Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

Hamzehei, Asso, Chung, Edward, & Miska, Marc
(2013)
Pre-crash traffic flow trend analysis on motorways.
In Kieu, L M & Hamzehei, A (Eds.) *Proceedings of OPTIMUM 2013 - International Symposium on Recent Advances in Transport Modelling.*
Smart Transport Research Centre, Queensland University of Technology, Australia, pp. 1-8.

This file was downloaded from: https://eprints.qut.edu.au/63216/

© Copyright 2013 [please consult the author]

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

PRE-CRASH TRAFFIC FLOW TREND ANALYSIS ON MOTORWAYS

Asso Hamzehei, Smart Transport Research Centre, Queensland University of Technology, Australia

Edward Chung, Smart Transport Research Centre, Queensland University of Technology, Australia

Marc Miska, Smart Transport Research Centre, Queensland University of Technology, Australia

ABSTRACT

Crashes on motorway contribute to a significant proportion (40-50%) of non-recurrent motorway congestions. Hence reduce crashes will help address congestion issues (Meyer, 2008). Crash likelihood estimation studies commonly focus on traffic conditions in a Short time window around the time of crash while longer-term pre-crash traffic flow trends are neglected. In this paper we will show, through data mining techniques, that a relationship between pre-crash traffic flow patterns and crash occurrence on motorways exists, and that this knowledge has the potential to improve the accuracy of existing models and opens the path for new development approaches. The data for the analysis was extracted from records collected between 2007 and 2009 on the Shibuya and Shinjuku lines of the Tokyo Metropolitan Expressway in Japan. The dataset includes a total of 824 rear-end and sideswipe crashes that have been matched with traffic flow data of one hour prior to the crash using an incident detection algorithm. Traffic flow trends (traffic speed/occupancy time series) revealed that crashes could be clustered with regards of the dominant traffic flow pattern prior to the crash. Using the k-means clustering method allowed the crashes to be clustered based on their flow trends rather than their distance. Four major trends have been found in the clustering results. Based on these findings, crash likelihood estimation algorithms can be fine-tuned based on the monitored traffic flow conditions with a sliding window of 60 minutes to increase accuracy of the results and minimize false alarms.

Keywords- Traffic Flow Fluctuations; Traffic Flow Trends; Motorway Crashes; Clustering;

INTRODUCTION

Crashes can occur on any part of a road network. However, among different types of roads, motorways (also referred as expressways, highways, and freeways) have received more attention from governments and researchers. Motorways play an important role in the traffic networks as arteries do in the human bodies blood vessel networks. Motorways transport a huge number of passengers and goods inter and intra cities. The economies of countries depend heavily on the flow of cars in motorways with less congestion and high speed. So, a crash on a motorway could have adverse effects on both the health of the people and can be detrimental to the economy. In this regard, authorities have tried to better control the motorways' traffic. Many motorways are equipped with a number of different kind of specialised sensors such as cameras, magnetic sensors, infrared sensors, microwave sensors, laser sensors, and inductive loop detectors (Lawrence A. Klein, 2006). In addition to these sensing technologies, there have been many traffic and transportation systems developed for monitoring vehicles, network traffic flows, transport infrastructure, and transport operators. The large volumes of data gathered from flow of vehicles have provided the opportunity for authorities and researchers to analyse this data and find new ways to reduce the motorway traffic risks factors as well as speed harmonisation and congestion reduction.

There is a necessity for suitable techniques to extract knowledge from huge and multi-dimensional road traffic flow data. In this regard, data mining has become an active research area. Data mining generally referred to as knowledge discovery in database (KDD) is a combination of statistical and Artificial

Intelligence (AI) techniques for extraction of patterns and knowledge stored in massive databases and data repositories.

