

Improving Safety and Security by Developing a Traffic Accident Prevention System

Mohamed Abdel-Aty, Nizam Uddin and Anurag Pande

University of Central Florida
Department of Civil Engineering
Orlando, FL 32816
(407) 823-5657
Fax: (407) 823-3315
mabdel@mail.ucf.edu

Topics

Transportation problems
Risk analysis, assessment and management

ABSTRACT

This research aims at developing accident prediction models to improve freeway safety. Also by reducing incident related congestion on freeways, response and evacuation times would also be reduced in emergency situations. There are considerable amounts of data that are collected and stored for ITS applications. This data includes speed, volume and occupancy provided by loop detectors. Most of these variables are known to be related to accident occurrence and patterns. Previous work has shown the effect of speed variation and volume on traffic safety. In most of the previous work, average or historical speed and volume data have been used. This study looks at detailed real-time data and its relationship to accident occurrence with the objective of determining whether it is possible to predict the potential of accidents before they occur. The data from a 13.25 mile segment of Interstate 4 in Central Florida equipped with loop detectors has been used. Preliminary analysis of detailed real-time speed data showed changes in speed upstream of accidents. Substantial variation in speed before the accident (both space and time) are found significant when compared to cases that experienced no accidents. The matched case-control logistic regression has been adopted and showed that the 5-minute average occupancy observed at the upstream station during 5-10 minutes prior to the accident along with the 5-minute coefficient of variation in speed at the downstream station during the same time have been found to affect the accident occurrence most significantly. This paper proves that real-time freeway loop detector data could be used in predicting accident likelihood 5-10 minutes before they occur. Therefore, Advanced Traffic Management Centers could attempt to prevent accidents by disseminating warning messages or adopting Variable Speed Limit techniques. The practicality of using real-time freeway loop detector data in predicting accident likelihood and hence prevent accidents is addressed.

Improving safety and security by developing a traffic accident prevention system

M. Abdel-Aty¹, N. Uddin² & A. Pande¹

¹*Department of Civil and Environmental Engineering*

²*Department of Statistics and Actuarial Science*
University of Central Florida, USA

Abstract

This research aims at developing real-time accident prediction models to be incorporated in Advanced Traffic Management Systems (ATMS). By reducing incident related congestion on freeways, response and evacuation times would also be reduced in emergency situations. The data from a 13.25 mile segment of Interstate 4 in Central Florida equipped with loop detectors have been used. Preliminary analysis of detailed real-time speed data showed changes in speed upstream of accidents. Substantial variation in speed before the accident (both space and time) are found significant when compared to cases that experienced no accidents. Logistic regression has been adopted and showed that the 5-minute average occupancy observed at the upstream station during 5-10 minutes prior to the accident along with the 5-minute coefficient of variation in speed at the downstream station during the same time have been found to affect the accident occurrence most significantly. This paper proves that real-time freeway loop detector data could be used in predicting accident likelihood 5-10 minutes before they occur. Therefore, Advanced Traffic Management Centers could attempt to prevent accidents by disseminating warning messages or adopting Variable Speed Limit techniques.

Keywords: real-time traffic accident prediction, advanced traffic management, driver warning, variable speed limits, traffic speed

1 Introduction

There are considerable amounts of data that are collected and stored for ITS applications. This data includes speed, volume and occupancy provided by loop detectors. Most of these variables are known to be related to accident occurrence and patterns. Previous work has shown the effect of speed variation and volume on traffic safety. In most of the previous work, average or historical speed and volume data have been used. This study looks at detailed real-time data and its relationship to accident occurrence with the objective of determining whether it is possible to predict the potential of accidents before they occur. This provide significant potential for drastically reducing incident related congestion, accident severity, and in case of an accident the response and clearance time, which would all ultimately benefit the flow of traffic and reduce emergency response time in normal and emergency situations.

