parameters well above the other rain and non-rain conditions.

<table>
<thead>
<tr>
<th>Weather Condition</th>
<th>Kolmogorov Smirnov</th>
<th>Anderson Darling</th>
<th>Chi Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1ST</td>
<td>2ND</td>
<td>3RD</td>
</tr>
<tr>
<td>No-Rain</td>
<td>Burr</td>
<td>Frechet</td>
<td>PT6</td>
</tr>
<tr>
<td>Light Rain</td>
<td>Burr</td>
<td>Frechet</td>
<td>PT6</td>
</tr>
<tr>
<td>Medium Rain</td>
<td>Burr</td>
<td>Frechet</td>
<td>GP</td>
</tr>
<tr>
<td>Heavy rain</td>
<td>Burr</td>
<td>Frechet</td>
<td>GEV</td>
</tr>
<tr>
<td></td>
<td>4TH</td>
<td>5TH</td>
<td>6TH</td>
</tr>
<tr>
<td>No-Rain</td>
<td>GP</td>
<td>LNM</td>
<td>GEV</td>
</tr>
<tr>
<td>Light Rain</td>
<td>GP</td>
<td>GEV</td>
<td>LNM</td>
</tr>
<tr>
<td>Medium Rain</td>
<td>PT6</td>
<td>GEV</td>
<td>LNM</td>
</tr>
<tr>
<td>Heavy rain</td>
<td>GP</td>
<td>PT6</td>
<td>LNM</td>
</tr>
</tbody>
</table>

The Generalised Extreme Value and the Generalised Pareto have decreases in the shape parameter for increase rain intensity and decrease in the scale parameter with decreasing rain intensity. There are no consistent trends observed in the Pearson 6 model parameters. The Lognormal parameters of shape and scale both have a consistent trend. Decrease in shape parameters values with rainfall and further decreases with higher rain intensities.

The Generalised Extreme Value and the Generalised Pareto have decreases in the shape parameter for increase rain intensity and decrease in the scale parameter with decreasing rain intensity. There are no consistent trends observed in the Pearson 6 model parameters. The Lognormal parameters of shape and scale both have a consistent trend. Decrease in shape parameters values with rainfall and further decreases with higher rain intensities.

VI. CONCLUSIONS

This study explored continuous probability distributions that best describes vehicle time headways under rainfall conditions. Six models were initially selected from numerous others and were further tested for goodness of fit. The Burr model, the Frechet model and the Pearson type 6 were rated as the best models to apply to headway data under rainfall conditions. The Burr model was rated first by all the goodness of fit tests and for all conditions; rain and non-rain. The Frechet model performed second under all tests except the heavy rain condition where the Generalised Extreme Value (GEV) model performed better under the Anderson Darling (A2) test. The Pearson type 6 model was consistently the third best model under the A2 and C-S tests for no-rain, light rain and medium rain respectively. Under the K-S test it out performed the others in no-rain and light rain conditions. There was a mixture of performance for the remaining three models under the prevailing conditions.
However, the highly rated lognormal model performed poorly among the remaining models.

The Headway characteristics showed decreased headways values with increase in rain intensity. Thus between no-rain to light rain condition there was 15.66% reduction in the mean headways. Also the headway reduction between no-rain and medium rain condition is 19.97% while the reduction between no-rain and heavy rain condition is 25.65%. About 50% of the headways were less than the mean values, an indication of the Skewness of the headways to the left of the mean. This trend increased with increase in rain intensity. The Skewness and coefficient of variation showed that clusters of vehicles headways increased as the rain intensified.

The model scale parameters could be related to the rain and no-rain conditions as well as the various rainfall intensity regimes explored. Threshold values would need to be established for each rain condition and this will require more time headway rain-conditioned analysis for confidence to be built into them.

The Burr model lends itself to model headway data under rainfall conditions.

ACKNOWLEDGMENT

The authors wish to thank the Department of Irrigation and Drainage of Malaysia (JPS) for providing the rainfall data.

REFERENCES