Crash related studies have been aiming to reveal influential factors that impact on motorway crashes. Traffic flow data observed from inductive loop detectors has been the source data for such studies. Therefore, traffic flow variables such as speed, flow, and occupancy and their variance analysed to discover their relationships with crash occurrence. Data limitation and/or methodological shortcomings resulted in contradiction in findings of studies and sometimes incompatible conclusions.

In this study, relevant data such as traffic flow data (vehicle count, speed, and occupancy) from installed sensors in the study area and crash dataset with detailed information about crashes occurred in the site is used. The objective of this paper is to find dominant traffic flow patterns that lead to crashes. In this regard, speed is selected as the main factor to observe traffic flow fluctuations. One hour time window before crash occurrence is selected, so, for each crash one speed series is available. The speed series contain information of pre-crash traffic flow. The series are clustered using a non-hierarchical clustering algorithm (K-Means) to cluster different 1 hour pre-crash speed series.

The remainder of the paper is organized as follows: Section 2 presents the background and literature of crash traffic condition studies. Section 3 describes the main research aims and methodologies of the study. Results are presented in section 4 and section 5 provides the conclusion and summary.

LITERATURE REVIEW

Studies on motorway crashes can be divided into aggregate and disaggregate studies. Aggregate studies use traffic flow data aggregated hourly or longer while disaggregate studies use minutely traffic flow. Disaggregate studies which mainly conducted prior to 2002, discovered relationship between crashes and traffic flow. For example Martin (2002) examined the effect of traffic flow on crashes. He discovered severe crash rates are higher on light traffic condition and motorways with 3 lanes received more crashes than motorways with 2 lanes.

However, in more recent studies, disaggregate studies, Golob et al. developed a tool to monitor traffic safety by assessing traffic flow changes in real-time. They demonstrated 21 traffic flow regimes in three different times of day and weather conditions. As a part of their conclusion, they found that congestion strongly influence traffic safety (Golob & Recker, 2004; Golob, Recker, & Alvarez, 2004).

Zheng (2012) shows that Crash Occurrence Likelihood (COL) is not the same in different traffic conditions. The risk of crash occurrence was less for free flow condition while transition and congestion traffic condition received higher COL respectively. Zheng applied Logit model to study relationship between traffic condition and crash occurrence.

The main influential factor of crash occurrence on motorways is traffic states (Yeo, Jang, Skabardonis, & Kang, 2012). Yeo et al (2012) investigated the involvement of motorway crashes in four traffic states: Free Flow (FF), Back of Queue (BQ), Bottleneck Front (BN), and congestion (CT). Traffic data is being measured for upstream and downstream detectors of crashes in order to specify the traffic states. By plotting the speed of downstream and upstream stations of a crash they segmented the crashes into the four defined traffic states.

In addition, another aspect of crash studies is studying normal situation and mapping the crashes into the recognised regimes based on normal traffic situations. Pham (2011) clustered all the non-crash traffic flow data in order to identify traffic regimes and considered the traffic regimes as the non-crash situations (M. H. Pham, Bhaskar, Chung, & Dumont, 2010; M.H. Pham, El Faouzi, & Dumont, 2011) .

Although some research are conducted to study crashes in accordance with traffic states, but still this area of research requires more investigation. In the literature of traffic condition leading to crashes, traffic flow is considered just around the time of crash occurrence. These studies have tried to find relationships between traffic flow variables or traffic condition and crashes just before crash occurrence (5 minutes time window). In other word, the majority of literature has focused on impacts of traffic characteristics on crash occurrence or just a particular traffic condition. There is lack of a thorough research on traffic conditions that ended with crashes. Moreover, crash likelihood estimation studies commonly focus on traffic conditions in a short time window and longer-term pre-crash traffic flow trends are neglected**.** As a result, in this study we aim to fill the current gap in the study of traffic condition of crashes.

