QUANTIFICATION OF BUS ACCIDENT ON RURAL ROADWAYS

By

MUAID ABDULKAREEM ALNAZIR AHMED

A dissertation submitted to the Department of Civil Engineering,

Lee Kong Chian Faculty of Engineering and Science

Universiti Tunku Abdul Rahman,

in partial fulfilment of the requirements for the degree of

Master of Engineering Science in

May 2017

ABSTRACT

Accidents are the outcomes of vastly complex random processes, whose general characteristics can be modelled statistically. Therefore, an accident risk is a quantifiable product probability and is a corresponding consequence. The causes of an accident, if successfully reproduced, can lead to another similar accident. Bus accidents, in particular, is a crucial issue to tackle as it relates to the public safety. Many existing studies have looked into the contributing factors of the occurrence of bus accidents. This study intends to investigate the relationship of bus accident occurrence with various factors namely driver characteristics, environmental conditions, and bus characteristics. Of these three factor categories, 13 variables were driven and studied to achieve the mentioned objective. Furthermore, these variables were then modelled to quantify the risk of an accident occurrence. A quantifiable model based on Bayesian network principal is found to be a useful tool to model the causality relationship among the studied variables.

The findings in this dissertation indicated presence of a gap between driver perception and road geometry via using Lamm model. Significant gap is found between operational speed and limit speed, and between developed and assumed friction values at various road sections. This result implies design inconsistencies which can lead to accidents. Moreover, based on Pearson linear correlation it was evidenced that the driver characteristics, bus function, working conditions and surrounding factors significantly influence the longitudinal acceleration and speed. This finding is further evidenced via Bayesian network model which presented that an accident occurrence is

directly influenced by driving conditions, and years of driving experience at investigated road. Motion parameters including speed and acceleration have no direct impact on accident risk.

ACKNOWLEDGEMENT

This study is supported by the Ministry of Science and Technology (MOSTI) Malaysia under the Science Fund, Project Number: 06-02-11-SF0175. We would like to express our sincere thanks to Professor Komiya and Mr. Daniel Leow for providing technical advises in data collection. In addition, we would like to thank the group of students who had assisted us to collect the data on sites.

APPROVAL SHEET

This dissertation entitled "**QUANTIFICATION OF BUS ACCIDENT ON RURAL ROADWAYS**" was prepared by MUAID ABDULKAREEM ALNAZIR AHMED and submitted as partial fulfilment of the requirements for the degree of Master of Engineering Science at Universiti Tunku Abdul Rahman.

Approved by:

____

(Dr. KHOO HOOI LING) Date: ………………….
Associate Professor/Supervisor
Department of Civil Engineering
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

____

(Mr. OOI EAN HUAT) Date: ………………….
Lecturer/Co-supervisor
Department of Internet Engineering and Computer Science
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

LEE KONG CHIAN FACULTY OF ENGINEERING AND SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

Date: ______________________

SUBMISSION OF DISSERTATION

It is hereby certified that MUAID ABDULKAREEM ALNAZIR AHMED (ID No: 1401377) has completed this dissertation entitled "QUANTIFICATION OF BUS ACCIDENT ON RURAL ROADWAYS" under the supervision of DR. KHOO HOOI LING (Supervisor) from the Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, and Mr. OOI EAN HUAT (Co-supervisor) from the Department of Internet Engineering and Computer Science, Lee Kong Chian Faculty of Engineering and Science.

I understand that the University will upload softcopy of my dissertation in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

____

(MUAID ABDULKAREEM ALNAZIR AHMED)

DECLARATION

I MUAID ABDULKAREEM ALNAZIR AHMED hereby declare that the dissertation is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

(MUAID ABDULKAREEM ALNAZIR AHMED)

Date: ______________________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

AASHTO	American association of state highway and transportation officials
DUI	Driving under influence
GDP	Gross domestic product
IDB	Irresponsible driving behaviour
PRT	Perception reaction Time
REAM	Road Engineering Association of Malaysia
SSD	Stopping sight distance
HSE	Health, safety and environment
TRB	Transportation Research Board

CHAPTER 1

INTRODUCTION

1.1 Background

The road network is seen as an important element for the overall economic development and social welfare for a country. In fact, the success and progress of a society strongly depends on capability of nation's physical infrastructure and transportation system to provide an adequate level of distributing resources and essential services to the public (Hudson et al., 1997; Deublein, 2013). A major component of the public transportation system is the buses which is agreed by many researchers to be effective measure to reduce the traffic congestion and accident on roads (Steg and Gifford, 2005; Mohamed and Kiggundu, 2007; Jayaraman et al. 2011).

Safety and comfort are major concerns of the commuters in the decision making of using a public transport service. Therefore, a bus accident gets much attention on media channel as the associated casualties and damages are often exceed that related with other mode of land transport. In Malaysia, the number of accidents involving a bus is about of 1.6% of the total road accident cases over the years (Ministry of Transport, 2016). Nonetheless, if the accident rate is compared, i.e. number of accident per registered vehicle, the accident rate involving bus is about three to four times higher compared to other mode of transport. Figure 1.1 shows the accident rate of bus, motorcar, and motorcycle in Malaysia from year 2004 to 2014 (Road Transport Department, 2016, Malaysian Institute of Road Safety Research, 2016).

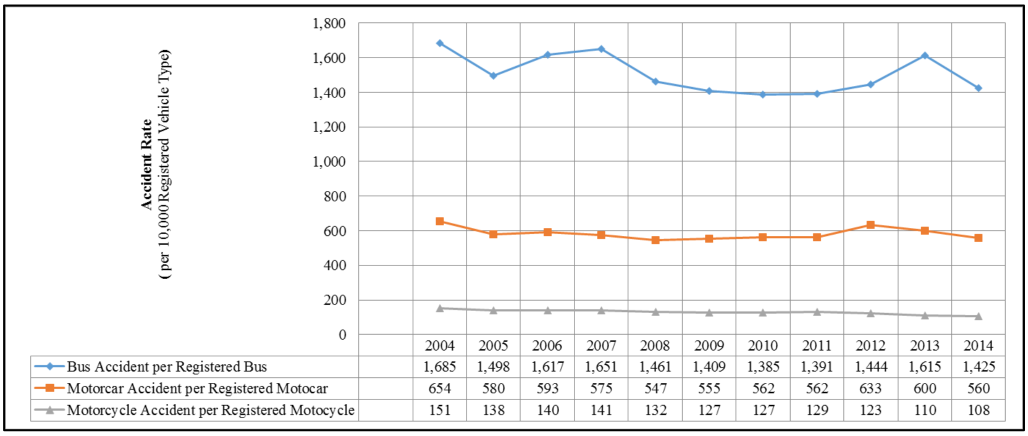

Figure 1.1: Bus accident statistics for the years 2000 to 2014

It is clearly observed that the annual accident rate for motorcycle is about 110 to 150 per 10,000 registered motorcycles while the annual accident rate for motorcar is about 550 to 650 per 10,00 registered motorcars. Bus accident rate is much higher which stands in the range of 1,440 to 1,650 per 10,000 registered buses from Year 2004 to 2014.

Based on accident statistics mentioned in Abidin et al. (2012), 89% of bus accidents are due to vehicle's overloading, mechanical failure (brake and tyre defects), road defects, and speeding as in Figure 1.2.

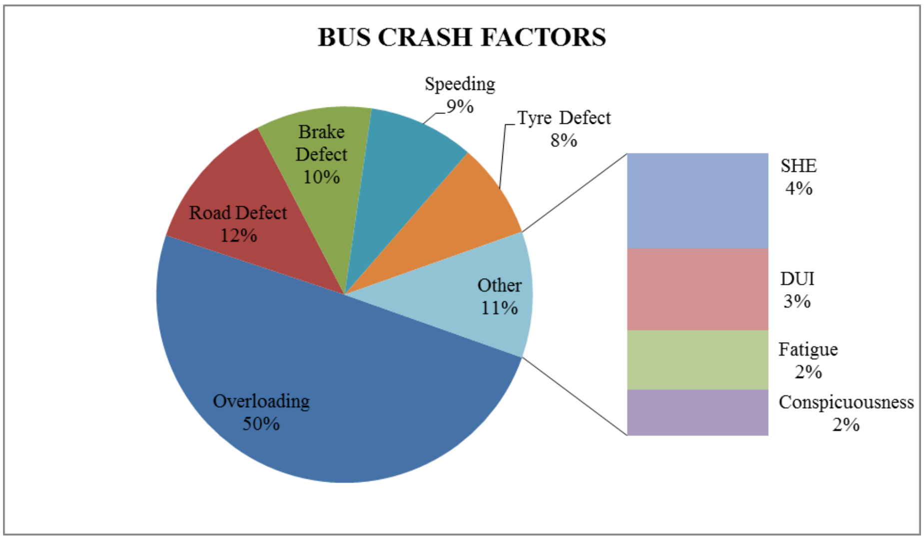

Figure 1.2: Bus crash factors in Malaysia

The years of 2007, 2010 and 2013 represent tragic years in Malaysia's road accidents. In 2007, a rollover accident accounted the lives of 23 people at Bukit Gantang, Perak. In 2010, a double-decker bus accident killed 28 people out of the 37 passengers, mostly tourists, on its way back from Cameron Highlands to Kuala Lumpur (Li Leen et al., 2016). The bus driver lost control of the bus as it was going down an incline. The bus crashed into divider, and overturned before it landed on its roof on a ditch at the 15KM of the Cameron Highlands-Simpang Pulai Road. While the worst bus accident in Malaysia's traffic history in the year 2013 when a bus plunged into a ravine near Genting Highlands, killing 37 of its 53 passengers (Gangopadhyay and Ng, 2013).

Such accidents tarnish the bus service and safety and such unfortunate events create stigma among bus users, travellers and tourists. This has seriously affected the public confident and willingness to use it. Furthermore, there are number of challenges which stand in the face of improving the bus safety in Malaysia including implementation of law and policy enforcement, and deficiency in domestic design standards compared to international standards, In late 2008, the department of Road Transport (RTD) made it compulsory requirement for the bus builders to install the seat built. As of 2013, only 5% of the 224 investigated buses by Solah et al. (2013) were equipped with seatbelts. Based on investigation carried by Solah et al. (2013) for a 224 buses, 41.4% of the buses failed to comply with the requirement of 'no obstruction at emergency path' and 73.3% of the sample population had no warning audible device to driver when emergency door is open. In terms of bus design and structure, the Malaysian regulation standard for bus design known as C&U1,

is not up to date and some rules do not comply with the international bus design standards UN R36 (Solah et al. 2013).

In the field of traffic safety and accident risk, many studies have been carried out on reducing or preventing the occurrence of an accident risk. Accident risk is defined as the product occurrence probability and the corresponding consequences (Kaplan and Garrick, 1981; Deublein, 2013). The occurrence of an accident is a complete combination of factors which, if reproduced, would result in another identical accident (Davis, 2001). The cause of accident is attributed to three factors, namely the driver, the vehicle, and the environment. The major factors of bus accidents identified are: driver, environment, and vehicle. The driver factor refers to personality of drivers such as his aggressiveness, experience, socio-economics characteristics, health condition, and working conditions. The environment factor refers to the road geometric such as road alignment design, cross-section, and pavement surface, traffic flow and type of control, roadside environment and infrastructure, as well as weather conditions. The vehicle factor refers to the bus condition, advanced technology adopted to assist driving, and the maintenance program. Stating these factors lead to the concept of causality in traffic accidents. Hence, the cause of an accident is a combination of factors which, if reproduced, would result in another identical accident. A causal factor is defined by Baker (1975) as “any circumstance contributing to a result without which the result could not have occurred” (Davis 2001).

These research findings are useful to plan and design the countermeasure strategies in tackling the bus accident. However, the major weakness of modelling studies is that most of them are post-accident models. They are formulated or modelled based on the accident data obtained from various sites or variables concern such as in Davis (2001), Sun (2006). De Oña et al., (2011). Zhang et al. (2013), Chen (2014), Zhao and Deng, (2015), and Deublein et al. (2015). Based on Hauer et al. (2002) (cited in Deublein et al. 2015) the methods which are only based on the observed numbers of accident events might lead to inaccurate modelling results; whether due to large variance on the dispersion or due to a systematic bias in predications. Perhaps, an accident preventive model would be more useful which could predict the accident risk with various factors. Accident preventive strategies can then be designed to reduce the occurrence of an accident.

A proper understanding of the bus driver behaviour and their interaction with the environment (such as road alignment) is crucial when formulating the bus driver training program. Furthermore, this understanding is also important for the road engineer or designer to avoid placing sharp curves or steep slopes that might increase the bus accident risk.

1.2 Problem Statement

Accident prediction models have to be provided to enable better understanding of accident causality and to improve road safety infrastructure. There are several deficiencies in accident risk modelling data source. The data used is normally based on post-accident reports. This is a restrictive approach in many ways. Contributing factors to crash occurrence and resulting severity are not collected e.g vehicle speed, driver braking, manoeuvring responses and other driving behaviour parameters. That lead to considerable unobserved heterogeneity which complicates modelling and precludes important information which could be used to make significant inferences.

1.3 Study Objectives

The objectives of this study are:

1. To investigate the relationship of bus accident risk with various factors inclusive bus driver, roadway, bus characteristics, and others.
2. To model a bus accident risk model using Bayesian principle.

1.4 Scope of Study

The study is limited to the bus routes that are running on the rural roadways. This is because the road alignment on rural highways are more challenging which impose higher workload on the driver. How the drivers behave and perform is one of the interesting elements which worth to be studied. Since the choice of bus routes is based on the site (rural area), the type of bus routes is not limited. They could be routine transit routes or interstate routes (served by tour bus or express bus). The types of buses studied are not restricted as well because the research was carried out on-board based on bus routes chosen. The buses board in the study include single decker (high & low floor) and double decker buses.

1.5 Organisation of Thesis

The thesis is divided into five chapters. In Chapter 1, introduction, the relevant background, problem statement and objectives are presented and discussed. The relevant and comprehensive literature review is included in Chapter 2. Chapter 3 comprises the detailed presentation of the novel methodology approach concerning this research. In Chapter 4; discussion and analysis, presents the development of the Bayesian forecasting model for the bus accident risk in rural roadways. In Chapter 5; conclusion, provides a summary of the main findings in this thesis, limitations of the current risk model. Besides that, recommendations are made for futuristic studies.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews the previous studies related to this research study. It provides the necessary background of the proposed methodologies adopted in this research study. First, a review of the accident contributing factors is carried out in which three major factors are identified including driver, environment, and vehicle. The review is performed for both private vehicle drivers and specifically on bus driver. Then, a review of the proposed methodology, i.e. Bayesian Network, is presented. The traffic studies that utilize Bayesian Network as the modelling or analysis tools are reviewed. Lastly, the limitations on the existing methodology in predicting bus accident occurrence are highlighted.

2.1 Factors of Accident Risk

Accident risk is defined as the product occurrence probability and the corresponding consequences (Kaplan and Garrick, 1981; Zurich, 2013). Researches in the probabilistic concepts have proved that accidents are the outcomes of a vastly complex random process, whose general characteristics can be modelled statistically (Elvik and Vaa, 2002; Chen 2014). The accident risk analysis is performed based on defined set of explanatory risk variables and dependent response variable.

The major factors of bus accidents identified are: driver, environment, and vehicle. The driver factor refers to personality of drivers such as his aggressiveness, experience, socio-economics characteristics, health condition, and working conditions. The environment factor refers to the road geometric such as road alignment design, cross-section, and pavement surface, traffic flow and type of control, roadside environment and infrastructure, as well as weather conditions. The vehicle factor refers to the bus condition, advanced technology adopted to assist driving, and the maintenance program. The following sub-sections elaborate the details of these factors.

2.1.1 Driver Behaviour Factor

Bus driver is one of the contributing factors to the occurrence of an accident. Factors such as personality, socio-economic, and working conditions have significant impact on driving behaviour.

Mallia et al. (2015) define personality as a relatively stable human characteristic which is not easily amenable by road safety interventions. Personality has control on individual behaviour that directly influences work performance and productivity (Ivancevich et al., 2005). In literature the role of personality characteristics on risky driving behaviour and accident risk is studied extensively. Some research studies focus on the impact of a single personality dimension on traffic accidents, while others estimate the crash accident on multivariate combination of different personality dimensions.

Mallia et al. (2015) investigated the role of personality traits and attitudes in bus crash risk in Italy. The three hundred and one (301) male driver subjects held driver license for about 15 years and worked as bus drivers for about 12 years within urban area. The subjects were asked to complete a structured questionnaire measuring personality traits, attitudes toward traffic safety, self-reported apparent driving behaviours including errors, lapses, and traffic violation, and accident risk in the past 1 year. Despite the findings indicated that the self-reported traffic violations were only factor related to bus accident risk, the personality traits were associated with driver's attitudes. Similar finding regards the association of driver personality traits-attitudes and attitudes-accident risk is found in Ulleberq and Rundmo (2003), Fishbein (2009), Ajzen (2011), Nordfjaern et al. (2010), and Lucidi et al. (2014).

Motivational factors such as risk taking and sensation seeking have major influence in driving at higher speeds and acceptance of shorter gap headways. The influence of motivational factors decreases with increasing age factor (Staplin et al., 1997). Moreover, human capability such as vision and reaction time usually decline with age.

Analysis of crash data involving franchised public buses in Hong Kong found the safest drivers were the older group aged 58–60 years, although explanatory factors were not analysed (Evans & Courtney, 1985). However, a study on the effects of age and driving experience on crashes among British bus drivers reported that experience had the strongest effect on crashes in the first year of driving, while age had a u-shaped association with accidents, that is, young and old drivers had more accidents (Dorn & af Wahlberg, 2008). On the other hand, a cross-sectional study among mini bus drivers in Jordan showed that driving experience and the age of drivers were negatively correlated with crash risk (Hamed et al., 1998).

Feng et al. (2016) investigated the effect of driver age and traffic violation history on bus involvement in a fatal accident in the United State between the years 2006 to 2010. The authors clustered the population driver to three categories including “middle-aged drivers with history of driving violations”, “young and elderly drivers with history of driving violations”, and “drivers without history of driving violations”. The findings indicated that under the same risk factors. The young and elderly drivers with history of driving violations are more likely to involve in severe accidents. The second in rank is the group with no driving violations, while the middle-aged drivers to be the safest category.

Kaplan and Prato (2012) developed a generalized ordered logit model to analyse the factors which affect the accident severity of bus accidents in the United States. The factors included driver characteristics, driving behaviour, environmental conditions, type of collisions, infrastructure characteristics, and interaction with other road users. The findings indicated that driver’s population of age group below 25 years and beyond 55 years possess higher accident severity. Moreover, female gender, over speed, low speed below speed limit, and intersection factors are positively correlated with accident severity.

Tseng (2012) studied the relationship between driver characteristics and at-fault accident for tour bus in Taiwan. Total of 2,023 drivers subject of whom 4.1% had at least one accident in a year period, of these, 68.7% were at fault for the accident. The findings revealed that a driver’s driving experience was the most crucial factor contributing to at-fault accident rate. Drivers with less than 3 years of experience have the highest at-fault accidents rate (12.4%). The drivers with more than 20 years of experience were not only the 2nd in rank of at-fault accident but their rating has doubled the overall average at-fault accidents of 2.8%. Drivers with driving experience between 6 to 8 years possessed the lowest at-fault accident rate (0.9%). Educational level is not significantly correlated with at-fault accident.

The stressful and draining working environment for bus drivers, expose them to the serious risk of occupational stress, driver fatigue and accident risk. In this literature some of the key themes related to occupational stress and fatigue are identified including the management support, working salaries and incentives, tight work schedules, working hours, turn-around and shift irregularity, interaction with passengers, and interaction with other road users, and travelling distance.

Biggs et al. (2009) conducted a survey study to evaluate the fatigue factors affecting the urban bus drivers in Australia. The findings found the tight route schedules can cause fatigue by producing time pressures, and therefore reducing the ability of the driver to utilize lay-over breaks. Moreover, extended shift cycles, and irregularity of shifts can lead to driver fatigue as a result of accumulation of sleep dept. In Tehran, Iran, Razmpa et al. (2011) reported that driver sleep problem affects accident rate. The National Sleep Foundation reported that 10% of bus drivers sleep less than 6 hours on their workday (Hege et al. 2015).

In developing countries the issue of sustainability becomes more complex to include issues of holidays, salaries and incentives. Bathija et al. (2014) found that more than 80% of government city bus drivers were under varying amount of stress in Hubli, India. Working conditions including hectic job schedules, difficulty in getting holiday, low salaries, and bad environment in bus have negatively contributed on driver's stress, driving performance and accident rate. Jayatilleke et al. (2009) investigated the effect of working conditions on bus accident rates in Sri Lanka. The authors found that the factors of disagreement about working hours and low salaries are significant factors for private bus crashes.

La et al. (2016) conducted a qualitative study to investigate the factors related to bus accidents in Hanoi, Vietnam using focus groups discussions and in-depth interviews. The subjects involved in the study include 75 participants who are, bus drivers, bus company managers, motorcycle users, bus passengers, traffic policemen and local authorities. The bus crash factors include variables of human factor, road and environment and vehicle factors. The findings revealed that supervision and penalty polices of bus companies influence the driving performance. Besides that, rapid increase in population due to urban migration, poor transport infrastructure, traffic behaviour of road users, and law enforcement have important impact on accident risk.

Chang and Yeh (2005) studied the factors affecting the safety performance of bus companies in Taiwan using a questionnaire based on the theory of organizational accidents. Environmental, organizational and driver factors were studied to determine the safety performance and accident risk. Forty two companies participated in the study. The findings of the study indicated significant association between accident risk and vehicle-specific and general management factors. The safety of large-sized bus companies compared to small or medium sized companies. Also, accident risk increases with old fleet buses compared to the newer ones.

Chang and Yeh (2005) investigated the link between accident rates and company characteristics such as rolling stock age, capital, vehicle mileage, and maintenance. The authors found those factors are statistically significant factor that contribute to accident.

Distraction is defined as a diversion of a driver’s attention away from the activities critical for safe driving (Lee eta al., 2008, Liang and Lee, 2014). Biggs et al. (2009) stated four possible scenarios of Interaction with passengers including (1) the passenger demand such as requests for directions, (2) aggressive passenger behaviour such as verbal or physical aggression, (3) noise and exposure to passenger vigilance, and (4) tasks of ticketing issuing and cash handling at every bus stop can cause driver distraction, fatigue, and acute stress respectively. The usage of technologies in vehicles such as navigation systems, smart phones, and internet-based devices exacerbate driver distraction from 0.6% to 0.9% (Liang and Lee, 2014). af Wåhlberg (2007) studied the effect of passengers on bus acceleration behaviour and the applicability of acceleration as incident prediction for transit buses. The author found that the incident records for those drivers who had many passengers increased the correlation between the driver acceleration behaviour and accident records compared to the drivers who had constant number of passengers.

Staplin et al. (1997) indicated the presence of a correlation between the annual travelled distance and accident rate. Drivers who had travelled more than 60,000 km/annum had at-fault accident rate of 4.3%. Similar findings for bus type of vehicle indicated by Tseng (2012) who found a positive correlation between annual travelled distance (more than 60,000km) and at-fault accident rate.

According to Hege et al. (2015) commercial drivers have been linked to wide range of health conditions including, among others, musculoskeletal and pulmonary disorders, cardiometabolic, overweight and obesity disorder. These health issues are associated with shorter life expectation than general population (Hege et al., 2015). Shift workers have a 40% increased risk of cariovasular disease compared to day workers (Bøggild and Knutsson, 1999). the direct effect of long shift work on cardiovascular disease (Haupt et al., 2008). Long working hours can cause irritability; physical and mental fatigue, excess sleepiness or insomnia, and inattention at work. Driver health condition could influence the accident rate and probability as well. Sumer (2003) showed that driver depression, anxiety, hostility and psychoticism has relationship to the accident rate, while af Wahlberg et al. (2009) mentioned that driver absenteeism has impact on predicting bus crash accidents.

Acceleration profile is a replica of the driver integration with the road environment and traffic status. A driver tends to decelerate, accelerate or maintain speed depending on his/her driving task and perception for the roadway environments. af Wåhlberg (2004) has found that the driver acceleration behaviour is a reliable measure to predict traffic accident. The assessment of driver acceleration behaviour (speed change) based on passenger comfort is studied extensively in bus literature. Borhan et al. (2014), Kamaruddin et al. (2012), Jayaraman et al. (2011), Mahmud (2010) studies aim to understand the factors which contribute to the willingness of a customer to use the public transportation especially buses. Majority, if not all, agrees that the issue of on-board safety is at least among the main influencing factors to use bus.

2.1.2 The Environment Factor

Despite the fatigue is usually associated as time-on-task factor, fatigue is likely to manifest itself in under-demanding driving conditions. Design consistency refers to geometry's conformance to driver expectancy (De Oña and Garach, 2012). Drivers make fewer mistakes in the vicinity of geometric features which fall within their expectation and more likely to involve in an accident at features which violate their expectations (De Oña and Garach, 2012). Wang et al. (2013) showed that 80% of traffic accidents were directly or indirectly related to the driver where road characteristics and environment have a significant impact on the driver's subjective perception of road safety. Under 'expected' situation, where drivers were aware of a hazard, the perception reaction time (PRT) was consistently 0.5 seconds faster than the 'unexpected' situation (Olson et al., 1984).

The driving environment impacts the development of driver fatigue and hence accident risk. This is because a poor design creates driver misperception of roadway and causes error that lead to accident.

Heger (1998) investigated the driver mental work as criteria of highway geometric design value. The study aimed to (1) quantify the mental workload requirements and the limitations of human information processing capabilities, and (2) combine operating-speed and workload for evaluating highway geometric design. The study results have revealed that the most psychophysiological parameters reach their maximum (1 to 2 seconds) behind the inconsistent design feature. This can explain the accidents occur on tangents after passing inconsistent design feature. The study recommends the use of mental work load as index to detect the inconsistencies in alignment and possible accident-related trends.

Thiffault and Bergeron (2003) conducted a simulation study to evaluate the impact of monotony of roadside visual simulation on driver's fatigue using a steering wheel movement (SWM) analysis procedure. Two types of road side stimuli were used in the study, the first stimuli case was repetitive and monotonous, while in the second case the environment contained disparate visual elements aiming to disrupt monotony without changing road geometry. The finding implies greater fatigue and vigilance decrements for the monotonous roadside compared to the disrupted environment.

Hassan and Easa (2003) showed that the horizontal curve looks consistently sharper when it overlapped with a crest curve and consistently flatter when it overlaps with a sag curve. Such misperception of horizontal curve has caused the speed variation to occur when driver navigating through the curves (Hassan and Sarhan, 2012).

Road alignment design is found to be statistically significant as a contributing factor of accident. This is because a poor design creates driver misperception of roadway and causes error that lead to accident. Hassan and Easa (2003) showed that the horizontal curve looks consistently sharper when it overlapped with a crest curve and consistently flatter when it overlaps with a sag curve. Such misperception of horizontal curve has caused the speed variation to occur when driver navigating through the curves (Hassan and Sarhan, 2012). Spacek (2005) showed that excessive steering corrections made by drivers could increase the centrifugal accelerations that cause the occurrence of loss-of-control or single-vehicle accident. Horizontal curves which require greater speed reductions from the approach tangent are more likely to have higher accident frequencies than horizontal curves requiring lower speed reductions (De Oña and Garach, 2012).

Spacek (2005) mentioned that drivers tend to correct his/her steering when navigating through the curves and such correction increase centrifugal accelerations that increase the risk of loss-of-control and single-car accident. Xu et al. (2015) recorded the lateral acceleration value that exceeded the comfort limit when they are navigating through curves on rural highways.

Similar findings are found in Kee et al. (2012) who conducted driving simulation study on the effect of long and monotonous driving task with different climate conditions on 25 bus driver subjects. The results included that prolonged driving had significantly induced fatigue level exclusively on monotonous roadways. Lui et al. (2009) stated that drivers ended with higher driving errors in monotonous driving environment.

Roadway Geometric Design and Traffic Flows are found to have impact on accident occurrence. For example, crash frequency is found to increase with traffic volume per lane (Miaou and Lump, 1993; Garber and Ehrhart, 2000; Chimba et al., 2010). Albertsson and Falkmer (2005) stated that 73% of buss crashes occur in urban roadways.

Chimba et al. (2010) studied the roadway and traffic factors in relation to transit bus crashes in the United States. The findings showed that the traffic volume per lane, presence of on-street parking, increase in number of lanes, and higher posted speed limit (excluding freeways) increase the probability of accident risk. In contrast, wider lanes and shoulder decrease the accident frequency. Kaplan and Prato (2012) stated that majority of bus accidents occur at two-way traffic roads.

