

DEVELOPMENT OF A CRASH RISK PROBABILITY MODEL FOR FREEWAYS BASED ON HAZARD PREDICTION INDEX

Md. Mahmud Hasan¹, Shamas Bajwa²

^{1,2} School of Civil, Environmental and Chemical Engineering, RMIT University, Melbourne, Australia

Received 21 June 2013; accepted 1 October 2014

Abstract: This study presents a method for the identification of hazardous situations on the freeways. The hazard identification is done using a crash risk probability model. For this study, about 18 km long section of Eastern Freeway in Melbourne (Australia) is selected as a test bed. Two categories of data i.e. traffic and accident record data are used for the analysis and modelling. In developing the crash risk probability model, Hazard Prediction Index is formulated in this study by the differences of traffic parameters with threshold values. Seven different prediction indices are examined and the best one is selected as crash risk probability model based on prediction error minimisation.

Keywords: crash risk, risk index, road hazard, probability model, Eastern freeway.

1. Introduction

Traffic safety can be impaired for several reasons which can be classified into three broad categories such as driving factors, vehicle factors and environment factors. Driving factors such as fatigue, excessive drugs or alcohol consumption, training of drivers, pressure to save time etc. cause road incidents (Hasan, 2012). Vehicle factors consisting of manufacturing or mechanical faults also result road incidents. Similarly, weather conditions such as rainfall, snowfall, fog or smog etc. have effects on road safety (Goodwin, 2002). Other road environment factor i.e. traffic parameters may influence hazardous situations on the roadway leading to accidents. After the accident, these parameters are changed dramatically as accidents usually disrupt the normal flow of traffic resulting in congestion. Due to accident, vehicles on the roadway are stuck in congestion for long durations. This

congestion results in increased travel time, vehicle emissions and fuel usage etc. (Golob et al., 2004).

Previous studies have developed several accident prediction models. Hong et al. (2005), Kalokota and Seneviratne (1994) developed regression method based crash prediction model using road geometry. Eisenberg (2004) and Shankar et al. (1995) formulated accident frequency prediction model on the basis of negative binomial regression. In order to make accident prediction model, Greibe (2003) used road geometric variables as well as traffic parameters whereas and Pham et al. (2010) considered only disaggregated traffic data as predictor variables.

This study evaluates the relationship between traffic parameters and road hazards by formulating crash risk probability model. Two important indices: traffic flow and speed are

¹ Corresponding author: mdmahmud.hasan@outlook.com

used to quantify the traffic performance while number of accident is used as a performance index for roadway safety. In this study, a model named as Hazard Prediction Index is formulated to evaluate how traffic parameters influence road hazard occurrence. The prediction index is built by the difference of traffic parameters with boundary condition.

2. Test Bed and Data

Eastern Freeway in Melbourne (Australia), one of the important urban freeways for commuting to city from eastern suburbs of Melbourne, is selected as the test bed for this study. The section for the study is approximately 18 km long, from Hoddle Street to Springvale road, consisting of three to five lanes in each direction.

For this study, traffic data and crash data are needed to develop the model. Traffic data consist of 5 minute interval traffic flow and speed data, and crash data include crash time and location. Both the dataset are from September 2007 to June 2010, and during this time 138 accidents occurred on the selected freeway. In order to formulate the index, the dataset are arranged such a way that there is equal number (138) of hazard and non-hazard cases included.

3. Hazard Prediction Index

Hazard Prediction Index is a representative model for predicting the hazardous situations on a roadway by evaluating the traffic performance indicators. The prediction index is built by the difference of traffic parameters with boundary conditions using the dataset described in previous section.

3.1. Variables Used for Developing the Hazard Prediction Index

The main objective of the model building is to predict the crash hazard based on traffic conditions. So, the independent variables should have the capability to represent the traffic condition. Traffic flow and vehicle speed are such two important variables that can represent traffic condition. To identify variables q is taken as traffic flow (veh/min), v is taken vehicle speed (km/hr), l is chosen for location and t is chosen for time.