STUDY SITE

The study sites are two Tokyo inner city expressways of 24 kilometres length in total which including 3180 crashes over two years (2007-09). There are 201 loop detectors spread along the study site and data are available for this two years period. The data include average speed, volume, and occupancy aggregated over the lanes into five minutes intervals. The crash dataset includes reported crash time, location of the crash, type of the crash, number of cars involve in the crash, and type of cars. The accuracy of time of crashes is checked and fix with incident detection algorithm developed by the author.

METHODOLOGY

The objectives of this research are understanding traffic patterns that end-up with a crash, finding dominant traffic patterns that cause crashes, exploring existence of relationship between pre-crash traffic flow patterns and crash occurrence on motorways, increasing Crash Likelihood Estimations accuracy, and categorising crashes according to their pre-crash traffic flow trends.

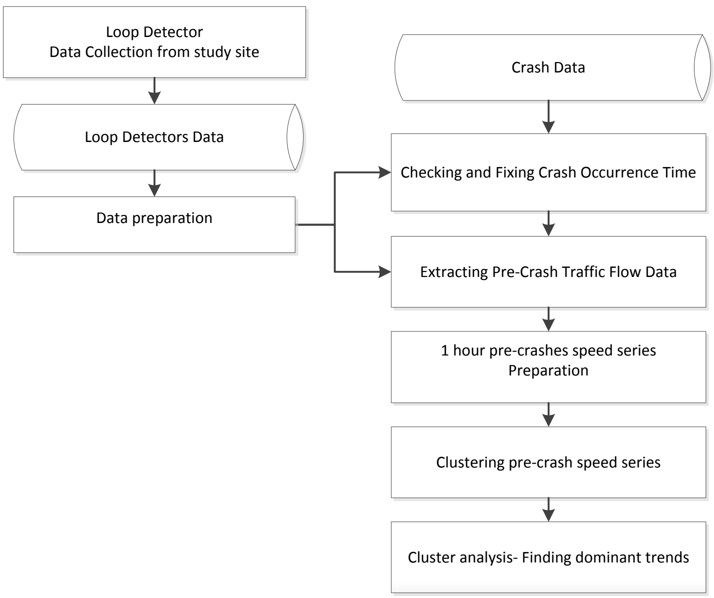

Figure 1 Methodology of the study

A methodology proposed to discover dominant risky traffic flow patterns. The skeleton of the methodology includes several steps. First, loop detector data is being collected from the study site. The data requires major pre-processing. Second, among the crashes, rear-end and sideswipe crashes are selected. The next step is checking the accuracy of reported time of crash occurrence. The extracted pre-crash traffic flow data are being pre-processed and speed series for 1 hour before crashes is prepared for analysis. The speed series of crashes are clustered using K-Means algorithm to discover dominant trends existent prior to crash occurrence on motorways. Consequently, in the last step, the clusters profiles are being examined to check the further differences between clusters in terms of time of crash occurrence, day of the week.

Traffic situation is the state of the traffic flow which is being measured by traffic flow detectors. The aim of traffic situation is to explain the safety level of traffic flow in a specific road section and time period. This research divides traffic situations into pre-crash situation and non-crash situation. A pre-crash situation refers to traffic state in a period of time prior to a crash in the crash location. Traffic flow in this period of time is considered as a risky state. As it is mentioned in the scope of the research, crashes would be studied that are caused by unstable or risky traffic situations. Therefore, any variation in traffic flow variables can reveal the cause and mechanism of crash occurrence. In this regard, speed is selected to study the dynamics and fluctuation of traffic flow. As the objective is discovering dominant risky traffic flow patterns, the time window should be long enough to observe traffic fluctuations over time. The observation time period that traffic flow might have had influence on the crash occurrence is set to 1 hour. It means, for each crash 60 minutes traffic flow data prior to the crash occurrence from selected loop detectors will be extracted. However, the challenge might be why 60 minutes? Why not 45 0r 30 minutes. Shortening the time window causes a few of the clusters merged. For example there would be no difference between crashes that happened under 1 hour congestion traffic condition or 20 minutes congestion traffic condition. In other words, shortening the time frame scarifies some information about pre-crash traffic flow dynamics.