Prato and Kaplan (2014) stated that the bus accidents in Denmark are almost equally distributed across season, across weekdays, and across time of day. Yang (2007) showed that major transit bus accidents occurred in benign environment conditions namely clear weather, daylight hours, and dry road surface. The evidence of this stated result can be found in earlier and later studies as in the following paragraphs. In support for these findings, Albertsson and Falmer (2005) found that majority of bus and coach accidents in eight European countries took place on urban roads and in dry conditions. Zahrah and Law (2004) indicated that accidents involved heavy vehicles most likely to occur in good weather compared to rainy weather. Kee et al. (2012) found that high vigilance level presented among bus drivers during night-time. af Wahlberq (2008) found that high temperature and rain in Sweden do not increase the risk of bus accident.

2.1.3 The Vehicle Factor

Despite the driving behaviour and its consequences are often the focus of research on bus safety, bus type has potential effect on accident occurrence. Blower et al. (2008) found that bus operation type is statistically significant factor on fatal bus crashes.

Bus Operation type: Long buses present operational challenges on roads with tight geometry (Chimba et al., 2010). Bus with lift movements were found to have higher accident risks which was attributed to be the likelihood of bus drivers running late, hence increasing accident risk (Strathman et a., 2010). Bus with low centre of mass namely double-deck becomes instable at high speeds, hill climbing's and acute corners.

Cabin Ergonomics: Several of the vehicle ergonomics are found to be associated with driver performance. Exposure to heat and glare, inadequacy of thermostatic control, neck, back and shoulder are all results from seating fit-out. Continuous exposure to chronic lower back pain leads to elevate level of physical distortion over time and deteriorates driving performance for heavy vehicle drivers (Nakata and Nishiyama, 1998). Exposure to heat, noise and vibration are linked to impaired driving performance with exposure to heat having the most negative effect (Wylie et al., 1997).

The deployment of **intelligent transportation system (ITS) technologies** such as driver assistance system can improve safety (Cafiso et al., 2013). Zegeer et al. (1993) mentioned that older bus has higher risk in getting an accident compared to the new ones. Tseng (2012) found that the use of automatic vehicle location system on tour buses was associated with lower at-fault bus accident rates. Adaptations of sleepiness-monitoring system such as image analyses are proven to be useful tool (Häkkänen et al. 1999).

2.2 Fundamentals of Bayesian Networks (BNs)

The cause of an accident is a complete combination of factors which, if reproduced, would result in another identical accident. This relationship of causality in traffic accident was first introduced in 1975 by Baker who defines a causal factor as "any circumstance contributing to a result without which the result could not have occurred" (Davis 2001). Causal assessment prior to an accident usually involves determining if a change in some prior condition could have prevented the accident. This causation mechanism is rather difficult and nearly impossible to reconstruct an accident because of the enormous amount of information and data prior to a crash which are often unobserved, intangible, and shadowed to clearly count in the accident report.

Uncertainty assessment in accident reconstruction can be seen as a special case of a method for causal analysis that has been formalized by Pearl in 1986, and later was given the name of Bayes network. This modelling approach has been widely accepted in other scientific fields and is gradually getting more and more space in the accident and safety studies.

The simplest form of Bayes theorem relates the marginal and conditional probabilities of events X and Y, provided event Y is not equal to zero is given as:

$$P(X|Y) = \frac{P(Y|X) * P(X)}{P(Y)} \qquad (2.4)$$

The qualitative aspect in this causality relationship can be represented by a graphical structure known as directed acyclic graph (DAG). The DAG represents a joint probability distribution (JDP) over a set of variables which are indexed by the vertices of the graph (Ben-Gal, 2007). The following Figure 2.1 illustrates a simple DAG which has three (3) variables.

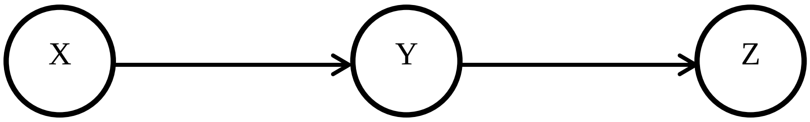

Figure 2.1: DAG for a network for three variables

The structure of DAG is comprised of two sets including the nodes (vertices) and arrows (directed edges). The nodes represent the random variables and are, usually, drawn as circles labelled by the variable names. The 'descendants' term is called upon the nodes that can be reached on a direct path from the node, and inversely the 'ancestors' term is called upon the nodes from which the node can be reached on a direct path (Ben-Gal, 2007). The terms ‘descendants’ and ‘ancestors’ are interchangeable with ‘child’ and ‘parents’ respectively.

The directed edge represents the dependency between two nodes, and is drawn as an arrow from the independent variable (parent node) to the dependent variable (child node). Therefore, the conditional probability of a child only depends on its parent node(s). The sets of conditional or joint probabilities are presented in conditional probability tables (CPT).

With a lager model and many variables, the conditional joint probabilities and structure of the network become computationally difficult. In this context, algorithms are developed to in model structure.

2.2.1 Structural Learning and Inference Algorithm

The learning algorithms are used to determine every possible edge orientation in a network and which direction to order the edge. The possible number of structures grows exponentially with the number of variables as every possible subset is a potential edge in the final model (Fast, 2015).

The structure learning of the Bayesian network is aimed to find the dependencies among the variables. There are two distinct methods to build a structure including expert knowledge and inference algorithm. Knowledge judgment of conditional dependences for structuring a network can influence problem formulation. If the variables are ordered carelessly, the resulting network may fail to reveal the right conditional independencies among the variables. In fact the casual semantics of a Bayesian network are in large part responsible for its success (Heckerman, 1996). Moreover, with large number of variables the possible formations of a network structure can be computationally tedious and difficult. Besides, in certain conditions, background knowledge of dependences among the variables can be hidden or unclearly understood. In this context algorithms for structure learning are developed to assist decision making such as Howard and Matheson (1981), Olmsed (1983), Pearl (1986), Shachter (1988),.Lauritzen and Spiegelhalter (1988), Jensen et al. (1990), and Dawid (1992). The computation of a probability of interest given a model is known as probabilistic inference (Heckerman, 1996).

There are two broad classes of structure learning algorithms including search-and-score, and constraint-based algorithms (Fast, 2010). The search-and-score or heuristic search technique class searches over possible Bayesian network structures to find the best factorisation of the joint probability distribution implied by the training data (Buntine, 1996, Heckerman et al. 1995). Search-and-score algorithms are generally very flexible and find high likelihood structure (Acid and Campos, 2001, Teyssier and Kollor, 2012). All possible edge additions, deletions, and reversals are considered. Greedy search, greedy search with restarts, best-first search, and Monte-Carlo methods are among the heuristic search algorithm class (Heckerman, 1996). The highest scoring edge is applied to the network and the algorithm continues until the highest score is achieved. Popular scoring functions include Bayesian Dirichlet (BDeu) (Buntine, 1991, Heckerman et al., 1995), Akaike Information Criterion (AIC) (Akaike, 1978) and Bayesian Information Criterion (BIC) (Schwarz, 1978).

The second class, constraint-based algorithms, learns the structures by first running local hypothesis tests to identify a dependency model M containing independence assertions that hold in training data (Cheng et al., 1997, Pearl, 2009; Fast, 2010). This class type, unlike search-and-score function, is more efficient and it is capable of enforcing conditional independence relationships and accurately produces the BN structure. The PC algorithm (Cheng et al., 1997), Sparse Candidate (SC) (Friedman et al., 1999), Max-Min Parents Childern (MMPC) (Tsamardinos et al. 2006), Max-Min Hill Climbing (MMHC) and Fast Adjacency Search (FAS) (Spirtes et al., 2000) are among this class of inference algorithm.

The third class of algorithms (known as hybrid algorithms) combines the techniques from both classes namely constraint-based and search-and-score algorithms. This class type is designed to take the advantage of the accuracy of the constrained-based algorithm while maintaining the flexibility of search-and-score algorithms (Tsamardinos et al., 2006).

Regardless of which approach is being used for structural learning, an optimal structure for a given set of variables is computationally demanding problem. Once the structure is specified, the user can make any kind of posterior inference, this is known as propagation of or posterior probabilities.

The advantage of using leaning algorithm is that it can accommodate for incomplete data. Information loss is not uncommon and can affect the model performance. There are two types of incomplete variables including observed variables with missing data cases, and non-observed (hidden) variables (Heckerman, 1996).

Incomplete data is caused by malfunctions or measurements errors in data collection, recording systems and processing errors (Sun, 2006). Bayesian network methods are suited to analyse both situations of incomplete data. Parameter learning algorithms are able to deal with missing information in the data set. Such algorithms including Expectation-Maximisation (EM) Gibbs Sample, and Metropolis-Hastings algorithms can exploit the partial information in the incomplete cases without affecting other cases (Deublein, 2013). Upon new data is available, the information can be updated to the conditional probability in the network, making the model performance can be continuously be improved.

Discrete and continuous variables can be used in BN model. Discretization is a common method because most of the learning algorithms are based on discrete variables (Xie et al., 2007). The BN models associated with continues variables are those limited to Gaussian variables and linear relationships (Xie et al., 2007).

2.2.1.1 Data Discretisation

Discretisation is the process of transferring continuous values of variables into interval (discrete) valued variable. Research shows that the discretisation increases the performance of the classification (Kaya et al., 2011). Moreover, discretisation increases the speed of induction algorithms (Dougherty et al. 1995).

Discretization methods have different types. There are 2 types including supervised and non-supervised learning. Unsupervised learning, unlike supervised learning, do not utilise instance labels in setting partition boundaries (Doughetry et al., 1995).

Supervised learning includes R algorithm (decision stumps), K-nearest neighbour classifier, Support vector machine classier (SVM) are among the supervised learning methods. Holte (1993, cited in Doughetry et al., 1995) sort the observed values of continues feature and attempts to greedily divide the domain of the feature into bins that each contain only instances of a particular class. K-nearest neighbour classifier (k-NN) is aimed to find the k-neighbourhood parameter which is determined in the initialization stage of the k-NN. The class of the value is determined according to the closest k-values among the data. Support vector machine classier (SVM) is based on statistical learning theory which aims to separate two classes optimally by finding the maximum margins of the hyperplanes (Kaya et al., 2011).

Examples of unsupervised learning include equal width intervals, equal frequency interval, and Bracket Medians methods. Equal width interval is the simplest method of discretisation. It involves sorting the observed values of continuous feature and dividing the range into equally sized bins. The unsupervised learning method is vulnerable to outliers which may drastically skewed the range and possible information loss (Clarke and Barton, 2000). Equal frequency method, as the name implies, partitions the values to intervals depending on the frequency in the data. Either equal width interval or equal frequency interval, the number of classes can be determined by the user. In

Bracket Medians Method, the mass in continues probability distribution function is divided into equally spaced intervals, Then a bracket median d1 is computed based on the endpoint of each interval. The discrete variable is defined within the same probability of the bracket median (Chen, 2014).

In comparison, equal frequency method, can give a fair approximation of the continuous variables if the right number of intervals are chosen. Otherwise, over-partioning, splitting relevant groups or combine separate groups of values might affect the model accuracy.

2.2.1.2 Data Partitioning

A common practise is to use the cross-validation method to divide data to training and validation sets. To insure the data of the training and validation sets are correlated, statistical computation are performed. The measure of goodness of fit for the discretised partitioned data includes linear correlation coefficient, comparison of means, computation of deviation between actual and predicted, and confidence intervals (Xie et al., 2007). SPSS, SAS and Matlab are common software which are utilised for data portioning.

2.2.2 Performance Evaluation of Structural Inference

The learning algorithm looks for patterns in the training data such discovered patterns might be valid for the whole population. Therefore, a high accuracy on the training samples than on the whole population. Only accuracy test on an independent test data is a fair estimate of the whole population (Elkan, 2012). The phenomenon of relaying on patterns which are strong in the training data is called overfitting. In practice overfitting is an omnipresent danger (Eklan, 2012).

There are number of techniques which can be used to evaluate the structural inference. Among which is the cross-validation, Structural Hamming distance (SHD), mean absolute deviation (MAD), and mean squared prediction error (MSPE).

Elkan (2012) defines cross-validation as a confusion matrix based on using each labelled example as a test example only once. Meaning an example is used for training, if and only if it has not been used for training. The largest possible number of folds is equivalent to the number of data being tested. This special case of cross-validation known as leave-one-out crosses validation. However, this type is associated with time complexity. Acceptable and common choice for number of folds is 10 (Elkan, 2012). The confusion matrix obtained by cross validation is intuitively a fair indicator of the performance of the learning algorithm on independent test data set. The confusion matrix for a 2 possible values of a variable is illustrated in the following Table 2.5.

Table 2.1: Confusion matrix for a 2x2 classifier

Confusion matrix		**Predicted Value**	
		Positive	Negative
Actual Value	True	TP	TN
	False	FP	FN

According to the above Table there are four s associated metrics with a confusion matrix are true positive, true negative, false positive, and false negative. True positive is an edge in the model which appears on true model.

False positive is a successful excluded edge in both models. False positive error is an edge in the model that does not occur in the true model. False negative error is an excluded edge from the learned model which appears in the true model. False positive error and false negative error are known as skeleton errors. Skeleton errors are form of errors in the binary edge decision for a pair of variables that is whether to add an edge in the model between those two variables (Fast, 2015). False negative errors are practically harmful for the model because they lead to edges being omitted form the model. The largest source of errors are the false negative errors (Fast, 2015). The possible sources of such errors are small sample size, high variance due to large degree of freedoms. Despite, no satisfactory solution for false negative error, possible correcting procedure is to use the scoring functions such as BDeu (Fast, 2015).

Depending on the application, there are number of evaluative statistics which are computed from these metrics including accuracy, precision, recall, and specifity (Xie et al., 2007)

Accuracy is the percentage of the correct edges (Peter et al., 2000, Bromberg and Margaritis, 2007). The following equation is used to compute this metric.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2.5}$$

Precision is defined as the number of correct edges divided by the total number of edges in the leaned model. The following equation is used to compute this metric (Elkan, 2012).

$$\text{Precission} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.6}$$

Recall or sensitivity is the number correct edges retrieved by learned model divided by the total number of existing relevant correct edges. The following equation is used to compute this metric (Elkan, 2012).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.7}$$

Specifity is defined as the number of wrong edges retrieved by learned model divided by the total number of existing relevant correct edges. The following equation is used to compute this metric

$$\text{Specifity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2.8}$$

Structural Hamming distance (SHD) is a graph edit distance and is equal to the number of edge deviations between the model and the true model. According to Fast (2010) to consider the decomposition of the SHD into skeleton errors i.e false positive and false negative errors.

Two evaluation criteria proposed by Oh et al. (2003, cited in Xie et al., 2007) evaluate the performance of the predicted and observed results including the mean absolute deviation (MAD), and mean squared prediction error (MSPE). MAD is employed to estimate the prediction deviation. While, MSPE is used for determining the variance of the difference between predicted and observed results. These evaluation criteria are given in the following equations.

$$\text{Mean Absolute Deviation (MAD)} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{2.9}$$

$$\text{Mean Squard Prediction Error (MSPE)} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{2.10}$$

The values of $\hat{y}$ and y correspond to the predicted and observed values respectively, and n is the size of the validation or training sets. Evaluation values closer to zero indicate better model performance (Xie et al., 2007).

2.2.3 Sensitivity Analysis

Forecasting models are used in decision making, and these decisions can be sensitive to small variations in probabilities. Marzban and Witt (2001) neural networks suffer from over-fitting problem. Besides that, the neural network models including Bayesian network have been long criticized for not being able to generate interpretable parameters for each explanatory variables for modelling crash frequency (Xie et al., 2007). To minimise these limitations, Fish and Blodgett (2003) and Delen et al. (2006) proposed a sensitivity approach to solve the problem of over fitting and to analyse the sensitivity of each explanatory variable. There are two types of sensitivity analyses including one-way analysis and n-way analyses.

The n-way analysis investigates the effects of simultaneous variation of n parameters compared to the one-way analysis which considers the study of one parameter at a time. Kjaerulff and Gaag (2000) stated that the n-way analysis serves to investigate the joint effects of inaccuracies in a set of parameters; therefore, this method is practical sound to measure the performance of the BN model. Conducting sensitivity analysis assists in conducting reasonable modification(s) to any variable within the BN model.

2.2.4 Model Testing and Validation

The validity of the model is established by comparing prediction results with historical records (Chen, 2014). The comparison should imply that the predicted results are close to historical records or both trend lines are similar (Chen, 2014).

2.2.5 Modelling Software

There are number of commercial and open source software which made the computational complexity associated with Bayesian network much easier. Thomas et al. (1992) created the first system called BUGS which takes a learning problem specified as a Bayesian network and compiles this problem into a Gibbs-sampler computer program. Despite that BUGS software is no longer available, another open source alternative is JAGS which uses Gibs sampler is available. Other software including TETRAD II, Uninet, BayesiaLab, Netica, OpenMarkov, Netica, BNGenerator and GeNIe are among many others (Heckerman, 1996, Fast, 2015, Chen, 2010, Deulein, 2010).

Despite some of the mentioned software are either open source or commercially available for the researchers to use, a researcher needs to decide a suitable software which suits the research aspect. Some restrictions are applied by software developer such as the size of the network being modelled, partial access to analytical tools (associated with commercial software), availability of inference algorithms, type of variables used in model construction (i.e continuous and discrete) and others. For instance, TETRAD II (developed by Scheines et al., 1994) can be used for constraint algorithms (Fast, 2015). JAGS uses Gibs-sampler algorithm. Bayeslab is a commercial software with a restricted downloadable trial version. Uninet deals with continues Bayesian network. GeNIe and Netica software develops models based on discrete variables. However, Netica has limitation to the size of the network which can be structured.

The GeNIe software is open source software which is developed by the Decision Systems Laboratory at the University of Pittsburgh is made available for the researchers and the community who are keen in developing and testing models. The advantage of this software is that the GeNIe is capable of building models of any complexity and size, the only limitation is the capacity of the operating memory of a computer (GeNIe, 2015). Moreover, the software is continuously updated. The availability of learning algorithms, user friendly, offered tutorials and continuous updates, online community support (forum) among others made the software the best choice for this thesis work.

2.3 BN in Traffic Engineering Studies

BN models began in to gain popularity in late 1990s and have been used even more since the 2000, only limited utilization of BN in traffic accidents were identified. The application of the Bayesian network for the analysis of accident risks and severity is still scarce (Xie et al., 2007, Zurich, 2013). According to Xie et al. (2007) the development of the BN was first initiated by Macaky (1992) and further improved by Neal (1995). The application can be seen in accident construction modelling, modelling of injury severity, and real-time accident risk prediction. The following case studies are found in the literature

Davis (2001) study aimed to develop an accident model to reconstruct vehicle/pedestrian collisions due to speeding factor. The study included 8 types of collision cases. The study shows a promising result, that adhering to speed limit could have positive impact on accident reduction.

Zhang et al. (2013) applied the Bayesian network to the identification of the accident severity in China in 2010. The results indicated, the probability of increment of property damage is associated with poor vehicle condition and irregular section of road and intersection. The latter testimony is justified as the driver decelerates when approaching the intersections and abnormal sections. The bus or truck involved and the poorer the vehicle condition variables do have direct impact on the number of injuries resulted from an accident.

Deublein et al. (2015) developed a Bayesian model to predict the number of accidents involving personal injury on the Swiss highway. The data were obtained from the FEDRO for the Swiss highway network for 3 years span; 2010 to 2012. A backward prediction is used to test the model. The model was used to predict the accidents for the year 2009 and compared with the accident data records for the same year. The number of accidents is correctly predicted on 86.53% of the road segment with a tolerance of 25%. Furthermore, the incorrect predictions were due to other many more factors which could affect the accident occurrence, such as presence of road works, ice or fog on highway, confusing sections with frequent congestion events.

Zhao and Deng (2015) developed a BN model for traffic fatalities and injuries at urban intersections in China. Total of 3,584 recorded crashes collected from the urban intersections of Changshu, China. The BN topological structure is developed to reflect the hierarchical characteristic of crash variables. The parameter learning process is completed with Dirichlet prior distribution. The results suggest the efficacy of BN approach in the prediction accuracy. For instance, the inferred probabilities of frontal collision at urban intersection crashes involving bicycles and electric bikes are 43.16% and 40.44% respectively. The average learned probability of illegal driving stands at 40.83%, which is found to be much higher than other learned probabilities of human factors. Heavy vehicles have a higher inferred probability in side collision than light vehicles, whose inferred side collision probability is 41.02%.

Chen (2014) developed a Bayesian network model. The study used accident data from Transport Canada's National Collision Database (NCDB) during the period of 1999 to 2010. There are 28 risk variables used in the study which were categorised to external environment conditions, operational conditions, driver conditions and vehicle conditions. The findings indicate that the BN model can be integrated with safety instrumented system (SIS) which acts as a risk-inferred warning system. The integral warning system acts as indicator of highway accident risk and attaching safety functions which helps creating an intelligent system to effectively prevent accidents and makes highway more safer for users.

De Oña et al. (2011) aimed to validate the possibility of using Bayesian network to classify traffic accidents according to their injury severity, and to measure the performance of the relevant variables which affect the injury severity of traffic accident on rural Spanish highway. The findings of the study are that the variables which best associated with a killed or severely injured accidents include accident type, driver age, lighting and number of injuries.

Sun (2006) modelled the traffic flow among adjacent road links in a transportation network using Bayesian network. The subjects used in the study are the urban vehicular traffic flow data of Bejing, China. The author adopted the Gaussian Mixture model (GMM) to approximate the joint probability distribution in the Bayesian network due to the small data in hand and difficulty to collect large amount of data concerning the traffic flow. The competitive expectation maximisation (CEM) algorithm is used to construct the Bayesian network. The results indicated that Bayesian network is very promising and effective approach for traffic flow modelling and forecasting, for both complete and incomplete data.

2.4 Advantage of BN as a Modelling Tool

The advantage of Bayesian network as an accident forecasting model includes (1) the applicability to model the complex relationships between the variables of traffic system, (2) the ability to accommodate for uncertainty and discovery for the unknown domains. Besides that, a key feature of BN is the reasoning of the uncertainties and discovery for the unknown domains (Poole, 2011). By representing the interactions, the BN yields a deep understanding and knowledge, not only of past events, but also allow practitioners to anticipate how a domain will behave under hypothetical circumstances (Conrady and Jouffe, 2011).

2.5 Limitations of Existing Studies

According to the conducted review in this literature, there is not yet a comprehensive model for accident prediction. The literature indicates the current researchers depend on the accident reports to build a forecasting model as in Davis (2001), Zhang et al. (2013), Deublein et al. (2015), Zhao and Deng, (2015), Chen (2014), De Oña et al., (2011), and Sun (2006). According to Hauer et al. (2002) (cited in Deublein et al., 2015) the methods which are only based on the observed numbers of accident events might lead to inaccurate modelling results; whether due to large variance on the dispersion or due to a systematic bias in predications. Mannering and Bhatl (2014) identified several deficiencies in depending on accident reports to develop a model among these mentioned disadvantages are undefined and under evaluated parameters especially those associated with the human factor. It is believed that the work result can be improved and different research related areas can be investigated.

CHAPTER 3

METHODOLOGY

3.1 The Methodology Framework

To monitor a driving behaviour, it is assumed that the highways and mountainous roads are suitable driving environments. Such roads assure freedom for a driver to cruise at speed of choice, besides that, a mountainous terrain provides a challenging geometric environment which requires much workload and attention from a driver. Thereof, five rural roadways are chosen for this research including B66 Jalan Batang kali, North-South expressway, Jalan Batu Feringghi, while Karak Highway and Route 59.

The research methodology consists of five stages. Stage 1: to determine the selection of data collection roads, and the data collection of risk factors including driver, passenger, bus and road geometric. Stage 2: to identify the suitable variables for this study. Stage 3: to evaluate the design consistency based on Lamm model to examine the current geometric alignment of studied roads to accommodate the driver behaviour in a quantitative measure. Stage 4: Pearson correlation is employed to identify the possible linear relationships among the variables of the study. Stage 5: involves step by step procedure to develop a predictive bus accident risk model. The following Figure 3.1 presents the flow chart of the followed stages in the research methodology

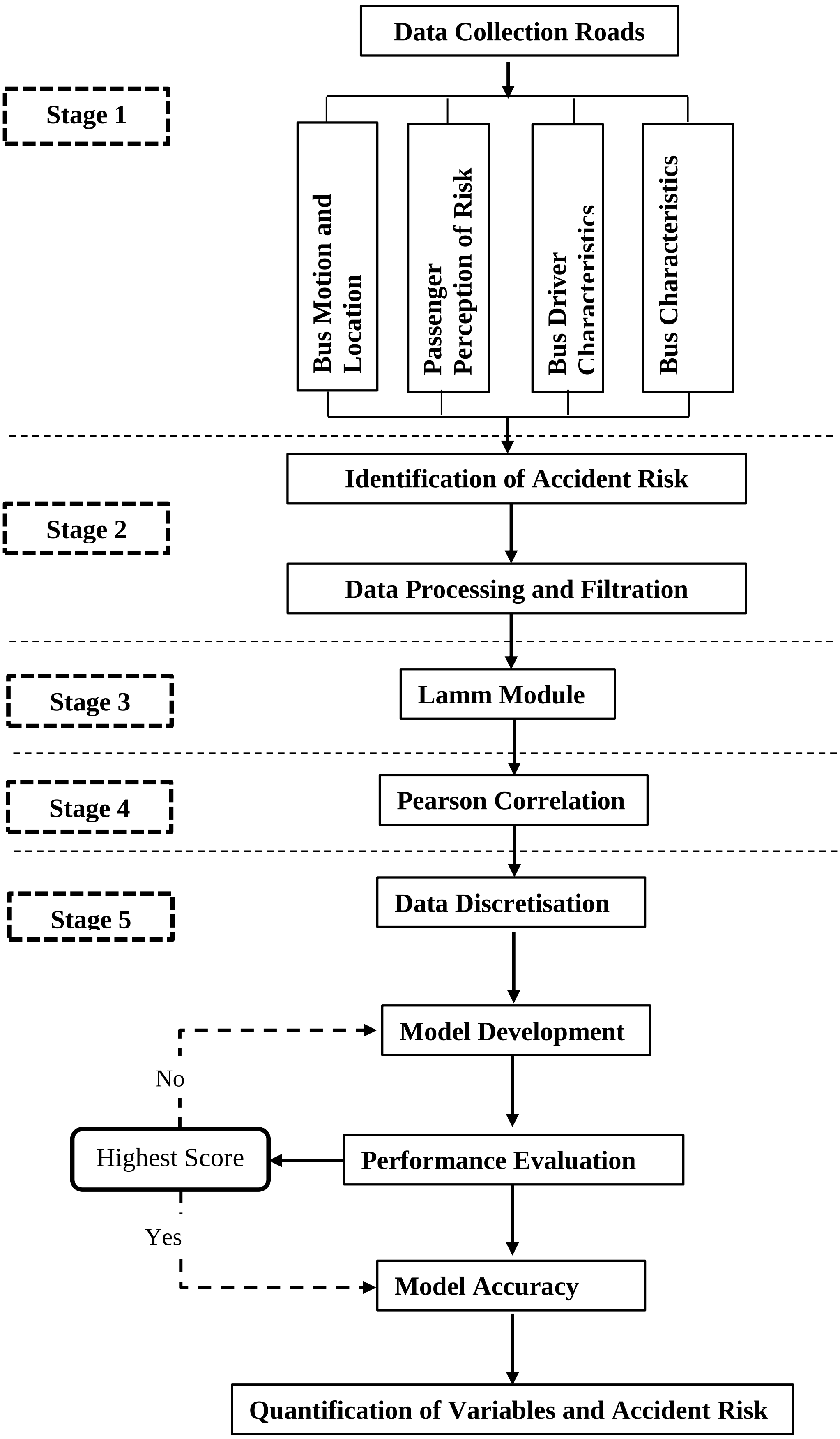

Figure 3.1: Flow chart of the followed procedures in the research methodology

3.2 Data Collection Roads

The data collection was conducted at off-peak hours to insure the least traffic congestions, and minimal surrounding distraction for the bus driver. The highways and the mountainous roads with none or few bus stops are of prime interest for the study. Highways are characterised with the freedom for the driver to cruise at speed of choice, though there is a posted speed limit. While, mountainous terrain provides a challenging geometric environment and requires much workload from the driver. This terrain would give a very good insight of the acceleration profile which is needed to establish safe limits for driving behaviour.

A total of five roads including B66 Jalan Batang Kali, Jalan Batu Feringghi, Route 99, Karak Highway and North-South Expressway were selected for this research as in the following subsections.

3.2.1 B66 Jalan Batang Kali

The B66 Jalan Batang Kali is located at Genting Highland in Pahang State. The road is rural roadway which adheres to category R3 based on the Malaysian road design standards known as REAM. The road is a dual-2 lane carriageway and it has speed limit of 50km/hr. Along the investigated road, there are no signalised intersection and only two roundabouts. The length of the investigated stretch of road is about 28km of length The following Figure 3.2 presents the stretch of road where data collection is carried out.

