Developing Accident-Speed Relationships Using a New Modelling Approach

by

Maria-Ioanna Imprialou

A Doctoral Thesis

Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University

August, 2015

Abstract

Changing speed limit leads to proportional changes in average speeds which may affect the number of traffic accident occurrences. It is however critical and challenging to evaluate the impact of a speed limit alteration on the number and severity of accidents due primarily to the unavailability of adequate data and the inherent limitations of existing approaches. Although speed is regarded as one of the main contributory factors in traffic accident occurrences, research findings are inconsistent. Independent of the robustness of their statistical approaches, accident frequency models typically use accident grouping concepts based on spatial criteria (e.g. accident counts by link termed as a link-based approach). In the link-based approach, the variability of accidents is explained by highly aggregated average measures of explanatory variables that may be inappropriate, especially for time-varying variables such as speed and volume. This thesis re-examines accident-speed relationships by developing a new accident data aggregation method that enables improved representation of the road conditions just before accident occurrences in order to evaluate the impact of a potential speed limit increase on the UK motorways (e.g. from 70 mph to 80 mph).

In this work, accidents are aggregated according to the similarity of their pre-accident traffic and geometric conditions, forming an alternative accident count dataset termed as the condition-based approach. Accident-speed relationships are separately developed and compared for both approaches (i.e. link-based and condition-based) by employing the reported annual accidents that occurred on the Strategic Road Network of England in 2012 along with traffic and geometric variables. Accident locations were refined using a fuzzy-logic-based algorithm designed for the study area with 98.9% estimated accuracy. The datasets were modelled by injury severity (i.e. fatal and serious or slight) and by number of vehicles involved (i.e. single-vehicle and multiple-vehicle) using the multivariate Poisson lognormal regression, with spatial effects for the link-based model under a full Bayesian inference method.

The results of the condition-based models imply that single-vehicle accidents of all severities and multiple-vehicle accidents with fatal or serious injuries increase at higher speed conditions, particularly when these are combined with lower volumes. Multiple-vehicle slight injury accidents were not found to be related with higher speeds, but instead with congested traffic. The outcomes of the link-based model were almost the opposite; suggesting that the speed-accident relationship is negative. The differences between the results reveal that data aggregation may be crucial, yet so far overlooked in the methodological aspect of accident data analyses. By employing the speed elasticity of motorway accidents that was derived from the calibrated condition-based models it has been found that a 10 mph increase in UK motorway speed limit (i.e. from 70 mph to 80 mph) would result in a 6-12% increase in fatal and serious injury accidents and 1-3% increase in slight injury accidents.

Acknowledgements

Although only my name appears on the cover of this thesis, it would not be possible to complete this work without the contribution of many people. I would like to take this opportunity to thank all those who have supported me throughout the development of this project.

Foremost, I would like to express my gratitude to my supervisors, Professor Mohammed Quddus and Dr David Pitfield for their immense support and guidance during the last three years. Professor M. Quddus has been an excellent mentor to me who dedicated time and effort in helping me to develop my ideas and also motivated and guided my broader development as a researcher. Dr D. Pitfield has been always there to listen my concerns and provided me with sound advice and encouragement every time I needed it.

I would like to thank Dr. Dominique Lord for providing me with useful feedback during my two-week student visit in the Zachry Department of Civil Engineering of Texas A&M University, US. Moreover, I thank Professor Lin Zhao and his research team as well as Dr. Liang Li for hosting me as an exchange student in the School of Automation in Harbin Engineering University, China, where I developed a part of my research. I am also grateful to Professors Benjamin Heydecker and Mike Maher from University College London for their valuable comments on my project.

My sincere thanks also go to Mr Martin Lynch, Mr Stuart Lovatt, Ms Elizabeth Martin and Mr Roy Jones from Highways England, Ms Ann Hayes from Capita, Dr. Ramesh Perera and Mr Chris Page from AECOM for providing me with the necessary data for my analysis and their support during the course of this work.

Special thanks to my parents, my sister Martha and my partner (and proofreader of this thesis!) Anastasis for their patience, love and constant support. I also thank all my fellow researchers from the Research Hub and especially my good friends Dr Efi Spentzou, Elli Nikolaidou, Dr Ali Alsaygh, Nafsika Drossou, Vivi Vatougiou and Stavros Kassinos for making my PhD experience much more enjoyable and interesting.

Last but not least, I must acknowledge my former supervisor during my studies at the National Technical University of Athens, Matthew Karlaftis (1969-2014) who believed in my potential as a researcher, inspired me and truly encouraged me to pursue this doctoral degree. I dedicate this thesis to his memory.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Road transport infrastructure is an important growth and productivity indicator for national economies (Mačiulis et al., 2009). Despite the undoubtable benefits to society, road transport systems have also negative social and economic impact. Traffic congestion, pollution (i.e. emissions and noise) and accidents are defined as the main externalities of road transport systems (Maddison et al., 1996). Relative comparisons between these three problems are not particularly meaningful as all of them have different and severe impact on road transport systems. Congestion mainly affects road networks' operation, causing disruption and delays, pollution is linked with environmental changes that have adverse consequences on human health and traffic accidents, which are the focus of this thesis, have impact on the physical integrity of road users and the networks' operation.

Accidents are defined as unwanted or unintended sudden events or a specific chain of such events which have harmful consequences (ITF et al., 2009). They cause serious traffic delays, congestion and property damage but more importantly they are linked with road traffic injuries that remain an unsolved global public health problem (WHO, 2013). The multiple problems that are linked with accidents and their inherent complexity make them one of the most challenging problems that policy makers and other stakeholders have to address.

During 2014 in the UK there were 194,477 reported traffic accidents and 12.6% of them had at least one killed or seriously injured casualty (Department for Transport, 2015*b*). A significant proportion of these accidents occurred on the motorway network. More specifically, although the length of motorways is only 1% of the total road network of the country, motorways account for almost 5% of accidents on the entire UK road network (Department for Transport, 2015*b*,*c*). This is certainly related with the fact that motorways carry approximately 21% of the total road traffic but it might be also due to the presence of potential accident triggering factors such as speed.

There is an endless list of potential accident contributory factors that are related either to the road environment or the users. Traffic conditions, road configuration, weather, driver demographic characteristics are only a few of the factors that are believed to be linked with accidents. The relationships between these factors and accidents are examined from different perspectives (e.g. behaviour, external conditions, infrastructure) especially since the middle of the 20^{th} century; however, the complexity and individuality of road accidents make the explanation of these phenomena ambiguous.

Speed, along with alcohol consumption and failure to use a seatbelt, are the top three factors related with 65% of fatal accidents in the UK (Clarke et al., 2010). Speed has been reported to be a contributory factor for 36% of fatal and 14% of all accidents in the UK during 2013 (Department for Transport, 2014). Driving with excessive speed is a potential contributory factor that is particularly interesting for high-speed road environments such as motorways. The percentage of fatal accidents on high speed environments is more than double than the corresponding percentage on roads with lower average speeds (Keep, 2013). This fact implies that a causal relationship between serious accidents and speed might exist. To control the negative impact of speeding, traffic authorities set and enforce speed limits on motorways and other main roads. The national motorway speed limit in the UK is 70 mph for all types of vehicles except for goods vehicles with laden weight over 7.5 tonnes and cars that tow caravans or trailers that should not exceed 60 mph (Department for Transport, 2007).

1.2 Problem Statement

Speed limits should not be considered as the target speed, however a great proportion of motorway drivers in the UK systematically exceed the speed limits; in 2014 46% of cars violated the 70 mph speed limit, leading to an annual average speed and an 85th percentile as high as 68 mph and 77 mph respectively (Department for Transport, 2015*a*). One could argue that these figures indicate the necessity of an update to the current speed limit that was firstly set in 1966. It is a fact that many factors related to road safety have been changed over time. For instance, vehicle technology, vehicle-based active and passive safety systems, driver training and education, emergency responses and medical services have been significantly improved. Based on the above and arguing that safety cannot be the only consideration when setting speed limits the Department for Transport expressed their intention to increase motorway speed limits from 70mph to 80 mph in 2011 (Department for Transport, 2011*a*). The reason behind this idea was that a speed limit increase could lead to reduced travel times and less congestion that can be translated to economic benefits for the region.

As expected, this announcement raised questions concerning the possible negative consequences that a regulation like this could bring to the number of traffic accidents on the UK motorways. To assess whether such an increase in the speed limit is sensible it is necessary to estimate its future effect on road safety. The relationship between speed and accidents is the key to quantify the impact of speed limit changes. Increases in speed limit are related with average speed raises (Freedman and Williams, 1992) that could potentially cause more accidents on the network. Previous studies on the impact of speed limit changes primarily concluded that the changes in accident frequency and severity are proportional to the speed limit alterations (e.g. Elvik, 2009). Nevertheless, the individual examination of the relationship of speed with accidents provides less clear results. Speeding is confirmed to be related with higher accident severity but it is not clear whether this is true for accident frequency (e.g. Aarts and Van Schagen, 2006; Kockelman and Ma, 2007; Quddus, 2013). The lack of a generalised conclusion and the absence of British studies on the impact of speed limit increases show that the relationship of speed with accidents should be further explored.

The dissimilar findings of existing research on the relationship of accident frequency with speed may be due to various methodological limitations of accident analyses that reduce the accuracy of representation of the actual circumstances that are related with, and probably caused, accidents. In conventional accident analyses accidents are typically aggregated using spatial criteria such as road links. Link-based analyses use variables that are by default highly aggregated as they represent the conditions on an entire link with one characteristic value (e.g. time-varying measures are usually represented by annual averages). In this way it is likely that the spatial and temporal variations within the link are not captured, making the representation of the pre-accident conditions practically impossible. Moreover, analysing all accident types together may reduce the capability of models to reveal the actual accident contributory factors as those are found to vary between different accident generation processes (Geedipally and Lord, 2010).

The speed limit increase that was proposed by the Department for Transport after the replacement of the Secretary of State for Transport (former Secretary Philip Hammond was replaced by Patrick McLoughlin) did not continue to be a priority and did not reach the implementation stage (Chorley, 2013). Nevertheless, the relationship of speed and traffic accidents is still an interesting and relatively unexplored subject for the motorway network of England. This thesis examines speed-accident relationships on the UK motorway attempting to overcome the current methodological limitations. This includes the exploration of alternative accident data aggregation approaches that will enable the accurate representation of the pre-accident traffic conditions in statistical models and the evaluation of the effects of speed and other contributory factors on different accident types. The results intend to increase the understanding of accident occurrences and the methods for analysing them.

1.3 Research Importance

Although over the last decades accidents have a decreasing trend especially in the Western world, their number is still unacceptably high (WHO, 2013). In addition to the devastating personal losses, the annual loss to society due to accidents in the UK is estimated to be £15 billion (the value for preventing a fatal accident is approximately £2 million) (Department for Transport, 2012*b*). As a consequence, mitigation of traffic accidents remains one of the top priorities of traffic management agencies in the UK and all over the world (e.g. Whitelegg and Haq, 2006). To decrease the number of accidents effectively by introducing new policies or technologies, it is necessary to firstly understand the mechanisms that lead to these events. Road safety research aims to reveal causal relationships and contributory factors of accidents so as to develop targeted preventive measures in the future.

The majority of accidents are related with combinations of human error with defects of the road environment (e.g. Wagenaar et al., 1990). While controlling drivers' errors is not always possible, designing and managing roads in a manner that hazardous situations are avoided is more feasible. Developing road infrastructure of high safety standards is one straightforward way for securing effective accident mitigation. Accident modelling is a key method for achieving this as it provides information on the relationships of road characteristics with accidents. This research project is meaningful because it provides qualitative results on the relationships of several contributory factors with accidents and also provides new methodological insight on accident modelling.

1.4 Aim and Objectives

This thesis aims to examine the relationship between motorway accidents and speed. This will be accomplished through the following objectives:

- To review the impact of speed and other contributory factors on accidents
- To examine existing statistical approaches in accident modelling
- To refine and merge data from multiple sources so as to enhance the quality of the analysis
- To develop accident-speed relationships using a new condition-based modelling approach

- To compare and contrast the results between the conventional and the condition-based modelling approach
- To evaluate the impact of a potential speed limit increase on accidents

1.5 Thesis Outline

This thesis consists of seven chapters. An outline of the chapters is provided below:

- Chapter 2 conducts an extensive literature review on the relationship of accidents with traffic and geometric contributory factors, the main statistical approaches in accident analyses and the existing accident location refinement methods;
- Chapter 3 presents the methodology of this thesis. The chapter starts with the description of an accident mapping algorithm that is applied to the accident data. Following are the two accident aggregation methods that are examined and the statistical models that are employed. The final section of this chapter explains the modelling strategy that was followed in this work;
- Chapter 4 illustrates the accident, traffic and geometric data and the outcomes of the data pre-processing methods that are used before proceeding to the main analysis;
- Chapter 5 shows and explains the results of the statistical models that are developed. The chapter also discusses methodological implications for accident analyses that were derived from the modelling outcomes;
- Chapter 6 provides estimations on the impact of a potential speed limit increase and provides policy recommendations for mitigation of traffic accidents;
- Chapter 7 summarises this research project and outlines its contributions to knowledge and limitations. Finally, some thoughts for future research directions are provided.

Chapter 2

Literature Review of Findings and Methods of Accident Analyses

2.1 Motivation

Accidents impose social and personal costs on drivers, passengers and generally the road network users. Traffic managers and local authorities for at least the last century are working on decreasing the number of road accidents with an emphasis on accidents with casualties (Norton, 2015). Prevention relies on the in-depth understanding of the factors and the mechanisms related with accidents. There are numerous factors associated with driving attitudes, traffic or external conditions (e.g. alcohol consumption, traffic congestion, adverse weather) and often combinations of such factors that may lead to accidents (Brown, 1982; Montella, 2011). The first systematic accident analyses emerged approximately seven decades ago and since then there have been constant significant advances on our understanding of accident causation (Hagenzieker et al., 2014). The randomness and the complexity that characterises road accidents though have not yet permitted a full explanation of these phenomena.

In order to develop an appropriate methodology for this research it is necessary to understand the existing knowledge and to identify the methodological limitations of previous accident analyses. The focus of this literature review is the main outcomes and the quantitative approaches that have been employed in accident modelling. The first part of this chapter outlines the findings of previous research on the relationships of accidents with speed, traffic volume and road geometric features that are the contributory factors that will be examined in this thesis. Following that, there is a discussion on the main statistical approaches that have been used in accident analyses and their strengths and limitations. The final section of this review outlines the significance of reported accident locations in accident analyses and the existing accident mapping techniques.

2.2 Speed and Accidents

Speed is related with a large proportion of traffic accidents (Aarts and Van Schagen, 2006; Clarke et al., 2010). This is explainable considering that speed is a potential accident contributory factor that is always present on the network in contrast to many others that have random (e.g. rainfall) or periodic character (e.g. darkness) (Elvik et al., 2004). As drivers do not always succeed in choosing appropriate speeds, speed regulation measures such as speed limits are necessary because they offer some guidance for correct speed choices (Elvik, 2010). Speed limits can therefore indirectly represent the typical traffic conditions on the roadway. Accepting that an accident-speed relationship exists, speed limit changes are expected to have impact on the safety levels of a road network. Network-level accident-speed relationships have been examined either by estimating the impact of speed limit changes or by developing statistical models that explain the number of accidents as a function of speed as a traffic variable.

High speeds are proven to increase the severity of accidents and there are also indications that they have a negative impact on accident frequency. Although the number of studies that attempted to quantify the accident-speed relationship is significant, the high complexity of accidents and some methodological restrictions have not permitted the generation of a relationship with general acceptance and application. The findings of the most significant studies on this topic are presented in the sections below.

2.2.1 Speed limits and accidents

2.2.1.1 Speed limit increases

The aim of setting speed limits is to maintain the equilibrium between road safety, traffic flow and energy consumption in road networks (TRB, 1998; Department for Transport, 2006). A speed limit does not necessarily represent a speed that is safe at all conditions and is definitely not a target speed. Except for denoting what is legal, properly set speed limits reflect the range of speeds that are considered sensible for a specific road environment according to its characteristics (Department for Transport, 2006).

Setting speed limits requires a thorough data collection and analysis of various factors related with the road environment. The most common factors that are taken into consideration are the design speed, the 85^{th} percentile, the mean speed, the road function, previous accident and enforcement experience, roadway geometry and the impressions a road gives to its users (TRB, 1998). The existence of appropriately set speed limits though does not ensure the keeping of speeds at safe levels; violating the speed limit is one of the most common road traffic law offences (Department for Transport, 2015*a*). As a consequence, speed limit enforcement is particularly important for effective speed regulation (Wilson et al., 2006). Among the numerous different methods for enforcing speed, speed cameras and speed guns are considered to be some of the most efficient. For instance, average speed drop and accident mitigation have been reported at roadway locations that are equipped with speed cameras (e.g. Mountain et al., 2005; Soole et al., 2013; Li et al., 2013).

In some cases, road management authorities decide to use special types of speed limits when it is believed that they will suit better the roads' characteristics and needs. As an example, some roadway sections have variable speed limits that adapt to the external conditions (e.g. weather, traffic) or differential speed limits that apply for some specific types of vehicles such as trucks (TRB, 1998). Since 2005, variable speed limits are implemented on some sections of the Strategic Road Network of England forming the so-called smart motorway network that is continuously expanding (Highways England, 2014). The smart motorway is a dynamic response to the continuously increased traffic demand that is also cost-effective compared to an expansion of the existing network (Highways Agency, 2010). For that purpose, except for using hard shoulders as lanes variable, mandatory and highly enforced speed limits are used. Variable speed limits are set in order to correspond efficiently to the traffic and environmental conditions on the motorway (Highways Agency, 2006; Department for Transport, 2012*a*).

Speed limits are measures that generally have long-term or even permanent character. Traffic authorities usually decide to change a speed limit only if they consider it as necessary for promoting safety or relaxing congestion. The distribution of speeds on the roadway is typically a function of the posted speed limit. As a consequence speed limit changes will lead to changes in the average driving speeds. Research has shown that these changes are proportional but comparatively moderate and that is possibly because new speed limits reflect better the current speed choices of the users (Rock, 1995). More specifically, the average speed change equals approximately to one quarter to half of the speed limit difference (e.g. Freedman and Williams, 1992; Finch et al., 1994; Baruya, 1998*b*; Aljanahi et al., 1999; Ossiander and Cummings, 2002; Elvik et al., 2004; Vadeby and Forsman, 2010; De Pauw et al., 2014). Assuming that speed is related with accidents, a speed limit change, if all other factors remain unchanged, should have an impact on the number of accidents on a road network and probably on the adjacent roads too (i.e. spill-over effect) (e.g. Garber and Grahman, 1990; Wagenaar et al., 1990). According to a meta-analysis by Elvik et al. (2004) 70.5% (weighted percentage) of the before-after included studies found a proportional change in accidents after a speed limit alteration.

There are quite a few examples from countries all over the world where the speed limits were changed (increased or decreased) and the differences in road accidents shed more light on the effect of speed limits on road safety. To be in line with the aim of this thesis, this review will focus on studies that examined the impact of increased speed limits on accidents.

There are a considerable number of before and after studies that are focused on speed limit changes in the United States. Almost a decade after the general speed limit reduction (to 55 mph) due to the petrol crisis, the US government in 1987 gave the permission to the States that they wished to, to increase their speed limits up to 65 mph. Following this, in 1995 the responsibility of setting speed limits completely returned to the States and from that point until today each State selects its own speed limit (TRB, 1998).

After the 10-mph increase of speed limits (from 55 to 65 mph), in 40 of the States in 1988; most of the state and national studies report a general trend of increased fatal accidents and fatalities (TRB, 1998; Houston, 1999). In their before and after study Brown et al. (1990) found that in the State of Alabama a year after the speed limit increase, 19% more accidents were reported on rural interstates. Wagenaar et al. (1990) applying an ARIMA time series model estimated that there was a 19% increase in fatalities and a 25-40% increase in injuries in rural interstates of Michigan. They suggested that the frequency of fatalities also rose in limited access freeways that did not experience a speed limit change, as a side effect of higher driving speeds in neighbouring roads, the so-called spill-over effect. Using ARIMA models Rock (1995) also suggested the existence of spill-over effects due to increased speed variance in the State of Illinois in addition to a substantial monthly increase of accidents on rural inter-states (345 more accidents and 15 more fatalities).

Baum et al. (1991) examined the impact of the speed limit rise in 40 states using as a reference eight states where the speed limit was unchanged for their before and after analysis. Controlling for the increased vehicle miles travelled and vehicle occupancy, they found a 29% increase in fatalities of rural interstates of the states that adopted higher speed limits (Baum et al., 1991). On the other hand, the eight states that did not apply the measure had a 12% reduction in fatalities that according to the confidence interval of the odds ratio that was used for the estimation was not found to be statistically significant. However, when accidents were examined disaggregated by state, the impact of speed limit increases was not uniform. Although the median of fatal accidents increased by approximately 15%, Garber and Grahman (1990) found that only 28 out of the 40 states experienced more fatal accidents. Chang et al. (1993) suggested that the increase was more significant in the comparatively smaller states noting the existence of several unobserved exogenous factors that contribute to these changes.

Arguing that the majority of the before and after studies focus on local effects only

Lave and Elias (1994) opposed the view that speed limit increases have a negative impact on road safety. Setting as their dependent variable state-wide instead of interstate fatalities, they estimated a 3.4% to 5.1% decline. They explained that this might was the effect of traffic diversion (i.e. faster drivers tended to avoid non-interstates), more efficient distribution of resources for road safety by the authorities and speed variations decrease despite the fact that such data were not available to them to confirm these explanations.

The effect of the speed limit relaxation in several main roads in Hong Kong was negative for accidents of all severities. The impact on fatal and serious accidents though was considerably more significant when the speed limit increased from 70 to 80 km/h ($\sim$ 43.50 to 49.71 mph) (fatal and serious accidents increased by 18%) than from 50 to 70 km/h ($\sim$ 31.07 to 43.50 mph) (fatal and serious accidents increased by 1%)(Wong et al., 2005). The European examples of speed limit increases in rural highways and motorways have also shown with some consistency that the impact of this measure is rather negative for road safety (OECD/ECMT, 2006). A 10km/h ($\sim$ 6.2 mph) speed limit increase lead to 13% more fatalities in Hungary and in considerably more personal injury accidents in Sweden and Denmark (OECD/ECMT, 2006; Hels, 2012; Vadeby, 2015). The impact on speed limits increase on Greek motorways varied based on their geometric and traffic characteristics. More specifically, motorways with lower geometric standards had more fatal accidents, but this was not true for motorways with better geometrical features and relatively lower traffic (Yiannis et al., 2015) .

2.2.1.2 Meta-analyses

Regardless of their results, the common feature of the above studies is that they examine local effects of speed limit changes having available accident data before and after the period of the measure implementation. In order to take strategic policy decisions though, it is useful to know in advance the potential impact of a speed limit increase. There have been several attempts in the literature to define general rules for the impact of speed limit changes; most of them are based on the combination of the outcomes of previous studies (i.e. meta-analyses).

The effect of changing speeds after the implementation of traffic measures on road safety was examined in-depth in the meta-analysis conducted by Elvik et al. (2004). The study primarily aimed to identify a generic relationship that could estimate the manner in which speed changes influence the number and severity of traffic accidents and additionally to test to what extent the Power Model (Nilsson, 2004) describes this relationship. Nilsson (2004) introduced the Power Model as a group of six power functions (see equations 2.1-2.6) that estimate the expected number of accidents or casualties by accident severity after a certain change in mean speeds on a road network assuming that the number of occurrences and casualties is always proportional to speed. Although the power functions are obviously quite simplistic to provide very accurate results, Nilsson (2004) supported that they are applicable for all network environments.

$$\text{Fatal Accidents: } y_1 = \left(\frac{v_1}{v_0}\right)^4 y_0 \tag{2.1}$$

$$\text{Fatalities: } z_1 = \left(\frac{v_1}{v_0}\right)^4 y_0 + \left(\frac{v_1}{v_0}\right)^8 (z_0 - y_0) \tag{2.2}$$

$$\text{Fatal and Serious Accidents: } y_1 = \left(\frac{v_1}{v_0}\right)^3 y_0 \tag{2.3}$$

$$\text{Fatal and Seriously Injured Casualties: } z_1 = \left(\frac{v_1}{v_0}\right)^3 y_0 + \left(\frac{v_1}{v_0}\right)^6 (z_0 - y_0) \tag{2.4}$$

$$\text{All Accidents: } y_1 = \left(\frac{v_1}{v_0}\right)^2 y_0 \tag{2.5}$$

$$\text{All Casualties: } z_1 = \left(\frac{v_1}{v_0}\right)^2 y_0 + \left(\frac{v_1}{v_0}\right)^4 (z_0 - y_0) \tag{2.6}$$

Where:

y_0 and y_1: the number of accidents before and after the change respectively

v_0 and v_1: the mean speed before and after the change respectively

z_0 and z_1: the casualties before and after the change respectively.

For their meta-analysis Elvik et al. (2004) identified 98 valid projects from all over the world conducted between 1960 and 2004. The topic of these studies was mainly about the consequences due to mean speed changes that have occurred due to a new road safety measure. As a part of their exploratory analysis Elvik et al. (2004), after standardising the explanatory power of the studies, found that average speeds were decreased 95% after a speed limit reduction and this was consistent with the Power model, which means that accident number and speed changed proportionally. The respective figure for the estimates of increasing mean speeds was 70.5%. Applying a meta-regression analysis they estimated new exponents applicable for equations 2.1-2.6 that are summarised in Table 2.1. The authors appeared to be confident about the results of their meta-analyses suggesting that the existing limitations such as study inclusion biases, omission of other factors are unlikely to have drastically affected the outcomes.

Figure 2.1: Best estimates of exponents for different accident or injury severity groups (source: Elvik et al., 2004)

Accident or injury severity	exponent	interval
Fatalities	4.5	(4.1 – 4.9)
Seriously injured road user	3.0	(2.2 – 3.8)
Slightly injured road user	1.5	(1.0 – 2.0)
All injured road users (severity not stated)	2.7	(0.9 – 4.5)
Fatal accidents	3.6	(2.4 – 4.8)
Serious injury accidents	2.4	(1.1 – 3.7)
Slight injury accidents	1.2	(0.1 – 2.3)
All injury accidents (severity not stated)	2.0	(1.3 – 2.7)
Property-damage-only accidents	1.0	(0.2 – 1.8)

Source: TØI report 740/2004

Nevertheless, the Power Model has a drawback that is unavoidably reflected on its results. The use of power functions is straightforward and transferable but the exponents provided are independent of the baseline speed and that might lead to inaccurate estimations (Hauer and Bonneson, 2006). As an example, the estimated proportional changes on the frequency of accidents are equal when mean speed increases from 30mph to 33mph and from 60mph to 66mph, because in both the cases there is a 10% increase of speeds. Hauer and Bonneson (2006) using the data from the meta-analysis by Elvik et al. (2004) proved the dependence of accident frequency on the baseline speed and examined whether other factors apart from mean speeds should be taken into consideration. Hauer and Bonneson (2006) and Hauer (2009) also developed new exponential prediction models that incorporated manoeuvre time and distance for collision avoidance.

Elvik (2009) suggested that the power model is simpler than the exponential and modified the power model in a way that the initial speeds were indirectly taken into consideration. Using an updated and richer dataset he estimated two different sets of exponents for high and low speeds that can be seen in Table 2.2. Although this approach provides improved results, the baseline speeds are still not taken into account and that is why Elvik (2013) re-parametrised these exponents in order to compare it with the exponential model. The exponential model had slightly better fit for injury and PDO accidents but for accidents with fatalities the power model fitted better. Elvik (2013) concluded that the analysis supports exponential models as they are more suitable for modelling the impact of speed changes on road safety. Despite that, Elvik's (2009) approach due to its simplicity is used as a point of reference by many transportation agencies in the world (e.g. Hels, 2012; Vadeby, 2015).

Figure 2.2: Best estimates of exponents for different accident or injury severity groups by road type (source: Elvik, 2009)

	Summary estimates of exponents by traffic environment					
	Rural roads/freeways		Urban/residential roads		All roads	
Accident or injury severity	**Best estimate**	**95 % confidence interval**	**Best estimate**	**95 % confidence interval**	**Best estimate**	**95 % confidence interval**
Fatal accidents	4.1	(2.9, 5.3)	2.6	(0.3, 4.9)	3.5	(2.4, 4.6)
Fatalities	4.6	(4.0, 5.2)	3.0	(-0.5, 6.5)	4.3	(3.7, 4.9)
Serious injury accidents	2.6	(-2.7, 7.9)	1.5	(0.9, 2.1)	2.0	(1.4, 2.6)
Seriously injured road users	3.5	(0.5, 5.5)	2.0	(0.8, 3.2)	3.0	(2.0, 4.0)
Slight injury accidents	1.1	(0.0, 2.2)	1.0	(0.6, 1.4)	1.0	(0.7, 1.3)
Slightly injured road users	1.4	(0.5, 2.3)	1.1	(0.9, 1.3)	1.3	(1.1, 1.5)
Injury accidents – all	1.6	(0.9, 2.3)	1.2	(0.7, 1.7)	1.5	(1.2, 1.8)
Injured road users – all	2.2	(1.8, 2.6)	1.4	(0.4, 2.4) #	2.0	(1.6, 2.4)
PDO- accidents	1.5	(0.1, 2.9)	0.8	(0.1, 1.5)	1.0	(0.5, 1.5)

Confidence interval specified informally
Source: TØI-report 1034/2009

2.2.2 Speed and speed variance

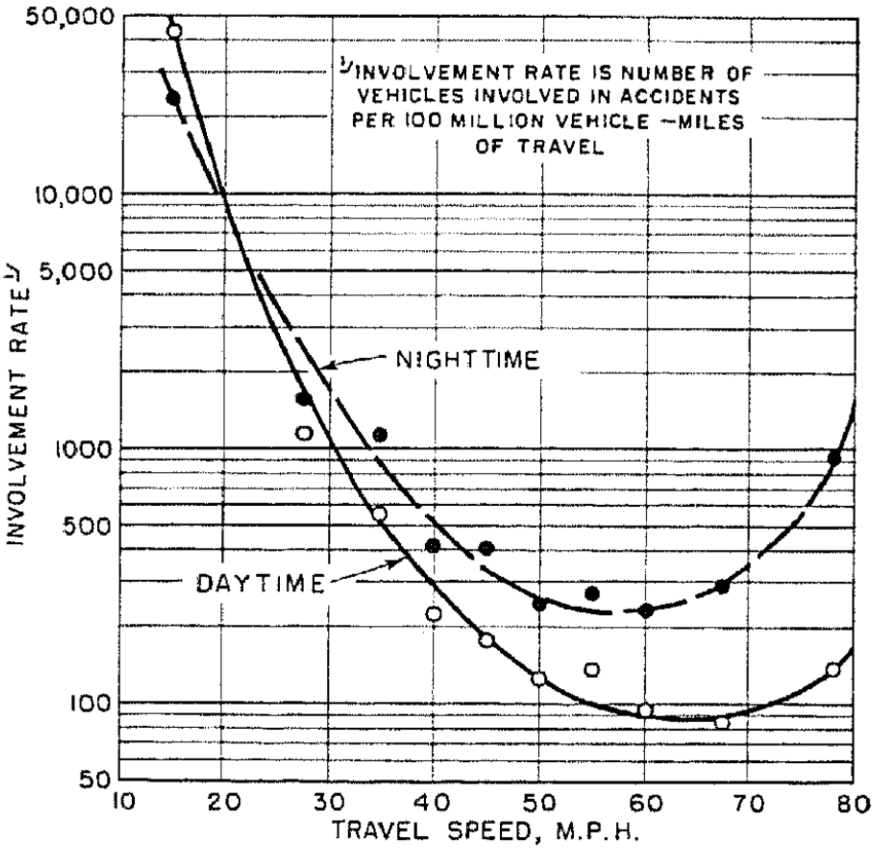
Meta-analyses' results are useful for identifying general data patterns, but are not accurate enough for predicting the effect of a speed limit increase on a particular road network as they cannot take into account the area-specific characteristics (geographic, cultural etc.) that may differentiate the outcomes. Consequently, to predict the impact of a potential speed limit increase on a road network it is necessary to define its accident-speed relationship. As has already been discussed, the majority of the before and after studies report proportional changes in accident frequency following speed limit alterations (e.g. Elvik et al., 2004; Aarts and Van Schagen, 2006). This effect is always attributed to the increase of average speeds on the roadway; however, the individual examination of the relationship of speed with accidents does not always support this idea.

Based on the amount of the kinetic energy that is released during a collision ($E_{Kinetic} = \frac{mV^2}{2}$), accidents that occur under high speed conditions are definitely more likely to lead to more serious outcomes (e.g. Joksch, 1993; Kloeden et al., 1997; Aarts and Van Schagen, 2006; Pei et al., 2012). High travel speeds are also associated with many accident triggering factors such as lower reaction times, longer decisions, breaking and stopping distances, reduction of manoeuvrability and increased possibilities of manipulation error, loss of control and exceeding the critical speed on a curve (Solomon, 1964; Godwin, 1984; Hale, 1990; Fildes and Lee, 1993; Patterson et al., 2000; Navon, 2003; Aarts and Van Schagen, 2006). On the other hand, higher speeds are also related with more uniform distribution of speeds (i.e. lower speed variance) that is considered to be beneficial for road safety (Lave, 1985; Graves et al., 1993; Navon, 2003).

There is a considerable amount of research on the accident-speed relationship but several points of disagreement between studies. Most of the studies found driving speeds to be linearly or exponentially related to accidents (e.g. Fildes et al., 1991; Baruya and Finch, 1994; Kloeden et al., 1997; Quimbly et al., 1999; Taylor et al., 2000; Kloeden et al., 2002). A few studies contradicted the common belief though, proposing that speed is inversely proportional with accidents (e.g. Baruya, 1998*a*; Stuster, 2004) and others reported statistically insignificant relationships (Lave, 1985; Garber and Gadiraju, 1989). Some of the most recent papers that explored the impact of speed on accidents using advanced statistical modelling did not find a statistically significant relationship between speed and accidents (e.g. Garber and Ehrhart, 2000; Kockelman and Ma, 2007; Quddus, 2013). Pei et al. (2012) attempted to explain the results' inconsistencies suggesting that the estimated accident-speed relationship by models strongly depends on the selected measure of exposure; the relationship was shown to be negative for distance-based exposure (i.e. vehicle miles travelled) but positive for time-based exposure (i.e. vehicle hours travelled). The inconsistent results between research papers in fact might be related with a variety of methodological and data limitations that do not permit the accurate evaluation of the speed-accident relationship. This will be further discussed in section 2.4.

The work of Solomon (1964) on the relationship of speed and accidents is one of the widely cited, replicated and criticised studies; maybe more than any other relevant study. This is because one of its main outcomes was that speed dispersion and not speed is related with accident frequency; contrasting what was believed so far and the findings of many subsequent studies. Speed dispersion (or speed variance) is defined as speed differences between or within lanes between individual vehicles or in a road section level (Aarts and Van Schagen, 2006). Speed differences lead to more vehicle passes that are related with more accident prone interactions (Lave, 1985; Navon, 2003) In his case-control study Solomon (1964) employed 10,000 accident reports that included information on the speed before the accident (based on police officers' estimations and a combination of other relevant data in accident reports). The control data included spot speed observations for 290,000 drivers that were measured with concealed, speed measuring devices. Solomon (1964) comparing the two speed distributions that stemmed from the case and the comparison groups (defining an accident involvement rate) observed that that accident-involved drivers were travelling at speeds that deviated from the average speed and more specifically were considerably lower than it. Accident involvement rates were calculated by dividing the number of accident-involved drivers by the respective vehicle-miles of travel and it was plotted against speed forming a U-shaped curve (see Figure 2.3) that according to Solomon (1964) represents the chance of a driver being involved in an accident as a function of their speed. It can be noticed that the involvement rate reaches a peak at very low speeds (15 mph) and its lowest value is for a speed approximately 5 mph above the average speed that was 50 mph. From that point and onwards, involvement rate increases constantly. As a consequence, Solomon (1964) suggested that the greater the variation (positive or negative) from the average in a vehicle's speed, the more likely it is for this vehicle to be involved in a road accident or differently put, relatively high driving speed is safer than either low or excessively high speed driving. His finding for accident severity was different and consistent with literature, as he suggested that accident severity and speed are directly proportional.

Figure 2.3: Accident involvement rate as a function of driving speeds for day (solid line) and night (dashed line) (source: Solomon, 1964).



Solomon's (1964) work was followed by some similar studies, in terms of methodology and results, that supported the existence of a U-shaped curve between the chance of accident involvement and speed (Cirillo, 1968; Munden, 1967; RTI, 1970). The validity of this result however is questionable if the limitations of these studies are considered. The inclusion of vehicles whose speed was not chosen such as turning vehicles and vehicles in congested conditions (46% of the sample) is one factor that could distort the results of these studies (Hauer, 1971; Frith and Patterson, 2001; RTI, 1970). Also, the use of the ratio of distributions of data that originate from dissimilar in terms of accuracy measurement methods (i.e. vehicle speeds before the accidents and driving speeds) by default leads to a U-shaped curve even if the two distributions were equal (Hauer, 2009).

Speed variance has been examined as a potential contributory factor several times and with different approaches ever since. One of the main challenges is the quantification of speed variance (Wang et al., 2013); speed differences cannot be directly measured like other traffic variables (e.g. volume, mean speed) so the researchers employed several different surrogate measures. For instance, Lave (1985) used the difference of the mean speed and the 85^{th} percentile of the speed distribution and estimated a positive relationship between speed variations and accidents. This finding agrees with the majority of studies where speed variance was defined as the standard deviation of speed (Baruya and Finch, 1994; Baruya, 1998*a*; Taylor et al., 2000; Quddus, 2013). Pei et al. (2012) however reported that there is not a statistically significant relationship between the standard deviation of speed and accidents. This is consistent with the findings of Kockelman and Ma (2007) who used more complex expressions for defining separately speed variance between and across lanes aiming to reflect disaggregate estimates of instant variation. All these results should be seen with some caution though as they probably reflect default mathematical properties of the data rather than actual causal relationships between speed variance and accident frequency (Davis, 2002). In fact, the mechanism of the impact of speed variance cannot be explicitly explained until individual vehicle-level second-by-second data are available (Kockelman and Ma, 2007).

2.3 Other Explanatory Variables

Speed is not an independent traffic characteristic; vehicles' speed is usually a function of the surrounding traffic conditions and infrastructure features. Therefore the relationship of speed with accidents cannot be defined without controlling for the simultaneous effect of other road characteristics (Aarts and Van Schagen, 2006). Various traffic characteristics have been examined and found that they have significant impact in traffic accidents such as traffic flow, traffic density, vehicle-capacity ratio and others. The effects of road geometry (e.g. curvature, gradient) and construction quality (e.g. pavement conditions) on road safety have also been widely studied in the literature. The relationship of traffic volume, horizontal and vertical alignment and the number of lanes will be mainly discussed here as these are variables that are included in the available datasets.

2.3.1 Traffic characteristics and accidents

Traffic flow is considered to be one of the most important accident precursors in the literature. Naturally, the number of vehicles on the roadway is directly proportional with the number of vehicle interactions that can potentially lead to collisions (Navon, 2003). Most studies represent traffic flow with the Annual Average Daily Traffic (i.e. AADT) or Average Daily Traffic (i.e. ADT) mainly because these variables are normally available in traffic datasets. AADT has been found by a large number of researchers to have a proportional relationship with accident frequency (e.g. Miaou and Lum, 1993; Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000; Chang, 2005; Anastasopoulos and Mannering, 2009), that in other words means that the busier a road is, the more accidents are expected to occur on it. In addition, studies that examined Levels of Service (LOS) as a surrogate measure for traffic conditions and density, confirmed that accident frequency is the highest for the lowest LOS and vice versa (Frantzeskakis and Iordanis, 1987).

There are however some exceptions to these findings; the study of Garber and Ehrhart (2000) reports slight decreases in accident rates when the flow per lane increases for the range of 90 to 100 vehicles per hour per lane. Gwynn (1967) and Ceder (1982) also found that the relationship of traffic flow and accident rates form a U-shaped curve meaning that accident probability rises at very high or very low traffic conditions and it is at its lowest levels under moderate traffic. The form of a U-shaped curve also takes the relationship of the vehicle-capacity ratio (i.e. v/c ratio) that is also a measure that represents, probably with higher precision, the traffic flow conditions on the roadway (Zhou and Sisiopiku, 1997).

At first glance the statement that the traffic flow and accidents are proportional seems valid, but it has been proven to be too generic and not representative of all accident types. There are studies that show that the effect of traffic volume has different impacts on different collision types. Specifically, single vehicle accidents have been found to be related to lower traffic flows while the number of multiple vehicle accidents increases at higher traffic volumes (Gwynn, 1967; Ceder, 1982; Martin, 2002; Qin et al., 2004; Lord et al., 2005; Kim et al., 2006; Ye et al., 2009; Bham et al., 2012). This finding can be the explanation for the U-shaped curves that were found to represent the relationship of accidents with traffic flow when all accidents are examined aggregated (Ceder, 1982; Martin, 2002). The left part of the U-curve might be related with single vehicle accidents that occur at low flow conditions and the right one refers to multiple vehicle accidents that are more likely to occur under more congested conditions.

Although at least for multiple vehicle collisions accident frequency and traffic flow have a positive relationship, more serious accidents tend to occur under lower flows and especially at off-peak times (Martin, 2002). This is probably because lower flows are indirectly related with high speed and speed variance (as traffic builds up) that are also thought to be significant accident precursors (e.g. Garber and Ehrhart, 2000; Elvik et al., 2004). The fact that the latter two traffic characteristics are considered to trigger accident frequency complicates their examination along with traffic flow and the interpretation of the findings. Recent research on real-time accident prediction models suggests that accidents occur due to turbulences of the traffic flow under particular combinations of traffic conditions attempting to provide a more complete definition for the role of traffic flow in accidents (e.g. Abdel-Aty and Pande, 2005; Hossain and Muromachi, 2013).

2.3.2 Road geometry and accidents

The geometrical characteristics of a road are thought to be related with accident frequency and severity in various ways (AASHTO, 2010). Particular road configurations can contribute to restricted visibility (e.g. sharp curves), higher stopping distances (e.g. downhill sections) or even indirectly encourage inappropriate driving behaviour (e.g. speeding on long straight segments). The effect of the road design on accidents is a very popular subject in road safety research due to the increased interest from the road managing agencies and data availability. It is a fact that in some cases the results of the studies are not always consistent. This is because of the variations in the land, construction, traffic and demographic characteristics (e.g. Haynes et al., 2007, 2008) that are difficult to be controlled and the differences in the statistical approaches and the surrogate measures that are used to represent geometry. The core literature findings for the relationship of curvature, gradient and number of lanes are presented here.

2.3.2.1 Curvature

One of the commonly accepted ideas about road geometrical characteristics is that sharp curves are related with higher accident rates (e.g. Miaou et al., 1992; Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000; Caliendo et al., 2007; Anastasopoulos and Mannering, 2009; Gitelman et al., 2014) and more severe injuries (Ma and Kockelman, 2006). Curves with small radiuses that often do not provide the necessary sight distance, are related with more driving errors and raise the lateral acceleration that can cause loss of vehicle control (Peters and Iagnemma, 2009). Chang (2005) however found that accident likelihood reduces at sharper curves. His counterintuitive result was explained as the effect of more careful driving when the road configuration is more challenging.

Radius is an important but not the only determinant of hazardous horizontal alignment; the frequency of curves on a road segment has also been found to be related with the number of accidents. According to Milton and Mannering (1998) the higher the length of a straight section just before a curve, the higher is the accident frequency. That is because when drivers traverse long tangent segments they are less likely to expect curves and might enter the curve with inappropriate speed. Shankar et al. (1995) found that fewer curves per mile are related with more serious accidents. This explains to a certain extent the finding of Caliendo et al. (2007) who found that apart from sharp curves, very long straight segments are also hazard prone. Some area-wide studies proposed that areas with more curves experience overall less accidents (Haynes et al., 2007, 2008; Wang et al., 2009*b*; Li et al., 2015). Wang et al. (2009*b*) explained this outcome suggesting that in more curved configurations drivers tend to be more cautious and aware of the road. However, since the sharpness of the curves was not separately available in the data (curvature was expressed with bend density), it is not possible to determine whether that was the effect of the number or the radius of the curves.

According to the findings outlined above, changes in the horizontal alignment should normally lead to some changes in the number of accidents and that has been confirmed from the study of Vogt and Bared (1998) who found that decreased curvature lead to less traffic accidents. In contrast, a panel-data analysis by Noland and Oh (2004) suggested there was not a significant effect on accidents related with area-wide changes neither in the sharpness nor in the number of curves.

2.3.2.2 Gradient

High vertical grades are associated with high accident frequency (Shankar et al., 1995; Milton and Mannering, 1998; Chang, 2005; Anastasopoulos and Mannering, 2009; Gitelman et al., 2014). Milton and Mannering (1998) refer that the increase in accidents on upgrades is related to speed decreases, especially for heavy vehicles that might lead to more overtaking from faster cars. Downgrades, on the other hand, lead to higher vehicle speeds and therefore stopping distances. Yuan et al. (2008) suggest that not only the grade but also the length of downgrades is related with higher accident frequency.

Road geometrical characteristics do not act independent of one another. It has been suggested that the combination of poor horizontal and vertical curvature is mainly related with more traffic accidents (May, 1994; AASHTO, 2010). A recent simulation-based study has shown that the lateral acceleration (that was used as the proxy for accident probability) is high on downhill curved segments as well as crests or sags that are combined with curves (Wang et al., 2015). Uphill segments are not related with increased lateral acceleration. This study provides a new insight on the impact of road geometry on safety but it examines the impact of only one risk factor related with geometrical combinations so its results cannot be considered as generic.

2.3.2.3 Number of lanes

The number of lanes is linked with lane changes and increased vehicle interactions that can be potentially dangerous. Kononov et al. (2008) defined the possible vehicle conflicts as a function of the number of lanes as follows:

$$C_n = n \cdot (n-1) \text{ for } n = 2 \tag{2.7}$$

and

$$C_n = n \cdot (n-1) + \frac{n!}{3!(n-3)!} \text{ for } n > 2 \tag{2.8}$$

where C_n: the number of possible conflicts and n: the number of lanes.

Most of the researchers who examined the influence of the number of road lanes on accidents have found that the number of lanes is proportional with accident frequency (Persaud, 1992; Milton and Mannering, 1998; Chang, 2005). Noland and Oh (2004) suggested that a higher number of lanes is increasing the fatality rates too. Milton and Mannering (1998) refer that the number of lanes as a variable might act as a proxy for separate road categories meaning that the estimated coefficients might additionally reflect the differences between roads rather than the number of lanes alone.

Ma and Kockelman (2006) estimated that the number of lanes is related with decreases in non-fatal accidents and that it has no effect on fatal accidents. The authors did not provide an explanation, but this result might imply that roads with more lanes are less congested and so they tend to have fewer accidents related with congested conditions, that are usually non-fatal. Park et al. (2010) suggested that the relationship of the number of lanes with accidents forms a U-shaped curve; 6-lane roads were found to be the least accident prone compared to 4 and 8-lane roads. The authors explained that roads with more lanes provide more space for accident avoidance manoeuvring.

2.4 Statistical Approaches in Accident Modelling

2.4.1 Evolution of the statistical approaches

Each accident is the outcome of a unique sequence of events related to the involved driver(s), vehicle(s) and the road environment. The in-depth examination of all accidents of a network individually though, is rarely possible due to the high number of accidents and the limited data availability. As a consequence, the accidents of a road network are usually analysed aggregated in a way that their volume is reduced while they remain informative (Lord and Mannering, 2010). The dominant accident aggregation method is based on topological and temporal criteria. Accident counts that occurred on pre-defined road links (i.e. link-based models) or areas (i.e. area-wide models) during a certain time period are aggregated and modelled against selected explanatory variables. From a methodological perspective, during the last decades the modelling approaches which are employed for modelling accident data have evolved reaching high levels of sophistication. Table 2.1 presents many of the papers presented in this literature review, that include link-based models, grouped by statistical approach.