Loop detectors data randomly contain noises that may result in unreasonable values for speed, volume, and occupancy. Moreover, they might be out of order and not measure the flow values. In the noisy case, traffic flow values can be evaluated and discarded when the values for volume, occupancy, and speed are not reasonable. For example, there is a non-zero value for speed but the flow or occupancy is zero. Additionally, accidents should be checked whether the traffic data is available. The crashes that their corresponding traffic data is unavailable or noisy should be discarded.

As crashes are reported and recorded by humans, it might not be accurate. For instance, traffic flow data is used to develop models for predicting crashes. In the analysis or prediction of crashes, it is important to know the exact traffic flow situation. Therefore, the time of the crash should be precise. Incident detection algorithms can help to check the accuracy of crashes and find the exact time based on the traffic flow data.

When an incident happens, the road partially or fully becomes blocked. In any kind of road with different numbers of lanes, there are traffic pattern changes for upstream and downstream road sections of the incident location. This post incident trend is expected to continue for a period of time until the road is cleared. In the upstream of the incident, fewer cars can pass the incident location. Therefore, it is expected that the occupancy for the upstream section increases rapidly, while speed and flow decrease. On the other hand, flow and occupancy decrease and speed increases in the downstream section of the road.

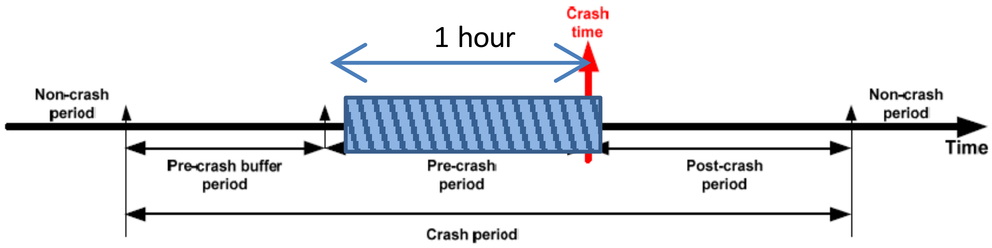

Figure 2 Crash and non-crash time periods

This research exploits the K-means clustering method to cluster extracted 1 hour pre-crash speed series. K-means clustering is a method of clustering which aims to partition N observations into K clusters in which each observation belongs to the cluster with the nearest mean. Normal evaluation of a proper K is to minimize the inner-cluster variation and maximize the among-cluster variation. K-means clustering is sensitive to outliers; therefore outliers must be deleted before running clustering algorithm on the data(Han & Kamber, 2006; Witten & Frank, 2005). Several distance functions can be used with K-Means clustering to calculate the distance between objects. The suitable distance function in this study is Euclidean distance function. It basically is the geometric distance in the multidimensional space. The following equation is showing distance between two vector of x and y:

Distance(x,y) = $\sqrt{\sum(xi - yi)^2}$ (Han & Kamber, 2006)

The obtained clusters represent different groups of risky traffic patterns. Dominant trends are frequent fluctuation trends which have been observed between many Speed series. In order to recognise such trends, clusters should have a considerable number of customers to be regarded as dominant trends. The number of clusters is set by Dunn index. The obtained clusters are being further analysed by creating their profiles to investigate the common similarities inside each cluster.

RESULTS

Speed series of rear-end and sideswipe crashes clustered for 2 to 30 number of clusters. The Dunn index value was maximum for 11 clusters (figure 4). Five different regimes are recognizable among the 11 clusters: Crashes where traffic flow was in free flow state during one hour prior to crash, crashes where traffic flow was in free flow but changed to congestion, crashes where traffic flow was in congestion but changed to free flow, crashes when traffic flow was in congestion during one hour before the crash, and crashes when traffic flow was unstable.

