National Technical University of Athens
School of Civil Engineering
Department of Transportation Planning and Engineering

An advanced multi-faceted statistical analysis of accident probability and severity exploiting high resolution traffic and weather data

By
Athanasios Theofilatos

A Doctoral Thesis

Submitted in partial fulfilment of the requirements
for the award of
Doctor of Philosophy

Advisory board:
G. Yannis
J. Golias
G. Kanellaidis

November 2015

To my parents and Lily...

Preface

The completion of the present thesis has been an intensive challenge requiring a tremendous amount of effort, discipline and time throughout the last four years. However, it would not have been completed without the elements of passion, devotion and love, as I never considered science and research simply as another “job”. It became part of who I am. When talking to my friends and family I sometimes compared my PhD with the journey of one of Stephen King’s most famous heroes who was searching for the dark tower. So, when I look at the past four years of my life, what I see is a journey to find my own dark tower. Even if there were some really hard times when I simply felt like “throwing the towel”, in the end I feel I finally have managed to find my own dark tower (at least for now!). In this journey, many individuals have considerably helped in completing the work undertaken or overcoming difficulties (of any kind). Without them I would not have been able to accomplish the present thesis. So, I would like to acknowledge their contribution and express my deepest gratitude from the depth of my heart.

I wish to thank, first and foremost, my supervisor, Professor George Yannis, who has inspired my interest and love towards road safety since my undergraduate studies and has supported me so much all these years. It was a great pleasure to work with him and I have learned a great deal from him, in terms of both academic and everyday life. I would also like to thank Professor John Golias and Professor George Kanellaidis for the encouragement, support and for the constructive comments. In addition, Professor Golias helped me “complete the puzzle” when I was totally snowed under, while the advices of Professor Kanellaidis especially at the early stages of the PhD, proved invaluable and served as a great guidance.

Furthermore, I would like to thank Assistant Professor Eleni Vlahogianni and Associate Professor Costas Antoniou, for their contribution and for the very fruitful discussions we had from the early days of my PhD years. They both encouraged me to keep searching and always improve my work. I am also grateful to Professor Magda Pitsiava-Latinopoulou and Professor Panos Papaioannou for their comments and support.

I would like of course, to thank Dr. Eleonora Papadimitriou for her help and support and also for her encouragement to start playing music again after almost 20 years, something that turned out to be very important to me.

I feel I have to mention another very important Professor, who believed in me and helped me so much from the first time I met him, when I was a student at the School of Civil Engineering. His name was Matthew Karlaftis. I owe him my passion for statistics. Thank you. You will always be missed and never be forgotten.

Furthermore, I have to express my gratitude to NTUA Research Committee (ELKE) for providing funding for my PhD. Without this financial support, it would be impossible for me to do it. I am also very thankful to Mrs. Lia Theofanidou, Mrs.

Maria Kouaki and Mr. George Kardamylakis from NTUA, for the continuous support and assistance regarding all administrative procedures.

I express my gratitude to those who assisted in collecting and processing the enormous amount of required data. More specifically, I would like to thank Antonis Chaziris, Theodoros Vorvolakos and Dimitris Serbis from Athens Traffic Management Centre (KDK), as well as Dr. Pantelis Kopelias and Fanis Papadimitriou from the Attica Tollway Traffic Management Centre, who provided me the opportunity to have access to rich and precious data and also for their very pleasant cooperation. A big thanks goes to my good friend Apostolos Ziakopoulos, who put great effort in helping me collecting these massive data. I am also grateful to Dimitris Pikoulas for his invaluable contribution during the data processing and to Antony Koutsourelis for the fruitful discussions about statistical models.

I am very grateful to Lily Koutsogiannopoulou for her patience, love, inspiration, tolerance and understanding. She carried me in my “highest highs” and saved me “in my lowest lows”. I would like to thank my parents, Tom and Zoe who always loved me, supported me and taught me to be such a dreamer. I really owe them. Thanks!

Last but not least, I express my gratitude to all the people of the research group of the Department of the Transportation Planning and Engineering (Petros, Dimos, Panagiotis, Titi, Anna, Katerina, Elli, Areti, Catherine) for their support and for being who they are! However, I really have to offer a special thanks to all my beloved colleagues of the 3rd floor (our big cold cave!) for the great time we have spent together all these years: Alexandra, Dimitris, Manos, Nadia, Chrysa, Foteini, Iosif, Efi, Brad. All my close friends (Panos, George-my best man!, Aristotelis, Nodas-the little one, Panagiotis, Apostolos, Nick, Fotis, Nodas-the big one, Dimitris, Joseph, Jim, Jacek, Alex) were always there for me in both good and bad times. I have of course to mention, that during my army days which was a much-needed break for me, I was fortunate to meet some wonderful people who became beloved friends of mine (Apostolis, Vasilis P., Vasilis M., Vasilis V.). Thank you all!

“Luck's the word those with poor hearts use for ka...”
— Stephen King

And now, the end is near;
And so I face the final curtain.
My friend, I'll say it clear,
I'll state my case, of which I'm certain.

I've lived a life that's full.
I've travelled each and every highway;
And more, much more than this,
I did it my way.

Regrets, I've had a few;
But then again, too few to mention.
I did what I had to do
And saw it through without exemption.

I planned each charted course;
Each careful step along the byway,
And more, much more than this,
I did it my way.

Yes, there were times, I'm sure you knew
When I bit off more than I could chew.
But through it all, when there was doubt,
I ate it up and spit it out.
I faced it all and I stood tall;
And did it my way.

I've loved, I've laughed and cried.
I've had my fill; my share of losing.
And now, as tears subside,
I find it all so amusing.

To think I did all that;
And may I say - not in a shy way,
"Oh no, oh no not me,
I did it my way".

For what is a man, what has he got?
If not himself, then he has naught.
To say the things he truly feels;
And not the words of one who kneels.
The record shows I took the blows -
And did it my way!

Yes, it was my way.

— My way, Frank Sinatra

Abstract

The objective of this PhD thesis is the investigation of accident probability and severity exploiting high resolution traffic and weather data from urban roads and motorways, collected on a real-time basis, with specific focus on Powered-Two-Wheelers. For that purpose, an advanced mesoscopic multi-faceted statistical analysis was conducted in order to expand previous road safety work and contribute to the further understanding the complex issues of accident probability and severity. Linear models as well as non-linear models were developed on the basis of 6-year accident data from urban roads as well as an urban motorway in Greater Athens area (Attica Tollway). Empirical findings indicate that high resolution traffic and weather data are capable of opening new dimensions in accident analysis in urban roads and urban motorways. The multi-faceted statistical analysis conducted in the thesis has revealed a consistent and strong influence of traffic parameters on accident probability and severity. It is interesting that weather parameters were not found to influence accident probability and severity when linear relationships are considered. Lastly, the application of cusp catastrophe models demonstrated that it is likely that even small traffic and weather changes may have a critical impact on road safety in urban roads as sudden transitions from safe to unsafe conditions (and vice versa) may occur.

Περίληψη

Ο στόχος της παρούσας διδακτορικής διατριβής είναι η διερεύνηση της πιθανότητας και σοβαρότητας ατυχήματος σε αστικές οδούς και αστικούς αυτοκινητοδρόμους με έμφαση στους δικυκλιστές, αξιοποιώντας κυκλοφοριακά και μετεωρολογικά δεδομένα υψηλής ευκρίνειας τα οποία έχουν συλλεχθεί σε πραγματικό χρόνο. Για τον σκοπό αυτό, εφαρμόστηκε μεσοπρόθεσμου χαρακτήρα σύγχρονη πολυ-επίπεδη στατιστική ανάλυση, με σκοπό να συμβάλει στην περαιτέρω κατανόηση των πολύπλοκων ζητημάτων της πιθανότητας και της σοβαρότητας του ατυχήματος. Γραμμικά μοντέλα καθώς και μη γραμμικά μοντέλα αναπτύχθηκαν αξιοποιώντας έξι ετών δεδομένα ατυχημάτων από αστικούς δρόμους καθώς και από την Αττική οδό. Τα αποτελέσματα δείχνουν ότι τα υψηλής ευκρίνειας κυκλοφοριακά και μετεωρολογικά δεδομένα είναι ικανά να ανοίξουν νέες διαστάσεις στην ανάλυση των τροχαίων ατυχημάτων στις αστικές οδούς και στους αστικούς αυτοκινητόδρομους. Η πολυ-επίπεδη στατιστική ανάλυση που διενεργήθηκε στο πλαίσιο της διατριβής, έδειξε ότι υπάρχει σημαντική επιρροή των κυκλοφοριακών παραμέτρων στην πιθανότητα και στη σοβαρότητα ατυχήματος. Είναι ενδιαφέρον το γεγονός ότι οι μετεωρολογικές παράμετροι δεν βρέθηκαν να επηρεάζουν την πιθανότητα ατυχήματος και τη σοβαρότητα, όταν εξετάζονται οι γραμμικές σχέσεις. Τέλος, η εφαρμογή της θεωρίας καταστροφής έδειξε ότι ακόμα και μικρές αλλαγές στις κυκλοφοριακές και μετεωρολογικές παραμέτρους μπορεί να έχουν σημαντική επίδραση στην οδική ασφάλεια σε αστικές οδούς, καθώς ξαφνικές μεταβάσεις από ασφαλείς σε μη ασφαλείς καταστάσεις (και το αντίστροφο) είναι δυνατό να συμβούν.

Extended Abstract

The effective treatment of road accidents and the improvement of road safety level is a major concern to societies due to the losses in human lives and the economic and social cost. Tremendous efforts have been dedicated by transportation researchers and practitioners to improve road safety. Recently, high resolution real-time traffic and weather data started to be used when analysing road safety in freeways. Regardless of modelling techniques, a major gap is that **very limited research has been conducted so far for urban roads**. Moreover, there is **no specific focus on Powered-Two-Wheelers (PTWs)**, which constitute a vulnerable type of road users and are affected by the interaction with other motorized traffic. Taking also into account the speeding and the manoeuvring capabilities of PTWs, investigation of PTW safety by incorporating traffic conditions would be of particular interest. It should be noted, that **an integrated methodology** was needed in order to understand accident probability and severity, due to the complex nature of these phenomena.

In that context, **the main research question of the thesis was whether and how traffic and weather parameters affect accident probability and severity in urban roads and urban motorways.** The thesis objectives were achieved through the utilization of high resolution traffic and weather data collected on a real-time-basis in order to conduct **a multi-faceted statistical exploration of accident probability and severity.**

To this end, **a number of five main research activities** were carried out:

1) A literature review of relevant research
2) Data collection and processing
3) Statistical analysis of accident probability in urban roads and motorways
4) Statistical analysis of accident severity in urban roads and motorways
5) Consideration of PTWs in the aforementioned statistical analyses

The **first research activity** included an extensive literature review, investigating the research topics examined: the effect of traffic and weather characteristics on road safety and afterwards the critical parameters of PTW behaviour and safety. More specifically, a systematic review of the effect of traffic and weather characteristics on road safety was conducted firstly, having a specific focus on recent studies featuring high resolution traffic and weather data. Then studies related to rider behaviour, PTW interaction with other motorized traffic, accident frequency, accident rates and accident severity were examined. The extensive literature review led to the identification of the research gaps and open research questions.

The **second research activity** concerns the data collection and processing. Empirical data have been collected for the period 2006-2011 to investigate the relationship between traffic, weather and other characteristics and road accidents. The road axes chosen were the Kifisias and Mesogeion avenues in Athens, Greece, mainly due to the fact that they had very similar characteristics. Secondary, Attica Tollway (“Attiki Odos”) was also chosen to be investigated separately.

The required **accident data** were collected from the Greek accident database SANTRA provided by the Department of Transportation Planning and Engineering of the National Technical University of Athens, based on data collected by the Police and coded by the Hellenic Statistical Authority. A 6-year period was considered, from 2006 to 2011. **Traffic data** were extracted from the Traffic Management Centre (TMC) of Athens for Kifisias and Mesogeion avenues, and from the Traffic Management Centre of Attica Tollway for the urban motorway. **Weather** data were collected from the Hydrological Observatory of Athens (HOA), operated by the Laboratory of Hydrology and Water Resources Management of the National Technical University of Athens.

Data collection led to the **data processing**. In this step, data quality was ensured (e.g. false values of traffic measurements were removed). Concerning the **accident cases**, the raw 5-min traffic and the 10-min weather data were further aggregated into 1-hour intervals in order to obtain averages, standard deviations and so on, following a more **mesoscopic analysis approach**. For accident probability examination purposes, data from **non-accident cases** were also collected, following the usual procedure described in international literature (Abdel-Aty and Pande, 2005; Abdel-Aty et al., 2007; Ahmed and Abdel-Aty, 2012; Yu and Abdel-Aty, 2013a).

In order to achieve the aims of the thesis through the **aforementioned research activities (third, fourth and fifth activity)**, a set of statistical analyses were carried out:

- Combined utilization of time series data and machine learning techniques (Support Vector Machine models) to predict PTW accident involvement and PTW accident type (Chapter 5),
- Finite mixture cluster analysis to identify traffic states and then explore the effect of traffic states on accident probability, accident severity, PTW accident severity and PTW accident involvement (Chapter 6),
- Investigation of the effect of individual traffic and weather parameters on accident probability and severity, by applying Random Forests (to detect potential significant variables) and then by applying finite mixture and Bayesian logit models (Chapter 7),
- Development of finite mixture, Bayesian and rare-events logit models to explore the factors affecting accident probability and severity in Attica Tollway (Chapter 8) and
- Application of the cusp catastrophe theory to estimate accident probability and accident severity in urban roads (Chapter 9).

This PhD thesis deals with accident probability and accident severity having specific focus on Powered-Two-Wheelers. For that reason, separate PTW severity and probability models were developed. Different models and datasets were used throughout the thesis to achieve the aim of the research. Figure 1 illustrates the general methodological framework of the PhD thesis.

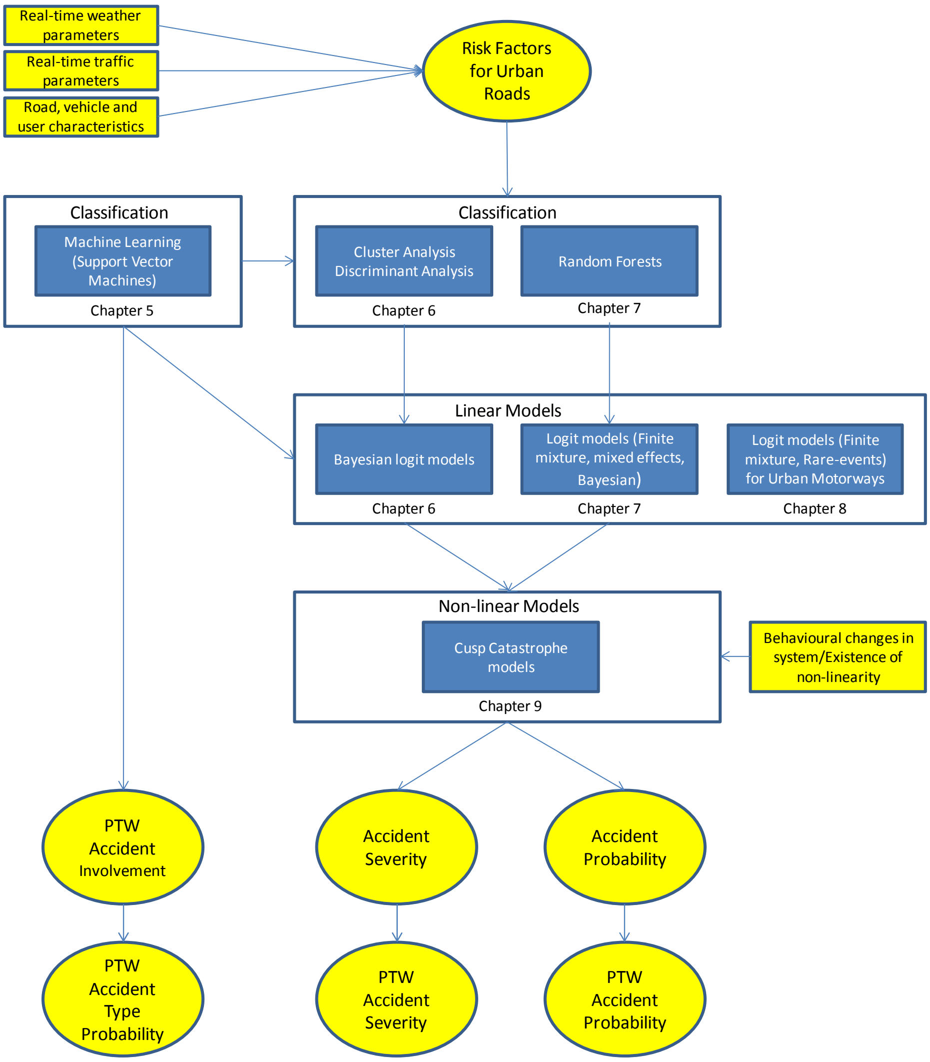


Figure 1: Overview of the methodological framework.

The present PhD thesis resulted in a number of original scientific contributions which are presented at the following sections. Section 11.2.1 demonstrates the main methodological contributions and conclusions, whilst the key research findings are presented on section 11.2.2. The original scientific contributions are the following:

i. Utilization of high resolution traffic and weather data in urban roads.
ii. Specific research focus on Powered-Two Wheelers in urban roads and motorways.
iii. Development of an integrated multifaceted approach to model accident probability and severity.
iv. Introducing advanced methods of analysis in exploring high resolution traffic and weather data and in road safety.
v. Investigating the existence of non-linear relationships when analysing accident probability and severity.

The utilization of high resolution traffic and weather data in urban roads and the simultaneous co-consideration of Powered-Two-Wheelers, covered several gaps of knowledge, as indicated by the extensive literature review that was conducted. Due to the fact that the large majority of similar studies considered freeway data, it was needed to investigate urban environments. This is considered of great importance, since traffic and safety dynamics in freeway and urban environments are very different. For example, in urban environments the road users are more vulnerable to interactions with other motorized traffic and of course the presence of intersections plays a critical role. The focus on PTWs in such studies is essential, because they are very vulnerable to interaction especially in urban environments. Therefore, the effect of flow conditions on PTW safety had to be investigated.

This thesis proposed **an innovative approach to investigate accident probability and severity in urban roads and motorways with the use of differently oriented advanced modelling approaches** in order to acquire the larger picture of the accident severity and probability phenomena. For that reason, several probability and severity definitions were used. It was aimed to acquire the larger picture of accident severity and probability phenomena. Various data sources (e.g. real-time traffic data, real-time weather data and traditional accident data) have been obtained, processed and utilized. Although the core part of the thesis concerned urban roads, analyses of urban motorway data were also performed to complement the research design.

It is noted, that **some of the methods were applied for the first time** when such data are utilized (finite mixture logit, cusp catastrophe) or for the first time in road safety (rare-events logit model). Moreover, the time series data mining techniques through the combined application of original and transformed time series with Support Vector Machines, provided promising results and should be expanded in more relevant studies.

The mesoscopic accident analysis approach of this thesis, has a lot to contribute to the better understanding of the road accident phenomenon. To be more specific, this approach possesses some advantages over both macroscopic methods and the real-time microscopic data analysis, mainly because of the following reasons: 1) it enables the provision of sufficient time that allow authorities to develop a proactive safety management system without losing information of critical variables caused by large time interval measurement and 2) provides much more information than aggregate measures of traffic parameters (e.g. hourly traffic or annual average daily traffic). Modelling accident probability and severity in urban environments is a highly complex procedure and that fact should always be taken into serious consideration when analytical models are developed. The various statistical models that were developed can either predict or explain the accident probability and severity phenomena.