The initial statistical model that has been used was Linear Regression. Linear Regression is very simple to apply and interpret and thus was employed by numerous researchers (e.g. Ceder and Livneh, 1982; Ceder, 1982). However, some of the properties of accident data, namely non-negativity, heteroscedasticity and non-normality of the error term, make linear regression unsuitable (Jovanis and Chang, 1986). These characteristics of accident data may lead to inaccurate predictions and invalid tests of significance. Instead of Linear Regression modelling Poisson regression has been applied by a number of researchers (e.g. Jovanis and Chang, 1986; Jones et al., 1991; Miaou and Lum, 1993; Miaou, 1994).

Poisson regression is applicable for non-negative integers that are usually characterised by low mean values, over-dispersion and heteroscedasticity like accident count data (see also Mannering and Bhat, 2014). One of the main assumptions of a Poisson model is equidispersion (i.e. the variance is equal to the mean) in the dependent variable. Accident data are actually rarely equidispersed. In fact, accident counts are usually overdispersed (i.e. the variance is higher than the mean) and in some cases underdispersed (i.e. the vari-

Table 2.1: Link-based accident analyses grouped by statistical approach employed.

Model	Research Papers
Linear Regression	Ceder (1982); Ceder and Livneh (1982); Lave (1985); Fildes et al. (1991); Quimbly et al. (1999); Taylor et al. (2000); Garber and Ehrhart (2000)
Poisson Regression	Jovanis and Chang (1986); Jones et al. (1991); Miaou et al. (1992); Miaou and Lum (1993); Miaou (1994); Baruya (1998*a*); Ivan et al. (1999, 2000); Kim et al. (2006); Caliendo et al. (2007);Qin et al. (2004) (zero-inflated Poisson)
Negative Binomial Regression (Poisson-Gamma)	Miaou (1994); Shankar et al. (1995); Vogt and Bared (1998); Milton and Mannering (1998); Abdel-Aty and Radwan (2000); Martin (2002); Chang (2005); Lord et al. (2005); Kim et al. (2006); Haynes et al. (2007); Caliendo et al. (2007); Haynes et al. (2008); Wang et al. (2009*b*); Park et al. (2010); Pei et al. (2012); Gitelman et al. (2014)
Poisson-Lognormal Regression	Lord and Miranda-Moreno (2008); Aguero-Valverde and Jovanis (2008)
Random Effects (including spatial correlation)	Quddus (2008); Guo et al. (2010); Quddus (2013)
Random Parameters	Anastasopoulos and Mannering (2009); El-Basyouny and Sayed (2009*a*)
Multivariate Poisson Regression	Ma and Kockelman (2006); Ye et al. (2009)
Multivariate Poisson Lognormal Regression	Park and Lord (2007); Ma et al. (2008); Aguero-Valverde and Jovanis (2009); El-Basyouny and Sayed (2009*b*)
Multivariate Poisson Lognormal Regression with Spatial Correlation	Aguero-Valverde (2013); Barua et al. (2014)

ance is lower than the mean) (Lord and Mannering, 2010). To control for overdispersion the negative binomial or Poisson-gamma regression has been proposed. A negative binomial model assumes that the Poisson parameter follows a Gamma distribution and it is probably the most widely applied statistical model in safety analyses (e.g. Miaou, 1994; Shankar et al., 1995; Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000). Another model that is suitable for over-dispersed data is Poisson lognormal regression, where the Poisson parameter is lognormally distributed. Although Poisson lognormal regression is more flexible as a model (Lord and Mannering, 2010) its estimation is more challenging because the Poisson lognormal distribution does not have a closed form and has been used in less studies such as Lord and Miranda-Moreno (2008) and Aguero-Valverde and Jovanis (2008).

More recent studies control for heterogeneity caused by unobserved correlations between the models' observations. As most of the models represent accident data as counts aggregated on a spatial basis it is reasonable to believe that observations from neighbouring regions have some characteristics in common (Lord and Mannering, 2010). Also, when panel data are available observations of the same area over different periods of time are likely to be correlated. Adding to the models random effects, that are assumed to follow a specific distribution, spatial or temporal correlations are controlled (e.g. Quddus, 2008; Guo et al., 2010). A more complex extension of the random-effects model are the random-parameters models which allow the models' parameters to vary among the observations (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009*a*).

Most of the papers in the literature that examine accident frequency employ the total number of accidents as a dependent variable. However, different accident mechanisms could be by definition related with different combinations of circumstances (Kim et al., 2006). As a consequence, the examination of accident contributory factors to an aggregate level might distort the results of the analyses. Researchers who studied the effects of accident contributory factors by accident type confirmed that there are indeed significant variations in the estimated coefficients by collision type and severity (Ivan et al., 1999, 2000; Qin et al., 2004; Kim et al., 2006; Ye et al., 2009; Geedipally et al., 2010; Bham et al., 2012). A limitation of these studies is the use of separate models for each accident type that cannot take into account the existing correlations between them as it was highlighted by Park and Lord (2007) and Geedipally and Lord (2010). Multivariate Poisson (Ma and Kockelman, 2006) and Multivariate Poisson lognormal models (Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009*b*) have been proposed for modelling simultaneously different accident types (e.g. by level of severity) and controlling for the unobserved heterogeneity that arises from the correlations between them. Multivariate models can incorporate multivariate spatial correlation random effects to control for the relationships between their spatial entities. Aguero-Valverde (2013) and Barua et al. (2014) have applied multivariate Poisson lognormal regression with multivariate conditional autoregressive (CAR) random effects that was shown to perform better than separate CAR models.

2.4.2 Limitations

The inconsistencies of the outcomes between studies that have been reviewed here can be related to data quality, the complex nature of accidents and methodological limitations. The rapid evolution of the accident modelling statistical approaches over the last decades allows some confidence on their explanatory potential. On the contrary, accident data aggregation approaches have a less dynamic trend; current methods are based on spatial adjacency of accidents. This approach is logical and effective from a practical point of view as the traffic data are usually available at the link level. However, a source of variability in the research outcomes might be due to the information losses caused by the conventional accident data aggregation approaches that employ spatial criteria such as the link-based method.

Deterministic link-based accident models imply that accident frequency on a particular link can be explained by its most frequently observed conditions. This can be proved invalid for two reasons. Firstly, the assumption of homogeneity within links may not necessarily be true. A typical link includes up to several miles of roadway and in some instances both directions of traffic. It is a fact that traffic and geometric conditions at the roadway may vary significantly even for adjacent parts of the same link (e.g. due to road topography and on-off ramps). Therefore, the observations used for representing the conditions on a road link might not be representative of all the parts of it. Secondly, the assumption that a relationship that applies at a group level (link-level) necessarily applies at an individual level is not always true. The characteristic values used in link-based models for representing each of the examined variables are usually central tendency statistics (e.g. annual average speed). These characteristic values can be proved to be totally unrelated with the actual conditions at the time and location of the accidents on this link. This situation is defined as an ecological fallacy (or aggregation bias) (Clark and Avery, 1976; Davis, 2002, 2004; Black et al., 2009) and it is suggested that when studies analyses are based on such an assumption they might produce erroneous outcomes.

Studies focusing on proactive accident prediction confirm that accidents are related to suddenly developing and often extreme traffic conditions (e.g. high and low speeds) that cannot be captured from aggregated measures such as hourly or annual averages (e.g. Pande and Abdel-Aty, 2005; Abdel-aty et al., 2005; Hossain and Muromachi, 2013). The use of these variables therefore leads to loss of information and under-representation of extreme conditions that may be crucial in explaining accident occurrences. These limitations of link-based accident modelling are likely to be reflected in the results of the analysis leading to, possibly, erroneous and inconsistent conclusions found in the literature especially for time-varying measures that might be more sensitive to aggregation bias.

To overcome this limitation, the data aggregation approach that should be developed must minimise the information losses as much as possible (Clark and Avery, 1976). This means that the data aggregation method should enable the representation of the actual conditions related with accidents and the risk that is related with these conditions (Davis, 2002). Considering the form of the available data developing an aggregation method that will represent the conditions just before accidents is a challenging and data demanding task, compared to the conventional approach. However, this approach will probably lead to deeper understanding of the relationships between potential contributory factors and accidents that is expected to contribute to the mitigation of their negative impact.

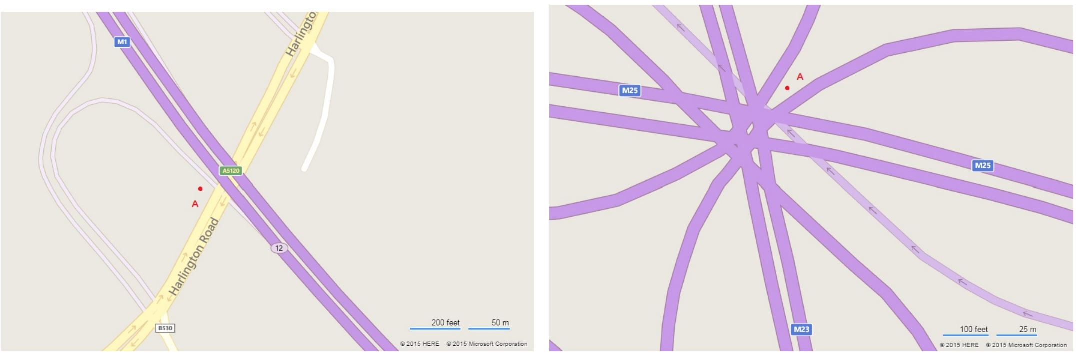
2.5 Accident Mapping

2.5.1 The role of accident mapping

Analysis of accident data aims to identify and explain the factors that lead to traffic accidents in order to mitigate the negative impact through the development of effective countermeasures. Accident analyses rely on the reporting of key variables such as accident time and location, road, driver and vehicle characteristics and contributory factors related to drivers' errors, vehicles' defects and problems with the road environment. The quality and reliability of the accident data is closely related to the validity of the analyses' outcomes (Austin, 1995; Loo, 2006; Tarko et al., 2009; Tegge and Ouyang, 2009; Deka and Quddus, 2014) .

The spatial nature of accidents makes location one of the primary attributes of accident databases (Koike et al., 2000; Tegge and Ouyang, 2009) which at the same time is very likely to be unreliable (Austin, 1995; Loo, 2006; Tarko et al., 2009; Deka and Quddus, 2014). Accidents are geographic events that can be represented by applying appropriate geocoding methods (Thill, 2000; Kam, 2003). Police officers who visit the accident scene just after an accident occurrence are usually in charge for reporting the accident locations with several different methods around the world (e.g. linear referencing, address, coordinates etc.). A common practice followed by the UK Police and by other authorities worldwide is recording the coordinates of the collision spot, obtained by grid maps or Geographic Information Systems (GIS), so as to achieve a higher level of accuracy.

However, even when the accident location coordinates are recorded, it is not guaranteed that the location of the accident can correctly be identified on a road network map. This is due to the errors that these measurements may include as they are collected mainly for administrative reasons (Loo, 2006) and the inconsistency between the network databases and the actual road network. For example, there are many simplified digital road maps in which roads are represented only by their centrelines and in some cases omitting features of the actual road geometry. As the majority of accident locations do not fall exactly on these centrelines, they have to be transferred and matched to the correct road segments. This is a quite challenging process, especially when they occur at areas with complex road



(a) Junction (b) Complex fly-over

Figure 2.4: Examples of complex accident mapping cases (source:BingMaps)

configuration where many road segments intersect, such as at junctions or flyovers (see Figure 2.4) that may lead to inaccurate accident allocation.

It is not known to what extent these errors are acceptable because the differences of road characteristics and the traffic conditions between two network points that are physically very close to each other may be considerable (e.g. the difference between a main motorway and its slip road). However, there are indications that the estimated coefficients of a safety model differ significantly when the corrected accident locations are used (Tegge and Ouyang, 2009). As a result, spatial accident data need to be enhanced through some special processing in order to be confident about their quality. Accident mapping is the process of the identification of the location (i.e. road section and coordinates) where a collision has occurred. Accident mapping can be useful for many types of accident analyses but it is particularly important if their purposes are:

- Locating hazardous segments within a network so as to design effective engineering countermeasures (e.g. altering road geometry) (e.g. Karlaftis and Golias, 2002; Bíl et al., 2013).

- Statistical modelling of traffic accidents with the aim of identifying the factors affecting accident frequency (e.g. curvature, gradient, traffic density and flow) and accident prediction (e.g. Miaou and Lum, 1993; Wang et al., 2009*a*).

- Estimation of the spatial distribution of safety risk across the network. This can be employed for risk mapping that may lead to the introduction of targeted and specialised accident prevention measures(e.g. Steenberghen et al., 2004; Loo, 2006).

2.5.2 Review of accident mapping methods

Although accurate accident locations can be particularly useful for accident analyses the majority of existing studies proceed to the analysis of accident location data without reporting any sort of prior processing (e.g. Koike et al., 2000). The number of accident mapping methodologies that can be found in the literature to date is relatively low as accident mapping is not a common area of research in road safety. Existing accident mapping techniques use all types of spatial data that are included in accident records in order to increase the possibilities of accurate matching. The approaches vary according to the aim of each study and the type of their locational input data. Accident locations are reported either using linear referencing, offset from intersections, addresses, or GIS coordinates. Table 2.2 provides the key features of existing algorithms found in the literature.

The linear referencing method is a straightforward and relatively accurate method of accident location reporting that can be used on numbered roads (i.e. typically sub-urban and rural networks). Studies that have available the indication of the closest milepost to the accident define this point of the network as the accident location (Geurts et al., 2006; Monsere et al., 2006). This approach demands minor processing from the researchers' behalf but tends to be insensitive to reporting inaccuracies. Moreover, its location error is equal to the half of the interval between two mileposts that can be as high as half a mile.

Accident locations are also reported using an offset from nearby junctions. To convert this type of accident locations to coordinates it is needed to combine the attributes of the accident and the network data. Dutta et al. (2007) and Qin et al. (2013) developed two algorithms for identifying accident locations in urban and sub-urban environment using On-At tables that demonstrate all the roads and their directions at each intersection. These algorithms are strongly dependent on the accuracy of the information included in the On-At tables and cannot be applied when junction information is missing. Qin et al. (2013) report a relatively high overall matching percentage of 83% that reaches 89.7% for local roads.

Address is an easily obtained spatial variable that can be used for identification of accident locations mainly for urban networks. Burns et al. (2013) tested some of the online geocoding APIs (Application Programming Interfaces) in terms of their capability to identify accident locations, when they are given the reported addresses. Google Maps API was found to have the highest matching rate (78%) but the accuracy of matching could not be quantified due to the ambiguity of the cases where the addresses include spelling or other mistakes. Another accident mapping method that is indirectly related to addresses is the method developed by Tarko et al. (2009) that attempted to link accident and network records. The method was theoretically founded on probabilistic linking techniques (Fellegi and Sunter, 1969) used for matching hospital data. Although their method succeeded in matching all accidents with the correct roads, in some cases accidents were matched with multiple roads making the final output ambiguous.

When GIS coordinates of accident locations are superimposed on digital maps of the road network, they rarely fall exactly on road sections. Thus, in order to identify accurate locations additional accident related information should be employed. When the location coordinates are available accident mapping can be seen as a special case of map-matching in the sense that the aim is the identification of the correct road segment on which a vehicle is travelling as well as a specific position on the segment (Quddus et al., 2003, 2007). The individual characteristics of accident mapping are that the process is static, independent (i.e. location is not related with prior locations) and offline (i.e. post processing).

Despite their differences in function and purposes, concepts used in map-matching can be employed in accident mapping. Some of the GIS-based studies use straightforward but simplistic approaches such as selection of the closest junction (Levine et al., 1995) or closest road section with road name filtering, (Loo, 2006). These techniques resemble the point-to-point matching and the point-to-curve matching (see: Bernstein and Kornhauser, 1996) respectively. A similar approach includes the use of restrictive, pre-defined buffer zones along with some descriptive variables such as road name, class, speed limit and junction details (Austin, 1995). These approaches may be effective for large datasets, but not very precise.

A variable that can reinforce the matching accuracy of accident mapping algorithms is vehicle direction. Wang et al. (2009*a*) introduced the use of vehicle direction in the form of the angular difference between the intended direction of the involved vehicles and a road segment. Wang et al. (2009*a*) used a maximum weighted score to combine the distance and the angular difference to identify accident locations on the M25 motorway in the UK. The weighted score approach, is not suitable though for more dense and complex networks due to its strong dependency on vehicle direction. Direction difference was later used by Deka and Quddus (2014) who developed an artificial neural network for matching accidents within the entire primary road network of the UK that considered the distance, vehicle direction, and the reported road name and type (accuracy level: 98.4%). One of the main shortcomings of this method is the expression of the direction of an accident by a single measure (i.e. the average of all the intended directions of all the involved vehicles) that may result to information losses if the examined accidents include multiple vehicles. A vehicle direction-based accident mapping algorithm that overcomes this limitation and enables the identification of road accidents without taking into consideration the reported road name was developed by Imprialou et al. (2015). This algorithm also employs an error circle, a transformed map-matching technique, for the initial candidate road segments selection (Zhao, 1997; Quddus et al., 2007). The final accident location selection is based on a matching score that was estimated based on a multilevel logistic regression model. The accuracy of the matches is high (97.1%), but the overall approach is relatively complex and time-consuming.

Table 2.2: Characteristics of existing accident mapping techniques

Author	Accident Location Form	Method	Advantages	Disadvantages
Austin (1995)	Coordinates	Buffer Zone	Simple and fast to implement; Transferable.	Strict buffer zone limits; Ambiguous results for junction accidents.

Levine et al. (1995)	Coordinates	Accidents allocated to the closest junction (point-to point matching)	Efficient for large datasets; Transferable.	All accidents were allocated to junctions (53% inaccurate matches).
Loo (2006)	Coordinates	Accidents allocated to the closest junction or road with road name filtering(point-to-curve matching)	Relatively reliable results; Transferable.	Lack of precision for the location of junction accidents; Road types are not considered.
Dutta et al. (2007)	Offset from junction	Conversion of accident location to a set of coordinates using On-At information	Accurate results.	Strong dependence on the accuracy of On-At location information; Moderate matching percentage;
Wang et al. (2009)	Coordinates	Maximum weighted score	Straightforward; efficient for selection between the two directions of a motorway.	Not suitable for dense road networks; Strong dependence on angular difference.
Tarko et al. (2009)	Address	Probabilistic linking technique	Accident location coordinates not necessary .	Identification of more than one matching segments.

Qin et al. (2013)	Offset from junction	Conversion of accident location to a set of coordinates using On-At information	Relatively accurate results (intersecting road information enhance the validity).	Strong dependence on the accuracy of On-At location information; Inflexible if junction information is missing.
Burns et al. (2013)	Address	APIs comparison	Application of freely available software.	Moderate matching percentage; Unknown accuracy;
Deka & Quddus (2014)	Coordinates	Artificial Neural Networks	Accurate results;	Black-box technique; Use of one measure for angular difference
Imprialou et al. (2015)	Coordinates	Multilevel Logistic Regression	Accurate results; Efficient for dense, urban networks.	Relatively high Complexity.

Summary

This literature review presented the most interesting research findings on the impact of speed, other traffic characteristics and road geometry on accidents. For all the variables the results are not entirely consistent between studies. This highlights the inherent complexity of accident analyses and some potential methodological limitations that might lead to erroneous outcomes. In summary, based on the most frequent findings of the reviewed literature:

- Speed limit raises generally lead to increased average speeds and higher accident frequency;
- Higher traffic speed is definitely associated with increased accident severity and possibly with increased accident frequency too;
- High speed variance (especially when it is represented by the standard deviation of speed) is related to higher accident frequency;
- Higher traffic flow is connected with higher (especially multiple-vehicle) accident frequency.
- Sharp and sparse curves are related with higher accident frequency and severity;
- Steep vertical (especially negative) grades are associated with higher accident frequency;
- Road segments with more lanes are considered to have higher accident frequency.

Among the various statistical models that have been employed in order to develop explanatory and predictive accident models, Poisson has been proved to be the family of models that has the most appropriate statistical properties for modelling accident counts. Negative binomial (Poisson-gamma) regression is by far the most commonly employed statistical model in the current literature. More recent accident analyses add in their models random effects that control for spatial and temporal correlations between observations. Analysing accidents divided by type (severity or collision type) is gaining attention as this approach seems to produce more informative and detailed results. The most suitable approach for modelling simultaneously different accident types and controlling for the correlations between them is multivariate Poisson-lognormal regression.

One of the most important limitations of current accident analyses is the aggregation bias that is related with the link-based aggregation method that is employed for grouping accidents. This aspect of accident analyses has not yet gained the attention of researchers but it might have significant impact on the findings of accident analyses. Another relatively overlooked and potentially very important subject is the accuracy of the accident locational data. Accident mapping is the process of correcting the reported accident location so as to improve the overall data quality and consequently the outcomes' accuracy.

Existing accident mapping techniques vary with respect to the type of the input accident data. When accident locations are defined with coordinates, the vehicle direction just before the collision and the road name have been found to be crucial variables for the algorithms' accuracy.

Chapter 3

Methodology

3.1 Introduction

The aim of this thesis is to examine the relationship of traffic speed with accidents through the development of robust statistical models. Modelling outcomes rely on the inherent quality and the aggregation of the data as well as the statistical approach that is used for the analysis. To derive accurate results, the methodology of this work includes new data pre-processing techniques, that are expected to maximise the representation of the available data, and appropriate and sophisticated statistical techniques.

Accident locations are often less accurate than desired, so in order to improve the overall input data quality an accident mapping algorithm was developed. The algorithm, that was designed for the study area, is based on a transformed map-matching technique combined with an Artificial Intelligence concept. Recognising the major limitation of the link-based accident data aggregation approach, which will be outlined in this chapter, one alternative data aggregation method was developed termed as condition-based. The condition-based approach enables the representation of the pre-accident traffic and geometric conditions and is expected to define more accurate relationships between accidents and the potential precursors.

Accident frequency split by accident type was modelled using multivariate count regression. By forming an appropriate framework, the models that will be developed will respond to the research questions and will facilitate comparisons between the accident data pre-processing approaches.

3.2 Research Design

The aim of this thesis is divided into six objectives that will be achieved through methods that are outlined below. Table 3.1 shows the objectives and the methods along with the corresponding chapter(s) in this thesis.

Table 3.1: Research objectives and methods.

Objective	**Method**	**Chapter**
To review the impact of speed and other contributory factors on accidents.	Literature review	Chapter 2
To examine existing statistical approaches in accident modelling.	Literature review	Chapter 2
To refine and merge data from multiple sources so as to enhance the quality of the analysis.	Development of an accident mapping algorithm; identification of pre-accident conditions	Chapter 3 and Chapter 4
To develop accident-speed relationships using a new, condition-based modelling approach.	Development of multivariate count regression models for datasets that are aggregated according to the pre-accident conditions	Chapter 5
To compare and contrast the results between the conventional and the condition-based modelling approach.	Development of multivariate count regression models for link-based datasets and evaluation of the two aggregation methods (i.e. link-based and condition-based) according to their results	Chapter 5
To evaluate the safety impact of a potential speed limit increase on accidents.	Estimation of the expected changes on accidents related with speed limit alterations	Chapter 6

3.3 Accident Mapping Algorithm

3.3.1 Introduction

As it has been mentioned in section 2.5.1, accident coordinates are usually not compatible with the digital networks that are used for conducting accident analyses. Thus, in order to identify the road segment where each accident occurred, it is necessary to use an appropriate accident mapping method. The allocation of accidents to the closest segment or junction (Levine et al., 1995; Loo, 2006) is the simplest approach but it is not expected to provide very accurate results. The accuracy levels of the accident mapping algorithm is crucial for this project because the accident datasets that will be used for modelling rely on the quality of accident locations. As a consequence, an accident mapping algorithm should be sophisticated enough to provide precise locations with the minimum possible manual intervention.

To obtain accurate accident locations for the next steps of the analysis, a new accident mapping algorithm was developed. The accident mapping algorithm is based on Fuzzy logic, an Artificial Intelligence concept that estimates the degrees of truth of a statement (e.g. Did this accident occur on this road section?) by aggregating a series of partial truths (MathWorks, 1999). The algorithm will be henceforth referred as to AMF (i.e. Accident Mapping Fuzzy logic). The AMF algorithm consists of three steps (Figure 3.1):

1. The network segments were firstly filtered in respect to their adjacency to the reported accident location and their road name and type so as to form a set of candidate segments;

2. Each of the candidate segments was then evaluated for its goodness of matching with the reported location and vehicle movement direction just before the accident using a Fuzzy Logic (FL) Inference System;

3. Each accident was allocated to a suitable point on the selected segment that is considered to be the actual location of the first impact.

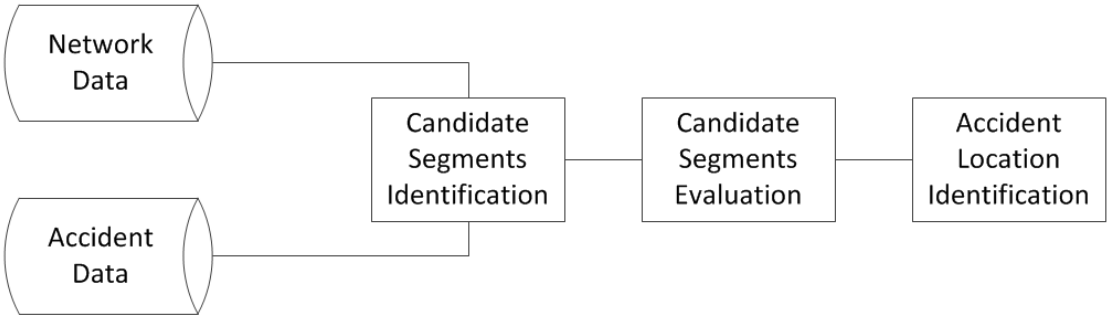


Figure 3.1: Flow chart of the main three steps of AMF

3.3.2 Candidate segment identification

This algorithm was developed to identify the road section that had the highest likelihood to be the one where an accident occurred. Since the network map is represented by 211,247 piece-wise straight segments, it was essential to apply an initial selection process where only the most relevant segments to the accident location would be kept for further processing and for the final matching (hence termed as candidate segments).

Candidate segment identification followed the method outlined in Quddus et al. (2007) and implemented by Imprialou et al. (2015) in which an error circle is formed around the reported accident point. The radius of the circle is subject to the quality of the network and accident location data. This was empirically derived from a sample.

Road segments that fell within the error circle or physically intersected with the error circle were considered to be the candidate segments. This can be termed as a classical circle-line intersection problem in which a segment can be considered as a candidate segment when one of the two following conditions is satisfied: a) both of the nodes of a segment fall within the circle, or b) there exists at least one intersecting point between a segment and the circumference of the circle. In other words a road segment (N_1N_2) (Figure 3.2) is candidate for an accident with reported location $O(x, y)$ (error circle (O, r)) when:

$(OP) \leq (r)$

Unless:

$\vec{v_{1n}} * \vec{v_{2n}} > 0$ for $n = 1, 2$;

and

$|\vec{v_{mn}}| > (I_1 I_2)$ for $m = 1$ and $n = 1, 2$ or $m = 2$ and $n = 1, 2$.

Where: $\vec{v_{mn}} = I_m \vec{N_n}$ for $m = 1, 2$ and $n = 1, 2$.

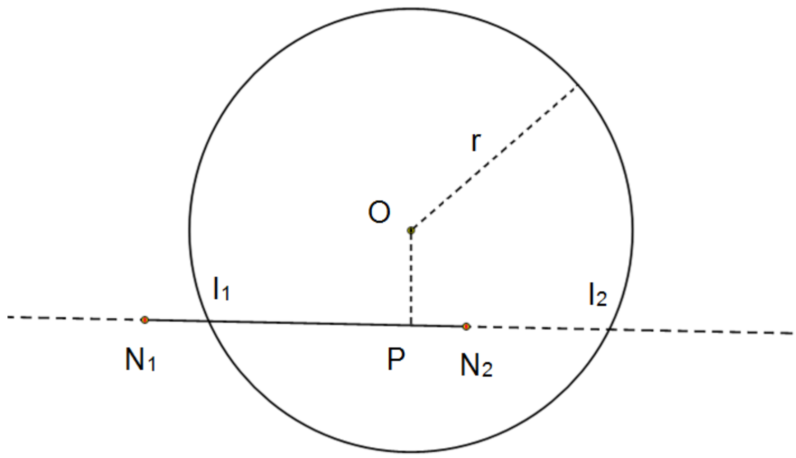


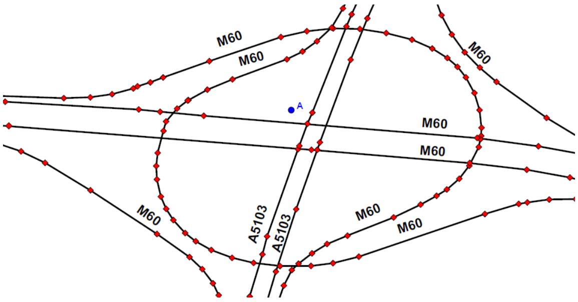
Figure 3.2: Error circle and a candidate segment (N_1, N_2 : segment's nodes, I_1, I_2: intersections with the circle, OP: the perpendicular distance from the centre of the circle to the segment)

Since road configuration close to junctions can be very complex, the number of candidate segments with different characteristics was also high. To avoid to the extent possible the number of mismatches the candidate segments were filtered based on their road name and road type. Segments that had different road name and road type from that reported in the accident database were excluded from the next step of the process. However, due to some inconsistency that appears to exist between the databases, this filtering process was proven to be very restrictive where no match was found. In these cases, the filtering rules were relaxed and the segments that remained for further evaluation were those that had either the same road name or road type with the reported.

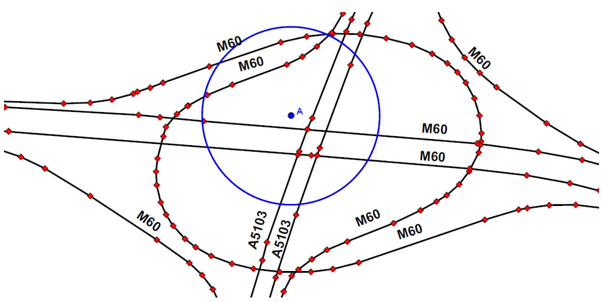
This process is displayed at Figure 3.3 in which the accident location is denoted by the round symbol A. An error circle of radius 100m is formed around the accident point, A (Figure 3.3(b)). The initial set of candidate segments are shown in Figure 3.3(c). According to STATS19 accident database, the accident (A) occurred on A5013 with "main carriageway" road type. Based on this information, Figure 3.3(d) shows the final set of candidate segments on A5103.

The radius of the error circle that was used in this research was set to 100 m based on empirical observations and the characteristics of the network data. From some initial manual matching of accidents using the accident database for the year before the study period (i.e. 2011) it was found that the perpendicular distance of the reported accident location to the matching segment was usually considerably less than 50 m. As a consequence, and taking also into consideration the segment lengths' descriptive statistics, an error circle with 100 m radius was expected to include a reliable number of candidate segments, eliminating both the probabilities of false alarm and missed detection.

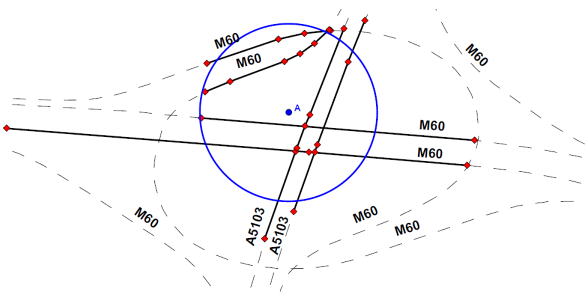
If no segments were found within the 100 m circle then the radius gradually expanded up to 200 m with a 50 m step. In addition, the filtering rule also changed to allow the inclusion of segments with either the same road name or the road type as depicted in the flow chart (see Figure 3.4). However, if still no candidate segments could be identified, the accident remained unmatched and was flagged in order to be manually matched.



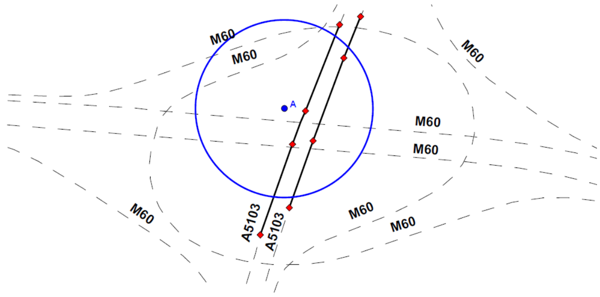
(a) Reported accident location



(b) Error circle formation

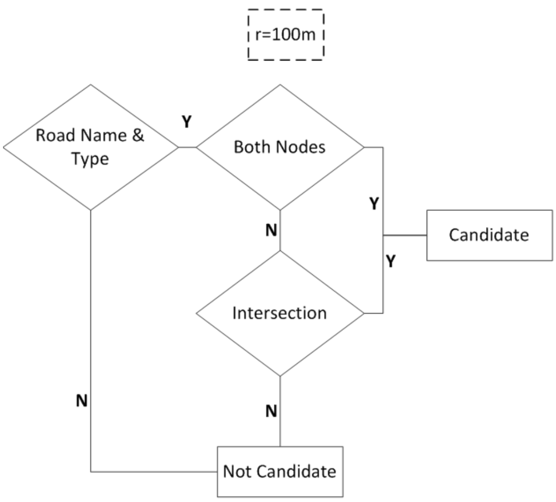


(c) Potential candidate segments

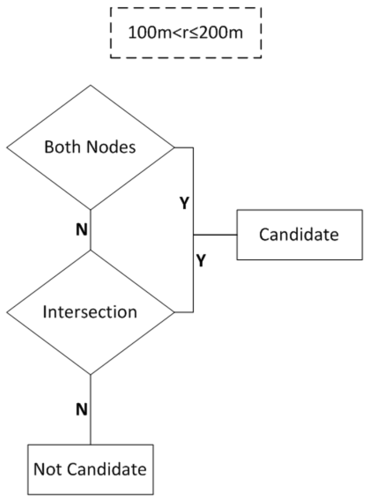


(d) Final set of candidate segments

Figure 3.3: Candidate segment identification process in four steps



(a) Initial error circle radius ($r = 100m$)



(b) Expanded error circle radius ($100m < r \leq 200m$)

Figure 3.4: The sequence of checks for each road segment of the network for the candidate segment set formulation.

3.3.3 Segment selection

For the final selection of the matching segment, the goodness of matching of each of the filtered candidate segments with the accident information was evaluated by applying a

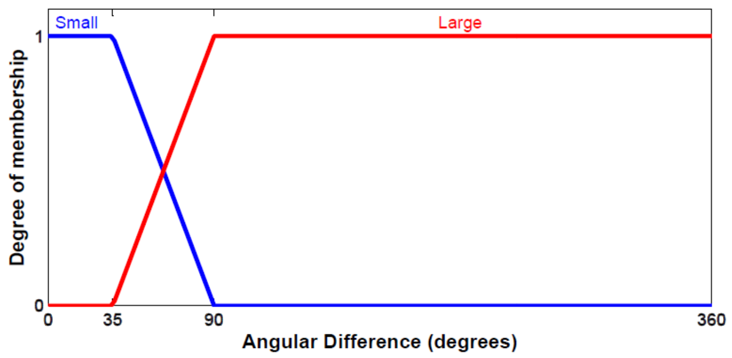
Mamdani Fuzzy Inference System (FIS) using the Fuzzy Logic Design Application included in Matlab 13. In this study it would not be easy, and possibly not successful, to set some pre-defined thresholds that would indicate which segment among the candidates is the most likely for the accident to occur. Fuzzy logic is a technique that is used when reasoning is not determined by exact rules but from different levels of truth (from completely true to completely false) and for interpretation of linguistic terms (such as short distance). Additionally, fuzzy logic is suitable for analyses that include data likely to be imprecise (MathWorks, 1999). The two input variables of the FIS were: the distance (D) from the accident point to a candidate segment and the angular difference ($\Delta\vartheta$) of the intended vehicle direction and the direction of a candidate link. D was defined as the perpendicular distance from the reported location point to the candidate link, if the reported location's projection point was between the two nodes of the link; otherwise it was the minimum of the distances between the accident location and the link's nodes.

After selecting the set of input and output variables, the remaining two major components of a FIS were: (1) fine-tuned membership functions and (2) fuzzy rules (MathWorks, 1999). Since membership functions can take different shapes and forms, one of the challenging aspects of designing a FIS is the development of membership functions in terms of their shapes and fine-tuning. The most commonly used membership functions (e.g. triangular, trapezoidal, sigmoidal and Gaussian) were empirically explored in this study. The results indicated that the performance of the FIS does not significantly vary by the shape of membership functions. Triangular and trapezoidal functions however offer slightly better accuracy (about 1% higher). Therefore, these two were considered in the FIS.

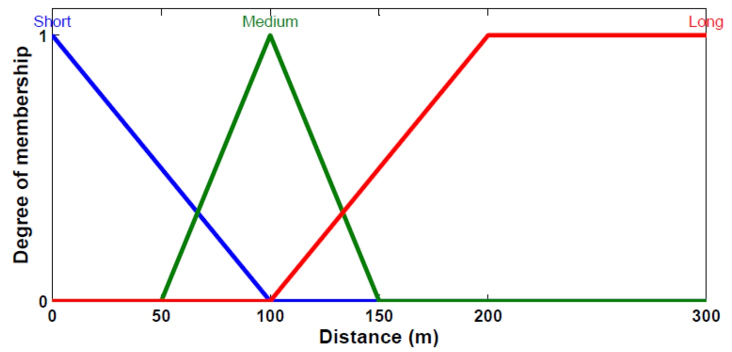
Membership functions were fine-tuned based on empirical observations and the nature of the data that are used for calculating distance (D) and the angular difference ($\Delta\vartheta$) respectively. The threshold values of each membership function were determined empirically after matching manually 200 accidents obtained from an independent database; the STATS19 accident reports of 2011. The exploratory manual matching process was essential for understanding the range of the two input variables and consequently for determining the number of the membership functions for each of the input variables as well as their thresholds. The angular difference ($\Delta\vartheta$) included two membership functions (i.e.

small and large) that represented how acceptable an angular difference was. The membership function was set considering that the angular difference could only slightly exceed the measurement error (22.5^o). The distance (D) included three membership functions (i.e. short, medium, long) expressing the level of the relative distance of the candidate segment to the reported accident location. The fuzzy rules (1-6) are all the possible combinations between the membership functions. Figure 3.5 shows the fine-tuned input and output membership functions. The following six rules were applied to the FIS:

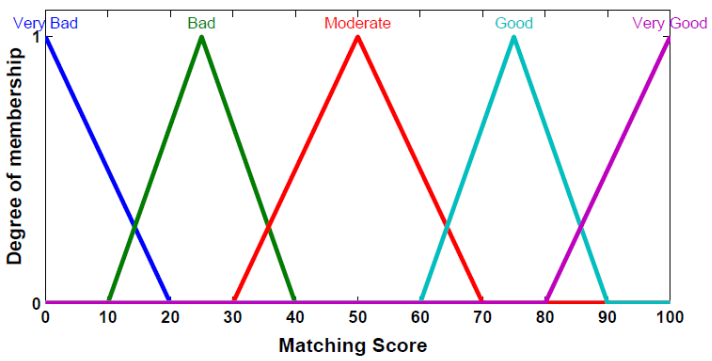
1. If (D is short) and($\Delta\vartheta$ is small), then (Matching Score is very good)
2. If (D is medium) and ($\Delta\vartheta$ is small), then (Matching Score is good)
3. If (D is long) and ($\Delta\vartheta$ is small), then (Matching Score is moderate)
4. If (D is short) and ($\Delta\vartheta$ is large), then (Matching Score is moderate)
5. If (D is medium) and ($\Delta\vartheta$ is large), then (Matching Score is bad)
6. If (D is long) and ($\Delta\vartheta$ is large), then (Matching Score is very bad)



(a)



(b)



(c)

Figure 3.5: FIS input (a and b) and output (c) membership functions

For the defuzzification of the rules the method of centroid was applied and the output variable was a matching score that ranged from 0 to 100. The candidate segment with the highest Matching Score was considered to be the most likely to be the segment where the examined accident should be located (Correct Link). In the case of more than one segments having equal Matching Scores the segment with the smaller D from the accident location was selected.

3.3.4 Accident location

After the identification of the correct segment the exact location of the accident on the selected segment was determined. The actual location was a point on the correct segment that is the closest to the reported accident location. If the perpendicular projection of the reported accident location fell between the two nodes of the correct segment, that was selected to be the accident location. If the perpendicular projection fell outside the segment, the closest node to the reported location was selected. Following, the distance between the reported and the projected location was calculated (henceforth: *Distance*).

Aiming to ensure the results' maximum accuracy, the three-step procedure described above was applied three times separately for: all the accidents, accidents that were reported to occur on roundabouts and accidents that were reported to occur on slip roads. By separating roundabout and slip road accidents and adding some additional steps where appropriate, mismatches were avoided to the maximum level. As it is mentioned at section 3.3.2, from the initial manual checks it was observed that Distance was at the majority of the cases lower than 50 m. That is why is when AMF selected a segment that was placed 50 m or more from the reported location a manual check on the accuracy of this result was considered to be useful.

- All accidents
  Network Database: HAPMS
  Additional steps:

  1. If *Distance* $\geq$ 50 m, the accident location was manually checked and changed if necessary
  2. If *Candidate Segments* $= 0$, the accident location was manually determined

- Roundabout/Slip Road accidents
  Network Database: HAPMS (Roundabouts/Slip Roads only)
  Additional step:

  3. If *Distance* $\geq$ 50 m or *Candidate Segments* $= 0$, the accident location was substituted with the respective correct segment estimated by All accidents.

3.3.5 Algorithm evaluation

Finally, AMF was evaluated in terms of its accuracy levels. For this purpose, a sample of the examined cases were matched manually to road sections, and the results were compared to the respective sections that are identified by the developed method. For the manual accident mapping additional variables from the STATS 19 were used such as speed limit, junction details, 2nd road class and number. The inclusion of the additional variables to the manual checks and the fact that during a manual accident mapping process it was possible to treat each case individually when this is necessary (e.g. accidents near complex configurations), makes the results of this process highly reliable. Therefore, manually mapped accidents formed a suitable reference set for the examined algorithm. The sample was obtained by dividing the entire road network into 70 exhaustive and mutually exclusive clusters of equal areas and selecting accidents randomly with the quota sampling approach (Figure 3.6). The size of the sample (N_s) was 716 cases and it was estimated using the equation of sample size for categorical data (see Bartlett et al., 2001) as follows:

$$n_0 = \frac{Z^2 p(1-p)}{d^2} \tag{3.1}$$

$$N_s = \frac{n_o}{1 + \frac{n_o}{N_p}} \tag{3.2}$$

Where:Z: Z value (here:1.96 for 95% confidence level), p: percentage of expected error (here: 2.5%), d : acceptable margin of error of the estimated proportion (here: 1.1%), N_p: population size (here: 10,520) and N_s: sample size.

The similarity of the proportions of the reported road types (roundabout, slip road and main carriageway) between the reference and the full dataset (see Table 3.2) suggest that the reference set is representative and suitable for the algorithm evaluation.

Table 3.2: Comparison of the proportion of accidents included to the reference set split according to the three main road types

Road Type	N	Roundabout	Slip Road	Carriageway
STATS 19 Reports	10520	8.2%	7.9%	83.9%
Reference set	716	7%	8.5%	84.5%

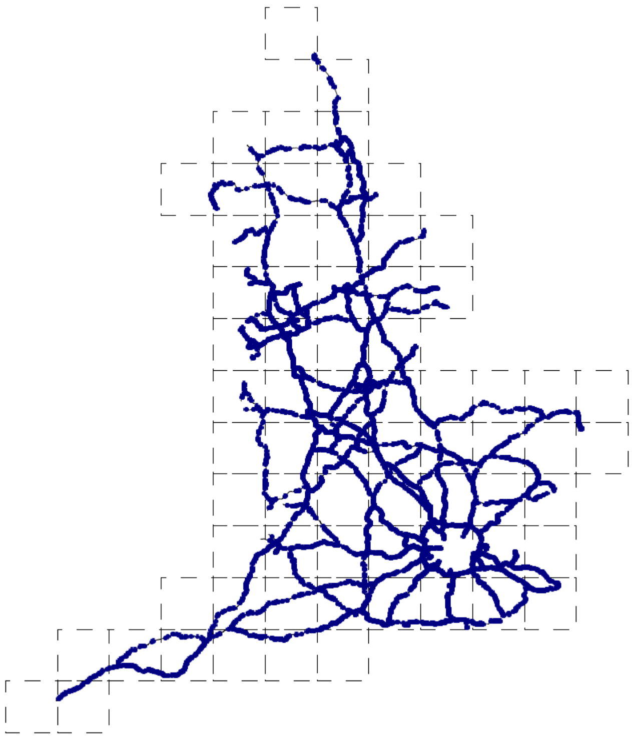


Figure 3.6: Map of the road network divided by the 70 equal and mutually exclusive clusters.

To assess the utility of AMF, it is interesting to compare it in terms of accuracy with existing accident mapping methods that are applicable to the study area. The three methods that fulfilled this criterion were:

1. Accident Mapping Method 1 (AMM1) based on Levine et al. (1995): Accidents were allocated to the closest segments of the network;

2. Accident Mapping Method 2 (AMM2) based on Loo (2006): Accidents were allocated to the closest segments of the network that had the same road names and types as the reported accident.

3. Accident Mapping Method 3 (AMM3) based on Wang et al. (2009*a*): Accidents were allocated to the segment that has the highest Weighting Score (WS_i) that was calculated using the following equation:

$$WS_i = \frac{1}{d_i} + \cos(\Delta\vartheta_i), d_1 \neq 0 \tag{3.3}$$

Where d_i: distance of the reports accident location from the road segment (i.e. D)

and $\Delta\vartheta_i$: angular difference between the intended vehicle direction and the link's direction (i.e. $\Delta\theta$).

3.4 Link-Based Approach

The link-based accident data aggregation approach is the conventional approach for accident modelling (e.g. Miaou and Lum, 1993; Lord and Mannering, 2010; Barua et al., 2014) and it has a very straightforward concept that is based on the accident locations. The sampling frame of a link-based dataset consists of actual spatial entities which are all the pre-defined road links of the examined network. A link-based dataset enlists the links of the network and the total number of accidents per link. The accidents of a link are accidents that occurred at different time points over a fixed time period. Each link contains information that represent the conditions on the road defined by characteristic values such as descriptive statistics (e.g. mean, median, maximum etc.). In link-based models accident counts are modelled against these links' characteristic variables to reveal any significant relationships. Figure 3.7 is an illustration of the link-based approach. Suppose that ABCDA is a road network that consists of five junction-to-junction links (AB, BC, CD, CA, DA). The red x's represent the locations of the accidents that occurred on the network during an entire study period (e.g. one year). In the link-based approach accident counts are aggregated by link and are modelled using the annual average speed and flow on the link forming a dataset like this that is presented in Table 3.3.

Although this aggregation approach is both convenient and simple and perhaps the only option for researchers due to limited and not detailed data sources, it suffers from aggregation bias. That is because in a link-based model it is assumed that the triggering factors for accidents that occurred on the same link are similar, which of course might not be true for all the cases. For instance, accidents 1 and 2 (Figure 3.7) that are assigned on the same linked might have occurred at different periods of time and under entirely different traffic conditions. This essentially means that link-based models show relationships of accidents with the average conditions, but not with the actual conditions just before the occurrence.

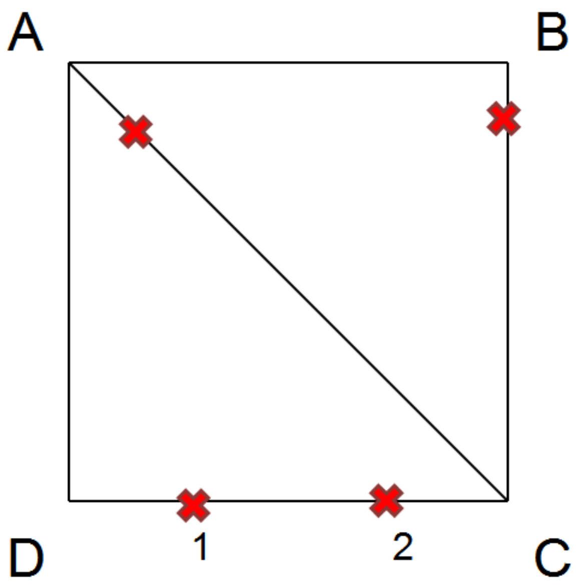


Figure 3.7: Example of a road network and the total accidents that occurred on it during a study period.

