Accident Analysis and the Suggestion of an Accident Prediction Model for Guwahati city

Rupanjan Chakraborty 1, Sarbajit Bhattacharyya 2, Mrinal Roy 3, Pinak Paul 4
U.G. Students, Department of Civil Engineering, Assam Engineering College, Guwahati, Assam, India1, 2, 3, 4

ABSTRACT: The ultimate purpose of scientifically designing a road is to avoid the occurrence of accidents on it. Even after the construction of a road, several measures should be taken to make the road sufficiently safe from accidents. The possibility of accidents occurring on a road depends upon numerous factors. In this paper, our study involved the analysis of those factors affecting number of accidents for Guwahati city by collecting accidental data for various road stretches connected to Jalukbari area of Guwahati which can be considered as a representative of almost the whole city. To avoid the demerits associated with a linear regression model, a log-linear model was developed which could fit the data obtained and was validated by using the model to predict the number of accidents per stretch length for a similar test with considerable accuracy.

KEYWORDS: Designing a road, Guwahati city, Jalukbari, Accidents, Factors affecting number of accidents, Log-linear model

I. INTRODUCTION

Guwahati, one of the most picturesque and important cities of India, is a gateway to Northeast India. The city is well connected to various parts of India. So, the development of the city is not only vital for the state of Assam but for the whole of Northeast India. Guwahati is a fast growing metropolis and with the population of 9,62,334 (according to 2011 census) is the most populous city of Northeast India. The South-western part of Guwahati city is one of the busiest areas and constantly growing in terms of population and infrastructure development. Jalukbari can be considered as the heart of the road network of this section of the city and can be rightly termed as the gateway to Guwahati metropolitan because of various reasons. This is mainly because it is the intersecting point of a number of important roads which can be considered vital from the point of view of transportation needs of not only the citizens of Guwahati but also to the visitors of the city. Jalukbari is almost the ending point of NH-37 which is one of the most important National Highways of India as it provides the entrance to the various states of North-eastern India and connects the city with other states of Eastern India like Bihar, Jharkhand and West Bengal. NH-37 connects Assam with Arunachal Pradesh by running from Pancharatra (near Goalpara, Assam) to Roing (in Arunachal Pradesh). The NH-37 also provides access to Lokapriya Gopinath Bordoloi International Airport situated at Borjhar, Guwahati as well as to the Inter State Bus Terminus situated at Lokhra, Guwahati. The NH-37 runs up to Jorabat which lies on the Assam-Meghalaya border before bifurcating into NH-54 which goes through Shillong, the capital city of Meghalaya. It also connects important educational institutions like the Indian Institute of Technology (Guwahati), Gauhati University, Assam Engineering College and Govt. Ayurvedic College to the more urbanised areas of the main city through the Assam Trunk Road. Due to the ease of connectivity, Jalukbari area has witnessed a huge growth of population as people have started to migrate to this relatively unexplored area in comparison to the other developed parts of the city, for the purpose of setting up industries, schools, residential apartments, luxurious hotels, etc. As a result there has been an exponential growth in the number of vehicles plying through this area which has subsequently accelerated the number of road accidents. Considering the aforementioned points it is very easy to gauge the importance of the area which deems it sufficiently suitable to be studied from the perspective of road safety in order to ensure the overall health and safety of the people residing in the Guwahati city who in one way or another, are dependent on the road network that has Jalukbari as its central area of importance. Four stretches of roads connected to this area have been considered for the purpose of conducting accidental studies and various trends have been observed.
that illustrates the dependency of the number of accidents on various factors which have ultimately helped in forming an accident model for the considered roads which can be considered as a representative of the whole city of Guwahati.