2 Background

There have not been many studies in the area of real-time accident prediction. Lee et al. [1] introduced the concept of “accident precursors” and hypothesized that the likelihood of an accident is significantly affected by short-term turbulence of traffic flow. They came up with factors like speed variation along the length of the roadway (i.e., difference between the speeds upstream and downstream of the accident location) and also across the three lanes at the accident location. Another important factor identified by them was traffic density at the instant of the accident. Weather, road geometry and time of the day were used as external controls. With these variables, an accident prediction model was developed using log-linear analysis. In a later study, Lee et al. [2] continued their work along the same lines and modified the aforementioned model. They incorporated an algorithm to get a better estimate of time of the accident and the length of time slice (prior to the accident) duration to be examined. It was found that the average variation of speed difference across adjacent lanes doesn't have direct impact on accidents and hence was eliminated from the model. They also concluded that variation of speed has relatively longer-term effect on accident potential rather than density and average speed difference between upstream and downstream ends of roadway sections.

Although these studies do indicate the potential of applying real-time loop detector data to identify “alarming” traffic patterns on freeways, the biggest shortcoming of their analysis is that the data used in these studies were coming from just one station downstream and/or upstream of the accident location. Alarming conditions leading to accidents on a freeway might actually originate far upstream and “travel” with traffic platoons until they culminate into an accident at certain downstream location. It is also important to note that if an accident prediction model has to be useful in preventing accidents we need to identify the accident prone conditions ahead of the accident occurrence time and not within just 5-minutes before the accident; so that

the Advanced Traffic Management Center (ATMC) has some time for analysis, prediction, dissemination of the information and then take the appropriate action.

To account for this possibility here, data from several stations upstream of the accident location at several time periods leading to the accident was examined. This will also serve the purpose of identifying how far in advance (in terms of both time and distance) of an accident occurrence certain freeway segment may be flagged in real-time due to high potential of an accident.

3 Loop detector data

Data used in this study is collected in the corridor of Interstate-4 (I-4) in the Orlando metropolitan area in the state of Florida. The corridor has underground dual-loop detectors installed, which provide the real time measurement for traffic flow variables.

The following data on I-4 is collected every 30 seconds: average vehicle counts, average speed, and average lane detector occupancy. This data used in the analysis is collected for each lane on I-4 in both directions and at stations spaced at approximately $\frac{1}{2}$ of a mile for approximately 13.25 mile stretch that passes through the city of Orlando.

The accident data was collected from the Orlando Police Department (OPD) for the year 1999. Using this data, the following information about each accident was obtained:

- Date of the accident.
- Time of the accident.
- Last upstream loop detector station.

The location and time of accidents are not known with precision due to the inherent inaccuracy of police reports. Therefore, a combined approach of shock wave analysis and rule-based model was developed to determine the time and location of freeway accidents using loop detector data. The model enables theoretical consistency with the underlying effects of freeway accidents on traffic flow patterns which are revealed through the real-time traffic data collected from loop detectors. Traffic flow patterns change temporarily at both the upstream and downstream sides of the accident location after an accident occurs, which motivates the use of the accident shock waves in the model. The rule-based model captures the progression of the accident shock waves from the accident site by determining the arrival times of the backward forming and forward recovery shock waves at nearby loop detector stations. The estimated speeds of these two shock waves are used to determine the location and time of accidents. For more information about this model, the reader is referred to Yu and Abdel-Aty [3].

A total of 670 accidents were recorded in 1999 in the study segment. For every accident, its time and location was identified initially using the police report, and accurately adjusted if needed using the algorithm developed by the authors using the shock wave speed (see above). For each accident, the nearest loop detector to the location of the accident was named the station of the accident. Traffic data at the real-time of the accident were extracted for a period of 30 minutes just before the time of the accident. This 30-minute data was extracted over a stretch of 3 miles, 2.5 miles upstream and 0.5 miles downstream of the location of the accident.