Table 3.3: Example of the link-based accident dataset for the network of Figure 3.7

Link ID	Accident Counts	Average Speed	Average Flow
AB	2	80	100
BC	0	78	120
CD	0	79.5	111
DA	1	76	300
AC	1	85	220

3.4.1 Exposure

In order to enable meaningful comparisons in terms of accident risk between the observations of safety data models it is necessary to take into account one exposure variable. The use of an offset in a count model indirectly transforms the dependent variable from a number of events to a rate of events per the exposure measure. Exposure in link-based approaches attempts to express the total amount of travel on each link. The most appropriate measures of exposure for link-based modelling have been broadly discussed in the literature (e.g. Qin et al., 2004; Lord et al., 2005; Pei et al., 2012) as there is a plurality of surrogate measures of exposure. In this analysis the link length was used and it is one of the most commonly employed measures of exposure.

3.5 Condition-Based Approach

The condition-based is a new accident data aggregation approach that aims to address the main limitation of the link-based aggregation by enabling the representation of the actual pre-accident conditions. The sampling frame of a condition-based dataset comprises all the possible combinations of traffic and geometric conditions on the examined network; a set of non-physical attributes that co-existed at the time and the location of an accident. The number of the possible condition combinations (henceforth: *scenarios*) depends on the number of examined variables, their specifications and the empirically defined separation intervals.

Each scenario is matched with a number of accidents (from zero to, theoretically, all the accidents of the database) that are found to occur under this particular combination of traffic and geometry conditions. Condition-based modelling attempts to represent the actual accident-related traffic and geometry conditions. In contrast to the link-based approach, the accidents that belong to the same condition scenario do not necessarily have a spatial or temporal relationship. Instead, they are similar in the sense that when they occurred the external circumstances on the road were approximately the same. Assuming that some or all of these circumstances might be related with the accident occurrences, the concentration (or absence) of accidents in some particular scenarios should provide useful information about the accident triggering factors.

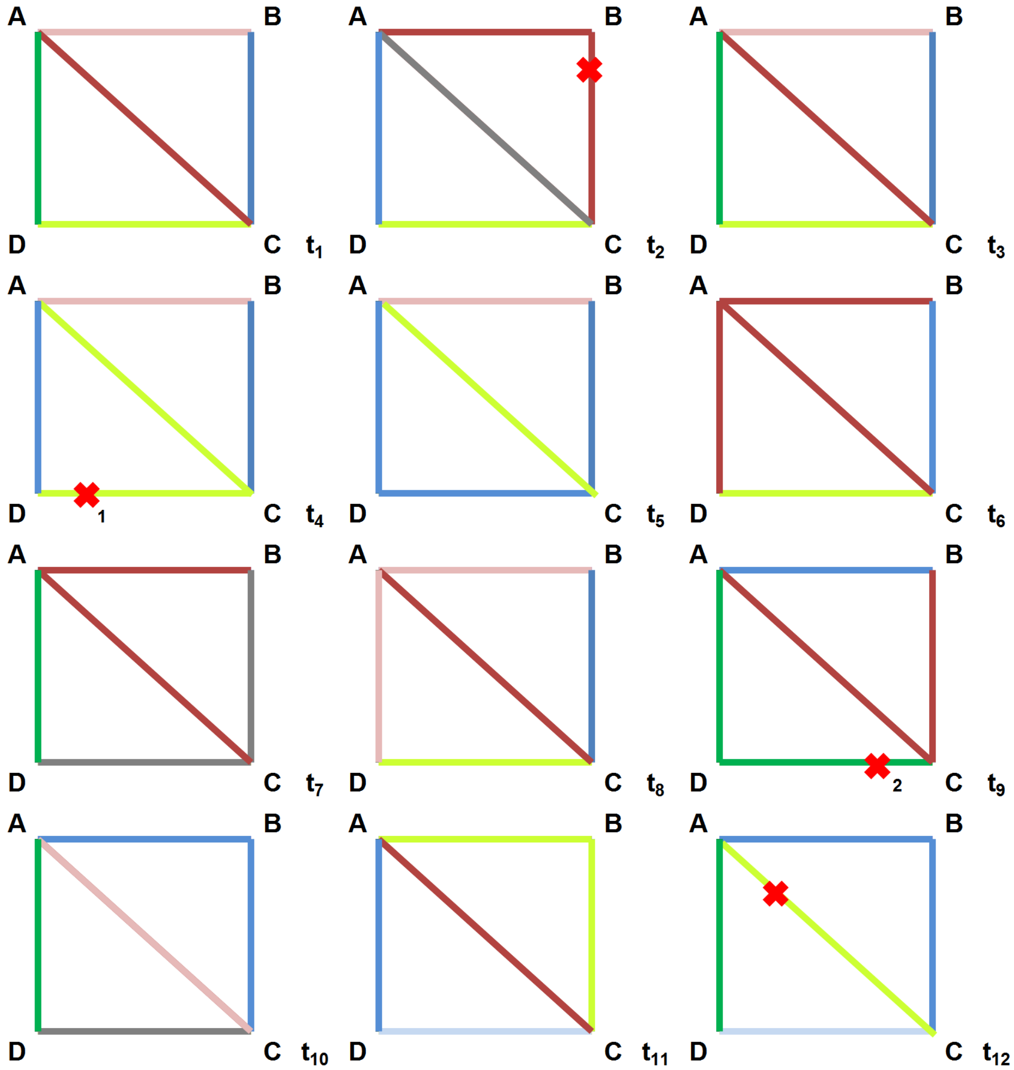


Figure 3.8: Example of the traffic conditions on a road network at subsequent time intervals ($t_1 - t_{16}$) and the total accidents that occurred during a study period.

Table 3.4: Example of the condition-based accident dataset for the network of Figure 3.8

Conditions	Accident counts	Speed	Flow
	2	80	100
	1	69	100
	0	55	200
	0	44	200
	0	30	300
	1	19	300
	0	7	300

Figure 3.8 graphically represents the aggregation of accidents with the condition-based approach. Suppose that ABCDA is a road network that consists of five links (AB, BC, CD, CA, DA). The study period is divided in to 16 ($t_1 - t_{16}$) subsequent time intervals [1] and the network is presented separately for each one of them. The colours of the links illustrate the seven distinct traffic condition scenarios of the study network. Again, the red x's represent the accidents on the network. It is clear that accidents that occurred on the same location at different points in time (accidents 1, 2) occurred under different traffic conditions. Accidents are aggregated based on the traffic conditions that were present when they happened as it can be seen in Table 3.4. For example, during the study period there were two accidents that occurred under "light green" conditions, but zero under "blue" conditions.

The formation of a condition-based dataset is quite complex comparatively to a link-based dataset. To specify the conditions prior to each of the accidents, accident location and time should be combined with the traffic and geometric data. Figure 3.9 displays a simple flowchart describing the main processes to develop the condition-based dataset consisting of Nmax accidents. The main steps (traffic and geometric conditions identification) are explained in detail at the following sections (see 3.5.1 and 3.5.2).

[1] The number of the time intervals is very small only to facilitate the presentation. Normally to represent a one-year study period the number of subsequent time intervals is very high (over 35,000 15-minute time intervals).

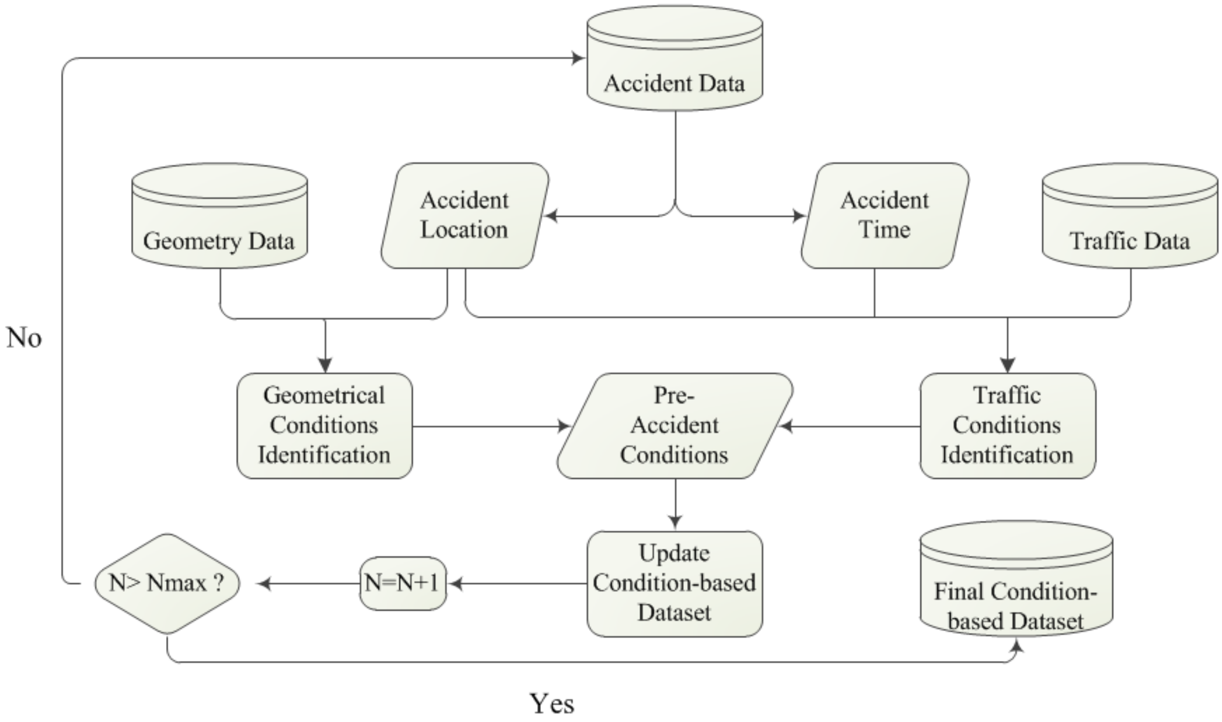


Figure 3.9: Flow chart of the Condition-Based dataset development process.

3.5.1 Traffic conditions identification

The identification of the pre-accident traffic conditions when an accident occurred was based on the reported spatial and temporal information for this accident. Firstly, each accident was matched with the unique set of traffic measurements that includes the location, date and time of the accident (i.e. speed (S_{acc}) and volume (V_{acc})). The traffic measurements of this interval are not equally representative for all accidents. For instance, if an accident occurred at the beginning of a time interval the traffic measurements are probably affected by the congestion caused after the accident. In this case, the traffic measurements just before the interval that includes the accident time (i.e. speed (S_{before}) and volume (V_{before})), are probably more related to the pre-accident conditions. In order to form a representative and comparable set of measurements for all accidents, each accident was matched with traffic data equivalent to one full traffic measurement interval (T). Therefore, the final accident speed (S_w) and volume (V_w) were estimated using a weighted average (see Equations 3.4 and 3.5) of the T-minute interval that included the time of the accident (second interval) and its precedent (first interval).

$$S_w = \left(\frac{t}{T}\right) S_{acc} + \left(1 - \frac{t}{T}\right) S_{before} \tag{3.4}$$

$$V_w = \left(\frac{t}{T}\right) V_{acc} + \left(1 - \frac{t}{T}\right) V_{before} \tag{3.5}$$

Where: S_w and V_w: Weighted average of Speed (mph) and Volume (vehicles), S_{acc} and V_{acc}: Speed (mph) and Volume (vehicles) measurements of the second interval, S_{before} and V_{before}: Speed (mph) and Volume (vehicles) measurements of the first interval, t: time difference between the start of the second interval and the reported accident time (minutes) and T: traffic data measurement interval (minutes).

3.5.2 Geometry conditions identification

The configuration of the roadway just before the accident location can be related with the accident occurrence. The precise critical length of a road segment upstream an accident location is not known though. Assuming that the road geometry is particularly important from the moment that a driver decides to stop the vehicle until the vehicle eventually collides, the length of the road that was considered was the stopping distance upstream of the identified accident location on the link. The stopping distance was estimated based on the annual average speed of motorways and A-roads separately using equation 3.6 (Elvik et al., 2004):

$$S_D = R_D + B_D = t_r v_0 + \frac{{V_0}^2}{2 f_k g} \tag{3.6}$$

Where S_D: Stopping distance (m), R_D: Reaction distance (m), B_D: Braking distance (m),t_r: reaction time (here: 1.5 sec),v_0: average speed (m/s), V_0: average speed (km/h), f_k: friction (here: 0.8, average tire on dry pavement), g: gravity acceleration (here: 9.8 m/sec^2).

Based on the above equation, the stopping distance was estimated 97 and 75 metres (0.06 and 0.047 miles) for motorways and A-roads respectively. To correct for errors in the accident location, the final road segment for each accident included the length of the stopping distance upstream of the accident location and 20 metres (0.012 miles) downstream (error distance). The length of the error distance was empirically defined. Figure 3.10 illustrates an example of the road segment that is considered for obtaining the geometrical conditions of each accident.

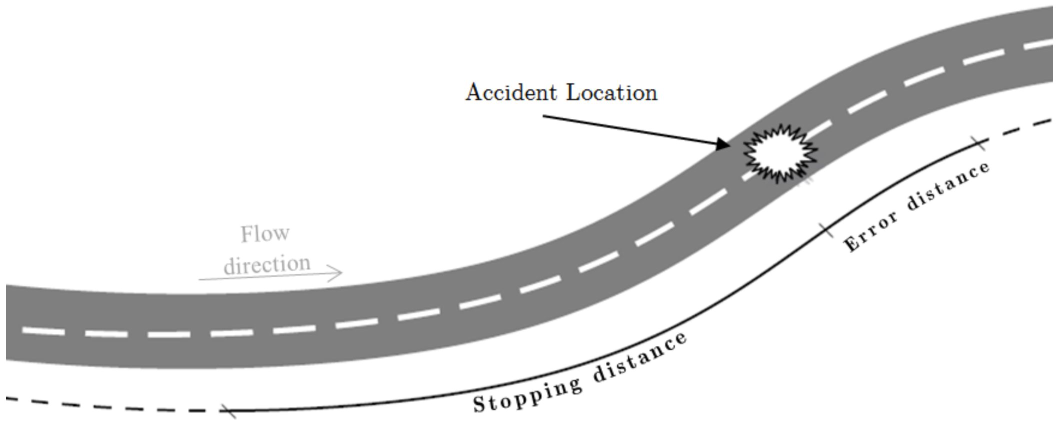


Figure 3.10: Road length upstream and downstream of an accident location for defining the road geometry that is considered for each accident.

The configuration of each road segment was specified by its corresponding geometric measurements. Each of the final road segments included a number of successive radius and gradient measurements that were converted to categorical variables so as to keep the number of scenarios of the final dataset relatively low. Thus, accidents were considered to occur on straight segments if the majority of the radius measurements of the segment were equal to the maximum horizontal curve (i.e. 2000 m) (Highways Agency, 2002) and on curves otherwise. Similarly, accidents that occurred on uphill segments were considered those that included mainly positive grades, on downhill those that include mainly negative grades and otherwise on level segments. The road width was represented with another dummy variable that separated road segments with more than two lanes from segments with up to two lanes.

3.5.3 Condition-based dataset formation

Apart from the accident data, to generate a condition-based dataset it is necessary to employ the entire range of data that represent the conditions on the network. The scenarios of a condition-based dataset should represent all the conditions that existed on the network independent of whether these were associated with accidents or not. That is why before generating a condition-based dataset the range and the distribution of the measurements of the variables that will be used to represent the network should be known. The process of the development of the condition-based dataset that is presented here might not be the only way for constructing a condition-based dataset. However, the presentation, analysis and comparison of different data combination methods are out of the scope of this thesis.

To facilitate controlling for the exposure, all the scenarios of the condition-based dataset were chosen to have an equal likelihood of occurrence during the examined study period. To achieve this, the continuous variables that are included in the dataset (i.e. speed and volume) were divided into groups of equal frequency defined by percentile ranges with a pre-specified constant step n (e.g. N^{th} percentile - $(N+n)^{th}$ percentile, $(N+n)^{th}$ percentile - $(N+2n)^{th}$ percentile, $(N+2n)^{th}$ percentile - $(N+3n)^{th}$ percentile etc.). Each group was represented in the dataset by an appropriate representative value (e.g. a central tendency statistic). In this way, for every continuous variable C_i there was a number of K_i equally likely distinct groups of observations (where $K_i = \frac{100}{n}$). Every discrete variable D_j had by default a number of categories L_j. To develop a dataset that includes every possible combination between all the variables the number of scenarios (S) that should be generated equals:

$$S = \prod_{i=1}^{I} K_i \prod_{j=1}^{J} L_j \tag{3.7}$$

The number of the scenarios of the dataset is not fixed and can be adjusted to serve the analysis' needs by selecting a smaller step n to decrease the number of scenarios and vice versa. Considering that a condition-based model will be analysed using count modelling techniques it is good to avoid generating too many scenarios that may result in a dataset with an excessive number of zeros. A count dataset with too many zeros is likely to lead to underestimation of the standard errors of the regression parameters (Hilbe, 2011).

After the scenarios of the dataset were specified, the number of accidents per scenario was estimated. Using the traffic and geometric conditions, each accident was matched with the most similar scenario, generating a dependent count variable.

3.5.4 Exposure

The fact that all the scenarios of the condition-based dataset are equally likely to occur, does not mean that they have equal accident probability. Consequently, the exposure cannot be considered as uniform among the condition scenarios. To enable comparisons between the scenarios, an exposure variable was employed.

Accident probability is proportional with the probability of accident prone interactions between vehicles on the network (Chipman et al., 1992; Navon, 2003). The number of such vehicle encounters at a particular condition scenario increases as the number of vehicles and the duration of their stay under these conditions rise. In order to control for this effect, the offset variable for the condition-based dataset was set to be the average vehicle-hours spent per scenario (i.e. average travel time per mile multiplied by average volume).

3.6 Modelling Accident Counts

Count variables refer to the number of times that an event occurred during a pre-defined time period (e.g. annual number of crimes per block or annual number of accidents per condition scenario) and they consist of nonnegative integer values. Count data often have relatively low mean values and are heteroscedastic. Typically their distributions are left skewed and kurtotic. Ordinary Linear Regression (OLS) models are usually not suitable for modelling count data as they might lead to biased standard errors and tests of significance. Moreover, an OLS model might produce negative predicted values, which are theoretically impossible (Hilbe, 2011; Cameron and Trivedi, 2013). The most suitable model , in terms of statistical properties, for modelling counts is Poisson regression and its extensions (Maher and Summersgill, 1996; Lord and Mannering, 2010).

The following sections describe the statistical models that were employed for this analysis. Despite their different data generation mechanisms, both the link-based and the condition-based datasets are cross sectional count datasets and will be modelled using the same approaches. Most of the models presented in this thesis were derived using multivariate regression (i.e. more than one dependent variables) that is employed to analyse accidents disaggregated by type (e.g. severity, collision type etc.). Sections 3.6.1 and 3.6.2 give an overview of the univariate Poisson and the Poisson lognormal regression modelling.

3.6.1 Poisson regression

In a Poisson regression model the dependent variable (i.e. number of accidents) Y_i is assumed to be Poisson distributed ($Y_i \sim Poisson$) with a mean λ_i that is a function of the M independent variables of the model x_{im} and their parameters β_m. The probability distribution function of Y_i is:

$$P(Y_i = y_i) = \frac{e^{\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{3.8}$$

Where:

$$\lambda_i = z_i e^{\beta_0 + \sum_{m=1}^{M} \beta_m x_{im}}, m = 1, 2, 3, ...M \tag{3.9}$$

Or

$$ln(\lambda_i) = \beta_0 + \sum_{m=1}^{M} \beta_m x_{im} + ln(z_i) \tag{3.10}$$

Where:

Y_i: the number of accidents at the i_{th} observation (either link or scenario)

λ_i: the expected number of accidents at the i_{th} observation

x_{im}: the value of the m_{th} explanatory variable at the i_{th} observation

β_m: the coefficient of the m_{th} explanatory variable

β_0: intercept

z_i: the exposure/offset variable at the i_{th} observation

The parameters β_0 and β_m can be estimated either with a frequentist approach using the Maximum Likelihood Estimation (MLE) or with a Bayesian approach using the Markov Chains Monte Carlo (MCMC) algorithm.

Poisson regression has been used for explaining accident occurrences in the literature (e.g. Jovanis and Chang, 1986; Jones et al., 1991; Miaou and Lum, 1993) although its specification is very restrictive. In Poisson regression it is assumed that the mean is equal to the variance ($E(Y_i) = Var(Y_i)$); a condition also known as equidispersion. Count variables are rarely actually equidispersed though. A common way to control for overdispersion (i.e. variance is higher than the mean) in the data is to apply Poisson-mixture models (e.g. Persaud, 1994; Poch and Mannering, 1996; Lord and Miranda-Moreno, 2008; Lord and Mannering, 2010). In these models the mean is considered to be a random variable drawn from a known distribution such as the gamma distribution (Poisson-gamma or negative binomial regression) or the lognormal distribution (Poisson-lognormal regression). The latter model of these models is discussed in the next section.

3.6.2 Univariate Poisson lognormal regression

Overdispersion in count datasets mainly arises from unobserved heterogeneity. To control for heterogeneity it is possible to add a random effect to the Poisson regression model. When the random effect is log-normally distributed the regression model transforms to a Poisson lognormal (PLN). After this addition equation 3.10 becomes:

$$\ln(\lambda_i) = \beta_0 + \sum_{m=1}^{M} \beta_m x_{im} + ln(z_i) + \varepsilon_i \tag{3.11}$$

The random term ε_i is normally distributed:

$$\varepsilon_i \sim N(0, \sigma^2) \text{ or } e^{\varepsilon_i} \sim LN(0, \sigma^2) \tag{3.12}$$

Where σ^2: variance parameter of the unobserved heterogeneity.

3.6.3 Multivatiate Poisson lognormal regression

The univariate PLN model is developed to model single count variables (i.e. all accidents on a road network). To examine the relationships of the independent variables with accidents by severity or collision type it is necessary to take into account the heterogeneous correlations between them. Different accident types cannot be considered independent of each other and modelled separately as such because they are subsets of the total accidents on a road network (e.g. Ma and Kockelman, 2006; Park and Lord, 2007). Multivariate Poisson lognormal (MVPLN) regression has been proposed for modelling simultaneously two or more accident categories. Having a lognormally distributed random effect, like in the PLN regression, MVPLN controls both for over-dispersion and the correlations between accident types (e.g. Park and Lord, 2007; Lord and Miranda-Moreno, 2008; Ma et al., 2008; El-Basyouny and Sayed, 2009*b*; Aguero-Valverde and Jovanis, 2009; Barua et al., 2014). The correlations between accident types can be either positive or negative. In MVPLN the number of accidents for each type K is Poisson distributed:

$$Y_{ik} \sim Poisson(\lambda_{ik}) \text{ for } k = 1, 2, 3, ...K \tag{3.13}$$

Where:

$$\ln(\lambda_{ik}) = \beta_{k0} + \sum_{m=1}^{m} \beta_{km} x_{ikm} + ln(z_i) + \varepsilon_{ik} \tag{3.14}$$

Where:

Y_{ik}: the number of accidents at the i_{th} observation (either link or scenario) for type k

λ_{ik}: the expected number of accidents at the i_{th} observation for type k

x_{ikm}: the value of the m_{th} explanatory variable at the i_{th} observation

β_{km}: the coefficient of the m_{th} explanatory variable for type k

β_{k0}: intercept for type k

z_i: the exposure/offset variable at the i_{th} observation

ε_{ik}: random term (unobserved heterogeneity) for type k which is multivariate normally distributed:

$$\varepsilon_{ik} \sim MVN(\mathbf{0}, \Sigma^2) \tag{3.15}$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2K} \\ \dots & \dots & \ddots & \dots \\ \sigma_{K1} & \sigma_{K2} & \dots & \sigma_{KK} \end{pmatrix} \tag{3.16}$$

Where Σ is the variance-covariance matrix of the unobserved heterogeneity.

The direct computation of the marginal distribution of accident counts is not possible to be obtained directly, because it requires the computation of a K-variate integral of the Poisson distribution with respect to the distribution of ε_{ik} (Ma, 2006). This means that the coefficients cannot be estimated using a maximum likelihood function (MLE). Therefore, the parameter estimation was done via Markov chain Monte Carlo (MCMC) in a Bayesian framework (e.g. Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009). The prior distribution for the parameters β_{km} is:

$$\beta_{km} \sim MVN(\beta_{k0}, R_{\beta_{k0}}) \tag{3.17}$$

The conjugate prior distribution of the inverse of the variance-covariance matrix (i.e. precision matrix) is usually Wishart (e.g. Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009):

$$\Sigma^{-1} \sim Wishart(R, d) \tag{3.18}$$

Where β_{k0}, $R_{\beta_{k0}}$ and R are known non-informative hyperparameters and d is equal to the degrees of freedom (number of the examined accident types: $d = K$).

3.6.4 Spatial correlation

In the models presented above it is assumed that the observations are spatially independent. This is theoretically valid for the condition-based dataset as its observations (i.e. scenarios) are not spatial entities and therefore they have not any defined spatial relationship. However, the observations of the link-based dataset cannot be considered as spatially independent. Consequently, the unobserved spatial relationships between adjacent road links can be controlled by adding a random effect using the conditional autoregressive priors (CAR) model in a hierarchical Bayesian approach (e.g. Wang et al., 2009*a*; Aguero-Valverde, 2013; Quddus, 2013; Barua et al., 2014). As the regression model here is multivariate, the multivariate CAR model will be used (Aguero-Valverde, 2013; Barua et al., 2014). The spatial effect includes a contiguity-based weighting scheme with first-order neighbours (e.g. Aguero-Valverde and Jovanis, 2008; Quddus, 2008). Contiguity-based weights are considered to be more suitable for link-based analyses than distance-based as the latter tend to provide high weights for opposite direction links which do not necessarily have common traffic conditions (Wang, 2010). The equations below present the expression that was used to models the link-based datasets only:

$$\ln(\lambda_{ik}) = \beta_{k0} + \sum_{m=1}^{m} \beta_{km} x_{ikm} + ln(z_i) + \varepsilon_{ik} + u_{ik} \tag{3.19}$$

Where:

u_{ik}: random effect for the spatial correlation between the i_{th} observation (i.e. link) and its neighbours for type k and

Y_{ik}, λ_{ik}, x_{ikm}, β_{km}, β_{k0}, z_i, ε_{ik}: as explained above. The random effect u_{ik} is multivariate normally distributed as proposed by Thomas et al. (2004):

$$u_{ik} \mid u_{jk} \sim MVN\left(\frac{\Sigma_{jk} u_{jk} w_{ij}}{\Sigma_{ik} w_{ij}}, \frac{\Omega}{\Sigma_i w_{ij}}\right), i \neq j \tag{3.20}$$

Where:

w_{ij}: Adjacency weight matrix that denotes $w_{ij} = 1$ if the link i and j are first order neighbours (they share a common boundary) or $w_{ij} = 0$ otherwise.

$$\Omega = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1K}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2K}^2 \\ \dots & \dots & \ddots & \dots \\ \sigma_{K1}^2 & \sigma_{K2}^2 & \dots & \sigma_{KK}^2 \end{pmatrix} \tag{3.21}$$

Ω: is the variance-covariance matrix for the spatial correlation.Where:

$$\Omega^{-1} \sim Wishart(S, d) \tag{3.22}$$

Where S is known non-informative hyperparameters and d is equal to the degrees of freedom (number of the examined accident types: $d = K$).

3.7 Modelling Strategy

This section describes the main steps that were taken to select the models that will be presented in this work. This firstly includes the definition of the dependent variable combinations that are the most relevant to the study. Following, is the generation of the candidate specifications of the independent continuous variables of the models and finally is the development of the final models.

3.7.1 Dependent variable expressions

With multivariate regression it is possible to express and model the dependent variables in multiple ways. A multivariate model can provide different information through its outcomes, as a result of the classification of its dependent variable. Having available accident data by severity and by collision it is possible to explore the accident relationships with potential contributory factors under different perspectives. Considering the data availability and the research questions of the study the statistical models are developed in order to:

- Understand the relationship of traffic and geometric variables with all traffic accidents on the network

- Understand the relationship of traffic and geometric variables with traffic accidents on the network by severity level
- Understand the relationship of traffic and geometric variables with traffic accidents on the network by collision type
- Understand the relationship of traffic and geometric variables with traffic accidents on the network by collision type and severity levels
- Compare the modelling results of the link-based and the condition-based aggregation approaches
- Test the significance of an advanced accident mapping technique in accident modelling

Addressing these issues requires the development of multiple different dependent variable combinations or in other words multiple different models. Table 3.5 presents a list of the six main dependent variable expressions that were formed along with their reference names, a description and the Poisson regression type that is required for modelling.

Table 3.5: Dependent variable combinations

Reference Name	Dependent variable(s)	Regression
All	All the accidents on the network combined	PLN
K-S-Sl	All accidents disaggregated by severity into three categories: Killed (K), Serious injuries (S) and Slight injuries (Sl)	MVPLN
KS-Sl	All accidents disaggregated by severity into two categories: Killed & Serious injuries (KS) and Slight injuries (Sl)	MVPLN
SV-MV[2]	All accidents disaggregated by collision type into two categories: single vehicle (SV) and multiple vehicle (MV)	MVPLN
SV_KS-Sl	Single vehicle accidents disaggregated by severity into two categories: Killed & Serious injuries (KS) and Slight (Sl) injuries	MVPLN
MV_KS-Sl	Multiple vehicle accidents disaggregated by severity into two categories: Killed & Serious injuries (KS) and Slight (Sl) injuries	MVPLN

Both models K-S-Sl and KS-Sl are examining the relationship of accidents split by severity with the dependent variables. Model K-S-Sl provides a more detailed insight on this relationship and is particularly useful for the impact estimation by severity category. The reason for developing model KS-Sl in addition to model K-S-Sl was the high number of zeros in the variable counting fatal accidents. As the number of accidents with fatal casualties was relatively low (this will be shown at section 4.2.2) the observations with zeros were many. Excessive number of zeros in count models is a distributional assump-

[2]Intersection MV accidents defined as accidents where the colliding vehicles had different intended directions were eliminated from the analysis because: a)intersection accidents are assumed to have have significantly different generation processes than the main carriageway ones and b) the small number of observations (4.6% of all accidents) did not permit the formation of an individual category.

tion violation that is associated with erroneous parameter and standard error estimation (Hilbe, 2011). To avoid such errors, fatal and serious accidents were aggregated into one category. This specification was adopted in models SV_KS-S and MV_KS-Sl for the same reason.

Some of the specifications of the Table 3.5 were used to model additional datasets so as to further explore some interesting topics. A way to evaluate the significance of accident mapping in safety modelling is to examine whether datasets based on accident mapping algorithms with different accuracies provide different results. To that end, model KS-Sl was applied on a dataset that was developed using the output of AMF (henceforth: KS-Sl AMF) and a dataset in which the accident locations assigned to the closest segment of the reported locations (henceforth: KS-Sl AMM1). One of the aims is to estimate the impact of a speed limit increase on the motorways of the study network. To isolate the motorways from the A-road sections of the network and give an accurate estimation on the expected changes in accidents after a potential speed limit increase models SV_KS-Sl and MV_KS-Sl were applied to the full network dataset and a separate dataset for motorways only (henceforth: *SV_KS-Sl moto and MV_KS-Sl moto*).

Each of the dependent variable combinations described above was applied separately on the link-based and the condition-based datasets, resulting in a total of 18 (including the additional datasets) models.

3.7.2 Independent variable expressions

The functional form of the relationships of the continuous independent variables of the models (i.e. speed and volume/AADT) with the dependent variable(s) (i.e. accidents) is not known (Qin et al., 2004). The assumption of linear relationships might be inaccurate and lead to false conclusions. In addition, there is no clear evidence on whether the interaction between these traffic variables is related with the number of accidents. In order to avoid any rough assumption on the form of the relationships and to control for a possible interaction between speed and volume, 20 different independent variables specification combinations were tested. The expressions included combinations of speed and volume in linear, squared, logarithmic and quadratic forms with and without an interaction term. The rest of the independent variables of the models are dummy variables and so they remained the same for all the expressions. The specification combinations that were examined are listed in Table 3.6 that shows for each combination the specification of speed and volume and whether an interaction term was included. The final column presents the part of the link function that includes speed and volume to indicate the number of variables per specification combination.

Table 3.6: Independent variable specification combinations.

Variable Specification Number	Speed Specification	Volume /AADT Specification	Interaction	Expression
1	linear	linear	No	$\beta_1 speed + \beta_2 volume$
2	linear	squared	No	$\beta_1 speed + \beta_2 volume^2$
3	linear	logarithmic	No	$\beta_1 speed + \beta_2 ln(volume)$
4	squared	linear	No	$\beta_1 speed^2 + \beta_2 volume$
5	squared	squared	No	$\beta_1 speed^2 + \beta_2 volume^2$
6	squared	logarithmic	No	$\beta_1 speed^2 + \beta_2 ln(volume)$
7	logarithmic	linear	No	$\beta_1 ln(speed) + \beta_2 volume$
8	logarithmic	squared	No	$\beta_1 ln(speed) + \beta_2 volume^2$
9	logarithmic	logarithmic	No	$\beta_1 ln(speed) + \beta_2 ln(volume)$
10	quadratic	quadratic	No	$\beta_1 speed + \beta_2 speed^2 + \beta_3 volume + \beta_4 volume^2$
11	linear	linear	Yes	$\beta_1 speed + \beta_2 volume + \beta_3 speed \cdot volume$
12	linear	squared	Yes	$\beta_1 speed + \beta_2 volume^2 + \beta_3 (speed \cdot volume)$
13	linear	logarithmic	Yes	$\beta_1 speed + \beta_2 ln(volume) + \beta_3 (speed \cdot volume)$
14	squared	linear	Yes	$\beta_1 speed^2 + \beta_2 volume + \beta_3 (speed \cdot volume)$
15	squared	squared	Yes	$\beta_1 speed^2 + \beta_2 volume^2 + \beta_3 (speed \cdot volume)$
16	squared	logarithmic	Yes	$\beta_1 speed^2 + \beta_2 ln(volume) + \beta_3 (speed \cdot volume)$
17	logarithmic	linear	Yes	$\beta_1 ln(speed) + \beta_2 volume + \beta_3 (speed \cdot volume)$
18	logarithmic	squared	Yes	$\beta_1 ln(speed) + \beta_2 volume^2 + \beta_3 (speed \cdot volume)$
19	logarithmic	logarithmic	Yes	$\beta_1 ln(speed) + \beta_2 ln(volume) + \beta_3 (speed \cdot volume)$
20	quadratic	quadratic	Yes	$\beta_1 speed + \beta_2 speed^2 + \beta_3 volume + \beta_4 volume^2 + \beta\_5(speed \cdot volume)$

The above specification combinations were tested for each of the 18 dependent variable expressions resulting to a total of 360 models. The functional form with the best goodness-of-fit statistics is considered as the most accurate representation of each dependent variable expression and thus for brevity only these 18 models will be presented (either in Chapter 5 or in the Appendix C). Each of the models will be named according to its dependent and independent variable expressions and the modelling approach that was used as follows:

- The first part of the name will indicate the accident data aggregation approach that was used (i.e. Link-based or Condition-Based);
- The middle part of the name will show the dependent variable expression and
- The last part of the name will show the best fitting dependent variable specification.

For example if the best fitting specification for the KS-Sl model for the condition-based dataset is specification 17 the model will be presented with the name *Condition-Based KS-Sl (17)*.

3.7.3 Deviance information criterion

The Deviance Information Criterion (DIC) is a goodness of fit statistic that is used for comparisons of models estimated on a full Bayesian inference approach. It is a generalisation of the Akaike Information Criterion (AIC) that is used for frequentist approaches (see Hilbe, 2011). DIC assesses models in terms of goodness-of-fit (deviance) and complexity (deviance of posterior means). The best fitting model is considered to be the most parsimonious; a model that accomplishes a good level of explanation of the data with the lowest number of independent variables possible. This model will have the smallest DIC among all the possible models (Spiegelhalter et al., 2002). The mathematical formulation of DIC is:

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \tag{3.23}$$

Where:

$D(\bar{\theta})$: deviance of the θ posterior means of the model parameters

p_D : the effective number of parameters in the model

$\bar{D}$: is the posterior mean of the deviance, $D(\bar{\theta})$.

3.8 Summary

This chapter presented the employed data processing and analytical methods. The correspondence between the objectives of this thesis and the methods were shown in the research design section.

To correct accident locations which are likely to be inaccurate, an advanced accident mapping algorithm (termed as AMF) was developed. The algorithm that is based on a fuzzy inference system and road name filtering will be evaluated and compared with less advanced accident mapping algorithms to confirm that its results are satisfying for the study dataset. The chapter provides a detailed description of the link-based accident data aggregation approach, that is the conventional approach in accident analyses. The condition-based accident data aggregation approach was also introduced for the first time. The condition-based approach enables the representation of the pre-accident conditions, addressing in this way the aggregation bias that is linked with the link-based approaches.

Following, the statisitcal methods that were used in order to model both the link-based and the condition-based datasets were presented. Generally, accident count datasets are preferably modelled using Poisson regression or one of its many variations. To control for the unobserved heterogeneity and to enable modelling of accidents split by accident type (i.e. severity or collision type) Multivariate Poisson lognormal regression models will be employed. The link-based models will include an additional multivariate random effect that will control for spatial correlation. The final part of this chapter included the modelling strategy that was followed in order to select and develop the final models that will be presented in the subsequent chapters.

Chapter 4

Data Description and Pre-Processing

4.1 Introduction

In data-driven studies, like this thesis, data quality and availability play a key-role in the validity and the clarity of the outcomes. To conduct the analysis, secondary datasets collected from multiple sources were processed, combined and analysed.

In this chapter, the features and the challenges of the datasets that were employed for the analysis will be described. Following, the results of the accident mapping algorithm will be presented and evaluated. Finally, the chapter will introduce the final link-based and condition-based datasets that were used for the statistical models.

4.2 Data Description

4.2.1 The study road network

The network that was used for this analysis is the Strategic Road Network of England (SRN). The SRN is the busiest network of the country as it consists of all the motorways and the major A-roads (see Figure 4.1). Although its total length is 4,272 miles (2.4% of all roads), the SRN carries 30% of all traffic and more than 65% of road freight annually. The SRN includes routes that have a strategic role for the country and connects most of the key locations in the UK. It is considered that at least a part of every national-level journey that takes place in the country is on the SRN (Department for Transport, 2011*b*).

The size, coverage and significance of the SRN make it a representative and therefore suitable study area for analysing the relationship of accidents with traffic and geometric characteristics.

Figure 4.1: Map of the Strategic Road Network (SRN) of England. (source:(Highways Agency, 2015))

In this work, the SRN is represented by two separate digital models: HATRIS (Highways Agency Traffic Information System) and HAPMS (Highways Agency Pavement Management System). Each of the models has been built for different purposes and is associated with different types of data related to the network; HATRIS with traffic measurements and HAPMS with the geometric characteristics.

HATRIS is the base network for the Traffic Flow Data System (TRADS) and the Journey Time Database (JTDB). HATRIS represents the SRN using a system of 2,505 junction-to junction links (average link length 3.25 miles) and nodes (Highways Agency, 2011). This representation of the network cannot provide details on the road configuration as it can be seen at Figure 4.2. The main features of HATRIS links are:

- Link ID
- Coordinates of starting node of the link
- Coordinates of ending node of the link
- Road Name
- Link length

HAPMS is a computer-based model of the SRN that is used for recording network, construction, definitive inventory, traffic, accident and condition data on a single database. HAPMS information is used for national, regional and area reports. HAPMS is represented by a section referencing system that divides the network into 20,734 sections with consistent road characteristics (road type, name, number of lanes etc.) and specified starting and ending points (Highways Agency, 2008).

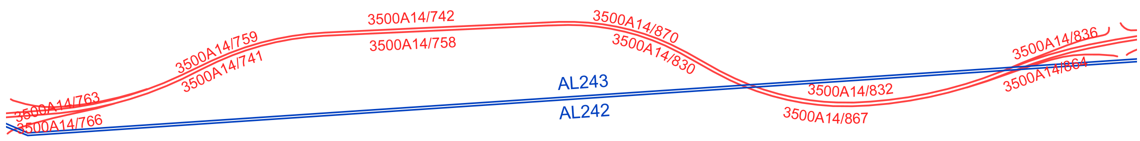
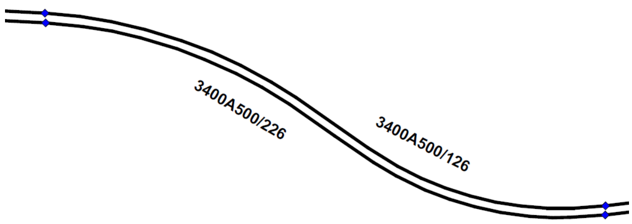


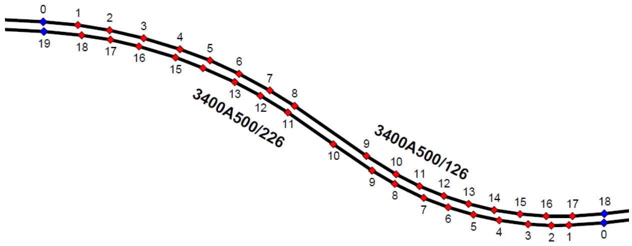
Figure 4.2: Representation of a road section of the A14 according to HATRIS (blue) and HAPMS (red).

The representation of the actual road configuration is quite realistic in the HAPMS model (see Figure 4.2) therefore the HAPMS network map is more suitable for accident mapping. In order to decrease the length of individual road sections so as to improve the road direction accuracy and consequently to increase the probability of mapping each accident to the accurate location, HAPMS sections were divided into smaller segments defined by their shape-points. Shape-point coordinates by section label were extracted in a GIS environment and a new set of nodes and segments was developed using the method proposed by Quddus (2006). In this way, the network that initially comprised 20,734 road sections with average length 0.462 miles, was divided into 211,247 piece-wise straight segments with average length: 0.045 miles. As an example of this process, Figure 4.3(a) represents two road sections of the HAPMS network on the A500 main carriageway (3400A500/126 and 3400A500/226). The start and the end nodes of these sections are displayed by the rhombuses. Figure 4.3(b) represents these two sections divided by their shape points into 18 and 19 smaller road segments respectively. The boundaries of each segment are illustrated by the numbered points (0 to 1, 1 to 2, and so forth). The features of HAPMS segments that are used are:

- Coordinates of starting node of the segment
- Coordinates of ending node of the segment
- Road Name
- Road Type
- Section Label
- Segment ID
- Speed Limit
- Number of lanes



(a)



(b)

Figure 4.3: Road sections on the A500 (a) full and (b) divided into their shapepoints

4.2.2 Accident data

Information about the accidents that occurred on the study area was obtained from the national accident database, STATS19. STATS19 data are collected by the Police and include all accidents that accounted for at least one injured casualty (Department for Transport, 2011*c*). The accident database for this study comprises all 10,520 STATS19 reports of accidents that occurred during 2012 on the SRN. The most important variables of the database that were used for the analysis are:

- Accident Reference Number: A unique seven-digit sequence that is used to distinguish road accidents.
- Accident Date (see Table 4.1)
- Accident Time
- Location: A pair of six-digit coordinates (easting and northing) obtained by the Ordnance Survey Grid map.
- Number of vehicles involved (see Table 4.2)
- Accident Severity: Indicates the most serious outcome of the accident. It can be *Fatal* if at least one of the involved casualties was killed in less than 30 days as a result of the accident and *Serious* or *Slight* if at least one of the casualties was seriously or slightly injured respectively. (see Table 4.4 and 4.3)
- 1st Road Class: The class of the road where the accident occurred. In the SRN Road Class can be either motorway (M) or main single carriageway (A).
- 1st Road Number: The number that corresponds to the road where the accident occurred.
- Road Type: Roundabout, one way street, dual carriageway, single carriageway, slip road or unknown.
- Speed Limit: The posted speed limit on the road where the accident occurred.
- Junction Detail: This variable expresses the proximity of the accident location to a junction. If the accident took place on a roadway section that is located less

than 20 metres from a junction, then the junction type should be reported (e.g. Roundabout, slip road, junction with more than four arms etc.).

- 2nd Road Class: The class of the intersecting road (if any).
- 2nd Road Number: The number that corresponds to the intersecting road.
- Vehicle Movement Compass Point: The intended direction of the vehicle just before the incident measured with compass and reported using the four cardinal points and their intermediates (N, NE, E, SE ...etc.)

Table 4.1: Percentage of accidents on the SRN by month.

Month	Accident Percentage (%)
January	7.47
February	7.32
March	7.72
April	7.96
May	7.82
June	7.78
July	9.33
August	9.05
September	8.66
October	8.63
November	9.55
December	8.72

Table 4.2: Percentage of accidents by number of vehicles involved.

Number of involved vehicles	Accident Percentage (%)
1	22.97
2	54.09
3	15.15
4 or more	7.78

Table 4.3: Percentage of accidents by severity.

Accident Severity	Accident Percentage (%)
Fatal	1.91
Serious	11.65
Slight	86.43
Total	100

Table 4.4: Percentage of accidents by number of casualties.

Number of casualties	Accident Percentage (%)
1	64.91
2	22.09
3	7.7
4	3.08
5 or more	2.22

4.2.2.1 Accident data limitations

STATS19 data are collected for organisational reasons rather than scientific and consequently they may include inaccuracies (Loo, 2006). The reason for this is that accident reports are often completed by police officers who arrive at the accident scene after some time and are possibly unable to evaluate the actual conditions or to identify precisely accident location or time. They might also have to complete the form in minimal time as they need to manage their other duties in parallel (e.g. opening blocked lanes) and so, the chance of making mistakes is relatively high (e.g. road name or type misreporting). Before proceeding to analyses that rely on STATS19 data, the inherent limitations should be known and addressed when this is possible. The two most significant limitations of the accident data that affect this analysis is the accident location and the accident time.

The importance of accident locations for safety statistical analyses has been discussed extensively in Section 2.5.1. In STATS19 reports, accident locations coordinates were less accurate than desired; the point that represents an accident rarely falls exactly onto a HAPMS section and when it does it is not guaranteed that this section is correct. To overcome this limitation the fuzzy logic accident mapping algorithm (AMF) that is outlined at Section 3.3 was implemented. The results of the algorithm are presented in the following Section 4.3.1.

Accident time is relatively difficult to be exactly determined by the police officers and consequently the accident time in STATS19 tends to be rounded. This is a known problem of accident databases that has been also reported by Kockelman and Ma (2007). In STATS19, accident time is reported with an hours-minutes format (i.e. HH:MM). Figure 4.4 presents the distribution of the second part of the reported time (i.e MM from 00 to 59) of the examined accidents. It can be seen that the distribution is clustered around the 5's which is theoretically invalid. The actual distribution of accidents is expected to be uniform for all the minutes as it is presented by the horizontal bar in Figure 4.4. It is not possible to correct accident time, however being aware of this limitation is important for defining the traffic measurements' interval that is suitable for identifying the pre-accident traffic conditions.