II. LITERATURE REVIEW

Accident model analysis is one of the most widely used methodologies adopted worldwide to study the various causes contributing to the number of accidents. Charles V. Zeeger et al. (1980) took the initiative of studying the effect of lane and shoulder widths on accidents which probably paved the path of modern day accident model analysis. M G. Karlaftis et al. (2001) studied the relationship between rural road geometric characteristics, accident rates and their prediction, using a rigorous non parametric statistical methodology known as ‘hierarchical tree-based regression’. P. Pramada Valli (2005) developed various models at an all India level as well as for major metropolitan cities by analysing the road accident data in order to understand the nature and extent of various causes of accidents using the concept of the widely accepted Smeed’s law and Andreasson’s equations. P. Bhat et al. (2013) developed an accident model for Bangalore city including the factors like road characteristics (viz. carriageway width, shoulder type, number of minor crossings, road condition, etc.). A. Kusumawati et al. (2013) developed an accident prediction model that showed a positive correlation between accidents at rural toll road sections with annual average daily traffic (AADT), number of horizontal curvatures and a negative correlation with lane width. P. Chengeye et al. (2013) investigated motorway safety by developing accident prediction models linking accident frequencies and their contributing factors which included traffic conditions, geometrical and operational characteristics of road and weather conditions. S. Rokade et al. (2013) suggested through an accident prediction model that the when the dependent variable was the number of accidents, the independent variables may be quantitative variables viz. cross section dimensions, horizontal curvature, traffic volume, speed and qualitative variables such as type of terrain, condition of road shoulders and medians etc. S.U. Bobade et al. (2015) did accident black spot analysis on Pune-Solapur National Highway with the help of various methods like severity index and ranking method.

III. DATA COLLECTION

We selected four stretches along the two of the busiest roads of south western Guwahati viz. A.T. Road and NH-37 (Jalukbari-Khanapara bypass) for our study. The stretches selected for our study are mentioned in Table 1-

<table>
<thead>
<tr>
<th>STRETCH</th>
<th>DESCRIPTION</th>
<th>LENGTH (KM)</th>
<th>POLICE STATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bharalumukh to Maligaon(A.T. Road)</td>
<td>3.6</td>
<td>Maligaon</td>
</tr>
<tr>
<td>2</td>
<td>Maligaon to Jalukbari(A.T. Road)</td>
<td>2.6</td>
<td>Maligaon</td>
</tr>
<tr>
<td>3</td>
<td>Jalukbari to Radisson Blu Hotel(NH37)</td>
<td>2.1</td>
<td>Jalukbari</td>
</tr>
<tr>
<td>4</td>
<td>Radisson Blu Hotel to Boragaon(NH 37)</td>
<td>2.9</td>
<td>Jalukbari</td>
</tr>
</tbody>
</table>

For analysing the road accidents of these stretches, the accidents records for a period of 5 years (2010-2014) were collected from the respective police stations viz. Jalukbari and Maligaon. Accident data (Number of accidents, persons killed/injured in the accidents) were collected from police station records of the period ranging from 2010 to 2014. The type of vehicles involved in accidents as recorded in the FIR was also noted down. The categories of vehicles involved varied from multi-axle Trucks to two-wheelers. The overall accidents of Guwahati city (detailing the number of fatal and non-fatal accidents) and the fatalities and injuries of the accidents in Guwahati from 2010 to 2014 is shown in the Fig. 1(a) and Fig. 2(a).

P.W.D. (Public Works Department) and NHAI records are the main source of traffic volume data & cross sectional elements data. In addition to the above, information about the road conditions, year of maintenance works etc. were also extracted from PWD and NHAI records of Guwahati. Our stretches 1 & 2 falls under PWD jurisdiction while 3 & 4 falls under NHAI. Traffic Volume data were collected from the NH sub division office of PWD at Panbazar, Guwahati. It is assumed that the traffic volume is uniform throughout the stretch.
As the yearly traffic census data was not available for all the years, the available data (2014 data for Stretches 3 and 4 and 2012 data for Stretches 1 and 2) were used to predict traffic volume on a road in each year for the period 2010-2014 as shown in Table-3. A growth rate (r) of 7.5% per year was assumed for this purpose (with due advice of experienced officials of PWD). A formula was used for forecasting as follows:

\[ A = P \left( 1 + \frac{r}{100} \right)^n \]

where, \( A \) = Predicted year traffic volume, \( P \) = Present year traffic volume, \( n \) = Years of count and minus sign indicates before the present year traffic count and plus sign indicates after the present year traffic count. The AADT data and annual number of accidents for the four stretches during 2010-2014 are shown in Table 2(a) and Table 2(b) respectively.