Flow at the current location at current time ($q_{l,t}$), Speed at the current location at current time ($v_{l,t}$), Flow at nearest upstream of the current location at current time ($q_{u,t}$), Speed at upstream of the current location at current time ($v_{u,t}$), Flow at current location at 5 minutes before current time ($q_{l,t-5}$), Speed at current location at 5 minutes before current time ($v_{l,t-5}$), Flow at nearest upstream of the current location at 5 minutes before current time ($q_{u,t-5}$), Speed at nearest upstream of the current location at 5 minutes before current time ($v_{u,t-5}$), Flow at current location at 10 minutes before current time ($q_{l,t-10}$), Speed at current location at 10 minutes before current time ($v_{l,t-10}$), Flow at nearest upstream of current accident location at 10 minutes before current time ($q_{u,t-10}$), Speed at nearest upstream of the current location at 10 minutes before current time ($v_{u,t-10}$).

3.2. Model Formulation

The formulation of the hazard prediction index consists of five steps. The process of this model formulation is described as follows (Fig. 1):

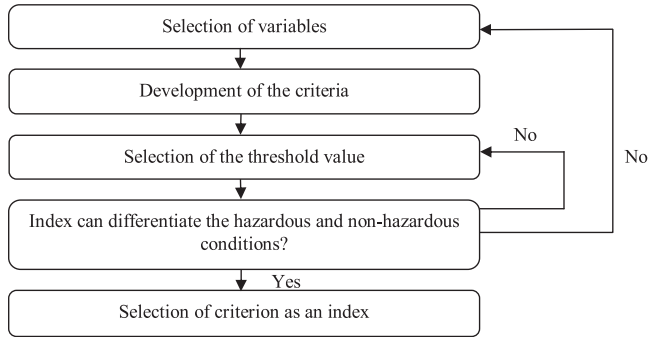


Fig. 1.
Accident Prediction Index Formulation Steps

3.2.1. Selection of Variables

The first step of model formulation is to select the independent variables from the list of variables mentioned in previous section. For each index, three different traffic flow variables and three different speed variables are selected by considering the location and time parameters. Among these three traffic flow and three speed variables, one variable is considered as the ‘controlling variable’ which is exclusive for each index. The controlling variable is used in an index to set the index criteria by measuring the variations of the other selected variables in the index from controlling variable.

3.2.2. Development of the Criteria

Every accident probability index consists of four criteria, two of them are for speed and the other two are for traffic flow. The criteria are developed by calculating absolute percentage variation of one variable from the controlling variable.

3.2.3. Initial Selection of Threshold Value

Threshold values were used to define the separation between hazardous and non-

hazardous traffic conditions. The threshold value for each of the criteria and combination is selected using an iterative method i.e. trial and error. First, two threshold values; one for the traffic flow criteria and other for the speed criteria are selected.

3.2.4. Differentiate the Hazardous and Non-Hazardous Conditions

To differentiate the hazard and non-hazard conditions, the criteria are checked in the model building dataset which includes equal number of hazard and non-hazard cases. If the variables in any case meet the condition of the criteria, it is marked as 1, otherwise 0. In this way, the values against all four criteria are summed up. The process is repeated for each case (accident and non-accident). Later, these values are separated and the average value for hazard and non-hazard cases is calculated. Next, the numerical differences of the average predicted hazard and non-hazard cases are calculated.

3.2.5. Selection of Criterion as an Index

If the difference of the average values that can separate the hazardous from non-hazardous situations significantly, the criteria are taken

as accident prediction index. The main objective is to maximise the difference of the average values of hazardous and non-hazardous cases. The objective function is to maximise the difference between the average value of hazard cases and average value of non-hazard cases. In this study, significant values are considered as ratio (average predicted value of non-hazard cases / average predicted value of hazard cases) that are greater than 50%. If the difference is not significant, then the threshold values are needed to be changed until highest difference can be found. After achieving the highest difference, the criterion is considered as the index which can be able to filter the hazardous and non-hazardous traffic conditions. If several changes of threshold

values do not differentiate hazard and non-hazard cases significantly, the criterion is not considered as an index and another criterion is selected for building the index.

4. Selected Hazard Prediction Index

For this study, seven different hazard prediction indexes are developed which are capable to differentiate hazardous and non-hazardous conditions based on different independent variables. The criterion for each of these indices is mentioned in Table 1. Each index includes four criteria and these are divided into two subsets with “and” which means there will only be a hazard if the traffic condition meets the criteria of the two subsets.