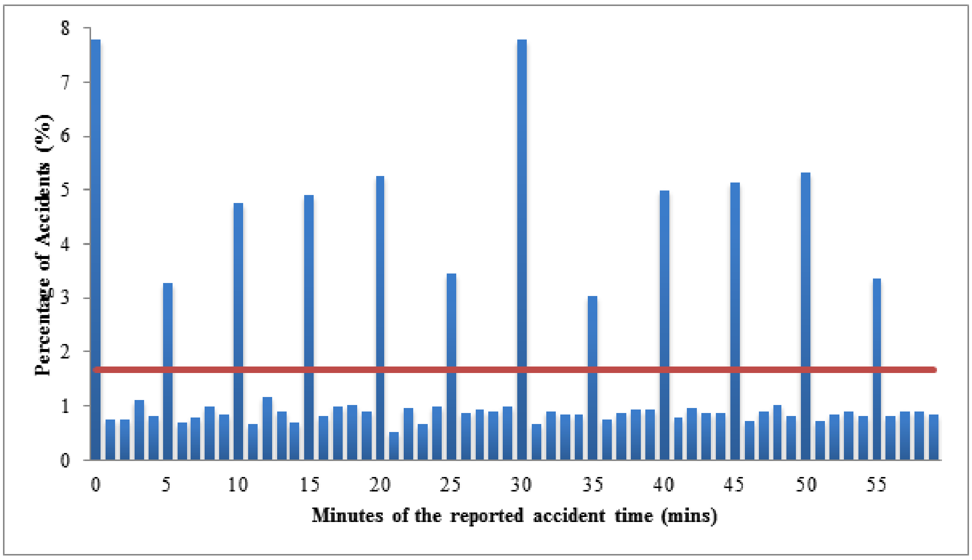


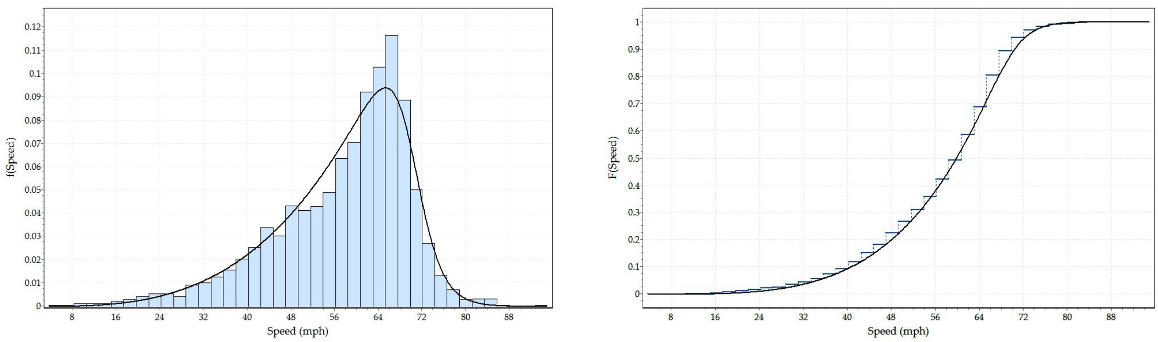
Figure 4.4: Accident distribution per minutes of the reported accident time (the horizontal bar shows the expected percentage per minute group (1.67%) if the distribution of accidents was, as expected, uniform)

4.2.3 Traffic Data

Traffic data were extracted from the UK Highways Agency Journey Time Database (JTDB) which is based on the HATRIS network and includes link-level traffic information obtained by inductive loop detectors for the entire SRN. The measurement interval is 15 minutes resulting in a dataset of approximately 88 million observations (Highways Agency, 2011). The variables used for this analysis are:

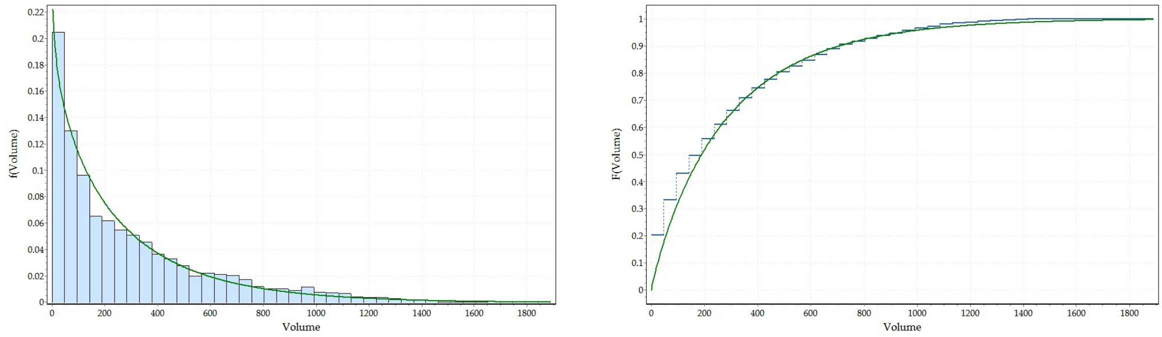
- 15-minute average speed (mph)
- 15-minute volume (vehicles)
- 15-minute average travel time (seconds).

Figures 4.5-4.7 show the histograms and the cumulative distributions of speed, total volume and volume per lane of all links of the SRN and Figures 4.8-4.10 for all the motorway links of the SRN. Additionally the figures present the best fitting probability distribution line among 55 probability distributions according to the Kolmogorov-Smirnov test (Massey, 1951) as it was estimated by the EasyFit software.



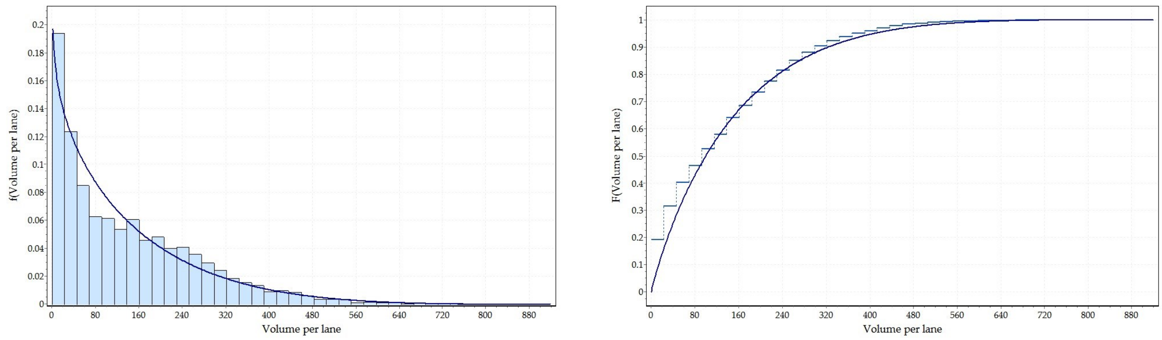
(a) (b)

Figure 4.5: (a) Annual distribution and (b) cumulative distribution of speed on the SRN ($\sim Dagum(0.15, 28.65, 113.64)$)



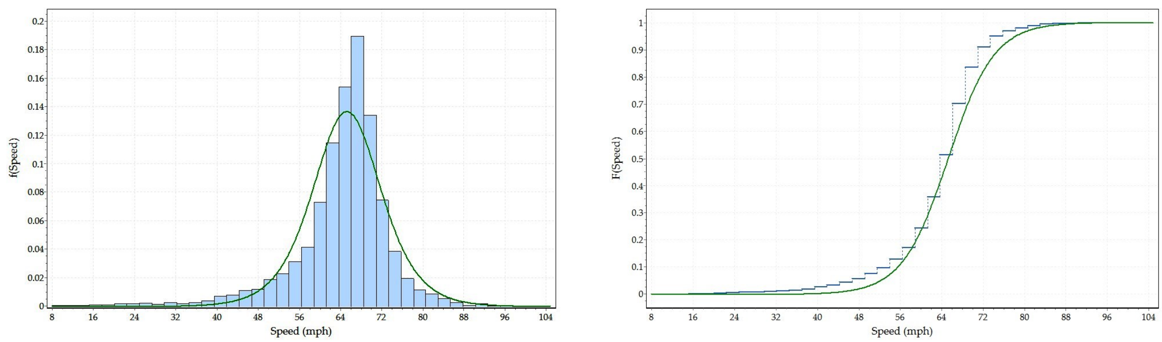
(a) (b)

Figure 4.6: (a) Annual distribution and (b) cumulative distribution of volume on the SRN ($\sim Weibull(0.91, 282.61, 0.25)$)



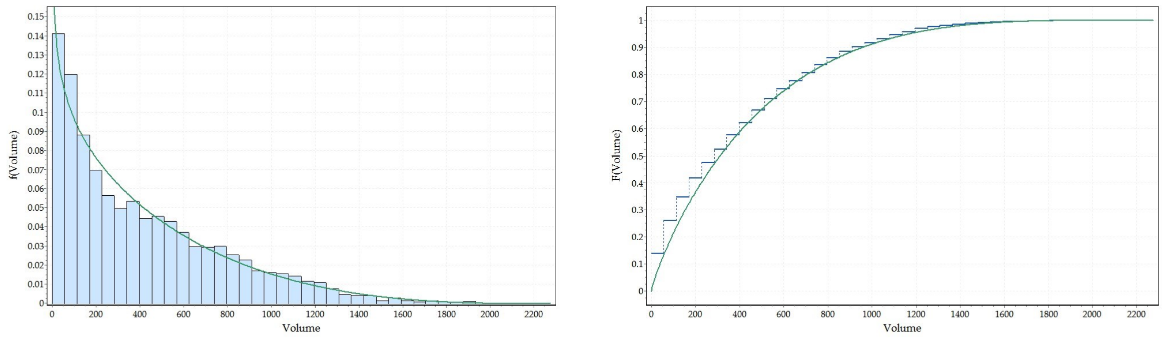
(a) (b)

Figure 4.7: (a) Annual distribution and (b) cumulative distribution of volume per lane on the SRN ($\sim Kumaraswamy(0.91, 5.97, 0.25, 1141.8)$)



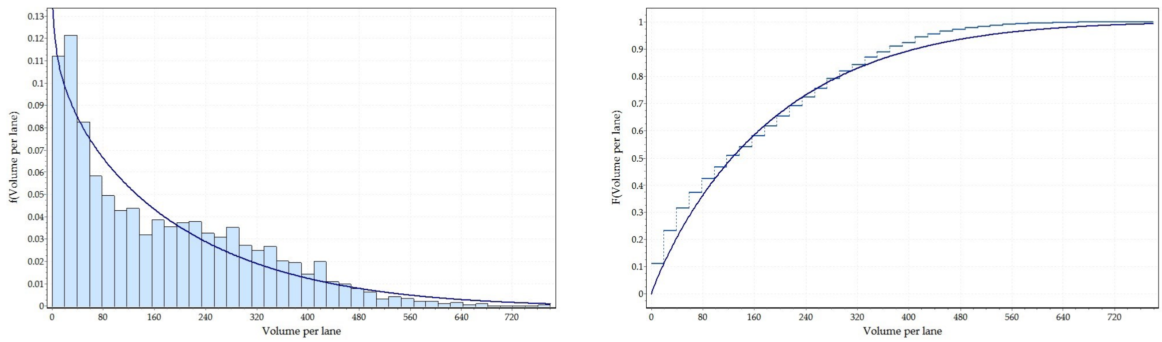
(a) (b)

Figure 4.8: (a) Annual distribution and (b) cumulative distribution of speed on motorways ($\sim four-parameter Dagum(0.40, 803.91, 3639.4, -3527.2)$)



(a) (b)

Figure 4.9: (a) Annual distribution and (b) cumulative distribution of volume on motorways ($\sim Beta(0.85, 3.78)$)



(a) (b)

Figure 4.10: (a) Annual distribution and (b) cumulative distribution of volume per lane on motorways ($\sim Kumaraswamy(0.92, 7.07, 0.13, 1653.2)$)

4.2.3.1 Traffic data limitations

Spatial and temporal aggregation of traffic data enables the reduction of the data volume but at the same time causes data losses. Within a quarter of an hour traffic conditions on the roadway can change more than one time and these changes might not be possible to be represented by a 15-minute average. The resolution of the traffic data is not ideal for defining precisely the traffic conditions just before accidents that is needed for condition-based modelling and it can also affect the accuracy of the average traffic measures that are used for link-based modelling.

Despite that, this aggregation level counterbalances the existing error in accident time data (see Section 4.2.2.1). Since the exact accident time is not known, to ensure that the traffic conditions prior to an accident are captured it is necessary to use traffic data that cover the span of the accident time error. Thus, even if more disaggregated traffic data were available (e.g. every 1 minute or 30 seconds), it would be needed to be aggregated at least to the 5-minute level so as to represent the pre-accident period.

4.2.4 Geometry data

Road configuration was determined based on the UK Highways Agency Traffic Speed Condition Survey database (TRACS). TRACS contains measurements of the geometric characteristics (i.e. curvature, gradient and crossfall) of all HAPMS sections with a 10-metre measurement interval. The measurements are obtained from survey vehicles instrumented with lasers, video image collection and inertia measurement apparatus (Highways Agency, 2008). The variables extracted from this database are:

- Road Curvature expressed by Radius (miles) (Maximum Radius=1.243 miles (2000 metres))
- Gradient expressed by slope percentage.

4.3 Data Refinement and Pre-processing

4.3.1 Accident mapping algorithm results

AMF was implemented to all the 10,520 STATS19 accidents. It has been found that the time required to process 10,520 accidents is 230 minutes (by using a laptop PC with 4GB RAM and 3.4GHz processing speed). This suggests that the developed method can process 46 accidents per minute. The accuracy of the four algorithms (AMM1, AMM2, AMM3, AMF) was then evaluated using the reference 716 accidents discussed in section 3.3.5; the accuracy that was estimated for each algorithm was the percentage of reference cases that were assigned to the same road segment with the one that was selected manually. The percentage of accuracy was estimated at the 95% confidence level with a confidence interval of± 1.1%. All the results can be found at Table 4.5. The total accuracy levels were found to be 81.6%, 87.7%, 85.0% and 98.9% for the AMM1, AMM2, AMM3 and AMF algorithms respectively. The percentages of accuracy for the three main reported road types (roundabouts, slip roads and main carriageways respectively) were also estimated in order to identify possible weaknesses of the method in the identification of specific categories of road accidents. It was revealed that AMM1 and AMM3 face particular difficulty in identifying the correct location for accidents that occurred on roundabouts and that AMM2 has the same problem for main carriageway accidents. In contrast, AMF gives accurate results for roundabout and slip road accidents whereas, for main carriageway accidents the results are less accurate. The mismatches on main carriageways are mainly due to the errors of the reported road name and type of the accident database. The mean Distance for each of the methods was calculated and found to be: 4.43 m for AMM1, 6.47 m for AMM2, 14.54m for AMM3 and 8.53 m for AMF.

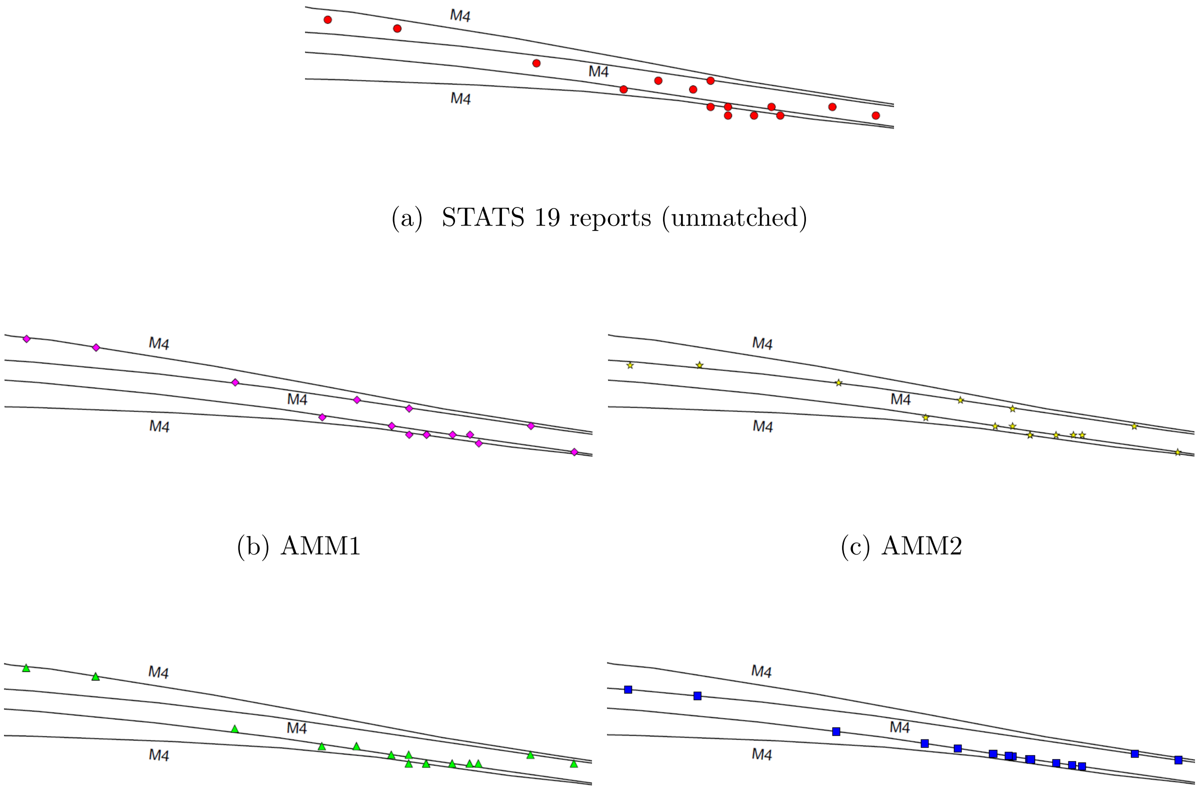
It is clear that the AMF method with error just above 1% gives the most reliable matching results among the examined methods. An interesting outcome is that the correct segment is not always the closest to the reported accident location that is in line with other studies (Loo, 2006; Deka and Quddus, 2014; Imprialou et al., 2015). An additional result to that was that even the closest segment that has the same road name and road type with the examined accident can be erroneous. This highlights the importance of the intended vehicle direction as a variable for an accident mapping algorithm. However, from the results of the AMM3 method that considers the vehicle's intended direction it is revealed that the inclusion of this variable in an inflexible formula does not guarantee the accuracy of the results. This supports the selection of fuzzy inference systems for accident location identification that provide a flexible framework that adapts to the reported data of each case individually. For the AMF method, the 99th percentile of the Distance was found to be 56.8 m and the 98.5th percentile 49.9 m confirming the validity of the selection of the 50 m threshold boundary. In other words, the fact that 98.5% of the cases have Distance less than 49.9 m justifies the need for a manual check in order to confirm the accuracy of the segment selection (as it is described at the Additional Steps 1 and 3 in section 3.3.4)) when the Distance from the selected segment is over 50 m.

Table 4.5: Estimation of accuracy and average distance for the four examined accident mapping methods

Method	Roundabouts (%)	Slip Roads (%)	Carriageways (%)	Total Accuracy (%) $(95\%, 1.1)$	Average Distance (m)
AMM1	74.0	80.3	82.3	81.6	4.63
AMM2	98.0	96.7	86.0	87.7	6.47
AMM3	52.0	96.7	88.3	85.0	14.54
AMF	100.0	100.0	98.7	98.9	8.53

From the 10,520 cases that were matched with the AMF method, there were 266 (2.5%) that needed manual checking according to the additional steps 1 and 2. 14 of them were accidents that could not be matched with any segment of the database due to simultaneous road name and road type mismatch with all the potential candidate segments. From the 266 cases that were checked manually, 107 (1%) needed manual correction. From the entire database, there were overall 206 (2%) cases of accidents that were matched to road segments that had different road names than the reported, and 557 (5.3%) segments that had different road types than reported. After the 107 manual corrections, there were 36 (0.3%) cases where neither the road name nor type of the selected road segment was the same with those referred to on the accident report. This situation indicates the existence of some inconsistency between the network and the accident database that is mostly responsible for the estimated error of the developed method.

Figures 4.11 and 4.12 represent graphically the accident locations when they are superimposed on the digital road network; before accident-mapping (a) and after the implementation of the AMM1(b), AMM2 (c), AMM3(d) and AMF(e) algorithms respectively. It can be easily noticed that the majority of the reported accident locations (Figures 4.11(a) and 4.12(a)) do not fall exactly onto a road segment and some of them are placed between two or more road sections. The accident locations indicated by the four methods (Figures 4.11(b)-4.11(e) and 4.12(b)-4.12(e)) have both similarities and differences. The locations of AMM1 and AMM2 (Figures 4.11(b), 4.11(c), 4.12(b) and 4.12(c)) are very similar to each other, as it was expected, but they are quite different from the AMF (Figures 4.11(e) and 4.12(e)) locations as the accidents are not necessarily assigned to the closest road segment, but the one that has the highest Matching Score. The locations of the AMM3 are almost identical to those indicated by AMF on the main carriageway accidents (Figure 4.11(d)) however the locations for the roundabout accidents (Figures 4.12(d)) are very different as AMM3 was found to have only 52.0% of accuracy for roundabouts.



(a) STATS 19 reports (unmatched)

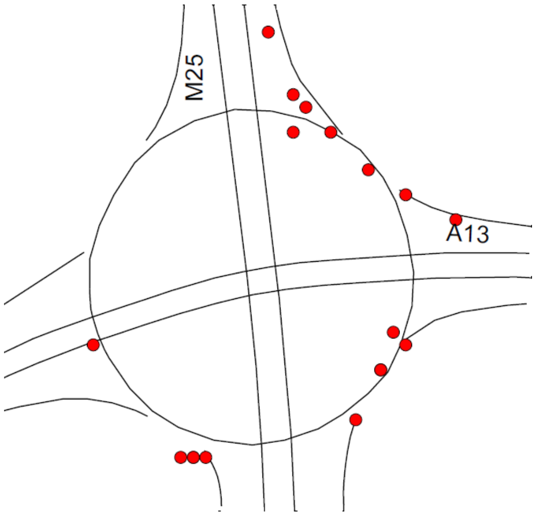
(b) AMM1

(c) AMM2

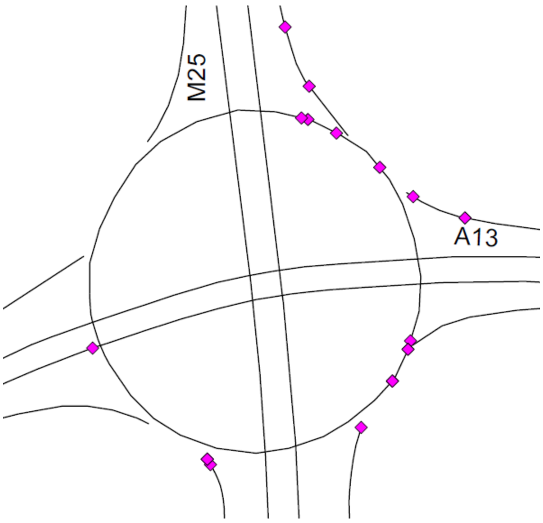
(d) AMM3

(e) AMF

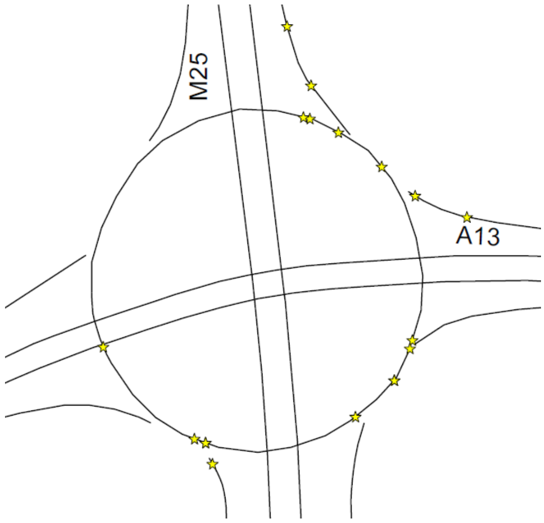
Figure 4.11: Accident locations at a segment of the M4 motorway



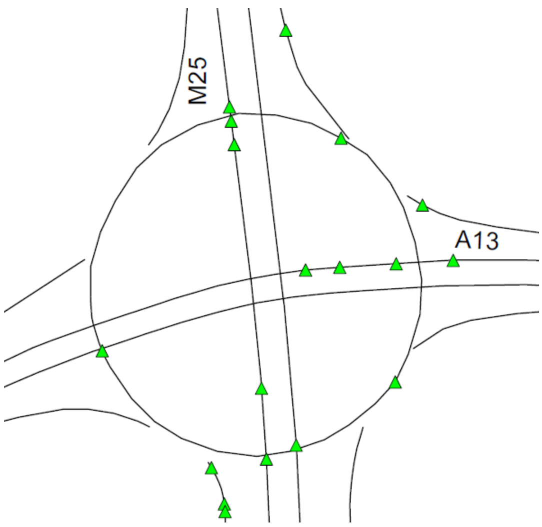
(a) STATS 19 reports (unmatched)



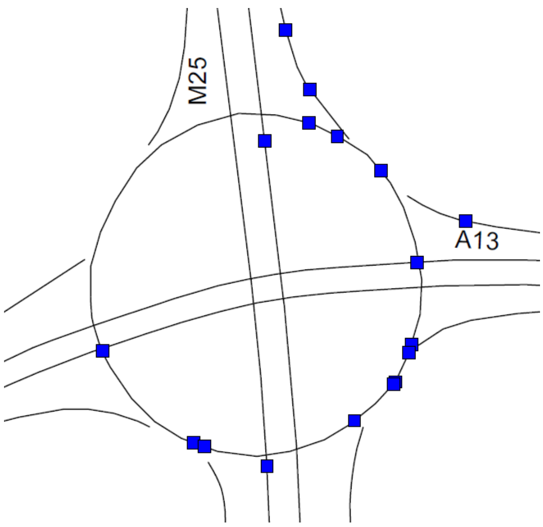
(b) AMM1



(c) AMM2



(d) AMM3



(e) AMF

Figure 4.12: Accident locations at the M25-A13 junction

4.3.2 Link-based dataset

The link-based dataset includes individual traffic and geometric characteristics and the total accidents per link. Traffic conditions are expressed by the annual average of speed and AADT, while road geometry was represented by categorical variables for curvature, gradient and number of lanes. A more detailed description of the variables can be found in Table 4.6. The number of accidents per road link was estimated using the output of AMF that indicates a HATRIS road link for each accident.

Considering the dynamic nature of the traffic variables (i.e. speed and volume) as well as the fact that a road link typically covers a considerable road length, it can be understood that both the traffic conditions and the geometric configuration of each link can only be partially represented by single measures per link. This can be proved from Figures 4.13 and 4.14. Figure 4.13 shows the frequency and the cumulative distribution of the ratio of the actual speed at the accident location to the annual average speed on the corresponding road link for all 2012 motorway accidents in England. Figure 4.14 is the same for traffic volume. It is obvious that the ratios are considerably different from one for a high proportion of accidents (ratio=1 means the equality of accident speed or volume with the respective annual average) confirming that representation of time-varying measures by annual averages is often rather inadequate.

After the exclusion of the links with missing traffic or geometry data the final link-based dataset included 2,356 observations (i.e. links) that represent overall 9,028 accidents. Accident counts were divided by severity into accidents with Fatal, Serious and Slight injuries and by collision type into single vehicle and multiple vehicle accidents. The descriptive statistics of the link-base dataset are shown at Table 4.7.

Table 4.6: Definition of variables of the link-based dataset

Variable	Link-based dataset
Speed	Annual average of measured speeds on each link (averaged over 35,040 records)
Volume	Annual average daily traffic per link (AADT)
Curvature	C1. Links with multiple and/or sharp curves (Curve) C2. Links that above 50% of their radius measurements are equal with 2000 m (Straight)
Gradient	G1. Links with median gradient above 0.5% (Uphill) G2. Links with median gradient below -0.5% (Downhill) G3. Links with median gradient between 0.5% (Level)
Lanes	L1. Links that above 50% of their sections include more than two lanes (Lanes above 2) L2. Links that above 50% of their sections include up to than two lanes (Lanes up to 2)
Length	Total link length

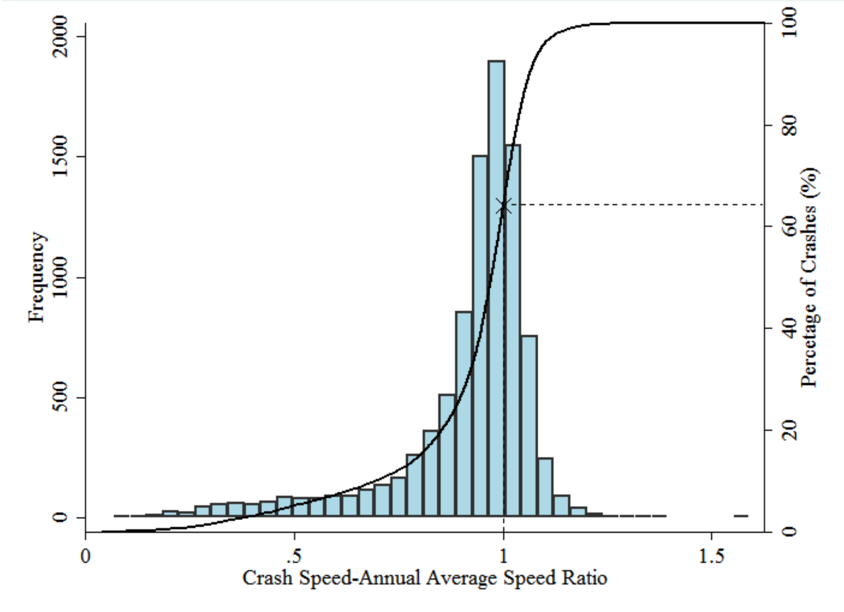


Figure 4.13: Frequency and cumulative distribution of the 15-minute speed at the time and the location of the accident by the annual average of the speed on this link.

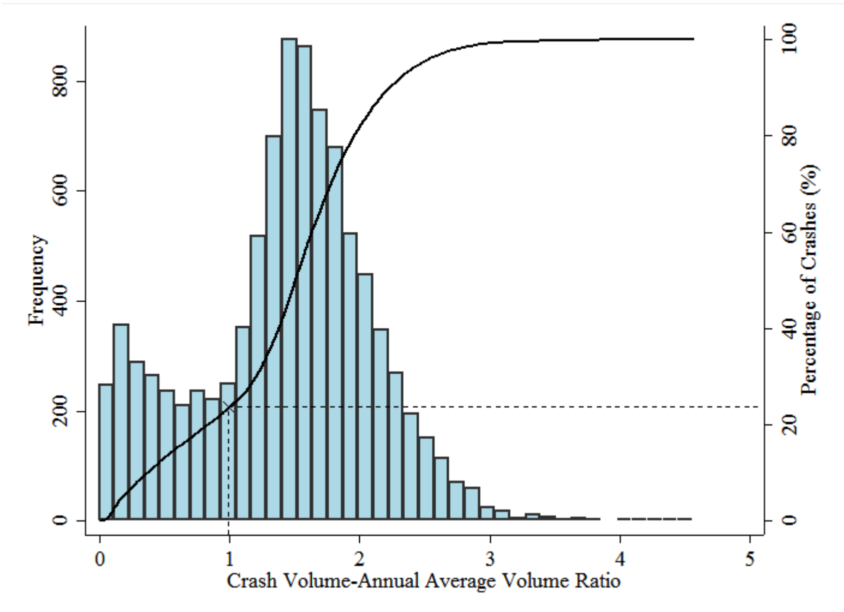


Figure 4.14: Frequency and cumulative distribution of the 15-minute volume at the time and the location of the accident by the annual average of the speed on this link.

Table 4.7: Descriptive statistics of the variables of the link-based dataset

Variable	**Mean**	**SD**	**Min.**	**Max.**
Dependent variables				
All accidents	3.83	4.34	0	36
Fatal accidents	0.08	0.29	0	2
Serious accidents	0.46	0.85	0	7
Slight accidents	3.29	3.88	0	36
Single vehicle accidents	0.94	1.46	0	13
Fatal single vehicle accidents	0.03	0.16	0	2
Serious single vehicle accidents	0.16	0.46	0	4
Slight single vehicle accidents	0.76	1.24	0	13
Multiple vehicle accidents	3.01	3.76	0	32
Fatal multiple vehicle accidents	0.06	0.25	0	2
Serious multiple vehicle accidents	0.31	0.69	0	6
Slight multiple vehicle accidents	2.64	3.39	0	31
Independent variables				
Speed (mph)	58.53	10.3	16.91	79.73
AADT (in thousands)	28.83	17.99	0.03	107.1
Curvature:				
Curve	0.46	0.5	0	1
Straight	0.54	0.5	0	1
Gradient:				
Uphill	0.11	0.31	0	1
Downhill	0.48	0.5	0	1
Level	0.41	0.49	0	1
Lanes:				
Lanes above 2	0.32	0.47	0	1
Lanes below 2	0.68	0.47	0	1
Length (miles)	3.85	3.18	0.02	27.95

4.3.3 Condition-based dataset

The condition-based dataset includes all the possible condition scenarios on the examined network and the number of accidents that occurred under these conditions. Following the traffic and geometry conditions identification (outlined in Sections 3.5.1 and 3.5.2), the initial 10,520 accidents of the database decreased to 9,310 due to missing or illogical values in one or more variables. Accidents left in the analysis were classified to a spreadsheet that included all the condition scenarios.

Tables 4.8 and 4.9 present the characteristics of the traffic and geometric conditions just before accidents respectively. In Table 4.8 it can be seen that the average speed of single vehicle accidents tends to be higher than that of multiple vehicle accidents. Also, the average speed is higher for more severe multiple vehicle accidents. The average volume per lane is significantly lower for single vehicle accidents compared with the multiple vehicle ones and is lowest for the highest severities. This is an initial indication that specific accident types are associated with different traffic conditions. Table 4.9 shows that single vehicle accidents occur more frequently on curved segments while multiple vehicle accidents on straight. The majority of the accidents regardless of the collision type occur on downgrades and on segments with up to two lanes.

Table 4.8: Descriptive statistics of the traffic conditions just before an accident split by accident type.

Accident Type	Average Speed (St.Deviation) (mph)	Average Volume per Lane (St.Deviation) (vehicles)
All	54.61 (14.34)	195 (127)
K	56.2 (11.14)	106 (101)
S	55.19 (13.88)	157 (118)
Sl	54.49 (14.47)	203 (127)
SV	59.63 (10.95)	138 (104)
SV_K	58.85 (9.86)	104 (100)
SV_S	59.1 (12.09)	132 (104)
SV_Sl	59.77 (10.73)	140 (104)
MV	54.01 (14.95)	233 (125)
MV_K	56.83 (11.39)	130 (117)
MV_S	55.5 (13.99)	203 (126)
MV_Sl	53.83 (15.07)	237 (124)

Table 4.9: Percentages of the geometrical features of the segments upstream accidents split by accident type.

Accident Type	Curve (%)	Straight (%)	Uphill (%)	Downhill (%)	Level (%)	Lanes above 2 (%)	Lanes below 2 (%)
All	46.96	53.04	31.92	64.60	3.48	43.64	56.36
K	49.74	50.26	26.94	69.43	3.63	33.16	66.84
S	49.42	50.58	29.54	67.14	3.31	35.72	64.28
Sl	46.55	53.45	32.38	64.13	3.5	45	55
SV	51.39	48.61	32.12	64.69	3.19	38.45	61.55
SV_K	54.84	45.16	30.65	64.52	4.84	33.87	66.13
SV_S	51.95	48.05	27.79	68.83	3.38	33.25	66.75
SV_Sl	51.15	48.85	33.11	63.8	3.09	39.74	60.26
MV	42.65	57.35	31.67	64.61	3.72	51.78	48.22
MV_K	43.84	56.16	31.51	63.01	5.48	45.21	54.79
MV_S	42.45	57.55	27.55	69.18	3.27	48.98	51.02
MV_Sl	42.65	57.35	32.06	64.21	3.73	52.13	47.87

The scenarios of the condition-based dataset were developed according to the process that was presented in Section 3.5.3. Traffic characteristics were grouped into categories of equal frequency. The speed groups were defined by dividing the cumulative speed distribution of the entire network into 50 ($K_{speed} = 50$) equal parts with a 2-percentile step ($n_{speed} = 2$). Following, the volume, for each speed group separately, was split into to the quartiles of its cumulative distribution ($K_{volume} = 4$ and $n_{volume} = 25$). Speed and volume per group were represented by their medians. Other measures were also tested such as the mean and the 85th percentile that did not exhibit any statistical difference in the modelling results. To keep the number of combinations relatively low, all the geometric variables were represented by categorical variables. As it is shown in Table 4.10, curvature and lanes have two categories each ($L_{curvature} = 2$ and $L_{lanes} = 2$) and gradient has three ($L_{gradient} = 3$). The number of scenarios (S) of this dataset can be estimated using equation 3.7:

$$S = K_{speed} \cdot K_{volume} \cdot L_{curvature} \cdot L_{gradient} \cdot L_{lanes} = 50 \cdot 4 \cdot 2 \cdot 3 \cdot 2 = 2400 \quad (4.1)$$

Each of the 2,400 scenarios represented a unique combination of traffic and geometric attributes (e.g. *Speed is between the 40th and the 42nd percentile with the median value of 58 mph, the volume is between the 50th and the 75th percentile for these speed conditions with median 112veh/lane, on a straight and downhill section with up to two lanes*). The distinct values of each categorical or continuous variable had equal frequency with the other values of this variable (e.g. 800 scenarios were on uphill segments, 800 scenarios on downhill and 800 scenarios on the level). Each accident was classified to one of these scenarios with respect to its traffic and geometric conditions and the severity of its consequences. The final output of this process was a dataset with 2,400 observations that represent the all accident counts by severity and by the number of vehicles involved for the pre-accident-condition scenarios. Table 4.11 presents the summary statistics of the explanatory variables of both the datasets.

Table 4.10: Definition of variables of the condition-based dataset

Variable	Condition-based dataset
Speed	S1. Speed up to 2nd percentile S2. Speed between the 2nd and the 4th percentile S3. Speed between the 4th and the 6th percentile ... S50. Speed between the 98th and the 100th percentile
Volume	Separately for each of the 50 speed scenarios: V1. Volume up to the 25th percentile V2. Volume between the 25th and the 50th percentile V3. Volume between the 50th and the 75th percentile V4. Volume over the 75th percentile
Curvature	C1. Segments that above 50% of their radius measurements are lower than 2000 m (Curve) C2. Segments that above 50% of their radius measurements are equal with 2000 m (Straight)
Gradient	G1. Segments that have more gradient measurements above 0.5% than below 0.5% (Uphill) G2. Segments that have more gradient measurements below -0.5% than above -0.5% (Downhill) G3. Segments that have more gradient measurements between 0.5% than below 0.5% and above -0.5% (Level)
Lanes	L1. Sections with more than two lanes (Lanes above 2) L2. Sections with up to two lanes (Lanes up to 2)
Vehicle-Hours	Average vehicle hours travelled per condition scenario

Table 4.11: Descriptive statistics of the variables of the condition-based dataset.

Variable	Mean	SD	Min	Max
Dependent variables				
All accidents	3.88	6.01	0	84
Fatal accidents	0.08	0.3	0	3
Serious accidents	0.47	0.93	0	8
Slight accidents	3.33	5.38	0	78
Single vehicle accidents	0.93	1.4	0	11
Fatal single vehicle accidents	0.03	0.17	0	2
Serious single vehicle accidents	0.16	0.44	0	3
Slight single vehicle accidents	0.74	1.18	0	8
Multiple vehicle accidents	2.95	5.31	0	83
Fatal multiple vehicle accidents	0.05	0.25	0	2
Serious multiple vehicle accidents	0.31	0.73	0	6
Slight multiple vehicle accidents	2.59	4.84	0	78
Independent variables				
Speed (mph)	58.53	11.78	21.28	80.49
Volume (vehicles/lane)	114.36	95.53	6.07	304.23
Curvature:				
Curve	0.5	0.5	0	1
Straight	0.5	0.5	0	1
Gradient:				
Uphill	0.33	0.47	0	1
Downhill	0.33	0.47	0	1
Level	0.33	0.47	0	1
Lanes:				
Lanes above 2	0.5	0.5	0	1
Lanes below 2	0.5	0.5	0	1
VehicleHours	5.31	5.5	0.35	43.61

4.4 Summary

This chapter presented the datasets that will be employed to conduct the analysis. The study area comprises the entire Strategic Road Network of England (SRN) that is represented by two digital maps with different detail level (i.e. HATIRS and HAPMS). The data that were employed represented the accidents (STATS19), traffic and the geometry conditions on the SRN during 2012.

The accident mapping algorithm that was outlined to the previous chapter was found to provide 98.9% ($\pm$ 1.1 %) accurate locations and that it clearly outperforms other existing algorithms. The link-based dataset that was developed consists of 2,356 observations (i.e. links), the number of accidents and variables that represent the average conditions on each link. The condition-based dataset includes 2,400 equally likely scenarios and the accident counts that occurred under each of the condition combinations in the study area and period.

Chapter 5

Modelling Results and Discussion

5.1 Introduction

Employing the methods and the datasets that have been discussed in Chapters 3 and 4 respectively, a series of accident count models have been developed. Through these models it is possible to quantify the relationships of speed, volume and road geometry with accidents and to examine the impact of high quality geo-coded accident data on the modelling results. This chapter presents and discusses the results of the developed models and their methodological implications.

The results of the link-based and the condition-based models will be presented in the second and the third section of this chapter respectively. The models that are included in these sections are:

- Link-based models

    1. All accidents (SRN)
    2. Accidents by severity type (SRN)
    3. Accidents by collision type (SRN)
    4. Single-vehicle accidents by severity type (SRN)
    5. Multiple-vehicle accidents by severity type (SRN)

- Condition-based models

    1. All accidents (SRN)

2. Accidents by severity type (SRN)
3. Accidents by collision type (SRN)
4. Single-vehicle accidents by severity type (SRN & motorways)
5. Multiple-vehicle accidents by severity type (SRN & motorways)

Each model is represented with a table that shows the coefficient estimates of the examined independent variables. Multivariate models also include covariance-correlation matrices which present the level of correlations between the examined accident types. The use of several different specifications for the traffic variables and the inclusion of interaction terms does not permit straightforward comparisons between the models. To facilitate understanding and comparing the models, the relationships of the traffic characteristics (i.e. speed and volume) with accidents are represented graphically with three-dimensional contour plots. These graphs show the traffic conditions that are more likely to be related with accidents according to the outcomes of each model.

The results clearly show that significant differences exist between models that originate from different accident data aggregation approaches. In the fourth section of this chapter these differences will be outlined and discussed in order to identify the most advantageous aggregation approach for analysing accident data. The fifth section of this chapter will discuss the effect of accident location accuracy on the modelling outcomes. This will be done by comparing the results of identical models that employ accident data with different location accuracies.

5.2 Link-Based Models

The univariate and multivariate models 3.11 and 3.19 were initially fitted to the link-based datasets using WinBUGS 1.4.3 (Spiegelhalter et al., 2003), an open-source software that is suitable for full Bayes model estimation using the Markov Chains Monte Carlo (MCMC) method. The estimations were derived from 50,000 iterations of two chains with a burn-in sample of 20,000 iterations. Convergence was visually detected by observing the Markov chain history graphs of the models' coefficients which are provided by the WinBUGS software. The first model that it is presented here is the link-based univariate

Poisson lognormal spatial model that examines the relationship of all accidents with the traffic and geometric independent variables. Next are the multivariate Poisson lognormal spatial models that examine accidents disaggregated by injury severity (i.e. fatal, serious or slight) and by collision type (i.e. single-vehicle, multiple-vehicle). This is followed by the presentation of the models for single-vehicle and multiple-vehicle accidents disaggregated by severity.

The output of all multivariate models includes variance-covariance matrices of the heterogeneity and the spatial effects (from equations 3.13 - 3.18 and 3.19 - 3.22). Correlations between different accident types are estimated using the following equation (Field, 2009):

$$Correl(x, y) = \frac{Cov(x, y)}{s_x \cdot s_y} \tag{5.1}$$

Where: $Cov(x, y)$: the covariance between variables x and y and s_x and s_y: the standard deviations of variables x and y respectively.

Correlations between the frequencies of different accident types may be caused by the omission of variables which are potentially significant for accident occurrence (e.g. Ma et al., 2008; Aguero-Valverde, 2013; Barua et al., 2014). The correlation coefficients can be from -1 to 1, with positive values obviously representing positive correlations and vice versa. Although negative correlations are possible to be estimated, in multivariate accident models they are rather unlikely because accidents tend to be mainly related with similar unobserved variables (Ma et al., 2008).

5.2.1 All accidents

The variable estimates of the univariate Poisson lognormal link-based model (*Link-based All (12)*) are presented in Table 5.1. As it can be seen by the model's name the best fitting functional relation to variables was shown to be specification 12 (see Table 3.6) in which speed was included to the model without transformation and AADT was squared. In addition, a speed-AADT interaction term was added to the model. The outcomes for the traffic variables of the model are attempted to be presented in a more comprehensive way though a three-dimensional contour plot that presents the simultaneous relationship of the accident rate (i.e. accidents/exposure) with average link speed and AADT (Figure 5.1). To develop this graph and all the similar graphs throughout this chapter, the equation that was derived from the model was used. To facilitate the estimation this equation refers to the reference cases of the independent variables (i.e. when Curve=0, Uphill=Downhill=0 and Lanes above 2=0). The purpose of this equation is not to predict specific accident rates but to visualise the accident trends as a function of the traffic variables (i.e. speed and AADT) so the selection of the reference variables is the most convenient approach. Therefore the equation that was used to develop Figure 5.1 was:

$$\frac{All\ accidents}{miles} = exp(-0.0569 \cdot Speed - 0.0002 \cdot AADT^2 + 0.0006 \cdot Speed \cdot AADT - 2.2393) \quad (5.2)$$

Table 5.1: Parameter estimates for the link-based univariate Poisson-lognormal model for all accidents (*Link-based All (12)*)

All accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	-0.0569 **	0.00294	0.00010	-0.0627	-0.0617	-0.0521	-0.0511
AADT squared	-0.0002 **	0.00004	0.00000	-0.0003	-0.0003	-0.0001	-0.0001
Speed * AADT	0.0006 **	0.00006	0.00000	0.0005	0.0005	0.0007	0.0008
Curve	-0.0270	0.03590	0.00080	-0.0975	-0.0862	0.0320	0.0439
Uphill	0.0595	0.06103	0.00122	-0.0609	-0.0409	0.1590	0.1797
Downhill	0.0572 *	0.03429	0.00068	-0.0102	0.0006	0.1134	0.1238
Lanes	0.0140	0.05134	0.00159	-0.0871	-0.0710	0.0985	0.1149
Intercept	2.2393	0.14097	0.00464	1.9636	2.0053	2.4723	2.5154
ln(Length)	1	-	-	-	-	-	-
$\bar{D}$	8630	**statistically significant at the 95% credible interval					
p_D	540	*statistically significant at the 90% credible interval					
DIC	9171						

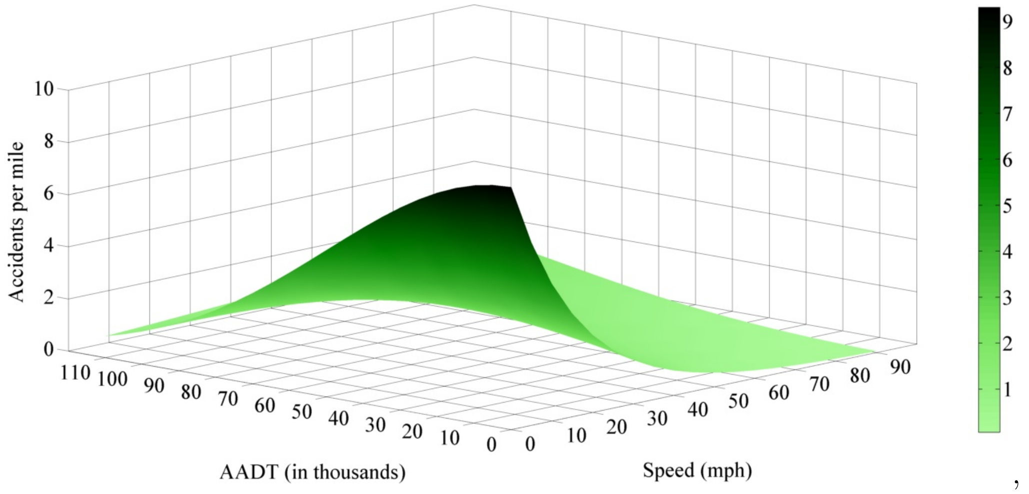


Figure 5.1: Contour plot of the predicted accidents per mile as a function of speed and AADT (model: *Link-based All (12)*).

Overall, the results were hard to interpret and counterintuitive to a certain extent. Speed was found to be inversely proportional with all accidents for links with AADT lower than 95,000 (estimated from the slope of speed in equation 5.2). The finding that high speeds increase accident rates for links with very high average volume refers only to a small proportion of links with AADT above 95,000 that is only 0.34%. For the rest of the links of the network the accident-speed relationship is negative. Although some other studies have suggested that higher speeds are associated with less accidents (e.g. Lave, 1985; Baruya, 1998*a*), none of the researchers has given a clear explanation of why higher average speeds could be overall safer. Some of the main arguments to support this idea are the increased design standards of high speed motorways and the longer available distances between vehicles at high speed conditions. However, the vast majority of studies that examine the number of accidents before and after speed limit changes (consequently changes in average speed) suggest that higher speeds are related to more accidents (e.g. Elvik et al., 2004).