A methodological remark that derived from the analyses carried out in the thesis, **is the general superiority in terms of goodness of fit, of the non-linear modelling techniques** when accident severity and probability are examined. Although a number of linear models achieved to adequately describe the aforementioned phenomena, the cusp catastrophe models were proved to be considerably promising and fruitful. In a sense, these results may be considered as a first trial and a first step towards the incorporation of chaos theory in accident research. While one cannot definitely say that these methods outperform the traditional statistical analysis methods, it is without doubt that new directions are opened. It is very promising that the application of non-linear cusp catastrophe models, produced new original results and it is therefore suggested that more research should be conducted towards that direction in order to accurately predict the points of catastrophe.

A remark is worth of discussion. This concerns a number of similarities between cusp catastrophe and chaos theory, such as the presence of strong nonlinear relationships, the fact that control factors govern the system and the potential tremendous impact that small changes in control factors have on the system. Consequently, **these results may be used as a first step towards the incorporation of chaos theory in accident research**. While one cannot definitely say that these methods outperform the traditional statistical analysis methods, it is without doubt that new directions are opened.

Furthermore, it is **suggested that an advanced multi-faceted statistical analysis of accident probability and severity exploiting high resolution traffic and weather data, can be proved as a very useful tool** for accident and injury causation analysis, but also for support of real-time road safety decision making.

This PhD thesis aimed to unveil the influence of high resolution traffic and weather parameters on accident probability and severity. Despite the emphasis given on such kind of data, traditional accident information was also used to enrich the interpretability of models. Overall, the findings of the thesis suggest that **high resolution traffic and weather data are capable of opening new dimensions in road accident analysis in urban roads and motorways**. In addition, the combination of traffic and weather data leads to a clearer picture of the road accident phenomenon, in terms of both probability and severity.

The multi-faceted statistical analysis conducted in the thesis has revealed a **consistent and strong influence of traffic parameters on accident probability and severity**. This finding suggests that similar accident studies and investigations should always consider and incorporate the traffic conditions before the occurrence of an accident. If effective real-time measures are implemented then accident probability and accident severity will be reduced. Nevertheless, the **statistical significance of some other specific accident attributes**, such as accident type, suggests that more data should be utilized, as they provide useful and important information.

It is without doubt that accident probability and severity and are two entirely distinct phenomena, which were found to be influenced by a number of common and non-common parameters. Each phenomenon was found to have different characteristics for different types of vehicles (passenger cars, Powered-Two-Wheelers) involved in the accident, but this does not always happen. For example, **the effect of traffic states on overall accident severity was found to be similar to PTW accident severity** (i.e. accidents were a PTW is involved). However, the importance of separate models for PTWs was justified in the rest of the chapters.

In general, it was found that **traffic parameters have mixed effects on accident severity**. For example, speed and flow variations had different effect on accident severity, depending on the latent class to which the accident was assigned (see table 7.1). Similar findings were revealed when accident severity and PTW accident severity were explored in the urban motorway.

When urban roads are analysed, accident occurrence with PTWs could be a matter of the behavioural interaction of PTWs with other motorized traffic, rather than PTW errors. This may be attributed to the fact that high fluctuations in traffic flow and multi-vehicle collisions (except for rear-end collisions), were found to have a strong association with accidents involving a PTW. On the other hand, PTWs are less likely to be involved in single-vehicle accidents. However, in urban motorways, PTW accident involvement was found to be correlated only with traffic flow and not with accident type. More specifically, **a non-linear relationship between traffic flow and accident with PTWs was observed in urban motorways**.

It is interesting that **weather parameters were not found statistically significant when linear relationships are considered.** This trend was observed regardless of the analysis method, the dependent variable of interest (i.e. severity or probability) or the area type (i.e. urban or motorway). However, the cusp catastrophe models indicated a strong significant effect of a number of weather parameters on the asymmetry and/or bifurcation factors, which determine the transition of safe to unsafe regimes and vice versa.

Lastly, the development of cusp catastrophe models implied that it is likely that **even small traffic and weather changes may have a critical impact on safe and unsafe traffic conditions in urban roads**. This regards not only overall accident probability and accident severity, but PTWs as well.

For example, severe accidents could be very easily turned into slight accidents in the future (and vice versa). Therefore, it should be further investigated if traditionally linear relationships are not appropriate in investigating accident probability and accident severity. When following this approach, the assumption that a dynamic system exists is required.

Πολυ-επίπεδη στατιστική ανάλυση της πιθανότητας και της σοβαρότητας ατυχήματος αξιοποιώντας κυκλοφοριακά και μετεωρολογικά δεδομένα υψηλής ευκρίνειας

Εκτεταμένη Περίληψη

1. Εισαγωγή

Η οδική ασφάλεια αποτελεί ένα θέμα μείζονος σημασίας, καθώς τα οδικά ατυχήματα έχουν ιδιαίτερες επιπτώσεις τόσο στην οικονομική όσο και στην κοινωνική ζωή. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (WHO, 2013), ο συνολικός αριθμός των νεκρών σε οδικά ατυχήματα παγκοσμίως είναι 1.24 εκ. ετησίως. Το 2013, ο αριθμός των νεκρών στην Ε.Ε. ήταν 25,900, ενώ ο αριθμός των σοβαρά τραυματιών ανήλθε σε 313,000 (ETSC, 2013). Σε ό,τι αφορά στους δικυκλιστές, το 18% του συνολικού αριθμού των νεκρών στην Ε.Ε. είναι δικυκλιστές (ERSO, 2011), ενώ στην Ελλάδα το αντίστοιχο ποσοστό για το 2012 ήταν 29% (ELSTAT, 2012). Συνεπώς, η έρευνα με σκοπό τη βελτίωση της οδικής ασφάλειας, ιδιαιτέρως σε ό,τι αφορά στους ευάλωτους χρήστες της οδού είναι αναγκαία και επιτακτική.

Επίσης, πρέπει να τονιστεί πως η οδική ασφάλεια είναι ένας ιδιαιτέρως πολύπλοκος επιστημονικός τομέας. Για το λόγο αυτό, η ευρεία χρήση παραδοσιακών μεθόδων και προσεγγίσεων ανάλυσης, πιθανώς να μην είναι ο καταλληλότερος τρόπος για τον σωστό εντοπισμό των κρίσιμων παραγόντων κινδύνου μέσα από τον οποίο θα πραγματοποιηθεί μια εις βάθος κατανόηση κρίσιμων ζητημάτων οδικής ασφάλειας όπως η πιθανότητα και η σοβαρότητα των ατυχημάτων.

Μέσα σε μια πληθώρα εν δυνάμει κρίσιμων παραμέτρων, ιδιαίτερη έμφαση έχει δοθεί τα τελευταία χρόνια στη διερεύνηση της επιρροής κυκλοφοριακών και μετεωρολογικών παραμέτρων στα ατυχήματα. Αυτή η σύγχρονη τάση οφείλεται στη δυνατότητα που υπάρχει πλέον για εύκολη πρόσβαση σε αξιόπιστα κυκλοφοριακά και μετεωρολογικά δεδομένα υψηλής ευκρίνειας (high resolution). Αυτά τα δεδομένα παράγονται και συλλέγονται σε πραγματικό χρόνο (real-time data) και προσφέρουν περισσότερες πληροφορίες στους ερευνητές και στους συγκοινωνιολόγους μηχανικούς, καθώς επιτρέπουν μικροσκοπικές αναλύσεις αποφεύγοντας τη χρήση μακροσκοπικών δεικτών όπως για παράδειγμα η ετήσια μέση ημερήσια κυκλοφορία (ΕΜΗΚ), η μηνιαία βροχόπτωση κλπ.

Στο πλαίσιο αυτό, ο στόχος της παρούσας διδακτορικής διατριβής είναι η πολυεπίπεδη στατιστική ανάλυση της πιθανότητας και της σοβαρότητας ατυχήματος αξιοποιώντας κυκλοφοριακά και μετεωρολογικά δεδομένα υψηλής ευκρίνειας. Ιδιαίτερη έμφαση δόθηκε στους δικυκλιστές μέσω της ανάπτυξης ξεχωριστών στατιστικών μοντέλων για τη διερεύνηση της πιθανότητας εμπλοκής δικυκλιστών σε ατύχημα καθώς και της σοβαρότητας ατυχήματος που εμπλέκεται δίκυκλο. Για το σκοπό αυτό, αντλήθηκαν δεδομένα ατυχημάτων τόσο από αστικές οδούς αστικών περιοχών, όσο και από αστικούς αυτοκινητοδρόμους. Διερευνάται η επίδραση των κυκλοφοριακών και μετεωρολογικών παραμέτρων αλλά και άλλων πιθανών παραγόντων. Αξίζει να σημειωθεί ότι εξετάζονται διαφορετικοί ορισμοί της πιθανότητας και της σοβαρότητας ατυχήματος, προκειμένου να επιτευχθεί μια πιο ευρεία κατανόηση του φαινομένου και των υποκείμενων παραγόντων κινδύνου. Επιπροσθέτως, όπου είναι δυνατό, γίνεται εφαρμογή εναλλακτικών μεθοδολογιών για τη βελτίωση της κατανόησης των ζητημάτων της πιθανότητας και της σοβαρότητας ατυχήματος.

Για να επιτευχθεί ο στόχος της παρούσας διδακτορικής διατριβής, μια σειρά από ερευνητικές δραστηριότητες έλαβαν χώρα:

1) Εκτεταμένη βιβλιογραφική ανασκόπηση
2) Συλλογή και επεξεργασία δεδομένων
3) Στατιστική ανάλυση της πιθανότητας ατυχήματος
4) Στατιστική ανάλυση της σοβαρότητας ατυχήματος
5) Έμφαση στους δικυκλιστές

2. Βιβλιογραφική ανασκόπηση

Στο πλαίσιο της πρώτης ερευνητικής δραστηριότητας, πραγματοποιήθηκε εκτεταμένη βιβλιογραφική ανασκόπηση, η οποία εστίαζε τόσο στην επιρροή των κυκλοφοριακών και μετεωρολογικών παραμέτρων στην οδική ασφάλεια και δεύτερον, όσο και στους κρίσιμους παράγοντες ασφάλειας των δικυκλιστών. Αρχικά, εξετάστηκαν οι συναφείς επιστημονικές εργασίες σχετικά με τους κυκλοφοριακούς και μετεωρολογικούς παράγοντες και την επιρροή τους στην πιθανότητα, τη σοβαρότητα και τη συχνότητα των ατυχημάτων.

Οι προαναφερθείσες εργασίες χωρίζονται μεθοδολογικά σε δύο υποκατηγορίες: στις έρευνες οι οποίες ακολουθούν μια μακροσκοπική προσέγγιση ανάλυσης καθώς και στις πιο πρόσφατες οι οποίες ακολουθούν μικροσκοπική προσέγγιση ανάλυσης, αξιοποιώντας υψηλής ευκρίνειας κυκλοφοριακά και μετεωρολογικά δεδομένα τα οποία έχουν συλλεχθεί σε πραγματικό χρόνο. Στη συνέχεια εξετάστηκαν οι επιστημονικές εργασίες οι οποίες εστίαζαν στους χρήστες δικύκλων και οι οποίες διερευνούσαν τη συμπεριφορά των δικυκλιστών, την αλληλεπίδρασή τους με τους άλλους χρήστες της οδού, τη συχνότητα και τη σοβαρότητα των ατυχημάτων με δικυκλιστές, καθώς η παρούσα διδακτορική διατριβή δίνει έμφαση και στους χρήστες δικύκλων.

Η βιβλιογραφική ανασκόπηση ανέδειξε εν δυνάμει κρίσιμες παραμέτρους οι οποίες παίζουν σημαντικό ρόλο στην οδική ασφάλεια και εντοπίστηκαν κενά γνώσης:

- Σε πολλές περιπτώσεις ύπαρξη αντιφατικών αποτελεσμάτων σχετικά με την επιρροή του κυκλοφοριακών παραμέτρων στα ατυχήματα.
- Ανάγκη να αξιοποιηθούν υψηλής ευκρίνειας κυκλοφοριακά και μετεωρολογικά δεδομένα σε αστικές οδούς αστικών περιοχών (Wang et al., 2013), καθώς όλες οι σχετικές έρευνες αφορούν αυτοκινητοδρόμους.
- Ύπαρξη πολύ περιορισμένου αριθμού σχετικών εργασιών ιδιαίτερα σε σχέση με τη σοβαρότητα ατυχήματος (Theofilatos and Yannis, 2014).
- Έλλειψη συναφών εργασιών που να εξετάζουν την επιρροή των κυκλοφοριακών και μετεωρολογικών παραμέτρων στους δικυκλιστές.
- Ανάγκη για εφαρμογή πολυ-επίπεδης στατιστικής ανάλυσης για πληρέστερη κατανόηση του φαινομένου.
- Ανάγκη να εξεταστούν διάφορες εναλλακτικές μεθοδολογίες (Frazier and Kockelman, 2004; Christoforou et al., 2010).

3. Συλλογή και επεξεργασία δεδομένων

Η δεύτερη ερευνητική δραστηριότητα αφορά στη συλλογή και την επεξεργασία των δεδομένων ατυχημάτων, όπως και των κυκλοφοριακών και μετεωρολογικών δεδομένων, για το χρονικό διάστημα 2006-2011. Τα στοιχεία αφορούσαν κατά βάση αστικές οδούς αστικών περιοχών (Λ. Κηφισίας και Λ. Μεσογείων) όπως επίσης και στην Αττική οδό.

Τα δεδομένα οδικών ατυχημάτων συλλέχθηκαν από τη βάση δεδομένων ΣΑΝΤΡΑ που διατηρεί ο Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής της Σχολής Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου ΕΜΠ, και η οποία βασίζεται στα δεδομένα που καταγράφονται από την Τροχαία και κωδικοποιούνται από την Ελληνική Στατιστική Αρχή. Τα δεδομένα αυτά περιελάμβαναν γενικά χαρακτηριστικά ατυχημάτων (τύπος ατυχήματος, τύπος οχήματος, συνθήκες φωτισμού, κατάσταση οδοστρώματος, κυβισμός οχημάτων κλπ.) αλλά και χαρακτηριστικά παθόντων (φύλο, ηλικία, εθνικότητα).

Τα κυκλοφοριακά δεδομένα συλλέχθηκαν από το Κέντρο Διαχείρισης Κυκλοφορίας της Αθήνας (ΚΔΚ), όπως και από το Κέντρο Διαχείρισης Κυκλοφορίας της Αττικής Οδού. Αντλήθηκαν δεδομένα από τους κοντινότερους ανιχνευτές κυκλοφορίας και εξετάστηκαν οι εξής παράμετροι κυκλοφορίας οι οποίες καταγράφονται σε πραγματικό χρόνο: κυκλοφοριακός φόρτος (αριθμός οχημάτων ανά πέντε λεπτά), κατάληψη (σε ποσοστό %), μέση ταχύτητα χρόνου (χιλιόμετρα ανά ώρα). Στα κυκλοφοριακά δεδομένα της Αττικής Οδού περιλαμβάνεται και το ποσοστό βαρέων οχημάτων στην κυκλοφορία (σε ποσοστό %).

Τα μετεωρολογικά δεδομένα αντλήθηκαν από το Υδρολογικό Παρατηρητήριο της Αθήνας που αναπτύχθηκε από τον Τομέα Υδρολογίας και Υδάτινων Πόρων του Εθνικού Μετσόβιου Πολυτεχνείου, και περιλαμβάνει 10 πλήρως αυτοματοποιημένους μετεωρολογικούς σταθμούς οι οποίοι ανήκουν στο δίκτυο METEONET. Τα μετεωρολογικά δεδομένα συλλέχθηκαν από τους κοντινότερους μετεωρολογικούς σταθμούς και περιλαμβάνουν τις εξής παραμέτρους με τις αντίστοιχες μονάδες μέτρησης: θερμοκρασία (σε βαθμούς κελσίου), ύψος βροχόπτωσης (σε χιλιοστά), σχετική υγρασία (σε ποσοστό %), ταχύτητα ανέμου (σε μέτρα ανά δευτερόλεπτο), διεύθυνση ανέμου (σε μοίρες) και ηλιακή ακτινοβολία (σε watt ανά τετραγωνικό μέτρο).

Σημειώνεται ότι με βάση τη βιβλιογραφία, έγινε συλλογή κυκλοφοριακών και μετεωρολογικών δεδομένων που αφορούσαν χρονικές περιόδους εντός και εκτός της περιόδου των ατυχημάτων, προκειμένου να μελετηθεί η πιθανότητα ατυχήματος (Abdel-Aty and Pande, 2005; Abdel-Aty et al., 2007; Ahmed and Abdel-Aty, 2012; Yu and Abdel-Aty, 2013a). Μετά τη συλλογή των δεδομένων ακολούθησε η προκαταρκτική επεξεργασία των δεδομένων προτού δημιουργηθούν οι τελικές βάσεις δεδομένων και εφαρμοστεί η πολυεπίπεδη στατιστική ανάλυση της διδακτορικής διατριβής. Στο σημείο αυτό τυχόν λανθασμένες ή μη ρεαλιστικές μετρήσεις αφαιρούνταν (πχ. αρνητικές τιμές κατάληψης) από τη βάση δεδομένων. Στην τελική επεξεργασία των δεδομένων έλαβε χώρα μια μεσοπρόθεσμου χαρακτήρα προσέγγιση. Πιο συγκεκριμένα, έγινε αναγωγή των αρχικών 5-λεπτων και 10-λεπτων μετρήσεων των κυκλοφοριακών και μετεωρολογικών δεδομένων αντίστοιχα, σε επίπεδο ώρας.

Έτσι υπολογίστηκαν τα παρακάτω στατιστικά μεγέθη:

- μέση τιμή
- τυπική απόκλιση
- διάμεσος
- συντελεστής μεταβλητότητας (τυπική απόκλιση προς τη μέση τιμή).