Table 1

Criteria of Hazard Prediction Indices

If the any of these two criteria fulfilled, there will be a hazard on the roadway; otherwise there will be no-hazard.

Index No.	Index criteria
1	{[abs(q _{lt,5} - q _{lt})*100 / (q _{lt})] >= 25 or [abs (v _{lt,5} - v _{lt})*100 / (v _{lt})] >= 10} and {[abs(q _{lt} - q _{lt})*100 / (q _{lt})] >= 25 or [abs (v _{lt} - v _{lt})*100 / (v _{lt})] >= 10}
2	{[abs(q _{lt,5} - q _{ut,10})*100 / (q _{ut,10})] >= 25 or [abs (v _{lt,5} - v _{ut,10})*100 / (v _{ut,10})] >= 10} and {[abs(q _{lt} - q _{ut,10})*100 / (q _{ut,10})] >= 25 or [abs (v _{lt} - v _{ut,10})*100 / (v _{ut,10})] >= 10}
3	{[abs(q _{lt,10} - q _{lt,5})*100 / (q _{lt,5})] >= 20 or [abs (v _{lt,10} - v _{lt,5})*100 / (v _{lt,5})] >= 5} and {[abs(q _{lt} - q _{lt,5})*100 / (q _{lt,5})] >= 20 or [abs (v _{lt} - v _{lt,5})*100 / (v _{lt,5})] >= 5}
4	{[abs(q _{ut,10} - q _{ut})*100 / (q _{ut})] >= 20 or [abs (v _{ut,10} - v _{ut})*100 / (v _{ut})] >= 10} and {[abs(q _{lt,5} - q _{ut})*100 / (q _{ut})] >= 20 or [abs (v _{lt,5} - v _{ut})*100 / (v _{ut})] >= 10}
5	{[abs(q _{lt,5} - q _{lt,10})*100 / (q _{lt,10})] >= 20 or [abs (v _{lt,5} - v _{lt,10})*100 / (v _{lt,10})] >= 5} and {[abs(q _{lt} - q _{lt,10})*100 / (q _{lt,10})] >= 20 or [abs (v _{lt} - v _{lt,10})*100 / (v _{lt,10})] >= 5}
6	{[abs(q _{ut} - q _{ut,5})*100 / (q _{ut,5})] >= 25 or [abs (v _{ut} - v _{ut,5})*100 / (v _{ut,5})] >= 5} and {[abs(q _{lt,5} - q _{ut,5})*100 / (q _{ut,5})] >= 25 or [abs (v _{lt,5} - v _{ut,5})*100 / (v _{ut,5})] >= 5}
7	{[abs(q _{ut,5} - q _{ut,10})*100 / (q _{ut,10})] >= 15 or [abs (v _{ut,5} - v _{ut,10})*100 / (v _{ut,10})] >= 5} and {[abs(q _{lt,10} - q _{lt,10})*100 / (q _{lt,10})] >= 25 or [abs (v _{lt,10} - v _{lt,10})*100 / (v _{lt,10})] >= 5}

5. Results and Discussion

Total dataset was divided into two subsets; one each for building and validation of the hazard prediction index. Model building subset consists of 80% of the total data and it is used for index formulation while model validation subset, consists of 20% of the total data and is used for validation of the index. Table 2 compares the prediction capability

of each index. As predicting hazardous situations leading to accidents is the main objective of the model building, strength of the each of the indices is evaluated both for total cases (accident and non-accident cases) and accident-only cases. The measurement of prediction capability of the indices is done with the percentage of error in predicting the actual situation, the higher the error percentage, the less the capability.