Table 1 Distribution of crashes inside each cluster and regimes

	C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11
Cluster members	126	112	132	24	6	120	38	54	40	104	67
Regimes	Free Flow		Transitoin-FF2C		T-C2FF	Congestion				Unstable traffic	
Regimes members	238		156		6	248				171	
Percentage	29%		19%		1%	30%				21%	

FF2C: Free Flow to Congestion
T-C2FF: Tansition-Congestion to Free Flow

TRAFFIC REGIMES

- FREE FLOW: clusters one and two contain speed series of crashes when traffic flow has been in free flow state since 1 hour before crash occurrence. The speed has been constant during one hour but was varied for different crashes. Traffic speed for crashes in cluster 1 varies from 50 to 70 and for cluster 2 varies from 70 to 90. So, both clusters have the same pattern, however cluster two has higher speed. Table 1 shows that Free Flow regime contains 238 crashes out of 824 which means %29 of crashes have occurred in free flow state. Among the weekdays, Saturday has received more crashes. Moreover, 7 am is the pick hour for Free Flow regime.
- TRANSITION FREE FLOW TO CONGESTION: clusters 3 and 4 contain crashes that happened in a transition traffic state from Free Flow to Congestion. The main factor of crash occurrence is congestion in downstream of crash location. While traffic is Free Flow in upstream and suddenly downstream turns to congestion condition, traffic in upstream face a fast deceleration. This fast deceleration is recognized as the influential factor in crash occurrence in this traffic regime. Table 1 shows that 156 crashes happened in this regime which is %19 of the whole crashes. Moreover, the pick hour for these crashes was in 6am and 3pm and weekday profile reveals that Sunday has received double of crashes than other weekdays while distribution of crashes in other weekday is almost in a same level.
- TRANSITION CONGESTION TO FREE FLOW: cluster 5 is the only cluster belongs to the transition regime from Congestion to Free Flow. The population of this cluster is very low about %1 of all crashes. Although this unique cluster is not as common as other clusters, but its uniqueness is important for Crash Likelihood Estimation studies. The special patterns like this regime might be able to be detected easier by COL models. So, we decided to not neglect them for further investigation in future studies.

- CONGESTION: four clusters (6, 7, 8, and 9) are belongs to the congestion regime. Cluster 6 is carrying crashes that have been in congestion condition for one hour before the crash time. Fatigue and tiredness of drivers during too many deceleration and acceleration can be a reason for crashes in this cluster. The traffic flow in the rest of clusters in this regime has turned into congestion condition from 40 to 10 minutes before the crash time. Thirty percent of crashes are located in congestion regime with population of 248 crashes. Among the weekdays, Friday received the most number of crashes in Congestion regime. Moreover, pick hour for crashes in this regime are 12pm and 6pm.