To compare traffic characteristics that lead to an accident (based on time and space) with corresponding normal traffic conditions that did not lead to an accident, traffic data were also extracted on all corresponding days to the day of every accident. The correspondence here means that, for example, if the accident had happened on Tuesday June/1/1999 at 4:00 PM at station 40. Then data were extracted for all Tuesdays of the same season from 3:30 PM to 4:00 PM and from station 35 to station 41 if the accident happened on the east bound or from station 39 to station 45 if west bound. To account for seasonal variation in traffic characteristics, two different seasons were defined as summer and non-summer. Summer season includes months May to August, while non-summer season includes April and September to November. It is important to note that the summer season in Orlando experience heavy rain conditions.

Based on the data availability, on average each accident case has 11 corresponding non-accident cases. The first step in the analysis was to filter the traffic data for the accident and non-accident cases. Loop detectors' data are known to suffer from inaccuracies due to intermittent hardware problems and other random errors. These errors manifest in the form of false speed, flow, and occupancy. Most of the times, the errors can be identified from the unreasonable values of traffic parameters. In this study, all unrealistic values were eliminated from the raw 30-second data. The unrealistic parameters include; Occupancy > 100 , speed = 0 or > 100 , flow $> 25 / 30$ sec., and flow = 0 with speed > 0 . Then, the 30 minutes were divided into six 5-minute intervals, named time-slices 1, 2, 3, 4, 5, and 6, with slice 1 being the last 5 minutes before the time of the accident. Also, The 3-mile distance covers 7 stations, named as stations A, B, C, D, E, F, and G, with F being the station of the accident (i.e., station closest to the accident location). Stations A to E are upstream stations while station G is downstream.

4 Preliminary speed analysis

The careful examination of the traffic data plots led to the conclusion that half an hour period before each accident and also five stations upstream and one station downstream of the accident would be more than enough to analyze. It also provided the indication to analyze the speed data values in short term as well as in long term. The short term consisted of ten minutes leading to the accident at one-minute

increments and the long term consisted of half an hour leading to the accident at five-minute increments.

The accident day loop data was compared with data at the same time and location on a similar day of the week that had no accident occurrence. The non-accident day should be the same day of week as the corresponding accident day so that the daily traffic variation between the accident and non-accident days could be excluded. Also, the non-accident day was picked close to the corresponding accident day in order to avoid the seasonal variation as well. Choosing the same location ensures controlling for the geometric factors.

Table 1: The comparison between short-term speed variance for an accident and non-accident day

Accident Number	TYPE	Variance of Spot Speeds on the accident day	Variance of Spot Speeds on a non-accident day	F-Statistic	p-value
1	RE	271.42	130.34	2.08	0.014
2	RE	334.74	37.84	8.84	0.00
3	RE	221.67	109.88	2.01	0.036
4	RE	25.04	8.95	2.79	0.00
5	RE	116.37	20.69	5.62	0.00
6	RE	53.39	45.45	1.17	0.60
7	RE	49.36	7.68	6.42	0.00
8	RE	137.93	13.49	10.23	0.00
9	RE	18.42	14.01	1.31	0.365
10	RE	289.00	11.88	24.32	0.00
11	RE	107.49	7.56	14.21	0.00
12	RE	38.08	15.20	2.50	0.029
13	RE	75.28	21.49	3.50	0.00
14	RE	254.15	11.47	22.14	0.00
15	RE	154.61	7.00	22.09	0.00
16	RE	184.15	8.70	21.15	0.00
17	SS	14.98	23.65	0.63	0.096
18	RE	90.88	15.53	5.85	0.00
19	RE	377.01	41.95	8.98	0.00
20	RE	87.93	96.81	0.91	0.674
21	AN	56.43	92.22	0.61	0.074
22	RE	101.61	187.69	0.54	0.018
23	RE	344.71	106.54	3.23	0.00
24	RE	4.89	7.15	0.68	0.532

To determine whether or not the variation in speed data recorded by the loop detectors may be used to predict the accidents, the variance in the accident and non-accident days' speed data was tested to investigate whether they differ by a statistically significant margin.