**Table 2(a)** Year-wise AADT of Stretches

<table>
<thead>
<tr>
<th>YEAR</th>
<th>STRETCH 1</th>
<th>STRETCH 2</th>
<th>STRETCH 3</th>
<th>STRETCH 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>74653</td>
<td>75112</td>
<td>73441</td>
<td>73441</td>
</tr>
<tr>
<td>2013</td>
<td>69445</td>
<td>69875</td>
<td>68798</td>
<td>68798</td>
</tr>
<tr>
<td>2012</td>
<td>64600</td>
<td>65000</td>
<td>65193</td>
<td>65193</td>
</tr>
<tr>
<td>2011</td>
<td>59755</td>
<td>60125</td>
<td>59909</td>
<td>59909</td>
</tr>
<tr>
<td>2010</td>
<td>55273</td>
<td>55615</td>
<td>54925</td>
<td>54925</td>
</tr>
</tbody>
</table>

**Table 2(b)** Year-wise number of accidents in stretches

<table>
<thead>
<tr>
<th>YEAR</th>
<th>STRETCH 1</th>
<th>STRETCH 2</th>
<th>STRETCH 3</th>
<th>STRETCH 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>41</td>
<td>40</td>
<td>29</td>
<td>47</td>
</tr>
<tr>
<td>2013</td>
<td>39</td>
<td>41</td>
<td>27</td>
<td>39</td>
</tr>
<tr>
<td>2012</td>
<td>30</td>
<td>37</td>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td>2011</td>
<td>29</td>
<td>40</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>2010</td>
<td>27</td>
<td>30</td>
<td>13</td>
<td>18</td>
</tr>
</tbody>
</table>

As the accidental data was collected manually from various Police Stations of Guwahati, there exists a chance of exclusion of many accidents as all accidents are not reported to the police station. Hence, in this paper, the accidental data were increased suitably wherever there was a scope of accidental data obtained being lower than the actual value. For example, before 2012, Jalukbari area was under the jurisdiction of Maligaon police station. But, in 2012, a new police station was set up at Jalukbari and thus there was a transfer of jurisdiction of traffic control and thus any data not reported was assumed by averaging accordingly.

Accidental analysis have shown that accident rate has a direct relation with the traffic volume. Accident rate i.e. number of accidents per km per year of the four different stretches for the 5 years (2010-2014) are found out. A trend line between the accident rate and traffic volume (AADT) is plotted for the different stretches using MS EXCEL. The following Fig. 2(a), Fig 2(b), Fig 2(c) and Fig 2(d) show the variation of accidents per kilometer per year with respect to AADT in Stretches 1,2,3 and 4 respectively.
IV. DETERMINATION OF AN ACCIDENT PREDICTION MODEL

Wide range of research work has been done considering linear models which assume the variation of the dependent variable (which is an indicator of the number of accidents) with respect to other independent variables such as AADT, road/shoulder conditions, etc. as linear. On the other hand, ample amount of literature is also dedicated to non-linear accident models with more popular ones based on Smeed’s law and Andreassen’s equation like that of Road Accident Models For Large Metropolitan Cities Of India by P. Pramada Valli (2003).

If number of accidents is dependent on variables $x_i$, then a general form of such linear models could be expressed as-

$$F(N) = \sum a_i x_i + b$$

where, $F(N)$ is a function of the number of accidents and $a_i$ and $b$ are the coefficients and constant respectively to be determined through suitable linear data analysis.