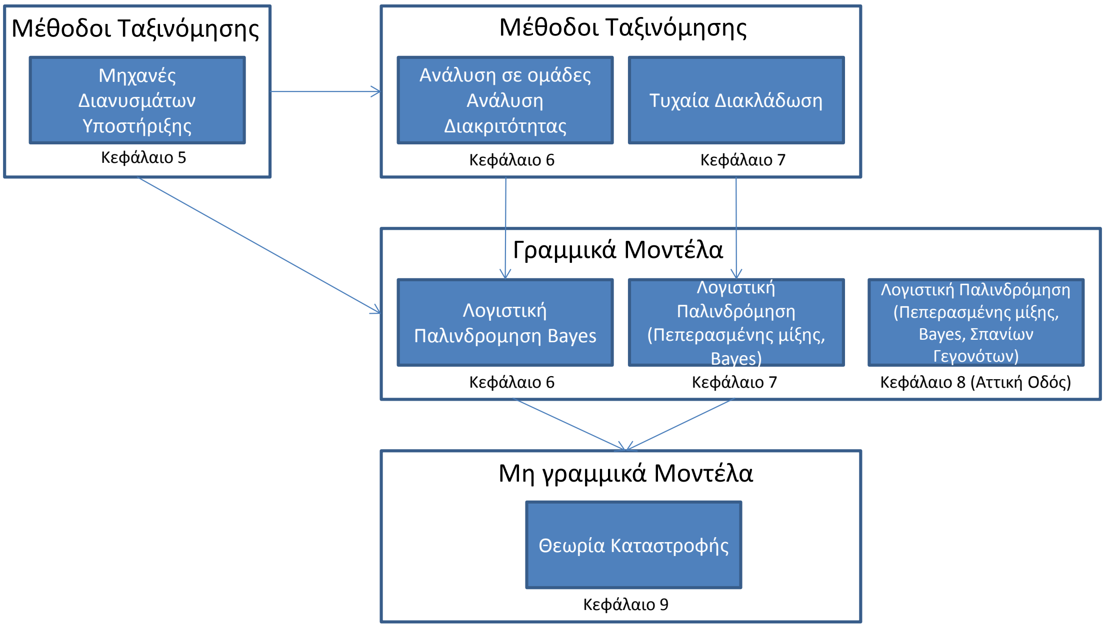
Ο σκοπός της αναγωγής των αρχικών δεδομένων μέσα στην ώρα αλλά και ο υπολογισμός των παραπάνω στατιστικών παραμέτρων, πραγματοποιήθηκε διότι τόσο οι κυκλοφοριακές όσο και οι μετεωρολογικές παράμετροι είναι δυναμικά μεγέθη. Έτσι ως δυναμικά μεγέθη ήταν επιθυμητό να μελετηθεί η χρονική τους εξέλιξη και επιρροή και όχι η σημειακή. Επίσης, η τυπική απόκλιση και ο συντελεστής μεταβλητότητας εκφράζουν τη διακύμανση. Για παράδειγμα, η τυπική απόκλιση του κυκλοφοριακού φόρτου εκφράζει τη διακύμανση των 5-λεπτων του κυκλοφοριακού φόρτου στην ώρα. Σημειώνεται επίσης, πως η τυπική απόκλιση του ύψους βροχής προσεγγίζει την ένταση της βροχόπτωσης. Έτσι, υψηλές τιμές τυπικής απόκλισης καταδεικνύει την πιθανή ύπαρξη έντονης βροχόπτωσης.

Από την τελική επεξεργασία των δεδομένων προέκυψαν συνολικά:

- 527 ατυχήματα με παθόντες σε Λ. Κηφισίας και Λ. Μεσογείων
- 326 ατυχήματα με δικυκλιστές σε Λ. Κηφισίας και Λ. Μεσογείων
- 285 ατυχήματα με παθόντες στην Αττική οδό
- 387 παθόντες στην Αττική οδό
- 141 ατυχήματα με δικυκλιστές στην Αττική οδό

4. Μεθοδολογική προσέγγιση

Προκειμένου να επιτευχθεί ο στόχος της διδακτορικής διατριβής μέσω των προαναφερθεισών ερευνητικών δραστηριοτήτων, μια σειρά από στατιστικές αναλύσεις έλαβαν χώρα στα αντίστοιχα κεφάλαια (Κεφάλαια 5-9). Οι παρακάτω στατιστικές αναλύσεις εντάσσονται στο πλαίσιο μιας πολυεπίπεδης στατιστικής ανάλυσης, η οποία εφαρμόζει έναν αριθμό διαφορετικών στατιστικών μοντέλων, τα οποία έχοντας άλλους επιμέρους στόχους, επιτελούν διαφορετικό ρόλο στην συνολικότερη κατανόηση των ζητημάτων της πιθανότητας και της σοβαρότητας ατυχήματος. Στο Γράφημα 1 απεικονίζεται η συνολική μεθοδολογική προσέγγιση της διδακτορικής διατριβής, ενώ στο Γράφημα 2 απεικονίζεται η πολυεπίπεδη στατιστική ανάλυσης.



Γράφημα 1: Επισκόπηση της μεθοδολογικής προσέγγισης της διατριβής.

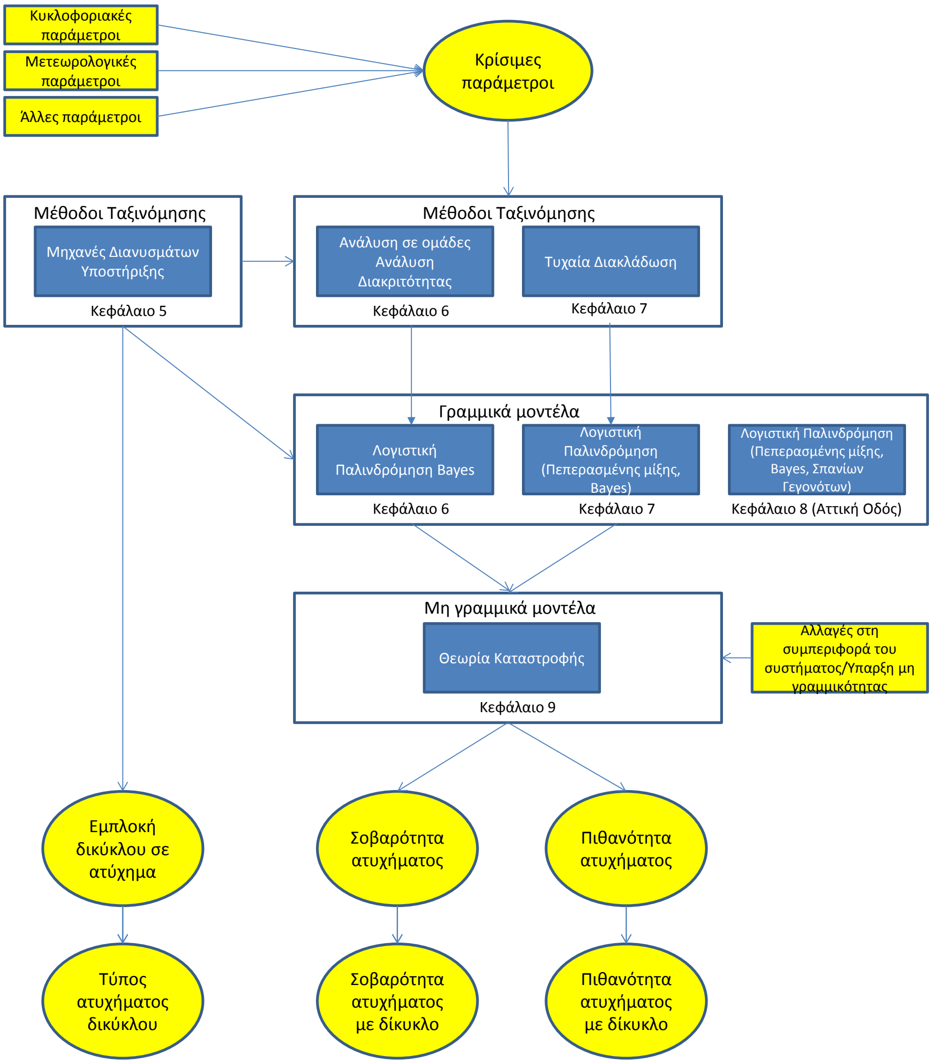
a) Συνδυαστική εφαρμογή χρονοσειρών (time series) και μηχανών διανυσμάτων υποστήριξης (Support Vector Machines). Αυτό το κεφάλαιο είναι αφιερωμένο αποκλειστικά στους δικυκλιστές, όπου διερευνάται η πιθανότητα ατυχημάτων με δίκυκλα και ο τύπος ατυχήματος με τη χρήση μιας συνδυασμένης προσέγγισης (Zhao, 2012). Σε αυτή την προσέγγιση, οι μηχανές διανυσμάτων υποστήριξης αναπτύχθηκαν μέσα από την αξιοποίηση τόσο των πρωτότυπων όσο και μετασχηματισμένων χρονοσειρών του κυκλοφοριακού φόρτου, της ταχύτητας και της κατάληψης (Κεφάλαιο 5).

b) Πεπερασμένης μίξης ανάλυση σε ομάδες (finite mixture cluster analysis), ανάλυση διακριτότητας (discriminant analysis) με σκοπό να καθοριστούν κυκλοφοριακές καταστάσεις (traffic states) καθώς και λογιστικό μοντέλο παλινδρόμησης Bayes, ώστε να διερευνηθεί η επιρροή τους στην πιθανότητα και τη σοβαρότητα ατυχήματος. Πιο συγκεκριμένα, το κεφάλαιο αυτό έχει στόχο να δημιουργήσει σημαντικές ομάδες των παρατηρήσεων σύμφωνα με τα χαρακτηριστικά της κυκλοφορίας (κυκλοφοριακές καταστάσεις) με βάση τις μετρήσεις κατάληψης. Στη συνέχεια, η ανάλυση διακριτότητας κατανέμει σωστά τυχόν νέες παρατηρήσεις στις υπάρχουσες ομάδες και τρόπον τινά δρα συμπληρωματικά στην ανάλυση σε ομάδες. Τέλος, διερευνάται η επιρροή των κυκλοφοριακών καταστάσεων στην πιθανότητα ατυχήματος και σοβαρότητα ατυχήματος, με ανάπτυξη ξεχωριστών μοντέλων ειδικά για τους χρήστες των δικύκλων (Κεφάλαιο 6).

c) Διερεύνηση της επιρροής των επιμέρους κυκλοφοριακών και μετεωρολογικών παραμέτρων στην πιθανότητα και τη σοβαρότητα ατυχήματος μέσω εφαρμογής μοντέλων τυχαίας διακλάδωσης (Random Forests) και πεπερασμένης μίξης λογιστικής παλινδρόμησης και λογιστικής παλινδρόμησης Bayes. Τα μοντέλα τυχαίας διακλάδωσης εφαρμόζονται αρχικώς για να καταδείξουν εν δυνάμει σημαντικές μεταβλητές, μέσα από ένα μεγάλο πλήθος ανεξαρτήτων μεταβλητών, χωρίς όμως να ποσοτικοποιήσουν την επιρροή τους. Στη συνέχεια, οι σημαντικές μεταβλητές χρησιμοποιούνται στα παραπάνω λογιστικά μοντέλα προκειμένου να ποσοτικοποιηθεί η επιρροή τους στην πιθανότητα ατυχήματος και σοβαρότητα. (Κεφάλαιο 7).

d) Πεπερασμένη μίξης λογιστική παλινδρόμηση, λογιστική παλινδρόμηση Bayes και λογιστική παλινδρόμηση σπανίων γεγονότων (rare-events), με σκοπό τη διερεύνηση της πιθανότητας και τη σοβαρότητας ατυχήματος στην Αττική οδό. Συγκεκριμένα, το λογιστικό μοντέλο σπανίων γεγονότων (King and Zeng, 2001), εφαρμόστηκε ώστε να διερευνήσει την πιθανότητα ατυχήματος όταν ατυχήματα θεωρούνται σπάνια γεγονότα (Κεφάλαιο 8).

e) Εφαρμογή θεωρίας καταστροφής (catastrophe theory) στη διερεύνηση της πιθανότητας και τη σοβαρότητας ατυχήματος στις αστικές οδούς. Τα μοντέλα καταστροφής υποθέτουν την ύπαρξη ενός δυναμικού συστήματος και εξετάζουν την ύπαρξη μη γραμμικών σχέσεων, και πιο συγκεκριμένα την ύπαρξη ξαφνικών μεταβάσεων από τη μια κατάσταση (ασφαλής κατάσταση) στην άλλη (μη ασφαλής κατάσταση) και το αντίστροφο. Στη διδακτορική διατριβή, ως ασφαλείς καταστάσεις ορίζονται η μη ύπαρξη ατυχήματος, η χαμηλή σοβαρότητα και η μη εμπλοκή δικυκλιστών σε ατύχημα.



Γράφημα 2: Επισκόπηση της πολυεπίπεδης στατιστικής ανάλυσης.

5. Ανάπτυξη στατιστικών μοντέλων

Η συνδυασμένη ανάλυση χρονοσειρών και μηχανών διανυσμάτων υποστήριξης είναι μια πρόσφατη μέθοδος πρόβλεψης η οποία χρησιμοποιείται για πρώτη φορά στην οδική ασφάλεια στα πλαίσια αυτής της διδακτορικής διατριβής. Ο Πίνακας 1 παρουσιάζει τα συνολικά αποτελέσματα πρόβλεψης της εμπλοκής δικυκλιστών σε ατύχημα με χρήση σταυρωτής επικύρωσης (10-fold cross validation). Για πρώτη φορά η εφαρμογή μηχανών διανυσμάτων υποστήριξης έγινε με βάση τις βέλτιστες τιμές των παραμέτρων του μοντέλου (C, γ). Γενικά τα μοντέλα δίνουν ικανοποιητικά ποσοστά πρόβλεψη της εμπλοκής δικυκλιστών σε ατύχημα.

Total SVM performance	Speed downstream	Speed upstream	Flow downstream	Flow upstream	Occupancy downstream	Occupancy upstream
Original Time series						
10-fold cross validation mean accurracy %	62.17%	65.64%	61.90%	64.15%	60.72%	65.38%
(C, γ)	(100, 0.002)	(10, 0.01)	(100, 0.001)	(100, 0.001)	(100, 0.0001)	(100, 0.001)
DWT Time series						
10-fold cross validation mean accurracy %	60.65%	58.38%	60.82%	61.39%	59.20%	61.54%
(C, γ)	(10, 0.001)	(100, 0.01)	(100, 0.002)	(100, 0.0001)	(100, 0.01)	(100, 0.0001)

Πίνακας 1: Συνολικά αποτελέσματα μηχανών διανυσμάτων υποστήριξης.

Για τον καθορισμό κυκλοφοριακών καταστάσεων με βάση την κατάληψη (Xu et al., 2012) εφαρμόστηκε η ανάλυση κατά ομάδες (περίοδος ατυχημάτων). Ο βέλτιστος αριθμός ομάδων καθορίστηκε σε 5 σύμφωνα με το κριτήριο BIC (Bayesian Information Criterion).

Traffic states	State 1	State 2	State 3	State 4	State 5
Percentage of total cases (%)	26%	25.24%	12.71%	23.72%	12.33%
Average Occupancy upstream (%)	11.28	4.92	21.1	26.53	24.28
Average Occupancy downstream (%)	11.49	6.01	30.16	12.03	27.06
St.deviation of Occupancy upstream (%)	1.66	0.75	8.52	8.17	4.38
St.deviation of Occupancy downstream (%)	2.45	0.84	6.29	1.94	7.88

Πίνακας 2: Αποτελέσματα ανάλυσης κατά ομάδες για την περίοδο των ατυχημάτων.

Η ανάλυση διακριτότητας εφαρμόστηκε για τη σωστή κατανομή νέων παρατηρήσεων στις υπάρχουσες ομάδες. Το 80% των δεδομένων χρησιμοποιήθηκε για εκπαίδευση and 20% για “τεστ”.

Traffic states	Predicted group membership using discriminant analysis				
	State 1 (%)	State 2 (%)	State 3 (%)	State 4 (%)	State 5 (%)
State 1	**67.65%**	21.74%	0.00%	0.00%	0.00%
State 2	17.65%	**78.26%**	0.00%	0.00%	0.00%
State 3	0.00%	0.00%	**86.67%**	0.00%	0.00%
State 4	5.88%	0.00%	0.00%	**94.44%**	0.00%
State 5	8.82%	0.00%	13.33%	5.56%	**100.00%**

Πίνακας 3: Αποτελέσματα ανάλυσης διακριτότητας για την περίοδο των ατυχημάτων.

Η επιρροή των κυκλοφοριακών καταστάσεων στη σοβαρότητα διερευνήθηκε μέσω της λογιστικής παλινδρόμησης Bayes. Το κριτήριο σημαντικότητας μιας μεταβλητής είναι να έχει το ίδιο πρόσημο στο διάστημα εμπιστοσύνης (2.5%-97.5%). Τονίζεται ότι χρησιμοποιήθηκαν ασαφείς πρότερες κατανομές για τις παραμέτρους του μοντέλου (non-informative priors).

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-2.589	0.357	0.075	-3.353	-1.936
traffic state 1 (ref)	-	-	-	-	-
traffic state 2	1.377	0.419	3.963	0.5839	2.242
DIC	331				

Πίνακας 4: Αποτελέσματα λογιστικού μοντέλου Bayes για τη σοβαρότητα ατυχήματος.

Το αποτελέσματα του μοντέλου δείχνουν ότι η χαμηλότερη κυκλοφοριακή συμφόρηση (κατάσταση 2) σχετίζεται με αυξημένη σοβαρότητα ατυχήματος, καθότι η κυκλοφοριακή κατάσταση 1 αντιπροσωπεύει κυκλοφοριακή συμφόρηση (κατηγορία αναφοράς). Σε ό,τι αφορά στη διερεύνηση της επιρροής των κυκλοφοριακών στη σοβαρότητα ατυχήματος με δικυκλιστές, τα αποτελέσματα είναι αρκετά παρόμοια με αυτά το προηγούμενου μοντέλου για τη σοβαρότητα ατυχήματος όπως φαίνεται και από τον Πίνακα 5.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-2.561	0.4546	0.077	-3.545	-1.753
traffic state 1 (ref)	-	-	-	-	-
traffic state 2	1.095	0.5702	2.989	0.01806	2.257
DIC	176.52				

Πίνακας 5: Αποτελέσματα λογιστικού μοντέλου Bayes για τη σοβαρότητα ατυχήματος με δικυκλιστές.

Το μοντέλο της επιρροής των κυκλοφοριακών καταστάσεων στην πιθανότητα ατυχήματος με δικυκλιστές φαίνεται στον Πίνακα 7. Το αρνητικό πρόσημο της κυκλοφοριακής κατάστασης 2 δηλώνει ότι η χαμηλότερη κατάληψη μειώνει την πιθανότητα ατυχήματος με δίκυκλα. Στην κυκλοφοριακή κατάσταση 3 (υψηλότερη κατάληψη από την κυκλοφοριακή κατάσταση 1) αυξάνεται η πιθανότητα ατυχήματος με δίκυκλα.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	0.53	0.1864	1.699	0.1679	0.902
traffic state 1 (ref)	-	-	-	-	-
traffic state 2	-0.6467	0.2613	0.524	-1.165	-0.138
traffic state 3	0.8263	0.3671	2.285	0.1274	1.564
DIC	625.51				

Πίνακας 6: Αποτελέσματα λογιστικού μοντέλου Bayes για την πιθανότητα ατυχήματος με δικυκλιστές.

Για τη μελέτη της πιθανότητας ατυχήματος, η ανάλυση σε ομάδες περιελάμβανε τόσο την χρονική περίοδο των ατυχημάτων (accident cases) όσο και την χρονική περίοδο εκτός περιόδου ατυχημάτων (non-accident cases). Ο βέλτιστος αριθμός ομάδων βρέθηκε ίσος με 9 (Πίνακας 8), ενώ η ανάλυση διακριτότητας πραγματοποιήθηκε με παρόμοιο τρόπο έχοντας πολύ ικανοποιητικά αποτελέσματα (Πίνακας 7).

Traffic states	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
Percentage of cases (%)	12.02%	21.38%	17.43%	11.30%	10.28%	1.90%	6.45%	9.49%	9.74%
Average Occupancy upstream (%)	10.81	4.35	10.52	23.52	18.89	30.82	28.59	12.87	29.74
Average Occupancy downstream (%)	14.15	5.45	7.95	11.24	14.89	6.07	34.78	27.9	24.53
St.deviation of Occupancy upstream (%)	1.32	0.67	1.45	3.75	7.86	11.81	10.41	1.57	5.67
St.deviation of Occupancy downstream (%)	2.82	0.76	1.03	1.18	3.99	1.44	9.55	6.69	4.15

Πίνακας 7: Αποτελέσματα ανάλυσης κατά ομάδες για την περίοδο εντός και εκτός των ατυχημάτων.