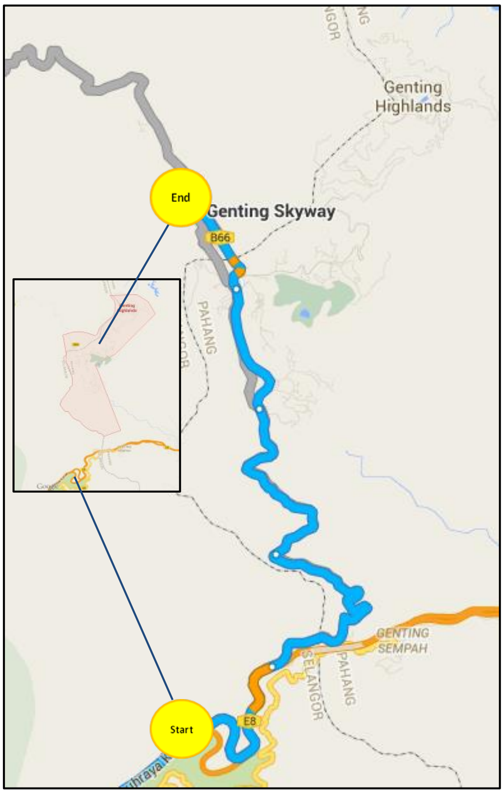

Figure 3.2: B66 Jalan Batang Kali, Genting Highland, Pahang state (Google maps a, 2015)

The data collection took five (5) days between the months of June and July 2014. In total, there are 28 trips were performed for the mentioned site with 16 different drivers. The bus superstructure is high floor single deck bus which accommodates around 50 passengers and no standing passenger is allowed. The following Table 3.1 presents a summary of B66 Jalan Batang Kali.

Table 3.1: Summary of data site information

Data Collection Period	**REAM Design Standard**	**Speed Limit (km/hr)**	**Data Collection (Start/End)**	**Bus**	**Total No. of Trips**	**Total No. of Drivers**
03, 17 Jun 2014 & 1, 5, 6 July 2014	R3	50	B66 Jalan Batang Kali - Genting Highlands/Skyway Terminal	Single Deck High Floor	28	16

3.2.2 North-South Expressway

The North-South expressway is the longest in Malaysia with the length of 772Km. The road runs from Bukit Kayu Hitam in Kedah state in the northern region of Malaysian-Thai border to southern state; Johor Bahru of Peninsular Malaysia and to Singapore. The dual carriageway expressway adheres to the design standard of rural roadway category R6 of the REAM road design standards.

In this research, only the stretch of expressway at Perak State between Jelapang and Kampung Menora is investigated as the road becomes highland, winding and dangerously cornered. In fact, the speed limit at this location of the expressway is 80km/hr compared to 110km/hr at the remaining stretch of the expressway. The length of the investigated stretch of road is about 20km of length. The following Figure 3.3 presents the stretch of road where data collection is carried out.

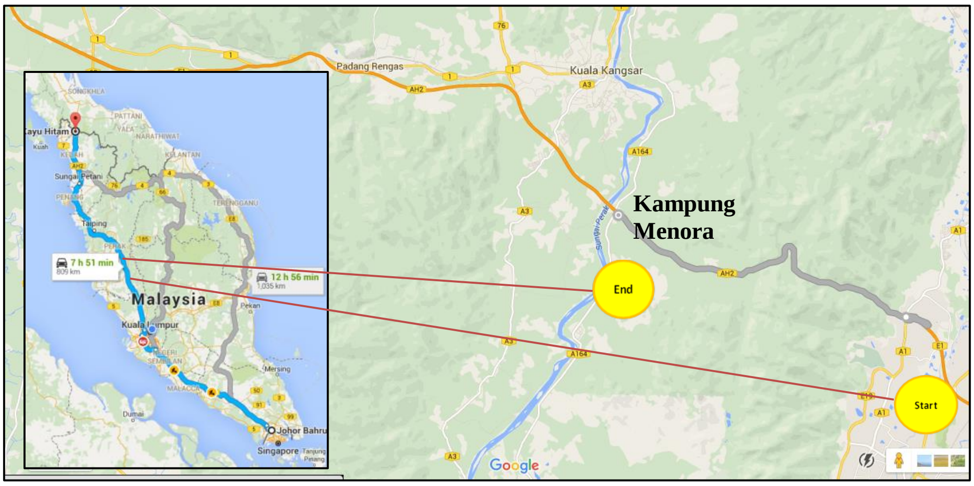

Figure 3.3: North-South expressway, Perak (Google maps b, 2015)

The data collection took four (4) days on the months of June and July 2014. In total, four (4) trips and four (4) different drivers were recorded. No more trips are accomplished for this site because of time and resource restrictions. The bus superstructure is high floor single deck bus which accommodates not more than 50 passengers and no standing passenger is allowed. Another bus type is a double-deck bus but the lower compartment is reserved for the crew and no passengers are allowed. The following Table 3.2 summarise the data for the North-South expressway.

Table 3.2: Summary of N-S expressway site Information

Data Collection Period	REAM (2002) Design Guidelines	Speed Limit (km/hr)	Data Collection (Start/End)	Length (km)	Bus Structure	Total No. of Trips	Total No. of Drivers
19, 22 Jun 2014 & 11, 13 July 2014	R6	80	Jelapang/ Kampung Menora	20	Single Deck High Floor & Double Decker	4	4

3.2.3 Jalan Batu Feringghi

The Batu Feringghi road is located at George Town in Penang state. The Jalan Batu Feringghi is a rural roadway of category R4 based on REAM design standards. The road is a single carriageway which varies between a 2lane-2way at steep climbing spots to 4lane-2way lane at some locations especially at flat non-climbing locations. The speed limit is 70 km/hr.

Moreover, as the road runs parallel to the coast, the road becomes unsignalised T-junctions at certain locations of the city area. Also, 2 main signalised intersections are located at the main hub of the Batu Feringghi area. Those intersections provide access to residential and commercial areas as well as other facilities within the vicinity.

The operator of the low floor public bus is RapidPenang. The bus allows for standing passengers with approximated full combined capacity not exceeding 50 passengers. Also, the transit bus has twelve (12) bus stops between departure and arrival terminals.

The length of the investigated stretch of road is about 10km of length. The following Figure 3.4 presents the stretch of road where data collection is carried out.

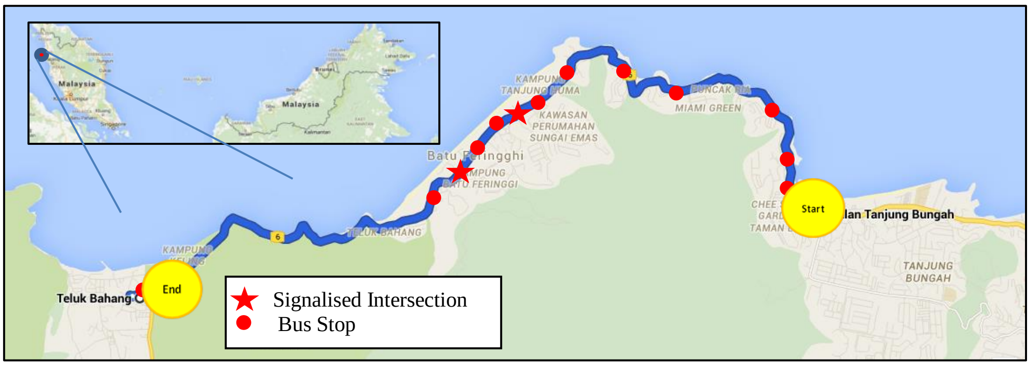

Figure 3.4: Jalan Batu Feringghi, George town, Penang state (Google maps c, 2015)

The data collection took four (4) days on the months of June and July 2014. In total, twenty seven (27) trips were accomplished with twenty two (22) different drivers. The bus superstructure is low floor single deck and with a maximum capacity of not more than 50 passenger including the standing passenger. The following Table 3.3 presents a summary of Jalan Batu Feringghi.

Table 3.3: Summary of Batu Feringghi site information

Data Collection Period	REAM (2002) Design Guidelines	Speed Limit (km/hr)	Data Collection (Start/End)	Length (km)	Bus Structure	Total No. of Trips	Total No. of Drivers
20, 21 Jun 2014 & 11, 12, 13 July 2014	R4	70	Jalan Tanjung Bungah/Teluk Bahang	10	Low Floor Single Deck	27	22

3.2.4 Karak Highway

The Karak Highway is a 60Km in length and 90Km/hr speed limit motorway connects the Federal territory of Kuala Lumpur, and the eastern state of Peninsular Malaysia; Karak town at Pahang. The expressway is dual-3lane carriageway from Kuala Lumpur to Genting Sempah Tunnel, and then narrows to dual-2 lane carriageway for the remaining stretch of the expressway. This research investigates about 30Km of the Karak highway between the Genting Sempah tunnel and the toll exit to Bentong town. This stretch of the expressway is chosen because of its unique geometric sections which involve winding lanes and steep terrain at some locations.

The expressway adheres to the design standard of rural roadway category R5 of the REAM road design standards. The length of the investigated road section is 30Km. The following Figure 3.5 presents the stretch of road where data collection is carried out.

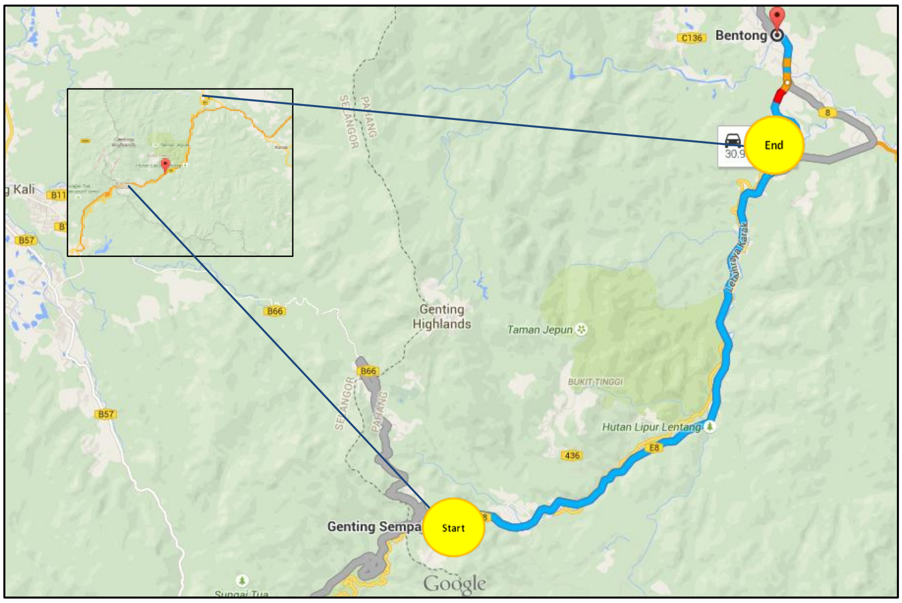

Figure 3.5: Karak Highway, Pahang state (Google maps d, 2015)

The data collection took three (3) days on the month of May 2015. In total, twelve (12) trips and seven (7) different drivers were recorded. The bus superstructure is stage high floor single deck bus and with a maximum capacity of not more than 50 passengers and no standing passenger is allowed. The following Table 3.4 presents a summary of Karak Highway.

Table 3.4: Summary of data site information

Data Collection Period	REAM (2002) Design Guidelines	Speed Limit (km/hr)	Data Collection (Start/End)	Length (km)	Bus Structure	Total No. of Trips	Total No. of Drivers
9, 15,16 May 2015	R5	90	Genting Sempah/Bentong	30	Stage	12	7

3.2.5 Route 59

The Route 59 is located at Pahang state. The route provides access from the Tapah town, Perak state to Tanah Rata town at Cameron Highland, Pahang state. The Route 59 is a rural roadway of category R5 based on REAM design standards. The road is a 2lane-2way single carriageway with steep climbing spots and winding curves. The mountainous terrain has speed limit of 70 km/hr. The length of the investigated road section is nearly 53Km and about 1.5 hours journey time.. The following Figure 3.6 presents the stretch of road where data collection is carried out.

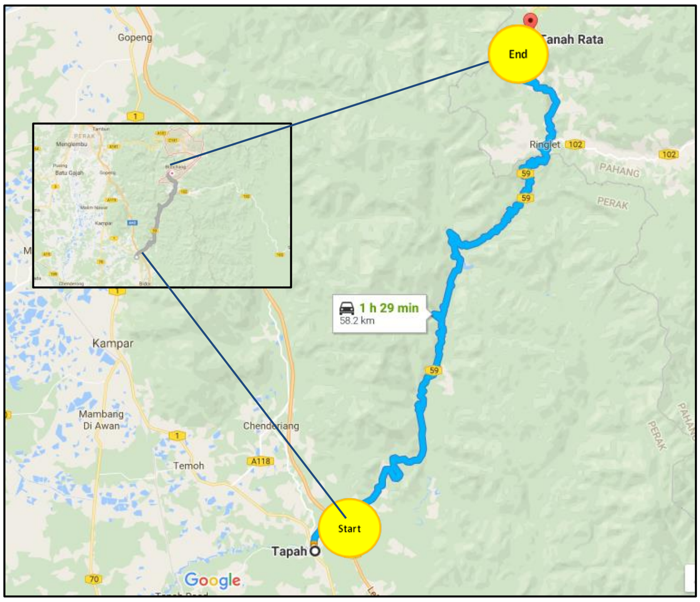

Figure 3.6: Route 59, Pahang state (Google maps e, 2015)

The data collection took three (3) days on the month of May 2015. In total, eight (8) trips and four (4) different drivers were recorded. The bus superstructure is stage high floor single deck bus and with a maximum capacity of not more than 50 passengers and no standing passenger is allowed. The following Table 3.5 presents a summary of Route 59.

Table 3.5: Summary for the sites for data collection

Data Collection Period	REAM (2002) Design Guidelines	Speed Limit (km/hr)	Data Collection (Start/End)	Length (km)	Bus Structure	Total No. of Trips	Total No. of Drivers
25, 26, 27 May 2015	R5	70	Route 59	53	Stage; Single Deck High Floor	8	4

3.3 Bus Motion and Location Measurements

The bus dynamic is measured by its three-axis acceleration (i.e. lateral, longitudinal and vertical) using the USB accelerometer of type X8M-3 Marine. It has a built-in 2 GB flash memory which records a constant stream of accelerometer data at frequency of 60Hz. The data are saved in a .csv file which is readable using the Excel spreadsheet (Accelerometer/Magnetometer Data Logger X8M-3mini, 2014)..

A GPS receiver of GMI-86 USB model is used to record the speed and positioning of the bus in NMEA file format. In certain obstructed environments such as forested areas and urban canyons, the GPS receiver might not have a clear view of the sky, therefore, the dilution of precision (DOP) can be large and position accuracy will suffer (Langely, 1999). This methodical error is encountered during the data collection, and the data were discarded from the analyses. The GPS and accelerometer are connected to a portable computer (laptop) and is laid on flat surface in the bus such as floor or on a levelled platform. The following Figure 3.7 shows equipment setup inside a bus.

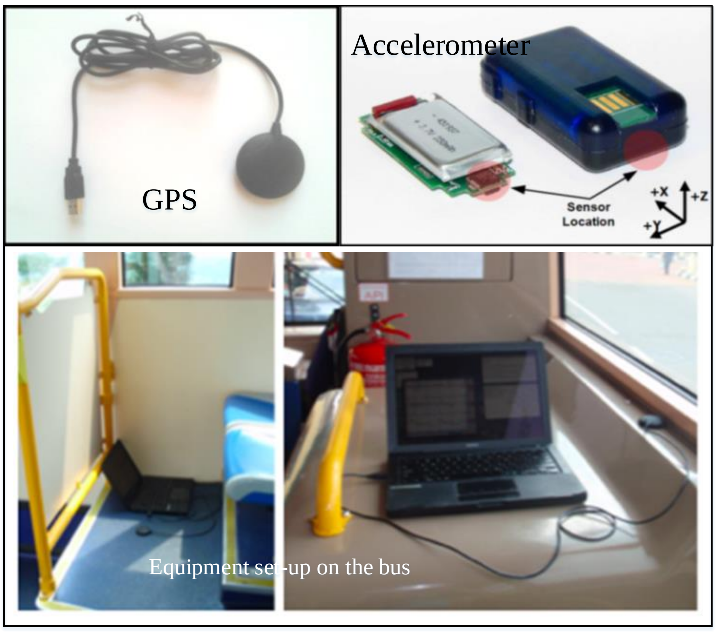

Figure 3.7: Equipment setup inside a bus

The following Figure 3.8 presents the row data from a NMEA file and when it is converted to readable spread sheet format.

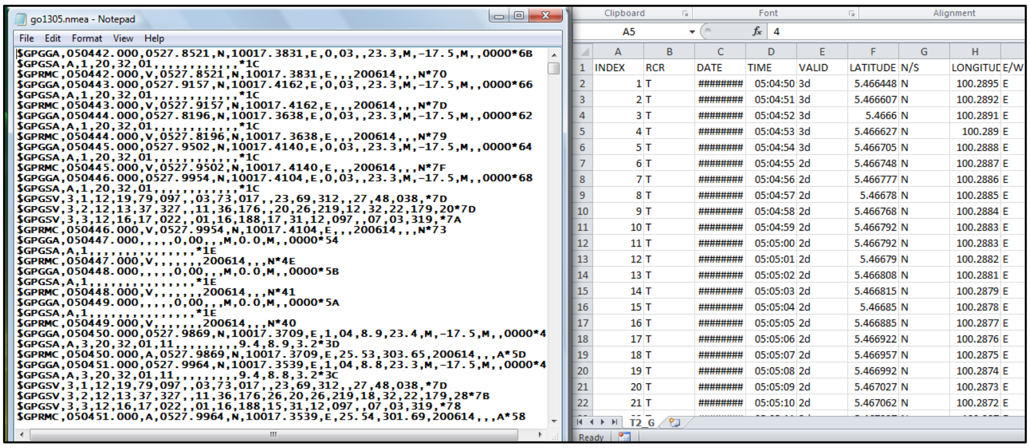

Figure 3.8: The NMEA format file before conversion (left) and the .csv file in spreadsheet format (right)

3.4 Passenger Risk Perception Measurement

A survey is used to measure of passenger risk perception. The survey is a designed intranet survey which allows its users to rate their perception of risk level on a scale of 5. The rating value of 1 presents the worst level of risk to rating value of 5 presents the highest level of safety or comfort. The survey is only accessible for a user via local communication network. The advantages of using such a surveying method include:

1. No data loss as every single rating will be transferred and saved immediately to the host server.
2. Every single rating is recorded based on time format i.e HH:MM:SS which will assist later to synchronise and match the recorded risk perception rating with the other variables collected from independent devices including GPS and accelerometer.

To establish an intranet surveying local network, two devices are needed including a laptop to act as a host server, and a mobile router. The router is used to provide the medium platform which allows participants to access the survey application in the host server; in this case a laptop. The number of survey participants depends on the type of router used. In this research a mobile router of model M5350 TP-LINK is used. This model of router accommodates up to 9 participants simultaneously. The Figure 3.9 shows the M5350 TP-LINK router.

Figure 3.9: TP-LINK Mobile Router Model M5350

A participant can access the survey address via a smartphone, a tablet or a laptop. A training session on how to access and on how to use the survey was provided for the participants. Moreover, the participants were recommended to rate their perception of the ridership as much as possible during the journey. All of the participants in this survey are university students aged between 20 to 25 years old.

Once the survey is accessed, there are four steps which need to be completed. These steps are aimed to collect data regards the passengers' name, position and seat location inside the bus, and their perception of level of risk.

The first step is to key-in the participants name as in Figure 3.10.

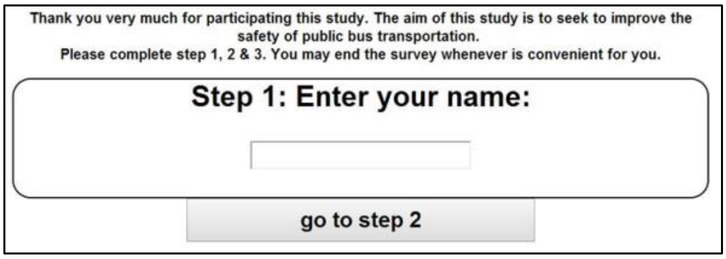

Figure 3.10: Name input, Passenger risk perception survey (Step 1)

The second step is the passenger's position i.e standing or sitting inside the bus as in Figure 3.11. This information will not be used in the analyses though it has a vital importance in rating. All of the bus trips do not allow for standing passenger, except for Jalan Batu Feringghi where some of the trips where accomplished with standing passengers.

Figure 3.11: Traveling posture, Passenger risk perception survey (Step 2)

The third step is the passenger's location in the bus where the participant can approximately choose the appropriate circle which indicates the location of the seat as in Figure 3.12. The information concerns the location of seat is not found to be parallel to the objective of data collection. The location of the seat has vital importance during an accident but has very minimal (almost void) effect on the forces which act on the commuter's body in case of buses from one seat to another.

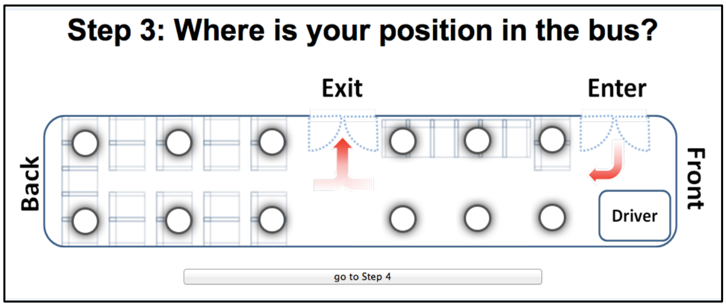

Figure 3.12: Seat location, passenger risk perception survey (Step 3)

The fourth step is the passenger's perception about the driver's driving behaviour at turning, braking and accelerating. The rating is in the scale of five units from 1 'Extremely Dangerous' to 5 'Extremely Safe'. This step can be repeated as much as the participant wants till he/she presses the ‘End Survey’ button as in Figure 3.13.

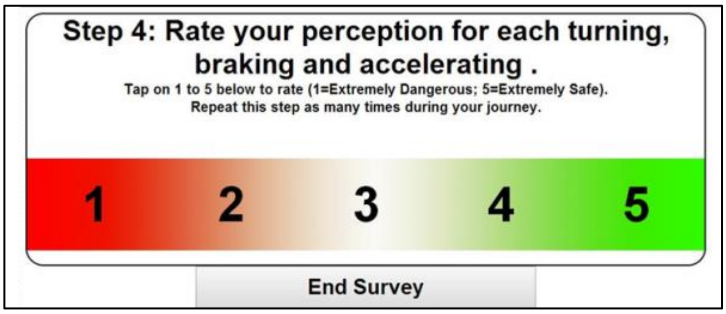

Figure 3.13: Risk perception level, Passenger risk perception survey (Step 4)

3.5 Bus Driver Characteristics

A driver should not be aware of the running data collection process throughout a journey in order to avoid any potential monopolisation in driving behaviour. At the end of the trip, a bus driver is approached to fill-in a survey. The survey included questions about (1) driver age, (2) total years of driving experience, (3) total years of driving experience at the investigated road, (4) number of working years for the bus company (5) shift working hours, (6) presence of assistant driver, and (7) driver perception of hazards at road. Refer to the APPENDIX A for a sample of the survey form.

There are 53 drivers who were surveyed in this study. Among those drivers, three (3) drivers were excluded from the analyses as the refused to answer the survey questions. These eliminated drivers include one driver from each of the following data collection roads; Jalan Batu Feringghi, Karak Highway and Route 99. Furthermore, for the trips which had a repeated driver, only one trip is used for the analyses by averaging all other trips of this particular driver. In total, there are 50 drivers who were considered for the analyses.

Total of 50 different drivers are surveyed in this research. All the subjects are male gender. 46% of the driver subjects are 41 to 50 years old, followed by elder age group of 51 to 60 years old. Young driver age groups (<40 years old) presents 28% of the studied population. The following Figure 3.14 presents the characteristics of age variable

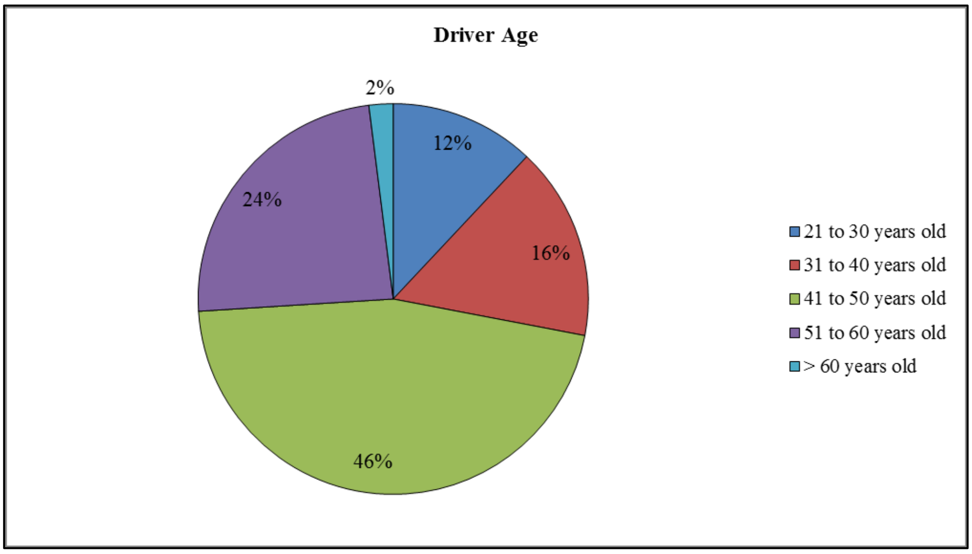

Figure 3.14: Driver age variable

Based on Figure 3.13, 28% of the driver subjects have less than 5 years of experience (junior drivers), while 24% have 6 to 10 years of driving experience. Surprisingly, drivers with more than 21 years of driving experience present a decent 18% of the driver population. The following Figure 3.15 presents the characteristics of age variable

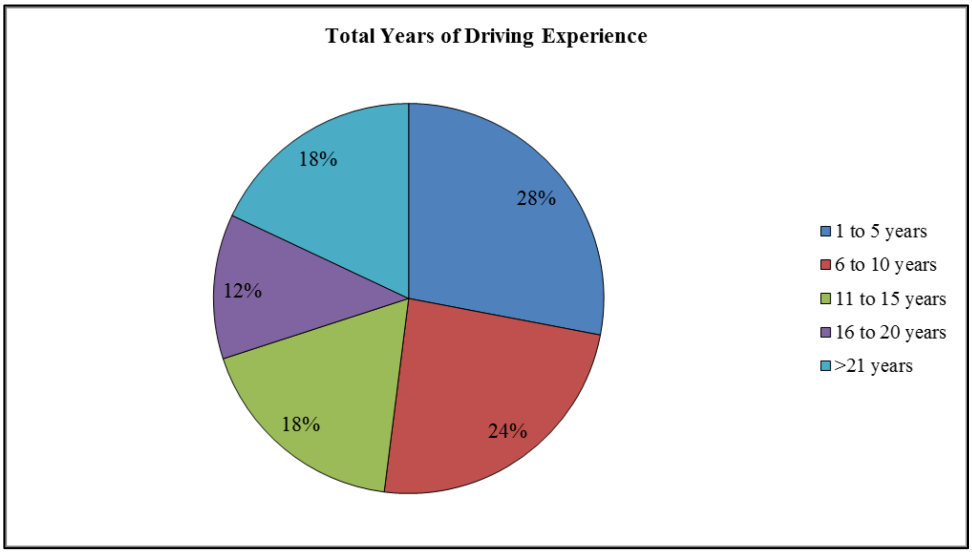

Figure 3.15: Total years of driving experience variable

Based on Figure 3.16, 51% of the drivre population has 1 to 5 years of specfic driving experience. Specific driving experience is interchangable term used for total years of driving experience at investigated road. This factor is found to be associated with accident risk as presented earlier in the literature (Tseng, 2012). The second in rank is 6 to 10 years specific experience at 25%, and the remaining 24% represent the other classes of the specific experience.

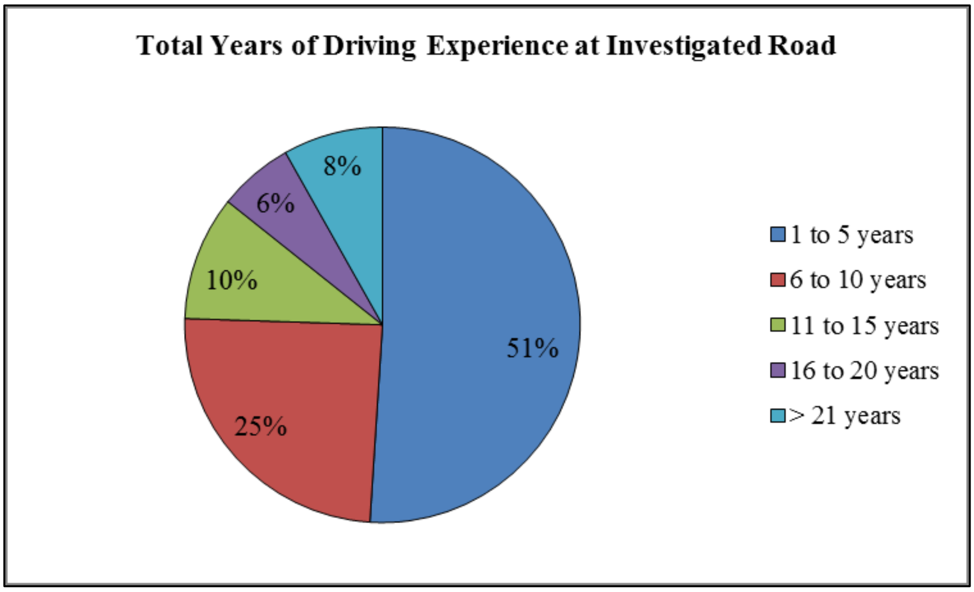

Figure 3.16: Total years of driving experience at investigated road variable

Based on Figure 3.17, 50% of the drivre population has 1 to 5 years of working for the bus company. The second in rank is 6 to 10 years specific experience at 28%, and the remaining 22% represent the other classes of the working experience for the bus company.