But such a linear model has one intrinsic flaw which renders such a model somewhat inefficient. An often occurring important independent variable in accident models is AADT. Other parameters include various other road elements and traffic behavior. Under any road condition however, if AADT is zero, we statistically don’t expect any accidents on the road. Therefore, practically, number of accidents should be zero when AADT is zero. But in a linear model for any road condition, the other terms will always contribute a finite value through $\sum a_i x_i + b$ where $x_i$ do not include AADT. In a linear model, it is impossible to make $\sum a_i x_i + b = 0 \forall x_i$ when AADT=0.

Hence to avoid such an anomaly in our study, we have, instead of a linear one, adopted a log-linear model of the form shown below-

$$F(N') = k \prod x_i$$

i.e. $\log\{F(N')\} = \sum a_i \log(x_i) + \log(k)$
Selection of variables -

The next step is to find the suitable dependent variable which is a function of the number of accidents and also the independent variables to which the increase or decrease of the number of accidents in a stretch can be properly attributed. The following variables were ultimately selected considering adequate reasoning as described below:

i) **Number of accidents per unit stretch length (λ)**

The stretches that have been selected for our study are sufficiently short (at most 3km). Under such circumstances, the chances of occurrence of multiple black-spots is pretty low and even if they do, the number of accidents in stretch/length, i.e. an averaging cannot be too difficult to consider, since the road parameters for a short stretch do not vary adequately over the length of the stretch to render such averaging useless. So, it is assumed that the total number of accidents in a stretch is directly proportional to the length of the stretch. Hence, \[ \frac{\text{Number of accidents}}{\text{stretch length}} \] is considered by us as the dependent variable which is a function of the number of accidents and is represented by \( \lambda \).

ii) **Annual Average Daily Traffic (AADT)**

One of the most comprehensive way of expressing the traffic volume is Annual Average Daily Traffic (AADT). It is obtained by determining the average daily traffic volume recorded for all the 365 days of the year. Now, to deal with the first mentioned issue as to why we selected a non-linear model, we presume \( \lambda = \frac{\text{Number of accidents}}{\text{stretch length}} \) (which represents the linear accident density) to be directly proportional to AADT as it is evident from our available data that the occurrence of accidents will be more frequent if AADT is more for a given stretch. i.e. \( \lambda \alpha AADT \)

(iii) **Shoulder Condition Rank (SC)**

Rakesh Kumar Singh & S. K. Suman (2002) (Accident Analysis and Prediction of Model on National Highways), based on recommendations of Chandra and Dev Raj (1999), suggested that the shoulders can be divided into different categories based on their degree/ease of utility by the vehicles which can be defined by the shoulders’ physical condition (characterized by the extent of damage after repairing work has been done) at time of data collection. Thus, shoulder may be ranked on a 5-point ranking scale which indicate the Shoulder Condition Rank was developed as shown below:

- Excellent = 5,
- Good = 4,
- Average = 3,
- Poor = 2 and
- Bad = 1

The two extremes of the rating system imply:

**Excellent** - It means that road and shoulders was maintained very recently

**Bad** - It means maintenance works were not carried within the last few years and hence these have deteriorated to the extent that vehicles cannot use them.

In our study, fractional rating for a stretch in a particular year indicates that maintenance work was carried out in the stretch in an earlier year/years, but the road has deteriorated to such a state that a whole number rating was not possible to be allocated to it when compared with other stretches. Hence, repair records were studied and an approximate decimal rating was assigned to it.