Traffic states	Predicted group membership using discriminant analysis								
	State 1 (%)	State 2 (%)	State 3 (%)	State 4 (%)	State 5 (%)	State 6 (%)	State 7 (%)	State 8 (%)	State 9 (%)
State 1	**78.25%**	0.00%	13.79%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State 2	0.00%	**74.39%**	10.34%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State 3	4.35%	25.61%	**58.63%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State 4	0.00%	0.00%	12.07%	**92.86%**	4.35%	0.00%	0.00%	0.00%	0.00%
State 5	8.70%	0.00%	5.17%	0.00%	**82.61%**	33.33%	0.00%	6.67%	0.00%
State 6	0.00%	0.00%	0.00%	0.00%	0.00%	**66.67%**	0.00%	0.00%	0.00%
State 7	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**92.86%**	6.67%	9.09%
State 8	8.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**80.00%**	0.00%
State 9	0.00%	0.00%	0.00%	7.14%	13.04%	0.00%	7.14%	6.67%	**90.91%**

Πίνακας 8: Αποτελέσματα ανάλυσης διακριτότητας για την πιθανότητα ατυχήματος.

Τα αποτελέσματα του λογιστικού μοντέλου Bayes που αφορά στην πιθανότητα ατυχήματος παρουσιάζονται συνοπτικά στον Πίνακα 9. Τα θετικά πρόσημα των κυκλοφοριακών καταστάσεων 5 και 6 δηλώνουν θετική συσχέτιση αυξημένων τιμών κατάληψης και πιθανότητας ατυχήματος σε αστικές οδούς.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-0.7932	0.1643	0.452	-1.121	-0.4779
traffic state 1 (ref)	-	-	-	-	-
traffic state 5	0.5862	0.2334	1.797	0.1323	1.047
traffic state 6	1.183	0.4332	3.264	0.3482	2.047
DIC	1821.37				

Πίνακας 9: Αποτελέσματα λογιστικού μοντέλου Bayes για την πιθανότητα ατυχήματος.

Στο επόμενο κεφάλαιο της διδακτορικής διατριβής (Κεφάλαιο 7) εξετάστηκε η επιρροή των κυκλοφοριακών και μετεωρολογικών παραμέτρων μέσω χρήσης γραμμικών στατιστικών μοντέλων. Αρχικώς, τα μοντέλα "τυχαίας διακλάδωσης" αποκάλυψαν εν δυνάμει κρίσιμες παραμέτρους (πχ. τύπος ατυχήματος, μέσος κυκλοφοριακός φόρτος, συντελεστής διακύμανσης του φόρτου, συντελεστής διακύμανσης της ταχύτητας). Μετά από περαιτέρω έλεγχο για συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών, αναπτύχθηκαν τα πεπερασμένης μίξης λογιστικά μοντέλα (Finite mixture logit models) με εφαρμογή στη σοβαρότητα ατυχήματος. Αυτή η μέθοδος ομαδοποιεί τις παρατηρήσεις και έτσι αντιμετωπίζει την ετερογένεια (unobserved heterogeneity). Τα πεπερασμένης μίξης λογιστικά μοντέλα ομαδοποίησαν τις παρατηρήσεις σε 2 λανθάνουσες ομάδες (latent classes). Ο Πίνακας 10 απεικονίζει συνοπτικά τα αποτελέσματα της λογιστικής ανάλυσης πεπερασμένης μίξης για τη σοβαρότητα ατυχήματος. Βρέθηκε ετερογενής επιρροή των κυκλοφοριακών παραμέτρων (ανάλογα με τη λανθάνουσα ομάδα).

Variables	Latent Class 1			Latent Class 2		
	Mean	t-statistic	p-value	Mean	t-statistic	p-value
Constant (random)	-1.823	-1.904	0.057	-0.402	-0.330	0.741
Acc.type0 (reference cat.)	-	-	-	-	-	-
Acc.type1 (fixed)	-1.099	-1.539	0.124	-1.099	-1.539	0.124
Acc.type2 (fixed)	-1.483	-2.246	0.025	-1.483	-2.246	0.025
Acc.type3 (fixed)	-16.942	-0.014	0.988	-16.942	-0.014	0.988
Acc.type4 (fixed)	-1.152	-2.264	0.023	-1.152	-2.264	0.023
Q_cv_1h_up (random)	10.936	1.977	0.048	-5.142	-0.619	0.536
MC.involvement.no (random)	-1.359	-1.733	0.083	1.289	1.312	0.190
V_cv_1h_up (random)	-0.387	-0.165	0.869	-32.089	-2.090	0.037
Log-likelihood at zero	-188.400					
Final Log-likelihood	-151.328					
Likelihood ratio test	74.144					
McFadden R^2	0.197					

Πίνακας 10: Αποτελέσματα της λογιστικής παλινδρόμησης πεπερασμένης μίξης για τη σοβαρότητα ατυχήματος σε αστικές οδούς.

Τα αποτελέσματα της λογιστικής παλινδρόμησης Bayes για τη μελέτη της πιθανότητας ατυχήματος φαίνονται στον Πίνακα 11, όπου βρέθηκε θετική συσχέτιση της διακύμανσης του φόρτου και της διακύμανσης της κατάληψης με την πιθανότητα ατυχήματος.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant (n.s.)	-0.328	0.2183	-	-0.7425	0.1074
Occ_stdev_1h_up	0.055	0.014	1.057	0.028	0.082
log(Q_cv_1h_up)	0.217	0.081	1.242	0.069	0.380
DIC	1980.64				

Πίνακας 11: Αποτελέσματα λογιστικού μοντέλου Bayes για την πιθανότητα ατυχήματος σε αστικές οδούς.

Σε ό,τι αφορά στην πιθανότητα εμπλοκής δικυκλιστών σε ατύχημα ο Πίνακας 12 δείχνει ότι η αύξηση του κυκλοφοριακού φόρτου και των διακυμάνσεων της ταχύτητας αυξάνουν την πιθανότητα ατυχήματος με δικυκλιστές σε αστικές οδούς αστικών περιοχών. Επιπροσθέτων, οι δικυκλιστές είναι πιθανότερο να εμπλακούν σε ατυχήματα με 2 ή περισσότερα οχήματα κάτι το οποίο υποδηλώνει την σημαντική αλληλεπίδραση με τους υπόλοιπους χρήστες της οδού.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-1.321	0.295	0.267	-1.905	-0.740
Q_avg_1h_up	0.001	0.000	1.001	0.001	0.002
V_cv_1h_up	1.249	0.603	3.487	0.077	2.441
Acc.type0 (ref)	-	-	-	-	-
Acc.type1	2.046	0.479	7.737	1.165	3.053
Acc.type2	0.435	0.260	-	-0.071	0.948
Acc.type3	2.021	0.417	7.546	1.244	2.888
Acc.type4	0.976	0.263	2.652	0.478	1.497
DIC	640.095				

Πίνακας 12: Αποτελέσματα λογιστικού μοντέλου Bayes για την πιθανότητα ατυχήματος με δικυκλιστές σε αστικές οδούς.

Η τελευταία μεγάλη κατηγορία γραμμικών στατιστικών μοντέλων που αναπτύχθηκαν στο πλαίσιο της πολυεπίπεδης στατιστικής ανάλυσης, αφορούσε στην Αττική Οδό. Στο κεφάλαιο αυτό της διδακτορικής διατριβής (Κεφάλαιο 9), εφαρμόστηκαν οι εξής στατιστικές μέθοδοι:

- Πεπερασμένης μίξης λογιστικά μοντέλα με σκοπό τη διερεύνηση της σοβαρότητας ατυχήματος.
- Λογιστική παλινδρόμηση Bayes με σκοπό τη διερεύνηση της πιθανότητας ατυχήματος με δικυκλιστές.
- Λογιστικά μοντέλα σπανίων γεγονότων (rare-events logit models) με σκοπό τη διερεύνηση της πιθανότητας ατυχήματος.

Η λογιστική παλινδρόμηση πεπερασμένης μίξης αποκάλυψε 2 λανθάνουσες ομάδες. Παρατηρείται διαφορετική επιρροή των κυκλοφοριακών παραμέτρων ανάλογα με τη λανθάνουσα ομάδα (Πίνακας 13).

Variables	Latent Class 1			Latent Class 2		
	Mean	t-statistic	p-value	Mean	t-statistic	p-value
Constant (random)	-1.720	-0.961	0.337	36.153	2.573	0.010
Acc.type0 (reference cat.)	-	-	-	-	-	-
Acc.type1 (fixed)	-2.338	-1.352	0.176	-2.338	-1.352	0.176
Acc.type2 (fixed)	-26.779	-2.262	0.024	-26.779	-2.262	0.024
Acc.type3 (fixed)	-7.196	-2.429	0.024	-7.196	-2.429	0.024
CC (fixed)	-0.002	-1.755	0.079	-0.002	-1.755	0.079
Tr.Prop_avg_30m_up (fixed)	-1.038	-2.429	0.015	-1.038	-2.429	0.015
Q_avg_30m_up (random)	0.027	1.362	0.173	-0.236	-2.354	0.019
Occ_stdev_30m_up (random)	154.060	2.597	0.009	-1629	-2.134	0.033
Log-likelihood at zero	-134.998					
Final Log-likelihood	-112.152					
Likelihood ratio test	45.692					
McFadden R^2	0.169					

Πίνακας 13: Αποτελέσματα της λογιστικής παλινδρόμησης πεπερασμένης μίξης για τη σοβαρότητα ατυχήματος στην Αττική οδό.

Σε ό,τι αφορά στη σοβαρότητα ατυχημάτων δικυκλιστών στην Αττική οδό η λογιστική παλινδρόμηση πεπερασμένης μίξης αποκάλυψε την ύπαρξη 2 λανθανουσών ομάδων σύμφωνα με το κριτήριο BIC. Και σε αυτή την περίπτωση διαπιστώθηκε ετερογενής επιρροή των κυκλοφοριακών παραμέτρων (μέσος κυκλοφοριακός φόρτος και συντελεστής μεταβλητότητας της ταχύτητας) στις δύο λανθάνουσες ομάδες.

Variables	Latent Class 1			Latent Class 2		
	Mean	t-statistic	p-value	Mean	t-statistic	p-value
Constant (random)	12.420	1.796	0.072	20.440	1.946	0.052
CC (fixed)	-0.009	-1.934	0.053	-0.009	-1.934	0.053
V_cv_30m_up (random)	16.812	1.106	0.037	-516.820	-2.081	0.037
Q_avg_30m_up (random)	-0.446	-1.647	0.099	-0.020	-0.602	0.547
Log-likelihood at zero	-64.212					
Final Log-likelihood	-40.927					
Likelihood ratio test	46.570					
McFadden R^2	0.363					

Πίνακας 14: Αποτελέσματα της λογιστικής παλινδρόμησης πεπερασμένης μίξης για τη σοβαρότητα ατυχήματος με δικυκλιστές στην Αττική οδό.

Η διερεύνηση της πιθανότητα ατυχήματος στην Αττική οδό έγινε με εφαρμογή του λογιστικού μοντέλου σπανίων γεγονότων, το οποίο εφαρμόστηκε για πρώτη φορά στην οδική ασφάλεια. Το λογιστικό μοντέλο σπανίων γεγονότων είναι μια πρωτότυπη και εναλλακτική μέθοδος η οποία θεωρεί τα ατυχήματα ως σπάνια γεγονότα. Αφότου ο ερευνητής πραγματοποιήσει στρωματοποιημένη δειγματοληψία για τον καθορισμό του στατιστικού δείγματος (stratified sampling), η ανάλυση αυτή εφαρμόζει κατάλληλες διορθώσεις σε σχέση με το κοινό λογιστικό μοντέλο.

Στην προκειμένη περίπτωση έγινε εξαγωγή δεδομένων από τρεις ανιχνευτές κυκλοφορίας με ίδιο αριθμό λωρίδων κυκλοφορίας, και πιο συγκεκριμένα τρεις. Ωριαίες μετρήσεις κυκλοφοριακών δεδομένων αντλήθηκαν για όλο το χρονικό διάστημα 2008-2011. Πραγματοποιήθηκαν τρεις δοκιμές για διαφορετικά δείγματα και παρήχθησαν παρόμοια αποτελέσματα (Πίνακας 15). Γενικά διαπιστώθηκε πως υπάρχει αρνητική σχέση μεταξύ του λογαρίθμου της μέσης ταχύτητας και της πιθανότητας ατυχήματος στην Αττική οδό. Αξίζει να σημειωθεί ότι σε παρόμοιο συμπέρασμα έχουν καταλήξει και πολύ πρόσφατες μελέτε της πιθανότητας ατυχήματος σε αυτοκινητοδρόμους (Ahmed et al., 2011 & 2012, Yu et al., 2013).

Trial 1	β	S.E.	z value	p-value	Odds Ratio
Constant	26.4158	11.3706	2.3232	0.0212	-
Truck.Prop.	-0.0394	0.1072	-0.3684	0.7129	-
log(Speed)	-7.4700	2.4369	-3.0653	0.0025	0.0006
Log-likelihood at zero		-113.9			
Final log-likelihood		-100.9			
Likelihood ratio test		26.0			
AIC		106.9			
McFadden R^2		0.1141			

Trial 2	β	S.E.	z value	p-value	Odds Ratio
Constant	33.2999	14.3741	2.3117	0.0216	-
Truck.Prop.	0.0157	0.0981	0.1597	0.8733	-
log(Speed)	-9.0004	3.0874	-2.9152	0.0039	0.0001
Log-likelihood at zero		-113.9			
Final log-likelihood		-100.6			
Likelihood ratio test		26.6			
AIC		106.6			
McFadden R^2		0.1168			

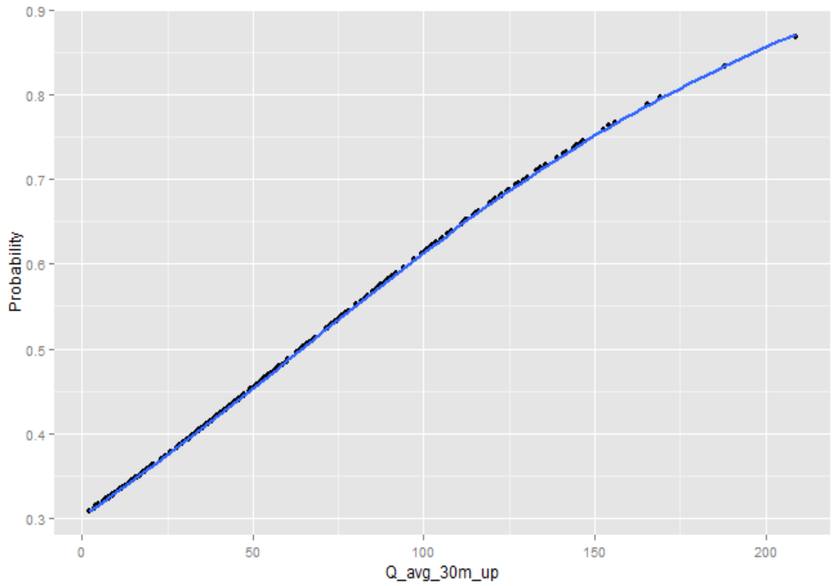
Trial 3	β	S.E.	z value	p-value	Odds Ratio
Constant	29.8363	12.6321	2.3619	0.0192	-
Truck.Prop.	-0.0444	0.0964	-0.4600	0.6460	-
log(Speed)	-8.2035	2.7063	-3.0311	0.0028	0.0003
Log-likelihood at zero		-113.9			
Final log-likelihood		-100.8			
Likelihood ratio test		26.2			
AIC		106.8			
McFadden R^2		0.1150			

Πίνακας 15: Αποτελέσματα της λογιστικής παλινδρόμησης πεπερασμένης μίξης για τη σοβαρότητα ατυχήματος με δικυκλιστές στην Αττική οδό.

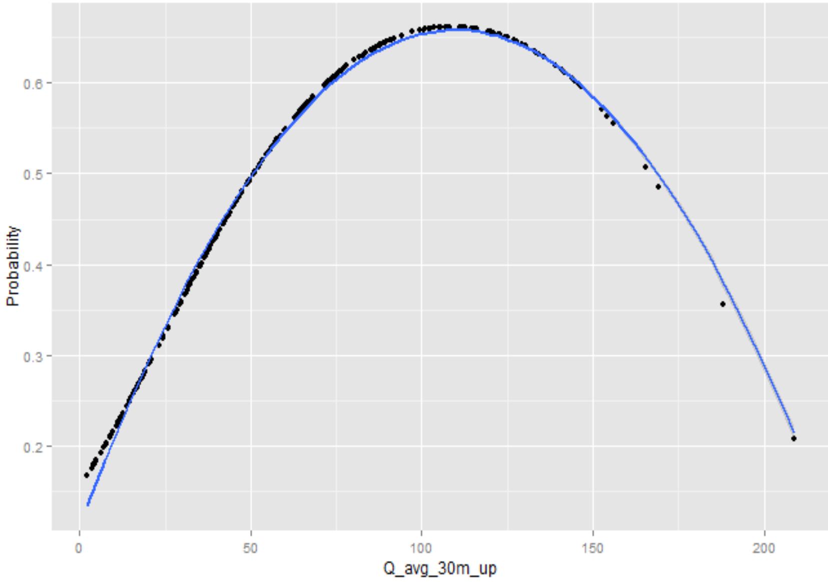
Η διερεύνηση της πιθανότητας ατυχήματος με δικυκλιστές πραγματοποιήθηκε με λογιστική παλινδρόμηση Bayes. Βρέθηκε τόσο γραμμική (θετική) όσο και μη γραμμική σχέση φόρτου-πιθανότητας ατυχήματος με δικυκλιστές με σχετικά καλύτερη προσαρμογή του μη γραμμικού μοντέλου. Ο Πίνακας 16 όπως και τα Γραφήματα 3 και 4, απεικονίζουν τα συνολικά αποτελέσματα της λογιστικής παλινδρόμησης Bayes.

Model1	Parameters Estimates			Credible Intervals		Model2	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%		Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-0.8357	0.2355	-	-1.2970	-0.3816	constant	-1.6950	0.3958	-	-2.5090	-0.9607
Q_avg_30m_up	0.0130	0.0033	1.0131	0.0068	0.0194	Q_avg_30m_up	0.0435	0.0114	1.0445	0.0223	0.0666
						$Q\_avg\_30m\_up^2$	-0.0002	0.0001	0.9998	-0.0003	-0.0001
DIC	376.926					DIC	367.84				

Πίνακας 16: Αποτελέσματα της λογιστικής παλινδρόμησης Bayes για την πιθανότητα ατυχήματος με δικυκλιστές στην Αττική οδό.



Γράφημα 3: Σχέση πιθανότητας ατυχήματος δικυκλιστών και κυκλοφοριακού φόρτου στην Αττική οδό για το γραμμικό λογιστικό μοντέλο Bayes.



Γράφημα 4: Σχέση πιθανότητας ατυχήματος δικυκλιστών και κυκλοφοριακού φόρτου στην Αττική οδό για το μη γραμμικό λογιστικό μοντέλο Bayes.