Higher AADT was related with more accidents at least for the majority of the links of the road network. The relationship of AADT with accidents can be described an inverse-U shaped curve with a peak that is a function of speed [1]. Considering that only 5.75% of the SRN links have average speed below 40 mph, in most of the cases accidents tend to increase as AADT increases (e.g. for the average speed that is 58.53 mph AADT in-

[1]Solving the first derivative of AADT (i.e. $\frac{\partial Allaccidents/miles}{\partial AADT} = 0$) of equation 5.2 it can be found that for speed=40, 55 and 70 mph the curve maximises at 60, 82.5 and 105 thousands vehicles

creases proportionally with accidents until it reaches 87.80 thousand vehicles). This result is in-line with most of the existing studies that suggest that higher AADT is associated with higher accident frequency (e.g. Abdel-Aty and Radwan, 2000; Anastasopoulos and Mannering, 2009).

All the links' geometrical features were found to be statistically insignificant except from negative grades which seem to be related with higher accident frequency. The use of dummy variables for the geometrical variables could have possibly affected the estimated coefficients of this and all the following models. However, the signs of the coefficients of the most important variables (i.e. speed) did not change even when the geometrical characteristics were represented by continuous variables in models that are not presented due to brevity. These results may be due to the examination of all accidents into one category that did not possibly permit the representation of the variations between different accident severities and types. However,they might also provide a first indication that the highly-aggregared time varying variables are likely to lead to inaccurate estimations.

5.2.2 Accidents by severity type

Accidents based on the severity of their outcomes, might have different inherent characteristics (e.g. Park and Lord, 2007; Ma et al., 2008). This was checked by using a multivariate regression model that examined accidents split into two severity categories: accidents with killed or seriously injured casualties (KS) and accidents with slight injuries (Sl) (*Link-based KS-Sl AMF (3)*)(A multivariate model that examines fatal (K), serious (S) and slight (Sl) accidents separately is presented in Appendix B). Table 5.2 shows the estimates of the model and Table 5.3 the covariance-correlation matrices. In the best fitting specification for this model speed is in a linear form and AADT logarithmically transformed. The correlation between KS and Sl accidents (marked with bold font in Table 5.3) was found to be high (0.812) and significant. This high correlation highlighted that there are indeed correlations between different accident types proving the suitability of multivariate modelling for accident analyses, a fact that has been already reported in a series of studies (e.g. Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009). The correlation for spatial effects between accident severities that was statistically significant but not very high (0.491). This is partially in line with the findings by Barua et al. (2014) that reported that both the correlation coefficients were high and significant. On the other hand, Aguero-Valverde (2013) found similar correlation coefficients for the random effects but the correlation due to spatial effects could be regarded as equal to zero; a finding that was explained by the presence of two random effects in the model.

The accident rate by severity for the reference cases of the categorical independent variables (i.e. Curve=0, Uphill=Downhill=0, Lanes above 2=0) for the *Link-based KS-Sl AMF (3)* is:

$$\frac{KS\,accidents}{miles} = exp(-0.0372 \cdot Speed + 0.1310 \cdot ln(AADT) - 0.5301) \tag{5.3}$$

$$\frac{Sl\,accidents}{miles} = exp(-0.0467 \cdot Speed + 0.6848 \cdot ln(AADT) + 0.1325) \tag{5.4}$$

Table 5.2: Parameter estimates for the link-based multivariate Poisson-lognormal model for fatal and serious (KS) and slight (Sl) accidents (*Link-based KS-Sl AMF (3)*)

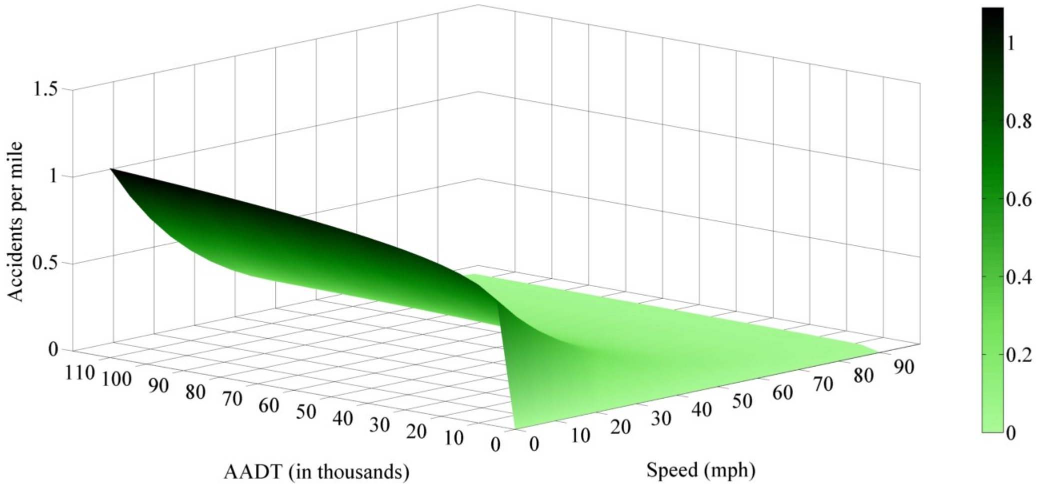
KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	-0.0372 **	0.0042	0.0002	-0.0450	-0.0438	-0.0301	-0.0287
ln(AADT)	0.1310 **	0.0636	0.0024	0.0046	0.0243	0.2355	0.2547
Curve	-0.0740	0.0720	0.0014	-0.2164	-0.1933	0.0439	0.0655
Uphill	-0.0763	0.1278	0.0016	-0.3285	-0.2889	0.1316	0.1700
Downhill	-0.0094	0.0680	0.0011	-0.1420	-0.1205	0.1028	0.1254
Lanes	0.2005 **	0.0913	0.0025	0.0226	0.0509	0.3502	0.3799
Intercept	-0.5301 **	0.2548	0.0105	-1.0453	-0.9570	-0.1110	-0.0292
ln(Length)	1	-	-	-	-	-	-
Sl accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
Speed	-0.0467 **	0.0024	0.0001	-0.0516	-0.0507	-0.0428	-0.0419
ln(AADT)	0.6848 **	0.0410	0.0019	0.6058	0.6201	0.7553	0.7699
Curve	-0.0271	0.0385	0.0008	-0.1022	-0.0903	0.0362	0.0481
Uphill	0.0728	0.0644	0.0010	-0.0537	-0.0330	0.1785	0.1989
Downhill	0.0814 **	0.0365	0.0006	0.0100	0.0209	0.1412	0.1528
Lanes	0.1396 **	0.0518	0.0017	0.0371	0.0538	0.2247	0.2405
Intercept	0.1325	0.1679	0.0081	-0.2034	-0.1478	0.4149	0.4709
ln(Length)	1	-	-	-	-	-	-
$\bar{D}$	12026	**statistically significant at the 95% credible interval					
p_D	669						
DIC	12695						

Table 5.3: Combined Covariance-Correlation matrix of the (A) random effect and (B) the spatial effect of the *Link-based KS-Sl AMF (3)* model.

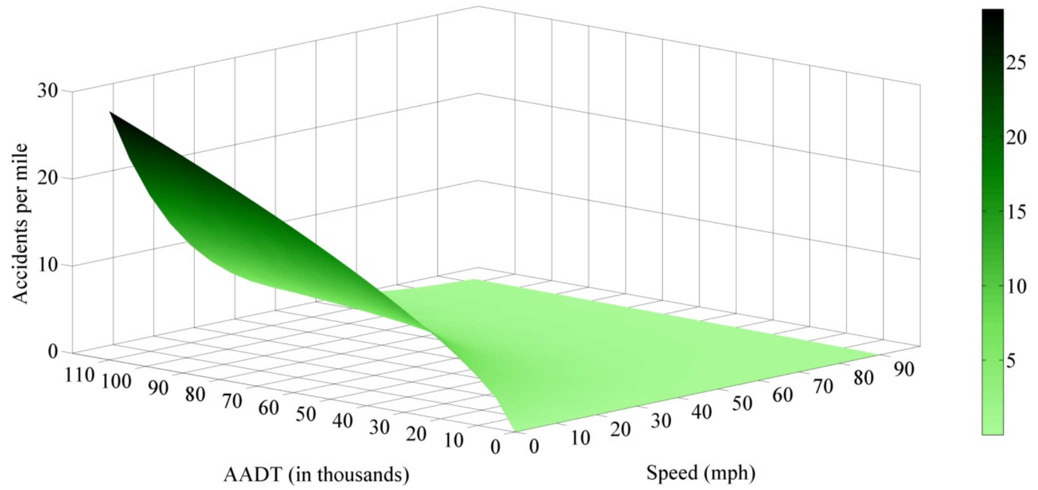
A	KS accidents	Sl accidents
KS accidents	0.276**	0.200**
Sl accidents	**0.812****	0.220**
B	KS accidents	Sl accidents
KS accidents	0.007**	0.004**
Sl accidents	**0.491****	0.011**

Correlation is marked with bold font.

**Statistically significant at the 95% credible interval



(a)



(b)

Figure 5.2: Contour plot of the predicted (a)fatal and serious (KS) and (b) slight (Sl) accidents per mile as a function of speed and AADT (model: *Link-based KS-Sl AMF (3)*).

Although accidents of different severity levels were modelled separately the results for speed are not different from those of the link-based univariate model (*Link-based All (12)* model at Section 5.2.1). The relationship of the average link speed with accidents was found to be negative even for KS accidents which according to the literature are more related with high speeds and speeding (e.g. Elvik et al., 2004; Pei et al., 2012). As it was mentioned above this result is counterintuitive and can be attributed to aggregation bias.

Higher AADT was again found to be related with more accidents of all severities, an outcome that is clearly more explainable and is confirmed by the literature (e.g Abdel-Aty and Radwan, 2000). Higher number of lanes was related with more KS and Sl accidents, a finding that also is in line with the findings of existing research (e.g. Milton and Mannering, 1998; Chang, 2005). Sl accidents were also associated with road links that tend to be downhill. Specifically, using the estimated coefficient (0.0814) it can be estimated that downhill links are expected to to have 8.48% (i.e. $(e^{0.0814} - e^{0}) \cdot 100\% = 8.48\%$) more accidents per mile, compared to uphill and level links, which is a quite interpretable outcome considering that negative grades are associated with longer braking distances and higher speeds (e.g. Milton and Mannering, 1998).

5.2.3 Accidents by collision type

Table 5.4 presents the modelling outcomes of the multivariate regression model for single-vehicle (SV) and multiple vehicle (MV) accidents (*Link-based SV-MV (3)*. Similarly to the *Link-based KS-Sl AMF (3)* model above, in the best fitting specification speed was in its linear form and AADT was logarithmically transformed. In the variance-covariance table for this model (Table 5.5) it can be seen that the correlation of the random effect is low (0.332) but the correlation of the spatial effect is quite high (0.720) meaning that particular locations are more related with accidents. The low random effect correlation coefficient can be explained by the fact that the generation processes of SV and MV accidents are quite different (e.g. Kim et al., 2006). However this outcome could also be related with the existence of two random effects in the model as it has been proposed by Aguero-Valverde (2013).

SV and MV accident rate for the reference cases of the categorical independent variables (i.e. Curve=0, Uphill=Downhill=0, Lanes above 2=0) for the *Link-based SV-MV (3)* is equal with:

$$\frac{KS\ accidents}{miles} = exp(-0.0134 \cdot Speed + 0.2218 \cdot ln(AADT) - 1.6989) \tag{5.5}$$

$$\frac{Sl\ accidents}{miles} = exp(-0.0575 \cdot Speed + 0.7093 \cdot ln(AADT) + 0.5705) \tag{5.6}$$

Table 5.4: Parameter estimates for the link-based multivariate Poisson-lognormal model for single-vehicle (SV) and multiple-vehicle (MV) accidents (*Link-based SV-MV (3)*)

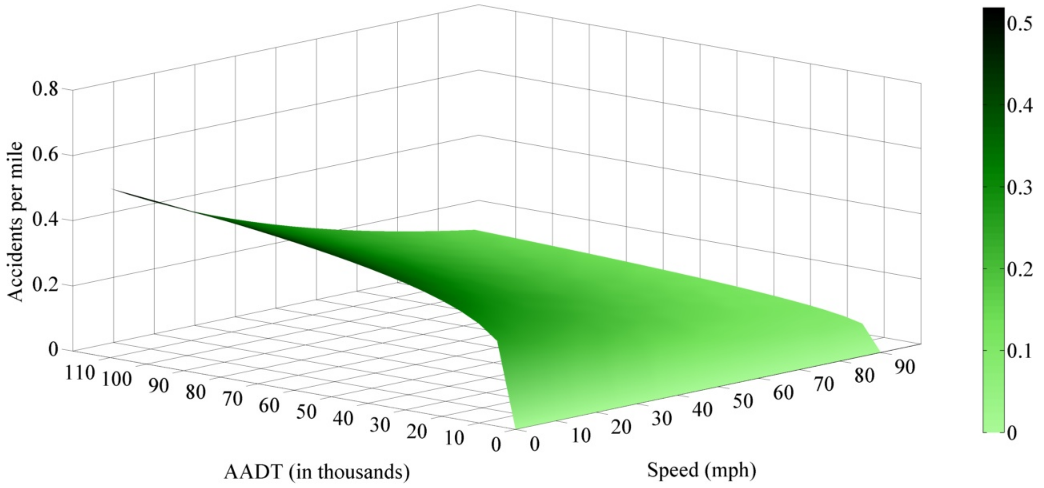
SV accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	-0.0134 **	0.0038	0.0002	-0.0208	-0.0196	-0.0074	-0.0062
ln(AADT)	0.2218 **	0.0581	0.0024	0.1064	0.1253	0.3165	0.3351
Curve	0.0262	0.0582	0.0013	-0.0878	-0.0690	0.1232	0.1405
Uphill	-0.0460	0.1002	0.0015	-0.2452	-0.2122	0.1177	0.1483
Downhill	0.0910 *	0.0540	0.0008	-0.0152	0.0019	0.1795	0.1975
Lanes	0.0579	0.0747	0.0023	-0.0884	-0.0644	0.1805	0.2032
Intercept	-1.6989 **	0.2471	0.0112	-2.1620	-2.0918	-1.2863	-1.1938
ln(Length)	1	-	-	-	-	-	-
MV accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
Speed	-0.0575 **	0.0023	0.0001	-0.0620	-0.0613	-0.0536	-0.0529
ln(AADT)	0.7093 **	0.0430	0.0020	0.6227	0.6386	0.7815	0.7938
Curve	-0.0532	0.0401	0.0009	-0.1317	-0.1188	0.0121	0.0249
Uphill	0.0056	0.0668	0.0010	-0.1241	-0.1039	0.1160	0.1376
Downhill	0.0790 **	0.0380	0.0006	0.0043	0.0163	0.1416	0.1531
Lanes	0.2102 **	0.0521	0.0017	0.1086	0.1237	0.2953	0.3124
Intercept	0.5705 **	0.1570	0.0073	0.2718	0.3207	0.8287	0.8763
ln(Length)	1	-	-	-	-	-	-
$\bar{D}$	12678	**statistically significant at the 95% credible interval					
p_D	838	*statistically significant at the 90% credible interval					
DIC	13516						

Table 5.5: Combined Covariance-Correlation matrix of the (A)random effect and (B) the spatial effect of the *Link-based SV-MV (3)* model.

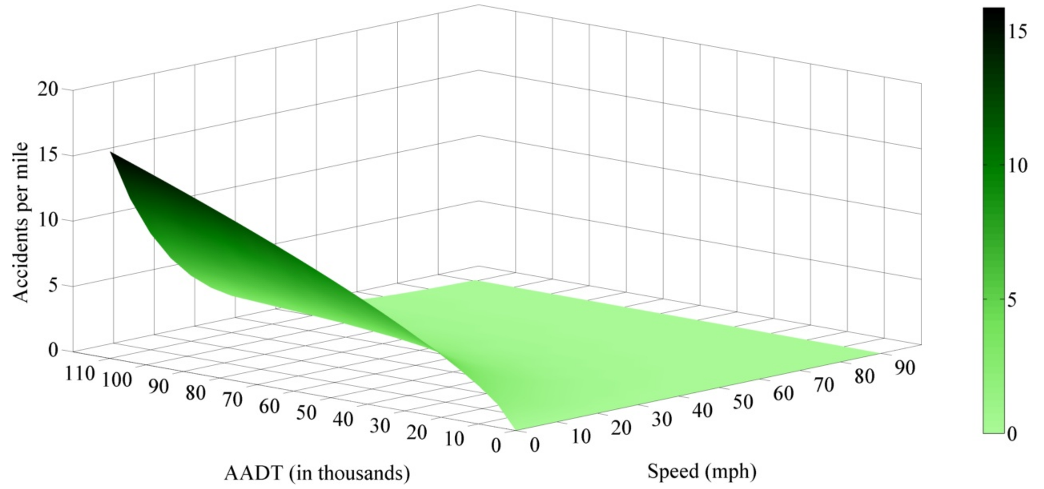
A	SV accidents	MV accidents
SV accidents	0.196**	0.071**
MV accidents	**0.332****	0.232**
B	SV accidents	MV accidents
SV accidents	0.015**	0.008**
MV accidents	**0.720****	0.009**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval



(a)



(b)

Figure 5.3: Contour plot of the predicted (a)fatal and serious (KS) and (b) slight (Sl) accidents per vehicle hours travelled per mile as a function of speed and AADT (model: *Link-based SV-MV (3)*).

Although SV and MV accidents are considered to have different generation processes according to the results of the *Link-based SV-MV (3)* model the traffic conditions that are associated with these collision types were identical. Exactly like in the previous two link-based models speed was negatively associated with accidents and AADT was positively related with them. This result, especially for SV accidents, is counterintuitive and in opposition with most of the literature that suggests that single vehicle accidents occur mainly at low density conditions so when speed is high and the traffic volume low (e.g. Xie et al., 2012; Kim et al., 2013). The outcome for MV accidents interpretable to a certain extent, as multiple-vehicle collisions are more likely to occur when the traffic is heavier (e.g. Ivan et al., 1999; Ivan, 2004). The only geometrical feature of the links that was found to be statistically significant for both collision types was negative grades. Moreover, links with more than two lanes were only found to be related with more MV accidents.

5.2.4 Single-vehicle accidents by severity

The modelling results for SV accidents by severity (i.e. *Link-based SV KS-Sl (9)*) can be found in Tables 5.6 and 5.7 and Figure 5.4. Speed and AADT were logarithmically transformed in the best fitting specification of the model. Table 5.7 presents the covariance-correlation matrices of this model. Both the correlation coefficients of the random and the spatial effects were found to be high (i.e. 0.900 and 0.693 respectively) showing the similarity of SV accidents independent of their severity.

The accident rate for SV_KS and SV_Sl accidents for the reference cases of the categorical independent variables (i.e. Curve=0, Uphill=Downhill=0, Lanes above 2=0) for the *Link-based SV KS-Sl (9)* equals:

$$\frac{SV\_KS\ accidents}{miles} = exp(-1.1271 \cdot ln(Speed) + 0.5015) \tag{5.7}$$

$$\frac{SV\_Sl\ accidents}{miles} = exp(-0.6032 \cdot ln(Speed) + 0.2732 \cdot ln(AADT) - 0.4820) \tag{5.8}$$

Table 5.6: Parameter estimates for the link-based multivariate Poisson-lognormal model for Fatal and serious (SV_KS) and slight single-vehicle (SV_Sl) accidents (*Link-based SV_KS-Sl (9)*)

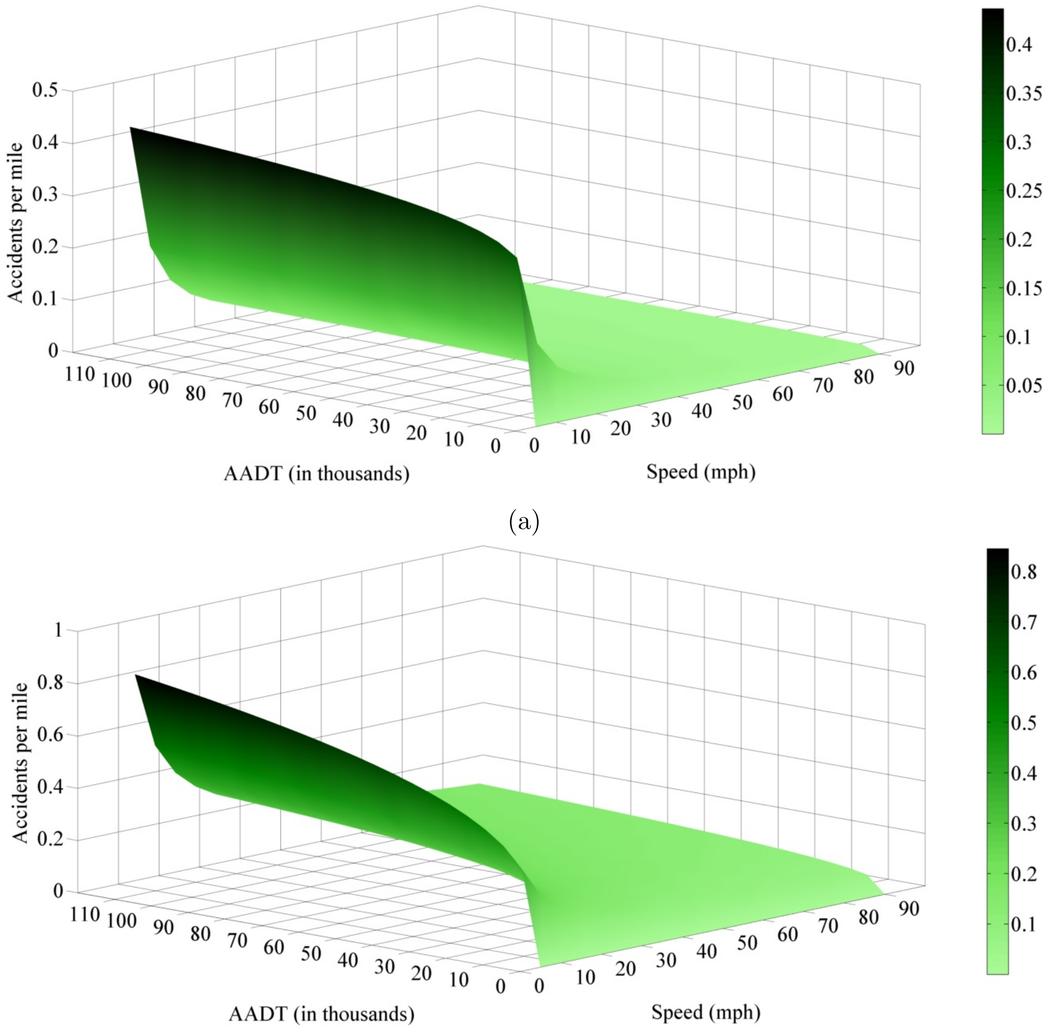
SV_KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
ln(Speed)	-1.1271 **	0.2788	0.0230	-1.6290	-1.5793	-0.6568	-0.5951
ln(AADT)	0.1028	0.1039	0.0073	-0.0997	-0.0788	0.2619	0.2836
Curve	0.0039	0.1134	0.0042	-0.2191	-0.1827	0.1886	0.2254
Uphill	0.1359	0.1941	0.0049	-0.2592	-0.1829	0.4500	0.5049
Downhill	-0.0202	0.1078	0.0031	-0.2317	-0.1983	0.1546	0.1868
Lanes	0.1034	0.1482	0.0079	-0.1732	-0.1296	0.3533	0.4021
Intercept	0.5015	1.0556	0.0870	-1.4893	-1.2710	2.2201	2.5223
ln(Length)	1	-	-	-	-	-	-
SV_Sl accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
ln(Speed)	-0.6032 **	0.1162	0.0095	-0.8053	-0.7752	-0.3581	-0.3289
ln(AADT)	0.2732 **	0.0586	0.0043	0.1619	0.1796	0.3677	0.3848
Curve	0.0286	0.0629	0.0022	-0.0956	-0.0749	0.1310	0.1529
Uphill	-0.1078	0.1118	0.0030	-0.3289	-0.2957	0.0742	0.1094
Downhill	0.1045 *	0.0604	0.0021	-0.0137	0.0048	0.2018	0.2208
Lanes	0.0344	0.0790	0.0041	-0.1195	-0.0954	0.1636	0.1890
Intercept	-0.4820	0.4538	0.0370	-1.6276	-1.4997	0.1595	0.2303
ln(Length)	1	-	-	-	-	-	-
$\bar{D}$	6417	**statistically significant at the 95% credible interval					
p_D	291	*statistically significant at the 90% credible interval					
DIC	6708						

Table 5.7: Combined Covariance-Correlation matrix of the (A)random effect and (B) the spatial effect of the *Link-based SV_KS-Sl (9)* model.

A	SV accidents	MV accidents
SV accidents	0.273**	0.213**
MV accidents	**0.900****	0.204**
B	SV accidents	MV accidents
SV accidents	0.012**	0.01**
MV accidents	**0.693****	0.017**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval



(a)

(b)

Figure 5.4: Contour plot of the predicted single-vehicle (a)fatal and serious (SV_KS) and (b) slight (SV_Sl) accidents per per mile as a function of speed and AADT (model: *Link-based SV_KS-Sl ()*).

The estimated coefficients were not particularly different in terms of signs from those of the *Link-based SV-MV (3)* model (see Section 5.2.3). Irrespective of their severity, SV accidents were found to decrease as speed increases. As it has been outlined in Section 5.2.3 this outcome is not expected and it cannot be interpreted. In fact, this finding confirms that aggregation bias that is associated with link-based approaches may indeed affect significantly the estimated coefficients. The rest of the examined variables (including AADT) for SV_KS accidents were not statistically significant. This might be the effect of the very low number of observations different than zero in this category (see Table 4.7). On the other hand, the estimated coefficients for AADT and downhill links were found to be significant and positively related with high SV_Sl accident frequency.

5.2.5 Multiple-vehicle accidents by severity

Tables 5.8 and 5.9 and Figure 5.5 present the results of the *Link-based MV KS-Sl (3)* model. In the best fitting model speed was in linear form and AADT was logarithmically transformed. The random effect correlation was high (0.837) and the spatial effect correlation relatively lower (0.509). The accident rate for the reference cases of the categorical independent variables (i.e. Curve=0, Uphill=Downhill=0, Lanes above 2=0) for the *Link-based SV KS-Sl (9)* that were estimated for fatal and serious and slight accidents are:

$$\frac{MV\_KS\ accidents}{miles} = exp(-0.0338 \cdot Speed + 0.5391 \cdot ln(AADT) - 3.1386) \qquad (5.9)$$

$$\frac{MV\_KS\ accidents}{miles} = exp(-0.0518 \cdot Speed + 1.0500 \cdot ln(AADT) - 1.2641) \qquad (5.10)$$

Table 5.8: Parameter estimates for the link-based multivariate Poisson-lognormal model for fatal or serious (MV_KS) and slight (MV_Sl) multiple-vehicle accidents (*Link-based MV_KS-Sl (3)*)

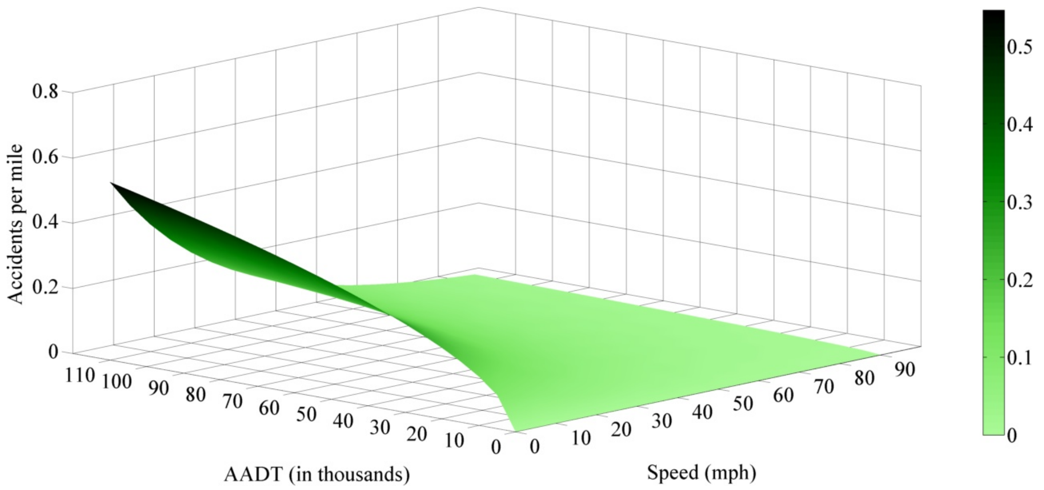
MV_KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	-0.0338 **	0.0064	0.0003	-0.0457	-0.0439	-0.0229	-0.0207
ln(AADT)	0.5391 **	0.0967	0.0037	0.3504	0.3805	0.6978	0.7255
Curve	-0.1542	0.1020	0.0020	-0.3554	-0.3227	0.0121	0.0440
Uphill	-0.6738 **	0.2254	0.0026	-1.1316	-1.0547	-0.3117	-0.2472
Downhill	0.0225	0.0912	0.0011	-0.1543	-0.1268	0.1739	0.2028
Lanes	0.1298	0.1250	0.0035	-0.1131	-0.0752	0.3356	0.3753
Intercept	-3.1386 **	0.4095	0.0172	-3.9532	-3.8235	-2.4802	-2.3413
ln(Length)	1	-	-	-	-	-	-
MV_Sl accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
Speed	-0.0518 **	0.0029	0.0001	-0.0573	-0.0565	-0.0471	-0.0459
ln(AADT)	1.0500 **	0.0472	0.0022	0.9607	0.9745	1.1265	1.1427
Curve	-0.0471	0.0434	0.0009	-0.1317	-0.1186	0.0242	0.0383
Uphill	0.0621	0.0735	0.0011	-0.0812	-0.0596	0.1827	0.2061
Downhill	0.0833 **	0.0416	0.0007	0.0019	0.0150	0.1519	0.1651
Lanes	0.1156 **	0.0562	0.0018	0.0049	0.0218	0.2070	0.2233
Intercept	-1.2641 **	0.1932	0.0092	-1.6456	-1.5850	-0.9371	-0.8794
ln(Length)	1	-	-	-	-	-	-
$\bar{D}$	9307.51	**statistically significant at the 95% credible interval					
p_D	551.633						
DIC	9859.14						

Table 5.9: Combined Covariance-Correlation matrix of the (A)random effect and (B) the spatial effect of the *Link-based MV_KS-Sl (3)* model.

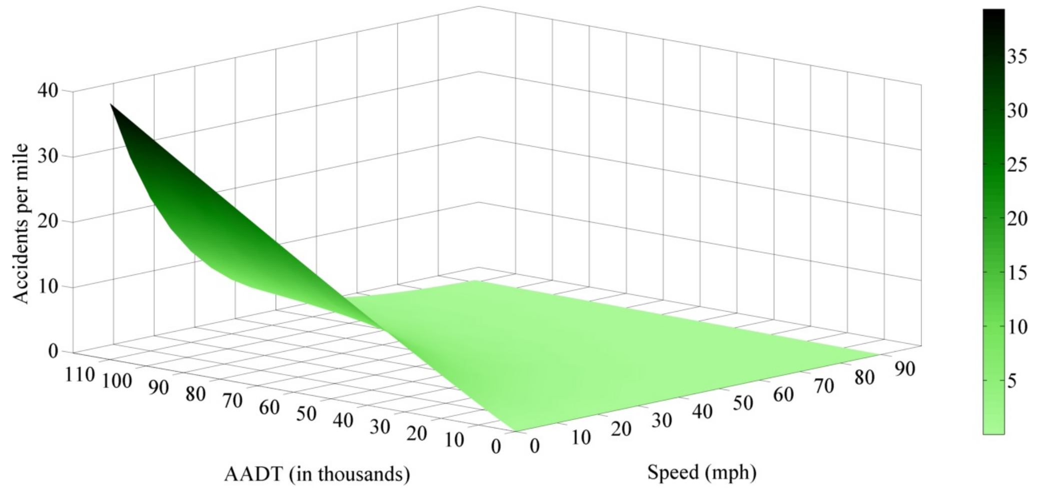
A	MV_KS accidents	MV_Sl accidents
MV_KS accidents	0.154**	0.157**
MV_Sl accidents	**0.837****	0.227**
B	**MV_KS accidents**	**MV_Sl accidents**
MV_KS accidents	0.007**	0.004**
MV_Sl accidents	**0.509****	0.008**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval



(a)



(b)

Figure 5.5: Contour plot of the predicted multiple-vehicle (a)fatal and serious (MV_KS) and (b) slight (MV_Sl) accidents per per mile as a function of speed and AADT (model: *Link-based MV_KS-Sl (3)*).

The results of this model are also very similar to those of the *Link-based SV-MV (3)* model (see Section 5.2.3). Speed was found to have a negative relationship with MV accidents of all severities, with the effect of lower speeds being higher for MV_Sl accidents (based on the estimated coefficients). AADT was linked with more MV accidents independent of the severity of their outcomes, but its impact is higher for MV_Sl accidents. The outcomes for MV_Sl accidents were more explainable as this type of accidents is linked with congested traffic and high density conditions, when speeds are lower (e.g. Ivan, 2004). However, it is not clear whether the estimations for MV_KS accidents are equally interpretable. This is because although MV accidents are related with intense traffic, on the other hand accidents with serious consequences are linked with higher speed conditions (e.g. Aarts and Van Schagen, 2006).

MV_KS accidents were the only accident type that has been associated with the presence of positive grades in link-based models. Specifically, uphill links were approximately 49% less likely ($(e^{-0.6738} - e^0) \cdot 100\% = -49\%$) to have MV_KS accidents. As uphill sections have been associated with lower speeds and braking (Milton and Mannering, 1998), this finding is not in line with this for speed. MV_Sl accidents though, increased at links with negative grades and more than two lanes. The latter indicates that multiple lane changes which can eventually lead in more vehicle interactions (Persaud, 1992; Milton and Mannering, 1998).

5.3 Condition-Based Models

The univariate and the multivariate models denoted by equations 3.11 and 3.14 were fitted to the condition-based datasets using WinBUGS 1.4.3 (Spiegelhalter et al., 2003). The posterior distributions were obtained from 50,000 iterations of two Markov chains. The first 20,000 iterations were discarded from the final estimations as the burn-in sample. Convergence was visually detected from Markov chain history graphs of the models' coefficients. The results are structured in a similar manner with the previous section. First is the univariate model for all the accidents presented, then the multivariate models for accidents split by severity group (i.e. fatal, serious or slight) and by collision type (i.e. single-vehicle and multiple-vehicle). Following, are the estimations of the multivariate models for single-vehicle accidents by severity and multiple-vehicle accidents by severity.

5.3.1 All accidents

The modelling outcomes of the univariate condition-based model (*Condition-based All (10)* can be seen in Table 5.10. The best fitting traffic variable specification was specification 10 which means that both speed and volume per lane were expressed in quadratic forms and without an interaction term. The accident rate for the reference cases of categorical independent variables (i.e. Curve=0, Uphill=Downhill=0, Lanes above 2=0) that was the equation used to plot Figure 5.6 is equal with:

$$\frac{All\ accidents}{VehHr\ per\ mile} = exp(0.05756 \cdot Speed - 0.00056 \cdot Speed^2 - 0.01147 \cdot Volume + 0.00003 \cdot Volume^2 - 2.977) \quad (5.11)$$

Table 5.10: Parameter estimates for the condition-based univariate Poisson-lognormal model for all accidents (*Condition-based All (10)*)

All accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	0.05756**	0.011070	0.0007824	0.04036	0.04208	0.05796	0.07217
Speed squared	-0.00056**	0.000106	0.0000075	-0.00075	-0.00072	-0.00056	-0.00043
Volume	-0.01147**	0.000777	0.0000504	-0.01298	-0.01275	-0.01149	-0.01049
Volume squared	0.00003**	0.000002	0.0000002	0.00002	0.00002	0.00003	0.00003
Curve	0.11350**	0.037130	0.0008276	0.04073	0.05201	0.11340	0.16110
Uphill	2.25000**	0.069270	0.0032730	2.11800	2.13900	2.24800	2.33900
Downhill	2.93200**	0.069010	0.0033860	2.80400	2.82400	2.93000	3.02400
Lanes	-0.37400**	0.036910	0.0007741	-0.44580	-0.43460	-0.37390	-0.32660
Intercept	-2.97700**	0.294900	0.0205500	-3.51700	-3.45500	-2.97000	-2.61600
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$	7624	**statistically significant at the 95% credible interval					
p_D	936						
DIC	8560						

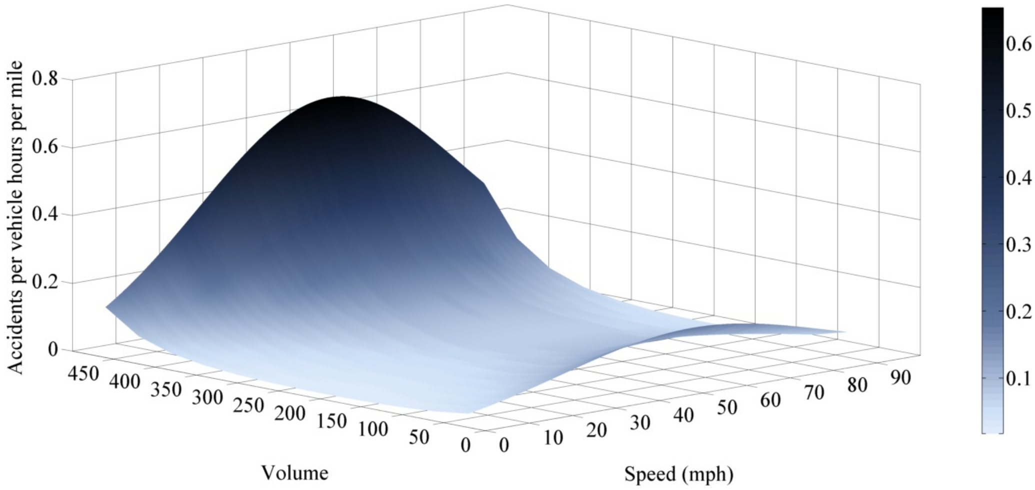


Figure 5.6: Contour plot of all the predicted accidents per vehicle hours travelled per mile as a function of speed and volume per lane (model: *Condition-based All (10)*).

Speed was found to have mainly a positive relationship with accidents. More specifically the accident-speed relationship according to this model can be described by an inverted U curve that reaches its peak at approximately 51.4 mph (estimated equating the first derivative of speed in equation 5.11 with zero). The shape of this curve reflects that the number of all accidents increases proportionally with speed until a certain point and then either it stabilises or decreases. The accident decrease at very high speeds could be potentially explained by the smaller number of accident prone reactions at very high speed conditions (Navon, 2003). However, another possible explanation is that the shape of this curve is due to the merger of accidents with dissimilar characteristics into a single dependent variable (all accidents). The results that are presented in the following sections confirm this statement.

Accident rate as a function of the volume per lane formed a U-shaped curve. As Figure 5.6 shows accident rate was highest at very high volume conditions and particularly when this was combined with relatively high speeds (approximately from 40 to 70 mph). Additionally, accident rate was higher for low volumes compared to the moderate ones. This result is in line with the findings of Gwynn (1967) and Ceder and Livneh (1982) who have also found a U-shaped relationship between accidents and volume. Similarly to speed though, this result might be the outcome of the examination of all the accidents combined. The left part of the U-curve might be mainly related with multiple-vehicle accidents that typically occur under heavier traffic while the right part might represent single-vehicle accidents that occur under less congested conditions.

The existence of curvature just before the accident location was found to increase the accident rate and specifically curved segment has a 12% (i.e. $(e^{0.1135} - e^{0}) \cdot 100\% = 12\%$) higher accident rate compared to straight segments. This finding is in line with the majority of the literature (e.g. Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000; Anastasopoulos and Mannering, 2009). Vertical alignment of the road section just before an accident was also associated with more accidents. The existence of both positive and negative slope seems to triggers accident occurrence although, based on the coefficient values, the latter has higher impact. This outcome is consistent with the findings of several studies (e.g. Shankar et al., 1995; Milton and Mannering, 1998). Finally, roads with more than two lanes were found related with lower accident counts. This is similar to the findings of Ma and Kockelman (2006) who reported that the number of lanes decreases accident counts for non-fatal accidents and the results by Park et al. (2010) who found that 6-lane freeways are less accident prone than 4 or 8-lanes but opposite to the majority of current literature (e.g. Chang, 2005; Milton and Mannering, 1998). A possible explanation for that could be that wider roads allow more manoeuvres for accident avoidance during a accident-prone encounter. Moreover, this result can also be explained by the inclusion of accidents that occurred on undivided (single) carriageways. Over half of the examined accidents occurred on A-roads that include some single carriageways which are related with hazardous vehicle interactions that may lead to accidents with severe consequences (e.g. head-on collisions). As it will be shown in sections 5.3.4 and 5.3.5 below, the results for the number of lanes are different in models that use accidents that occurred exclusively on motorways.

5.3.2 Accidents by severity type

Literature suggests that not all accident types can be attributed to the same precursors and thus a univariate model can be only partially informative about the relationships of accidents with their potential contributory factors (e.g. Ma and Kockelman, 2006; Park and Lord, 2007). In order to identify the differences between the contributory factors of accidents for each severity level a multivariate models has been applied. In the model fatal and serious accidents (KS) are merged into one category and are modelled along with slight injury accidents (Sl) (*Condition-based KS-Sl AMF (20)*). A model where accidents are divided into three severity categories (i.e. accidents with killed casualties (K), serious injuries (S) and slight injuries (Sl)) has also been applied but did not provide statistical significant results for speed probably due to excess zeroes. This model is presented in Appendix B.

Table 5.11 shows the estimates of the *Condition-based KS-Sl AMF (20)* model. The best fitting specification was quadratic for both speed and volume per lane and included an interaction term. The covariance-correlation matrix of the multivariate model is presented at Table 5.12. The values presented with bold font are the correlations between accidents with different severities. The correlation between KS and Sl accidents was very high (0.96). The high correlations between accidents with different severity confirm the necessity of the application of multivariate models. Similar correlations have been reported by previous research on multivariate models; for example Aguero-Valverde and Jovanis (2009) found 0.97 correlation between accidents with major and moderate injuries.

The accident rate by severity for the reference cases of categorical independent variables (i.e. Curve=0, Uphill=Downhill=0, Lanes above 2=0) for the *Condition-based KS-Sl AMF (20)* model is:

$$\frac{KS\,accidents}{VehHr\,per\,mile} = exp(0.03884 \cdot Speed - 0.00037 \cdot Speed^2 - 0.02037 \cdot Volume + 0.000035 \cdot Volume^2 + 0.00004 \cdot Speed \cdot Volume - 3.714) \quad (5.12)$$

$$\frac{Sl\,accidents}{VehHr\,per\,mile} = exp(0.05868 \cdot Speed - 0.00051 \cdot Speed^2 - 0.00759 \cdot Volume + 0.00002 \cdot Volume^2 - 0.00005 \cdot Speed \cdot Volume - 3.483) \quad (5.13)$$

Table 5.11: Parameter estimates for the condition-based multivariate Poisson-lognormal model for fatal and serious (KS) and slight (Sl) accidents (*Condition-based KS-Sl AMF (20)*)

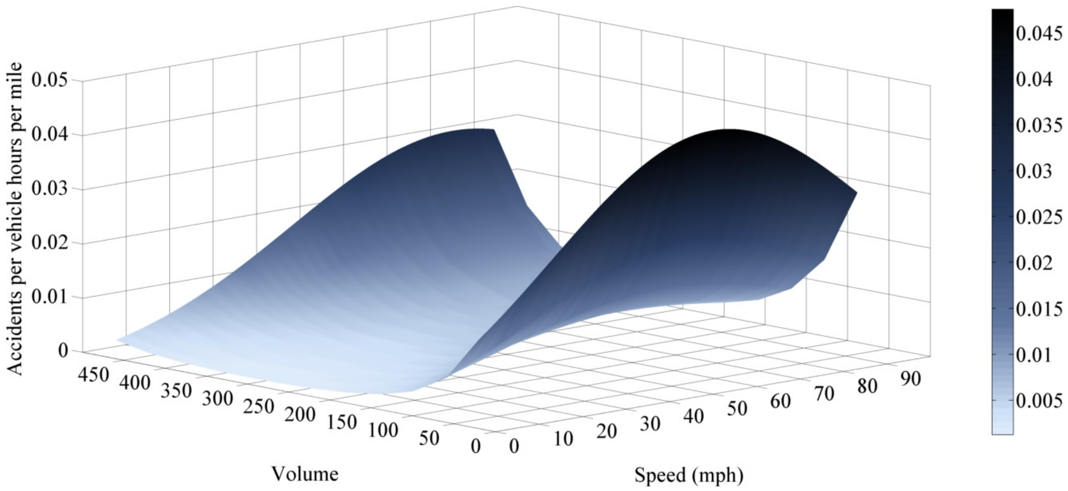
KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	0.03884 **	0.014430	0.0007483	0.01158	0.016782	0.06508	0.06888
Speed squared	-0.00037 **	0.000139	0.0000070	-0.00066	-0.000615	-0.00016	-0.00011
Volume	-0.02037 **	0.002008	0.0000889	-0.02417	-0.023630	-0.01695	-0.01621
Volume Squared	0.000035 **	0.000005	0.0000002	0.00003	0.000030	0.00005	0.00005
Speed*Volume	0.00004 *	0.000027	0.0000012	-0.00001	0.000006	0.00008	0.00009
Curve	0.08056	0.065340	0.0009904	-0.04698	-0.026650	0.18860	0.20870
Uphill	2.12500 **	0.158900	0.0044130	1.82500	1.870000	2.39400	2.45100
Downhill	2.95200 **	0.154600	0.0045080	2.65900	2.704000	3.21400	3.27000
Lanes	-0.64770 **	0.069540	0.0009894	-0.78550	-0.761900	-0.53300	-0.51090
Intercept	-3.71400 **	0.431000	0.0213500	-4.62600	-4.491000	-3.06100	-2.95400
ln(VehHr/mile)	1	-	-	-	-	-	-
Sl accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
Speed	0.05868 **	0.009759	0.0005183	0.03715	0.04005	0.07424	0.07678
Speed squared	-0.00051 **	0.000090	0.0000047	-0.00069	-0.00066	-0.00036	-0.00033
Volume	-0.00759 **	0.001200	0.0000603	-0.00986	-0.00959	-0.00552	-0.00518
Volume Squared	0.00002 **	0.000003	0.0000001	0.00002	0.00002	0.00003	0.00003
Speed*Volume	-0.00005 **	0.000016	0.0000008	-0.00008	-0.00008	-0.00002	-0.00002
Curve	0.11740 **	0.037730	0.0007983	0.04327	0.05534	0.17930	0.19080
Uphill	2.25700 **	0.071980	0.0022170	2.11600	2.13800	2.37500	2.39700
Downhill	2.91100 **	0.070250	0.0022510	2.77100	2.79300	3.02500	3.04600
Lanes	-0.32670 **	0.037900	0.0007605	-0.40100	-0.38910	-0.26460	-0.25260
Intercept	-3.48300 **	0.280600	0.0145800	-4.00800	-3.91400	-2.96000	-2.83700
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$	10683	**statistically significant at the 95% credible interval					
p_D	985	*statistically significant at the 90% credible interval					
DIC	11668						

Table 5.12: Combined Covariance-Correlation matrix of the random effect of the *Condition-based KS-Sl AMF (20)* model.