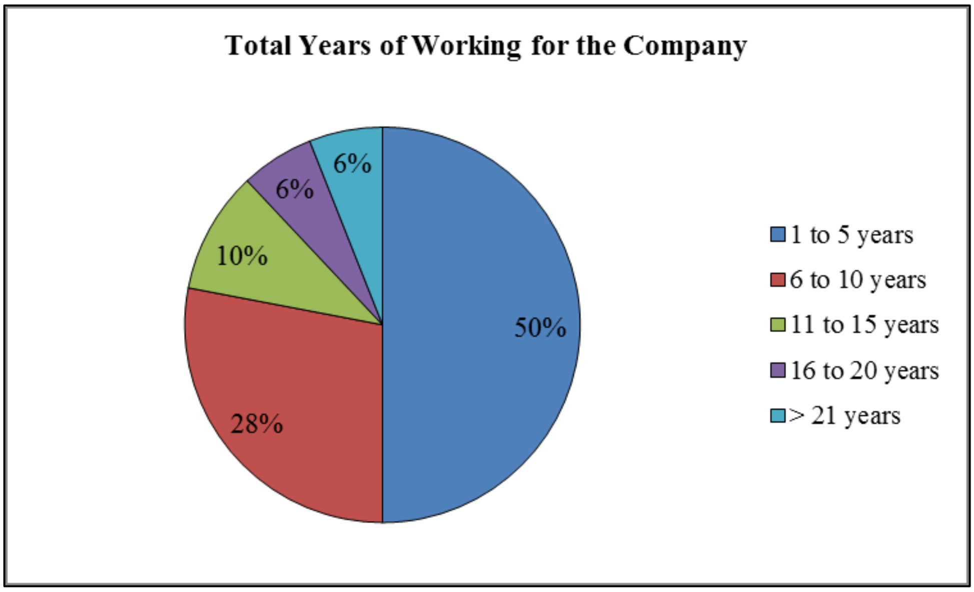

Figure 3.17: Total years of driving experience of working for the bus company variable

The survey has investigated the driver's perception about hazard at road. Based on Figure 3.18, 58% of the response was nominating the road geometry as a risk factor. This finding implies that driver perception differ from how road are designed. The second in rank is driver's behaviour at 11%, and the remaining 22% represent the other classes of the hazard factors based on driver's perception. Moreover, speeding and driver age are not considered by the bus drivers as risk factors a road.

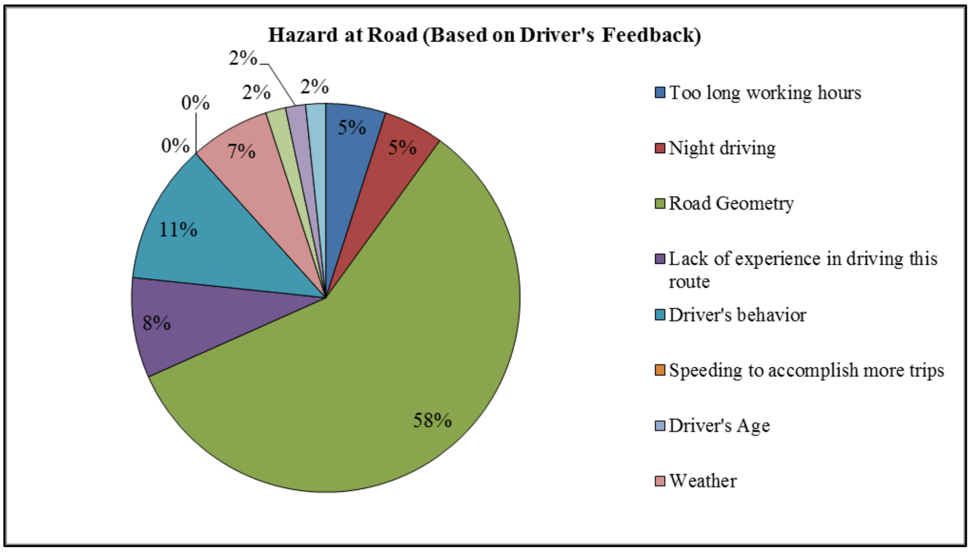

Figure 3.18: Hazard at road based on driver's perception

3.6 Vehicle Characteristics

There are three (3) types of buses covered in this thesis. The bus division is based on mechanical characteristics to low and high floors and to single and double-deck buses. According to the collected sample population, the stage bus is dominating the sample at 57% due to the nature of locations and routes. Stage buses are very common for compared to low floor buses (34%) which are normally associated with public transport for rapid transit usage. The double-deck buses (2%) share the stage bus for long haul travel distances. The following Table 3.6 justify the bus based on locations.

Table 3.6: Bus category and engine configuration for the collected data

Variable	Category	Data Site	Count	Weighted Sample
Bus Type	Low Floor Single Decker	Jalan Batu Feringghi	27	34%
	High Floor Single Decker (Stage)	B66 Jalan Batang Kali, North-South Expressway, Karak Highway & Route 59	45	57%
		Route 59	5	6%
	Low Floor Double Decker	North-South Expressway	2	3%
	Total		**79**	**100%**

The bus structure is further categorized based on factors of net weights and dimensions. These factors have direct influence on the centre of mass and the stability of the vehicle. Besides that, it is hypothesised that a passenger level of risk is related to the mentioned bus factors. The Scania bus manufacturer is used as a standard to represent other bus manufacturer. The choice of this manufacture as it is very popular manufacturer in Malaysia and is normally used as a transit city bus by Rapid transit operators namely RapidPenang and RapidKL. Nonetheless, the Cameron Highland bus is Mercedes and the model number was not applicable to be obtained even after a through looking at the world wide web, no close model was been able to obtained for this bus. The following Figure 3.19 shows the different buses used.

Figure 3.19: Bus used in Cameron Highland (top-left), Batu Feringghi (top-right), Karak Highway (bottom-left), and double-decker (bottom-right)

Based on Table 3.7, there are four (4) classes among the buses. These categorizations are based on engine configuration and dimensions. The four class classification is considered for the analyses are based on bus type.

Table 3.7: Bus classes based on engine configuration, dimension and total weight

Manufacturer	**Bus Type**	**Engine Configuration**	**Dimension (m)**						**Total Weight (Kg)**
			Wheel Base	Overhang Front	Overhang Rear	Length	Width	Height	
Scania	Low-floor single Deck (K270UB)	Rear	5.22	2.08	3.05	10.35	2.5	3.2	10,100
	High-floor Single Deck (F270IB)	Rear	6.3	1.9	3.42	11.62	2.5	3.5	13,400
	Double Deck	Rear	6	2.5	3.92	12.2	2.5	4.1	NA
Mercedes	High-floor Single Deck	Front	NA	NA	NA	NA	NA	NA	9,400

All in all, the categorisation of a bus is based on its structure and weight. The mentioned categorisation imply four types including low floor (10,100kg), stage (13,400kg), double-deck (13,400kg), and stage (9,400kg).

3.7 Identification of Accident Risk Variables

3.7.1 Passenger Risk Perception

This variable is obtained directly from the passenger survey. The rating results of five scale is averaged to rating of three to suit the accident risk category of rating 1: risk, rating 2: neutral, and rating 3: safety.

3.7.2 Driver Behavioural Factor

Out of the seven (7) question variables in the driver survey, only three (3) variables were chosen for this study including (1) drive age, (2) total years of driving experience, and (3) total years of driving experience at the investigated road. The carried out literature in chapter 2 indicates a significant relationship between driver demographics and accident risk (refer to 2.1 Factors of accident risk).

Another variables in belong to driver factor are the speed and acceleration. While speed is obtained from the GPS, the resultant acceleration is computed from the longitudinal, lateral, and vertical acceleration readings which were recorded in the accelerometer. The resultant acceleration, a_R , measures the resultant forces act on bus by combining the acceleration in three axis, as shown below:

$$a_r = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (3.1)$$

where a_x, a_y and a_z denote the lateral, longitudinal, and vertical acceleration respectively.

3.7.3 The Environmental Factor

Each investigated road is segmented to horizontal and vertical alignments sections. The horizontal alignment properties including radius of curvature and tangent sections are computed using AutoCAD (Autodesk, 2015). The map from Google Earth is extracted and insert into AutoCAD as the overlay. The radius and tangent length are then computed. This is an acceptable procedure as recommended by the Malaysian Road Safety Audit Guidebook (Malaysian Institute of Road Safety Research, 2013). Three variables are obtained from the road segmentation including tangent, radius of curvature, and grades.

The grade of the section (G_i) is computed using the elevation data collected from GPS and Theorem of Pythagoras as shown in equation 3.1.

$$G_i = \frac{e_e^i - e_s^i}{l_i} \times 100\% \quad (3.2)$$

where e_e^i and e_s^i are the start and end point elevation for section i; l_i is the length of section i . The degree of curvature (C_i) is computed as follow:

$$C_i = \tan^{-1}\left(\frac{G_i}{100}\right) \quad (3.3)$$

Despite Route 59 is not the longest among the investigated roads, there is 519 road sections associated with Route 59 because of it winding and curvy roads. Besides that, the maximum inclination and declination values are associated with Route 59. The maximum road curvature is at Jalan Batu Feringhhi with radius of curvature of 1,600m. The following Table 3.6 shows the number of segmented sections of each road.

Table 3.8: Summary of road segmentation

Road Name	Number of sections					Maximum curve radius (m)	Minimum curve radius (m)	Maximum upgrades (%)	Maximum downgrades (%)
	Tangent	Curve	Gradient		Total				
			Uphill	Downhill					
B66 Jalan Batang Kali	18	31	38	7	49	792	83	5.84	3.27
North-South Expressway	29	29	17	12	58	1,193	199	11.61	5.61
Jalan Batu Feringghi	23	49	31	30	72	1,612	16	16.28	18.18
Karak Highway	15	18	9	23	33	680	110	4.22	11.43
Route 59	145	374	303	159	519	373	14	19.5	20

3.7.4 Bus Type Factor

The bus variable will be divided to four categories based on its structure low floor, high floor single-deck (stage), double-deck, and undefined old stage bus. The last category came as a need to define this bus. Based on earlier section 3.5 the old stage bus and the new bus has a difference in superstructure and total mass. Therefore, a differentiating category is needed for both buses.

3.8 Data Processing and Filtration

This step depends mainly on the recorded time for a data in the GPS, accelerometer, and host server for the passenger risk survey. It is important to assure those three technical devices have the same timing before starting data recording. The outcome of this step is not only that all the variables from different sources can be matched into a single file but also match the data to the appropriate road section. Therefore, one master file compiles all the thirteen (13) variables.

It is important to eliminate any section which has an error or a missed data within its variables before compiling the data with the passenger's perception risk results. The accelerometer data are recorded in the three dimensions, though an error in recorded data could be not visible, thereof, the vertical acceleration is used to point out the false measurement. The vertical acceleration is associated with the gravitational acceleration of 1g. Therefore, if the recorded value for the mentioned acceleration is negative or too extreme from the 1g, the value is discarded and the associated values for the lateral and longitudinal accelerations for the same point. Such acceleration error does occur if the accelerometer device falls or flipped while recording the data.

While the accelerometer and GPS records can be matched precisely based on the recorded time, the passenger rating is basically taken within a maximum range of +2 seconds to allow for the decision and reaction time which are taken by a passenger. LOOKUP formula is designated to look up for a passenger rating at a maximum range of +2 seconds from the referenced GPS time, that is at exact reference time, reference time+1sec, and reference time +2secs to allow for the decision and reaction time. If there is more than 1 value for the rating within the given time frame, the worse rating is taken to cater for the safety. An Excel spread sheet was created to combine these data and import them to one master file.

The second step taken to filter the data is by taking the average of the data for the repetitive driver based on the direction of travel for each trip.

Finally, the master file was done by discarding the missed data for each road section. For instance, if a speed record is missed, then the associated acceleration and perception ratings are discarded. Therefore, in the analyses only a section is used when all the five (5) data values are available. The sixth data value; acceleration resultant, is computed and added for the analyses based on the filtered data.

The data from driver survey should not contradict within the overall driver feedback. There is only one case where a driver who gave incorrect statement about the years of driving experience at the Karak Highway which was at a range above the total years of driving experience. Nonetheless, in this case the recorded data were used in the analyses but not the character if the driving experience at the Karak highway. The following Table is a statistical summary of the filtered data.

Table 3.9: Statistical summary of the data after processing and filtration

Statistic	No. of Data	Minimum	Maximum	Range	Median	Mean	Variance	Standard Deviation
Mean Speed (km/hr)	2854	0.00	117	117	36	38.50	196.65	14.02
Mean Lateral Accel (g)	2854	0.000	0.524	0.524	0.126	0.134	0.005	0.072
Mean Longitudinal Accel (g)	2854	0.004	0.405	0.402	0.068	0.078	0.002	0.049
Mean Vertical Accel (g)	2854	0.000	0.319	0.319	0.026	0.032	0.001	0.026
Mean Resultant Accel (g)	2854	0.029	0.590	0.562	0.159	0.169	0.005	0.069
Mean Risk Perception Rating	2854	1.000	5.000	4.000	3.000	2.953	0.895	0.946
85th Speed	2854	16.289	117.800	101.511	40.341	43.218	183.284	13.538
85th Lateral Accel. (g)	2854	0.015	0.402	0.388	0.191	0.191	0.004	0.061
85th Longitudinal Accel. (g)	2854	0.004	0.405	0.402	0.106	0.119	0.003	0.054
85th Vertical Accel. (g)	2854	0.000	0.346	0.346	0.047	0.051	0.001	0.025
85th Resultant Accel. (g)	2854	0.037	0.590	0.553	0.217	0.225	0.003	0.059

3.8.1 Data Distribution

Normality test was conducted to evaluate the distribution of the collected data. These normality tests are carried for the bus motion variables including speed, and acceleration, and perception risk rating. Moreover, these variables were divided to their means and 85^{th} values.

In this context the hypothesis for the mentioned test includes:

H_0: The variable from which the sample was extracted follows a Normal distribution.

H_a: The variable from which the sample was extracted does not follow a Normal distribution.

Skewness is a measure for the data symmetric distribution. Based on this definition, the normal distribution has zero skewness value, and any symmetric data should have skewness near zero. Data with negative values for the skewness imply that the data are skewed left (the left tail of the distribution graph is long relative to the right tail), and data with positive values for the skewness indicate data are skewed right.(the right tail of the distribution graph is long relative to the left tail). A well-known measure coefficient for skewness is Fisher-Pearson. The following formula is used to compute the Fisher-Pearson coefficient (NIST/SEMATECH, 2012).

$$g_1 = \frac{\sum_{i=1}^{n}(y_i - \mathring{y})^3/n}{s^3} \tag{3.4}$$

where,
g_1= Fisher-Pearson skewness factor,
n= number of data points,
$\mathring{y}$=mean, and
s=standard deviation

The following Table 3.10 has the values for the skewness coefficient as computed using XLSTAT tool.

Table 3.10: Skewness values for the data

Statistic	**Average Value**					
	Speed	Acceleration				Rating
		Lateral	Longitudinal	Vertical	Resultant	
Skewness (Fisher-Pearson)	1.53	0.82	1.10	2.53	0.86	-0.14
Statistic	**85^{th} Value**					
	Speed	Acceleration				Rating
		Lateral	Longitudinal	Vertical	Resultant	
Skewness (Fisher-Pearson)	1.76	0.19	1.35	3.34	0.62	-0.70

The values for asymmetry and kurtosis between -2 and +2 are considered acceptable to prove normal univariate distribution (Trochim & Donnelly, 2006, George & Mallery, 2010, George, & Mallery, 2010, Gravetter & Wallnau, 2014). Therefore, the skewness values for the average and the 85th values indicate that the variables have normal distributions. Except for the vertical acceleration where the skewness values are 2.53 and 3.34 for average and 85th values respectively. The high value of skewness is explained by the heavy tails as the mean and standard deviation are distorted by extreme vertical acceleration values in the collected data. In fact, the vertical acceleration of 0.14 and above are associated with climbing and descending terrains. Therefore, these values cannot be considered as outliers and cannot be excluded from the data. The following Figure 3.19 supports the mentioned statement.

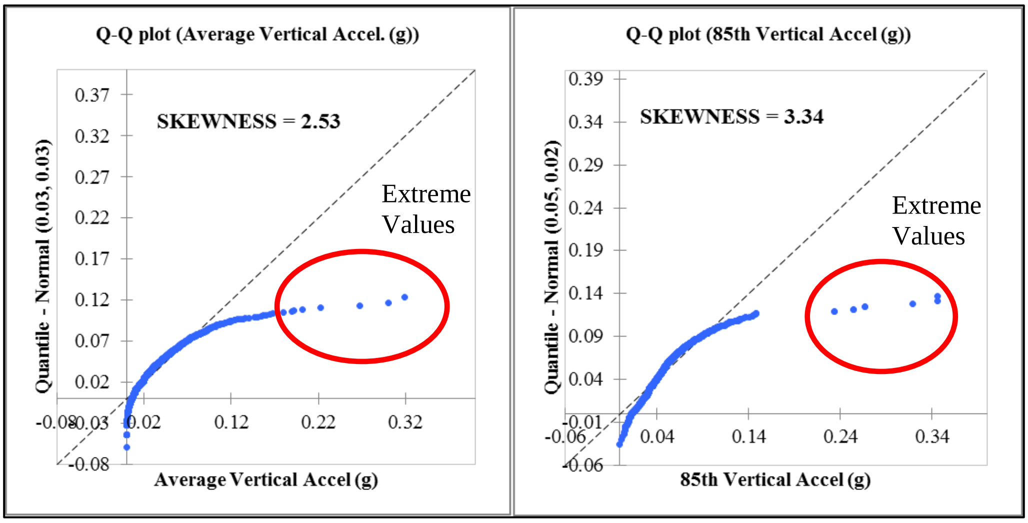

Figure 3.20: Extreme values of vertical acceleration distribution cause the positive tailing

3.8.2 Data Scaling

Numerical values were given to the continuous variables of bus type, driver characteristics, and road geometric elements. The statistical software only accepts the numerical values. Despite the bus type and driver characteristics are straight forward process of scaling, the road geometric elements are not.

The angle of inclination was divided into scale of 5 numbers. The grade is taken as below -6.5%, -6.5% to -1.6%, -1.6% to 1.6%, 1.6% to 6.5% and above 6.5%. These values correspond to the critical grade of a loaded truck of 180kg/kW and the maximum speed reduction of 15km/hr based on REAM (2012). The negative grades are indication for downgrading or descending.

The radii of horizontal curves are divided into scale of 6 with number 1 indicates very sharp curves of below 50m, and number 6 indicates large curves of 450m and above. Though literatures refer to below 100m measure of curvature as sharp curves, yet it is important to divide this class to 2 classes because there are too many short and sharp curves in this study which considers mountainous terrains. The tangent length is divided to scale of 7 with number 1 for short tangents; below 50m and number 7 for 600m and longer tangents.

The driver characteristics, bus type and road geometric elements were substitute to scale of 7 numbers as in Table 3.11.

Table 3.11: Driver characteristics and geometric elements and their equivalent scale

Scale Number	**Bus Type**	**Driver Characteristics**			**Road Geometric Elements**		
		Driving Experience (Years)	Driving at the Investigated Road (Years)	Age (Years)	Tangent (m)	Radius (m)	Grade (%)
1	Single Deck Stage	1 to 5	1 to 5	21 to 30	<=50	<=50	<=-6.5
2	Low Floor	6 to 10	6 to 10	31 to 40	50-100	50-100	-6.5--1.6
3	Double Deck	11 to 15	11 to 15	41 to 50	100-200	100-200	-1.6-1.6
4	Undefined Stage model in Cameron Highland	16 to 20	16 to 20	51 to 60	200-300	200-300	1.6-6.5
5		21 and above	21 and above	61 and above	300-450	300-450	>=6.5
6					450-600	450-600	
7					>=600		

3.9 Lamm et al. Model

Lamm et al. (1987) have developed a design consistency and safety model based on the operating speed, degree of curvature, and friction force measurements. The measurements were derived from 261 sites in New York, US (TRB, 2001). The model developed is divided into three criteria including adequate side friction, design speed and operating speed, and design consistency in horizontal alignment.

Lamm et al. (1988) defines criterion I (speed consistency) as the design speed (V_d) shall remain constant for the longer roadway section, and shall tuned at the same time with the actual driver driving behaviour, expressed by the 85th percentile speed (V_{85}).

Criterion II; dynamic driving design, is defined as a well-balanced driving dynamic sequence between and for individual design elements within a road section with the same design promotes a consistent and economic driving pattern.

Design criterion III compares the assumed side friction, f_{RA} stated in design guidelines, of circular curve with the actual friction demanded, f_{RD} at curve side. The maximum permissible assumed side friction coefficient in design is given by Lamm et al. (1988) and which is developed based on international research is as in the following equation.

For flat Terrain

$$f_{Rperm} = 0.25 - 2..04x10^{-3}x\, V_d + 0.63\, x\, 10^{-5}\, x\, (V_d)^2 \qquad (3.5)$$

And for Hilly and Mountainous Terrain

$$f_{Rperm} = 0.22 - 179x10^{-3}x\, V_d + 0.56x10^{-5}\, x\, (V_d)^2 \qquad (3.6)$$

Where,

f_{Rperm} = maximum side friction factor

V_d = design speed (km/hr)

The demanded side friction coefficient is as in the following equation.

$$f_{RD} = \frac{V^2}{127\ (R)} - e \qquad (3.7)$$

Where,

f_{RD} = demanded side friction

V= operating speed, V_{85th} (km/hr)

R= radius of curvature (m)

e = super-elevation (%)

Based on Lamm et al. (1988) all the three safety criteria are weighted equally for the design consistency and each criteria is categorised into good, fair and poor scale as in the following Table 3.12.

Table 3.12: Ranges for safety criteria I, II, and III and corresponding scale evaluation

SAFETY CRITERION		SCALE		
No.	**Equation**	**Good** (Permissible Difference)	**Fair** (Tolerated Difference)	**Poor** (Non permissible Difference)
I	$\lvert \boldsymbol{V_{85i} - V_d} \rvert$	<=10 km	10 km to 20 km	>20 km
II	$\lvert \boldsymbol{V_{85i} - V_{85i+1}} \rvert$	<=10 km	10 km to 20 km	>20 km
III	$\mathbf{f_{RA}}$ **-** $\mathbf{f_{RD}}$	≥ +0.01	+0.01 to -0.04	< -0.04

1) Related to the individual design element ‘i’ in the observer roadway segment

2) Related to two successive design elements ‘i’ and ‘i+1’ (curve to curve or independent tangent to curve)

3) V_d= design speed (km/hr)

V_{85}= operating speed (km/hr)

f_{RA} = assumed coefficient of friction as in equation

f_{RD} = demanded ceoffiecient of friciton as inequation

3.10 Pearson Correlation Test

Pearson Correlation is adopted to find the possible correlation between these variables in pairs. It is a measure of the linear relationship between a pair of variable X and Y, measuring by the correlation coefficient, $-1 \le r \le 1$ where $r = 1$ is total positive correlation, $r = 0$ is no correlation, and $r = -1$ is total negative correlation. The correlation coefficient r is computed as follow:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \tag{3.8}$$

where X_i and Y_i is the observed data i for variable X and Y respectively, $\overline{X}$ and $\overline{Y}$ is the mean of variable X and Y ; n is the total number of observation. In addition, the p-value of the correlation is observed for its statistical significance. The level of confidence level used is 95%. It means that if the p-value is lesser than 0.1, the correlation is deemed statistically significant.

The reliability of driver behaviour (both speed and acceleration) was calculated as Pearson correlations for the whole population. The correlation analysis is carried out for four groups including driver characteristic, road geometric section, bus type, and risk perception rating. The purpose is to find the significance effect (if any) between the driver behavior and every variable within the mentioned four groups.

3.11 Data Discretisation

The Bayesian network model is to be structured using discrete variables and not continuous ones. Therefore, the continuous variables including speed and acceleration should be discretised. The discretization is performed using the XLSTAT tool in Microsoft Excel is used. For discretisation, the XLSTAT tool is used. The mentioned tool is commercially available in the worldwide web. No prior knowledge is found to emphasize the appropriate intervals for the acceleration and the speed. Therefore, trials for the appropriate number of intervals are needed and best interval can be chosen after the Bayesian network is fully constructed.

The K-mean clustering method is used to divide the data to 3, 5, and 10 intervals. The comparison of these intervals is based on the Bayesian network accuracy test. Refer to Appendix C for the complete results of 3, 5 and 10 intervals and comparison.

The 10 intervals based on K-mean clustering are found to be the most appropriate for the model as in Table 3.13.

Table 3.13: The 10 interval for the motion variables based on K-mean clustering

Interval/Class	mean Speed		mean Lateral Acceleration		mean Longitudinal Acceleration		mean vertical Acceleration		mean Resultant	
	Range	Frequency	Range	Frequency	Range	Frequency	Range	Frequency	Range	Frequency
1	>31	544	<0.136	1006	<0.088	1144	0 to 0.037	1175	0.029 to 0.175	1038
2	31 to 39	624	0.136 to 0.208	537	0.088 to 0.137	434	0.037 to 0.062	448	0.175 to 0.242	508
3	39 to 47	344	0.208 to 0.254	164	0.137 to 0.170	127	0.062 to 0.084	135	0.242 to 0.283	153
4	47 to 57	194	0.254 to 0.308	75	0.170 to 0.20	61	0.084 to 0.104	42	0.283 to 0.330	80
5	57 to 68	37	0.308 to 0360	24	0.20 to 0.228	42	0.104 to 0.131	13	0.330 to 0.370	22
6	68 to 77	29	0.360 to 0.380	8	0.228 to 0.251	9	0.131 to 0.150	4	0.370 to 0.403	16
7	77 to 84	26	0.380 to 0.42	7	0.251 to 0.266	1	0.150 to 0.165	3	0.403 to 0.457	4
8	84 to 95	22	0.42 yo 0.439	1	0.266 to 0.289	4	0.165 to 0.223	2	0.457 to 0.47	1
9	95 to 96	2	0.439 to 0.524	1	0.289 to 0.321	1	0.223 to 0.3	1	0.47 to 0.557	1
10	96 and above	2	>=0.524	1	>=0.321	1	>=0.3	1	>=0.557	1

3.12 Development of Accident Risk Model using Bayesian Network (BN)

The GeNIe software is a free program which helps users to build decision theoretic models including the Bayes nets. The software which is developed by the Decision Systems Laboratory at the University of Pittsburgh is made available for the researchers and the community who are keen in developing and testing models. The advantage of this software is that the GeNIe is capable of building models of any complexity and size, the only limitation is the capacity of the operating memory of a computer (GeNIe, 2015).

The software offers a number of learning algorithms if the user has no prior knowledge on the structure of the model. Therefore, two algorithms were used to learn the model including the Bayes search and the Greedy Thick Thinning. The default software parameters were maintained. The advantage of using these learning algorithms is that the user has the option to provide a background knowledge which includes any forbidden and/or forced connection(s) to determine a particular conditional probability. In this research the background knowledge is set as to forbid the passenger rating to become a parent.

3.12.1 Training and Testing Datasets

The cross validation (rotation estimation) is used to divide the database to learning (training) set and validation (or testing set). The data is thereof divided to 75% and 25% to training and validation sets. The Matlab software is used to partition the data. For the detailed partitioning procedure in Matlab software refer to Appendix B.

To insure that data of the training and validation sets are correlated, statistical computation were performed. According to Table 3.14, the mean comparison only shows significant difference in the grade variable between both sets. Overall, it is concluded that there is a fair distribution and representation of data among both sets. Only the minimum values of mean rating, and the 85^{th} speed differ between both sets and the maximum values in case of the longitudinal and vertical accelerations.