Τέλος, εφαρμόστηκε μια μεγάλη κατηγορία μη γραμμικών και μη παραδοσιακών μοντέλων και πιο συγκεκριμένα της θεωρία καταστροφής (catastrophe theory). Η θεωρία καταστροφής εξετάζει φαινόμενα που χαρακτηρίζονται από απότομες αλλαγές στη συμπεριφορά τους οι οποίες προέρχονται από μικρές αλλαγές στις παραμέτρους. Επειδή θεωρείται το σύστημα ως ντετερμινιστικό, εφαρμόζονται κατάλληλοι μαθηματικοί μετασχηματισμοί, έτσι ώστε να μπορεί να εφαρμοσθεί σε στοχαστικά συστήματα. Η θεωρία καταστροφής εφαρμόστηκε στις αστικές οδούς για τα διερεύνηση της πιθανότητας και σοβαρότητας ατυχήματος. Προκειμένου να κατανοηθεί σε βάθος το ζήτημα της σοβαρότητας, χρησιμοποιήθηκαν 2 ακόμη ορισμοί για τη σοβαρότητα:

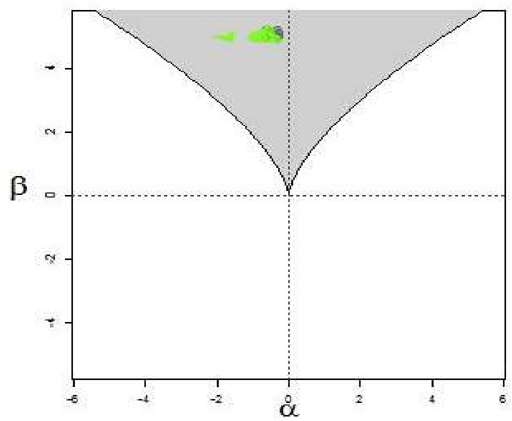
- αριθμός νεκρών και σοβαρά τραυματιών προς σύνολο παθόντων
- αριθμός παθόντων προς σύνολο εμπλεκομένων οχημάτων

Τα συνολικά αποτελέσματα του μοντέλου καταστροφής για τη σοβαρότητα ατυχήματος απεικονίζονται στον Πίνακα 17.

Cusp Catastrophe Models	
Catastrophe Model_1	**y** = - 2.32333 + 4.7048*Severity_1
	α = - 0.2899 - 0.6741*V_cv_1h_up - 0.3146*Acc.type1 - 0.2468*Acctype2 - 1.2944*Acc.type3 - 0.2851*Acc.type4
	β = 5.7148
pseudo R^2	0.9874
Linear model R^2	-
Logistic model R^2	0.0749
Catastrophe Model_2	**y** = - 2.3389 + 4.8234*Severity_2
	α = - 0.0006*Q_avg_1h_up - 0.2937*Acc.type2 - 1.2182*Acc.type3 -0.4165*Acc.type4 - 0.1481*MC.involvement.no + 0.0009*W.Dir_1h_avg
	β = 4.9843 + 0.2843*MC.involvement.no
pseudo R^2	0.8675
Linear model R^2	0.0938
Logistic model R^2	0.1576
Catastrophe Model_3	**y** = - 2.6870 + 1.446*Severity_3
	α = - 0.8214 - 0.9527*Acc.type1 - 2.7280*Acc.type2 - 3.6840*Acc.type3 - 3.0340*Acc.type4 - 0.0006*Sol_1h_max
	β = 1.0920 + 0.0003*Q_avg_1h_up + 0.0162*Rain_12h_sum
pseudo R^2	0.2282
Linear model R^2	0.2798
Logistic model R^2	0.2904

Πίνακας 17: Αποτελέσματα θεωρίας καταστροφής για τη σοβαρότητα ατυχήματος σε αστικές οδούς.

Τα αποτελέσματα του μοντέλου καταστροφής έδειξαν πολύ καλή προσαρμογή του μοντέλου. Στη δισδιάστατη απεικόνιση της επιφάνειας καταστροφής (Παράδειγμα μοντέλου 1) δείχνει την περιοχή διακλάδωσης (κρίσιμη περιοχή αστάθειας). Φαίνεται ότι όλες οι παρατηρήσεις βρίσκονται μέσα στην περιοχή αστάθειας. Συνεπώς, μικρές αλλαγές στους παράγοντες α, β οδηγεί σε ξαφνικές μεταβάσεις από ασφαλείς (χαμηλή σοβαρότητα) σε μη ασφαλείς καταστάσεις (υψηλή σοβαρότητα) και το αντίστροφο.

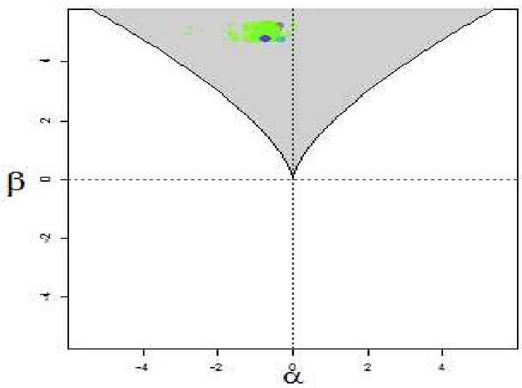


Γράφημα 5: Δισδιάστατη απεικόνιση της επιφάνειας καταστροφής της σοβαρότητας ατυχήματος σε αστικές οδούς (Παράδειγμα μοντέλου 1).

Με παρόμοιο τρόπο εξετάστηκε και η σοβαρότητα ατυχήματος με δικυκλιστές (Πίνακας 18).

Cusp Catastrophe Models	
Catastrophe Model_1	**y** = - 2.32390 + 4.7722*Severity_1
	α = - 0.4610 - 0.0019*Rain_12h_sum - 1.3155*Acc.type3
	β = 5.5490
pseudo R^2	0.9954
Linear model R^2	-
Logistic model R^2	0.1107
Catastrophe Model_2	**y** = - 2.3328 + 4.9400*Severity_2
	α = 1.1122 - 0.2729*log(Q_avg_1h_up) - 0.1949*Rain_12h_sum
	β = 4.7636 + 0.4835*Acc.type2+0.3808*Acc.type3
pseudo R^2	0.8041
Linear model R^2	0.0983
Logistic model R^2	0.0585
Catastrophe Model_3	**y** = - 2.7342+ 1.5773*Severity_3
	α = - 0.8214 - 1.1054*Acc.type1 - 2.5298*Acc.type2 - 3.3955*Acc.type3 - 3.0374*Acc.type4
	β = 1.0920 + 0.0007*Q_avg_1h_up
pseudo R^2	0.3479
Linear model R^2	0.3317
Logistic model R^2	0.3539

Πίνακας 18: Αποτελέσματα θεωρίας καταστροφής για τη σοβαρότητα ατυχήματος με δικυκλιστές σε αστικές οδούς.

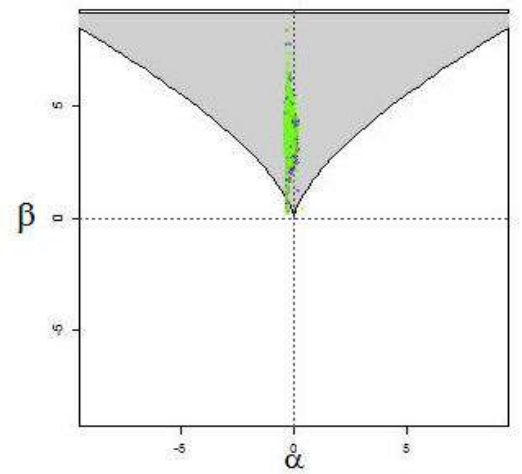


Γράφημα 6: Δισδιάστατη απεικόνιση της επιφάνειας καταστροφής της σοβαρότητας ατυχήματος με δικυκλιστές σε αστικές οδούς (Παράδειγμα μοντέλου 1).

Στον Πίνακα 20 καθώς και στο Γράφημα 7 φαίνονται τα αποτελέσματα του μοντέλου καταστροφής σε ό,τι αφορά στην πιθανότητα ατυχήματος σε αστικές οδούς. Η στατιστική προσαρμογή του μοντέλου καταστροφής υπερτερεί έναντι του απλού γραμμικού μοντέλου, καθώς εμφανίζει πολύ υψηλότερο R^2 (0.9258 έναντι 0.0136). Συνεπώς, συνίσταται η ύπαρξη μη γραμμικότητας στο σύστημα και η ύπαρξη απότομων μεταβάσεων από ασφαλείς καταστάσεις (όπου δε συμβαίνουν ατυχήματα) σε μη ασφαλείς καταστάσεις (όπου συμβαίνουν ατυχήματα) και το αντίστροφο.

Cusp Catastrophe Model	
Accident Probability	**y** = - 1.996 + 3.978*Accident_Occurrence **α** = - 0.2840 + 0.5587*V_cv_1h_up + 0.0004*Sol_1h_max **β** = 0.0004*Sol_1h_max + 0.0044*Q_avg_1h_up
pseudo R^2	0.9258
Linear model R^2	-
Logistic model R^2	0.0136

Πίνακας 19: Αποτελέσματα θεωρίας καταστροφής για την πιθανότητα ατυχήματος σε αστικές οδούς.

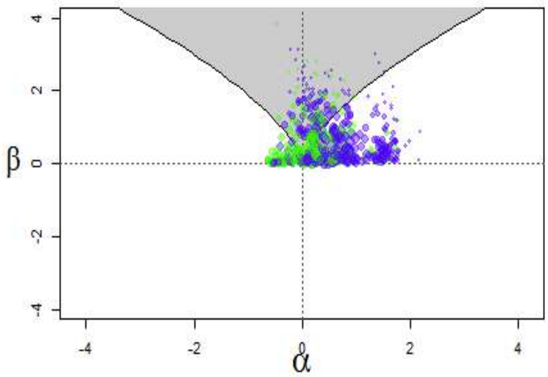


Γράφημα 7: Δισδιάστατη απεικόνιση της επιφάνειας καταστροφής της πιθανότητας ατυχήματος σε αστικές οδούς (Παράδειγμα μοντέλου 1).

Στη συνέχεια εξετάστηκε η πιθανότητα ατυχήματος με δικυκλιστές (πιθανότητα εμπλοκής δικυκλιστών σε ατύχημα). Και σε αυτή την περίπτωση τα μοντέλα καταστροφής υπερτερούν έναντι των απλών γραμμικών μοντέλων, καταδεικνύοντας ότι μικρές αλλαγές σε κυκλοφοριακές παραμέτρους είναι δυνατόν να επιφέρουν δραματικές αλλαγές στην πιθανότητα ατυχήματος με δικυκλιστές.

Cusp Catastrophe Model	
PTW Accident Probability	**y** = - 1.1511 + 2.4605*PTW_Accident_Involvement **α** = - 0.7148 + 0.0009*Q_avg_1h_up + 1.4027*Acc.type1 + 0.2717*Acc.type2 + 1.3054*Acc.type3 + 0.6211*Acc.type4 **β** = 4.3153*V_cv_1h_up
pseudo R^2	0.333
Linear model R^2	-
Logistic model R^2	0.1165

Πίνακας 20: Αποτελέσματα θεωρίας καταστροφής για την πιθανότητα ατυχήματος με δικυκλιστές σε αστικές οδούς.



Γράφημα 8: Δισδιάστατη απεικόνιση της επιφάνειας καταστροφής της πιθανότητας ατυχήματος με δικυκλιστές σε αστικές οδούς (Παράδειγμα μοντέλου 1).

6. Κριτική σύνθεση αποτελεσμάτων

Συνολικά, τα αποτελέσματα διδακτορικής διατριβής έδειξαν ότι η ολοκληρωμένη πολυ-επίπεδη ανάλυση που εφαρμόστηκε είναι δυνατόν να οδηγήσει σε πληρέστερη κατανόηση δύο πολύπλοκων ζητημάτων όπως είναι η πιθανότητα και η σοβαρότητα των ατυχημάτων τόσο σε αστικές οδούς όσο και στον αυτοκινητόδρομο. Επιπροσθέτως, τα υψηλής ευκρίνειας κυκλοφοριακά και μετεωρολογικά δεδομένα σε συνδυασμό με τη μεσοπρόθεσμου χαρακτήρα μεθοδολογική προσέγγιση, είναι δυνατόν να δώσουν σε νέες διαστάσεις στην ανάλυση και κατανόηση των οδικών ατυχημάτων. Είναι επίσης σημαντικό να δίνεται έμφαση στη μελέτη των ευάλωτων χρηστών της οδού όπως για παράδειγμα οι δικυκλιστές.

Από μεθοδολογικής απόψεως, εφαρμόστηκαν καινοτόμες και εναλλακτικές μέθοδοι, παρουσιάζοντας πολύ ικανοποιητική στατιστική προσαρμογή. και παράγοντας ενθαρρυντικά αποτελέσματα. Μάλιστα ορισμένες από αυτές τις μεθόδους εφαρμόζονται για πρώτη φορά σε συνδυασμό με τέτοιου είδους δεδομένα (πχ. μηχανές διανυσμάτων υποστήριξης, θεωρία καταστροφής) ή και για πρώτη φορά στο γνωστικό αντικείμενο της οδικής ασφάλειας. Ειδικά για τα μοντέλα της θεωρίας καταστροφής, τονίζεται ότι η υπόθεση για ύπαρξη δυναμικών συστημάτων πιθανώς να μπορούσε να οδηγήσει σε πιο γενικευμένη εφαρμογή παρόμοιων μεθόδων όπως για παράδειγμα εφαρμόζεται στη μελέτη των Frazier και Kockelman (2004).

Παρατίθενται συνοπτικά τα συνολικά αποτελέσματα της διδακτορικής διατριβής τα οποία χωρίζονται σε τέσσερα μέρη: πιθανότητα ατυχήματος, πιθανότητα ατυχήματος με δικυκλιστές, σοβαρότητα ατυχήματος, σοβαρότητα ατυχήματος με δικυκλιστές. Σε ό,τι αφορά στην πιθανότητα ατυχήματος τα συνολικά αποτελέσματα συνοψίζονται ως εξής:

- Η υψηλή κατάληψη είναι συσχετισμένη με αυξημένη πιθανότητα ατυχήματος.
- Υψηλές διακυμάνσεις της κατάληψης και του φόρτου αυξάνουν τον κίνδυνο ατυχήματος.
- Μη στατιστικά σημαντική επιρροή των μετεωρολογικών παραμέτρων όταν αναπτύσσονται γραμμικά μοντέλα.
- Στον αυτοκινητόδρομο, η μέση ταχύτητα έχει αρνητική συσχέτιση με την πιθανότητα ατυχήματος (Ahmed et al., 2011 & 2012, Yu et al., 2013).
- Η ανάλυση μέσω θεωρίας καταστροφής έδειξε πως υπάρχει πιθανότατα έντονη μη γραμμικότητα στο σύστημα.
- Πολύ ικανοποιητική προσαρμογή των μοντέλων καταστροφής.
- Ο μέσος κυκλοφοριακός φόρτος και οι διακυμάνσεις στην ταχύτητα βρέθηκαν να επηρεάζουν σημαντικά (με μη γραμμικό τρόπο).
- Σημαντική μη γραμμική επιρροή των μετεωρολογικών παραμέτρων.

Τα συνολικά αποτελέσματα της πιθανότητας εμπλοκής δικυκλιστών σε ατύχημα είναι:

- Η αξιοποίηση χρονοσειρών των κυκλοφοριακών παραμέτρων δίνει ικανοποιητική πρόβλεψη.
- Οι δικυκλιστές είναι πιθανότερο να εμπλακούν σε ατύχημα με 2 ή περισσότερα οχήματα και υπό συνθήκες αυξημένης κυκλοφορίας. Καταδεικνύεται δηλαδή ότι οι δικυκλιστές είναι ευάλωτοι στην αλληλεπίδραση με τους υπόλοιπους χρήστες της οδού.
- Μη σημαντική επιρροή των μετεωρολογικών παραμέτρων (σε κλασσικές γραμμικές σχέσεις).
- Ο μέσος κυκλοφοριακός φόρτος, οι διακυμάνσεις στην ταχύτητα και ο τύπος ατυχήματος επηρεάζουν την πιθανότητα εμπλοκής δικύκλου σύμφωνα με το μοντέλο καταστροφής μέσω μη γραμμικών σχέσεων.
- Καλύτερη προσαρμογή του μη γραμμικού έναντι του απλού γραμμικού μοντέλου Bayes στον αυτοκινητόδρομο.
- Μη γραμμική σχέση με τον κυκλοφοριακό φόρτο (πολυωνυμική σχέση β’ βαθμού).

Σε ό,τι αφορά στην σοβαρότητα ατυχήματος τα συνολικά αποτελέσματα συνοψίζονται ως εξής:

- Χαμηλά επίπεδα κατάληψης σχετίζονται με αυξημένη σοβαρότητα ατυχήματος.
- Διαπιστώθηκε ετερογενής επιρροή των κυκλοφοριακών παραμέτρων στις αστικές οδούς (ύπαρξη ετερογένειας) ανάλογα με τη λανθάνουσα ομάδα.
- Η κυκλοφοριακή συμφόρηση οδηγεί σε λιγότερο σοβαρά ατυχήματα στον αυτοκινητόδρομο (Christoforou et al., 2010).
- Καμία επιρροή των μετεωρολογικών παραμέτρων στη σοβαρότητα ατυχήματος εάν αναπτυχθούν γραμμικά μοντέλα.
- Σημαντική επιρροή του τύπου ατυχήματος.
- Ισχυρή ύπαρξη μη γραμμικότητας στο σύστημα.
- Οι διακυμάνσεις στην ταχύτητα και ο μέσος κυκλοφοριακός φόρτος επηρεάζουν μη γραμμικά τη σοβαρότητα ατυχήματος.
- Σημαντική μη γραμμική επιρροή μετεωρολογικών παραμέτρων (θερμοκρασία, βροχόπτωση, κτλ.).

Σε ό,τι αφορά στη σοβαρότητα ατυχήματος με δικυκλιστές βρέθηκαν τα εξής:

- Η χαμηλή κυκλοφορία βρέθηκε συσχετισμένη με την αυξημένη σοβαρότητα ατυχήματος με δικυκλιστές.
- Μη γραμμική συσχέτιση των κυκλοφοριακών και μετεωρολογικών παραμέτρων καθώς και του τύπου ατυχήματος με τη σοβαρότητα.
- Στον αυτοκινητόδρομο, ετερογενής επιρροή των κυκλοφοριακών παραμέτρων ανάλογα με τη “λανθάνουσα” ομάδα.
- Αρνητική συσχέτιση της σοβαρότητας με τον κυβισμό του δικύκλου.
- Δε βρέθηκε συσχέτιση της σοβαρότητας με τον τύπο ατυχήματος.

7. Συμπεράσματα και συνεισφορά της διδακτορικής διατριβής

Η επιστημονική και ερευνητική συνεισφορά της παρούσας διδακτορικής διατριβής περιλαμβάνει όπως αναφέρθηκε προηγουμένως, καινοτόμες και εναλλακτικές μεθοδολογίες καθώς και σημαντικά αποτελέσματα και ευρήματα τα οποία παρατίθενται παρακάτω.

Για παράδειγμα, σε αστικές οδούς, η εμπλοκή των δικυκλιστών σε ατυχήματα φαίνεται να οφείλεται περισσότερο στην αλληλεπίδραση με τους υπόλοιπους χρήστες της οδού παρά στα σφάλματα των ιδίων των δικυκλιστών. Αυτό το συμπέρασμα μπορεί να αποδοθεί στο γεγονός ότι όταν συμβαίνει ατύχημα με δύο και περισσότερα εμπλεκόμενα οχήματα ή όταν υπάρχουν υψηλές διακυμάνσεις του κυκλοφοριακού φόρτου, αυξάνεται η εμπλοκή δικύκλων σε ατύχημα.