Table 2
Prediction Capability of the Accident Prediction Indices

Index Number	Total cases (hazard and non-hazard)	Total wrong prediction	% Error in total cases
1	53	17	32.07%
2	53	14	26.41%
3	53	14	26.41%
4	53	13	24.52%
5	53	10	18.86%
6	53	10	18.86%
7	53	16	30.18%

In the context of overall prediction capability, index 5 and index 6 have the least prediction error whereas index 1 and index 7 have quite higher value than that of index 2, index 3 and index 4. So, considering the overall prediction capability, index 5 or index 6 can be the best choice. But for hazard-only cases, the scenario is quite different. In comparing the percentage error in prediction of hazard-only cases, it is found that Index 5 and Index 6 have percentage error of 52.94% and 35.3% respectively. So, it can be said that for hazard-only cases, index 6 can be the best predictor of the hazardous conditions on the roadway. After comparing both types of error, it is decided that index 6 is the best crash prediction index. The selected criteria mean that there will be a hazard if the two criteria are fulfilled at a time:

1. The difference between flow at upstream at current time and flow of upstream at 5 minute before current time is 25% of the flow at upstream at 5 minute before current time or the difference between speed at upstream at current time and speed at upstream at 5 minute before current time is 5% of the speed at upstream at 5 minute before current time.
2. The difference between flow at the current location at 5 minute before current time and flow at upstream at 5 minute before current time is 25% of the flow at upstream at 5 minute before current time or the difference between speed at the current location at 5 minute before current time and speed at upstream at 5 minute before current time is 5% of the speed at upstream at 5 minute before current time.

6. Conclusion

In order to predict the crash hazard situations, this research developed a crash risk probability model called Crash Risk Prediction Index. This model is formulated by using traffic parameters i.e. traffic flow and vehicle speed at the upstream and current location during current time, 5 minutes before and 10 minutes before the current time. After evaluation, it is found that the developed model can predict crash risks with appropriate accuracy and it can be said that the result of the study can be applied to similar roads in any area as the selected site and data are in accordance with global standards. Outcomes of this research can be used to aid traffic control strategies e.g. ramp metering, variable speed limit, co-ordinated traffic signal etc. in order to reduce the hazard risk on the roadway.

References

- Eisenberg, D. 2004. The mixed effects of precipitation on traffic crashes, *Accident Analysis & Prevention*. DOI: [http://dx.doi.org/10.1016/S0001-4575\(03\)00085-X](http://dx.doi.org/10.1016/S0001-4575(03)00085-X), 36(4): 637-647.
- Golob, T.F.; Recker, W.W.; Alvarez, V.M. 2004. Freeway safety as a function of traffic flow, *Accident Analysis & Prevention*. DOI: <http://dx.doi.org/10.1016/j.aap.2003.09.006>, 36(6): 933-946.
- Goodwin, L.C. 2002. Weather Impacts on Arterial Traffic Flow, *Prepared for the FHWA Road Weather Management Program*.
- Griebe, P. 2003. Accident prediction models for urban roads, *Accident Analysis & Prevention*. DOI: [http://dx.doi.org/10.1016/S0001-4575\(02\)00005-2](http://dx.doi.org/10.1016/S0001-4575(02)00005-2), 35(2): 273-285.
- Hasan, M.M. 2012. Development of a crash risk probability model and its application to mitigate hazardous conditions on freeways, RMIT University, 2012. Available from Internet: <<http://researchbank.rmit.edu.au/eserv/rmit:160149/Hasan.pdf>>. Accessed 5th November 2010.
- Hong, D.; Kim, J.; Kim, W.; Lee, Y.; Yang, H.C. 2005. Development of traffic accident prediction models by traffic and road characteristics in urban areas. In *Proceedings of the Eastern Asia Society for Transportation Studies*, 2046-2061.
- Kalokota, K.R.; Seneviratne, P.N. 1994. Accident prediction models for two-lane rural highways [online]. Available from Internet: <<http://www.mountain-plains.org/pubs/pdf/MPC94-32.pdf>>. Accessed 21st November 2011.
- Pham, M.H.; Bhaskar, A.; Chung, E.; Dumont, A.G. 2010. Random forest models for identifying motorway Rear-End Crash Risks using disaggregate data [online]. In *Proceedings of 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 19-22.
- Shankar, V.; Fred, M.; Woodrow, B. 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies, *Accident Analysis & Prevention*. DOI: [http://dx.doi.org/10.1016/0001-4575\(94\)00078-Z](http://dx.doi.org/10.1016/0001-4575(94)00078-Z), 27(3): 371-389.