Table 3.14: Mathematical comparison between the training and validation sets

Variable	Training						Validation					
	Mean	Variance	StdDev	Min	Max	Count	Mean	Variance	StdDev	Min	Max	Count
Age Code (Yrs)	2.83	0.75	0.87	1	5	2277	2.91	0.79	0.89	1	5	568
Tot. Driving Experience (Yrs)	2.80	1.70	1.30	1	5	2277	2.88	1.67	1.29	1	5	568
Driving Invest Route (Yrs)	1.68	1.41	1.19	1	5	2277	1.80	1.69	1.30	1	5	568
Mean Speed (km/hr)	38.51	190.87	13.82	10	115	2277	39.03	203.90	14.28	11	117	568
Mean Lateral Acceleration (g)	0.13	0.01	0.07	0.00	0.52	2277	0.13	0.01	0.07	0.00	0.49	568
Mean Longitudinal Acceleration (g)	0.08	0.00	0.05	0.00	**0.41***	2277	0.08	0.00	0.05	0.00	**0.27***	568
Mean Vertical Acceleration (g-1)	0.03	0.00	0.03	0.00	**0.32***	2277	0.03	0.00	0.02	0.00	**0.19***	568
Mean Resultant Acceleration (g)	0.17	0.00	0.07	0.03	0.59	2277	0.17	0.00	0.07	0.04	0.50	568
Mean Rating (scale out of 5)	2.94	0.90	0.95	0.02*	5.00	2277	2.98	0.93	0.96	0.15*	5.00	568
85^{th} Speed (km/hr)	43.09	179.53	13.40	**16.29***	117.80	2277	43.71	194.02	13.93	**23.29***	117.00	568
85^{th} Lateral Acceleration (g)	0.19	0.00	0.06	0.01	0.40	2277	0.19	0.00	0.06	0.03	0.40	568
85^{th} Longitudinal Acceleration (g)	0.12	0.00	0.05	0.00	0.41	2277	0.12	0.00	0.05	0.01	0.35	568
85^{th} Vertical Acceleration (g)	0.05	0.00	0.02	0.00	0.35	2277	0.05	0.00	0.03	0.00	0.35	568
85^{th} Resultant Acceleration (g)	0.22	0.00	0.06	0.04	0.59	2277	0.23	0.00	0.06	0.04	0.45	568
Tangent length (m)	185	54127	232	10	2078	2277	172	34942	186	10	1144	568
Radius of curve (m)	90	10060	100	0.00	562	2277	94	10979	104	0.00	585	568
Grade (%)	**0.26***	37.25	6.10	-20.00	19.75	2277	**0.61***	36.17	6.01	-20.00	18.45	568
85^{th} Passenger Rating (scale out of 5)	3.54	0.67	0.82	1.00	5.00	2277	3.55	0.73	0.85	1.00	5.00	568

3.12.2 Learning Algorithm and Scoring Functions

Two learning algorithms are used to lean the inference of the proposed forecasting model including Bayesian Search (BS) and Greedy Thick Thining (GTT). The scoring functions used include BDeu and K2. For inference update algorithm, three (3) algorithms are utilised including clustering, EPIS sampling and AIS sampling.

In total, 27 accident risk model based on Bayesian Network are produced, and evaluated to find the most suitable and accurate model for the collected data in this thesis.

3.13 Sensitivity Analysis

The n-way analysis serves to investigate the joint effects of inaccuracies in a set of parameters; therefore, this method is practical sound to measure the performance of the BN model. The analysis is the performed analysis via GeNIe software.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Lamm's Model for Design Consistency

The objective of using the Lamm's model is to examine the current geometric alignment of studied roads to accommodate the driver behaviour in a quantitative measure. The measure of consistency depends on comparing (1) the speed limit with the operation speed (85^{th} speed of driver population), (2) speed consistency between consequent sections, and (3) design and actual friction coefficients. The analyses are carried out for both directions of travel i.e northbound and southbound.

4.1.1 Speed Limit Consistency

The speed limit compared to the 85^{th} operational speed of based on each road segment indicate the presence of inconsistency between driver behaviour and speed limit. The 85^{th} speeds at B66 Jalan Batang Kali (62km/hr) and North-South expressway (101km/hr) are noticeably above the speed limits at 50km/hr and 80km/hr at these roads respectively. There are several factors which contribute to over-speeding of bus driver especially tight working schedules, compensation seeking upon early arrival to destinations, and risk taking and sensation seeking.

At Karak Highway the 85^{th} speed (91km/hr) is adequate and slightly above the speed limit of 90km/hr. The design is considered 'good' in respect to driver perception.

The 85th speed at Jalan Batu Feringghi (45km/hr) and Route 59 (48km/hr) are much lower than the speed limits at 70km/hr for both roads. This variation between the posted speed and actual operating (85th) speed is due to road winding and challenging steep terrain which does not allow drivers to speed up due to bus weight. Moreover, literature indicates positive relationship between accident occurrence and speed drops above 15km/hr of speed limit for heavy vehicles. The following Figure 4.1 presents the normal cumulative distribution 85th speed for the investigated roads.

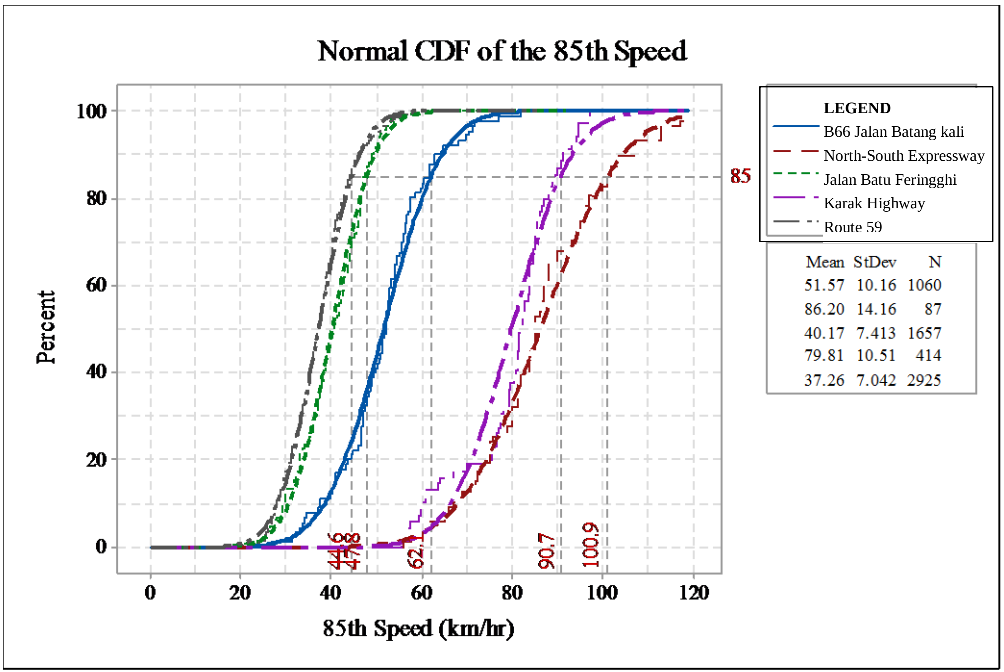

Figure 4.1: Normal cumulative distribution for the investigated roads

Moreover, the 85th speed for each section was compared to the speed limit. Evaluating each section on both direction of travel indicate that Route 59 has the highest 'poor' designed sections among other roads at 96% and 93% for southbound and northbound directions respectively. The second in rank is Jalan Batu Feringghi at 94% and 90% for northbound and southbound directions respectively. Both of these roads have 0% 'good' rated sections based on speed limit consistency criteria. In comparison, B66 Jalan Batang Kali, Karak highway and North-South expressway have higher percentage of ‘good’ and ‘fair’ sections for the mentioned criteria. The following Figure 4.2 presents those findings.

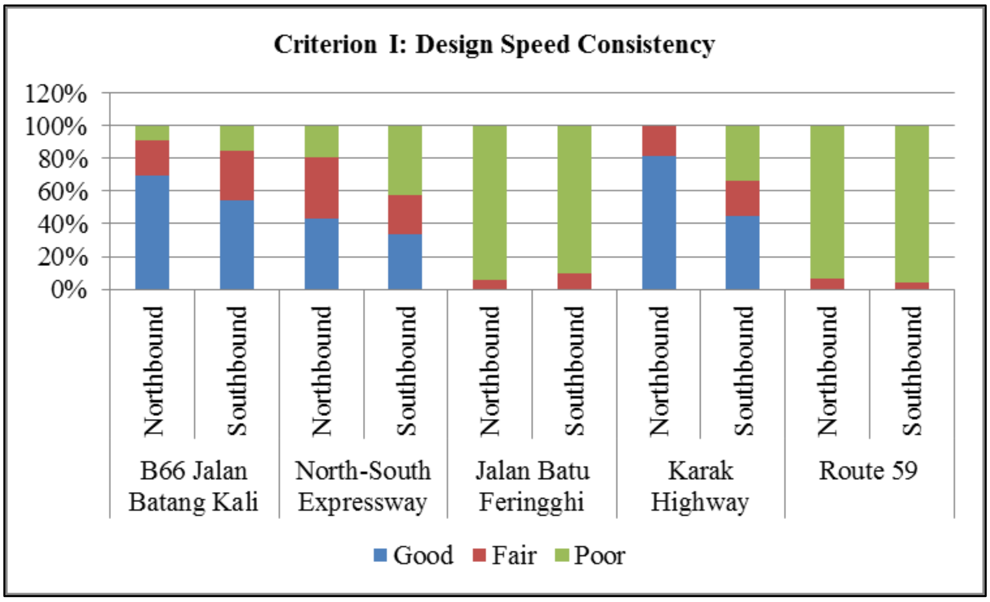

Figure 4.2: Criterion I: design speed consistency between the speed limit and 85th operational speed

4.1.2 Operating Speed Consistency

Overall, in any of the investigated roads, the transition speed between two consecutive sections is found to be consistent. A smooth transition based on speed difference from one segment to another was recorded along each road and travel direction. Minor ‘poor’ inconsistent sections are found in B66 Jalan Batang Kali, Jalan Batu Feringghi, Karak Highway and North-South Expressway. The following Figure 4.3 presents those findings.

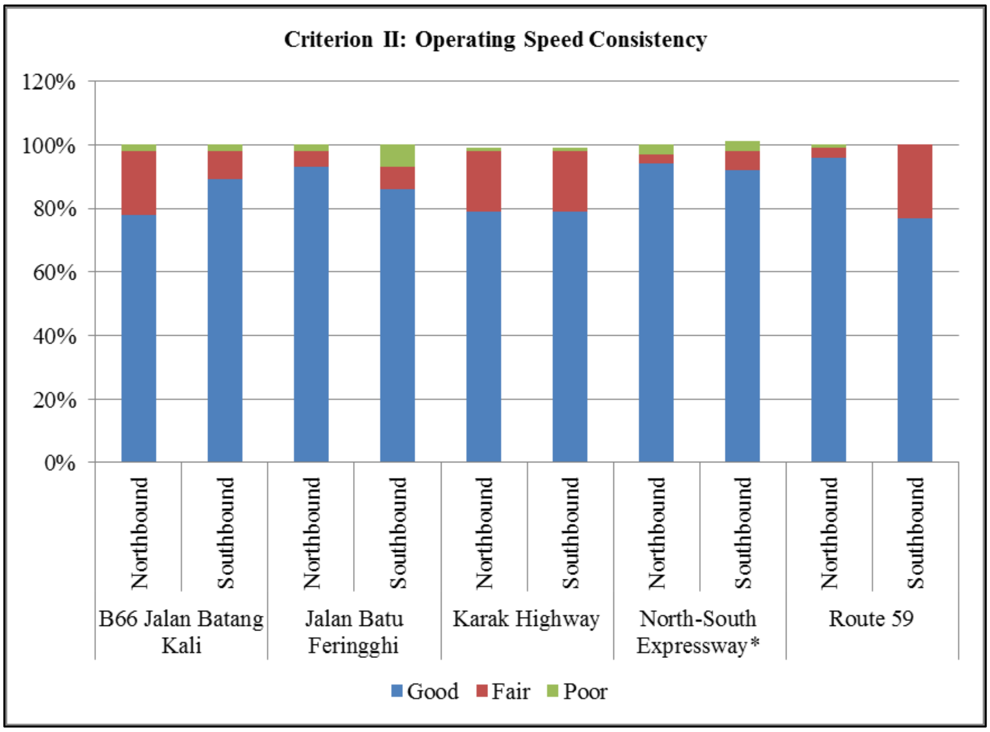

Figure 4.3: Criterion II: Operating speed consistency between subsequent sections

4.1.3 Friction Consistency

It is recommended by the Malaysian Road Geometry Design Guideline (The Road Engineering Association of Malaysia, 2002) that the side friction value considered for roadway design at a maximum of 0.10 for rural roadways. Overall, at B66 Jalan Batang Kali, Jalan Batu Feringghi and North-South expressway, the friction on section is equivalent or higher than the design friction by 0.04, thereof, the sections are ‘poor’ rated on Lamm scale. This is consistent with criteria I finding which showed that B66 Jalan Batang Kali and North-South expressway have higher speed than speed limit, hence lower frictions developed compared to the designed friction values. the bus mechanical condition and the environment play a very important role to prevent accident. Good conditions of tire and brake pedal as well as a good pavement condition are required to provide sufficient friction to the bus to prevent accident from happening. It is important to note that the radiuses of these curves are small and the speed of the bus is slow. Figure 4.4 presents the findings.

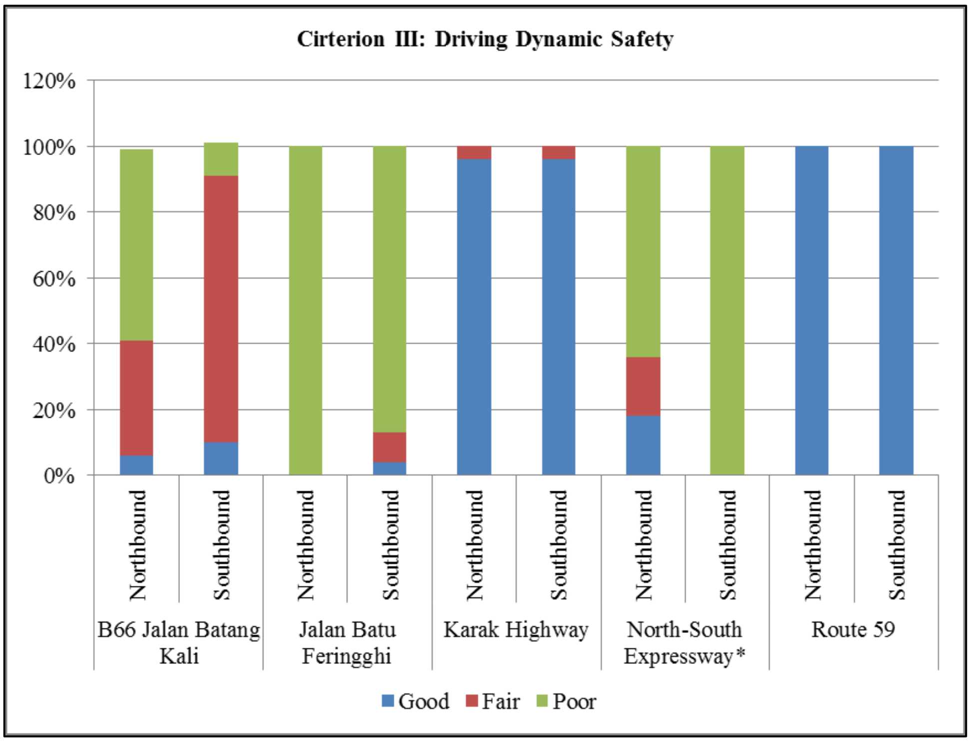

Figure 4.4: Criterion III: Driving dynamics (friction) consistency between subsequent sections

4.1.4 Model Evaluation

Based on the findings, the speed consistency between respective sections is found to be much easier to meet driver expectation. On the other hand, the operating speed tends either to be above the stated speed limit or well below the speed limits. Being below speed limit is usually associated with positive grades (climbing), the bus weight presents restrictions for a driver to speed up, and driving above the speed limit is associated with relaxed designs in terms of tangents, curves and non-steep terrains.

Lamm considers that each of the three criterions has equal quantity on the consistency evaluation. Thereof, despite that some sections do fail in only one criterion, the evaluation of that section is elevated via the other two criteria. Stating the above, a section design is considered inconsistent if at least two ‘poor’ values are measured. Moreover, it is found that such scenario is usually associated with sections which violate operational speed and friction criteria.

Based on the combination of the three criteria of consistency, it is found that 9% of B66 Jalan Batang Kali sections to be poorly designed and 3% for North-South expressway are in the same rating category. The following Figure 4.5 shows these percentages.

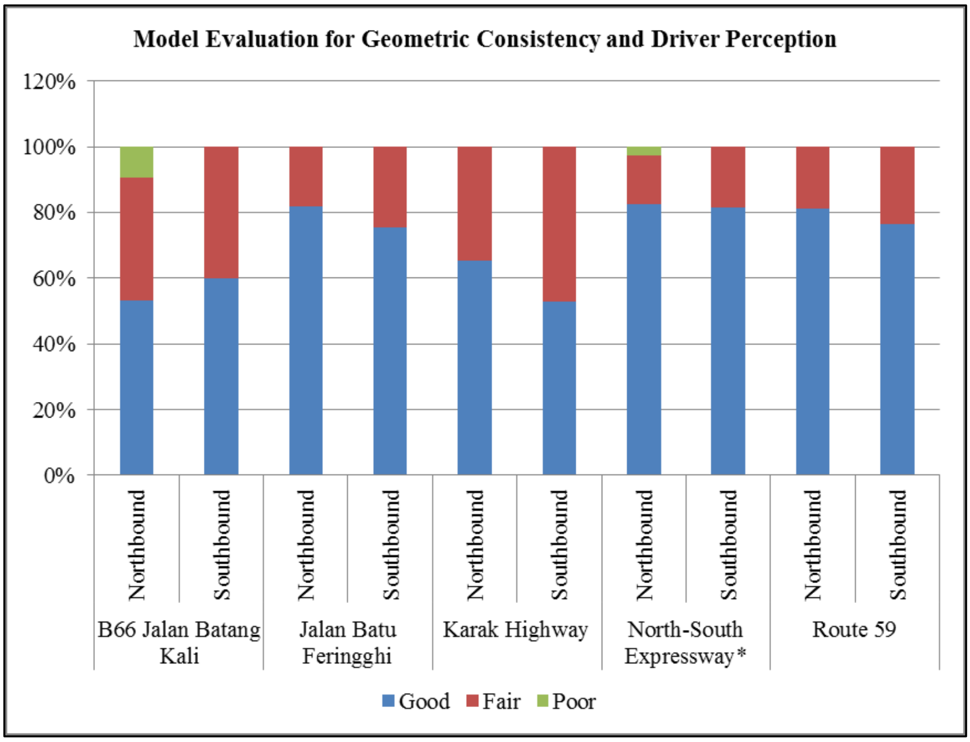

Figure 4.5: Model evaluation for geometric consistency and driver perception

In order to improve poor sections, domestic geometric standard needs to be modified to accommodate driver expectation. It is recommended to raise the maximum value for friction at rural roadways to 0.30 (similar to urban roads).

Such increment is expected to accommodate the difference in actual and assumed friction coefficients.

4.2 Pearson Correlation

Pearson correlation analysis was carried out to study the linear relationship among the variables, besides the Pearson correlation is necessary to structure a Bayesian network. The Pearson correlation is strong, moderate or weak as $r>=0.75$, $>=0.5$ or $>= 0.25$ respectively. It is of vital importance to understand that, those variables of less than 0.25 cannot be exclusively proven not to have correlation with driver behaviour. This is because Pearson's method can only identifies linear regressions, hence, other correlations or interactions could exist as well.

From Table 4.1, it can be observed that significant linear relationship exist across the variables. The Pearson correlations indicate low to strong linear relationships for value ranges 0.25 to 0.90 respectively. Moreover, for the sake of simplicity only those relationships which were found to be of linear significant importance ($>=0.25$) are presented and elaborated in the following subsections.

Table 4.1: Pearson correlation values across studied variables

Variables	Driving Invest Route	Average Speed	Tangent	Age	Grade	Tot. Driving Exper.	Average Vertical Accel.	Average Lateral Accel.	Average Longitudinal Accel.	Bus Type	Average Resultant Accel.	Radius of Curvature	Risk Rating Perceptio
Driving Invest Route	**1**	**0.136**	**0.089**	**0.575**	**-0.129**	**0.144**	-0.024	**-0.103**	**-0.163**	**-0.376**	**-0.158**	**0.074**	**0.116**
Average Speed	**0.136**	**1**	**0.432**	**0.079**	**-0.080**	-0.024	-0.024	-0.036	**-0.092**	**-0.312**	**-0.072**	**0.555**	**0.089**
Tangent	**0.089**	**0.432**	**1**	0.031	**-0.107**	0.010	-0.032	0.006	**-0.074**	**-0.353**	**-0.040**	**0.655**	**0.292**
Age	**0.575**	**0.079**	0.031	**1**	**-0.065**	**0.500**	0.017	**-0.108**	**0.059**	**0.056**	**-0.071**	0.034	**0.077**
Grade	**-0.129**	**-0.080**	**-0.107**	**-0.065**	**1**	0.003	-0.015	0.013	-0.015	**0.074**	0.000	**-0.106**	**-0.048**
Tot. Driving Exper.	**0.144**	-0.024	0.010	**0.500**	0.003	**1**	**0.138**	-0.020	**0.340**	**0.364**	**0.139**	-0.021	**-0.125**
Average Vertical Accel.	-0.024	-0.024	-0.032	0.017	-0.015	**0.138**	**1**	**0.223**	**0.230**	**0.085**	**0.415**	-0.038	**-0.134**
Average Lateral Accel.	**-0.103**	-0.036	0.006	**-0.108**	0.013	-0.020	**0.223**	**1**	-0.024	-0.030	**0.874**	-0.018	**-0.062**
Average Longitudinal Accel.	**-0.163**	**-0.092**	**-0.074**	**0.059**	-0.015	**0.340**	**0.230**	-0.024	**1**	**0.475**	**0.419**	**-0.084**	**-0.062**
Bus Type	**-0.376**	**-0.312**	**-0.353**	**0.056**	**0.074**	**0.364**	**0.085**	-0.030	**0.475**	**1**	**0.194**	**-0.306**	**-0.202**
Average Resultant Accel.	**-0.158**	**-0.072**	**-0.040**	**-0.071**	0.000	**0.139**	**0.415**	**0.874**	**0.419**	**0.194**	**1**	**-0.059**	**-0.149**
Radius of Curvature	**0.074**	**0.555**	**0.655**	0.034	**-0.106**	-0.021	-0.038	-0.018	**-0.084**	**-0.306**	**-0.059**	**1**	**0.219**
Risk Perception Rating	**0.116**	**0.089**	**0.292**	**0.077**	**-0.048**	**-0.125**	**-0.134**	**-0.062**	**-0.062**	**-0.202**	**-0.149**	**0.219**	**1**

Values in bold are different from 0 with a significance level alpha=0.05

The driver characteristic of age is positively correlated to (1) driving at investigated road, and (2) total years of driving experience are at values of 0.575 and 0.500 respectively. The moderate significance exists is intuitive and logic as an increment in driving experience is an increment in driver age. On the other hand, driving at investigated road and total years of driving experience do not have a linear relationship at Pearson value of 0.144. The total years of driving experience of a bus driver does not determine the driver's year of driving experience at a particular road. The driver's driving experience at particulate road is a factor of the working condition and assigned working routes. While transit bus is used for short haul trips, stage bus is utilised for long haul trips. Thereof, the bus type has almost equal yet opposite effect on driving years of experience variables. The Pearson value are -0.376 and 0.364 for bus type-driving at investigated road, and bus type-total years of driving experiences respectively. The following Figure 4.6 shows the mean value of both experience variables namely total years of driving experience and driving at investigated road.

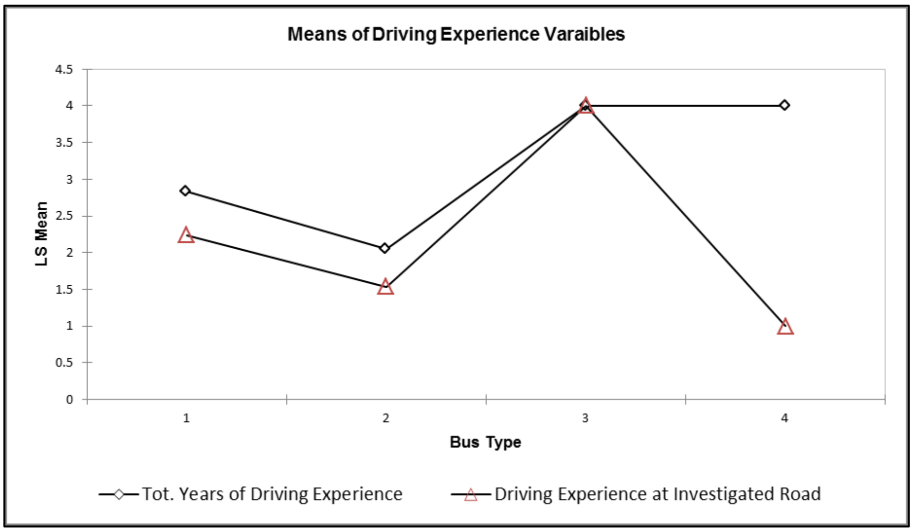

Figure 4.6: Mean value of total years of driving experience and driving at investigated road based on bus type

Overall, the transit bus population drivers have lower means of both categories of experience variables compared to the long haul stage bus (Figure 4.1). The younger driver population for the transit bus contribute to this finding. The collected data is random; hence, it is suitable to assume that the driver characteristic distribution results present the actual population. Statistically, it is difficult to explain why such distributions exist; the experienced drivers tend to work for long haul; compared to young drivers who have higher presence in transit buses. The double deck and stage (9,400kg) had one bus driver who was surveyed, the mean difference in this both categories do not present the actual population. The stage (9,400kg) bus at Rout 59 had the same drivers throughout the conducted seven trips. The double deck was used only once in the collected data. Resources and time restriction prevented from further data collection for this bus type.

The gap between the total years of driving experience to driving at investigated road is smaller for the transit short haul bus at 0.51, compared to 0.60 for stage bus. While stage bus drivers have a fixed schedule and routes, the stage bus drivers work at different trip roads which are assigned to drivers based on the company’s demand.

Another significant linear correlation is found between the bus type and longitudinal acceleration at Pearson value of 0.475. Moreover, this finding emphasises that driver behaviour can be much influenced by the driving task and environment conditions. The longitudinal acceleration is a product of the driving task. The bus function, working condition and surrounding factor influence the driver behaviour which can be monitored in longitudinal acceleration. The low floor bus is associated with frequent stops and short haul travels (500m between each bus stop) putting into account factors of (1) surrounding traffic at arterial roads compared to highways, (2) standing passengers, (3) interaction with embarking and disembarking passengers among other factors influence and (to certain extend) control the driver acceleration behaviour. In support to this statement are the presented mean difference values of longitudinal acceleration across different bus function and types. Low floor bus is associated with low longitudinal acceleration values compared to long haul buses as.in Figure 4.7.

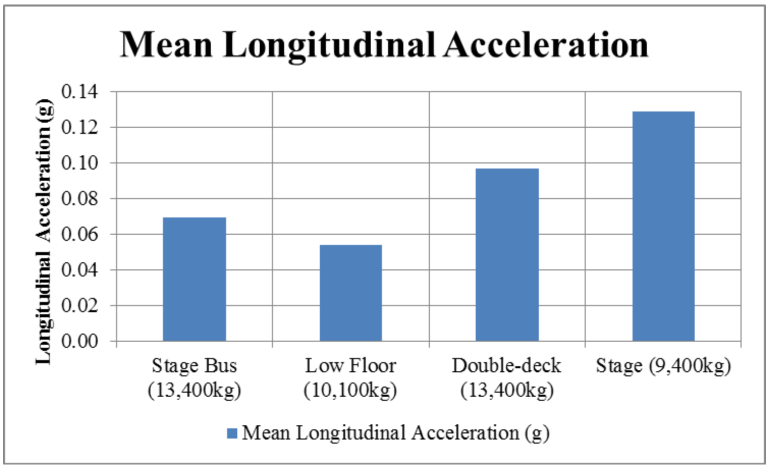

Figure 4.7: The difference in mean for stage bus (13,400kg) for different bus type

Despite the stage bus driver shows more aggressive longitudinal acceleration behaviour, there is a clear difference between the two types of stage buses based on net weight (13,400kg and 9,400kg). The 13,400kg stage bus is used in trips related to mountainous terrain including Route 59 and B66-Jalan Batang Kali, and flat terrain including North-South expressway and Karak Highway. Separating the data based on terrain type, had revealed that the road type significantly affect the longitudinal acceleration for the same bus type. The difference is significant at 0.01g between the flat and mountainous terrain.

The following Figure 4.8 show the mean values for the stage bus (13,400kg) on different terrain.