The speed values obtained from the loops were averaged over the three lanes. For the short term, 60 observations were obtained (a 10 minute period at 1 minute increment for 6 loop detector stations -- $10 \times 6 = 60$). For the long term, 30 minutes at 5 minute increments consisting of observations at 0, 5, 10, 15, 20, 25, and 30 minutes before the accident time were used for 6 loop detector stations. Hence, 42 observations ($7 \times 6 = 42$) for the long-term analysis were used. However, in some cases there were few missing values, so the number of observations might be less.

The short-term observations for the accident day were compared with the short-term observations in the non-accident day and similarly for the long-term observations. Table 1 presents the results of the F-tests on real time short-term spot speed variances of accident and non-accident cases. As shown in Table 1, in a sample of randomly drawn accidents: 22 accident cases were rear-end (RE), one side swipe (SS), and one angle collision (AN). Here, F-test uses a one-tailed test because the claim to be tested here is that the speed variance on the accident day is higher than that of a non-accident day. The p-values are also presented to provide further statistical information for decision making, because in a statistical test they provide a sense of the strength of the evidence against the null hypothesis (the lower the p-value, the stronger the evidence). Here, the p-value in each test indicates the probability that the true F-statistic is greater than or equal to the observed F-statistic value if the null hypothesis is true. It is clear from Table 1 that out of the 22 rear-end accident cases, used in this analysis, the null hypothesis of equal variances is rejected in 19 out of these cases with 95 % confidence level. Out of these 19 cases, the accident speed variance is found to be greater than the non-accident in all the cases but one (i.e., approximately 82% of the rear-end accidents followed the hypothesis – 18 out of 22 cases). The sides wipe and angle collisions were not expected to follow the hypothesis, since the variation in speed is likely caused by queue formation and differences in speed upstream and downstream of the accident location which affect mainly rear-end accident occurrence.

A similar analysis for long-term spot speed variance was conducted. Here the analysis focuses on a half hour of speed data before the accident at 5 minutes increments across six stations. Out of the 23 cases with complete data, the null hypothesis of equal variances is rejected in 17 cases with 95 % Confidence Level. Out of these cases the accident variance is found to be greater in all the accident cases. In these cases, where the accident variance is statistically higher than the non-accident variance, all the accidents, but one, were found to be rear end. The null hypothesis of equal variance was rejected in the case of Side Swipe (SS) accident but not in the case of the Angle (AN) accident. Therefore, there are only five cases, which have not conformed to the high accident variance hypothesis, out of 21

accidents that were in fact found to be rear end (i.e., about 77% of the cases followed the hypothesis).

From the analyses it is clear that before most of the rear-end accidents the speed data obtained from the loops upstream tend to show higher variance than a corresponding non-accident day. This provides the basis to further analyze the loop data in order to predict the probability of potential accidents.

To clarify the speed patterns in the loop data, two graphs are presented. These graphs show patterns in the speed values observed by the loop detectors prior to a typical rear-end accident, which occurred on April 21, 1999, 4:45 PM at station number 34. Figure 1 shows the speed variation for the 30 minutes period before the accident at the stations leading to the accident. Figure 2 shows a similar trend in the 5 minutes before the accident. Both figures depict the substantial variation in speed. It is worth mentioning that the same trend was found in most cases that met the hypothesis and that the cases of no accident had mostly constant speeds.