A general look at the shoulder conditions for a stretch and \( \lambda \) at once made it apparent that shoulder condition affects \( \lambda \) such that improvement of shoulder condition decreases \( \lambda \). So, conforming to a log linear model, we further consider:

\[ \lambda \alpha \frac{\text{Shoulder Condition Rank (SC)}}{\text{Road width}} \]

(iv) **Effective Road Width per Number of lanes (ERWPL)**

The width of the lane in a road affects the number of accidents because a car is given more space to maneuver if the lane width is more whereas the car’s movements get restricted if the lane width available is less. Deadly accidents in the form of rear end collisions are often seen to occur when vehicles at high speeds attempt to overtake other vehicles in roads with narrow lane width. The increase of road width available per lane is hence found to decrease the chance of accidents occurring on the roads. Thus through proper reasoning, we identify \( \text{Road width} \) as another variable and also infer that it varies inversely with \( \lambda \). As per IRC, the road width per lane for a multilane road is 3.5m. But, when we are dealing with contrasting kinds of road stretches, such as a local road and a part of a highway, the latter may have a
lesser effective road width and this narrowing of the effective road width can be attributed to some of the factors mentioned below-

a) The disturbing movement of pedestrians and animals on the edge of the roads
b) The illegal parking of vehicles on the side of the roads
c) Road side vendors who illegally occupy the driving space of vehicles on the side of the roads, etc.

Hence we identify the decrement to be provided to the road width due to the above causes and find out the Effective Road Width for each of our concerned stretches.

Thus we identify \( \frac{\text{Effective Road width}}{\text{Number of lanes}} \) (ERWPL) as a variable and conclude that-

\[ \lambda \propto \frac{1}{\text{Number of lanes \ (ERWPL)}} \]

v) Maximum Permissible Speed (v)-

Based on various factors such as the type of road, locality (residential, commercial, non-residential, non-commercial, etc.), geometric conditions of the road, etc. the maximum permissible speed is enforced by the traffic authority on the vehicles moving on the road. Now, it is generally found that if the maximum permissible speed on a road is higher, then the number of accidents is higher, which was tangible from our study. This may be due to various reasons-

a) All the drivers may not be experienced in driving safely at such high speeds and so they may crash their vehicles into others.

b) Many drivers are found to be driving illegally under the influence of alcohol which makes them susceptible to accidents and if the speed limit is higher, then they are often found to drive carelessly at high speeds while they are drunk.

Hence, another possible parameter is the maximum permissible speed of traffic ‘v’ in the stretches. So, our assumption from generalized observation was that-

\[ \lambda \propto v \]

Hence, by combining all the above mentioned proportionalities we should find-

\[ \lambda \propto \frac{\text{AADT x SC x ERWPL}}{v} \]

But the fluctuations in the value of constant was found to scale with the value of v

Suitable observations of the fluctuation behavior then forced us to assume that-

\[ \lambda \propto \sqrt{v} \]

This assumption helped us to minimize the fluctuations in the value of the constant

Hence, from the above assumptions, the final model becomes-

\[ \lambda \propto \frac{\text{AADT x } \sqrt{v}}{\text{SC x ERWPL}} \]

or,

\[ \lambda = \frac{\text{AADT x } \sqrt{v}}{\text{SC x ERWPL}} \times k \]

where k is a constant

Tabulation and calculation of the value of constant (k)-

The data collected for each of the dependent and independent variables of our proposed model for the three stretches (Stretch 1, Stretch 2 and Stretch 3) are represented in the following tables viz. Table 3(a), Table 3(b) & Table 3(c) along with properly calculated values of the constant k for each stretch in each year-

Table 3(a) Calculation of constant k for Stretch 1
Now, for the purpose of our study, we considered Stretch 4 as our test stretch and examined the validity of our model by utilizing the value of constant k obtained after averaging the values from the previous three stretches as found above (as when the data collected from the respective stretches for the different individual variables are put in the model equation, we get the value of the constant to be the approximately same for all the stretches with only 20% variation at most). The averaging provided us with the value of k as 0.000315. Now, the data originally collected for Stretch 4 and the expected value of \( \lambda \) along with the error (difference in value of original and predicted) for this stretch as derived by using the model are shown in Table 3(d) below –