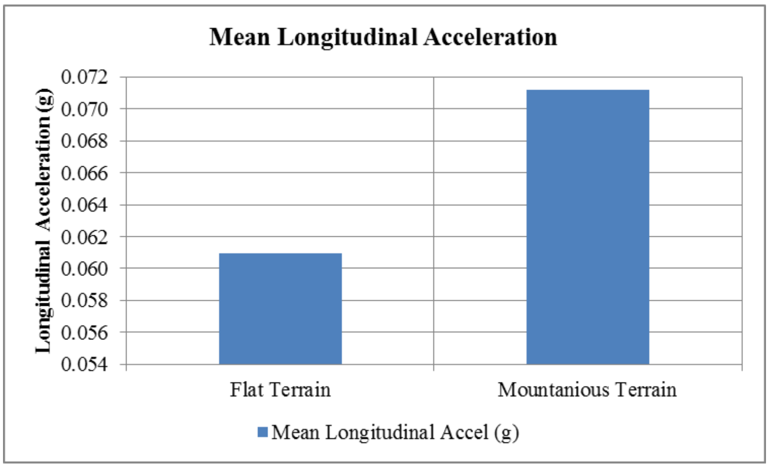

Figure 4.8: The difference in mean for stage bus (13,400kg) for different road type

Average speed and bus type are inversely proportional at -0.312 on Pearson scale. The speed variable depends mainly on the function of the bus and road terrain. For example, the long haul stage bus (13,400kg) is used in flat terrains including North-South expressway, and Karak Highway, and mountainous terrains including R66-Batang Kali and Route 59. The mean difference between the mentioned stage bus and the other long haul double-deck bus is about 37km/hr. This difference in mean speed is associated with the higher collected data frequency for mountainous terrains compared flat terrain. When categorising the data based on terrain type, the stage and double deck buses become almost equivalent in mean speed. The same finding is for the Route 59 when comparing the rear engine stage bus (13,400kg) and front engine stage bus (9,400 kg). Though both are different in net mass and engine configuration, seat arrangement, capacity, structural model and technological advancement, the mean speed is insignificant at 1km/hr. The following Table 4.2 presents these findings.

Table 4.2: Bus motion variables for different geometric terrain roads for the same type of bus

Bus Type 1	Mean Speed (km/hr)	Mean Lateral Accel (g)	Mean Longitudinal Accel (g)	Mean Vertical Accel (g)	Mean Resultant Accel (g)	Mean Risk Perception Rating
Expressway and Highway	76.03	0.10	0.06	0.03	0.13	3.01
Mountainous Terrain	38.94	0.14	0.07	0.03	0.17	3.00

The low floor bus is used as a transit bus possesses the lowest mean speed because of its frequent stops as a short-haul trip. Despite the findings indicate the significance of bus type of driving speed, the bus type in this case manifests itself as a representative variable of road terrain and trip type.

The average speed is positively proportional to radius of curvature and tangent at Pearson correlation of 0.655 and 0.432. In comparison, the grade is non-linear associated with the speed at -0.107. It is observed that drivers tend to drive at a higher speed when the radius of curvatures is more than 600m. A large radius curvature behaves as tangents, meaning that a driver can maintain cruise speed. The radius of curvature and tangent length are linearly significant at 0.655 can further support the mentioned statement. A larger radius of curve requires less cornering effort and skill which allow the drivers to drive at a higher speed.

The resultant acceleration is a function of the 3-axe accelerations; a change in any of acceleration factors leads to a change in the resultant acceleration. Moreover, the resultant acceleration is twice linearly significant to lateral acceleration at 0.817 compared to longitudinal and vertical accelerations at 0.419 and 0.415. The small change in lateral acceleration could result in higher change in resultant acceleration compared to other acceleration variables.

Furthermore, based on Pearson scale, the passenger perception rating is not found to be linearly correlated to risk variables except the tangent. The Pearson correlation is 0.292 for tangent and risk perception rating. This result implies that the risk perception of an occupant and accident risk is rather a situation which possibly occurs due to interaction of different variables.

4.3 Performance Evaluation of the BN Accident Risk Model

In all evaluated models, it is found that the inference update algorithms including Clustering, EPIS Sampling and AIS Sampling do not have an impact of the model’s number of edges (dependences) and evaluation parameters i.e accuracy, precision, false positive rate and true negative rates. Therefore, the following comparisons are based on inference learning algorithm and scoring functions. To validate these models the confusion matrix is used. The results are summarised and presented in the following Figure 4.9.

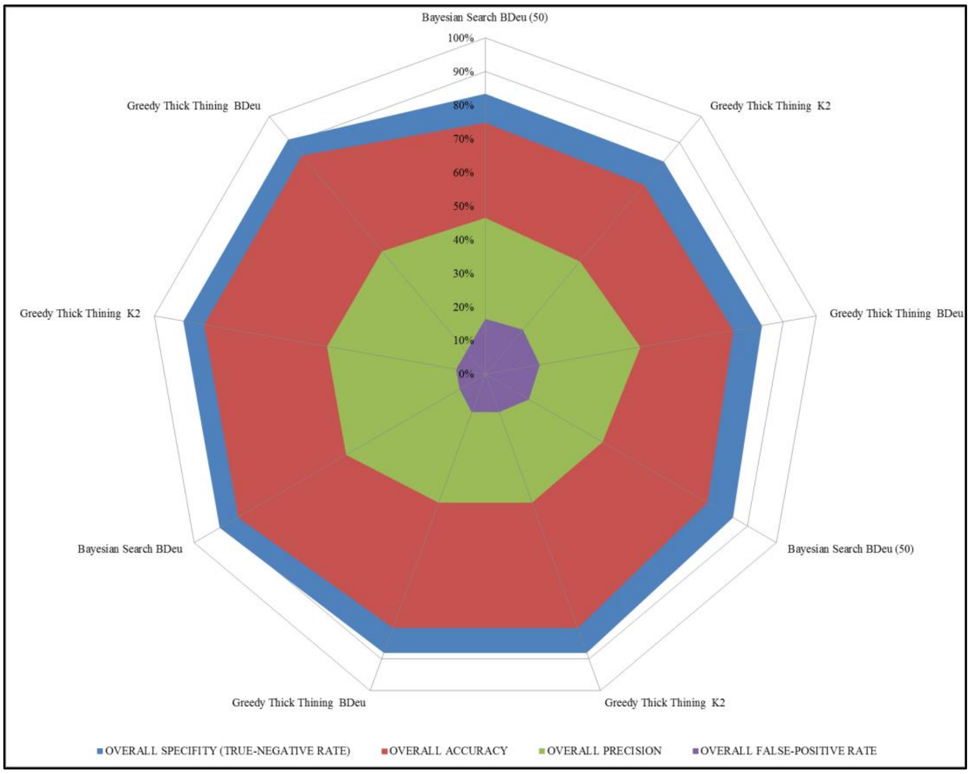

Figure 4.9: Score evaluation of the risk model based on different combinations of inference algorithms and scoring functions

According to Figure 4.9 the structural algorithm of Greedy Thick Thinning (GTT) has a better score in terms of accuracy, precision, specifity and false positive rate compared to Bayesian Search (BS) algorithm. Despite there is no difference in the performance of both scoring function (BDeu and K2) for GTT algorithm, the BDeu scoring function is chosen for the developed model as it distinguishes the least number of dependences among the variables. The BDeu scoring function has eighteen (18) edges or arrows compared to twenty four (24) edges for the K2 algorithm. Moreover, the higher the number of interval classes for continuous variables, the better the performance of the model. In this context the 10 interval class model is integrated in the final GTT model as well.

The following Table 4.3 presents a comparison summary of the accident risk model characteristics based on number of classes for continues variables, number of dependences in the developed model, and the performance evaluation. For full comparison of the 28 models refer to Appendix C.

Table 4.3: The summary of 3 BN models based on K-means clustering and Greedy Thick Thinning algorithm

Discretization Method	k-means clustering		
No. of Intervals	3	5	10
Learning Algorithm	Greedy Thick Thinning		
Scoring Function	BDeu		
Update Algorithm	Clustering		
Inference Algorithm	Policy Evaluation		
Network Properties			
Node count	13	13	13
Arcs (Edges)	22	21	20
Validation			
Validation Test	Leave-one out		
Model Checking (Confusion Matrix)			
Accuracy	74.87%	80.16%	84.92%
Recall (Precision)	46.85%	40.56%	47.80%
False-positive Rate	16.45%	11.91%	8.81%
Specifity (True-negative Rate)	83.55%	88.09%	91.19%

Based on Table 4.3, the Greedy Thick Thinning model of 10 intervals has a high accuracy at 85% compared to five and three intervals. The higher the number of intervals, the data is nearer to its actual distribution. On the other hand, the precision is almost 48%, meaning the model is capable of recalling 48% of the true answers and minimal false answers at 9% during the validation stage. There is no comparison basis is made for this results to be compared with other models. In fact, even empirical generalisation cannot be made because model fit may vary from one database to the other (Mannering and Bhat, 2014).

4.4 Accuracy of the Accident Risk Model

The Mean Absolute Percentage Error (MAPE) is employed to examine the accuracy of the forecasting model. The focus should be on the parent nodes; which have a direct impact on the passenger's rating. A measure draw packs in using the MAPE measure for accuracy checking is when the actual value of a parameter is zero as no division by zero. This disadvantage is encountered during this stage of the analyses. Some of the data are related to North-South expressway and Route 59. The double-deck bus (bus type state 3) was only used once thereof, only one driver data is available, and the local bus (bus type state. 3) used in Route 59 was also a one driver data, though there were a few trips accomplished but with the same driver. Another single case combination is missed is for a low floor bus with 16 to 20 years of driving experience at Jalan Batu Feringghi. The collection of data is random such a specific cases were not put into account during the data collection.

In total out of the 20 combination case scenarios, only eleven (11) cases were tested and evaluated as in Table 4.4. The other nine (9) cases were not validated as no such cases are matched in the validation set in particular and the whole dataset in general. Meanwhile, the Bayesian network forecast these events with equal predictions, this indicates the applicability of the BN to evaluate unobserved event

Table 4.4: Mean Absolute Percentage Error (MAPE) for the model accuracy

			Passenger Rating									Tot. Number of Validation Data
			1			2			3			
No	Bus Type	Driving at Investigated Road	Validated	Predicted Bayesian	Relative Absolute Error	Validated	Predicted Bayesian	Relative Absolute Error	Validated	Predicted Bayesian	Relative Absolute Error	
1	1	1	0.452	0.304	0.328	0.333	0.256	0.232	0.214	0.440	0.513	466
2	1	2	0.111	0.025	0.775	0.296	0.042	0.858	0.593	0.933	0.365	
3	1	3	0.136	0.071	0.479	0.455	0.141	0.690	0.409	0.788	0.481	
4	1	4	0.000	0.053	na	1.000	0.053	0.947	0.000	0.893	1.000	
5	1	5	0.367	0.078	0.788	0.408	0.404	0.010	0.224	0.518	0.567	
6	2	1	0.174	0.007	0.960	0.663	0.190	0.713	0.163	0.803	0.797	
7	2	2	0.089	0.001	0.989	0.768	0.093	0.879	0.143	0.907	0.842	
8	2	3	0.250	0.001	0.996	0.750	0.171	0.772	0.000	0.829	1.000	
9	2	4	Na	0.333	na	na	0.333	na	na	0.333	na	
10	2	5	0.700	0.118	0.831	0.200	0.558	1.790	0.100	0.324	0.691	
11	3	1	na	0.333	Na	na	0.333	na	na	0.333	na	
12	3	2	na	0.333	Na	na	0.333	na	na	0.333	na	
13	3	3	na	0.333	na	na	0.333	na	na	0.333	na	
14	3	4	0.000	0.003	na	0.000	0.003	na	1.000	0.994	0.006	
15	3	5	Na	0.333	na	na	0.333	na	na	0.333	na	
16	4	1	0.661	0.234	0.646	0.243	0.401	0.647	0.096	0.365	0.738	
17	4	2	Na	0.333	na	na	0.333	na	na	0.333	na	
18	4	3	na	0.333	na	na	0.333	na	na	0.333	na	
19	4	4	na	0.333	na	na	0.333	na	na	0.333	na	
20	4	5	na	0.333	na	na	0.333	na	na	0.333	na	

The Mean Absolute Percentage Error (MAPE) is computed using the following equation

$$MAPE = \frac{1}{N}\sum_{i-1}^{N}\left|\frac{A_i - P_i}{A_i}\right| \tag{4.1}$$

Where,

$\left|\frac{A_i - P_i}{A_i}\right|$ = Relative Absolute Error

$$MAPE = \frac{1}{466} X\ (6.79 + 7.536 + 7.00)\ \mathrm{X100}$$

= 4.58%

The MAPE value stands at 4.58% which is relatively small and is considered acceptable for this probabilistic model.

4.5 Quantification of Accident Risk Variables

The network consists of 13 nodes which fall under 5 groups including driver characteristics, motion variables, bus type, geometric elements and risk perception. The edges show the casual relationships among and across group variables. The link strength or arc weight is defined for specific edge and measure the strength of connection only along that single connection (Elbert-Uphoff, 2009). The link strength is presented as an Euclidean value. The Euclidean value is used to measure the symmetry of the overall difference between two distributions by identifying the significance of a node on another node. The main purpose is the visualisation of the strength of an edge connection in the learnt Bayesian network in order to learn more about the inherited properties of the system. The Euclidean value ranges between 0 to 1, those values nearer to 0 or 1 are more significant. The following Figure 4.10

shows the inference of the constructed model and influence of strength variables based on Euclidean measure.

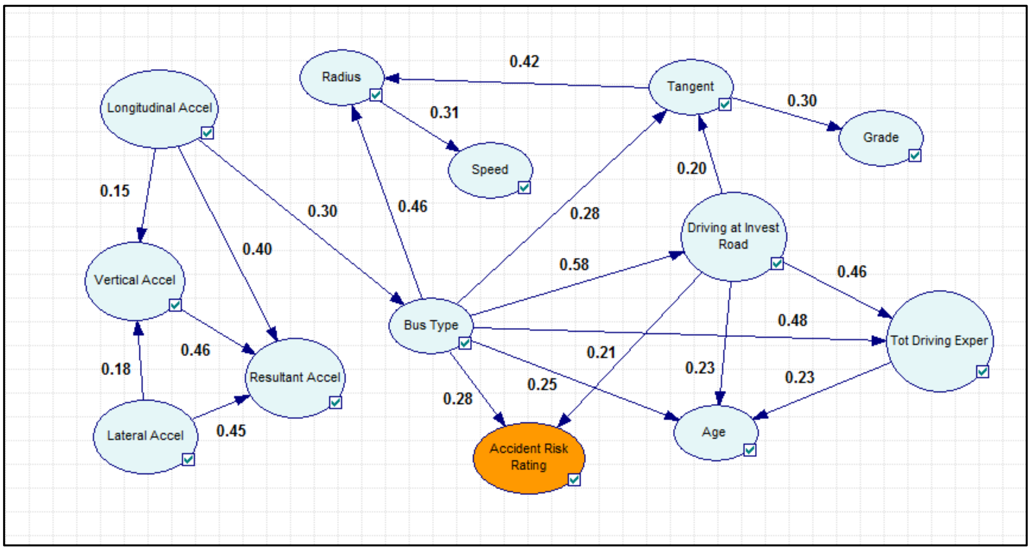

Figure 4.10: The structure of the Bayesian network based on GTT and BDeu algorithm

4.5.1 Driver Behaviour Factors

Despite the vertical acceleration is seemed to have the highest strength influence on resultant acceleration at 0.46, the lateral acceleration is the most influential factor on the resultant acceleration at 0.45 value. This is because that the vertical acceleration variable in the network is a converging node from both lateral and longitudinal acceleration variables (parent nodes). Thereof, any influence of the vertical acceleration is inherited from its mentioned parent nodes. The difference in strength influence values between lateral and longitudinal accelerations on vertical acceleration are 0.18 and 0.15 respectively. The small difference of 0.03 is caused by the difference in the distribution values of parent nodes. The interval range of lateral acceleration is 0.52g compared to longitudinal acceleration 0.41g as in Table 4.5.

Table 4.5: Statistics of acceleration variables

Variable	Range (g)	Mean (g)	Std. deviation
Average Lateral Acceleration	0.52	0.13	0.07
Average Longitudinal Acceleration	0.41	0.08	0.05
Average Vertical Acceleration	0.32	0.03	0.03
Average Resultant Acceleration	0.56	0.17	0.07

Based on following 3D Figure 4.11 the resultant acceleration distribution has similar probabilistic distribution across different ranges of vertical acceleration intervals. The Figure consists of four vertical acceleration ranges including (1) 0.00g to 0.06g, (2) 0.06g to 0.08g, (3) 0.08g to 0.10g, and (4) 0.10g to 0.30g. The small variation in vertical motion contributes less on the resultant acceleration. Most of the forecasted resultant acceleration values are less than or equal to 0.11g. The forecasted resultant acceleration value is most sensitive to small changes in acceleration parameter values (a_z<=0.08g, a_x<=0.14g, a_y<=0.17g). Nonetheless, extreme values of resultant acceleration (>=0.67g) are significantly associated with increment in lateral acceleration and longitudinal acceleration simultaneously. In comparison, the higher influence value of lateral acceleration on vertical acceleration and direct influence on the resultant acceleration cause the lateral acceleration node to have the higher impact on resultant acceleration compared to longitudinal acceleration.

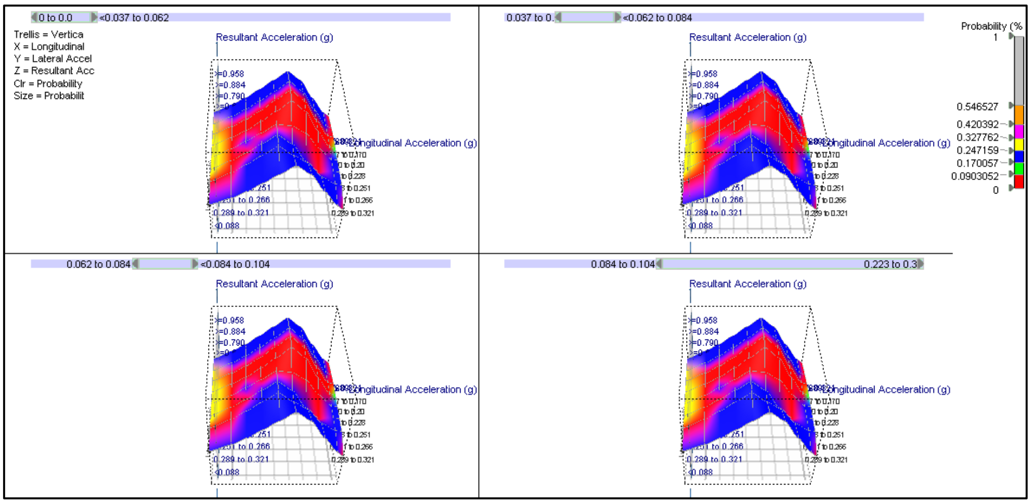

Figure 4.11: Resultant acceleration probabilistic distribution based on lateral, longitudinal and vertical acceleration values

The speed variable does not affect any other variables in the model. Moreover, the model could not connect the speed as a parent variable for accelerations, though these variables are computationally proven to be related to each other. The acceleration is a measure for speed change over a time span. Perhaps as the model does not have a time variable, such relationship between speed and acceleration are not discovered by the inference algorithm GTT. On the other hand, the model successfully distinguishes the speed as an influenced variable by the radius of curvature. Many author works have identified the relation of speed and radius of curvature (De Oña and Garach, 2012, Hassan and Sarhan, 2012, Kaplan and Prato, 2012).

The low speed is highly associated with shorter radius of curvatures and the speed value increases significantly with higher radius of curvature. The sensitivity analysis for speed indicates three group trends. The first group includes that at radius of curvature less than 200m, the bus driver speed is lower than 57km/hr is highly. The second group is speed range between 57km/hr to 77km/hr at radius of curvature between 200m and 450m. The third group is driving at higher speed exceeding 77km/hr at radius of curvature exceeds 450m. The following Figure 4.12 presents the probability of speed and radius of curvature. These probabilities are based on the cumulative probability of the speed variable as a factor of radius of curvature.

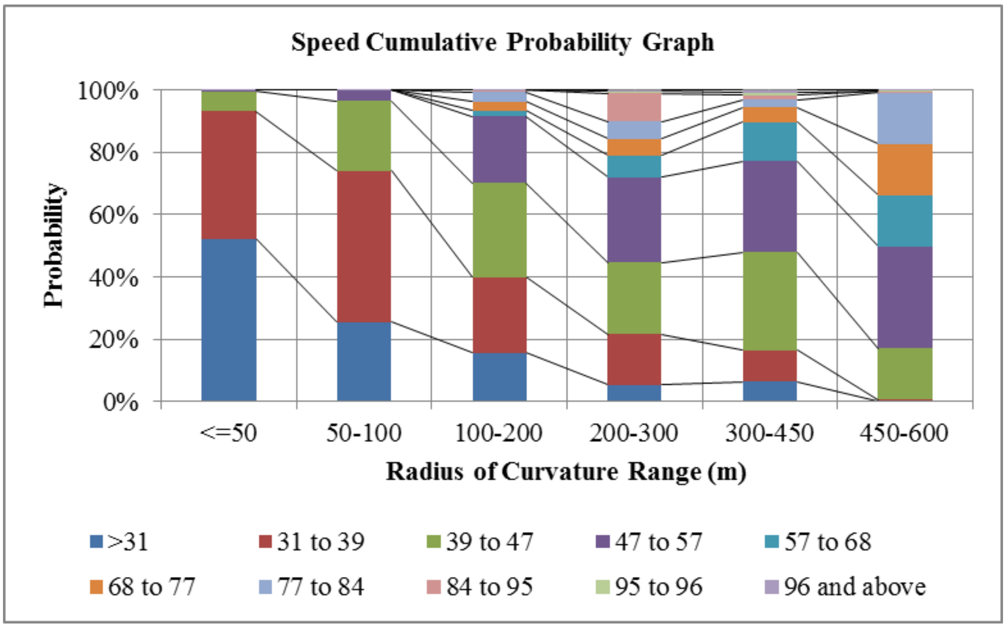

Figure 4.12: Speed variable probability based on radius of curvature range variable

4.5.2 Driving Conditions

Bus type is directly influenced by one variable namely the longitudinal acceleration with 0.30 as a measure of influence. Based on the cumulative probability distribution, two (2) trends are distinguished including the values of longitudinal acceleration below 0.09g and above 0.09g. In fact, 51% of the longitudinal acceleration values below 0.09g are associated with low floor, compared to 42% and 7% to be associated with stage bus (13,400kg), and stage bus (9,400kg) respectively. In comparison, the longitudinal acceleration values above 0.09g are associated with stage buses especially the 9,400kg stage bus. For example, 44% of the forecasted bus type is stage bus (9,400kg) which has longitudinal acceleration range between 0.09g to 0.14g, compared to 21% stage bus (13,400kg), and 17% low floor bus. The stage bus (9,400kg) dominates the forecasted bus type population when longitudinal acceleration exceeds 0.14g. The double-deck bus is clearly associated with longitudinal accelerations exceeds 0.25g yet putting into consideration that this bus type has the least presentation in the collected data cannot justify the mentioned association. The following Figure 4.13 presents the bus type cumulative probability graph based on the constructed Bayesian network.

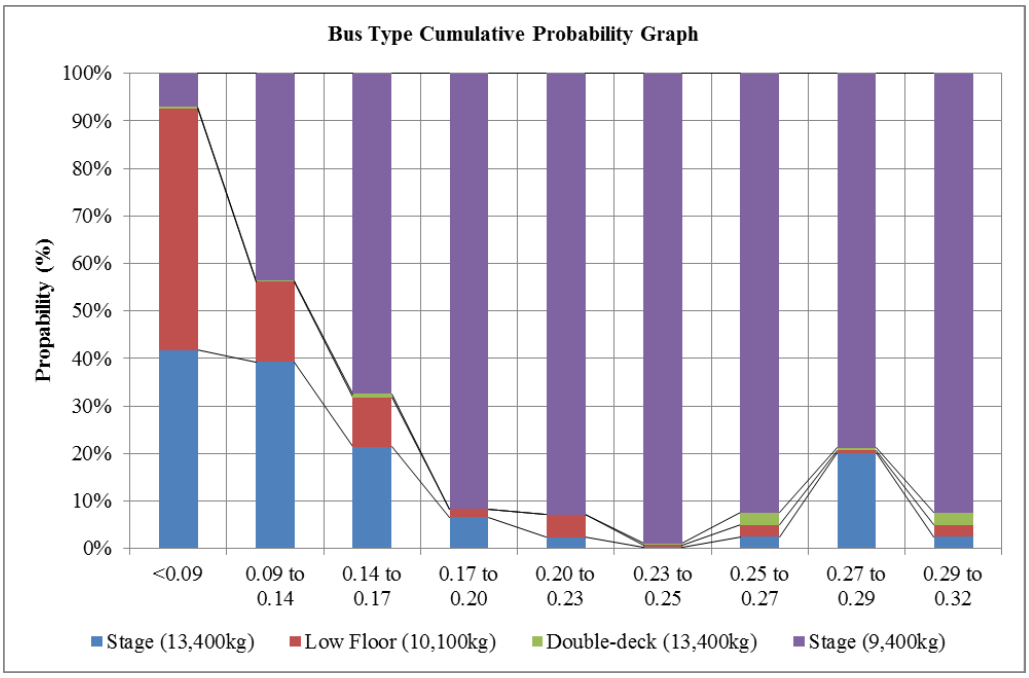

Figure 4.13: Bus type cumulative probability graph based on longitudinal acceleration variable

Moreover, those findings support the earlier results of Pearson correlation which indicated that driver longitudinal acceleration is much influenced by the driving task and environment conditions.

The bus type is in the centre of the network with converging influence on driver characteristics, geometric variables including tangent and radius, and accident risk perception variable.

Road geometric elements are directly connected by variables of bus type in case of radius of curvature and driving at investigated road to tangent. The grade and radius of curvature variables are conditionally dependent on the diverging tangent node Tangent length. Grade variable has no influence in the network. The radius of curvature has direct impact on the speed which is based on the model has no impact on other variables.

4.5.3 Driver Characteristics

Commercial vehicles are associated with certain demographics of drivers. Therefore, it is anticipated to find that the bus directly influence each variable of the driver characteristics.

Driver experience at investigated road (DIV) node is linear connected to the bus type node and a degree of influence at 0.58. The small variation in driver population of both double-deck and stage (9,400kg) influence the forecasting applicability of the model to predict the associated probability across a variable parameter. The DIV of less than 5 years of experience dominates the population of driver categories. The associated DIV is less than or equal to 5 years for stage bus (9,400kg) at 100%, for low floor bus is 62%, and 47% for stage (13,400kg). The DIV driver variable of a stage bus (13,400kg) has 42% chance to have 5 years or less experience, 27% between 6 to 15 years, 21% chance to be a 21 years or more, and only a 3% chance to be between 16 to 20 years of experience. The DIV driver variable of a low floor bus has 62% chance to have 5 years or less experience, 32% between 6 to 15 years, 5% chance to be 21 years or more, and 0% chance between 16 to 20 years of experience. The following Figure 4.14 presents those findings based on the constructed model.

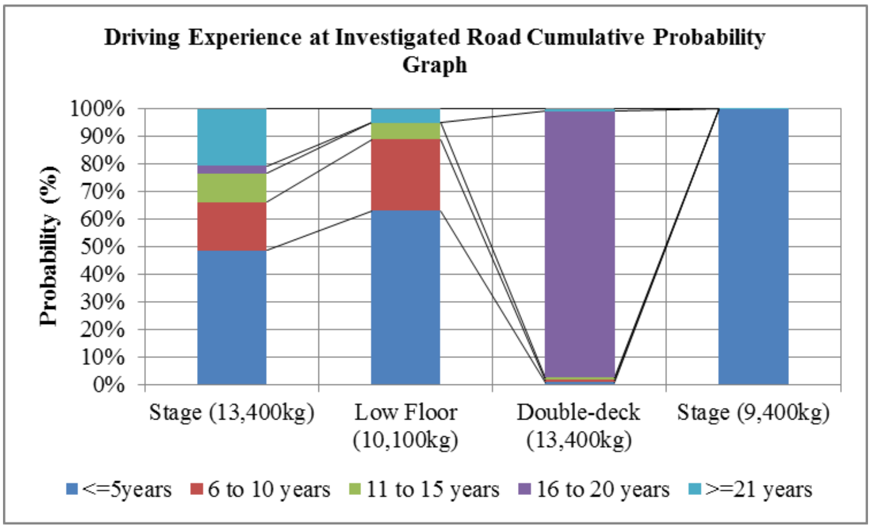

Figure 4.14: Driving at investigated road cumulative probability graph based on bus type variable

The variable node of total years of driving experience node inherits its properties from bus type and driving at investigated road nodes. The strength connection between total years of experience and bus type is 0.48 and 0.46 with driving at investigated road.