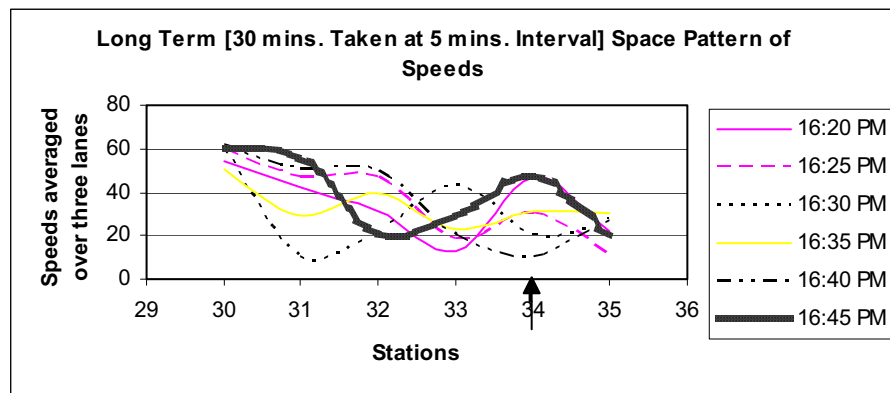


Figure 1: Long-term space pattern of average speeds over three lanes
Location of the accident: Station 34 (Shown by the arrow)

5 Accident-non-accident analysis

The purpose of the proposed accident-non-accident analysis is to explore the effects of traffic flow variables that affect both outcomes. A binary logistic model was applied.

For each of the seven loop detectors and six time durations, values of means (AS , AV , AO) and standard deviations (SS , SV , SO) of speed, volume and occupancy of all accidents and the corresponding non-accidents are available.

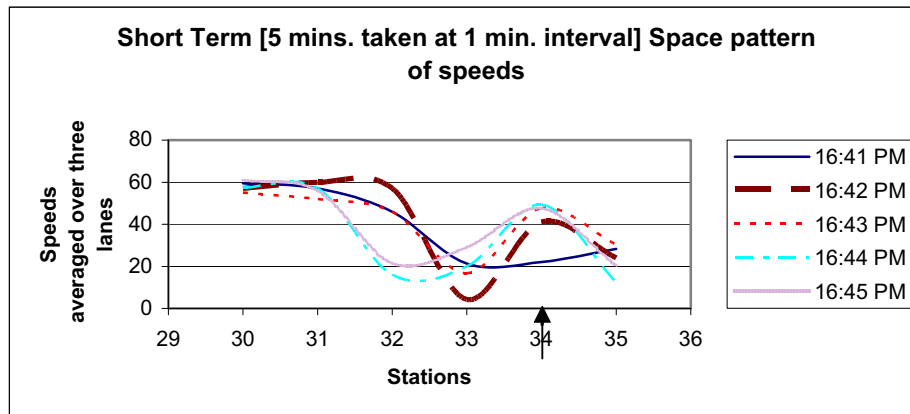


Figure 2: Short-term space pattern of average speeds over three lanes
 Location of the accident: Station 34 (Shown by the arrow)

A total of 252 explanatory variables are available. As a first step in the analysis of data with so many variables, we try to eliminate some (or reduce the number of) variables using simple summary results such as means and standard deviations and known results from published sources. Based on the literature review, the mean and standard deviation of speed was combined into one variable *CVS* (coefficient of variation of speed) as SS/AS . With five variables at each loop detector and each time duration, first conducted a stratified conditional simple (one variable at a time) logistic regression analysis was conducted to identify time duration(s) and loop detector(s) whose traffic characteristics are significantly associated with the binary outcome (accident, non-accident) variable Y . This was done by calculating hazard ratio using proportional hazard regression analysis of each of the 210 single variable models; one model for each of the five variables *CVS*, *AO*, *AV*, *SO*, *SV* over every station A-G and time slice 1-6.

It appears that only two variables out of the five are significant, $LOG(CVS)$ and *AO*. Hazard ratio is an estimate of the expected change in the risk ratio of having an accident versus non-accident per unit change in the corresponding factor. For example, a hazard ratio of 2.5 corresponding to coefficient of variation of speed (*CVS*) means the risk for an accident increases 2.5 times for each unit increase in *CVS*. Note that, for a continuous variable such as *CVS*, the hazard ratio is multiplicative in nature. This is to say that a two unit increase in *CVS* changes the risk by 2.5^2 or 6.25 when this risk is 2.5 for one unit increase in *CVS*.