Είναι ενδιαφέρον το γεγονός ότι οι μετεωρολογικές παράμετροι δεν βρέθηκαν στατιστικά σημαντικές, όταν εξετάζεται η επιρροή τους μέσω κλασσικών γραμμικών σχέσεων. Αυτή η τάση παρατηρήθηκε ανεξάρτητα από τη γραμμική μέθοδο ανάλυσης, την εξαρτημένη μεταβλητή (την πιθανότητα ή τη σοβαρότητα) και τον τύπο της οδού (π.χ. αστικές οδοί ή αυτοκινητόδρομοι). Αντιθέτως, τα μοντέλα καταστροφής έδειξαν ισχυρά σημαντική επίδραση κάποιων μετεωρολογικών παραμέτρων.

Η ανάπτυξη των μοντέλων καταστροφής καταδυκνύει ότι είναι πολύ πιθανό ότι ακόμα και μικρές αλλαγές των κυκλοφοριακών και των μετεωρολογικών παραμέτρων είναι δυνατόν να επιφέρουν σημαντική αλλαγή στο επίπεδο οδικής ασφάλειας σε αστικές οδούς. Αυτό αφορά όχι μόνο την πιθανότητα αλλά και τη σοβαρότητα ατυχήματος.

Τέλος, η κύρια συμβολή της διδακτορικής διατριβής συνοψίζεται στα πέντε παρακάτω σημεία:

1. Χρήση κυκλοφοριακών και μετεωρολογικών δεδομένων υψηλής ευκρίνειας για πρώτη φορά εκτός αυτοκινητοδρόμων.
2. Για πρώτη φορά έμφαση σε ανάλυση ατυχημάτων δικυκλιστών.
3. Ανάπτυξη και εφαρμογή πολυεπίπεδης στατιστικής ανάλυσης στην πιθανότητα και σοβαρότητα ατυχήματος.
4. Εφαρμογή εναλλακτικών και καινοτόμων μεθοδολογιών, ορισμένες από τις οποίες συνδυάζονται για πρώτη φορά με τέτοιου είδους δεδομένα ενώ άλλες εφαρμόζονται για πρώτη φορά γενικά στην οδική ασφάλεια.
5. Διερεύνηση και έμφαση στην μελέτη και ανάπτυξη μη παραδοσιακών μη γραμμικών σχέσεων σε ό,τι αφορά στην πιθανότητα και τη σοβαρότητα ατυχήματος.

Table of contents

List of tables

List of figures

Chapter 1 Introduction

1.1 Overview

Road safety is a very complicated scientific field of transportation research and has attracted huge efforts by researchers and practitioners. It is a major concern in the transportation industry, as accidents impose serious problems to society in terms of human costs, economic costs, property damage costs and medical costs. According to World Health Organization (WHO) (2013), the total number of road fatalities worldwide remains at 1.24 million per year. In 2013, some 25,900 people were killed in the European Union because of road accidents, around 313,000 were seriously injured and many more suffered slight injuries (ETSC, 2013). In 2013, 9,919 people were killed in traffic accidents on urban roads in the EU, corresponding to 38% of all traffic accident fatalities in 2013 (ERSO, 2015a). Almost 30,000 people were killed in traffic accidents on motorways in European Union countries between 2004 and 2013, corresponds to 8% of all traffic accident fatalities in those countries (ERSO, 2015b).

In Greece, according to ELSTAT data (ELSTAT, 2012), among the 988 persons killed in Greece in 2012, 489 were outside built-up areas and 499 were inside built-up areas. 52% of road fatalities outside built-up area occurred on national roads. More than three quarters of road accidents and half of fatalities occurred inside built-up areas. However, accident severity is more than 4 times higher outside built-up areas in total. Figure 1-1, illustrates the fatalities by area and road type for Greece in 2012.

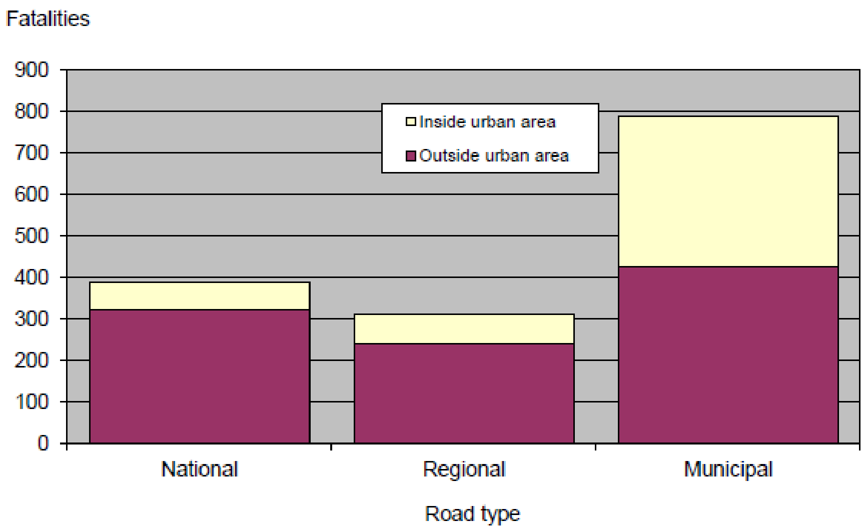


Figure 1-1: Road fatalities by area and road type, Greece, 2012. Source: Hellenic Statistical Authority (ELSTAT, 2012).

Accidents are basically resulted from a complex interaction of three fundamental factor groups: driver, vehicle, and environmental factors. Aside from these factors, other factors may also influence accident occurrence and severity such as socioeconomic factors, factors related to policy applied, legislation and of course randomness. Consequently, understanding the various factors that cause road accidents and their combined influence is very crucial.

The analysis of accident severity (or crash injury severity as stated in many studies), is also of great interest. The key part of such analysis is to understand the way which various contributing factors influence accident severity. Such factors may include driver and passengers attributes, geometric design characteristics, traffic conditions, weather, vehicle type etc. (Al-Ghamdi 2002; Sze and Wong 2007; Yamamoto and Shankar 2004; Yau 2004; Chang and Wang 2006; Milton et al. 2008; Quddus et al. 2002; Savolainen and Mannering 2007; Yu and Abdel-Aty, 2014b). Injury severity data are usually expressed by two or more discrete categories, according to the outcome of the accident. The most common measurements are the Abbreviated Injury Scale, the international classification of diseases (ICD), the five-level KABCO (K[fatal], A[incapacitating injury], B[non incapacitating injury], C[possible injury], O [property damage only]) scale and the injury severity score (ISS) (Lin & Kraus, 2008).

Although there has been a very considerable research effort so far, there is still much to be investigated, especially in order to acquire a better knowledge of detailed pre-accident conditions in order to have a better proactive safety management in major roads of the transport network. The advances in the field of Intelligent Transport Systems (ITS) and Meteorology enabled the constant and detailed monitoring of real-time traffic and weather conditions and have contributed to the safety assessment of major roads.

In international literature, the so-called “vulnerable road users” have attracted attention from researchers and practitioners as well. One group of vulnerable road users which is often involved in road accidents is motorcyclists. Moreover, there is a significant increase in motorcycling activities in many countries worldwide during the last years. Over the last two decades, the number of mopeds and motorcycles together referred to as powered-two-wheelers (PTWs) in Europe has doubled (Yannis et al., 2010). This mode shift is highly likely to be attributed to economic, mobility, flexibility and also environmental benefits that mopeds and motorcycles, together referred to as PTWs, offer to the users. A survey conducted by Jamson and Chorlton (2009) provided evidence that the nature of motorcycling seems to be changing as older riders appear and motorcycling becomes a leisure pursuit.

Due to the increased numbers in the percentage of PTWs in the motorised fleet and their lack of protection, it is not surprising that the motorised two-wheelers are considered a dangerous transport mode as the risk of being severely injured is significantly higher than the car occupants (Aare and von Holst, 2003; Wegman et al., 2008; Zambon and Hasselberg, 2006). Per vehicle mile travelled, motorcycle riders have a 34-fold higher risk of death in an accident than the motor vehicles users (Lin and Kraus, 2009).

Furthermore, PTW fatalities accounted for 18% of the total number of road accident fatalities in 2009 in the European Union-23 (EU-23) countries (ERSO, 2011). In spite of the number of measures that have been implemented in the last decade regarding PTW safety, the number of fatalities in accidents involving PTWs in EU is not reduced compared to traffic fatalities as shown in Figure 1-2.

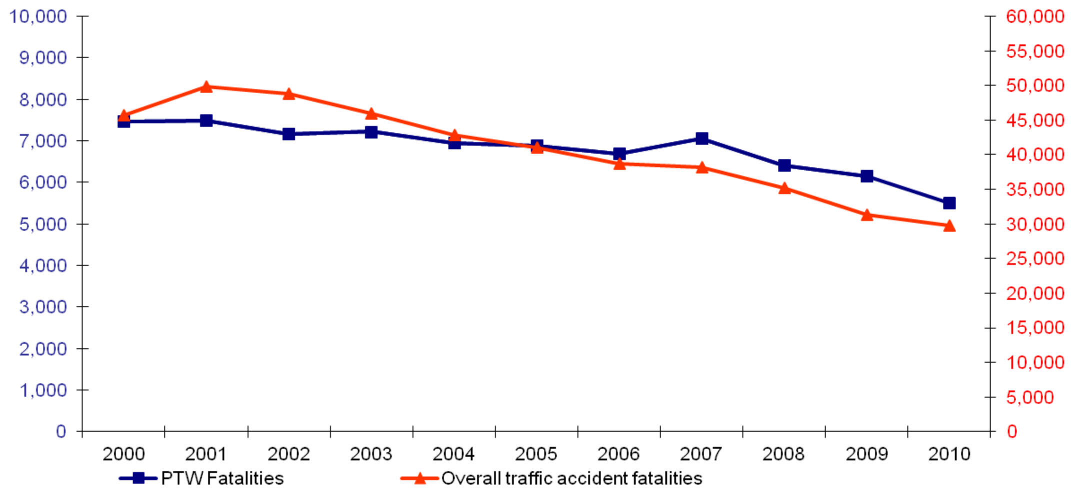


Figure 1-2: PTW rider fatalities compared to the total number of road accident fatalities in EU, 2000-2010. Source: European Road Safety Observatory (ERSO, 2011).

Regarding Greece, in 2012, 29% of road fatalities are motorcycle riders, whereas 39% of road fatalities are passenger car occupants. Most car occupant fatalities occur outside built-up areas while most motorcycle and pedestrian fatalities occur inside built-up areas. Figure 1-3 illustrates the road fatalities by transport mode in Greece for the year 2012 (ELSTAT, 2012). According to NRSO database (NRSO, 2009), while 90% of riders wear helmets outside built-up areas, only 41% of motorcycle passengers wear helmets inside built-up areas.

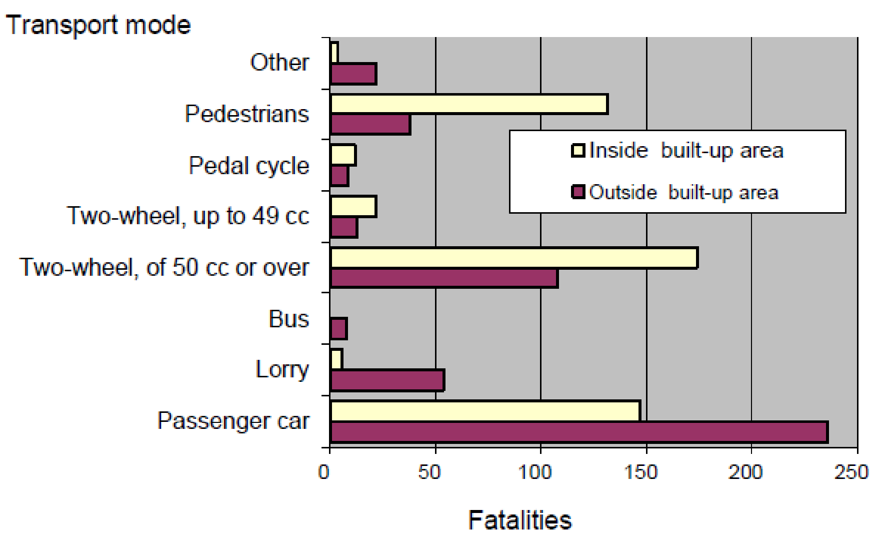


Figure 1-3: Road fatalities by transport mode, Greece, 2012. Source: Hellenic Statistical Authority (ELSTAT, 2012).

It is suspected that developing countries outside Europe are very likely to face a high number of PTW fatalities, mainly as a result of increased PTW use due to a generally low economic status. For example, in the western Pacific region, 36% of the deaths occurred among motorcyclists (Toroyan et al., 2013). Another example of such country is Nigeria, where not only motorcycling has become an increasingly popular means of transport but also riders seem to ignore safety measures (Solagberu et al., 2006). In the same study, it is also mentioned that there is an issue of underreported motorcycle injuries in developing countries.

Another study (Iamtrakul et al., 2003) states that a large portion of motor vehicles in Thailand are motorcycles and the majority of fatalities and injuries regard motorcyclists. Moreover, the reasons why motorcycle deaths have recently been increasing in the USA and why middle-aged riders are consistently over representing in fatal motorcycle accidents have to be investigated (Lin and Kraus, 2009). Table 1-1 that follows next, illustrates the number of PTWs, the fatalities and the fatality rates for year 2010 in Europe and in a few other countries as well.

Country	Mopeds			Motorcycles		
	Mopeds (x 1.000)	Road Fatalities	Fatalities per 10^6 mopeds	Motorcycles (x 1000)	Road Fatalities	Fatalities per 10^6 motorcycles
Austria	319	18	56	393	68	173
Czech Republic	473	4	8	430	95	221
Finland	260	9	35	227	16	70
France	1,121	248	221	1,436	704	490
Germany	2,104	74	35	3,763	635	169
Greece	1,389	36	26	1,499	372	248
Iceland	2	0	0	7	1	143
Israel	19	3	158	94	40	426
Italy	2,550	203	80	6,305	943	150
Japan	7,694	454	59	5,042	564	112
Luxemburg	26	0	0	16	1	63
Netherlands	500	44	88	623	60	96
New Zealand	27	0	0	73	50	685
Poland	922	83	90	1,013	259	256
Slovenia	41	6	146	41	17	415
Spain	2,290	100	44	2,707	386	143
Sweden	201	8	40	303	37	122
Switzerland	165	4	24	651	67	103
United Kingdom	84	10	119	1,182	403	341

Table 1-1: Number of PTWs, fatalities and fatality rates across European and other countries, 2010. Source: IRTAD, 2011.

Therefore, PTW safety has become a very important issue for societies either developed or developing and related studies are multiplied lately (Tiwari, 2013; Van Elslande and Elvik, 2012). However, existing literature regarding PTW safety and behaviour is still not very extensive compared to other road users' (passenger cars, etc.) and lacks from joint consideration of the various contributory factors. Consequently, there is a clear need for further research with respect to PTWs.

1.2 Research aim and objectives

The aim of the present PhD thesis is **the multi-faceted statistical exploration of accident probability and severity exploiting high resolution traffic and weather data.** Specific focus is being given on the analysis of Powered-Two-Wheelers. For that reason, data from urban roads as well as from urban motorways are considered. The effect of traffic and weather parameters is investigated, but other potential risk factors are investigated as well. It is also aimed to detect specific patterns in traffic characteristics and investigate how and to what extent they affect accident probability and severity. It is important to note that different definitions of severity are examined in order to achieve a more broad understanding of the phenomenon and of the underlying risk factors.

Consequently, an integrated approach featuring high resolution traffic and weather data and utilizing a series of advanced statistical and machine learning methods is applied in order to acquire a holistic exploration of accident probability and severity. This methodological toolbox aims to provide the broader picture and to result to the better understanding of accident risk, having dedicated specific emphasis and focus on PTWs. The above listed goals, were accomplished by the application of the following methodologies:

a. Utilizing a time series and machine learning techniques to predict Powered-two-wheelers accident involvement and accident type.
b. Detecting traffic states and exploring their effect on accident probability and severity. The accident probability and severity of PTWs are also explored.
c. Applying advanced statistical methods while all potential high resolution traffic and weather data were considered to model accident probability and severity.
d. Developing advanced models to explore accident probability and severity in urban motorways.
e. Considering the feasibility of cusp catastrophe theory to detect possible behavioural changes in the safety system and potential non-linearity while modelling accident probability and severity in urban roads.

1.3 Clarification of key terms used in this thesis

A handful of key terms used throughout the thesis, such as “accident severity”, “accident probability”, “PTW accident probability”, need to be clarified. It is noted that all the required definitions of traffic, weather and other relevant data are provided in Chapter 4.

The multi-faceted modelling approach of the thesis required different definitions for accident severity. More detailed definitions will be provided in the respective chapters. However, a more general description is required. There are three categories of accidents classified by their severity levels, namely fatal, serious and slight injury accidents. In the Greek database each accident is classified by the most severely injured person involved. More specifically, a fatal accident is defined as "an accident in which at least one person is killed"; and a serious injury accident is defined as an accident "in which at least one person is seriously injured but no person is killed". Similarly, a slight injury accident is defined as an accident "in which at least one person is slightly injured". It is noted that fatalities are defined as all persons killed within 30 days after the accident, serious injuries are all persons hospitalized for at least 24 hours, and slight injuries are all other persons reported by the police as injured without hospitalization. It is also noted that accidents resulting only in property damage are not included in the Greek database.

In the present thesis the term "PTW accident severity" is frequently used as well throughout the chapters, and is defined very similarly as "accident severity". The core difference is that it is measured only when a PTW is involved in the accident. Another term examined and analysed throughout the text is "occupant injury severity". This term describes the severity of every injured person involved (driver, rider or passenger), by classifying the involved occupants to the previously severity levels. "PTW occupant injury severity" refers to the injury severity levels of every injured person involved in the accident but was a PTW occupant (rider or passenger). Summing up, severity refers to the level of severity of an accident outcome given that an accident has occurred.

Aside from severity, accident probability is also investigated. Accident probability is the likelihood that an accident occurs. In order to investigate accident probability, a matched case-control approach is followed in the present thesis, where the accident database is enriched with a sample of non-accident cases. More details will be provided in Chapter 4 "Data Collection and Processing".

On the other hand, the definition of PTW accident probability is different than the definition of the traditional accident probability. PTW accident probability, is the likelihood that a PTW is involved in an accident given that this accident has occurred. The term "predicting PTW accident involvement", insinuates the prediction of whether a PTW is involved or not in an accident that has occurred. It is noted that in this thesis, PTW accident probability and PTW accident involvement is basically the same, but in order to be precise, the term "PTW accident involvement" is used in Chapter 5, where the prediction of whether a PTW is involved in accident is aimed.

1.4 Outline of the thesis

The present thesis is organised into **twelve** chapters. This section provides an overview of each following chapters. **Chapter 2** provides a literature review. The aim of this chapter is twofold. Firstly, to provide a critical review of the effect of traffic and weather on road safety. Specific focus is given on studies including high resolution traffic and weather data. Secondly, a detailed review of studies regarding PTW safety is provided in terms of data, methodology and results. All aspects of PTW safety are examined: rider behaviour, PTW interaction with other motorized traffic, accident frequency (as well as rates) and also accident severity. This chapter ends with a section dedicated to a summary of current trends and also to an identification and discussion of open research questions.

Chapter 3 provides a description of the concept of this thesis as well as the methodological approach followed in order to achieve the aims of the thesis.