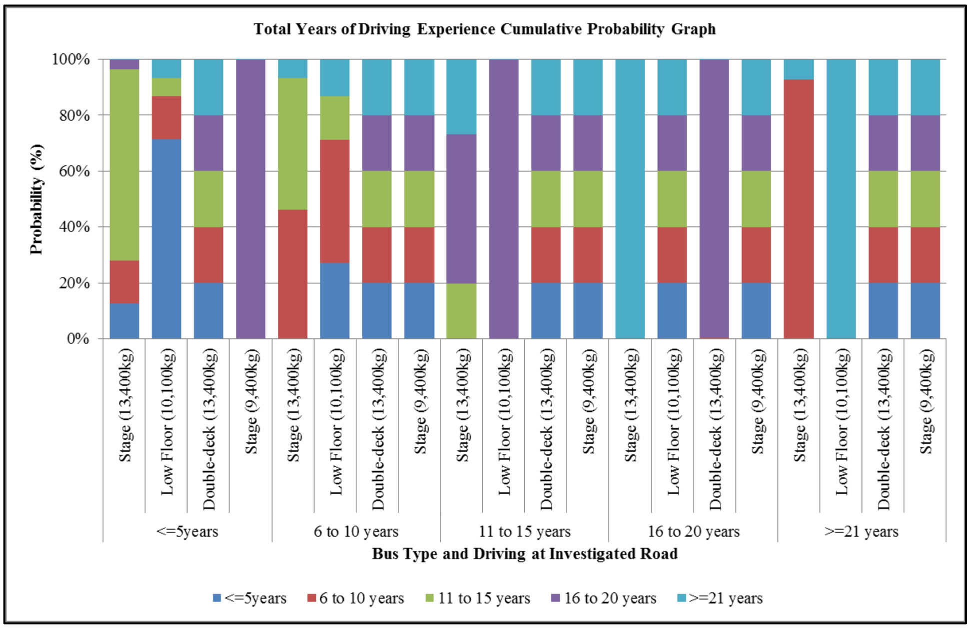

Figure 4.15: Probabilistic distribution of total years of driving experience based on bus type and driving at investigated road variables

Despite driver age is end node as no edge emerges from the mentioned node to any other nodes within the network, the age node gives important insight on bus driver population on road. The forecasted age depends on the variables of driving at investigated road, total years of driving experience, and bus type. The influence of strength of those parent variables on age is almost equal and range between 0.23 and 0.25.

The variation of the driver age groups is associated with low floor and stage (13,400kg) buses compared to double-deck and stage (9,400kg) buses. This is mainly due to the variation in the collected sample of the driver population across different bus groups that can affect the probability of the constructed network. The age is significantly correlated to both driving at investigated road and total years of driving experience.

Based on Figure most of the double-deck and stage (9,400kg) have driver population age between 31 and 50 years old. This result is influenced by associated with the collected data. The small sample of surveyed drivers of these two bus categories influences the probability distribution. The strength of influence indicates much stronger significance of driver age and driving experience at investigated road at 0.48. In comparison, only 0.23 strength of influence is measured for driver age and each of the other two variables including total years of driving experience and bus type. The age intuitively increases with experience variables. The following Figure 4.16 presents those findings.

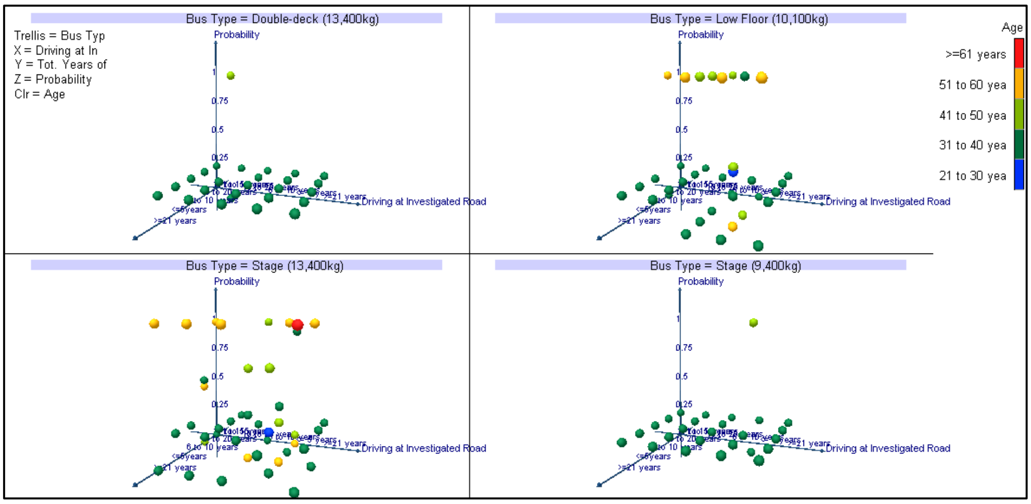

Figure 4.16: Probabilistic distribution of the driver age variable based on bus type, total years of driving experience and driving experience at investigated road

4.5.4 Accident Risk Perception

The accident risk perception is directly influenced by the driving at investigated road and bus type. The influence strength is 0.28 and 0.21 for bus type and driving at investigated road respectively. The latter is much more sensitive to the accident risk compared to bus type.

The results concerning the driver characteristics indicate that deterrent actions for risk perception should focus on year of driving experience at investigated road.

Table 4.6: CPT for accident risk based ant its parent variables including experience at driving experience at investigated road and bus type

	Driving Experience at Investigated Road	Bus Type	Forecasted Accident Risk		
			Danger	Moderate or Neutral	Safe
Variable State	1	1	30%	26%	44%
		2	1%	19%	80%
		3	33%*	33%*	33%*
		4	23%	40%	36%
	2	1	3%	4%	93%
		2	0%	9%	91%
		3	33%*	33%*	33%*
		4	33%*	33%*	33%*
	3	1	7%	14%	79%
		2	0%	17%	83%
		3	33%*	33%*	33%*
		4	33%*	33%*	33%*
	4	1	5%	5%	89%
		2	33%*	33%*	33%*
		3	0%	0%	99%
		4	33%*	33%*	33%*
	5	1	8%	40%	52%
		2	12%	56%	32%
		3	33%*	33%*	33%*
		4	33%*	33%*	33%*

*Unobserved cases in the collected data. There CPT values are forecasted by BN

According to the Tables 4.6 several findings are drawn. The worst risk perception is associated with drivers who have less than 5 years of driving experience at investigated road followed by the most senior group of 21 and above years of experience. This finding is parallel to the results of linear correlation. Regardless of the bus type the driver group (state 2) who had 6 to 10 years of driving experience have the highest level of safety perception at

91% to 93% compared to 79% to 83% for the second group in rank (state 3) who have 11 to 15 years of driving at investigated road. These findings are parallel to the literature such as in Tseng (2012).

It is noticeable that in some of the variable accident risk state combinations has uniform parameter forecasting i.e 33% for each state. This is justified as no background knowledge concerning these combinations is present in the collected data. As example of such cases is the double-deck bus. Only one way trip was performed using this type of bus at North-South expressway. The driver who had state 3 (11 to 15 years) driving experience at this expressway, recorded the highest safety rate at 99% in the studied driver population. Such a finding cannot be conclusive as it is based on 1 data records. Further investigation is needed. The second scenario which is also restricted to 1 driver characteristics is the stage bus (9,400 kg). This case is recorded at Route 59. Putting into account that several trips were made with the same driver as no other drivers are found to operate the local bus trips at Route 59. The averaged data records of the accomplished trips (3 northbound and 2 southbound) were used in the analyses. The third combination case is the 16 to 20 years of experience and low floor bus type. The sample of the driver characteristic is not present in the collected data.

4.6 Sensitivity Analysis

Based on the results of the sensitivity analysis in the following Table 4.6, three (3) variables including driving experience at investigated road (DIV), bus type (BT), and longitudinal acceleration (LA) are found to have potential effect the accident risk compared to other variables. Table 4.6 presents the five most significant scenarios which affect the risk variable. Table 4.7 has three values namely the lower bound, base and upper bound values. The lower bound value is the value of an uncertainty where there is only a 10% estimated chance that the forecasted value will be less than said base value. The base value is the measured probability from the developed model. The upper bound value represents the value of uncertainty where there is a 10% chance to be higher than the base value.

The longitudinal acceleration plays a role in occupant comfort. Though the mentioned variable is not jointly conditioned with the rating, yet a significant impact is found for this type of acceleration as a sole variable and as a combined variable with the bus type.

The combination 4 is forecasted as the worst combination at upper bound value of 26.6% which is almost twice the base probability of unsafe rating at 13%. A driver with minimal driving experience at a mountainous terrain such as Route59 and out-dated bus present a hazard for passenger and traffic safety (critical situation).

The bus type can improve the safety by as much as 3% when other parameters are held constant. The low floor bus (state 2) has the highest range of safety (Rate3) compared to stage buses (states 1 and 4).

Table 4.7: Sensitivity analysis based on n-way analyses for the BN model

Ranking of Combination	**DIR**	**BT**	**LA**	**Rate 1 (Danger) Probability**				**Rate 2 (Neutral) Probability**				**Rate 3 (Safe) Probability**			
				Lower Bound	Base	Upper Bound	Range	Lower Bound	Base	Upper Bound	Range	Lower Bound	Base	Upper Bound	Range
Combination 1 (C1)	state 1	state 2		0.1298		**0.1303**	0.0005	**0.2517**		0.2605	0.0088	**0.5953**		0.6325	0.0372
Combination 2 (C2)			Ay01	**0.124**		0.1357	0.0117	0.2496		**0.2625**	0.0129	**0.6017**		0.626	0.0243
Combination 3 (C3)	state 1	state 1		**0.1255**	0.13	0.1345	0.009	**0.2514**	0.2561	0.2607	0.0093	0.6085	0.6139	**0.6192**	0.0107
Combination 4 (C4)	state 1	state 4		**0.124**		0.2663	0.1423	**0.2458**		0.2663	0.0205	**0.6046**		0.6232	0.0186
Combination 5 (C5)		state 2	Ay01	**0.1245**		0.1356	0.0111	0.2538		**0.2584**	0.0046	**0.606**		0.6217	0.0157

DIR: Driving experience at Investigated Route, BT: Bus Type and LA: Longitudinal Acceleration.

Bold font values indicate the likelihood of the forecasted case; that is towards the upper or lower bounds.

CHAPTER 5

CONCLUSION

5.1 Summary of Study

Appropriate steps have to be taken to address the issue of bus accidents and driving behaviour on rural roadways. Driving behaviour is influenced by three factors, namely the driver, the vehicle, and the environment. The relationship between the driver behaviour and variables related to accident where investigated using the Lamm model, Pearson correlation and Bayesian network. Each of the three quantifiable measures has different approach to identify the interactions. Lamm model focuses on the geometric design and its consistency with driver perception in terms of speed behaviour and friction coefficient parameters. The model indicated that the driver perception differed significantly (20km/hr or more) when it comes to speed limit and operational speed in certain road sections. When the operational speed is higher than the speed limit, the cause is related to tight time schedules and working conditions. In comparison, the geometric terrain caused the cruising speed to drop well below speed limit. Moreover, those deficiencies between driving and speed limit led to inconsistencies in the developed friction coefficients especially at sections where the driving speed where higher than the speed limit. Both friction and speed variables play important role in accident occurrence. Presented literature studies in this dissertation indicated that the probability of accident increases when speed drops above 15km/hr for heavy vehicles. Moreover, low frictions can destabilise a vehicle causing it to overturn on corners. The evaluation of a geometric design to be ‘poorly’ rated if and only at least two of its three criteria fail the scale of Lamm model. The three criteria include speed consistency between subsequent sections, operational speed consistency and friction coefficients. Thereof, the overall evaluation of each of the five roads was capable to pass the Lamm model evaluation. Exception is found for 9% of B66 Jalan Batang Kali and 3% of North-South highway to match the criteria of ‘poor’ inconsistent designs.

Pearson correlation analysis was carried out to study the linear relationship among the variables. Despite the risk perception variable was not conclusively proven to be linear to accident variables using Pearson correlation, other significant linear relationships among variables were found. The results indicated that positive significant correlation is found between long haul bus and driving experience. The experienced drivers tend to work for long haul; compared to young drivers who have higher presence in transit buses. Low floor bus is associated with low longitudinal acceleration and speed values compared to long haul buses with much aggressive and higher speed. The bus function, working condition and surrounding factor influence the longitudinal acceleration and speed. The lateral acceleration is most significant factor for the resultant acceleration, though no significant relationship is conclusively proven for lateral acceleration with other variables.

Based on the literature, a methodological limitation in constructing an accident risk model is the dependence on post-accident reports. Such method approach under evaluate driving behaviour parameters. In this study, the limitation was overcome by collecting onsite data which puts into account the driver behaviour variables. Overall the model is developed using 13 variables. The model has a precision value of 48% and accuracy measure of about 85%. The mean absolute error (MAPE) is 4.58% which is relatively small and is considered acceptable for this probabilistic model. The model results could confirm some of the findings of the Pearson correlation and Lamm model including driver’s behaviour on curvatures, driver longitudinal acceleration is much influenced by the driving task and environment conditions, and relationship between driver demographics and working conditions. For accident risk, the impact of young drivers driving at investigated road is found to be similar to literature. In addition to this parameter, the quantification of accident risk is significantly associated with working condition and environment as well as longitudinal acceleration.

5.2 Limitation of Study

The limitation faced in the current research thesis appeared during the evaluation of the model. When only 55% or 11 cases out of the 20 cases were tested and evaluated. The other nine cases do not have supportive background knowledge in the collected data from bus trips. Despite the stated limitation, the model forecasting capability is not affected.

5.3 Recommendations

Research is still required in the field of bus accident risk modelling to support and improve the findings of this thesis. It is recommended to adapt the developed risk model to more explanatory variables which can affect the driving behaviour such as gender factor, weather conditions, night driving, passenger effect, and others. So far little literature about Bayesian modelling is done in literature in general and for bus type of transportation in particular.

Moreover, the outcome of the model can provide wide range of solutions for policy makers and traffic engineers to provide safer roads for bus users. For instance, the BN model could assist in designing characterised employment program. The program assists to best match variables of driver, bus type and road geometry to have the best performance in safe driving, and to train the bus drivers who might not be of a fit for certain circumference of geometric road sections.

REFERENCES

Abele, L. & Møller, M, n.d. The relationship between road design and driving behavior. RSS 2011: Road Safety and Simulation 2011 Conference.

Abidin, AN, Faudzi, SA, Lamin, F & Manap, AR 2012. *MIROS Crash Investigation and Reconstruction Annual Statistical Report 2007–2010*. Malaysian Institute of Road Safety Research (MIROS). viewed 21 April 2014, <http://www.miros.gov.my/c/document_library/get_file?uuid=1c2d3aa1-eb7e-40a3-8322-55e152beb5c4&groupId=26426>.

Acid, S. and de Campos, L.M., 2001. A hybrid methodology for learning belief networks: BENEDICT. *International Journal of Approximate Reasoning*, *27*(3), pp.235-262.

Akaike, H., 1978. A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical mathematics*, *30*(1), pp.9-14.

Albertsson, P. and Falkmer, T., 2005. Is there a pattern in European bus and coach incidents? A literature analysis with special focus on injury causation and injury mechanisms. *Accident Analysis & Prevention*, *37*(2), pp.225-233.

af Wåhlberg, A. E. 2004, The stability of driver acceleration behavior, and a replication of its relation to bus accidents, *Accident Analysis & Prevention*, 36**,** 83-92.

af Wåhlberg, A. E. 2007, Effects of passengers on bus driver celeration behavior and incident prediction, *J Safety Res*, 38**,** 9-15.

af Wåhlberg, A.E., 2007. Long-term prediction of traffic accident record from bus driver celeration behavior. *International journal of occupational safety and ergonomics*, *13*(2), pp.159-171.

Ahmad Noor Syukri, Z., Siti Atiqah, M., Fauziana, L. & Abdul Rahmat, A. 2012, MIROS crash investigation and reconstruction: annual statistical report 2007-2010, Malaysian Institute of Road Safety Research (MIROS), viewed 21 April 2014, <http://www.miros.gov.my/web/guest/reports?p_p_id=101_INSTANCE_ssIa&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=rightbar&p_p_col_pos=1&p_p_col_count=2&_101_INSTANCE_ssIa_struts_action=%2Ftagged_content%2Fview_content&_101_INSTANCE_ssIa_redirect=%2Fweb%2Fguest%2Freports&_101_INSTANCE_ssIa_assetId=430968>.

Ajzen, I., 2011. The theory of planned behaviour: reactions and reflections. Psychol. Health 26 (9), 1113–1127.

Akaike, H., 1998. A Bayesian analysis of the minimum AIC procedure. In *Selected Papers of Hirotugu Akaike* (pp. 275-280). Springer New York.

Arrifin, AH, Rahman, MK, Solah, MS, Kassim, KA & Voon, WS 2012, *Stability of high-deck buses in a rollover and contact-impact with traffic barriers*, Malaysian Institute of Road Safety Research (MIROS), viewed 11 May 2014, <http://www.miros.gov.my/c/document_library/get_file?uuid=363ebc08-8c4e-433b-83e5-68f19adedefb&groupId=26426>.

Baker, J. (t975) Traffic Accident Investigation Manual, Traffic Institute, Northwestern University, Evanston, IL.

Bathija, G.V., Bant, D.D., Itagimath, S.R., Lokare, L., Godbole, M., Nekar, M.S., Shidaraddi, K. and Kurigodiavar, M.D., 2014. A study on stress among government city bus drivers in Hubli. *International Journal of Biomedical Research*, *5*(2), pp.102-104.

Ben-Gal, I., 2007. Bayesian networks. *Encyclopedia of statistics in quality and reliability*.

Berndt, H, Emmert, J & Dietmayer, K 2008, *Continuous driver intention recognition method based on a driver model framework*, Proceeding of the 11th International IEEE Conference on Intelligent Transportation Systems, October 12-15.

Bromberg, F. and Margaritis, D., 2007, January. Efficient and Robust Independence-Based Markov Network Structure Discovery. In *IJCAI* (pp. 2431-2436).

Bøggild, H. and Knutsson, A., 1999. Shift work, risk factors and cardiovascular disease. *Scand J Work Environ Health*, *25*(2), pp.85-99.

Biggs, H., Dingsdag, D. and Stenson, N., 2009. Fatigue factors affecting metropolitan bus drivers: A qualitative investigation. *Work*, *32*(1), pp.5-10.

Blower, D., Green, P.E. and Matteson, A., 2008. Bus operator types and driver factors in fatal bus crashes: results from the buses involved in fatal accidents survey.

Borhan, MN, Syamsunur, D, Akhir, NM, Yazid, MR, Ismail, A & Rahmat, RA (2014). Predicting the use of public transportation: a case study from Putrajaya, Malaysia. *The Scientific World Journal*, 2014, 1-9.

Buntine, W. (1991). Theory refinement on Bayesian networks. *In Proceedings of Seventh Conference on Uncertainty in Artificial Intelligence*, Los Angeles, CA, pages 52{60. Morgan Kaufmann.

Buntine, W., 1996. A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on knowledge and data engineering*, *8*(2), pp.195-210.

Cafiso, S., Di Graziano, A. and Pappalardo, G., 2013. Using the Delphi method to evaluate opinions of public transport managers on bus safety. *Safety science*, *57*, pp.254-263.

Chang, H.L. and Yeh, C.C., 2005. Factors affecting the safety performance of bus companies—The experience of Taiwan bus deregulation. *Safety Science*, *43*(5), pp.323-344.

Chapter 4 vehicle dynamics n.d., viewed 5 May 201, <http://scholar.lib.vt.edu/theses/available/etd-5440202339731121/unrestricted/CHAP3_DOC.pdf>.

Chimba, D., Sando, T. and Kwigizile, V., 2010. Effect of bus size and operation to crash occurrences. *Accident Analysis & Prevention*, *42*(6), pp.2063-2067.

Chen, D., 2014. *Predictive accident modeling for highway transportation system using Bayesian networks* (Doctoral dissertation, Memorial University of Newfoundland).

Cheng, J., Bell, D.A. and Liu, W., 1997, January. Learning belief networks from data: An information theory based approach. In *Proceedings of the sixth international conference on Information and knowledge management* (pp. 325-331). ACM.

Clarke, E.J. and Barton, B.A., 2000. Entropy and MDL discretization of continuous variables for Bayesian belief networks. *International Journal of Intelligent Systems*, *15*(1), pp.61-92.

Conrady, S., Jouffe, L. 2011. *Introduction to Bayesian networks practical and technical perspective*, Condray applied science, LLC, viewed 5 September 2013, <http://www.bayesia.com/en/products/bayesialab/resources/tutorials/IntroBBN.php>.

Dawid, P. (1984). Statistical theory. The prequential approach (with discussion). *Journal of the Royal Statistical Society A*, 147:178{292.

Davis, G.A., 2001, August. Using Bayesian networks to identify the causal effect of speeding in individual vehicle/pedestrian collisions. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence* (pp. 105-111). Morgan Kaufmann Publishers Inc..

Dell'Olio, L., Ibeas, A. & Cecín, P. 2010, Modelling user perception of bus transit quality. *Transport Policy*, 17**,** 388-397.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accid. Anal. Prev. 38 (3), 434–444.

De Oña, J., Mujalli, R.O. and Calvo, F.J., 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. *Accident Analysis & Prevention*, *43*(1), pp.402-411.

De Oña, J. and Garach, L., 2012. Accidents prediction model based on speed reduction on Spanish two-lane rural highways. *Procedia-Social and Behavioral Sciences*, *53*, pp.1010-1018.

Deublein, M., Schubert, M., Adey, B.T. and de Soto, B.G., 2015. A Bayesian network model to predict accidents on Swiss highways. *Infrastructure Asset Management*, *2*(4), pp.145-158.

Deublein, M.K., 2013. *Roadway accident risk prediction based on Bayesian probabilistic networks* (Doctoral dissertation, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 21093, 2013).

Dimitropoulos, I, and Kanellaidis, G, 1998, *Highway Geometric Design: The Issue of Driving Behavior Variability*, Proceeding of International Symposium.

Dougherty, J., Kohavi, R. and Sahami, M., 1995, July. Supervised and unsupervised discretization of continuous features. In *Machine learning: proceedings of the twelfth international conference* (Vol. 12, pp. 194-202).

Dorn, L. and Af Wåhlberg, A., 2008. Work-Related Road Safety: An Analysis Based on UK Bus Driver Performance. *Risk analysis*, *28*(1), pp.25-35.

Ebert-Uphoff, I., 2009. Tutorial on how to measure link strengths in discrete Bayesian networks.

Elkan, C., 2012. Evaluating classifiers. *University of San Diego, California, retrieved [01-11-2012] from http://cseweb. ucsd. edu/elkan B, 250*.

Elvik, R., Hoye, A., Vaa T., 2002. The Handbook of Road Safety Measures (Second Edition). UK: Emerald Group.

Evans, W.A. and Courtney, A.J., 1985. An analysis of accident data for franchised public buses in Hong Kong. *Accident Analysis & Prevention*, *17*(5), pp.355-366.

Fast, A.S., 2010. *Learning the structure of Bayesian networks with constraint satisfaction* (Doctoral dissertation, University of Massachusetts Amherst).

Feng, S., Li, Z., Ci, Y. and Zhang, G., 2016. Risk factors affecting fatal bus accident severity: Their impact on different types of bus drivers. *Accident Analysis & Prevention*, *86*, pp.29-39.

Field, A. 2000, Discovering statistics using spss for windows. London- Thousand Oaks- New Delhi: Sage publications.

Field, A. 2009, Discovering statistics using SPSS. London: SAGE.

Fish, K.E., Blodgett, J.G., 2003. A visual method for determining variable importance in an artificial neural network model: an empirical benchmark study. J. Target. Meas. Anal. Market. 11 (3), 244–254.

Fishbein, M., 2009. An integrative model for behavioral prediction and its application to health promotion. In: Di Clemente, R.J., Crosby, R.A., Kegler, M.C. (Eds.), Emerging Theories in Health Promotion Practice and Research. John Wiley & Sons, New York, pp. 215–234

Friedman, N., Nachman, I. and Peér, D., 1999, July. Learning bayesian network structure from massive datasets: the «sparse candidate «algorithm. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* (pp. 206-215). Morgan Kaufmann Publishers Inc.

Gangopadhyay, A. and Ng, J. 2013. *Malaysia Bus Accident Claims 37 Lives*. [online] WSJ. Available at: http://www.wsj.com/articles/SB10001424127887323980604579026742660795498 [Accessed 12 Sep. 2016].

Garber, N. J., Hoel, L. A. 2009. *Traffic and highway engineering* (4th ed.). US: Thomson West.

Garber, N. J. and A. A. Ehrhart, 2000. The Effect of Speed, Flow, and Geometric Characteristics on Crash Rates for Different Types of Virginia Highways. Virginia Transportation Council.

Goh, K., Currie, G., Sarvi, M. & Logan, D. 2014, Factors affecting the probability of bus drivers being at-fault in bus-involved accidents. *Accid Anal Prev*, 66**,** 20-6.

GeNIe, 2015, *Introduction*. [online] viewed on 28 September 2016, <https://dslpitt.org/genie/wiki/Introduction:_GeNIe>

Glass, G.V., P.D. Peckham, and J.R. Sanders. 1972. Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. Rev. Educ. Res. 42: 237-288.

Google Maps a. 2015 [Genting Highland][Street Map]. Retrieved from https://www.google.com/maps/place/Genting+Highlands,+Pahang,+Malaysia/@3.3861353,101.7534098,13z/data=!4m2!3m1!1s0x31cc14067e9a39c3:0x1b8edcdc4d5bf56

Google Maps b. 2015. [North South Expressway] [Street map]. Retrieved from

<https://www.google.com/maps/place/AH2,+30020+Ipoh,+Perak,+Malaysia/@4.7141509,100.9024694,12z/data=!4m15!1m12!4m11!1m3!2m2!1d101.0711696!2d4.6677437!1m6!1m2!1s0x31cabe722b648505:0xc08d2f283c2344ba!2sKampung+Menora,+33000+Kuala+Kangsar,+Perak,+Malaysia!2m2!1d100.947168!2d4.711462!3m1!1s0x31ca949d8408357f:0x2c8d2a1af87d6533>.

Google Maps c. 2015. [Batu Ferringhi, George Town, Penang] [Street map]. Retrieved from <https://www.google.com.my/maps/@5.4692148,100.2410331,14>.

Google Maps d. 2015. [Karak Highway] [Street map]. Retrieved from <https://www.google.com/maps/place/Karak+Expy,+Malaysia/@3.3906883,101.7077357,11z/data=!4m2!3m1!1s0x31cc17ebb92363cf:0xf1e61726dc4d3646>.

Gravetter, F., & Wallnau, L. 2014, Essentials of statistics for the behavioral sciences (8th ed.). Belmont, CA: Wadsworth.

Häkkänen, H., Summala, H., Partinen, M., Tiihonen, M. and Silvo, J., 1999. Blink duration as an indicator of driver sleepiness in professional bus drivers. *Sleep*, *22*(6), pp.798-802.

Halliday, D., Resnick, R., & Walker, J. 1993. *Fundamentals of physics extended* (4th edn.). New York, US: John Wiley & Sons, Inc.

Hamed, M.M., Jaradat, A.S. and Easa, S.M., 1998. Analysis of commercial mini-bus accidents. *Accident Analysis & Prevention*, *30*(5), pp.555-567.

Howard, R. and Matheson, J., editors (1983). The Principles and Applications of Decision Analysis. Strategic Decisions Group, Menlo Park, CA.

Hassan, Y. and Easa, S.M., 2003. Effect of vertical alignment on driver perception of horizontal curves. *Journal of transportation engineering*, *129*(4), pp.399-407.

Hassan, Y. and Sarhan, M., 2012. Operational effects of drivers' misperception of horizontal curvature. *Journal of Transportation Engineering*, *138*(11), pp.1314-1320.

Haupt, C. M., et al. (2008), 'The relation of exposure to shift work with atherosclerosis and myocardial infarction in a general population', *Atherosclerosis*, 201 (1), 205-11

Harwell, M.R., E.N. Rubinstein, W.S. Hayes, and C.C. Olds. 1992. Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. J. Educ. Stat. 17: 315-339.

Heckerman, D. (1996). A Tutorial on learning with Bayesian networks. Microsoft Research Technical Report MSR-TR-95-06. Updated Nov. 1996.

Heckerman, D., Mamdani, A., and Wellman, M. (1995). Real-world applications of Bayesian networks. Communications of the ACM, 38.

Hege, A., Perko, M., Johnson, A., Yu, C.H., Sönmez, S. and Apostolopoulos, Y., 2015. Surveying the impact of work hours and schedules on commercial motor vehicle driver sleep. *Safety and health at work*, *6*(2), pp.104-113.

Heger, R. 1998. Driving behavior and driver mental workload as criteria of highway geometric design quality. *Transportation research circular***,** 43:1-10.

Highway Geometric Design Practices, Boston, Massachusetts, Aug. 30–Sept. 1, 1995, pp.41-1.

Hill, J. 2008, '*Getting ahead: returning Britain to European leadership in road casualty reduction'*, *Road Safety Foundation*, viewed 2 October 2013, <http://www.roadsafetyfoundation.org/media/14940/getting%20ahead.pdf>.