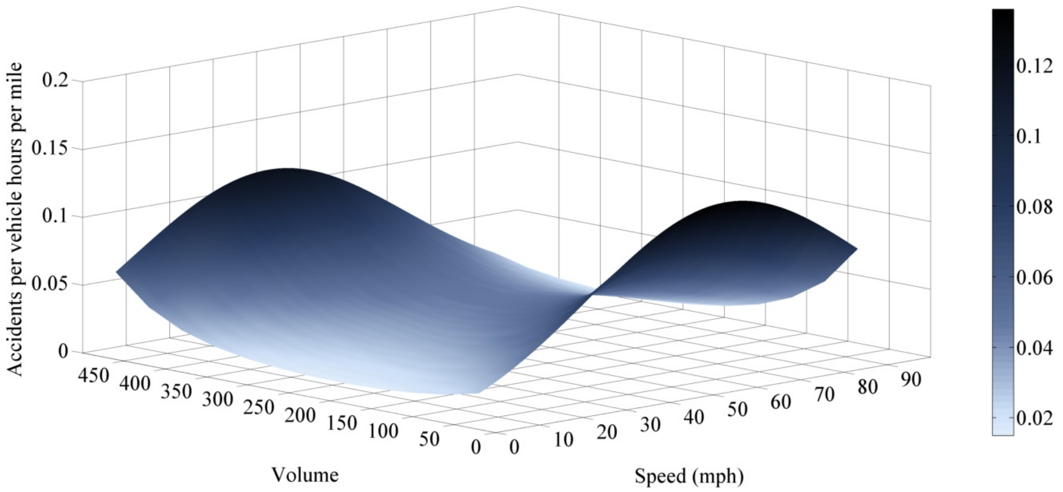
	KS accidents	Sl accidents
KS accidents	0.375**	0.323**
Sl accidents	**0.960****	0.302**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval



(a)



(b)

Figure 5.7: Contour plot of the predicted (a)fatal and serious (KS) and (b) slight (Sl) accidents per vehicle hours travelled per mile as a function of speed and volume per lane (model: *Condition-based KS-Sl AMF (20)*).

Observing Tables 5.11 and 5.12 and Figure 5.7 makes clear that accidents with different severity levels are associated with different traffic conditions. KS accidents (see Figure 5.7(a)) were found to have a generally proportional relationship with speed that for lower volume conditions after a certain maximum point either stabilises or decreases. Due to the presence of the interaction term in the model the peak of the accident-speed curve increased proportionally with volume per lane (e.g. the peak is 57.9 mph for 100 vehicles/lane and 74.1 mph for 400 vehicles/lane). This may suggest that at lower volume conditions there is more space available for manoeuvring so some accidents can be avoided, while when traffic is dense this is not possible. The frequency of Sl accidents (see Figure 5.7(b)) with speed was also found to increase until speed reaches 50 mph and after that point had a downward trend. Opposite to the KS accidents, the peak of the curve gradually decreased while volume increased (e.g. the peak is 52.6 mph for 100 vehicles/lane and 37.9 mph for 400 vehicles/lane). This outcome should not be interpreted as an indication that higher speeds are safer than the lower ones. The low frequency of Sl accidents at high speeds mainly shows that at higher speed conditions accidents tend to have usually more serious outcomes, a finding that is well-supported by the literature (e.g. Kloeden et al., 1997; Pei et al., 2012). The overall outcome from these results was that speed is a triggering factor for accident frequency and severity.

An interesting finding of this model was that the frequency of accidents is higher at low volume conditions than that of at high volume conditions, ceteris paribus. More specifically, the relationship between accident rate and volume was described as an approximate U-shaped curve with the minimum accident rates were found to be at 256 and 263 vehicles per lane for KS and Sl accidents respectively at average speed conditions. This outcome is consistent with the results for speed, because high volume is usually associated with congested, low speed conditions when accidents are less likely to be severe and reported (Lord, 2002). Another explanation for this finding could be that low volumes are related with higher speed variations (when traffic builds up) that may increase the probability of accidents (Lave, 1985; Baruya and Finch, 1994; Garber and Ehrhart, 2000; Aarts and Van Schagen, 2006). This is because when the volume decreases drivers have more freedom to choose their own speed and so speed patterns on the roadway tend to be less uniform leading to more encounters between vehicles (Elvik et al., 2004). Additionally, low volumes occur more often during off-peak periods, such as night time, that is related to insufficient light conditions and extreme driving behaviours (e.g. drinking and driving) that are also factors proved to trigger severe accident occurrence (Jonah, 1986;

Chang and Wang, 2006; Clarke et al., 2010).

Curvature was not shown to have a statistically significant relationship with KS accidents although it was related with increased Sl accident rates. A possible explanation for this may be that speeding that is a contributory factor for KS accidents is more unlikely to occur on curved sections where drivers tend to drive more carefully (e.g. Chang, 2005). Presence of uphill or downhill vertical grades and roads with two lanes or less were shown to increase the accident rates of all severity levels similarly to what was found in the univariate model presented in section 5.3.1.

5.3.3 Accidents by collision type

Table 5.13 and Figure 5.8 show the results for the *Condition-based SV-MV (10)* model. The best fitting specification for this model was again quadratic for both speed and volume per lane without an interaction term. The correlation between single-vehicle (SV) and multiple-vehicle (MV) accidents was 0.876 (see Table 5.14). The SV and MV accident rate for the reference values of the categorical independent variables is:

$$\frac{SV\ accidents}{VehHr\ per\ mile} = exp(0.08065 \cdot Speed - 0.00046 \cdot Speed^2 - 0.01764 \cdot Volume + 0.00003 \cdot Volume^2 - 5.2) \quad (5.14)$$

$$\frac{MV\ accidents}{VehHr\ per\ mile} = exp(0.03056 \cdot Speed - 0.00035 \cdot Speed^2 - 0.00517 \cdot Volume + 0.00002 \cdot Volume^2 - 3.487) \quad (5.15)$$

Table 5.13: Parameter estimates for the condition-based multivariate Poisson-lognormal model for single-vehicle (SV) and multiple-vehicle (MV) accidents (*Condition-based SV-MV (10)*)

SV accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	0.08065 **	0.011952	0.000694	0.060742	0.062596	0.101970	0.106481
Speed squared	-0.00046 **	0.000111	0.000006	-0.000701	-0.000659	-0.000296	-0.000275
Volume	-0.01764 **	0.001159	0.000053	-0.019787	-0.019480	-0.015620	-0.015253
Volume squared	0.00003 **	0.000004	0.001704	0.242022	0.254270	0.380300	0.390643
Curve	-0.04855	0.051957	0.000644	-0.151294	-0.134407	0.000004	0.000005
Uphill	2.27117 **	0.124568	0.004381	2.037040	2.071680	2.481940	2.525070
Downhill	2.97919 **	0.122469	0.004476	2.749740	2.783110	3.186530	3.230890
Lanes	-0.54559 **	0.054326	0.000669	-0.652050	-0.634800	-0.455914	-0.439609
Intercept	-5.20008 **	0.354946	0.019799	-5.972930	-5.827420	-4.660250	-4.587040
ln(VehHr/mile)	1	-	-	-	-	-	-
MV accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
Speed	0.03056 **	0.009444	0.0005522	0.01214	0.01478	0.04528	0.04905
Speed squared	-0.00035 **	0.000092	0.0000053	-0.00052	-0.00049	-0.00019	-0.00016
Volume	-0.00517 **	0.000956	0.0000517	-0.00694	-0.00666	-0.00356	-0.00320
Volume squared	0.00002 **	0.000003	0.0000002	0.00001	0.00001	0.00002	0.00002
Curve	0.25508 **	0.039271	0.0006518	0.17823	0.19091	0.31938	0.33234
Uphill	2.17001 **	0.077392	0.0024701	2.01937	2.04510	2.29740	2.32051
Downhill	2.84673 **	0.075486	0.0025174	2.69704	2.72194	2.97000	2.99523
Lanes	-0.04794	0.039475	0.0006163	-0.12558	-0.11329	0.01668	0.02877
Intercept	-3.48742 **	0.249639	0.0141333	-3.98386	-3.91887	-3.08287	-3.02694
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$	10823	**statistically significant at the 95% credible interval					
p_D	11690						
DIC	867						

Table 5.14: Combined Covariance-Correlation matrix of the random effect of the *Condition-based SV-MV (10)* model.

	SV accidents	MV accidents
SV accidents	0.33**	0.246**
MV accidents	**0.876****	0.239**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval

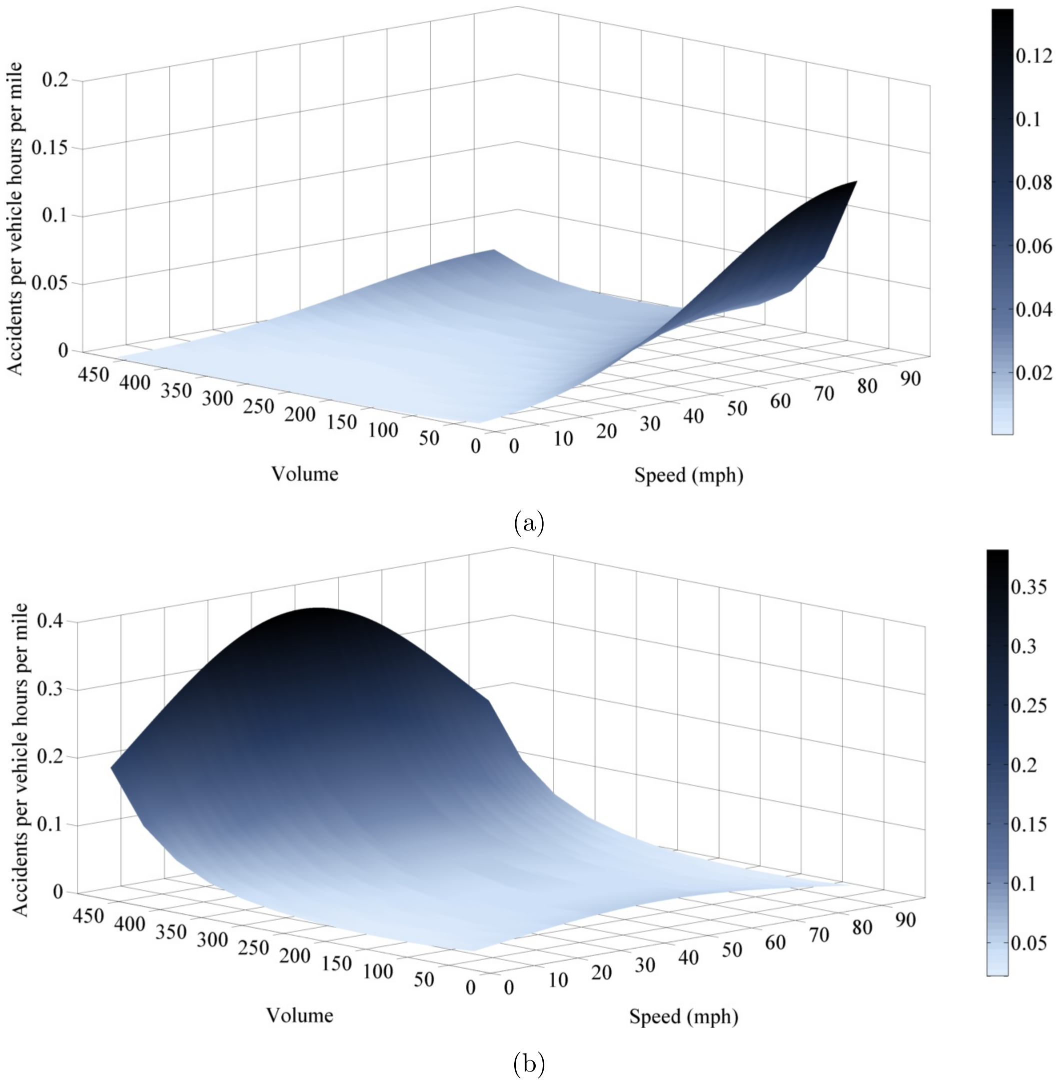


Figure 5.8: Contour plot of the predicted (a) single-vehicle (SV) and (b) multiple-vehicle (MV) accidents per vehicle hours travelled per mile as a function of speed and volume per lane (model: *Condition-based SV-MV (10)*).

Comparing the coefficient values of the SV with those of the MV it became clear that the two accident types tend to occur under entirely different traffic conditions. SV collisions increased at high speed and low volume conditions; in other words, they tend to occur more frequently when traffic density is low. On the other hand, MV collisions seemed to be related with lower speeds and higher volumes and consequently with more intense traffic. In principle, this outcome, that is in line with the existing literature (e.g. Ivan, 2004; Ivan et al., 2000, 1999), re-confirmed that modelling accidents by type is advantageous as it can be more informative about the circumstances that particular accidentaccidentsaccident types occur. However, interpreting these results without looking at different severity levels by accident type might be misleading, especially for MV accidents.

5.3.4 Single-vehicle accidents by severity

This section presents the coefficient estimates of the multivariate models that examined single-vehicle accidents by severity. Table 5.15 and Figure 5.9 present the results for the *Condition-based SV KS-Sl (9)* model that considers all the SV accidents on the SRN. The specification combination that fitted the data best was when speed and volume were logarithmically transformed. Table 5.17 and Figure 5.10 show the outcomes of the *Condition-based SV KS-SL moto (3)* model that includes SV accidents that occurred on motorways only. In the best fitting specification for this model speed was squared and volume was logarithmically transformed. In Tables 5.16 and 5.18 it can be seen that the correlations between SV_KS and SV_Sl accidents were found very high (0.903 for the *Condition-based SV KS-Sl (9)* and 0.812 for the *Condition-based SV KS-SL moto (3)*). The equations expressing the SV_KS and SV_Sl accident rate that was derived from the *Condition-based SV KS-Sl (9)* model for the reference geometric variables are:

$$\frac{SV\_KS\ accidents}{VehHr\ per\ mile} = exp(1.085 \cdot ln(Speed) - 0.8943 \cdot ln(Volume) - 5.3093) \quad (5.16)$$

$$\frac{SV\_Sl\ accidents}{VehHr\ per\ mile} = exp(1.505 \cdot ln(Speed) - 0.7843 \cdot ln(Volume) - 6.4418) \quad (5.17)$$

The corresponding equations for the *Condition-based SV KS-SL moto (3)* model are:

$$\frac{SV\_KS\ moto\ accidents}{VehHr\ per\ mile} = exp(0.00031 \cdot Speed^2 - 0.8037 \cdot ln(Volume) - 4.375) \quad (5.18)$$

$$\frac{SV\_Sl\ moto\ accidents}{VehHr\ per\ mile} = exp(0.00019 \cdot Speed^2 - 0.7578 \cdot ln(Volume) - 2.734) \quad (5.19)$$

Table 5.15: Parameter estimates for the condition-based multivariate Poisson-lognormal model for fatal and serious (SV_KS) and slight (SV_Sl) single-vehicle accidents (*Condition-based SV_KS-Sl (9)*)

SV_KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
ln(Speed)	1.0850 **	0.1777	0.0097	0.7493	0.7984	1.3618	1.4062
ln(Volume)	-0.8943 **	0.0450	0.0012	-0.9811	-0.9673	-0.8191	-0.8044
Curve	-0.0830	0.0995	0.0010	-0.2765	-0.2451	0.0817	0.1136
Uphill	1.9192 **	0.2443	0.0075	1.4550	1.5240	2.3261	2.3995
Downhill	2.8110 **	0.2355	0.0074	2.3666	2.4327	3.2051	3.2772
Lanes	-0.7147 **	0.1053	0.0009	-0.9218	-0.8887	-0.5415	-0.5075
Intercept	-5.3093 **	0.7406	0.0404	-6.6894	-6.4587	-4.1203	-3.9292
ln(VehHr/mile)	1	-	-	-	-	-	-
SV_Sl accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
ln(Speed)	1.5050 **	0.1200	0.0068	1.2831	1.3133	1.7068	1.7407
ln(Volume)	-0.7843 **	0.0251	0.0008	-0.8317	-0.8242	-0.7421	-0.7338
Curve	-0.0502	0.0562	0.0007	-0.1600	-0.1431	0.0420	0.0600
Uphill	2.3155 **	0.1428	0.0050	2.0414	2.0817	2.5532	2.6035
Downhill	2.9689 **	0.1394	0.0050	2.7007	2.7423	3.2031	3.2502
Lanes	-0.4619 **	0.0574	0.0006	-0.5743	-0.5557	-0.3674	-0.3489
Intercept	-6.4418 **	0.5185	0.0294	-7.4520	-7.3281	-5.6116	-5.4784
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$	6345	**statistically significant at the 95% credible interval					
p_D	6779						
DIC	434						

Table 5.16: Combined Covariance-Correlation matrix of the random effect of the *Condition-based SV_KS-Sl (9)* model.

	SV_KS accidents	SV_Sl accidents
SV_KS accidents	0.295**	0.262**
SV_Sl accidents	**0.903****	0.285**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval

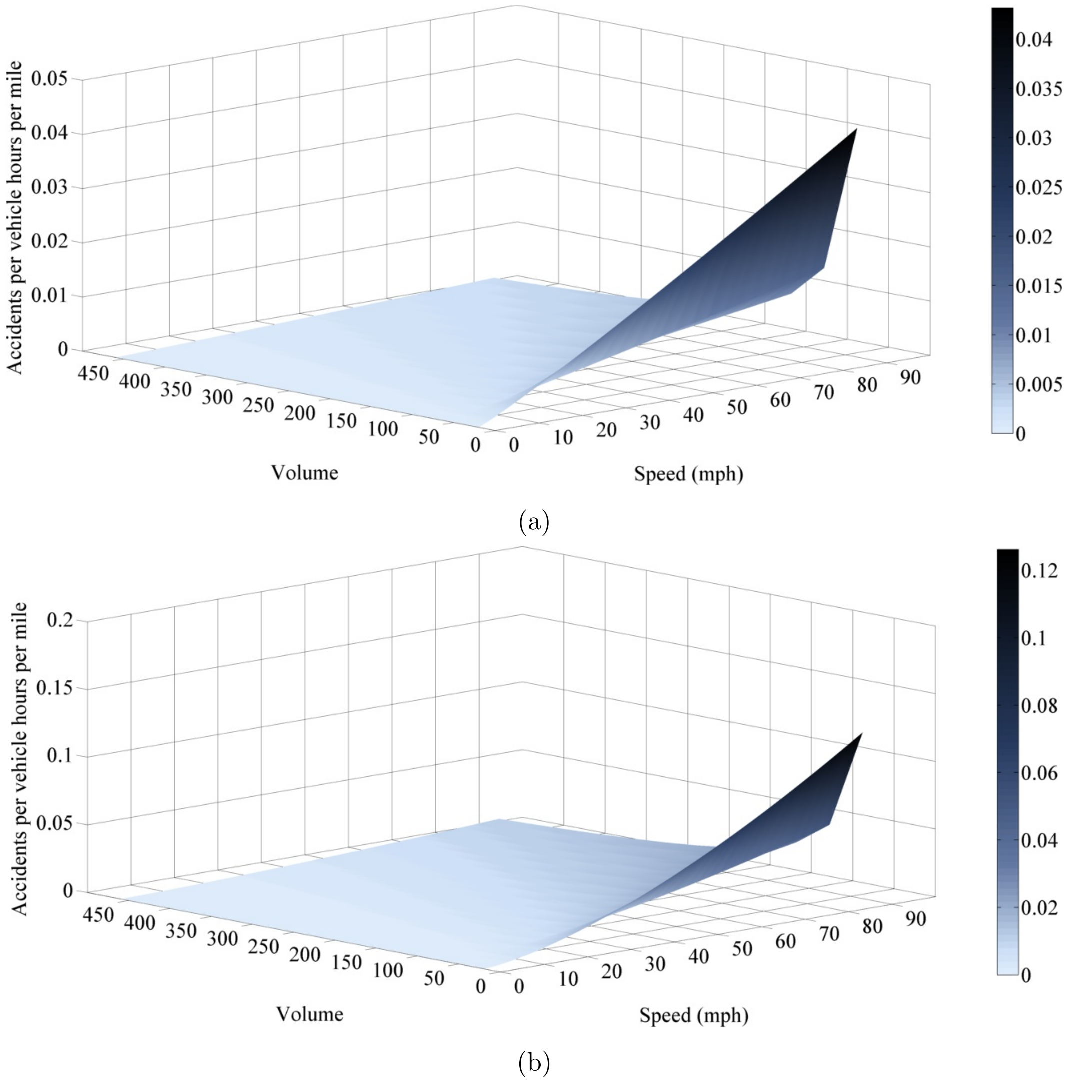


Figure 5.9: Contour plot of the predicted (a) fatal or serious (KS) and (b) slight (Sl) single-vehicle (SV) accidents per vehicle hours travelled per mile as a function of speed and volume per lane (model: *Condition-based SV_KS-Sl (9)*).

Table 5.17: Parameter estimates for the condition-based multivariate Poisson-lognormal model for fatal and serious (SV_KS) and slight (SV_Sl) motorway single-vehicle accidents (*Condition-based SV_KS-SL moto (3)*)

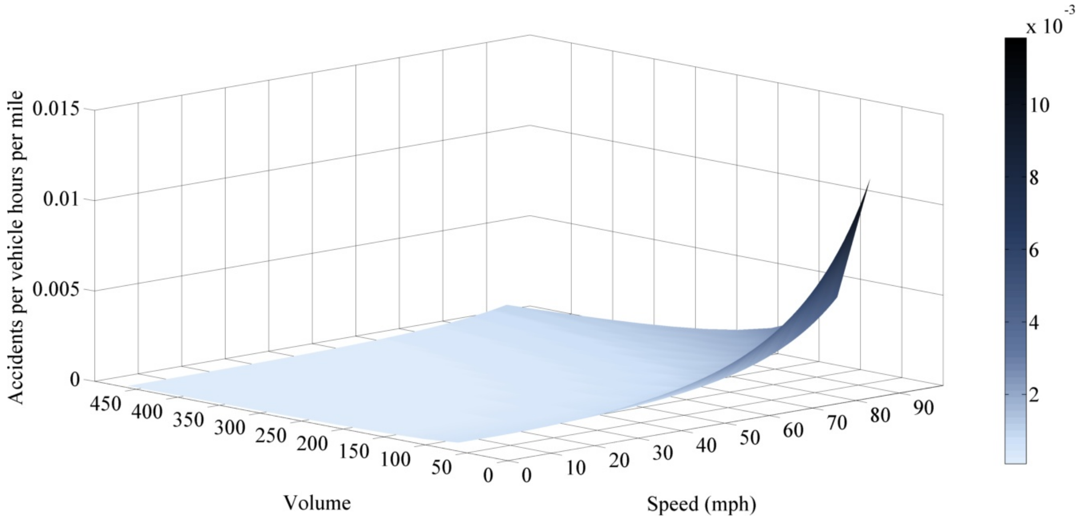
SV_KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed squared	0.00031 **	0.00008	0.000002	0.00016	0.00018	0.00043	0.00046
ln(Volume)	-0.80370 **	0.07473	0.002060	-0.95020	-0.92640	-0.68200	-0.65740
Curve	0.26960 *	0.15170	0.001377	-0.02645	0.02165	0.52200	0.56940
Uphill	1.77200 **	0.34860	0.007825	1.12900	1.22600	2.36800	2.49500
Downhill	2.60000 **	0.33390	0.007781	1.99000	2.08100	3.18000	3.30400
Lanes	0.82470 **	0.16360	0.001689	0.50910	0.55800	1.09800	1.15000
Intercept	-4.37500 **	0.60270	0.020980	-5.56300	-5.38000	-3.39200	-3.22300
ln(VehHr/mile)	1	-	-	-	-	-	-
SV_Sl accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed squared	0.00019 **	0.00003	0.000001	0.00013	0.00014	0.00025	0.00026
ln(Volume)	-0.75780 **	0.03620	0.001016	-0.82770	-0.81700	-0.69790	-0.68650
Curve	0.27369 **	0.11432	0.001108	0.04912	0.08487	0.46082	0.49686
Uphill	2.10300 **	0.18000	0.004502	1.76300	1.81200	2.40800	2.46900
Downhill	2.82500 **	0.17500	0.004513	2.49300	2.54300	3.12100	3.17900
Lanes	0.89620 **	0.07684	0.000822	0.74730	0.77050	1.02400	1.04900
Intercept	-2.73400 **	0.29220	0.010120	-3.30300	-3.21700	-2.25400	-2.16200
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$ 4204	**statistically significant at the 95% credible interval						
p_D 132	*statistically significant at the 90% credible interval						
DIC 4335							

Table 5.18: Combined Covariance-Correlation matrix of the random effect of the *Condition-based SV_KS-SL moto (3)* model.

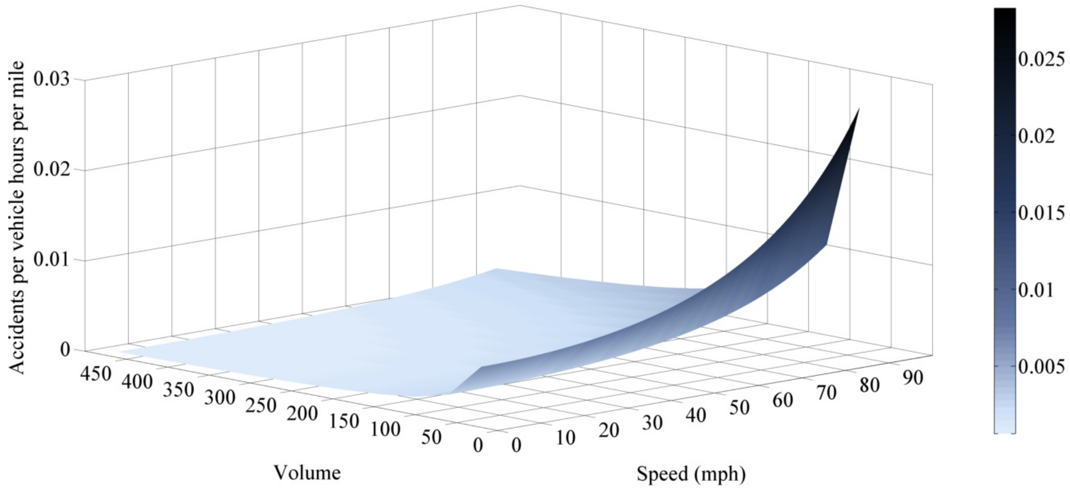
	MV_KS accidents	MV_Sl accidents
MV_KS accidents	0.212**	0.126**
MV_Sl accidents	**0.812****	0.113**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval



(a)



(b)

Figure 5.10: Contour plot of the predicted (a) fatal or serious (KS) and (b) slight (Sl) motorway single-vehicle (SV) accidents per vehicle hours travelled per mile as a function of speed and volume per lane (model: *Condition-based SV_KS-SL moto (3)*).

Regardless of the severity of their outcomes, SV accident frequency increases proportionally with speed (see Figures 5.9 and 5.10). The coefficients of speed for motorway SV_KS and SV_Sl accidents showed that, as expected, higher speed is also related with more serious injuries (see Table 5.17). However, the coefficient of speed was lower for SRN SV_KS than for SV_Sl (see Table 5.15). A possible explanation for this counterintuitive result might be that the proportion of the SRN SV_KS accidents that occurred on A-roads (58.8%) is higher than the proportion of the SV_Sl that occurred on A-roads (50.7%). Because of this, SRN SV_KS accidents might be more representative of A-road accidents that by definition are likely to have occurred under lower speed conditions. The SV accident-volume relationship was found entirely different than the one with speed.

More specifically, higher SV accident frequency and severity was found to be linked with lower traffic volume.

The results of these models are explainable and confirm what is intuitively believed: that SV accidents are probably the most speed-related accident type. SV accidents are associated with circumstances that are linked with speeding such as loss of control, alcohol or drug misuse, risk-taking actions, fatigue and sleepiness (Lang et al., 1996; Xie et al., 2012; Kim et al., 2013). They usually occur during off-peak times and especially at night time when density is at low levels and vehicle encounters are less likely (Ivan et al., 1999).

Road gradient was found to have negative impact on SV accidents. The presence of curves was shown to increase only motorway SV accidents though. This might be due to the higher average speeds on the motorways that in combination with speed could increase accident probability. Roads with more than two lanes are found to be more related with motorway accident occurrences. It is not clear whether this finding justifies that the number of lanes generally influences accidents due to the restrictive specification of the variable. The difference in the results between the two models however indicates a difference between the relationship of number of lanes for A-roads and motorways.

5.3.5 Multiple-vehicle accidents by severity

Tables 5.19, 5.20, 5.21 and 5.22 as well as Figures 5.11 and 5.12 present the results of the models for the SRN and motorway MV accidents by severity respectively (i.e *Condition-based SV KS-Sl (18)* and *Condition-based MV KS-SL moto (12)* models). The best fitting variable specification for the former model included the logarithm of speed and squared volume while in the latter model speed was linear and volume squared. Both the models included interaction terms. The correlation coefficient was very high for the *Condition-based SV KS-Sl (18)* model (0.897) and slightly lower for the *Condition-based MV KS-SL moto (12)* model (0.649) (see Figures 5.20 and 5.22). The SRN MV_KS and MV_Sl accident rates for the reference cases can be expressed by the following equations:

$$\frac{MV\_KS\ accidents}{VehHr\ per\ mile} = exp(0.82095 \cdot ln(Speed) + 0.00001 \cdot Volume^2 - 0.0001 \cdot Speed \cdot Volume^2 - 8.03) \quad (5.20)$$

$$\frac{MV\_Sl\ accidents}{VehHr\ per\ mile} = exp(0.44509 \cdot ln(Speed) + 0.00002 \cdot Volume^2 - 0.00008 \cdot Speed \cdot Volume^2 - 4.89) \quad (5.21)$$

The motorway MV_KS and MV_Sl accident rates are equal:

$$\frac{MV\_KS\ moto\ accidents}{VehHr\ per\ mile} = exp(0.03057 \cdot Speed + 0.00001 \cdot Volume^2 - 0.00011 \cdot Speed \cdot Volume^2 - 7.65) \quad (5.22)$$

$$\frac{MV\_Sl\ moto\ accidents}{VehHr\ per\ mile} = exp(-0.011316 \cdot Speed + 0.00004 \cdot Volume^2 - 0.000027 \cdot Speed \cdot Volume^2 - 4.23) \quad (5.23)$$

Table 5.19: Parameter estimates for the condition-based multivariate Poisson-lognormal model for fatal and serious (MV_KS) and slight (MV_Sl) multiple-vehicle accidents (*Condition-based MV KS-Sl (18)*)

MV_KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
ln(Speed)	0.82095 **	0.230523	0.0131319	0.3677650	0.421568	1.19561	1.28473
Volume Squared	0.00001 *	0.000004	0.0000002	-0.0000003	0.000001	0.00002	0.00002
Speed*Volume	-0.00010 **	0.000024	0.0000012	-0.0001470	-0.000140	-0.00006	-0.00005
Curve	0.27544 **	0.092098	0.0009709	0.0938857	0.124243	0.42679	0.45483
Uphill	1.93934 **	0.223761	0.0068636	1.5159800	1.581600	2.32307	2.40337
Downhill	2.80867 **	0.215267	0.0068404	2.4057700	2.466790	3.17936	3.25401
Lanes	-0.12876	0.092723	0.0011095	-0.3094220	-0.280351	0.02378	0.05316
Intercept	-8.03037 **	0.850118	0.0480741	-9.6859500	-9.383410	-6.58344	-6.33447
ln(VehHr/mile)	1	-	-	-	-	-	-
MV_Sl accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
ln(Speed)	0.44509 **	0.100003	0.0057987	0.2154290	0.252006	0.58510	0.60427
Volume Squared	0.00002 **	0.000002	0.0000001	0.0000111	0.000012	0.00002	0.00002
Speed*Volume	-0.00008 **	0.000012	0.0000006	-0.0001045	-0.000100	-0.00006	-0.00006
Curve	0.25126 **	0.040977	0.0006403	0.1706950	0.183615	0.31849	0.33210
Uphill	2.17826 **	0.082759	0.0026507	2.0201900	2.043090	2.31542	2.34303
Downhill	2.83694 **	0.081125	0.0027282	2.6810600	2.704490	2.97031	2.99630
Lanes	-0.02870	0.041998	0.0006899	-0.1110720	-0.098437	0.04009	0.05347
Intercept	-4.89359 **	0.377902	0.0218648	-5.49489	-5.421130	-4.19463	-4.00560
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$	7919	**statistically significant at the 95% credible interval					
p_D	617	*statistically significant at the 90% credible interval					
DIC	8537						

Table 5.20: Combined Covariance-Correlation matrix of the random effect of the *Condition-based MV KS-Sl (18)* model.

	MV_KS accidents	MV_Sl accidents
MV_KS accidents	0.31**	0.244**
MV_Sl accidents	**0.897****	0.239**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval

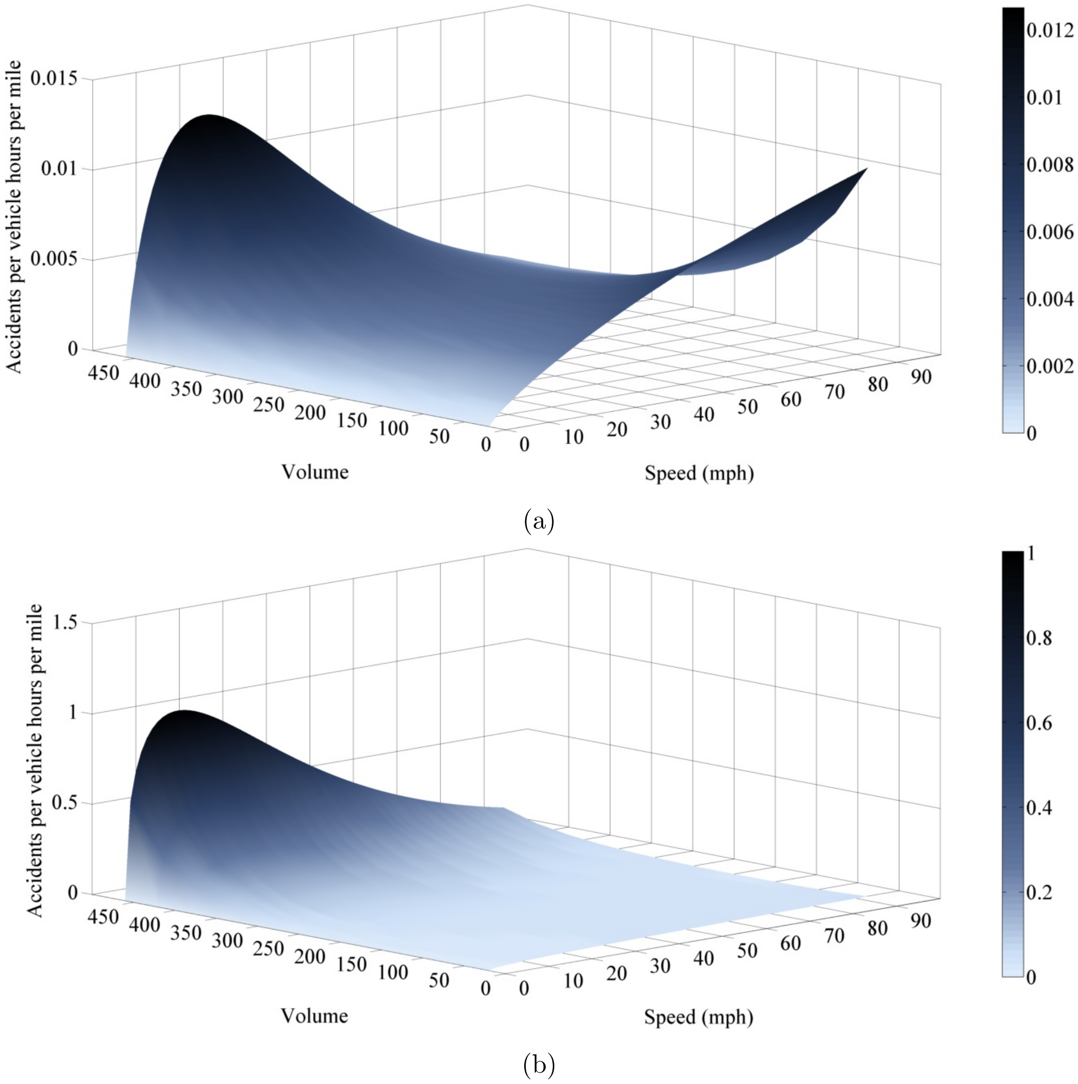


Figure 5.11: Contour plot of the predicted (a) fatal or serious (KS) and (b) slight (Sl) multiple-vehicle (MV) accidents per vehicle hours travelled per mile as a function of speed and volume per lane (model: *Condition-based MV KS-Sl (18)*).

Table 5.21: Parameter estimates for the condition-based multivariate Poisson-lognormal model for fatal and serious (MV_KS) and slight (MV_Sl) motorway multiple-vehicle accidents (*Condition-based MV_KS-SL moto (12)*)

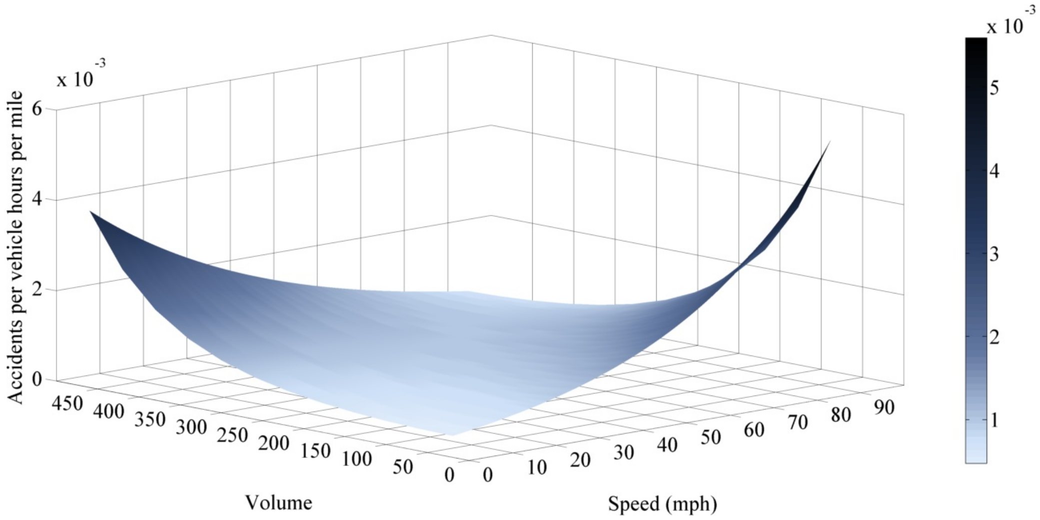
MV_KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	0.03057 **	0.00671	0.000277	0.01718	0.01961	0.04166	0.04401
Volume squared	0.00001 **	0.02502	0.000901	0.03514	0.04301	0.12550	0.13260
Speed*Volume	-0.00011 **	0.00002	0.000001	-0.00015	-0.00014	-0.00009	-0.00008
Curve	0.43220 **	0.09468	0.000995	0.24760	0.27780	0.58890	0.61980
Uphill	1.96400 **	0.24250	0.006220	1.50500	1.57500	2.37000	2.45900
Downhill	2.85200 **	0.23370	0.006246	2.41300	2.47800	3.24800	3.33000
Lanes	1.25300 **	0.11030	0.001199	1.03900	1.07200	1.43500	1.47300
Intercept	-7.65800 **	0.44410	0.018440	-8.54900	-8.39200	-6.94300	-6.80400
ln(VehHr/mile)	1	-	-	-	-	-	-
MV_Sl accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	-0.011316 **	0.00374	0.000173	-0.01839	-0.01726	-0.00542	-0.00413
Volume squared	0.000004 **	0.01363	0.000584	0.01531	0.01908	0.06357	0.06831
Speed*Volume	-0.000027 **	0.00001	0.000000	-0.00005	-0.00004	-0.00001	-0.00001
Curve	0.482100 **	0.04645	0.000624	0.39050	0.40520	0.55820	0.57290
Uphill	2.149000 **	0.10250	0.002764	1.95500	1.98500	2.32000	2.35600
Downhill	2.853000 **	0.09973	0.002750	2.66400	2.69300	3.02100	3.05600
Lanes	1.216000 **	0.05064	0.000778	1.11700	1.13300	1.30000	1.31600
Intercept	-4.238000 **	0.23220	0.010480	-4.67200	-4.60300	-3.85000	-3.76700
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$	6588	**statistically significant at the 95% credible interval					
p_D	267						
DIC	6855						

Table 5.22: Combined Covariance-Correlation matrix of the random effect of the *Condition-based MV_KS-SL moto (12)* model.

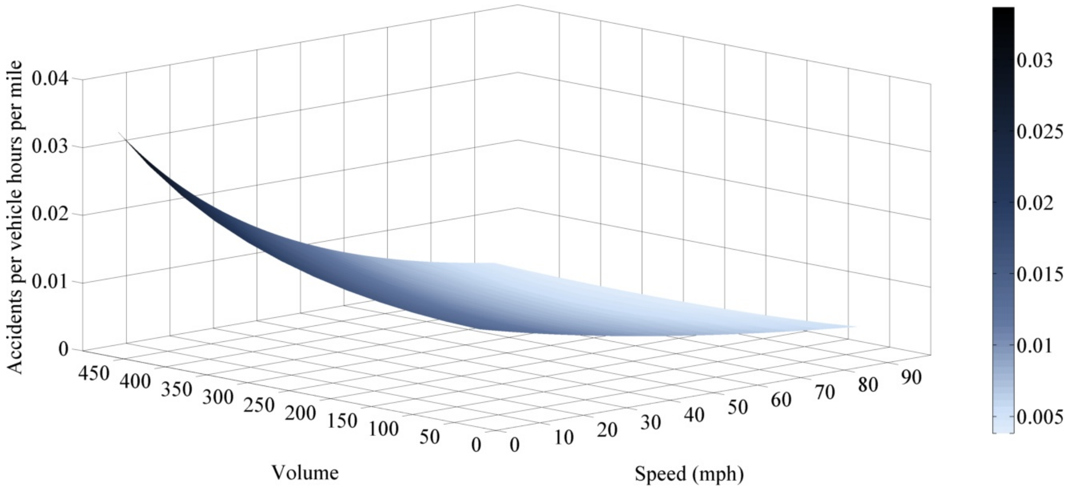
	MV_KS accidents	MV_Sl accidents
MV_KS accidents	0.056**	0.053**
MV_Sl accidents	**0.649****	0.122**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval



(a)



(b)

Figure 5.12: Contour plot of the predicted (a) fatal or serious (KS) and (b) slight (Sl) motorway multiple-vehicle (MV) accidents per vehicle hours travelled per mile as a function of speed and volume per lane (model: *Condition-based MV_KS-SL moto (12)*).

Figures 5.11 and 5.12 show that the traffic conditions that were linked with MV_KS were not exactly the same with those of MV_Sl accidents. MV_KS accidents were found to generally increase for higher speeds when the volume is relatively low and the opposite for high volume conditions (see Figures 5.11(a) and 5.12(a)). This outcome can be explained considering the characteristics of the two main collision types of MV same direction accidents: side and rear-end collisions. Side impacts are more likely to occur during overtaking manoeuvres that are more frequent under high speed and lower density conditions. On the contrary, rear-end collisions are linked with more dense traffic and might not be particularly related with speeding. This result might also be the effect of the merger of fatal and serious accidents into one category. If accidents with killed casualties were modelled separately speed would probably be positive for all volumes however that was not possible due to the low number of cases in the datasets.

MV_Sl accidents on the other hand had a negative relationship with speed but a positive relationship with volume. The maximum MV_Sl accident rate can be observed (see Figures 5.11(b) and 5.12(b)) at heavily congested conditions when traffic is very high and mobility is limited. This is consistent with past findings that suggest that high traffic intensity and peak hours are related with more MV accidents (Ivan et al., 1999; Ivan, 2004).

Road vertical alignment and curves were found to have negative impact on MV accidents of all severities. Motorway MV accident frequency was found to be higher for roads with more than two lanes, while the opposite happened for SRN MV accidents that were more likely to occur on sections with up to two lanes. As it was previously mentioned at Section 5.3.4, this result might be more related with the restrictive specification of the dummy variable for lanes.

5.4 Comparison of the Link-Based and the condition-based Approaches

The results of the link-based and the condition-based models that have been presented in Sections 5.2 and 5.3 have shown that there are significant differences between the two aggregation approaches. Tables 5.23 and 5.24 summarise the relationships of accidents by type with the examined traffic and geometric variables respectively. Upward pointing arrows represent positive relationships and downward pointing arrows negative relationships. Relationships that can be described by U-shaped or inverse-U-shaped curves are denoted by U and inverse-U respectively. When the curves tend to have a dominant tendency (positive or negative) this is shown in brackets. The statistical insignificant relationships between variables and accidents are represented with a dash.

Table 5.23: Qualitative relationships of the traffic variables (speed and volume) with accidents according to the outcomes of the link-based and the condition-based aggregation approaches.

Accident	Speed		Volume	
Type	Link-Based	Condition-Based	Link-Based	Condition-Based
All	↓	inverse-U (↑)	inverse-U (↑)	U
KS	↓	↑	↑	U
Sl	↓	inverse-U	↑	U
SV	↓	↑	↑	↓
SV_KS	↓	↑	–	↓
SV_Sl	↓	↑	↑	↓
MV	↓	inverse-U (↓)	↑	↑
MV_KS	↓	U (↑)	↑	U
MV_Sl	↓	↓	↑	↑

Table 5.24: Qualitative relationships of the geometric variables (curvature, gradient and lanes) with accidents according to the outcomes of the link-based and the condition-based aggregation approaches.

	Curve		Uphill		Downhill		Lanes	
Accident Type	Link Based	Condition Based	Link Based	Condition Based	Link Based	Condition Based	Link Based	Condition Based
All	–	↑	–	↑	↑	↑	–	↓
KS	–	–	–	↑	–	↑	↑	↓
Sl	–	↑	–	↑	↑	↑	↑	↓
SV	–	–	–	↑	↑	↑	–	↓
SV_KS	–	–	–	↑	–	↑	–	↓
SV_Sl	–	–	–	↑	↑	↑	–	↓
MV	–	↑	–	↑	↑	↑	↑	–
MV_KS	–	↑	↓	↑	–	↑	–	↓
MV_Sl	–	↑	–	↑	↑	↑	↑	↓

Observing Table 5.23 it is clear that at most of the cases the impact of speed and volume varies between the two examined aggregation approaches and sometimes it is exactly the opposite. For instance, SV accidents according to the link-based approach are negatively related with speed and positively related with traffic volume (represented by AADT in the model) but according to the condition-based, these accidents are related with higher speeds and lower volumes. Exceptions to this are the findings for MV and MV_Sl accidents that are shown to be negatively related to speed and positively related with volume. Geometrical variables are found to have some different (e.g. lanes) but also some similar results (e.g. downhill) between the two accident aggregation approaches. Considering the fact that the models originate from exactly the same data and were analysed with the same models the differences in the results are almost definitely related with the selected aggregation approach especially for time varying traffic variables. The differences of the geometric variables can be also due to the specification of the variables; as these variables are less sensitive to aggregation bias. Accident aggregation has been disregarded by most researchers, who mainly focused their research on developing more advanced statistical models; however it seems that the way accident data are grouped before the statistical analysis is also crucial.

Link-based and condition-based models cannot be directly compared to each other neither using goodness of fit statistics nor based solely on their outcomes. At this point it is important to highlight that link-based and condition-based models examine accident occurrence from entirely different perspectives, which leads to different interpretations of their outcomes. Link-based models examine the relationships of the average characteristics of pre-defined network areas (i.e. links) with accidents. Condition-based models examine the relationships of preceding traffic conditions with accidents. Considering this and the outcomes in Section 5.2 it can be said that link-based models practically show that the expected number of accidents (including SV and accidents with serious and fatal injuries) on faster links is lower than those with lower average speed. This result is probably valid but it might be due to other traffic or geometrical characteristics of these links rather than speed. Assuming that this outcome reflects the relationship of individual accidents with speed is an ecological fallacy. Link-based models cannot clearly define the accident-speed relationships and therefore are possibly not suitable for the estimation on of the impact of a speed limit increase.

On the other hand, condition-based approaches focus on the location of the accident and can provide more information on the actual circumstances which are related with accidents. Based on that, it can be argued that they give a significantly more accurate representation of accidents at an individual-level. That leads to more interpretable outcomes that are mostly in line with existing literature which is also an indication for higher reliability. Considering this, condition-based models are more appropriate for the definition of the accident-speed relationships and following the impact of a potential speed limit raise.

5.5 The Impact of Accident Location Accuracy in Accident Modelling

Both the examined accident aggregation approaches employ accident locations for grouping accidents. In the link-based approach accident location determines the link where each accident is assigned and in the condition-based approach accident location is the key element for specifying the pre-accident traffic and geometric conditions. The impact of the accuracy of accident location on accident models independent of their aggregation method is an interesting methodological issue which can provide useful insight about the usefulness of accident data refinement methods.