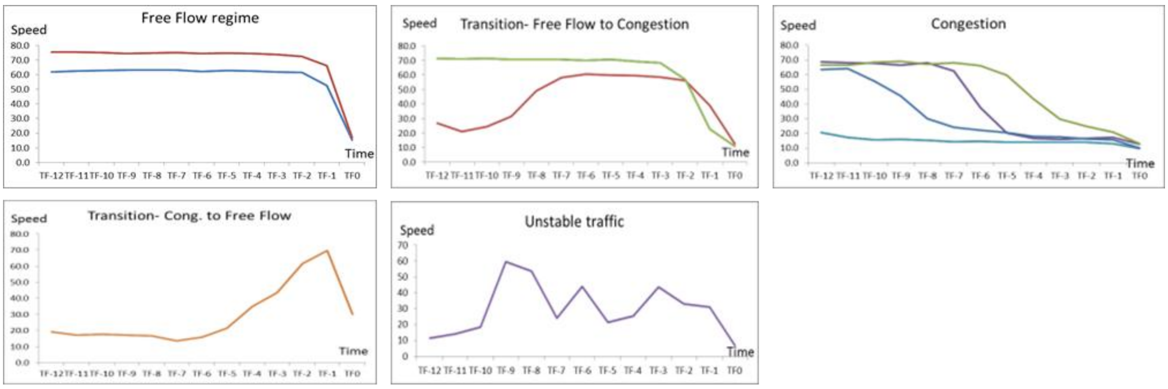

Figure 3 Traffic regimes

- UNSTABLE TRAFFIC: clusters 10 and 11 are labeled as Unstable Traffic regime because traffic has been fluctuating during the 1 hour before crash time. This regime contains most challenging crashes as crashes are not homogenous like other clusters. The values of traffic measures are changing continuingly and the changes are not following a same pattern. However, the profile of crashes in Unstable regime shows that 8am, 9am, 12pm, and 7pm are the pick hours for crashes with unstable traffic flow. Moreover, Saturdays among the weekdays have the most crashes with unstable traffic flow.