Although time duration 1 exhibits significant differences between mean values of some variables, this is too close to the actual time of the accident and thus not useful in practice for accident prediction models. This time duration is thus ignored from further consideration. For each of the remaining time durations, we thus have $p = 20$ traffic flow variables; *CVS*, *AV*, *SV*, *AO*, *SO* at each of the four loop detectors D, E,

F, G. Each time duration data for each matched data set (1:m, m = 1,2,3,4 and 5) is now analyzed separately. No significant differences have been observed when changing m. Therefore, here a detailed description of the analysis of 1:5 matched data sets for time duration 2 is presented (i.e. 5-10 minutes prior to the time of the accident). To identify all significant variables from these 20 variables, the binary outcome variable y is now modeled. The distributions of these explanatory variables are skewed and thus the logarithms of these variables are used instead.

Two significant variables resulted: $g2logcvs = \log_{10}(CVS)$ from station G (log of the coefficient of variation of speed 5-10 minutes before the accident at the downstream station) and $e2logao = \log_{10}(AO)$ of station E (log of the average occupancy 5-10 minutes before the accident the upstream station). All other variables are found to be statistically insignificant. No variable from station F (nearest to accident site) is found to be significant. This should not be surprising since we are looking at time duration five to ten minutes before the actual accident time. Thus, the final model includes variables $g2logcvs$ and $e2logao$. Both variables have positive beta coefficients, which mean that the odds of an accident increases as these variables increase. A few accidents have very low values and few non-accidents have very high values for these two variables. These few cases may be considered unusual (caused by factors other than traffic flow characteristics, such as careless driving leading to accidents even on normal traffic conditions) and may bias estimates of beta coefficients. These accidents and non-accidents are thus deleted using 3-sigma rule. Estimates of beta coefficients and the associated summary results obtained from the final model are presented in Table 2.

Table 2: Logistic Model Description

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
g2logcvs	1	0.72875	0.27720	6.9113	0.0086	2.072
e2logao	1	0.52689	0.36080	2.1326	0.1442	1.694

6 Conclusion

This paper confirmed the hypothesis that speed variance leads to accidents on freeways. The initial analysis presented at the beginning of this paper showed that speed variation in both short and long terms affect the possibility of an accident to occur, which indicate that a real-time accident predictive model could predict the possibility of accident occurrence up to 30 minutes and 3 miles in advance. However, more rigorous analysis presented later in the paper showed that this might be a little too ambitious, and that predicting accidents 5 or 10 minutes in advance, and about half a mile before the accident occurrence could be achievable.

The results show that using the stratified case-control analysis, the log odds of accident occurrence may be obtained for a given value of certain traffic flow parameters. Hence, the potential “accident location” created due to ambient traffic conditions may be identified to warn the motorists about the impending hazard, to attempt to influence the speed to reduce its variation (variable speed limits could also be employed). The coefficient of variation in speed at the downstream station (Station G) and average occupancy at the station upstream of accident location (Station E) during time duration 2 (5-10 minutes prior to the time of the accident) are found to be the parameters most significantly affecting the accident likelihood. The threshold value of 1.0 for the log odds ratio based on them leads to identification of 68% accidents. It is worth mentioning that this analysis considered only traffic conditions leading to accidents.

The main conclusion of this paper is that real-time loop detector data currently used for many ITS applications could have a very promising potential to be used in accident analysis and prevention. Also by reducing incident related congestion on freeways, response and evacuation times would also be reduced in emergency situations.

References

- [1] Lee, C., Saccomanno, F., and Hellinga, B. (2003) Real-time crash prediction model for the application to crash prevention in freeway traffic. Presented at *the 82nd annual meeting of Transportation Research Board*, Washington, D.C.
- [2] Lee, C., Saccomanno, F., and Hellinga, B. (2002) Analysis of crash precursors on instrumented freeways. Presented at *the 81st annual meeting of Transportation Research Board*, Washington, D.C.
- [3] Yu J. and Abdel-Aty M. (2003) Combining Shock Wave Analysis and Rule-based Models to Determine the Time and Location of Freeway Crashes Using Loop Detector Data, manuscript.