<table>
<thead>
<tr>
<th>Year</th>
<th>( \lambda ) (original)</th>
<th>AADT</th>
<th>SC</th>
<th>ERWPL (m)</th>
<th>v(kmph)</th>
<th>( \lambda ) (predicted)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>16.2069</td>
<td>73441</td>
<td>2.5</td>
<td>4.2</td>
<td>50</td>
<td>15.57918874</td>
<td>-0.627711265</td>
</tr>
<tr>
<td>2013</td>
<td>13.4483</td>
<td>68798</td>
<td>3.1</td>
<td>4.3875</td>
<td>60</td>
<td>13.51320343</td>
<td>0.064903226</td>
</tr>
<tr>
<td>2012</td>
<td>10.6897</td>
<td>65193</td>
<td>3.3</td>
<td>4.3875</td>
<td>60</td>
<td>10.8043154</td>
<td>0.114615403</td>
</tr>
<tr>
<td>2011</td>
<td>7.93103</td>
<td>59909</td>
<td>3.8</td>
<td>4.3875</td>
<td>60</td>
<td>8.586904085</td>
<td>0.655874085</td>
</tr>
<tr>
<td>2010</td>
<td>6.2069</td>
<td>54925</td>
<td>4.2</td>
<td>4.3875</td>
<td>60</td>
<td>7.282094992</td>
<td>1.075194992</td>
</tr>
</tbody>
</table>

**Validation of the model**

The Chi-Squared Test was performed to check the deviation of values of \( \lambda \) obtained from the model from the original values, i.e. ultimately to check the goodness of fit or validity of the model. The result obtained was found to be
excellent. The value of Chi-Square was found to be 0.99358 which was found to have level of significance of about 90 percent for the set of values with a Degree of Freedom = 4.

V. CONCLUSION

The study helped us to derive a statistical model which can be applied to find the vulnerability of a road stretch to accidents. But the model cannot be claimed to be exhaustive in the sense that it can only approximately predict the number of accidents in stretches which have similar characteristics i.e. type of locality, population, extent of urbanisation, etc. to that of our three considered stretches. Trials were made to predict the number of accidents for stretches with much different characteristics than the stretches that were used to construct the model and as expected, the trials met with failure to further strengthen the validity our aforementioned statement. Therefore, our model can suitably be used for many other similar stretches of Guwahati city. The following conclusions can be said to have been established with sufficient experimental evidence from our study-

1. The number of accidents is directly proportional to AADT. This fact has been further proved through linear regression in our study considering separate stretches along with log-linear regression. So, sufficient traffic regulatory measures should be imposed on roads with high AADT to reduce the number of accidents occurring on those roads.

2. The number of accidents is directly proportional to maximum permissible speed on the road raised to a fractional power (in our case the power being 0.5). Thus, maximum permissible speed for a stretch should be decided cautiously by the authorities as a higher value can contribute significantly to increase the number of accidents. Therefore, our model can suitably be used for many other similar stretches of Guwahati city.

3. The number of accidents is inversely proportional to Shoulder Condition Rank asserting the need of properly constructed and maintained shoulders for a road which is not susceptible to accidents.

4. The number of accidents is inversely proportional to the Effective Road Width Per Lane. So, in narrow roads traffic should be regulated in an efficient manner and also sufficient precautions should be taken so as not to allow the reduction in Effective Road Width Per Lane of roads.

5. The model can almost accurately predict the number of accidents per unit length with a high level of significance of about 90% in stretches with similar characteristics.

ACKNOWLEDGEMENT

The authors thank Bhaskarjyoti Das (Associate Professor, Department of Civil Engineering, Assam Engineering College, Guwahati), Sunip Mukherjee (U.G. Student, Department of Physics, Presidency University, Kolkata) and Souradeep Sengupta (P.G. Student, Department of Physics and Astrophysics, University of Delhi) for their valuable suggestions regarding the accident model. Also, the authors thank the Public Works Department (Guwahati), NHAI and Guwahati Traffic Police for their immense support during the preparation of this paper.

REFERENCES


International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015