Chapter 4 presents the data used in this thesis. The study areas are two major urban roads in Athens (Kifisias and Mesogeion avenues) as well as the Attica tollway ("Attiki odos"). The data collection methods are presented. This is followed by a description of the data for both urban and non-urban study areas. The data processing before used in the analysis is also described and basic descriptive statistics are presented.

Chapter 5 presents the results of the combined application of original and transformed time series data with Support Vector Machines to predict Powered-two-wheeler accident involvement and accident type.

Chapter 6 presents the clustering results of traffic states identification and then explores the effect of traffic states on accident probability and severity. Separate models for PTWs are demonstrated.

Chapter 7 aims to investigate the influence of high resolution traffic and weather data on accident probability and severity by using classification techniques as well as advanced statistical models.

Chapter 8 presents advanced statistical models to explore the factors affecting occupant injury severity, PTW occupant injury severity, accident probability and PTW accident probability in Attica Tollway.

Chapter 9 investigates the feasibility of applying the cusp catastrophe theory to estimate accident probability and severity, by considering separate models for PTWs as well.

Chapter 10 illustrates a critical synthesis of the findings of the advanced multi-faceted statistical analysis of accident probability and severity.

Chapter 11 presents the conclusions of the thesis and discusses the contribution to knowledge, the limitations as well as the recommendations for further research.

Chapter 12 provides a complete list of the references.

It is noted that Chapters 5, 6, and 9 are dedicated to urban roads, while Chapter 8 is dedicated to Attica Tollway. The thesis main text is followed by two appendices (A and B) which provide a number of tables summarizing the main findings of the literature review conducted.

Chapter 2 Literature review

2.1 Introduction

The review of literature is divided into two main sections. Firstly, a systematic review of the effect of traffic and weather characteristics on road safety is provided. There is a specific focus on recent studies featuring high resolution traffic and weather data. Since the thesis has a specific focus on Powered-Two-Wheelers, the second main section is dedicated to those studies related to rider behaviour, PTW interaction with other motorized traffic, accident frequency, accident rates and accident severity. At the end of the chapter, research gaps and open research questions are discussed.

The objective of this chapter is therefore to provide a systematic and critical review of current literature relating on one hand to the effect of traffic and weather parameters on safety and on the other hand to the various factors affecting PTW safety. This would benefit the development and interpretation of models relating traffic as well as weather characteristics and accidents, with specific focus on PTW accidents.

2.2 A review of the effect of traffic and weather characteristics on road safety

2.2.1 Introduction

Understanding the various factors that cause road accidents is very crucial and has attracted great attention in the literature. Although there has been a very considerable research effort so far, there is still much to be investigated, especially in order to acquire a better knowledge of detailed pre-accident conditions in order to have a better proactive safety management in major roads of the transport network. The advances in the field of Intelligent Transport Systems (ITS) and Meteorology enabled the constant and detailed monitoring of real-time traffic and weather conditions and have contributed to the safety assessment of major roads.

Stimulated by this emerging and promising approach, this section attempts to focus on traffic and weather characteristics and review their effect on road safety. Tables A1 and A2 summarize key findings and are presented at the Appendix A of the thesis. There is a very limited number of studies that provide a review of the impact of traffic or weather characteristics on road safety (Wang et al., 2013a). The present review goes one step further by reviewing both traffic and weather characteristics and by providing specific focus for the first time on recent studies which analyse real-time data. The combined effect of real-time traffic and weather on accident is also discussed.

The relevant and scientifically sound English-language studies were selected through a comprehensive search of major databases. Both international journals and conferences were considered. Moreover, reports from academic institutes or national research centres were also examined. In general, papers were reviewed if traffic and/or weather effects on safety were discussed.

This section is structured as follows: first, the effect of traffic characteristics (traffic flow, density/occupancy and speed) on road safety is discussed followed by a chapter discussing the effect of weather conditions. Next, recent trends in applying real-time traffic and weather data for the identification of their separate and/or combined effect on accident frequency and severity are investigated. In this section, road safety is expressed either as number of accidents (the absolute number of accidents according to thesis's area of examination, e.g. accidents per country, per road type, per segment), accident rates (the number of accidents per unit exposure of vehicle kilometres), fatalities, accident probability (Table A3 in the Appendix) (probability of accident occurrence usually on a road segment) and severity (of the accident or of each occupant involved) (Table A4 in the Appendix). Definitions of commonly used independent variables are presented at the beginning of each subsection.

It worth noticing, that the relationship between safety and approaching flows and speeds in urban junctions and intersections (Tanner, 1953; Colgate and Tanner, 1967; Leong, 1973; Golias, 1992; Papaioannou, 2007) has been excluded from the focus of the thesis, constituting a traffic phenomenon requiring a separate investigation.

2.2.2 Traffic characteristics

Traffic can be classified as flow, occupancy, density and speed. Traffic flow (or volume of traffic) is defined as the number of vehicles passing a cross-section of a road in a unit time, occupancy as proportion of time during which the loop detector is occupied, and density as the number of vehicles present per length of road at a given moment. Speed can be classified either as the space mean speed or the time mean speed. Hoogendoorn (2007) provides the next two definitions: "The mean speed is the arithmetic mean of the speeds of the vehicles passing a point on a roadway. The mean speed can be measured at a point (cross section) or at a given moment. The space-mean speed is the arithmetic mean of the speeds of the vehicles on a road section at a given moment".

From the literature review it was obvious that the effect of traffic characteristics on safety has gained considerable attention so far in order to understand the complex underlying relationships. However, despite the importance of the topic only a few studies were found to provide a review of the effect of traffic characteristics on road safety so far (Wang et al., 2013a).

2.2.2.1 Flow

The first steps in understanding this relation were set in the1950's, 1960's and 1970's. The common measures that were used were the hourly traffic volume and accident rates (number of accidents per year per million motor-vehicle kilometres).

The complexity of this phenomenon was evident from the beginning, resulting often in contradictory results. Belmont and Forbes (1953) found that for two-lane sections accident rates tend to increase almost in a linear way with hourly traffic volume. On the other hand, Smeed (1955) showed that accident rates have little variation for different annual hourly volumes, while later both Leutzbach (1966) and Gwynn (1967) found a U-shaped relationship. Ceder and Livneh (1978) attempted to analyse the relationships between accident density (accidents per km), rates and average daily traffic (ADT) focusing on interurban road sections. Power functions were applied, suggesting that the total accident density increases with the increase of ADT.

However, the first studies raised the need for more focused analyses. Ceder and Livneh (1982) made one step further to the investigation of the relationship between traffic flow and accident rates in terms of modelling by applying power functions and by using hourly traffic volumes instead of ADT. They authors also separated single from multi-vehicle accidents. More specifically, an inverse relationship between hourly traffic flow and accident rates was established in all single-vehicle accidents. On the other hand, mixed effects have been found in multi-vehicle accidents. Ceder (1982) carried on the same approach and extended it by separating traffic flow data into free flow and congested flow. The author found that the effect of flow on accidents is different for these two states. More specifically, for free flow data the well-known U-shaped curve was present for both single and multi-vehicle accidents, while in congestion the multi-vehicle accident rate was found to have an abrupt increase with hourly flow.

Linear relationships also existed as suggested by Turner and Thomas (1986) proposed models for both dual 2-lane and dual 3-lane motorways and found a negative relationship between traffic volumes and severe accidents. Similarly, Martin (2002) examined the relationship between traffic and severity (expressed as a percentage of accidents with injuries) based on 2000 km of French interurban motorways over a study period of 2 years and observed that property-damage accidents as well as accidents with injuries reach their maximum when the traffic volumes are under 400 vehicles per hour. The author argues that accident rates are higher in light traffic. On the other hand, Dickerson et al. (2000) found that high accident externalities are associated with high traffic flows, whilst these externalities are almost zero in light traffic.

Moreover, a positive relationship between traffic volume (in this case AADT—annual average daily traffic) and the number of accidents in both tangents and curves of four-lane median-divided Italian motorway based on 5-year accident data seems to exist (Caliendo et al., 2007). However, these results are not consistent with Vitaliano and Held (1991) who did not detect any significant increase in accident rates as the traffic flow increases.

However, only a few empirical studies have been found to analyse the effect of traffic flow separated for different groups of road users. One interesting approach was followed by Hiselius (2004) who attempted to investigate the effect of traffic flow on accident frequency in Swedish rural roads. The author separated traffic flows in two different ways: consisting of homogeneous vehicles and inhomogeneous (cars and lorries). It is stated that important information is lost if the differences between vehicle types are not considered when estimating the marginal effect of the traffic flow. Chang (2005) used interpolated daily traffic volumes for 1997and 1998 based on traffic surveys conducted earlier and found that accident frequency increases when the traffic per lane increases. Milton et al. (2008) applied a mixed logit model in order to model highway accident severity. It is very interesting that the effect of average daily traffic per lane increases injury severity on some segments whilst decreases it on others. The authors imply that the effect of traffic on accident severity cannot be assumed to be uniform across geographic locations. Similarly, Anastasopoulos and Mannering (2009) explored vehicle accident frequencies by using random-parameters count models. The authors argued that in the vast majority of road segments the number of accidents increase as AADT increases and only a few segments faced a decrease in accidents.

As it can be seen, the effects of traffic flow are more clearly investigated when the appropriate separations are considered (for example high or low volume, single or multi-vehicle accidents and so on). In that context, several researchers attempted to investigate the effect of high volumes or high volume to capacity ratios on accidents. For that reason, this relationship has been examined, but to a lesser extent than traffic flow itself.

The first hypothesis was set by Shefer (1994), suggesting a bell-shaped curve to associated volume to capacity (V/C) ratio and the total number of fatalities. Later, Shefer and Rietveld (1997) pro-posed that accident frequency increases in congestion because of the increased interactions among vehicles. On the other hand, accident severity decreases because lower speeds prevail in congested traffic and as a result the accident is not severe. Zhou and Sissiopiku (1997) considered a 26 km segment of Inter-state I-94 in the Detroit area. The authors obtained the volume to capacity ratios (V/C) from average hourly traffic volume counts collected in 1993 and 1994 and produced similar results. U-shaped relationships exist and in congestion accident rates are high but accidents tend to be less severe. Noland and Quddus (2005) used proxies in order to capture the effect of congestion on safety in Inner and Outer London, and suggest that there is little or even no effect of congestion on accident severity in urban areas. Kononov et al. (2008) explored the relation-ship between safety and congestion on urban freeways in Colorado, California and Texas. The authors found that increased AADT leads to an increase in fatal and injury accident rates. It is noted that congestion was reflected through the AADT.

Wang et al. (2009) conducted the study in M25 London orbital motorway, which was divided into 70 segments. In that study, the level of congestion was reflected by the congestion index instead of proxies. The congestion index (CI) was calculated as the difference between the actual travel time (T) and the free flow travel time (T_0) divided by the free flow travel time. The authors found no or little impact of traffic congestion on accident frequency. Similarly, Quddus et al. (2010) applied various advanced ordered models such as ordered logit models, heterogeneous choice models, and generalized ordered logit (partially constrained) models to measure the effect of congestion on safety while controlling for other contributory factors. The authors concluded that congestion has no impact on accident severity. The same authors carried on the research (Wang et al., 2013b) and concluded that increased traffic congestion is correlated with more KSI (killed and severely injured) accidents but has little impact on slight injury accidents.

2.2.2.2 Density/Occupancy

The literature review showed that traffic density has received considerably less attention than traffic flow with only a few studies having examined the effect of other traffic characteristics such as traffic density or occupancy on road safety. The main reason was that these type of data are not easily obtained (Lord et al., 2005). Moreover, density is found to have both negative and positive effects. Occupancy has received however far less attention than density. Only a few studies have been found (Garber and Subramanyan, 2002). In that study a non-linear relationship (U-shaped) is suggested to exist.

Brodsky and Hakkert (1983) stated that rural highway accident rates should theoretically increase with an increase in travel densities (vehicle miles per mile). Ivan et al. (2000) estimated Poisson regression models for predicting single and multi-vehicle highway accident rates as a function of traffic density and various parameters such as land use, ambient light conditions and time of day. The natural logarithm of the segment volume to capacity ratio (which in that study reflects density) was found to have a negative relationship with single-vehicle accident rates, but was found to be insignificant for multi-vehicle accidents.

Lord et al. (2005) studied the influence of volume, density and volume to capacity (V/C) ratios on the occurrence of accidents on rural and urban motorways. It is argued that in overall, density as well as V/C ratio has a bell-shaped relationship with accidents. This relationship exists also in single-vehicle accidents but not in multi-vehicle accidents where a linear positive relationship exists. More specifically, as density per V/C ratio increase, accident frequency increases. Then accident frequency reaches a maximum and then decreases again. The authors suggest that single and multi-vehicle accidents should be modelled separately. Another important conclusion is that predictive accident models should not rely solely on traffic flow, since density and V/C ratio provide a better explanation of this complex phenomenon.

Lastly, Kononov et al. (2012) examined the relationship between flow–density, speed, and accident rates. Data were obtained from free-ways in Colorado, U.S. The results show that as density and flow increase, the accident rate remains constant until it reaches a certain critical value (threshold). If this threshold is exceeded, the accident rate rises rapidly. The authors state that this rise can be attributed to the rise of flow without an important reduction in speed. This results in very little headways and any error made by drivers cannot be compensated, so an accident is difficult to be avoided.

2.2.2.3 Speed

The effect of speed on road accidents has gained considerable attention in international literature and much research has been done so far. Parameters such as the mean speed and the speed variance have been examined in order to define the underlying relationships and find which is the most influential. However, Baruya (1998) states that it is not straightforward to deploy a speed–accidents relationship because of the presence of other factors which affect speed distribution parameters and accidents.

Since this section does not only focus on the effect of speed on accidents, it is noted that a relevant comprehensive review was made by Aarts and van Schagen (2006). Authors stated that speed and accident rates are found to be positively related with an exponential or a power function. In general, increased speeds are associated with accident occurrence and increased severity (Nilsson, 2004; Elvik et al., 2004). Taylor et al. (2002) show that accident frequency rises rapidly with the mean traffic speed. According to this report, a 10% increase in the mean speed may result in a 30% increase in fatal and severe accidents.

On the other hand, other researchers resulted in different relationships between accidents and speed, and suggested that speed is negatively related with accidents (Garber and Gadirau, 1988). Taylor et al. (2000) found an inverse relationship, stating that as the mean speed increases, fewer accidents occur. In contrast to these studies, a recent study by Quddus (2013) showed that average speeds do not affect accident rates when controlling for other factors such as traffic volume and road geometry. But it is stated that speed variation is positively correlated with accident rates.

In fact, there are a handful of additional studies suggesting that speed variation is an important factor as well (Lave, 1985; Aljanahiet al., 1999). Solomon (1964) proposed a U-shaped relationship between accident rates and travel speed, but also stated that speed variation may have a larger effect on accidents than speed itself. Theoretically speaking, when traffic shifts from low to high speed or the opposite drivers may not adapt quickly to the new driving conditions. Similarly, if a vehicle or a number of vehicles deviate from the average traffic speed this may also contribute to an accident. In general, it is suggested that when there is less variance in speed between drivers fewer accidents are possible to occur, while greater variance leads to more accidents (Warren, 1982). Lassare (1986) suggested that a greater homogeneity in speed increases road safety. More research is needed in this topic since there were not many relevant studies. More effort were made later when real-time data became available (please see Section 4).

Another important topic is speed limits. The data required are more easily obtained and as a consequence more studies were found to examine this topic. Speed limits seem to have the clearest effect on accident fatalities as they reduce mean speeds and improve road safety (Aljanahi et al., 1999; De Pauw et al., 2014). It is important to note that this will not happen automatically, because of the presence of other behavioural factors such as travel habits and non-acceptance of measures which may counterbalance the effect of speed limit reduction (McCarthy, 1998). On the other hand, increased speed limits are considered to be related within creased number of accidents and fatalities (Rock, 1995; Ossiander and Cummings, 2002; Wong et al., 2005). Most of these studies were carried out in the U.S. In 1987 in Alabama, U.S., the speed limits on rural interstates was raised from 55 mph to 65mph. Brown et al. (1990) analysed the effect and suggested that a significant increase in accident frequency was observed (18.88%) (Table A5), but accident severity remained almost unaffected. Increased speed limits have been found also to be related with increased number of alcohol-related accidents (McCarthy, 1993).

The effects of speed limit reduction on Swedish motorways were examined by Johansson (1996), by using Poisson and negative binomial count data models and found that accidents with minor injuries and property damage were also decreased. However, the reduction in severe accidents was found to be statistically insignificant. Vernon et al. (2004) found that the increased speed limits in Utah, U.S., which took place in 1995 did not have a major overall effect, but different effects in the various road types. One possible limitation of the study was that the period of observation after the implementation of this measure was relatively short. In addition, the results may be affected by major reconstruction projects in the urban Salt Lake County, which may have affected accident occurrence.

2.2.2.4 Overall traffic considerations

International literature shows that traffic flow has gained much more attention than density and occupancy. Despite that, the complexity of the phenomenon leads sometimes to contradictory results. Some studies show a linear positive relationship between flow and accidents. On the other hand, U-shaped curves describe better the relationship between flow and accident rates. It is suspected, that one reason for potential differences may be attributed to the fact that the transferability of results is not straightforward. For example, regarding accident frequency, the use of different count regression models to examine the effect of traffic flow (e.g. Poisson, negative binomial, random effects models) may lead to different results.

Moreover, most of existing studies (at least not until recently when real-time data are widely applied) use too aggregate traffic data. This means that more short-term effects of traffic flow that may cause accidents cannot be captured, for instance, abrupt short-term traffic volume changes. AADT may be too aggregated (hiding variations in time and volume) to be directly linked with accidents and its effect may not be sufficiently detailed. The use of differently measured flow variables as AADT, hourly traffic volumes, volume per lane, V/C ratio and so on may also not lead to converging results. In addition, accidents are affected by various factors. Traffic characteristics may be related with other behavioural factors that cannot be directly captured. This hidden combined “factor” may not have the same effect in all cases. The interaction of traffic characteristics with other factors such as weather and behavioural factors need to be further examined.

When focusing on high traffic volumes and congestion, it can be observed that the impact on road safety is not always clear, especially in terms of accident severity. This topic has only recently been explored deeply by applying more advanced statistical models and along with the fact that the majority of present studies used proxies (involving varying definitions of congestion) in order to measure congestion seems to constitute potential reasons for inconsistency so far. Wang et al. (2013a) consider the wide application of proxies to capture congestion as a critical limitation of these studies.

On the other hand, the impact of traffic density has received far less attention in international literature. This fact has probably led to a degree of inconsistency regarding research results. Considering the impact of speed on road safety, still further research is needed, because results are not always consistent (even though more consistent than the impact of the rest of traffic parameters). It has to be noted though, that most studies state a positive correlation with accidents. Changes in speed limits are more likely to have an even more straightforward impact on safety.

Lastly, speed itself may not be the most influential factor since the impact of speed variation is likely to be more important. Speed variation has not excessively been investigated in the past, possibly due to data unavailability, however its effect is clearer when examined with high resolution data (see Section 2.2.4). Definitely, more research is needed to gain further understanding of this important topic.

2.2.3 Weather characteristics

International literature shows that weather and potential changes have an effect not only on road safety but on transport in general (Koetse and Rietveld, 2009). In fact, several weather parameters may have an influence as literature shows. The most common weather parameter that has been considered in literature is precipitation (rainfall, rainfall intensity, snowfall). It is noted that rainfall intensity is defined as the ratio between total amount of rain and rain duration). Precipitation can be measured annually, monthly, daily or hourly (in mm per unit time), depending on the type of data used in each study. Other weather parameters include air temperature, visibility and wind speed.