Assuming that higher accident location accuracy should enhance the accuracy of the modelling outcomes, the models that have been presented so far in this thesis employed the refined accident locations that were produced by the AMF method presented in Sections 3.3 and 4.3.1. Comparing the outcomes of identical models that are based on less precise location it is possible to examine whether this assumption is correct. This comparison will be done using accident locations that were provided from the output of the simplest accident mapping algorithm that has been presented in the literature, AMM1 (i.e. accidents are assigned to the closest section). As it was shown in Section 4.3.1 the accuracy of AMM1 reaches 81.6%, which is significantly lower than the 98.9% accuracy of AMF. Using these new locations of AMM1 the multivariate models for KS and Sl accidents were refitted using the same variable specifications of the previous KS-Sl models (see Tables 5.2 and 5.11).

Tables 5.25, 5.26, 5.27 and 5.28 present the outcomes of the link-based and the condition-based models for KS and Sl accidents (i.e. *Link-based KS-Sl AMM1 (3)* and *Condition-based KS-Sl AMM1 (20)*). The estimated coefficients of the models will not be discussed in terms of the validity and interpretability of the relationships they imply, but will only be compared to the corresponding outcomes of the models which are based on the AMF accident mapping algorithm (i.e. *Link-based KS-Sl AMF (3)* and *Condition-based KS-Sl AMF (20)*).

Table 5.25: Parameter estimates for the link-based multivariate Poisson-lognormal model for fatal and serious (KS) and slight (Sl) accidents- accident locations identified with the AMM1 (closest section) method (*Link-based KS-Sl AMM1 (3)*)

KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	-0.0379 **	0.0043	0.0002	-0.0463	-0.0450	-0.0307	-0.0293
ln(AADT)	0.1112 *	0.0646	0.0025	-0.0105	0.0092	0.2218	0.2434
Curve	-0.0662	0.0706	0.0013	-0.2049	-0.1826	0.0502	0.0724
Uphill	-0.0958	0.1276	0.0018	-0.3492	-0.3077	0.1115	0.1504
Downhill	-0.0117	0.0672	0.0009	-0.1442	-0.1229	0.0986	0.1187
Lanes	0.1821 **	0.0924	0.0026	0.0016	0.0299	0.3335	0.3620
Intercept	-0.4648 *	0.2764	0.0118	-1.0025	-0.9225	-0.0224	0.0669
Sl accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	-0.0481 **	0.0026	0.0001	-0.0532	-0.0524	-0.0436	-0.0428
ln(AADT)	0.6672 **	0.0437	0.0020	0.5785	0.5924	0.7355	0.7501
Curve	-0.0072	0.0401	0.0009	-0.0855	-0.0730	0.0588	0.0717
Uphill	0.0529	0.0661	0.0011	-0.0770	-0.0553	0.1613	0.1831
Downhill	0.0743 **	0.0376	0.0007	0.0007	0.0127	0.1364	0.1480
Lanes	0.1508 **	0.0534	0.0018	0.0467	0.0637	0.2388	0.2562
Intercept	0.2723	0.1765	0.0085	-0.0624	-0.0110	0.5805	0.6314
$\bar{D}$	12004	**statistically significant at the 95% credible interval					
p_D	721	*statistically significant at the 90% credible interval					
DIC	12724						

Table 5.26: Combined Covariance-Correlation matrix of the random effect of the *Link-based KS-Sl AMM1 (3)* model.

A	KS accidents	Sl accidents
KS accidents	0.266**	0.231**
Sl accidents	**0.903****	0.246**
B	KS accidents	Sl accidents
KS accidents	0.007**	0.006**
Sl accidents	**0.576****	0.013**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval

Table 5.27: Parameter estimates for the condition-based multivariate Poisson-lognormal model for fatal and serious (KS) and slight (Sl) accidents- accident locations identified with the AMM1 (closest section) method (*Condition-based KS-Sl AMM1 (20)*)

KS accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	0.02895 **	0.012453	0.0007123	0.00517	0.01047	0.04987	0.05296
Speed squared	-0.00027 **	0.000116	0.0000065	-0.00050	-0.00046	-0.00009	-0.00005
Volume	-0.01489 **	0.002050	0.0001058	-0.01891	-0.01824	-0.01160	-0.01067
Volume Squared	0.00003 **	0.000005	0.0000002	0.00002	0.00002	0.00004	0.00004
Speed*Volume	-0.00001	0.000024	0.0000012	-0.00006	-0.00006	0.00003	0.00003
Curve	-0.18081 **	0.063393	0.0008655	-0.30496	-0.28430	-0.07561	-0.05695
Uphill	2.04399 **	0.148110	0.0046915	1.75968	1.80401	2.29526	2.34377
Downhill	2.82774 **	0.144266	0.0047476	2.55310	2.59668	3.07094	3.12077
Lanes	-0.77462 **	0.068163	0.0009148	-0.90989	-0.88833	-0.66307	-0.64231
Intercept	-3.34391 **	0.389280	0.0213640	-4.12813	-4.01218	-2.71451	-2.57703
ln(VehHr/mile)	1	-	-	-	-	-	-
Sl accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
Speed	0.04691 **	0.009536	0.0005605	0.02821	0.03003	0.06177	0.06455
Speed squared	-0.00036 **	0.000083	0.0000048	-0.00052	-0.00050	-0.00023	-0.00021
Volume	-0.00348 **	0.001241	0.0000700	-0.00568	-0.00542	-0.00132	-0.00068
Volume Squared	0.00002 **	0.000003	0.0000001	0.00002	0.00002	0.00002	0.00003
Speed*Volume	-0.00010 **	0.000016	0.0000009	-0.00013	-0.00012	-0.00007	-0.00007
Curve	-0.19247 **	0.035716	0.0006198	-0.26185	-0.25087	-0.13373	-0.12225
Uphill	2.18968 **	0.070093	0.0024162	2.05664	2.07764	2.30832	2.33164
Downhill	2.83214 **	0.068432	0.0024936	2.70231	2.72208	2.94713	2.97010
Lanes	-0.46072 **	0.036458	0.0005357	-0.53258	-0.52085	-0.40094	-0.38979
Intercept	-3.11601 **	0.305059	0.0177031	-3.72361	-3.62815	-2.59004	-2.51469
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$	10643	**statistically significant at the 95% credible interval					
p_D	839						
DIC	11482						

Table 5.28: Combined Covariance-Correlation matrix of the random effect of the *Condition-based KS-Sl AMM1 (20)* model.

	KS accidents	Sl accidents
KS accidents	0.309**	0.26**
Sl accidents	**0.955****	0.24**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval

Comparing the significance and the signs of the coefficients of the two models (*Link-based KS-Sl AMM1 (3)* and *Link-based KS-Sl AMF (3)*) presented at Tables 5.25 and 5.2 it can be seen that the results of the two link-based models do not have significant differences. This means that the lower accuracy of accident locations did not lead in any substantial changes in the model. This result of the link-based model can be explained by the lower probability of erroneous allocation of accidents on links. Junction to junction links are typically quite long (average HATRIS link length: 3.25 miles) and thus they include multiple different road sections (average HAPMS section length: 0.46 miles) (see Figure 4.2). This fact makes unlikely the allocation of an accident to a false link even if the road section that has been selected by the AMM1 algorithm was not the correct one. Because of this attribute of the data in link-based analyses that use relatively long road links, accident location refinement techniques might not be necessary.

The coefficients of the *Condition-based KS-Sl AMM1 (20)* and *Condition-based KS-Sl AMF (20)* models (Tables 5.27 and 5.11 respectively) are generally similar but they have differences in significance tests for the dummy variable for curved road sections. More specifically, curved configuration according to the results of the *Condition-based KS-Sl AMM1 (20)* model seems to decrease KS accidents while from the results of the *Condition-based KS-Sl AMF (20)* model the impact of the curvature is not significant. This inconsistency in the results originates from the quality differences of the accident locations that were used in the two models. The geometrical pre-accident conditions are determined by the characteristics of the road section where an accident was allocated. The road sections that each accident was matched with which varies significantly between the AMM1 and AMF algorithms. These differences lead to the identification of different geometric pre-accident geometric conditions that were reflected in the estimated coefficients of the models. If the traffic measures were available to a section-level rather than a link-level probably these differences would be reflected in the coefficients of the traffic variables too. The direct comparison of the models' fit using DIC is not applicable, though, because the dependent variables have the same number of observations but different distributions. What is possible to comment about the two models is that accident location accuracy is possible to shift the outcomes of condition-based accident models. Also, taking into consideration that horizontal curves are mainly considered to be a road feature that triggers accidents (e.g. Milton and Mannering, 1998) and that the less accurate accident dataset gives the opposite result, the precision of the accident locations that are used as an input seems also to affect the validity of the modelling results. This is a clear indication that accident location accuracy is important for condition-based approaches and justifies the application of accident location refining techniques in safety analyses.

5.6 Summary

This chapter has presented the results of the models that have been developed to examine the relationships of traffic variables (i.e. speed and volume) and geometric variables (i.e. curvature, gradient, number of lanes) with accidents. Accidents were modelled all together and disaggregated by severity and collision type. All the models have been applied separately to link-based and the condition-based datasets.
The main findings of the link-based models are:

- Higher speed is related with decreased accident frequency for all accident types;
- Higher AADT is related with increased accident frequency for all accident types;
- Road links with downgrades are more likely to have higher accident frequency.

The main findings of the condition-based models are:

- Higher speed is related with increased frequency of accidents with killed or seriously injured casualties and single-vehicle accidents of all severities;
- Lower speed is related with increased frequency of multiple vehicle accident with slight injuries;

- Lower volume is related with increased frequency of single vehicle accidents;
- Higher volume is related with increased frequency of multiple vehicle accidents;
- Slight and multiple vehicle accident frequency increases on locations with curvature;
- Accident frequency increases on locations with vertical grade (especially downgrades).

The differences in the results of the link-based and the condition-based models show that accident data aggregation approaches play a significant role in the outcomes of safety models. The outcomes of the link-based models for speed are mainly counterintuitive, which proves that aggregation bias may affect drastically the reliability of these models. By contrast, condition-based models which express in more detail pre-accident conditions, provide more interpretable and thus more reliable results that are more suitable for the quantification of the impact of a potential speed limit increase.

The significance of accident location accuracy on accident modelling has been discussed in the last section of this chapter. The results show that accident models are sensitive to erroneous accident locations especially when they rely on network data that refer to relatively small road sections.

Chapter 6

Impact Estimation and Policy Implications

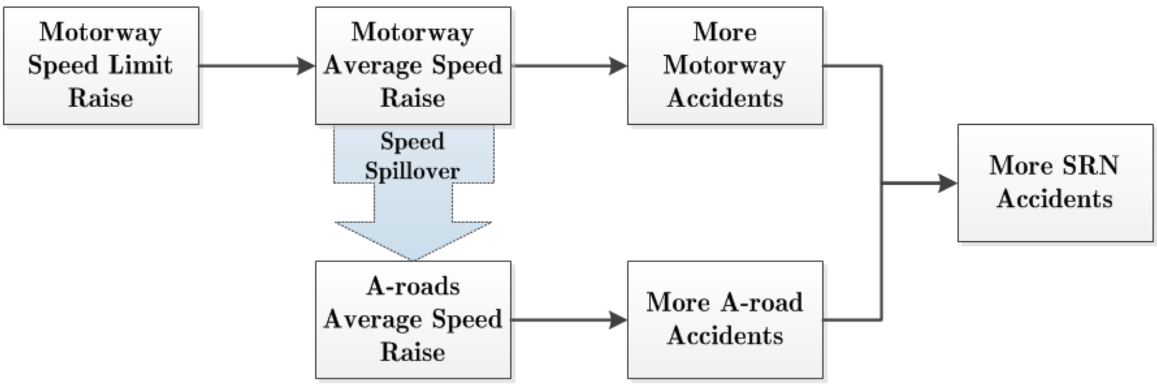
6.1 Introduction

The models that have been presented in Chapter 5 define the accident-speed relationships on the study network. Knowing how accidents are related with speed it is possible to estimate the impact of an average speed increase that could be caused by a potential speed limit increase on the UK motorway. The second section of this chapter will present the estimation of the impact of this traffic measure. The estimation is based on the speed elasticity of motorway accidents disaggregated by type and severity. The third section of this chapter will outline recommendations for accident prevention policies that are based on the outcomes of the models that have been developed in this thesis.

6.2 The Impact of a Speed Limit Increase

One of the objectives of this thesis is to evaluate the impact of a potential speed limit increase, from 70 to 80 mph, on accidents. Apart from explaining the relationship of accidents with traffic and geometry related variables, the developed models can be employed for impact estimation. Using the elasticity of accidents with respect to speed it is possible to estimate the expected changes in the number of accidents as a result of a motorway speed limit increase. For this purpose it is meaningful to employ the most detailed models for the accident-speed relationships which are the models that examine accident occurrence by collision type and severity (i.e. SV_KS-Sl and MV_KS-Sl models). As it has been explained in Section 5.4 the models that will be considered for the impact estimation are the condition-based models only.

Speed limit increases typically lead to proportional average speed changes that are believed to be related with more traffic accidents. It has been reported in the literature that this effect is often observed beyond the boundaries of the road network which had its speed limits changed. The term spillover effect expresses the tendency of the average speed of road networks with unchanged speed limits to be affected by the increase of average speeds on adjacent road networks which experienced speed limit raises (e.g. Rock, 1995; Richter et al., 2004). This effect can be attributed to drivers' attitude changes and speed acceptance that is transferred to other networks (Dutta and Noyce, 2005). This mechanism is illustrated in Figure 6.1. As a consequence, a speed limit increase on the motorway could lead to an increase in the number of accidents on the entire SRN and even beyond. However, as the extent of speed spillovers is not known, the potentially additional accidents on adjacent sections of the motorway cannot be quantified. Consequently, the estimated impact that will be presented here refers only to the accident increases on the motorway network, that is the minimum expected impact.



,

Figure 6.1: Graphical representation of the speed spillover mechanism on the SRN.

According to existing literature the average speed on a road is expected to be raised by 25% to 50% of the amount of the speed limit increase (e.g. Finch et al., 1994; Rock, 1995; Vadeby and Forsman, 2010). This means that if the speed limit of UK motorways increases from 70mph to 80mph (i.e. 10mph) the average motorway speed would be expected to increase by 2.5mph to 5mph. It is not clear how this change would affect the speed distribution of the networks. A speed limit change could cause a uniform shift to the speed distribution, or it could cause a more significant increase at higher speed conditions than at the lower ones. Considering that low speeds are normally caused by traffic congestion, the second case is more likely to be representative. Since it is not possible to predict the form of the new speed distribution though, the elasticity values that are presented here are estimated based on the expected changes on the average speed. The equation of the mean elasticity of the m^{th} variable of the k^{th} accident category is:

$$Elasticity = \frac{\partial E(y \mid x_{mk})}{\partial x_{mk}} \cdot \frac{x_{mk}}{y} \tag{6.1}$$

Table 6.1 shows the mean elasticity of accident with reference to speed and the estimated minimum and maximum percentage of increase for SV and MV motorway accidents based on the outcomes of the *Condition-based SV_KS-Sl (9)* and the *Condition-based MV_KS-Sl (18)* models respectively. As discussed, a 10 mph increase in the speed limit would result in 3.86% in average speed raise (i.e. the average 64.7mph speed would at least increase by 2.5 mph so the increase is $\frac{2.5}{64.7}$). Given that the mean elasticity of accidents with respect to speed is 2.595 for SV_KS accidents (see Table 6.1), the corresponding increase in SV_KS accidents would be at least 10.0% (i.e. $3.86 \cdot 2.595$). In a similar manner SV_Sl and MV_KS accidents would have a minimum increase of 6.14% (i.e. $3.86 \cdot 1.591$) and 3.57% (i.e. $3.86 \cdot 0.925$) respectively. The speed elasticity for the MV_Sl accidents was chosen not to be presented here. As the relationship of speed with this accident type is negative, the elasticity of speed is a negative, too. Having no evidence to support that a speed limit increase can be associated with decrease in particular types of accidents and to keep the results conservative it is considered that the number of MV accidents that lead to slight injuries will not change.

Assuming that all other variables remain the same, SV_KS are expected to increase by 10.0%-20.1% on motorways after one year of implementation of the speed limit increase. For SV_Sl this number will fluctuate from 6.1% to 12.3%. This means that after a speed limit increase there will be 73-146 more SV occurrences on the UK motorway per year. The increase of MV_KS motorway accidents will be from 3.6%-7.1% equivalent to 11-21 more MV_KS accidents. The overall predicted increase due to the anticipated average speed raise for all motorway KS and Sl accidents will reach 6.2%-12.1% (30-59 accidents) and 1.5% -2.9% (11-22 accidents) respectively. Considering the spillover effects this increase can be possibly even higher. The use of average elasticity may lead to underestimation of the impact because it does not take into account the unknown new speed distribution. However, these results provide clear evidence that a change on the 70mph current speed limit is expected to have a considerable and impact on road safety.

Table 6.1: Elasticity of speed and the minimum (mean speed increases by 2.5 mph) and maximum (mean speed increases by 5 mph) expected increase of motorway accidents by type.

Accident		Expected Accident Increase (%)		Additional Accidents	
Type	Elasticity	Min.	Max.	Min.	Max.
SV_KS	2.595	10.027	20.054	19	37
SV_Sl	1.591	6.141	12.283	54	108
MV_KS*	0.925	3.571	7.141	11	22

*Estimation based on the average volume conditions (i.e. 148 vehicles per lane)

6.3 Recommendations for Accident Prevention Policies

The findings of this project offer some new insight on the relationship of speed with accidents on highway environments that can contribute to the development of new and more efficient accident prevention policies. One of the main outcomes of the models that have been developed is that certain traffic conditions are associated with particular accident types. More specifically, accidents with fatal and serious injuries are mainly related with conditions that are characterised by higher speeds and lower volumes. Slight accidents depending on their collision type are associated with different traffic conditions. Single vehicle accidents with slight injuries tend to increase under high speed and low volume conditions, but multiple vehicle collisions with slight injuries occur at congested road sections. As a consequence, to achieve efficient accident prevention it is important to reduce the development of extreme traffic conditions on the roadway (i.e. either very high or very low average speed conditions). Appropriate use of the technological advances in traffic management can enable full control of the traffic conditions and thus it will be the key-element for future accident prevention measures.

Based on the fact that the most serious accidents are found to have a clearly proportional relationship with speed, reduction of speeding should be prioritised. Considering that a potential change of the national motorway speed limit from 70 to 80mph would encourage speeding and could lead in up to 167 more accidents per year on motorways, increasing the speed limit is probably not a rational decision. Instead, a reduction of the current number of speed limit violations would be beneficial as it could decrease the number and the severity of accidents. To achieve this, motorway speed limits should not only remain unchanged but they also need to be enforced more efficiently. Smart motorways (or Active Traffic Management systems) can play a significant role to that. The smart motorway is an approach for dynamic control and management of the traffic based on the available network capacity, using real-time data and predictive models (Kurzhanskiy and Varaiya, 2010). One of the main techniques of smart motorways is the use of variable speed limits which are enforced through the installation of digital cameras. In this way, speed limits can be adjusted appropriately according to the external conditions (i.e. adverse weather, road works etc.) and at the same time speed violations can be minimised. In 2005, smart motorway systems have been introduced in parts of the UK motorway and ever since the smart network of the country is fast expanding (Highways Agency, 2010; Highways England, 2014). Considering the benefits, the adoption of ATM systems in the entire SRN would certainly lead in significant decreases of serious accidents.

Congested traffic has been found to be related with multiple vehicle slight accidents. Although their severity is not high, these accidents are approximately the two thirds of all accidents on the SRN. Apart from the property damages, the disruption that is related with multiple vehicle accidents under congested conditions is very significant (e.g. queues, delays, secondary accidents). Consequently, it is particularly meaningful to develop the right policies to decrease accidents related with congestion and their impact. The smart motorway is currently one of the most efficient methods for congestion reduction. For example hard shoulder use as a normal running lane during peak times is one useful measure for immediate increase of the capacity of existing road networks. The use of variable warning signs that can provide useful information about the existence of queues or accidents downstream and variable speed limits can improve the levels of mobility on the road networks. Other ways for congestion reduction include car sharing schemes, improvement of the efficiency, attractiveness and coverage of public transport and promotion of alternative modes of freight (e.g. rail freight). Last but not least are the emerging technologies on vehicular communication (VC) that enable vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication. VC is one of the most promising technologies in transport and is expected to improve simultaneously road safety and mobility (Papadimitratos et al., 2009).

Based on the outcomes for the road geometric features that are most related with traffic accidents, some additional countermeasures can be developed. Curvature is one geometric characteristic that has been related with more slight and multiple vehicle accidents. Moreover, roads with negative grades have been found by most of the models that have been presented here (including the link-based models) to be related with higher accident frequency. These findings should be considered for improving the design characteristics of future road networks. Also, to improve the safety levels of the existing roads it is suggested to use appropriate warning signs upstream of locations with sharp curves and that will remind the drivers to avoid overtaking or other actions that can be affected by decreased visibility. Warning signs should be also used before steep downgrades. The location of these signs in relation to the hazard-prone road sections is critical; there should be sufficient distance so as the drivers to have enough time to adjust their speed and movements as appropriate. In addition to this, variable speed limits should be coordinated in accordance with the geometric characteristics on the network. For instance, variable speed limits should be lower on sections that have regular and steep negative grades where breaking distances are longer.

6.4 Summary

This chapter presented the impact of the results of the models that have been developed in this thesis. The impact of a potential speed limit increase on the UK motorway as it was estimated by the average speed elasticity of accidents is anything but negligible. After one year of implementation of the measure, there will be up to 167 more accidents on the UK motorway (59 of which will be fatal or serious) and considering potential speed spillover on the SRN, this number may be higher. These findings show clearly that a speed limit increase would have a negative impact on road safety.

To eliminate the number of accidents effectively it is necessary to recognise and prevent the traffic conditions that are mostly related with accidents. Condition-based modelling provides new insight on the traffic conditions that should be avoided. New technologies that enable traffic control and coordination should be the key feature of future accident prevention policies. Smart motorways are currently the most advanced and immediately applicable approach that can offer significant improvements on both mobility levels and road safety.

Chapter 7

Conclusions

7.1 Summary

Traffic accidents are one of the most serious problems in road transport as they cause serious traffic delays, congestion and property damage but more importantly they are linked with road traffic injuries. Driving with excessive speed is believed to be a dominant accident contributory factor, especially for high-speed road environments such as motorways. Speed limit increases are related with higher average speeds that can potentially lead to more accidents on the network. To quantify the impact of a potential speed limit increase first, it is necessary to define the current relationship of speed with accidents on a road network.

Current literature includes a number of inconsistent views on the accident-speed relationships. This could be due to methodological limitations that may affect the results of statistical models. Existing accident analyses usually examine the role of accident contributory factors employing advanced statistical techniques on accident datasets that are aggregated according to the link-based approach. This means that accident occurrences per road link (i.e. junction to junction section) are modelled against independent variables that represent the dominant conditions on the link over the study period. Link-based approaches have the benefit of simplicity but they are connected with aggregation bias issues that are likely to affect the accuracy of safety models' outcomes.

This research attempts to define accident speed-relationships on the UK motorway introducing a new accident aggregation method: the condition-based approach. In the condition-based approach accidents are not grouped according to spatial criteria, but instead based on the similarities of the traffic and geometric conditions on the roadway just before the accidents occurred. This approach allows for a more accurate representation of the pre-accident conditions and overcomes aggregation bias issues which are related with the link-based approaches. The implementation of the condition-based method, though, demands the identification of detailed pre-accident traffic and geometric conditions for each accident individually which increases the complexity of the data pre-processing.

The accidents in the analysis were all the accidents that occurred during 2012 on the SRN of England which comprises of all the motorways and some of the most important A-roads of the country. The traffic conditions on the network were represented using 15-minute traffic measurements and the characteristics of the road configuration were available in 10-metre intervals. To correctly allocate accidents on road links (for the link-based approach) and to identify the exact pre-accident traffic and geometric conditions (for the condition-based approach) the reported accident locations were refined using an accident mapping algorithm. This algorithm, that was based on a transformed map-matching technique and Fuzzy Logic, was developed exclusively for the study network and provided approximately 98.8% accurate accident locations.

The relationships of accidents with speed and other traffic and geometric variables were developed applying full Bayesian multivariate models which enable simultaneous modelling of accidents disaggregated by severity (i.e. fatal, serious or slight) and/or type of the collision (i.e. single-vehicle and multiple-vehicle). Parameter estimation was done with the MCMC method. To examine the impact of accident aggregation approaches on safety analyses, all the models were applied separately on link-based and condition-based datasets that were developed using identical accident, traffic and geometrical data.

The results of the link-based models showed that speed has a negative relationship with all accidents in contrast to AADT that has a positive relationship. The results especially for speed are counterintuitive and at most cases very different from the corresponding results of the condition-based models. More specifically, in condition-based models speed was found to have a positive relationship with fatal, serious and single-vehicle accidents. The latter accident type was also found to be related with very low volumes indicating that these accidents occur more frequently at low-density traffic conditions. On the contrary, slight, multiple-vehicle accidents were found to be more at lower speed conditions and under congested traffic. The outcomes of the condition-based models for all accident types are interpretable and some are also in line with existing literature. These results confirmed that particular traffic conditions are indeed associated with different accident types.

The significant differences between the results of the link-based and the condition-based models are probably due to the aggregation bias that is by default associated with the first approach. This finding revealed that the accident aggregation method is a critical element of accident analyses that, although it has been overlooked by previous studies, plays an important role on the outcomes. Considering that the condition-based models offer a more detailed representation of the accident-related circumstances and that their outcomes are more explainable, these approaches can be considered as more reliable. In addition, another methodological implication of this study was that the precision of the accident locations that are employed in accident analyses is related with the validity of their outcomes, especially for condition-based models.

The coefficients of the models provided some new insight on the relationships of some accident contributory factors with accidents which can lead to the development of improved accident prevention policies in the future. Estimating the mean speed elasticity of accidents from the condition-based models it was found that a speed limit increase on the UK motorway from 70 to 80 mph could compromise the safety levels on the network. During the first year of implementation of this measure up to 12.1% more fatal or serious and 2.9% more slight motorway accidents can be expected. To enhance mobility without increasing the already high number of accidents, instead of increasing the speed limits, it is suggested to promote the optimal use of existing and forthcoming technologies in road transport that will enable improved traffic control and management.

7.2 Contribution to Knowledge

This work has produced new qualitative and methodological outcomes that are useful to be considered for future analyses. The main contributions to knowledge of this research are:

1. Speed-accident relationships for the British motorway network

This research has examined extensively the role of speed in traffic accidents in the UK and the findings add to the debate of the current literature on this issue. The outcomes of this analysis increase the understanding about the impact of speed on different accident types and severities which is a relatively unexplored topic in the literature. Specifically, higher speeds were found to have a positive relationship with all fatal and serious accidents. Thanks to the interaction term that was used in some of the models it was also found that it is not speed per se, but the combination of high speeds with extremely high or low volumes that is related with more fatal or serious accidents. The findings for single-vehicle accidents were slightly different: single vehicle accidents independent of the severity of their outcomes are more frequent at higher speed and lower volume conditions. In other words, single vehicle are the most speed-related collisions as they tend to occur at low density conditions when speeds are by definition high. Multiple vehicle accidents with slight injuries though, were found to be mainly related with low speeds and high volumes. This means that these accidents can be mainly attributed to congestion rather than speed.

These findings show that the question whether high speeds are responsible for more accidents does not have a binary answer. Instead, it has been confirmed that different traffic patterns are associated with particular types of accidents. Some of them are related with high speeds and some are not. This is a balanced and explainable outcome that helps in understanding accident mechanisms in more depth and can lead to the development of better and targeted preventive measures in the future.

2. Significance of accident data aggregation in safety analyses

The most significant methodological implication of this study is related with accident data aggregation.It has been shown that the accident data aggregation approach that is applied can change dramatically the results of safety models. Until now, the majority of researchers focused their efforts in developing more sophisticated models from a statistical perspective. This trend has definitely upgraded accident analyses and has lead in new and interesting findings. However, it is a fact that regardless how advanced a statistical model is, the quality of its input data is also important. This work stressed that the conventional, link-based data aggregation approaches are likely to fail in representing the traffic conditions that are actually related with accidents. As an alternative to this method the condition-based aggregation approach has been developed. The condition-based approach provides a new non-spatial accident aggregation framework that overcomes aggregation bias problems and succeeds in representing pre-accident conditions more accurately.

The link-based approach that is related with aggregation bias produced many counterintuitive results, while the condition-based approach that represents pre-accident conditions provided more interpretable results. The differences between the outcomes of the two methods showed clearly that the accident aggregation approach in safety analyses matters as it can entirely change their results. Accident data aggregation approaches, that were so far been overlooked by researchers, might in the future become a key element for more comprehensive accident analyses.

3. The role of accident location accuracy in safety analyses

Another methodological aspect of accident analyses that has been highlighted in this work is the role of accident locations in safety modelling. Locations are crucial for accident analyses as they are the attribute that determines how accidents will be matched with specific observations of the explanatory variables. Nevertheless, the effect of accident location accuracy on the modelling outcomes is rather understudied. This research has shown that raw accident locations are very likely to include errors. To that end, a new and transferable method for correcting accident locations has been developed.

From modelling accident datasets with different location data (i.e. data from the developed method and from simpler methods that demand minimal pre-processing) it was found that accident locations' accuracy is possible to change the coefficient estimations. This is mainly true for analyses that use network data that consist of relatively small segments, where the probability of erroneous accident allocation is higher. Consequently, it has been shown that analysing accident data with inaccurate accident locations is likely to compromise the validity of the modelling outcomes.

7.3 Study Limitations

This study includes several data and methodological limitations. The most important of these limitations are outlined below:

- **Accident time inaccuracies:** The exact time when an accident happened it is not known and so the reported accident time was used for identifying the traffic conditions prior of accidents. However, the reported accident time is likely to be significantly different from the actual accident time which means that the traffic measurements that were used for some of the examined accidents might represent the traffic conditions after the occurrence which are typically characterised by lower speeds.

- **Accidents' underreporting:** The accidents that were analysed are all the reported accidents that had at least one injured casualty. Property damage and slight-injury accidents that have not been reported were excluded from the analysis. The number of these accidents it is very high and consequently their omission might have affected significantly the outputs of this analysis.

- **Traffic data aggregation:** Traffic data were provided in 15-minute averages for all the lanes of relatively long road sections. These data are not detailed enough to provide information about the exact conditions prior of accidents.

- **Geometric conditions:** The road section that was considered for determining the road configuration just before an accident was set to be equal with the average stopping distance (estimated using the average speed). This is assumption is apparently not always correct. Also, geometric variables in both the Link-based and the Condition-based models were represented roughly with the use of dummy variables. This representation is not complete and this is possibly reflected in the models' estimations.

- **Combination of motorways with A-roads:** Although all the roads of the SRN include routes with major commercial and social significance for the country, they are roads that have different speed limits, traffic characteristics, geometry, capacity and construction quality. Due to these variations, accident generation mechanisms on these roads might be different too. Analysing all SRN accidents together could have affected the results of the models.

- **Omitted variables:** The models that have been developed did not control for a number of potentially important accident contributory factors such as speed variance, weather and light conditions, time of the day, pavement condition/wetness, left/right curve and many others. The exclusion of these unobserved variables have possibly lead to erroneous estimations for the variables that were included (i.e. omitted-variable bias).

- **Grouping continuous traffic data into discrete scenarios:** To form the condition-based scenarios a range of (similar) traffic measurements were grouped together and represented in the model by one value that was the median of the range. This data aggregation was necessary to develop the regression model but it might compromise the accurate representation of the traffic conditions (especially for volume that had only four distinct categories per speed scenario).

- **Spatial independence of condition-based models:** The accidents of every scenario of the condition-based approach were assumed not to have any spatial relationship, which might not be true.

- **Mean elasticity:** The use of mean elasticity of accidents can provide only a crude estimation of the impact of a potential speed limit increase. As the new speed distribution is unknown this method could either overestimate or underestimate the new number of accidents. Additionally, the estimations assume that all other factors remain the same which is probably a strong assumption as a speed limit increase might be followed by other changes (e.g. stricter enforcement, higher volumes etc.).

7.4 Extensions and Suggestions for Future Research

The condition-based approach that has been presented in this thesis is a new and promising accident data aggregation approach that can contribute to the development of more accurate accident models. The method is flexible and transferable to other study areas and network environments. The use of condition-based approaches can increase the insight about accident triggering factors by indicating hazard-prone traffic conditions that should be avoided. Considering the aforementioned limitations of this study, there are several improvements that can be done in the future towards this direction.

Condition-based models rely on the quality of their input data. Improved data mean more accurate representation of the pre-accident conditions and thus more valid results. In the future, condition-based analyses should employ traffic measurements with higher spatial and temporal resolution that can be provided from fixed and mobile sensors or loop detectors. To ensure that the traffic and geometric conditions just before an accident are truly accurate, more disaggregated traffic data should be combined with more reliable accident reports. Accident reports in the future should provide precise accident time and the location. Moreover, additional information that will describe in detail the circumstances under which a collision occurred such as traffic conditions upstream and downstream, road configuration, weather and other would be extremely useful as it would facilitate the application of condition-based models. Future research should be focused on how new technologies can contribute to the development of integrated accident data collection methods.

In the future, condition-based models should also incorporate more variables that are likely to be linked with accident occurrences on the network so as to fully understand the unwanted conditions for road safety. Using the outcomes of these models in combination with traffic forecasting models it would be possible to explore alternative approaches for optimal coordination of the smart motorway variable speed limits.

Further research should also investigate the optimal number of pre-accident scenarios that should be developed with respect to the number of accidents that will be analysed. Moreover, the potential of using condition-based approaches without the use of discrete condition scenarios, but continuous observations so as to avoid data coarsening, should be also explored.

Finally, the assessment of link-based and condition-based approaches should continue beyond this study. Instead of employing condition-based as a substitute of link-based methods, it would be interesting to understand in more depth the strengths and weaknesses of these two approaches. Through this, it would be possible to answer whether and how these approaches could work complementary of each other towards the quantification of accident risk from different perspectives.

Bibliography

Aarts, L. and Van Schagen, I. (2006), 'Driving speed and the risk of road crashes: A review', *Accident Analysis & Prevention Prevention* 38(2), 215–224.

AASHTO (2010), 'Highway Safety Manual,American Association of state highway and Transportation Officials', *Washington, DC* .

Abdel-Aty, M. and Pande, A. (2005), 'Identifying crash propensity using specific traffic speed conditions.', *Journal of Safety Research* 36(1), 97–108.

Abdel-Aty, M. and Radwan, A. E. (2000), 'Modeling traffic accident occurrence and involvement.', *Accident Analysis & Prevention* 32(5), 633–42.

Abdel-aty, M., Uddin, N. and Pande, A. (2005), 'Split Models for Predicting Multivehicle Crashes During High-Speed and Low-Speed', *Transportation Research Record: Journal of the Transportation Research Board* 1908, 51–58.

Aguero-Valverde, J. (2013), 'Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica', *Accident Analysis & Prevention* 59, 365–373.

Aguero-Valverde, J. and Jovanis, P. P. (2008), 'Analysis of road crash frequency with spatial models', *Transportation Research Record: Journal of the Transportation Research Board* 2061(1), 55–63.

Aguero-Valverde, J. and Jovanis, P. P. (2009), 'Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking', *Transportation Research Record: Journal of the Transportation Research Board* 2136, 82–91.

Aljanahi, A. A., Rhodes, A. and Metcalfe, A. V. (1999), 'Speed, speed limits and road traffic accidents under free flow conditions.', *Accident Analysis & Prevention* 31(1-2), 161–8.

Anastasopoulos, P. C. and Mannering, F. L. (2009), 'A note on modeling vehicle accident frequencies with random-parameters count models', *Accident Analysis & Prevention* 41(1), 153–9.

Austin, K. (1995), 'The Identification of Mistakes in Road Accident Records: Part 1, Locational Variables', *Accident Analysis & Prevention* 27(2), 261–276.

Bartlett, J. E., Kotrlik, J. W. and Higgins, C. C. (2001), 'Organizational Research : Determining Appropriate Sample Size in Survey Research', *Information Technology, Learning and Performance Journal* 19(1), 43–50.

Barua, S., El-Basyouny, K. and Islam, M. T. (2014), 'A Full Bayesian multivariate count data model of collision severity with spatial correlation', *Analytic Methods in Accident Research* 3-4, 28–43.

Baruya, A. (1998*a*), MASTER: Speed-accident relationship on European roads, Technical report, Transport Research Laboratory, UK.

Baruya, A. (1998*b*), Speed-accident relationships on European roads, *in* '9th International Conference Road Safety in Europe', Vol. 1, Bergisch Gladbach, Germany.

Baruya, A. and Finch, D. J. (1994), Investigation of traffic speeds and accidents on urban roads, *in* 'Traffic Management and Road Safety. Proceedings of Seminar J held at the PTRC European Transport Forum,September 12-16, Volume P381', Warwick, UK.

Baum, H. M., Wells, J. K. and Lund, A. K. (1991), 'The Fatality Consequences of the 65 mph', *Journal of Safety Research* 22(4), 171–177.

Bernstein, D. and Kornhauser, A. (1996), 'An introduction to map-matching for personal navigation assistants.'.
URL: *http://www.njtude.org/reports/ mapmatchintro.pdf*

Bham, G. H., Javvadi, B. S. and Manepalli, U. R. R. (2012), 'Multinomial Logistic Regression Model for Single-Vehicle and Multivehicle Collisions on Urban U.S. Highways in Arkansas', *Journal of Transportation Engineering* 138(June), 786–797.

Bíl, M., Andrášik, R. and Janoška, Z. (2013), 'Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation.', *Accident Analysis & Prevention* 55, 265–73.

Black, J., Hashimzade, N. and Myles, G. (2009), *A dictionary of economics*, Oxford University Press.

Brown, D. B., Maghsoodloo, S. and McArdle, M. E. (1990), 'The safety impact of the 65 mph speed limit: A case study using Alabama accident records', *Journal of Safety Research* 21(4), 125–139.

Brown, I. D. (1982), 'Exposure and experience are a confounded nuisance in research on driver behaviour', *Accident Analysis & Prevention* 14(5), 345–352.

Burns, S., Miranda-moreno, L., Stipancic, J., Saunier, N. and Ismail, K. (2013), 'An Accessible and Practical Geocoding Method for Traffic Collision Record Mapping : A Quebec Case Study', *Transportation Research Record* .

Caliendo, C., Guida, M. and Parisi, A. (2007), 'A crash-prediction model for multilane roads', *Accident Analysis & Prevention* 39(4), 657–670.

Cameron, A. C. and Trivedi, P. K. (2013), *Regression Analysis of Count Data*, Econometric Society Monographs, Cambridge University Press.

Ceder, A. (1982), 'Relationships Between Road Accidents and Hourly Traffic Flow - 2. Probabilistic Approach', *Accident Analysis & Prevention* 14(1), 35–44.

Ceder, A. and Livneh, M. (1982), 'Relationships between road accidents and hourly traffic flowI: analyses and interpretation', *Accident Analysis & Prevention* 14(1), 19–34.

Chang, G.-L., Chen, C.-H. and Carter, E. C. (1993), 'Intervention analysis for the impacts of the 65 mph speed limit on rural interstate highway fatalities', *Journal of Safety Research* 24(1), 33–53.

Chang, L.-Y. (2005), 'Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network', *Safety Science* 43(8), 541–557.

Chang, L.-Y. and Wang, H.-W. (2006), 'Analysis of traffic injury severity: an application of non-parametric classification tree techniques.', *Accident Analysis & Prevention* 38(5), 1019–27.

Chipman, M. L., MacGregor, C. G., Smiley, A. M. and Lee-Gosselin, M. (1992), 'Time vs. distance as measures of exposure in driving surveys', *Accident Analysis & Prevention* 24(6), 679–684.

Chorley, M. (2013), 'Motorway speed limit could yet be increased to 80mph as minister reveals government division over the plan'.
URL: *http://www.dailymail.co.uk/news/article-2438551/Motorway-speed-limit-increased-80mph-minister-reveals-government-division-plan.html*

Cirillo, J. A. (1968), 'Interstate System Accident Research Study II, Interim Report II', *Public Roads* 35(3).

Clark, W. and Avery, K. (1976), 'The effects of data aggregation in statistical analysis', *Geographical Analysis* 8(4), 428–438.

Clarke, D. D., Ward, P., Bartle, C. and Truman, W. (2010), 'Killer crashes: fatal road traffic accidents in the UK', *Accident Analysis & Prevention* 42(2), 764–70.

Davis, G. A. (2002), 'Is the claim that 'variance kills' an ecological fallacy?', *Accident Analysis & Prevention* 34(3), 343–346.

Davis, G. A. (2004), 'Possible aggregation biases in road safety research and a mechanism approach to accident modeling', *Accident Analysis & Prevention* 36(6), 1119–1127.

De Pauw, E., Daniels, S., Thierie, M. and Brijs, T. (2014), 'Safety effects of reducing the speed limit from 90km/h to 70km/h.', *Accident Analysis & Prevention* 62, 426–31.

Deka, L. and Quddus, M. (2014), 'Network-level accident-mapping: Distance based pattern matching using artificial neural network', *Accident Analysis & Prevention* 65, 105–13.

Department for Transport (2006), Setting Local Speed Limits, Technical Report August, Department for Transport, London, England.

Department for Transport (2007), *The Official Highway Code*, 15 edn.

Department for Transport (2011*a*), 'Government proposes 80MPH motorway speed limit - Announcements'.

URL: *https://www.gov.uk/government/news/government-proposes-80mph-motorway-speed-limit*

Department for Transport (2011*b*), Road Network Policy Consultation, Technical report, London, England.

Department for Transport (2011*c*), 'STATS19 road accident injury statistics report form'. **URL:** *https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/230590/stats19.pdf*

Department for Transport (2012*a*), Managed Motorways All lans running, Concept of Operations (to accompany IAN 161/12), Technical report, Department for Transport, London, England.

Department for Transport (2012*b*), Reported Road Casualties in Great Britain: 2012 Annual Report: A Valuation of Road Accidents and Casualties in Great Britain in 2012, Technical report, London, England.

Department for Transport (2014), Reported Road Casualties Great Britain: 2013 Annual Report, Technical report, Deaprtment for Transport, London, UK.

Department for Transport (2015*a*), Free Flow Vehicle Speed Statisitics: Great Britain 2014, Technical report, London, England.

Department for Transport (2015*b*), Reported Road Casualties in Great Britain : Main Results 2014, Technical report, London, England.

Department for Transport (2015*c*), Road Lengths in Great Britain 2014, Technical report, London, England.

Dutta, A. and Noyce, D. (2005), Impact of raising speed limits on traffic safety, Technical report, Traffic Operations & Safety Laboratory, University of Wisconsin.

Dutta, A., Parker, S., Qin, X., Qiu, Z. and Noyce, D. A. (2007), A System for digitising Winsconsin crash location information, *in* '86th Annual Meeting of the Transportation Research Board', Washington, DC.

El-Basyouny, K. and Sayed, T. (2009*a*), 'Accident prediction models with random corridor parameters', *Accident Analysis & Prevention* 41(5), 1118–1123.

El-Basyouny, K. and Sayed, T. (2009*b*), 'Collision prediction models using multivariate Poisson-lognormal regression.', *Accident Analysis & Prevention* 41(4), 820–8.

Elvik, R. (2009), The Power Model of the relationship between speed and road safety: update and new analyses, Technical report, Institute of Transport Economics.

Elvik, R. (2010), 'A restatement of the case for speed limits', *Transport Policy* 17(3), 196–204.

Elvik, R. (2013), 'A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims.', *Accident Analysis & Prevention* 50, 854–60.

Elvik, R., Christensen, P. and Amundsen, A. (2004), Speed and road accidents: An evaluation of the Power Model, Technical report, Institute of Transport Economics.

Fellegi, I. P. and Sunter, A. B. (1969), 'A Theory for Record Linkage', *Journal of the American Statistical Association* 40, 1183–1210.

Field, A. (2009), *Discovering statistics using SPSS*, Sage publications.

Fildes, B. and Lee, S. (1993), The Speed review: Road Environment, Behaviour, Speed Limits, Enforcement and Crashes, Technical report, Monash University Accident Research Centre, Victoria.

Fildes, B. N., Rumbold, G. and Leening, A. (1991), Speed Behaviour and Drivers' Attitude to Speeding, Technical report, Monash University, Accident Research Centre, Victoria,Austalia.

Finch, D. J., Kompfner, P., Lockwood, C. R. and Maycock, G. (1994), Project report 58-Speed, Speed Limits and Accidents, Technical report, Safety Resource Centre, Transpot Research Laboratory, Crowthorne, England.

Frantzeskakis, J. M. and Iordanis, D. I. (1987), 'Volume-to-capacity ratio and traffic accidents on interurban four-lane highways in Greece', *Transportation Research Record* 1112.

Freedman, M. and Williams, A. F. (1992), 'Speeds Associated with 55- mph and 65-mph Speed Limits in NortheasternStates', *ITE Journal* 62(2), 17–22.

Frith, W. J. and Patterson, T. L. (2001), Speed Variantion, Absolute Speed and their Contribution to Safety, With Special Reference to the Work of Solomon, *in* 'IPENZ Transportation Group', pp. 1–8.

Garber, N. and Ehrhart, A. (2000), 'Effect of Speed, Flow, and Geometric Characteristics on Crash Frequency for Two-Lane Highways', *Transportation Research Record* 1717, 76–83.

Garber, N. J. and Gadiraju, R. (1989), 'Factors Affecting Speed Variance and Its Influence on Accidents.', *Transportation Research Record: Journal of the Transportation Research Board* 1213, 64–71.

Garber, S. and Grahman, J. D. (1990), 'The effects of the new 65 mile-per-hour speed limit on rural highway fatalities: a state-by-state analysis.', *Accident Analysis & Prevention* 22(2), 137–49.

Geedipally, S. R. and Lord, D. (2010), 'Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models', *Accident Analysis & Prevention* 42(4), 1273–1282.

Geedipally, S. R., Patil, S. and Lord, D. (2010), 'Examination of Methods to Estimate Crash Counts by Collision Type', *Transportation Research Record: Journal of the Transportation Research Board* 2165, 12–20.

Geurts, K., Wets, G., Brijs, T., Vanhoof, K. and Karlis, D. (2006), 'Ranking and selecting dangerous crash locations: correcting for the number of passengers and Bayesian ranking plots.', *Journal of Safety Research* 37(1), 83–91.

Gitelman, V., Doveh, E., Carmel, R. and Pesahov, F. (2014), The Relationship between road accidents and infrastructure characteristics of low-volume roads in Israel, *in* 'International Conference on Traffic and Transport Engineering-Belgrade'.

Godwin, S. (1984), 'International experience with speed limits during and prior to the energy crisis of 1973-74.', *Transportation Planning and Technology* 9, 25–36.

Graves, P., Lee, D. and Sexton, R. (1993), 'Speed variance, eforcement and the optimal speed limit', *Economics Letters* 42, 237–243.

Guo, F., Wang, X. and Abdel-Aty, M. a. (2010), 'Modeling signalized intersection safety with corridor-level spatial correlations.', *Accident Analysis & Prevention* 42(1), 84–92.

Gwynn, D. W. (1967), 'Relationship of accident rates and accident involvements with hourly volumes', *Traffic Quarterly* 21(3).

Hagenzieker, M. P., Commandeur, J. J. and Bijleveld, F. D. (2014), 'The history of road safety research: A quantitative approach', *Transportation Research Part F: Traffic Psychology and Behaviour* 25, 150–162.

Hale, A. E. (1990), 'Safety and speed. A systems view of determinants and control measures', *IATSS Research* 14(1).

Hauer, E. (1971), 'Accidents, overtaking and speed control', *Accident Analysis & Prevention* 3(1), 1–13.

Hauer, E. (2009), 'Speed and Safety', *Transportation Research Record* 2103, 10–17.

Hauer, E. and Bonneson, J. (2006), An Empirical Examination of the Relationship Between Speed and Road Accidents, Technical Report 38, Unpublished manuscript dated March 5, 2006.

Haynes, R., Jones, A., Kennedy, V., Harvey, I. and Jewell, T. (2007), 'District variations in road curvature in England and Wales and their association with road-traffic crashes', *Environment and Planning A* 39(5), 1222 – 1237.