Hudson, R. W., Haas, R. & Uffin, W. 1997. *Infrastructure Management*, New York, McGraw-Hill.

Huy, N. H. & Taneerananon, P. 2013, Reality of urban road safety in Ho Chi Minh City and suggested solutions. *Journal of Society for Transportation and Traffic Studies*, 3**,** 9-21.

Ivancevich, J. M., Konopaske, R., Matteson, M. T. 2005, Organizational Behavior and Management, New York, McGraw-Hill/Irwin, 7th Edition

Jayaraman, K, Susan, J., & Lin S. 2011, Robust models for the utilization of public bus transport services in Malaysia. *African Journal of Business Management*, 5(26), 10586-10600.

Jayatilleke, A.U., Nakahara, S., Dharmaratne, S.D., Jayatilleke, A.C., Poudel, K.C. & Jimba, M. 2009. "Working conditions of bus drivers in the private sector and bus crashes in Kandy district, Sri Lanka: A case-control study". Injury Prevention, vol. 15(2), pp. 80-86.

Jawi, ZM, Isa, MH, Sarani, R, Voon, WS & Sadullah, AF 2010, *Weather as a road safety hazard in Malaysia-an overview*, Malaysian Institute of Road Safety Research (MIROS), viewed 11 May 2014, <http://www.miros.gov.my/c/document_library/get_file?uuid=d2ff8ec9-5247-4cfd-81af-72d8cfe14c74&groupId=10124>.

Jensen, F., Lauritzen, S., and Olesen, K. (1990). Bayesian updating in recursive graphical models by local computations. *Computational Statisticals Quarterly*, 4:269.

Kaya, E., Findik, O., Babaoglu, I. and Arslan, A., 2011. Effect of discretization method on the diagnosis of Parkinson's disease. *Int. J. Innov. Comp. Inf. Control*, *7*, pp.4669-4678.

Kee, S., Tamrin, S.B.M. and Goh, Y., 2010. Driving fatigue and performance among occupational drivers in simulated prolonged driving. *Global Journal of Health Science*, *2*(1), p.167.

Kjærulff, U. and van der Gaag, L.C., 2000, June. Making sensitivity analysis computationally efficient. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence* (pp. 317-325). Morgan Kaufmann Publishers Inc..

Kaplan, S. and Garrick, B.J., 1981. On the quantitative definition of risk. *Risk analysis*, *1*(1), pp.11-27.

Kaplan, S. and Prato, C.G., 2012. Risk factors associated with bus accident severity in the United States: A generalized ordered logit model. *Journal of safety research*, *43*(3), pp.171-180.

Kamaruddin, R., Osman, I., & Pei, C. A. C. 2012. Public transport services in Klang valley: customer expectations and its relationship using SEM. *Procedia-Social and Behavioral Sciences*, 36, 431-438.

La, Q.N., Van Duong, D., Lee, A.H. and Meuleners, L.B., 2016. Factors underlying bus-related crashes in Hanoi, Vietnam. *Transportation research part F: traffic psychology and behaviour*.

Lamm, R., Choueiri, E. M. M. & Haryward, J. C., 1988. Tangent as independent design element. Transportation research board. 115,123-131.

Langely, BR 1999, *Dilution of precision*, viewed 2 March 2014, <http://www.nrem.iastate.edu/class/assets/nrem446_546/week3/Dilution_of_Precision.pdf>.

Lauritzen, S. and Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistical Society B*, 50:157.

Lee, P 2014, 'Call to scrap two-driver rule for long bus journeys', *The Star Online*, 24 April, viewed 24 April 2014, <http://www.thestar.com.my/News/Nation/2014/04/24/Call-to-scrap-twodriver-rule-for-long-bus-journeys-Operators-It-is-causing-more-fatigue-based-on-res/>.

Li Leen, C., Loh, I., P., A. and Kammed, K. 2016. *27 killed as tour bus overturns in Cameron Highlands*. [online] Sammyboy.com. Available at: http://www.sammyboy.com/showthread.php?82823-27-killed-as-tour-bus-overturns-in-Cameron-Highlands [Accessed 12 Sep. 2016].

Liang, Y. and Lee, J.D., 2014. A hybrid Bayesian Network approach to detect driver cognitive distraction. *Transportation research part C: emerging technologies*, *38*, pp.146-155.

Lee, J.D., Young, K.L., Regan, M.A., 2008. Defining driver distraction. In: Regan, M.A., Lee, J.D., Young, K.L. (Eds.), Driver Distraction: Theory, Effects, and Mitigation. CRC Press Taylor & Francis Group, Boca Raton, FL, USA, pp. 31–40.

Lix, L.M., J.C. Keselman, and H.J. Keselman. 1996. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. Rev. Educ. Res. 66: 579-619

Lucidi, F., Giannini, A.M., Sgalla, R., Mallia, L., Devoto, A., Reichmann, S., 2010. Young novice driver subtypes: relationship to

Lui, C. C., S. G. Hosking and M. G. Lenne. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of Safety Research,* doi:10.1016/j.jsr.2009.04.005.

Lum, H. & Reagan, J. A. 1995, Interactive highway safety design model: Accident predictive model, *Public Roads,* 58**,** 14-17.

Mahmud, M. A. (2010) *Studies on bus services and passenger demand sustainability*. Faculty of Civil Engineering & Earth Resources, Universiti Malaysia Pahang.

Mackay, D.J.C., 1992. Bayesian methods for adaptive models. Ph.D. Dissertation. California Institute of Technology, Pasadena, California.

Mallia, L., Lazuras, L., Violani, C. and Lucidi, F., 2015. Crash risk and aberrant driving behaviors among bus drivers: the role of personality and attitudes towards traffic safety. *Accident Analysis & Prevention*, *79*, pp.145-151.

Mannering, F.L. and Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. *Analytic Methods in Accident Research*, 1, pp.1-22.

Marzban, C., and A. Witt, 2001: A Bayesian neural network for severe-hail size prediction. *Wea. Forecasting*, 16, 600–610.

Ministry of Transport, 2016. *Statistics of Land Transport*. [Online] available at: http://www.mot.gov.my/en/resources/yearly-statistic [Accessed 01 September 2016].

Malaysian Institute of Road Safety Research (MIROS), 2014. *Annual Reports*. [Online] available at: https://www.miros.gov.my/1/publications.php?id_page=24 [Accessed 01 September 2016].

Miaou, S. and H. Lump, 1993. Modeling Vehicle Accidents and Highway Geometric Design Relationships. Accident Analysis and Prevention Vol. 25, 689–709.

Milton, J. and F. Mannering, 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation* Vol. 25, 395–413.

Mohamad, J. and Kiggundu, A.T., 2007. The rise of the private car in Kuala Lumpur, Malaysia: assessing the policy options. *IATSS research*, *31*(1), pp.69-77.

Murad, D 2014, 'Malaysia has 17th most dangerous roads in the world, according to Michigan university', *The Star Online*, 22 Feburary, viewed 12 May 2014, < http://www.thestar.com.my/News/Nation/2014/02/22/Nations-with-deadliest-roads-Malaysia-17th/>.

Mustafa, AA, & Hokao, Kazunori 2012, 'Database development of road traffic accident case study Johor Bahru, Malaysia', *Journal of society for transportation and traffic studies(JSTS)*, vol.3, no. 1, pp. 22-31.

Nakata, M. and Nishiyama, K., 1988. Fatigue and low-back pain of freight-container tractor drivers. Correlation with ergonomic factors. Sangyo igaku. Japanese journal of industrial health, 30(1), pp.28-45.

Neal, R.M., 1995. Bayesian learning for neural networks. Ph.D. Dissertation. University of Toronto, Toronto, Ontario

NIST/SEMATECH. 2012, *e-Handbook of Statistical Methods*. [online], [Accessed 1 September 2016]. Available from: http://www.itl.nist.gov/div898/handbook/, 2012.

Nordfjærn, T., Jørgensen, S.H., Rundmo, T., 2010. An investigation of driver attitudes and behaviour in rural and urban areas in Norway. Saf. Sci. 48 (3), 348–356.

Olmsted, S. (1983). *On representing and solving decision problems.* PhD thesis, Department of Engineering-Economic Systems, Stanford University.

Olson, P.L., Cleveland, D.E., Fancher, P.S., Kostyniuk, L.P. and Schneider, L.W., 1984. *Parameters affecting stopping sight distance* (No. HS-037 760).

Osman, MR, Abas, F, Othman, I, Ahmad, MS, Paiman, FN, Faudzi, SA, Batch, WA, Zulkipli, ZA & Noh, MS 2013, *Evaluation of HSE COP practices of a bus operator practicing one-man operation during Ops Bersepadu Chinese new year 2010*, Malaysian Institute of Road Safety Research (MIROS), viewed 11 May 2014, <http://www.miros.gov.my/c/document_library/get_file?uuid=cf9f2d31-3992-4ce0-81cd-17587da9cc60&groupId=10124>.

What is the acceptable range of skewness and kurtosis for normal distribution of data? - ResearchGate. Available from: https://www.researchgate.net/post/What_is_the_acceptable_range_of_skewness_and_kurtosis_for_normal_distribution_of_data [accessed Aug 31, 2016].

Pearl, J., 1986. Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, *29*(3), pp.241-288.

Pearl, J., 1995. Causal diagrams for empirical research. *Biometrika*, *82*(4), pp.669-688.

Pearl, J., 2009. *Causality*. Cambridge university press.

Pearl, J. and Verma, T.S., 1995. A theory of inferred causation. *Studies in Logic and the Foundations of Mathematics*, *134*, pp.789-811.

Poole, D., 2011, May. Logic, probability and computation: Foundations and issues of statistical relational AI. In *International Conference on Logic Programming and Nonmonotonic Reasoning* (pp. 1-9). Springer Berlin Heidelberg.

Prato, CG & Kaplan, S 2014, 'Bus accident severity and passenger injury: evidence from Denmark' *European Transport Research Review*, vol 6, no. 1, pp. 17-30. DOI: 10.1007/s12544-013-0107-z.

Razmpa, E., Niat, K.S., & Saedi, B. 2011. "Urban bus drivers' sleep problems and crash accidents". Indian Journal of Otolaryngology and Head and Neck Surgery, Vol. 63 (3), pp. 269-273.

Richard Scheines. An introduction to causal inference. In Causality in Crisis? Univ. of Notre Dame Press, 1997

Road Engineering Association of Malaysia (REAM), 2002, G*uide on geometric design of roads*.

Road transport department. 2016. *Land-statistic*. [online] Available at: http://www.mot.gov.my/en/resources/land-statistic [Accessed 12 Aug. 2016].

Roess, R., P., Prassas, E., S., & McShane, W., R. 2004. *Traffic Engineering* (3rd edn.). New Jersey, US: Pearson Prentice Hall

Ross, A & Goodge, M 2005, '*Arrive alive: ASEAN road safety strategy and action plan (2005–2010)*', Asian Development Bank, viewed 14 May 2014, <http://www.adb.org/sites/default/files/pub/2005/arrive-alive.pdf>.

Schwarz, G., 1978. Estimating the dimension of a model. *The annals of statistics*, *6*(2), pp.461-464.

Shachter, R. (1988). Probabilistic inference and in uence diagrams. *Operations Research*, 36:589.

Simoncic, M., 2004. A Bayesian network model of two-car accidents. Journal of Transportation and Statistics 7 (2, 3), 13–27.

Solah, M. S., Ariffin, A. H., Isa, M. H. M. & Wong, S. V. 2013, In-depth crash investigation on bus accidents in Malaysia, *Journal of Society for Transportation and Traffic Studies,* 3**,** 22-31.

Sumer, N. 2003. “Personality and behavioral predictors of traffic accidents: testing a contextual mediated model”. *Accident Analysis and Prevention*, vol. 35 96), pp. 949-964.

Sun, S., Zhang, C. and Yu, G., 2006. A Bayesian network approach to traffic flow forecasting. *IEEE Transactions on intelligent transportation systems*, *7*(1), pp.124-132.

Spacek, P., 2005. Track behavior in curve areas: attempt at typology. *Journal of transportation engineering*, *131*(9), pp.669-676.

Spirtes, P., Glymour, C.N. and Scheines, R., 2000. *Causation, prediction, and search*. MIT press.

Staplin, L, Gish, KW, Decina LE, Lococo, KH, Harkey, DL, Tarawneh, MS, Lyles, R, Mace, D & Philip, G 1997, *Synthesis of human factors research on older drivers and highway safety volume II: human factors and highway safety research synthesis*, Federal highway administration, viewed 10 December 2013, < https://www.fhwa.dot.gov/publications/research/safety/97094/97094.pdf>.

Steg, L. and Gifford, R., 2005. Sustainable transportation and quality of life. *Journal of transport geography*, *13*(1), pp.59-69.

Stradling, S., Carreno, M., Rye, T. & Noble, A. 2007, Passenger perceptions and the ideal urban bus journey experience, *Transport Policy,* 14, 4, 283-292.

Strathman, J.G., Wachana, P.H. and Callas, S., 2010. Analysis of Bus Collision Incidents Using Transit ITS and other Archived Operations Data.

Teyssier, M. and Koller, D., 2012. Ordering-based search: A simple and effective algorithm for learning Bayesian networks. *arXiv preprint arXiv:1207.1429*.

Thiffault, P. & Bergeron, J. 2003, Monotony of road environment and driver fatigue: a simulator study, *Accident Analysis & Prevention,* 35, 3, 381-391.

Thomas Richardson and Peter Spirtes. Ancestral graph Markov models. The Annals of Statistics, 30(4):962–1030, 2002.

Tseng, C.-M. 2012, Social-demographics, driving experience and yearly driving distance in relation to a tour bus driver’s at-fault accident risk, *Tourism Management,* 33, 3, 910-915.

Tsamardinos, I., Brown, L.E. and Aliferis, C.F., 2006. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine learning*, *65*(1), pp.31-78.

Transportation Research Board (TRP) 2001. *NCHRP Synthesis 299: Recent Geometric Design Research for Improved Safety and Operations*. National

Academy Press. Washington, D.C.. viewed on 5 May 2014, <http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_syn_299.pdf>.

Trochim, W. M., & Donnelly, J. P. 2006, The research methods knowledge base (3rd ed.). Cincinnati, OH:Atomic Dog.

Ulleberg, P., Rundmo, T., 2003. Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. Saf. Sci. 41 (5), 427–443.

Wang, Y., Zhang, C. and Mao, C., 2013. Fatal motor vehicle crashes on road segments in Harbin, China: combining rates into contributory factors. *Transport*, *28*(2), pp.117-129.

Wang, Y., Li, L., Feng, L. and Peng, H., 2014. Professional drivers' views on risky driving behaviors and accident liability: a questionnaire survey in Xining, China. *Transportation letters*, 6(3), pp.126-135.

Wylie, C.D., Shultz, T., Mitler, M.M. and Mackie, R.R., 1996. *Commercial motor vehicle driver fatigue and alertness study: Technical summary* (No. FHWA-MC-97-001).

Xie, Y., Lord, D. and Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. *Accident Analysis & Prevention*, *39*(5), pp.922-933.

Xu, J., Yang, K., Shao, Y. and Lu, G., 2015. An experimental study on lateral acceleration of cars in different environments in Sichuan, Southwest China. *Discrete Dynamics in Nature and Society*, *2015*.

Yang, C.Y., 2007. Trends in transit bus accidents and promising collision countermeasures. *Journal of Public Transportation*, *10*(3), p.7.

Zaharah, S and T. H. Law. (2004). The establishing of heavy vehicle and motorcycle accident model. Malaysian Universities Transport Research Forum Conference: 1st-2nd December 2004; The Pan-Pacific Glenmarie, Shah Alam.

Zegeer, C.V., Stutts, J.C., Huang, H., Zhou, M., Rodgman, E., 1993. Analysis of elderly pedestrian accidents and recommended countermeasures. Transport. Res. Rec. 1405, 56–63.

Zhang, G., Yau, K.K. and Chen, G., 2013. Risk factors associated with traffic violations and accident severity in China. *Accident Analysis & Prevention*, *59*, pp.18-25.

Zhao, J. and Deng, W., 2015. The use of Bayesian network in analysis of urban intersection crashes in China. *Transport*, *30*(4), pp.411-420.

Zong, F., Xu, H. and Zhang, H., 2013. Prediction for traffic accident severity: comparing the Bayesian network and regression models. *Mathematical Problems in Engineering*, *2013*.

Zou, Y., Zhang, Y., Lord, D., 2013. Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis. Accident Analysis Prevention 50, 1042–1051.

.

APPENDIX A

Survey Sample

Survey Question (Road Safety)

Surname: Mr

Please tick for correct answer for each question.

1) How many years experience in driving buses?
   ☐ 1-5 ☐ 6-10 ☐ 11-15 ☐ 16-20 ☐ 21 and above

2) How many years working as bus driver in this company?
   ☐ 1-5 ☐ 6-10 ☐ 11-15 ☐ 16-20 ☐ 21 and above

3) How many years of experience work as driver in driving this route?
   ☐ 1-5 ☐ 6-10 ☐ 11-15 ☐ 16-20 ☐ 21 and above

4) What is your age?
   ☐ 21-30 ☐ 31-40 ☐ 41-50 ☐ 51-60 ☐ 61 and above

5) How many working hours daily?
   ___ Hours.

6) Do you have any assistant driver on each bus trip?
   ☐ Yes ☐ No

7) What is the most common danger in driving this route in your opinion?
   - ☐ Working hour too long that causes fatigue
   - ☐ Driving at night
   - ☐ Geometry of road
   - ☐ Insufficient of experiences driving in this route
   - ☐ Insufficient of road barrier/safety features in this route
   - ☐ Behavior of driver when driving
   - ☐ Rushing to destination to get more trips per day
   - ☐ Age

APPENDIX B

Learning and Validation Datasets

The cross validation (rotation estimation) is used to divide the database to learning (training) set and validation (or testing set). The partition of the database is done using the function of the 'cvpartition' in Matlab software. The no of K-folds is set as 5 which is found to be recommended in different resources. The database consists mainly of the horizontal alignment of type curve, the sole tangent sections were not considered because of analyses purposes. The following command is used:

```
>> c= cvpartition(2854,'k',5)
c =
K-fold cross validation partition
   NumObservations: 2854
       NumTestSets: 5
         TrainSize: 2284  2283  2283  2283  2283
          TestSize: 570  571  571  571  571
>>traindata= find(training(c,5));
>>testdata= find(test(c,5));
```

APPENDIX C

	Network ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Training: Validation Sets	4:01									4:01				
	Discretization Method	K-mean Clustring									K-mean Clu				
	No. of Intervals	3									5				
	Learning Algorithm	**Bayesian Search**			**Greedy Thick Thining**						**Bayesian Search**				
	Scoring Function (Sample Size)	Bdeu (50)	Bdeu (50)	Bdeu (50)	K2	K2	K2	BDe u	BDeu	BDeu	Bdeu (50)	Bdeu (50)	Bdeu (50)	K2	K2
	Update Algorithm	Clustr ing	EPIS Samplin g	AIS Samplin g	Clust ring	EPIS Samplin g	AIS Samplin g	Clust ring	EPIS Samplin g	AIS Samplin g	Clustr ing	EPIS Samplin g	AIS Samplin g	Clust ring	EPIS Samplin g
	Inference Algorithm	Policy Evaluation									Policy Eval				
	Network Properties														
	Node count	13													
	Avg indegree	1.385	1.385	1.385	2.231	2.231	2.231	1.615	1.615	1.615	1.462	1.462	1.462	1.923	1.923
	Max indegree	3	3	3	4	4	4	3	3	3	3	3	3	4	4
	Avg outcomes	4.231	4.231	4.231	4.231	4.231	4.231	4.231	4.231	4.231	5	5	5	4.923	4.923
	Max outcomes	7	7	7	7	7	7	7	7	7	7	7	7	7	7
	Nodes	13	13	13	13	13	13	13	13	13	13	13	13	13	13
	Chance - General	13	13	13	13	13	13	13	13	13	13	13	13	13	13
	No. of Edges	18	18	18	29	29	29	22	22	22	19	19	19	25	25
	Validation														
Variabl e	**Validation Test**	Leave One													
Tangen t	Overall Accuracy	81.73 %	81.73%	81.73%	81.73 %	81.73%	81.73%	81.73 %	81.73%	81.73%	81.73 %	81.73%	81.73%	81.73 %	81.73%
	Overall Recall (This Is Referd As Accuracy In Genie)	36.05 %	36.05%	36.05%	36.05 %	36.05%	36.05%	36.05 %	36.05%	36.05%	36.05 %	36.05%	36.05%	36.05 %	36.05%
	Overall False-Positive Rate	10.66 %	10.66%	10.66%	10.66 %	10.66%	10.66%	10.66 %	10.66%	10.66%	10.66 %	10.66%	10.66%	10.66 %	10.66%
	Overall Precision	36.05 %	36.05%	36.05%	36.05 %	36.05%	36.05%	36.05 %	36.05%	36.05%	36.05 %	36.05%	36.05%	36.05 %	36.05%

	Overall Recall (This Is Referd As Accuracy In Genie)	63.73%	63.73%	63.73%	63.73%	63.73%	63.73%	63.73%	63.73%	63.73%	43.13%	43.13%	43.13%	43.13%	42.70%
	Overall False-Positive Rate	18.13%	18.13%	18.13%	18.13%	18.13%	18.13%	18.13%	18.13%	18.13%	14.22%	14.22%	14.22%	14.22%	14.32%
	Overall Precision	63.73%	63.73%	63.73%	63.73%	63.73%	63.73%	63.73%	63.73%	63.73%	43.13%	43.13%	43.13%	43.13%	42.70%
	Overall Specifity (True-Negative Rate)	81.87%	81.87%	81.87%	81.87%	81.87%	81.87%	81.87%	81.87%	81.87%	85.78%	85.78%	85.78%	85.78%	85.68%
Ax	Overall Accuracy	63.09%	62.80%	63.09%	63.09%	63.38%	62.09%	63.09%	62.95%	63.09%	71.93%	71.93%	71.33%	71.93%	71.59%
	Overall Recall (This Is Referd As Accuracy In Genie)	44.64%	44.21%	44.64%	44.64%	45.06%	43.13%	44.64%	44.42%	44.64%	29.83%	29.83%	28.33%	29.83%	28.97%
	Overall False-Positive Rate	27.68%	27.90%	27.68%	27.68%	27.47%	28.43%	27.68%	27.79%	27.68%	17.54%	17.54%	17.92%	17.54%	17.76%
	Overall Precision	44.64%	44.21%	44.64%	44.64%	45.06%	43.13%	44.64%	44.42%	44.64%	29.83%	29.83%	28.33%	29.83%	28.97%
	Overall Specifity (True-Negative Rate)	72.32%	72.10%	72.32%	72.32%	72.53%	71.57%	72.32%	72.21%	72.32%	82.46%	82.46%	82.08%	82.46%	82.24%
Ay	Overall Accuracy	67.10%	67.10%	67.10%	67.10%	67.10%	67.10%	67.10%	67.10%	67.10%	73.22%	71.85%	71.07%	73.30%	71.76%
	Overall Recall (This Is Referd As Accuracy In Genie)	50.64%	50.64%	50.64%	50.64%	50.64%	50.64%	50.64%	50.64%	50.64%	33.05%	29.61%	27.68%	33.26%	29.40%
	Overall False-Positive Rate	24.68%	24.68%	24.68%	24.68%	24.68%	24.68%	24.68%	24.68%	24.68%	16.74%	17.60%	18.08%	16.68%	17.65%
	Overall Precision	50.64%	50.64%	50.64%	50.64%	50.64%	50.64%	50.64%	50.64%	50.64%	33.05%	29.61%	27.68%	33.26%	29.40%
	Overall Specifity (True-Negative Rate)	75.32%	75.32%	75.32%	75.32%	75.32%	75.32%	75.32%	75.32%	75.32%	83.26%	82.40%	81.92%	83.32%	82.35%
Az	Overall Accuracy	73.68%	73.68%	73.68%	73.68%	73.68%	73.68%	73.68%	73.68%	73.68%	77.34%	77.34%	77.34%	77.34%	77.42%
	Overall Recall (This Is Referd As Accuracy In Genie)	60.52%	60.52%	60.52%	60.52%	60.52%	60.52%	60.52%	60.52%	60.52%	43.35%	43.35%	43.35%	43.35%	43.56%
	Overall False-Positive Rate	19.74%	19.74%	19.74%	19.74%	19.74%	19.74%	19.74%	19.74%	19.74%	14.16%	14.16%	14.16%	14.16%	14.11%
	Overall Precision	60.52%	60.52%	60.52%	60.52%	60.52%	60.52%	60.52%	60.52%	60.52%	43.35%	43.35%	43.35%	43.35%	43.56%
	Overall Specifity (True-Negative Rate)	80.26%	80.26%	80.26%	80.26%	80.26%	80.26%	80.26%	80.26%	80.26%	85.84%	85.84%	85.84%	85.84%	85.89%
Ar	Overall Accuracy	60.94%	59.37%	59.37%	34.76%	58.37%	60.37%	60.94%	60.09%	60.94%	71.76%	71.16%	70.56%	71.76%	71.50%
	Overall Recall (This Is Referd As Accuracy In Genie)	41.42%	39.06%	39.06%	2.15%	37.55%	40.56%	41.42%	40.13%	41.42%	29.40%	27.90%	26.39%	29.40%	28.76%
	Overall False-Positive Rate	29.29%	30.47%	30.47%	48.93%	31.22%	29.72%	29.29%	29.94%	29.29%	17.65%	18.03%	18.40%	17.65%	17.81%

	Rate)	%			%			%			%			%	
Tot. Driving Experience	Overall Accuracy	72.10%	71.93%	72.10%	72.10%	72.10%	72.19%	72.10%	72.10%	72.10%	72.10%	72.10%	72.10%	72.10%	72.10%
	Overall Recall (This Is Referd As Accuracy In Genie)	30.26%	29.83%	30.26%	30.26%	30.26%	30.47%	30.26%	30.26%	30.26%	30.26%	30.26%	30.26%	30.26%	30.26%
	Overall False-Positive Rate	17.44%	17.54%	17.44%	17.44%	17.44%	17.38%	17.44%	17.44%	17.44%	17.44%	17.44%	17.44%	17.44%	17.44%
	Overall Precision	30.26%	29.83%	30.26%	30.26%	30.26%	30.47%	30.26%	30.26%	30.26%	30.26%	30.26%	30.26%	30.26%	30.26%
	Overall Specifity (True-Negative Rate)	82.56%	82.46%	82.56%	82.56%	82.56%	82.62%	82.56%	82.56%	82.56%	82.56%	82.56%	82.56%	82.56%	82.56%
Driving Investigated Route	Overall Accuracy	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%	85.06%
	Overall Recall (This Is Referd As Accuracy In Genie)	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%
	Overall False-Positive Rate	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%	9.33%
	Overall Precision	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%	62.66%
	Overall Specifity (True-Negative Rate)	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%	90.67%
Bus	Overall Accuracy	67.70%	68.03%	67.92%	69.64%	68.88%	68.99%	69.64%	69.53%	69.64%	67.70%	69.10%	69.53%	69.64%	68.45%
	Overall Recall (This Is Referd As Accuracy In Genie)	35.41%	36.05%	35.84%	39.27%	37.77%	37.98%	39.27%	39.06%	39.27%	35.41%	38.20%	39.06%	39.27%	36.91%
	Overall False-Positive Rate	21.53%	21.32%	21.39%	20.24%	20.74%	20.67%	20.24%	20.31%	20.24%	21.53%	20.60%	20.31%	20.24%	21.03%
	Overall Precision	35.41%	36.05%	35.84%	39.27%	37.77%	37.98%	39.27%	39.06%	39.27%	35.41%	38.20%	39.06%	39.27%	36.91%
	Overall Specifity (True-Negative Rate)	78.47%	78.68%	78.61%	79.76%	79.26%	79.33%	79.76%	79.69%	79.76%	78.47%	79.40%	79.69%	79.76%	78.97%
MODEL	**Overall Accuracy**	74.73%	74.58%	74.65%	73.45%	74.69%	74.71%	74.87%	74.81%	74.87%	76.09%	75.98%	75.86%	80.16%	79.97%
	Overall Recall	46.55%	46.24%	46.37%	43.83%	46.47%	46.50%	46.85%	46.72%	46.85%	40.23%	39.95%	39.65%	40.56%	39.98%
	Overall False-Positive Rate	16.54%	16.64%	16.60%	17.39%	16.57%	16.56%	16.45%	16.49%	16.45%	14.94%	15.01%	15.09%	11.91%	12.02%
	Overall Precision	46.55%	46.24%	46.37%	43.83%	46.47%	46.50%	46.85%	46.72%	46.85%	40.23%	39.95%	39.65%	40.56%	39.98%
	Overall Specifity (True-Negative Rate)	83.46%	83.36%	83.40%	82.61%	83.43%	83.44%	83.55%	83.51%	83.55%	85.06%	84.99%	84.91%	88.09%	87.98%