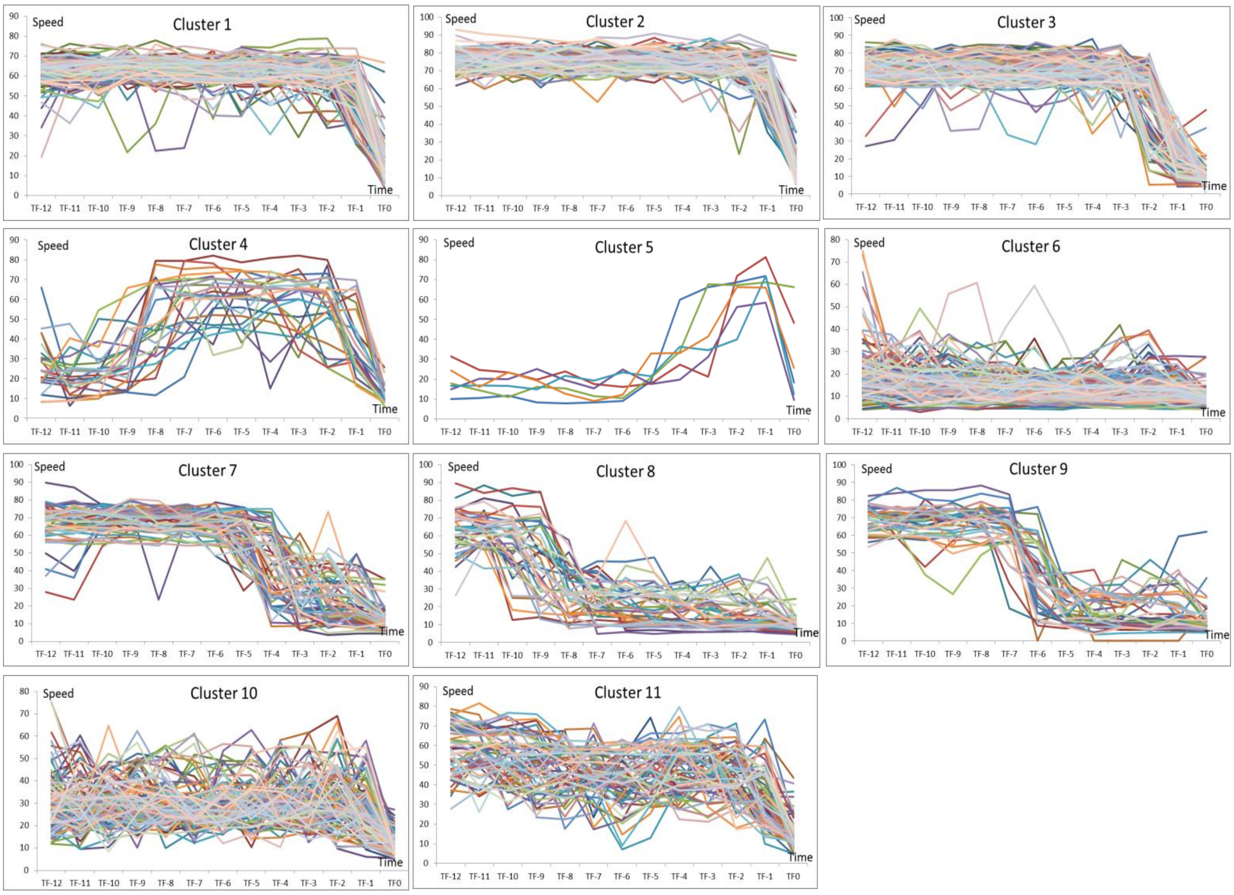

Figure 4 K-Means Clustering of Speed series with Euclidian distance function

CONCLUSION

Speed series clustered using non-hierarchical clustering algorithm (K-Means) and Dunn index is used to find the optimal number of clusters which was 11 clusters. Among the 11 traffic regimes, five major pre-crash traffic flow trends have been discovered in the clustering results: Crashes where traffic flow was in free flow state during one hour prior to crash, crashes where traffic flow was in free flow but changed to congestion, crashes where traffic flow was in congestion but changed to free flow, crashes when traffic flow was in congestion during one hour before the crash, and crashes when traffic flow was unstable. Future works of the current study will be to apply the clustering in crash estimation modeling and checking the accuracy of estimation models with and without having clustered crashes, investigating the Level of Service relationship with crash occurrence and the extracted traffic regimes. Moreover, taking segments of roads (ramps and basic flow segments) into account is another dimension of expansion the current study.

REFERENCES

Golob, T. F., & Recker, W. W. (2004). A method for relating type of crash to traffic flow characteristics on urban freeways. *Transportation Research Part A: Policy and Practice, 38*(1), 53-80. doi: 10.1016/j.tra.2003.08.002

Golob, T. F., Recker, W. W., & Alvarez, V. M. (2004). Freeway safety as a function of traffic flow. *Accident Analysis & Prevention, 36*(6), 933-946. doi: 10.1016/j.aap.2003.09.006

Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques* (Second Edition ed.): Morgan Kaufmann Publishers.

Lawrence A. Klein, M. K. M., David R.P. Gibson. (2006). Traffic Detector Handbook: Third Edition—Volume I (Third ed., Vol. 1, pp. 288). McLean, VA:

Turner-Fairbank Highway Research Center, Federal Highway Administration.

Martin, J.-L. (2002). Relationship between crash rate and hourly traffic flow on interurban motorways. *Accident Analysis & Prevention, 34*(5), 619-629. doi: http://dx.doi.org/10.1016/S0001-4575(01)00061-6

Pham, M. H. (2011). *Motorway Traffic Risks Identification Model - MyTRIM Methodology and Application.* Doctorate, École polytechnique fédérale de Lausanne, Lausanne (4998)

Pham, M. H., Bhaskar, A., Chung, E., & Dumont, A. G. (2010, 25-27 May 2010). *Towards a pro-active model for identifying motorway traffic risks using individual vehicle data from double loop detectors.* Paper presented at the Road Transport Information and Control Conference and the ITS United Kingdom Members' Conference (RTIC 2010) - Better transport through technology, IET.

Pham, M. H., El Faouzi, N. E., & Dumont, A. G. (2011). *Real-time identification of risk-prone traffic patterns taking into account weather conditions.* Paper presented at the 90th annual meeting of Transportation Research Board, Washington, DC.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Yeo, H., Jang, K., Skabardonis, A., & Kang, S. (2012). Impact of traffic states on freeway crash involvement rates. *Accident Analysis & Prevention*(0). doi: 10.1016/j.aap.2012.06.023

Zheng, Z. (2012). Empirical Analysis on Relationship between Traffic Conditions and Crash Occurrences. *Procedia - Social and Behavioral Sciences, 43*(0), 302-312. doi: http://dx.doi.org/10.1016/j.sbspro.2012.04.103