Haynes, R., Lake, I. R., Kingham, S., Sabel, C. E., Pearce, J. and Barnett, R. (2008), 'The influence of road curvature on fatal crashes in New Zealand', *Accident Analysis & Prevention* 40(3), 843–850.

Hels, T. (2012), Speed and Crash Risks : Analysis of Recent Monitoring in IRTAD Countries, *in* 'Transportation Research Board 91st Annual Meeting'.

Highways Agency (2002), Highway Link Design, *in* 'Design Manual for Roads and Bridges', Department for Transport, UK.

Highways Agency (2006), Controlled Motorways, M25, Technical report, Department for Transport, London, England.

Highways Agency (2008), Pavement Design & Maintenance, *in* 'Design Manual for Roads and Bridges', Department for Transport, UK.

Highways Agency (2010), Smart motorways, Technical report.

Highways Agency (2011), HATRIS JTDB Reference Manual, Technical report.
URL: *https://jtdb.hatris.co.uk/\JTDB Reference manual.pdf*

Highways Agency (2015), 'Our Road Newtork'.
URL: *http://webarchive.nationalarchives.gov.uk/20140603112028/http://www.highways.gov.uk/our-road-network/our-network/*

Highways England (2014), Strategic Bussiness Plan 2015-2020, Technical report.

Hilbe, J. (2011), *Negative Binomial Regression*, Cambridge University Press.

Hossain, M. and Muromachi, Y. (2013), 'Understanding crash mechanism on urban expressways using high-resolution traffic data.', *Accident Analysis & Prevention* 57, 17–29.

Houston, D. J. (1999), 'Implications of the 65-MPH Speed Limit for Traffic Safety', *Evaluation Review* 23(3), 304–315.

Imprialou, M.-I., Quddus, M. and Pitfield, D. (2015), Multilevel Logistic Regression Modeling for Crash Mapping in Metropolitan Areas, *in* 'Transportation Research Board 94th Annual Meeting', Washington DC.

ITF, Eurostat and UNECE (2009), *Illustrted Glossary for Transport Statistics*, Luxembourg: Publications Office of the European Union.

Ivan, J. N. (2004), 'New Approach for Including Traffic Volumes in Crash Rate Analysis and Forecasting', *Transportation Research Record* 1897, 134–141.

Ivan, J. N., Pasupathy, R. K. and Ossenbruggen, P. J. (1999), 'Differences in causality factors for single and multi-vehicle crashes on two-lane roads', *Accident Analysis & Prevention* 31(6), 695–704.

Ivan, J. N., Wang, C. and Bernardo, N. R. (2000), 'Explaining two-lane highway crash rates using land use and hourly exposure.', *Accident Analysis & Prevention* 32(6), 787–95.

Joksch, H. C. (1993), 'Velocity change and fatality risk in a crashA rule of thumb', *Accident Analysis & Prevention* 25(1), 103–104.

Jonah, B. A. (1986), 'Accident risk and risk-taking behaviour among young drivers.', *Accident Analysis & Prevention* 18(4), 255–71.

Jones, B., Janssen, L. and Mannering, F. (1991), 'Analysis of the frequency and duration of freeway accidents in Seattle', *Accident Analysis & Prevention* 23(4), 239–255.

Jovanis, P. P. and Chang, H.-L. (1986), 'Modeling the relationship of accidents to miles traveled', *Transportation Research Record* 1068, 42–51.

Kam, B. H. (2003), 'A disaggregate approach to crash rate analysis', *Accident Analysis & Prevention* 35(5), 693–709.

Karlaftis, M. G. and Golias, I. (2002), 'Effects of road geometry and traffic volumes on rural roadway accident rates.', *Accident Analysis & Prevention* 34(3), 357–65.

Keep, M. (2013), Reported Road Accident Statistics, Technical report.

Kim, D. G., Washington, S. and Oh, J. (2006), 'Modeling Crash Types: New Insights into the Effects of Covariates on Crashes at Rural Intersections', *Journal of Transportation Engineering* 132, 282–292.

Kim, J. K., Ulfarsson, G. F., Kim, S. and Shankar, V. N. (2013), 'Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender', *Accident Analysis & Prevention* 50, 1073–1081.

Kloeden, C., McLean, A. and Glonek, G. (2002), Reanalysis of Travelling Speed and the Risk of Crash Involvement in Adelaide South Australia, Technical report, Road Accident Research Unit, The University of Adelaide.

Kloeden, C., Mclean, A., Moore, V. and Ponte, G. (1997), Travelling Speed and the Risk of Crash Involvement Volume 1 - Findings, Technical report, South Australia.

Kockelman, K. M. and Ma, J. (2007), 'Freeway Speeds and Speed Variations Preceding Crashes , Within and Across Lanes', *Transportation Research Forum* 46(1), 43–61.

Koike, H., Morimoto, A., Hanzawa, Y. and Shiraishi, N. (2000), 'Development of Hazard Map Using GIS to Reduce Traffic Accidents', *Computing in Civil Engineering* pp. 217–224.

Kononov, J., Bailey, B. and Allery, B. K. (2008), 'Relationships Between Safety and Both Congestion and Number of Lanes on Urban Freeways', *Transportation Research Record: Journal of the Transportation Research Board* (2083), 26–39.

Kurzhanskiy, A. A. and Varaiya, P. (2010), 'Active traffic management on road networks: a macroscopic approach', *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 368(1928), 4607–4626.

Lang, S. W., Waller, P. F. and Shope, J. T. (1996), 'Adolescent driving: Characteristics associated with single-vehicle and injury crashes', *Journal of Safety Research* 27(4), 241–257.

Lave, C. (1985), 'Speeding , Coordination , and the 55 MPH Limit', *American Economic Association* 75(5), 1159–1164.

Lave, C. and Elias, P. (1994), 'Did the 65 mph speed limit save lives?', *Accident Analysis & Prevention* 26(1), 49–62.

Levine, N., Kim, K. E. and Nitz, L. H. (1995), 'Spatial Analysis of Honolulu Motor Vehicle Crashes: I, Spatial Patterns', *Accident Analysis & Prevention* 27(5), 663–674.

Li, H., Graham, D. J. and Majumdar, A. (2013), 'The impacts of speed cameras on road accidents: An application of propensity score matching methods', *Accident Analysis & Prevention* 60, 148–157.

Li, H., Graham, D. J. and Majumdar, A. (2015), 'Effects of changes in road network characteristics on road casualties: An application of full Bayes models using panel data', *Safety Science* 72, 283–292.

Loo, B. P. Y. (2006), 'Validating crash locations for quantitative spatial analysis : A GIS-based approach', *Accident Analysis & Prevention* 38, 879–886.

Lord, D. (2002), 'Issues Related to the Application of Accident Prediction Models for Computation of Accident Risk on Transportation Networks', *Transportation Research Record* 1784(1), 17–26.

Lord, D., Manar, A. and Vizioli, A. (2005), 'Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments.', *Accident Analysis & Prevention* 37(1), 185–99.

Lord, D. and Mannering, F. (2010), 'The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives', *Transportation Research Part A: Policy and Practice* 44(5), 291–305.

Lord, D. and Miranda-Moreno, L. F. (2008), 'Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective', *Safety Science* 46(5), 751–770.

Ma, J. (2006), Bayesian multivariate Poisson-Lognormal regression for crash prediction on rural two-lane highways, PhD thesis, University of Texas, Ausin.

Ma, J. and Kockelman, K. M. (2006), 'Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity', *Transportation Research Record* 1950, 24–34.

Ma, J., Kockelman, K. M. and Damien, P. (2008), 'A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods.', *Accident Analysis & Prevention* 40(3), 964–75.

Maddison, D., Pearce, D., Johansson, O., Calthrop, E., Litman, T. and Verhoef, E. (1996), *True Costs of Road Transport, Blueprint 5*, Earthscan Publications, London.

Maher, M. and Summersgill, I. (1996), 'A comprehensive Methodology for the Fitting of predictive accident models', *Accident analysis and prevention* 28(3), 281–296.

Mannering, F. L. and Bhat, C. R. (2014), 'Analytic methods in accident research: Methodological frontier and future directions', *Analytic Methods in Accident Research* 1, 1–22.

Martin, J. L. (2002), 'Relationship between crash rate and hourly traffic flow on interurban motorways', *Accident Analysis & Prevention* 34(5), 619–629.

Massey, F. J. (1951), 'The Kolmogorov-Smirnov Test For Goodness Of Fit', *Journal of the American Statistical Association* 46(253), 68–78.

MathWorks (1999), Fuzzy Logic Toolbox Users Guide, Technical report, TheMathWorks, Inc. Mail, Natick,MA.

Mačiulis, A., Vasiliauskas, A. V. and Jakubauskas, G. (2009), 'The impact of transport on the competitiveness of national economy', *Transport* 24(2), 93–99.

May, A. (1994), *Interchange Operations on the Local Street Side: State of the Art*, Transportation Research Board.

Miaou, S. P. (1994), 'The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions', *Accident Analysis & Prevention* 26(4), 471–482.

Miaou, S.-P., Hu, P. S., Wright, T., Rathi, A. K. and Davis, S. C. (1992), 'Relationship between truck accidents and highway geometric design: a Poisson regression approach', *Transportation Research Record: Journal of the Transportation Research Board* 1376.

Miaou, S. P. and Lum, H. (1993), 'Modeling vehicle accidents and highway geometric design relationships.', *Accident Analysis & Prevention* 25(6), 689–709.

Milton, J. and Mannering, F. (1998), 'The relationship among highway geometrics , traffic-related elements and motor-vehicle accident frequencies', *Transportation* 25(6), 395–413.

Monsere, C. M., Bertini, R. L. and Bosa, P. G. (2006), Comparison of Identification and Ranking Methodologies for Speed-related Crash Locations, Technical report, No. FHWA-OR-RD-06-14, Oregon Department of Transportation Reseach Unit, Washington DC.

Montella, A. (2011), 'Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types.', *Accident Analysis & Prevention* 43(4), 1451–63.

Mountain, L. J., Hirst, W. M. and Maher, M. J. (2005), 'Are speed enforcement cameras more effective than other speed management measures? The impact of speed management schemes on 30 mph roads.', *Accident Analysis & Prevention* 37(4), 742–54.

Munden, J. (1967), The Relation Between a Driver's Speed and His Accident Rate, Technical report, Road Research Laboratory, Ministry of Transport, Road Research Laboratory, Ministry of Transport,Crowthorne, England.

Navon, D. (2003), 'The paradox of driving speed: two adverse effects on highway accident rate.', *Accident Analysis & Prevention* 35(3), 361–7.

Nilsson, G. (2004), Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety, PhD thesis, Lund Institute of Technology and Society, Traffic Engineering, 2004.

Noland, R. B. and Oh, L. (2004), 'The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data', *Accident Analysis & Prevention* 36(4), 525–532.

Norton, P. (2015), 'Four Paradigms: Traffic Safety in the Twentieth-Century United States.', *Technology and Culture* 56(2), 319–334.

OECD/ECMT (2006), *Speed Management*, OECD/ECMT Joint Transport 767 Research Committee, Paris, France.

Ossiander, E. M. and Cummings, P. (2002), 'Freeway speed limits and traffic fatalities in Washington State.', *Accident Analysis & Prevention* 34(1), 13–8.

Pande, A. and Abdel-Aty, M. (2005), 'A Freeway Safety Strategy for Advanced Proactive Traffic Management', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 9(3), 145–158.

Papadimitratos, P., La Fortelle, A., Evenssen, K., Brignolo, R. and Cosenza, S. (2009), 'Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation', *Communications Magazine, IEEE* 47(11), 84–95.

Park, B.-J., Fitzpatrick, K. and Lord, D. (2010), 'Evaluating the Effects of Freeway Design Elements on Safety', *Transportation Research Record: Journal of the Transportation Research Board* 2195, 58–69.

Park, E. S. and Lord, D. (2007), 'Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity', *Transportation Research Record: Journal of the Transportation Research Board* 2019, 1–6.

Patterson, T. L., Frith, J. W. and Small, M. W. (2000), Down with speed: a review of the literature, and the impact of speed on New Zealanders, Technical report, Accident Compensation Corporation and Land Transport Safety Authority,Wellington,NZ.

Pei, X., Wong, S. C. and Sze, N. N. (2012), 'The roles of exposure and speed in road safety analysis.', *Accident Analysis & Prevention* 48, 464–71.

Persaud, B. (1992), 'Road Safety-A Review of the Ontario Experience and Relevant Work Elsewhere', *Ontario Ministry of Transportation* .

Persaud, B. N. (1994), 'Accident prediction models for rural roads', *Canadian Journal of Civil Engineering* 21(4), 547–554.

Peters, S. C. and Iagnemma, K. (2009), 'Stability measurement of high-speed vehicles', *Vehicle System Dynamics* 47(6), 701–720.

Poch, M. and Mannering, F. (1996), 'Negative binomial analysis of intersection-accident frequencies', *Journal of Transportation Engineering* 122(April), 105–113.

Qin, X., Ivan, J. N. and Ravishanker, N. (2004), 'Selecting exposure measures in crash rate prediction for two-lane highway segments', *Accident Analysis & Prevention* 36, 183–191.

Qin, X., Parker, S., Liu, Y., Graettinger, A. J. and Forde, S. (2013), 'Intelligent geocoding system to locate traffic crashes.', *Accident Analysis & Prevention* 50, 1034–41.

Quddus, M. (2013), 'Exploring the Relationship Between Average Speed, Speed Variation, and Accident Rates Using Spatial Statistical Models and GIS', *Journal of Transportation Safety & Security* 5(1), 27–45.

Quddus, M. A. (2006), High integrity map matching algorithms for advanced transport telematics applications, PhD thesis, Imperial College London, United Kingdom.

Quddus, M. A. (2008), 'Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data.', *Accident Analysis & Prevention* 40(4), 1486–97.

Quddus, M. A., Ochieng, W. Y. and Noland, R. B. (2007), 'Current map-matching algorithms for transport applications : State-of-the art and future research directions', *Transportation Research Part C* 15, 312–328.

Quddus, M. A., Washington, Y. O., Zhao, L. and Noland, R. B. (2003), 'A General Map Matching Algorithm for Transport Telematics Applications', *GPS Solutions* 7(3), 157–167.

Quimbly, A., Maycock, G., Palmer, C. and Buttress, S. (1999), The factors that influence a driver ' s choice of speed a questionnaire study, Technical report, Transport Research Laboratory, Crowthorne, England.

Richter, E. D., Barach, P., Friedman, L., Krikler, S. and Israeli, A. (2004), 'Raised Speed Limits, Speed Spillover, Case-Fatality Rates, and Road Deaths in Israel: A 5-Year Follow-Up', *American Journal of Public Health* 94(4), 568–574.

Rock, S. M. (1995), 'Impact of the 65 mph speed limit on accidents, deaths, and injuries in Illinois.', *Accident Analysis & Prevention* 27(2), 207–14.

RTI (1970), Speed and accidents. Vols. I & II, Technical report, Research Triangle Institute, North Carolina.

Shankar, V., Mannering, F. and Barfield, W. (1995), 'Effect of roadway geometrics and environmental factors on rural freeway accident frequencies.', *Accident Analysis & Prevention* 27(3), 371–89.

Solomon, D. (1964), Accidents on main rural highways related to speed, driver and vehicle, Technical report, U.S Department of Commerce & Bureau of Public Roads, Washington DC.

Soole, D. W., Watson, B. C. and Fleiter, J. J. (2013), 'Effects of average speed enforcement on speed compliance and crashes: a review of the literature.', *Accident Analysis & Prevention* 54, 46–56.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4), 583–639.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003), 'WinBUGS user manual'.

Steenberghen, T., Dufays, T., Thomas, I. and Flahaut, B. (2004), 'Intra-urban location and clustering of road accidents using GIS: a Belgian example', *International Journal of Geographical Information Science* 18(2), 169–181.

Stuster, J. (2004), Aggressive Driving Enforcement: Evaluations of Two Demonstration Programs. Report DOT HS 809 707., Technical report, National Highway Traffic Safety Administration, Washington DC.

Tarko, A. P., Thomaz, J. and Grant, D. (2009), 'Probabilistic Determination of Crash Locations in a Road Network with Imperfect Data', *Transportation Research Record: Journal of the Transportation Research Board* 2102, 76–84.

Taylor, M. C., Lynam, D. A. and Baruya, A. (2000), The effects of drivers speed on the frequency of road accidents, Technical report, Transport Research Laboratory, Crowthorne, England.

Tegge, R. and Ouyang, Y. (2009), 'Correcting erroneous crash locations in transportation safety analysis', *Accident Analysis & Prevention* 41, 202–209.

Thill, J.-C. (2000), 'Geographic information systems for transportation in perspective', *Transportation Research Part C* 8, 3–12.

Thomas, A., Best, N., Lunn, D., Arnold, R. and Spiegelhalter, D. (2004), 'GeoBugs user manual', *Cambridge: Medical Research Council Biostatistics Unit* .

TRB (1998), Managing Speed: Review of Current Practice for Setting and Enforcing Speed Limits. Special Report 254, Technical report, Transportation Research Board, Washington D.C.

Vadeby, A. (2015), Speed changes and crash risk - a case based study within IRTAD, *in* 'Transportation Research Board 94th Annual Meeting'.

Vadeby, A. and Forsman, A. s. (2010), Evaluation of the new speed limits-the effect on vehicle speeds, Phase 1, Technical report, VTI, Swedish National Road and Transport Research Institute, Linköping, Sweeden.

Vogt, A. and Bared, J. (1998), 'Accident models for two-lane rural segments and intersections', *Transportation Research Record: Journal of the Transportation Research Board* 1635(1), 18–29.

Wagenaar, A. C., Streff, F. M. and Schultz, R. H. (1990), 'Effects of the 65 mph speed limit on injury morbidity and mortality.', *Accident Analysis & Prevention* 22(6), 571–85.

Wang, C. (2010), The Relationship Between Traffic Congestion and Road Accidents: An Econometric Approach Using GIS, PhD thesis, Loughborough University, Loughborough, England.

Wang, C., Quddus, M. a. and Ison, S. G. (2009*a*), 'Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England.', *Accident Analysis & Prevention* 41(4), 798–808.

Wang, C., Quddus, M. and Ison, S. (2009*b*), 'The effects of area-wide road speed and curvature on traffic casualties in England', *Journal of Transport Geography* 17(5), 385–395.

Wang, H., Li, Z., Hurwitz, D. and Shi, J. (2013), 'Parametric modeling of the heteroscedastic traffic speed variance from loop detector data', *Journal of Advanced Transportation* 49, 279–296.

Wang, X., Wang, T., Tarko, A. and Tremont, P. J. (2015), 'The influence of combined alignments on lateral acceleration on mountainous freeways: a driving simulator study', *Accident Analysis & Prevention* 76, 110–117.

Whitelegg, J. and Haq, G. (2006), 'Vision zero: Adopting a target of zero for road traffic fatalities and serious injuries', *Stockholm Environment Institute* .

WHO (2013), *WHO global status report on road safety 2013: supporting a decade of action*, World Health Organization.

Wilson, C., Willis, C., Hendrikz, J. K. and Bellamy, N. (2006), 'Speed enforcement detection devices for preventing road traffic injuries', *Cochrane Database of Systematic Reviews* 2.

Wong, S. C., Sze, N. N., Lo, H. K., Hung, W. T. and Loo, B. P. Y. (2005), 'Would relaxing speed limits aggravate safety? A case study of Hong Kong.', *Accident Analysis & Prevention* 37(2), 377–88.

Xie, Y., Zhao, K. and Huynh, N. (2012), 'Analysis of driver injury severity in rural single-vehicle crashes', *Accident Analysis & Prevention* 47, 36–44.

Ye, X., Pendyala, R. M., Washington, S. P., Konduri, K. and Oh, J. (2009), 'A simultaneous equations model of crash frequency by collision type for rural intersections', *Safety Science* 47(3), 443–452.

Yiannis, G., Papantoniou, P. and Georgiadou, E. (2015), Investigation of the impact on road safety of increasing the speed limit on motorways, *in* '6th Greek National Road Safety Conference'.

Yuan, W., Fu, R., Guo, Y., Feng, H. and Shi, J. (2008), 'Influences of Longitudinal Gradient on Traffic Accident Rate Considering Length of Downgrade', *Journal of Highway and Transportation Research and Development (English Edition)* 3(2), 122–126.

Zhao, Y. (1997), *Vehicle Location and Navigation System.*, Artech House, Inc., MA.

Zhou, M. and Sisiopiku, V. P. (1997), 'Relationship between volume-to-capacity ratios and accident rates', *Transportation Research Record: Journal of the Transportation Research Board* 1581(1), 47–52.

Appendix A

Distribution of speeds just before accidents

As it has been explained in Section 3.5 every accident was matched with a speed observation that reflects the speed just before the accident time and location. Figures A.1-A.11 show the distributions of these speeds for all accidents together and then by accident severity and collision type.

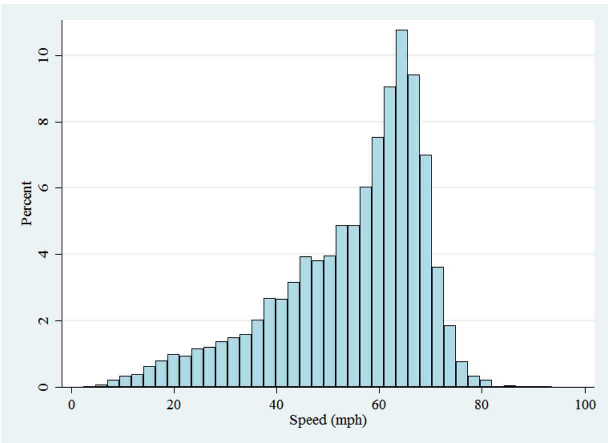


Figure A.1: Speed distribution of all accidents.

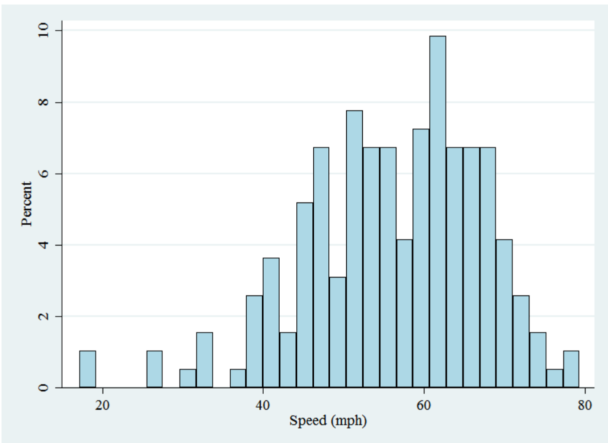


Figure A.2: Speed distribution of fatal accidents.

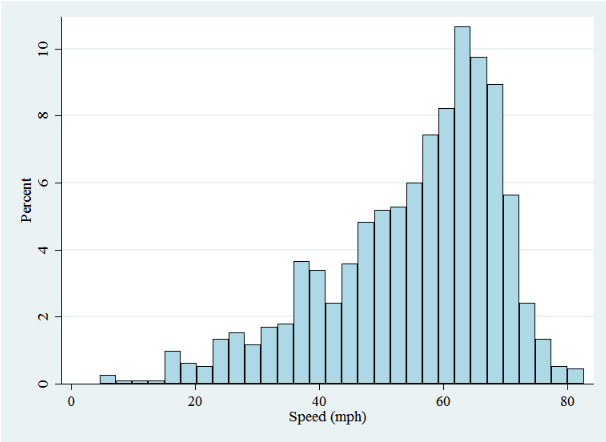


Figure A.3: Speed distribution of serious accidents.

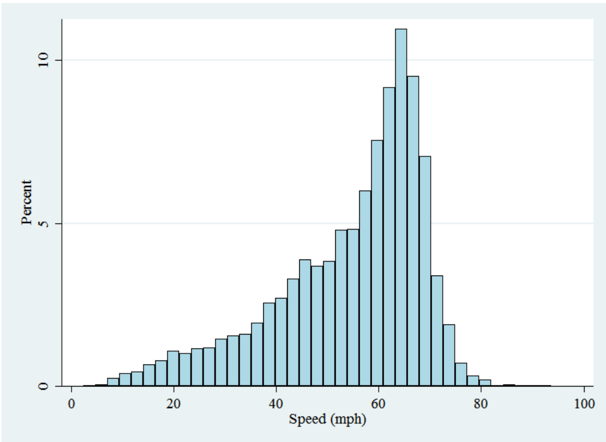


Figure A.4: Speed distribution of slight accidents.

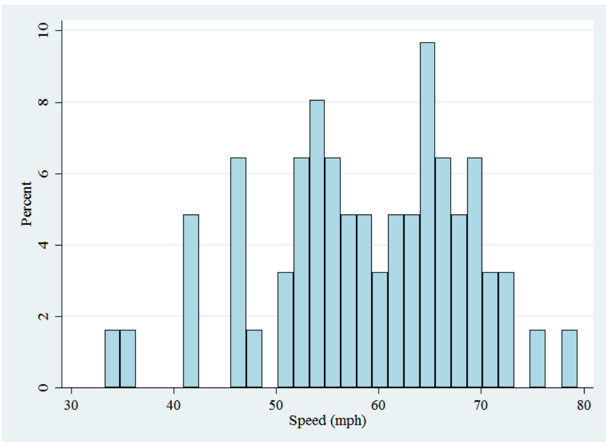


Figure A.5: Speed distribution of fatal single vehicle accidents.

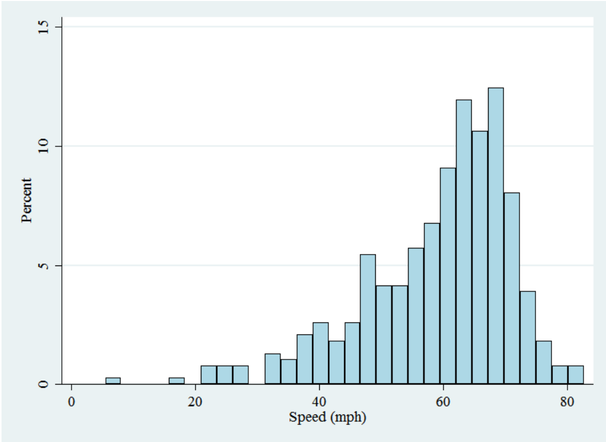


Figure A.6: Speed distribution of serious single vehicle accidents.

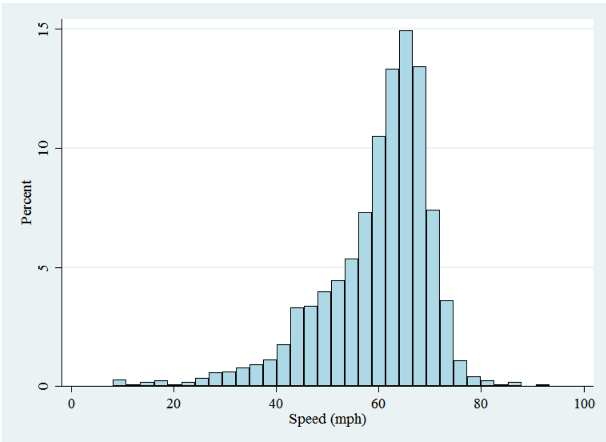


Figure A.7: Speed distribution of slight single vehicle accidents.

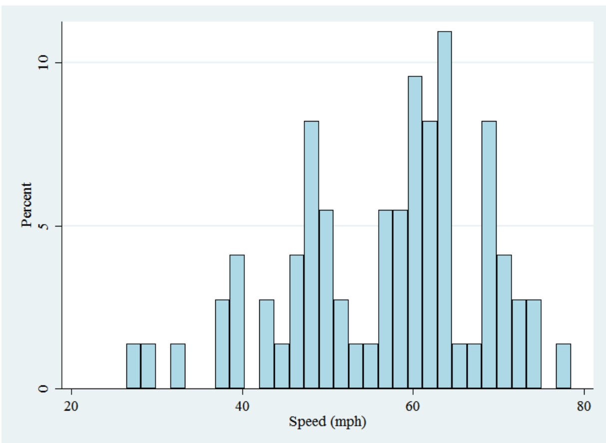


Figure A.8: Speed distribution of fatal multiple vehicle accidents.

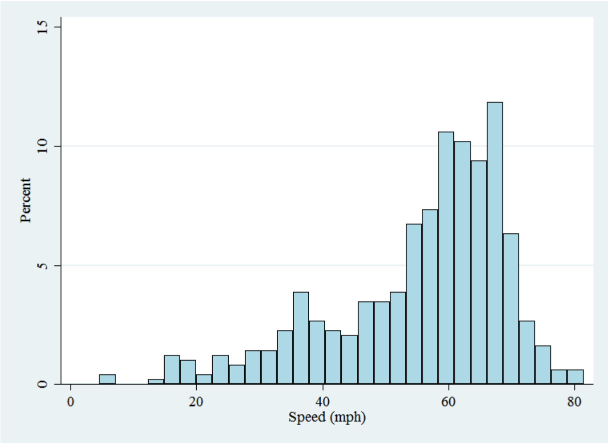


Figure A.9: Speed distribution of serious multiple vehicle accidents.

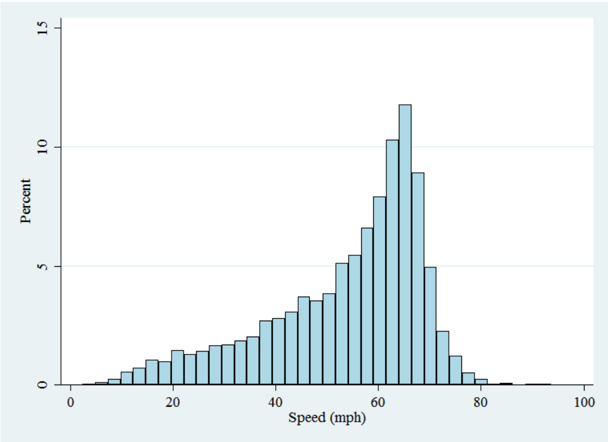


Figure A.10: Speed distribution of slight multiple vehicle accidents.

Distribution of volume per lane just before accidents

Similarly to the above section, Figures A.11-A.20 present the distribution of the volumes for all accidents and by accident severity and collision type.

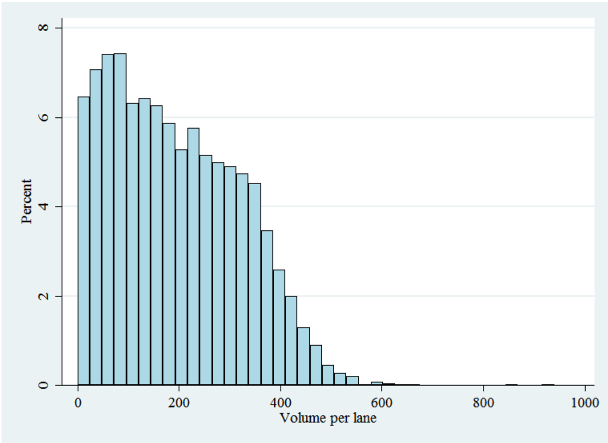


Figure A.11: Volume per lane distribution of all accidents.

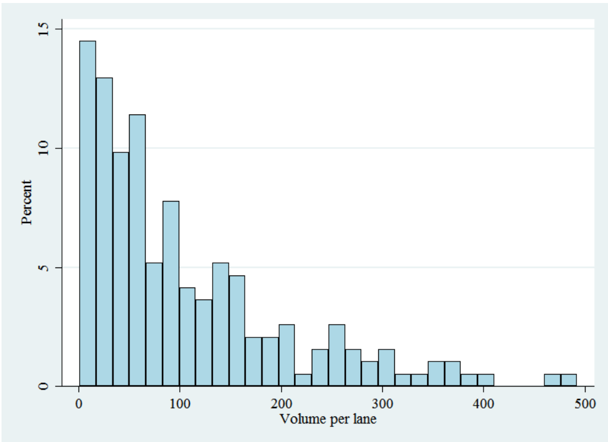


Figure A.12: Volume per lane distribution of fatal accidents.

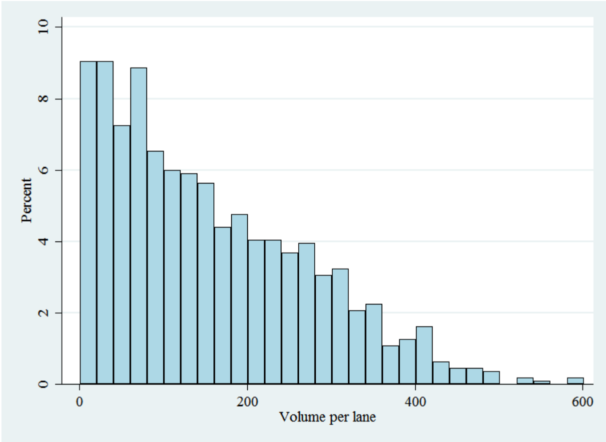


Figure A.13: Volume per lane distribution of serious accidents.

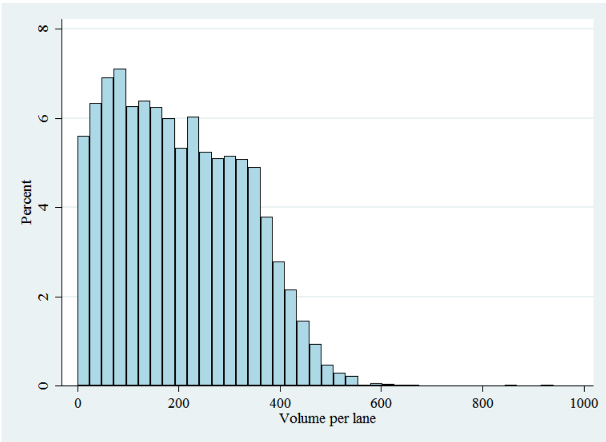


Figure A.14: Volume per lane distribution of slight accidents.

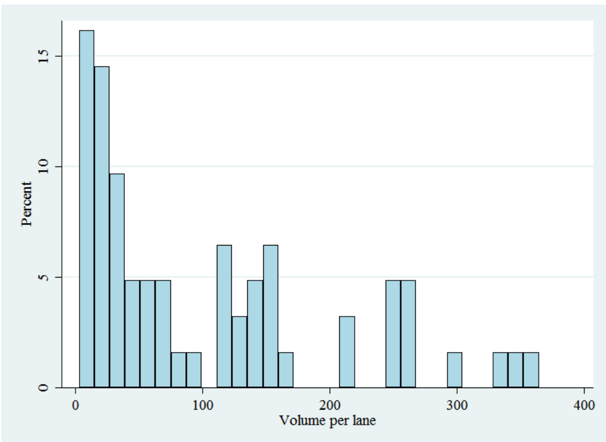


Figure A.15: Volume per lane distribution of fatal single vehicle accidents.

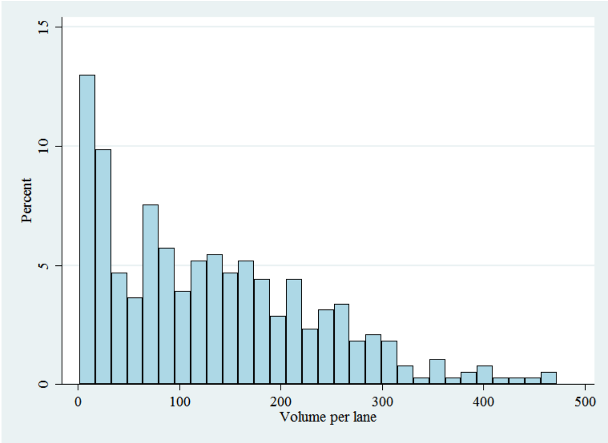


Figure A.16: Volume per lane distribution of serious single vehicle accidents.

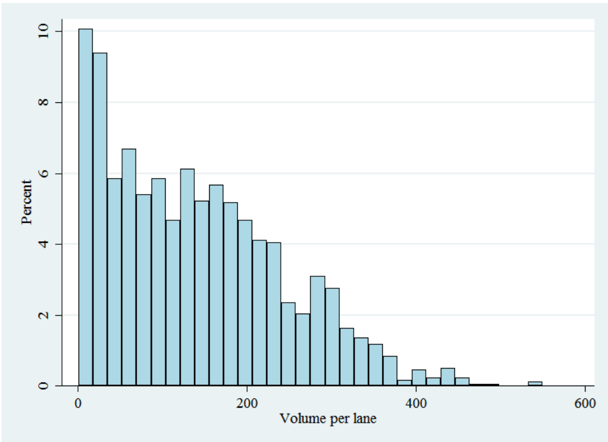


Figure A.17: Volume per lane distribution of slight single vehicle accidents.

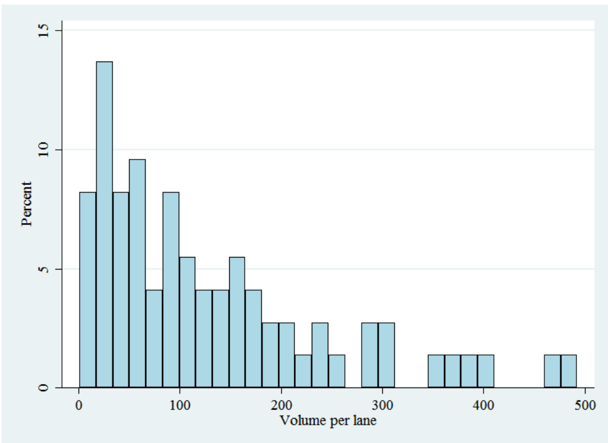


Figure A.18: Volume per lane distribution of fatal multiple vehicle accidents.

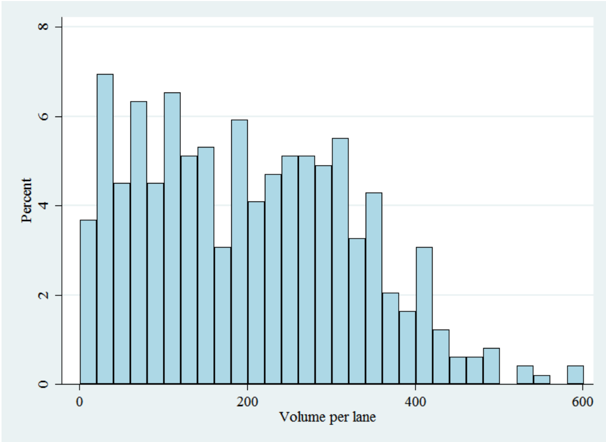


Figure A.19: Volume per lane distribution of serious multiple vehicle accidents.

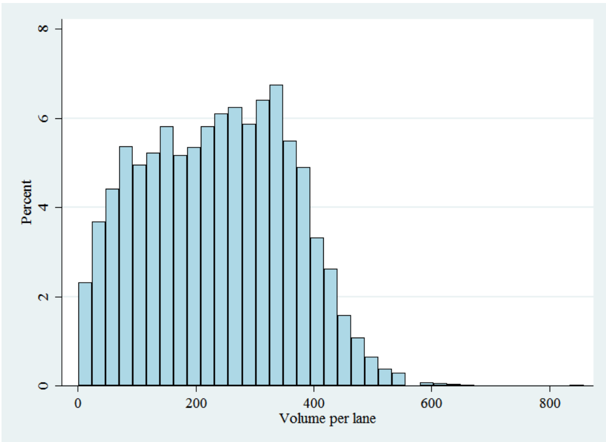


Figure A.20: Volume per lane distribution of slight multiple vehicle accidents.

Appendix B

Accidents by injury severity (K-S-Sl)

Both the Link-based and the Condition-based multivariate models for K, S and Sl accidents did not provide statistically significant coefficients for either speed or volume in none of the 20 variable specifications that were tested. That is why these models were not suitable for calculation of the impact of a speed limit increase as it was initially intended. A possible explanation for this is the very high number of zeroes in the K and S variables. The coefficient estimations of the best fitting specification of the K-S-Sl models for both the aggregation approaches are presented in Tables B.1 and B.3.

Table B.1: Parameter estimates for the link-based multivariate Poisson-lognormal model for fatal (K), serious (S) and slight (Sl) accidents (*Link-based K-S-Sl (7)*)

K accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
ln(Speed)	-1.2875 **	0.3512	0.0241	-1.9121	-1.8167	-0.6744	-0.4833
AADT	-0.0028	0.0058	0.0002	-0.0143	-0.0124	0.0067	0.0085
Curve	0.1825	0.1663	0.0037	-0.1459	-0.0945	0.4545	0.5061
Uphill	-0.1927	0.3251	0.0061	-0.8548	-0.7384	0.3303	0.4164
Downhill	0.1689	0.1594	0.0031	-0.1426	-0.0937	0.4324	0.4845
Lanes	0.1525	0.2257	0.0063	-0.2951	-0.2170	0.5224	0.5912
Intercept	0.4039	1.3785	0.0945	-2.7687	-2.0177	2.5082	2.8374
S accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
ln(Speed)	-1.6653 **	0.2168	0.0151	-2.1752	-2.0671	-1.3173	-1.2598
AADT	0.0089 **	0.0026	0.0001	0.0038	0.0046	0.0131	0.0138
Curve	-0.0657	0.0775	0.0023	-0.2172	-0.1917	0.0619	0.0883
Uphill	-0.0576	0.1371	0.0029	-0.3437	-0.2876	0.1629	0.2030
Downhill	-0.0416	0.0727	0.0017	-0.1858	-0.1610	0.0773	0.1005
Lanes	0.0827	0.1078	0.0043	-0.1319	-0.0949	0.2585	0.2920
Intercept	4.0538 *	0.8765	0.0610	2.4176	2.6467	5.7181	6.1395
Sl accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
ln(Speed)	-1.9287 **	0.1132	0.0080	-2.1773	-2.1456	-1.7587	-1.7352
AADT	0.0240 **	0.0013	0.0001	0.0214	0.0217	0.0262	0.0266
Curve	0.0043	0.0380	0.0012	-0.0699	-0.0579	0.0668	0.0786
Uphill	0.0765	0.0641	0.0016	-0.0513	-0.0306	0.1813	0.2026
Downhill	0.0718 **	0.0362	0.0010	0.0004	0.0127	0.1317	0.1428
Lanes	0.0439	0.0546	0.0024	-0.0629	-0.0472	0.1346	0.1519
Intercept	6.7303 **	0.4503	0.0316	5.9681	6.0534	7.5999	7.7345
$\bar{D}$	12790	**statistically significant at the 95% credible interval					
p_D	742						
DIC	13532						

Table B.2: Combined Covariance-Correlation matrix of the random effect of the *Link-based K-S-Sl (7)* model.

A	K accidents	S accidents	Sl accidents
K accidents	0.348**	0.28**	0.211**
S accidents	**0.908****	0.273**	0.178**
Sl accidents	**0.757****	**0.724****	0.223**
B	K accidents	S accidents	Sl accidents
K accidents	0.006**	0.002**	0.003**
S accidents	**0.344****	0.008**	0.005**
Sl accidents	**0.402****	**0.534****	0.01**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval

Table B.3: Parameter estimates for the condition-based multivariate Poisson-lognormal model for fatal (K), serious (S) and slight (Sl) accidents (*Condition-based K-S-Sl (10)*)

K accidents	Mean	SD	MC error	2.5%	5%	95%	97.5%
Speed	0.02811	0.01728	0.0005317	-0.00531	-0.00004	0.05680	0.06197
Speed squared	-0.00018	0.00016	0.0000048	-0.00050	-0.00045	0.00009	0.00014
Volume	-0.02470 **	0.00324	0.0000630	-0.03101	-0.03004	-0.01938	-0.01836
Volume squared	0.00004 **	0.00001	0.0000002	0.00002	0.00003	0.00006	0.00007
Curve	0.08182	0.14976	0.0008583	-0.21238	-0.16494	0.32741	0.37564
Uphill	1.69598 **	0.34149	0.0047644	1.05371	1.15406	2.27498	2.39824
Downhill	2.64576 **	0.32377	0.0047632	2.04827	2.13583	3.20088	3.31813
Lanes	-0.66454 **	0.16106	0.0006778	-0.98257	-0.93142	-0.40115	-0.35258
Intercept	-4.74510 **	0.59221	0.0159974	-5.91887	-5.73395	-3.78561	-3.59610
ln(VehHr/mile)	1	-	-	-	-	-	-
S accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
Speed	0.01706	0.011792	0.0004217	-0.00565	-0.00023	0.03706	0.04079
Speed squared	-0.00011	0.000116	0.0000041	-0.00035	-0.00031	0.00008	0.00011
Volume	-0.01658 **	0.001561	0.0000402	-0.01965	-0.01915	-0.01402	-0.01352
Volume squared	0.00003 **	0.000005	0.0000001	0.00002	0.00003	0.00004	0.00004
Curve	0.06776	0.069765	0.0004221	-0.06852	-0.04703	0.18290	0.20503
Uphill	2.14829 **	0.173792	0.0029489	1.81762	1.86881	2.44068	2.49627
Downhill	2.95681 **	0.168723	0.0029483	2.63720	2.68532	3.24092	3.29557
Lanes	-0.66081 **	0.073928	0.0004915	-0.80652	-0.78295	-0.53954	-0.51748
Intercept	-3.65365 **	0.344858	0.0111135	-4.33652	-4.22520	-3.09491	-2.98007
ln(VehHr/mile)	1	-	-	-	-	-	-
Sl accidents	**Mean**	**SD**	**MC error**	**2.5%**	**5%**	**95%**	**97.5%**
Speed	0.03648 **	0.008672	0.0003355	0.01873	0.02142	0.05014	0.05159
Speed squared	-0.00037 **	0.000084	0.0000032	-0.00052	-0.00050	-0.00022	-0.00020
Volume	-0.01014 **	0.000843	0.0000276	-0.01177	-0.01150	-0.00873	-0.00847
Volume squared	0.00002 **	0.000003	0.0000001	0.00002	0.00002	0.00003	0.00003
Curve	0.11894 **	0.037877	0.0003241	0.04461	0.05666	0.18114	0.19291
Uphill	2.25357 **	0.072534	0.0012971	2.11281	2.13579	2.37376	2.39738
Downhill	2.90723 **	0.071425	0.0013364	2.76866	2.79079	3.02584	3.04835
Lanes	-0.32547 **	0.037881	0.0002995	-0.40007	-0.38794	-0.26324	-0.25127
Intercept	-2.72188 **	0.234734	0.0086965	-3.15577	-3.10112	-2.33506	-2.26015
ln(VehHr/mile)	1	-	-	-	-	-	-
$\bar{D}$	11492.8	**statistically significant at the 95% credible interval					
p_D	965.937						
DIC	12458.8						

Table B.4: Combined Covariance-Correlation matrix of the random effect of the *Condition-based K-S-Sl (10)* model.

	K accidents	S accidents	Sl accidents
K accidents	0.41**	0.367**	0.326**
S accidents	**0.911****	0.397**	0.325**
Sl accidents	**0.924****	**0.938****	0.303**

Correlation is marked with bold font.

**statistically significant at the 95% credible interval

Appendix C

Publications related to this thesis

Imprialou, M.I.M., Quddus M., Pitfield D.E and Lord D. (2016), Re-visiting the speed-crash relationships: A new perspective in crash modelling., *Accident Analysis & Prevention*, 86, 173-185.

Imprialou, M.I.M., Quddus M., and Pitfield D.E. (2016), Predicting the impact of a speed limit increase using condition-based Multivariate Poisson Log-normal regression., *Transportation Planning and Technology (in Press)*.

Imprialou, M.I.M., Quddus M. and Pitfield D.E (2015), Multilevel logistic regression modelling for crash mapping in metropolitan areas., *Transportation Research Record: Journal of the Transportation Research Board*, 2515, 39-47.

Imprialou, M.I.M., Quddus, M. and Pitfield, D.E. (2015), Exploring the role of Speed in highway crashes: Pre-crash-condition-based Multivariate Bayesian modelling, in 94th TRB Annual Meeting, January 11-15, Washington DC.

Imprialou, M.I.M. (2015), A new modelling approach to develop accident-speed relationships using Multivariate Poisson Log-normal regression modelling, in 47th Annual UTSG Conference, January 5-7, London.

Imprialou, M.I.M., Quddus M. and Pitfield D.E (2014), High accuracy crash mapping using fuzzy logic., *Transportation Research Part C: Emerging Technologies* 42, 107-120.

Imprialou, M.I.M., Quddus, M., Pitfield, D.E. and Li, L. (2014), A high-accuracy generic method for automatic crash mapping using Fuzzy Logic, 93rd TRB Annual Meeting, January 12-16, Washington DC and in 46th Annual UTSG Conference, January 6-8, Newcastle.