2.2.3.1 Precipitation

Extensive research has focused on precipitation effects. Non-professional travel activities are affected the most by precipitation (Bijleveld and Churchill, 2009). Satterthwaite (1976) argued that single vehicle accidents were the ones which were affected the most by wet weather. The findings regarding precipitation are quite consistent. In most studies it was found that precipitation is associated with increased number of accidents (Smith, 1982; Scott, 1986; Andrey and Yagar, 1993; Fridstrøm et al., 1995; Caliendo et al., 2007; Levineet al., 1995; Edwards, 1996; Chang and Chen, 2005). When the examined independent variables are the number of rainy days in month, a similar effect is proposed (Hermans et al., 2006), Shankaret al. (1995) had consistent results with previous studies, but apart from maximum rainfall, they suggested that by including the number of rainy days as an indicator variable, the influence of other factors such as exposure and pavement condition can be captured.

This positive linear relationship between rainfall and accidents is likely to exist also when analysing pedestrian accidents (Grahamand Glaister, 2003). Haghighi-Talab (1973) found a positive effect of rainfall on accident rates, but observed no statistical difference between moderate and heavy rainfall, as these two weather conditions have similar effects. Andrey and Yagar (1993) observed that accident risk is reduced to normal levels immediately after the end of a rainfall event. The authors considered the risk posed by rainfall as a combination of poor road friction and low visibility, and they argue that even if drivers are able to compensate for wet road conditions, accident risk is increased by low visibility.

Relatively few studies have found negative or non-significant correlations between rain and accidents. For instance, Jones et al. (1991) found that presence of rainfall had little effect on accident frequencies on freeways of Seattle, but it has to be noted that wet surface condition had a consistent positive relationship. Aguero-Valverde and Jovanis (2006) analysed 5-year injury and fatal accidents in Pennsylvania, U.S., and concluded that, although total precipitation was found to have a positive linear relationship in the traditional negative binomial models, it was not statistically significant in the hierarchical Full Bayes models. Nonetheless, some contradictory results were also produced. Karlaftis and Yannis (2010) used 21 years of daily count data for Athens, Greece, and found that high amount of precipitation may reduce the number of accidents. This seems counterintuitive but since rainfall in Greece is relatively rare compared to examine regions of previous literature, drivers may be more cautious and reduce their driving speeds. Similar results for the Athens region are presented in Bergel-Hayatet al. (2013). It is likely that the speed adjustment of drivers in regions like Athens may be enough to counterbalance the adverse weather conditions in contrast to other studies stating the opposite (Edwards, 1999). It is noted however, that these are not the only studies implying drivers' compensation in adverse weather (Khattak et al., 1998; Andrey and Yagar, 1993).

An interesting contributing parameter was found in Eisenberg (2004) and in Keay and Simmonds (2006). This parameter is the lagged effect of rain which was first found significant in Brodsky and Hakkert (1988). The analysis showed that the risk rises rapidly when the time since last precipitation increases. Results show that 1 cm of precipitation increases the fatal accident rate about 3% if exactly 2 dry days have passed and by about 9% if more than 20 days have passed. It is interesting that this pattern applied for non-fatal accidents as well. However, Brijs et al. (2008) did not confirm this finding and argue that lagged rain was insignificant. Instead, it was suggested that rain intensity (defined as the ratio between total amount of rain and rain duration) was highly significant. Consequently, the lag-effect is not supported. It is also suspected that snowfall may have an effect similar to the lagged effect of rain when attempting to interpret the results of Eisenberg and Warner (2005).

However, the impact of rainfall on accident severity may not be similar as a few studies imply no effect (Sherretz and Farhar, 1978) or a negative relationship

(Theofilatos et al., 2012). Caliendo et al. (2007) analysed accident occurrence separately on tangents and curves, by using hourly rainfall data by means of various Poisson, negative binomial and negative multinomial regression models and concluded that wet pavements increased severe accidents more than accidents with property damage only.

A handful of studies suggest that snowfall was found to have a negative relationship with accidents (Fridstrøm et al., 1995). In that study it is illustrated that injury accidents in Denmark are reduced by circa 1.2% for each additional snowing day. This finding may be attributed to risk compensation by Nordic drivers or even reduced exposure in winter. It is interesting however, that the first snowing day of winter may catch drivers by surprise and increase the risk (although this variable is hardly significant). On the other hand, Andreescu and Frost (1998) found snowfall to be the more influential weather variable, in sense that the number of accidents increased dramatically with increased snowfalls. Eisenberg and Warner (2005) stated that snowy weather resulted in less fatal accidents than non-fatal and property damage accidents. It is interesting that authors indicate that there were more fatalities in the first snowy day of the year than other snowy days. Recently, El-Basyouny and Kwon (2012) concluded that both severe and property-damage-only collisions have a significant positive relationship with total snowfall and total daily precipitation.

2.2.3.2 Other weather characteristics

Aside from precipitation, other weather variables that were considered are low visibility, wind and air temperature. Relatively few studies have specific focus on the effect of visibility on road safety. According to Al-Ghamdi (2007), fog-related accidents have remarkably higher injury and fatality rates. Abdel-Aty et al. (2011) studied the impact of low visibility due to fog and smoke on accidents. Injury severity in low visibility conditions was found to be higher, while head-on and rear-end collisions were the prevailing types of accidents. The effect of wind is not extensively explored in literature, with related studies being relatively few (Baker and Reynolds, 1992; Levine et al., 1995; Hermans et al., 2006; Brijs et al., 2008). Baker and Reynolds (1992) found during heavy storms in the UK almost 50% of accidents are overturning accidents and 66% of accidents involved high trucks implying a risk. The other studies do not support the hypothesis of significant effect of wind on accidents. The impact of wind needs further research regarding particularly high vehicles but also motorcycles. Hermans et al. (2006) examined the influence of various factors such as temperature, wind, radiation and relative humidity on the hourly number of accidents in The Netherlands in 2002. The authors suggest that increased wind gusts and longer duration of precipitation are among the most significant variables and they are correlated with increased number of accidents. Increased radiation and sunshine hours seem to have the same effect. The effect of other variables such as relative humidity is not so straightforward.

Concerning air temperature, the effect is not clear while it is found that warmer temperatures are associated with fewer accidents (Scott, 1986). The opposite effect is suggested by Antoniou et al. (2013) and by Karlaftis and Yannis (2010). In Antoniou et al. (2013), the examined area is a typical Mediterranean city (Athens) where low temperatures and rainfall take place mainly in winter. The authors attribute this finding to reduced mobility in adverse weather. Bergel-Hayat et al. (2013) argue that 1◦C of additional average temperature during a month increases the number of injury accidents in that month by 1–2%. Not straightforward and negative non-linear relationship is proposed by Brijs et al. (2008). More specifically, lower temperatures related to the reference category (temperatures above 20ºC) lead to an increase in the number of accidents. Moreover, deviating from the monthly average temperature (when the daily mean temperature exceeds the monthly mean temperature) leads to more accidents as well.

2.2.3.3 Overall weather considerations

Adverse weather has a more consistent impact on accident risk than on accident severity which has been examined less extensively so far. The effects of rainfall and snowfall in particular have been discussed earlier in this section. It is noted that adverse weather is likely to increase accident severity of vulnerable road users like bicyclists (Kim et al., 2007) and motorcyclists (Majdzadeh et al., 2008). Malyshkina and Mannering (2009) applied a Markov switching multinomial logit model to explore injury severity, having specific focus on the methodological part. Weather variables such as amount of precipitation and snowfall, fog, temperature and wind gusts were examined. Results showed that the number of severe and fatal accidents as well as minor accidents increase in adverse weather, but their proportion is approximately the same as in good weather.

Some recent studies have found contradictory results concerning precipitation and accidents. It is suspected that the effect of rainfall is different in Mediterranean countries where rainfall occurs mainly in winter and is rarer than some U.S. states and other European countries. A number of studies (Karlaftis and Yannis, 2010; Theofilatos et al., 2012; Bergel-Hayat et al., 2013) suggest a negative relationship between adverse weather and road safety, mainly because drivers of these countries are not used to drive under adverse weather and consequently adjust their behaviour by driving more carefully especially when there is a sudden rainfall after a long dry period of time. Another possible explanation is the limited volumes of motorcyclists in adverse weather. Findings concerning the lag effect of rain do not seem to apply in data from Mediterranean countries, mainly because of the adjusted drivers' behaviour mentioned previously. In such cases, the sudden deterioration in weather does not surprise drivers but on the contrary they turned out to be more cautious. Concluding, it is high likely that the weather effects on safety and also on driver behaviour depend on the specific area characteristics which probably have influence on how the population perceives extreme weather or sudden changes. This probably causes contradictories and makes the transferability of results more complicated.

The literature review showed a notably limited number of studies which examine the combined effect of weather and traffic on safety. To be more specific, Keay and Simmonds (2006) examined the effect of rainfall in Melbourne area and normalized accident count by traffic volume. As most other studies, rainfall constantly represents a driving hazard and a strong impact of rainfall after dry spells was found (lagged effect). It is interesting to note the indirect effect of weather on traffic. Keay and Simmonds (2005) state that there is a negative relationship between traffic volume and the amount of rainfall, while Bergel-Hayat and Depire (2004) proposed a similar indirect effect of monthly rain on traffic volumes.

The combined effect of weather and traffic on safety is not adequately addressed, but recent studies which usually examine more weather and traffic parameters can be found to have investigated this combined effect at a more disaggregate level as it can be observed in the next chapter, where literature examining microscopic traffic and weather data is discussed.

One last remark concerns a factor that is closely related with weather and is the lack of sufficient friction. The skidding resistance of pavements has been shown to vary seasonally. Research was performed in Great Britain in the 1950s in order to quantify the accident risks associated with certain levels of side-force coefficients (Salt 1977). In the early 1990s, a study conducted in Puerto Rico found a statistically significant relationship between the minimum Mu-Meter skid number and the ratio of wet to dry 24 accidents in a section (Gandhi et al., 1991). The average friction coefficient in a section was found to be less related to accident rates than the minimum friction coefficient.

Rogers and Gargett (1991) conducted a large-scale national skid resistance survey in the UK in 1981, to examine the link between skid resistance and accidents with injuries. The network was divided into different categories of sites (e.g. junctions, traffic signals, pedestrian crossings and so on). The next figure (Figure 2-1) demonstrates examples of the relationships from the aforementioned study.

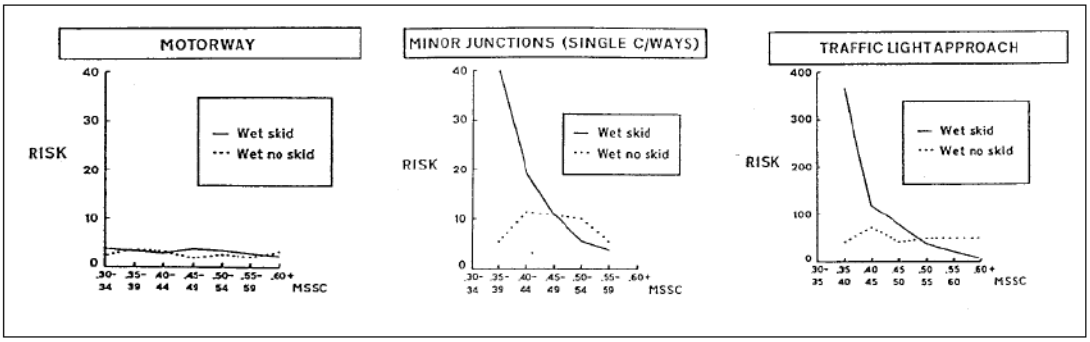


Figure 2-1: Risk of accidents on wet roads and mean summer SCRIM coefficient (MSSC).

Wilson (2006) and Kuttesh (2004), provide good reviews of skid resistance and accident risk. According to Jellie (2003), a number of statistical techniques are applied, in order to examine the relationship between road accidents and road pavement surface skid resistance, such as before and after studies, comparison with the Norm studies and regression studies. Kuttesh (2004) states that, although the relationship between surface friction and roadway safety has long been recognized, attempts to quantify the effect of pavement skid resistance on wet accident rates have produced inconsistent results. For example, while some studies have not found a statistically significant relationship between friction and wet accident rates, a clear relationship was found to exist in some other. For that reason, Kuttesh (2004) developed regression models and found that there is statistically significant effect of skid resistance on wet accident rate; the wet accident rate increases with decreasing skid numbers. The author also argues that skid resistance alone is not adequate to model the variability in the wet accident rates. In addition, the wet accident rate also decreases with increasing traffic volume. Moreover, much of this research has concluded that the risk of wet accidents increases significantly below certain friction levels on the pavement surface.

2.2.4 Real-time traffic and weather characteristics

Considerable efforts have been made recently to investigate road safety by incorporating weather and traffic characteristics recorded on a real-time basis but also to address the short-comings of the previous studies where mainly aggregated measures of traffic flow and speed were used as well as qualitative or aggregated measures of weather variables. Most of recent studies aimed to assess the real-time safety of major freeways. This is achieved through the identification of accident precursors, such as traffic volumes, speed variation between lanes, variation in the traffic volume and so on, in order to predict an accident or how severe it would be given that it occurs. Similarly, abrupt changes in traffic or speed variables indicate the existence of an incident. In these studies, data are very disaggregate. Traffic parameters are usually measured every 30sec and then usually aggregated over a 5-min period to obtain averages and standard deviations. For example, speed variation is usually expressed through the standard deviation of speed on a given segment. Weather parameters are commonly measured in an hourly basis (for example, 1 hour prior to an accident).

Concerning data collection, traffic data are usually collected from loop detectors both downstream and upstream of the accident location. In some cases, other data sources such as Automatic Vehicle Identification Data (AVI data) were used (Ahmed and Abdel-Aty, 2012, 2013). Weather characteristics were obtained from records stored in meteorological stations close to the area of interest and involve mainly precipitation and visibility. It is noted that since most of recent literature extracted both real-time traffic and weather data, it was decided to examine both factors together in this section and not separately. This section was preferred to be structured according to the aim of the present review (accident probability, frequency and severity) having followed a short introduction.

The first studies that used real-time traffic data were published recently (Oh et al., 2001; Golob and Recker, 2003). Lee et al. (2003) introduced the philosophy of accident precursors. A 10-km long part of Gardiner Expressway in Toronto was examined and data were extracted from 38 loop detector stations for a period of 13 months. Accident rates were investigated by using traffic characteristics upstream of the accident segment. Controlling for other factors such as weather and geometrical characteristics, the risk factors include increased density, congestion and speed variation. Golob et al. (2004) concluded that safety is affected by the mean volume and the median speed, but also by temporal variations in volume and speed by freeway lane.

An issue of this approach is that the actual time of the accident may be not always precisely known. This may be problematic since a short-term prediction is usually attempted, mainly by aggregating raw 30-s data into 5-min time slice to obtain averages of traffic parameters. For that reason, and also to avoid traffic conditions caused by the accident itself, Golob and Recker (2002) discarded2.5 min before the occurrence of each accident. This methodology is followed also by later studies (Kockelman and Ma, 2007). It is noted that Lee et al. (2003) proposed a methodology to estimate the actual time of the accident in freeways by detecting abrupt changes in traffic volumes and speed.

2.2.4.1 Accident probability

A common methodology adopted to predict accident probability is to include data for accident cases but also for random non-accident cases. This methodology is known as matched case-control study and is followed by several researchers (Abdel-Aty and Pande, 2005; Abdel-Aty et al., 2007; Ahmed and Abdel-Aty, 2012; Yu and Abdel-Aty, 2013a). More detailed information about the matched case–control logistic regression model can be found in Ahmed et al. (2012a). Results of the study imply that increased speed variation at any given (accident) segment combined with a decrease in average speed in the respective downstream segment, results in increased likelihood of rear-end accidents. The authors suggest that almost 73%of rear-end accidents could be correctly predicted. Ahmed and Abdel-Aty (2012), found that the probability of a accident increases when the variation in speed increases and the average speed decreases at the segment of the accident at 5–10 min before (prior to the accident occurrence). Speed variations are not always a risk factor. Kockelman and Ma (2007), found no connection between 30-sec speed changes and accident occurrence.

Studies examining accident probability usually include traffic and weather characteristics as well as the combined effect of traffic and weather. Ahmed et al. (2012b) attempted to investigate the impact of geometrical, traffic and weather variables on accident occurrence on freeways. These factors are considered critical, since the study is based on a mountainous freeway which faces adverse weather in winter. In winter, results are consistent with previous studies, suggesting that low visibility, high precipitation and speed variation increase the likelihood of accidents. Surprisingly, for dry season, low average speeds and low visibility increase the odds of an accident. On the basis of this difference, it is proposed that different traffic management strategies should be applied for dry and snow season. As it can be seen, the interaction of traffic and weather risk factors is different for dry and winter season, suggesting a different combined effect.

Xu et al. (2013a) performed a similar methodology and deployed three models for clear weather, rainy weather and low visibility weather. The authors argue that speed difference between upstream and downstream detectors increased the accident risk for low visibility and/or rainy weather. Rain intensity and occupancy variations downstream also increased accident risk for rainy weather and low visibility weather, respectively. Under clear weather, speed variations measured by the standard deviation of speed increased risk as well. In a similar study (Xu et al., 2013b), it is argued that accident probability is high when the traffic density upstream, the speed variance upstream and/or downstream, volume difference between upstream and downstream station and the occupancy difference between upstream and downstream station are high as well. Adverse weather conditions also tend to increase the probability of an accident.

Hassan and Abdel-Aty (2013) and Abdel-Aty et al. (2012) focused mainly on the risk factors under low visibility. Abdel-Aty et al. (2012) extracted data from loop detectors as well as from automatic vehicle identification (AVI) sensors and managed to predict more than 70% of accidents under low visibility conditions, concluding that both kinds of data can be used for safety applications such as low visibility accident prediction. Hassan and Abdel-Aty (2013) stated that significant traffic flow variables are different between visibility related accidents and accidents under clear visibility. More specifically, average speed and average occupancy (downstream), as well as average speed upstream, increased accident risk under low visibility conditions. On the other hand, the logarithm of average occupancy downstream and the coefficient of variation of speed upstream were the main risk factors in clear weather.

Some recent studies were very focused and aimed to measure a specific phenomenon of traffic and its effect on safety. For instance, the effect of traffic oscillations (stop and go driving) in queued traffic conditions on accident occurrence was examined by Zheng et al. (2010). It was found that traffic oscillations as expressed by the standard deviation of speed increased the probability of rear-end accidents. However, authors also state that the average traffic states prior to accidents were less significant than the speed variations in congestion. Hossain and Muromachi (2013) aimed to identify accident predictors on urban expressways. The main difference from previous studies is that freeway segments and ramp vicinities were analysed separately. Their hypothesis was supported by their findings which suggest that the accident mechanism is not the same for basic freeway segments and ramps. A limitation of the study was the lack of severity data. Xu et al. (2012) developed accident risk models for different traffic states. Traffic flow parameters were found to have different effects on safety for every traffic state. For instance, the average downstream occupancy seemed to reduce accident risk in two traffic states (in congested traffic as well as in transition from free flow to congested flow) but caused an increase in the overall model.

2.2.4.2 Accident severity

It has been showed, that most of existing studies aimed to predict accident probability according to short-term traffic parameters to develop proactive management strategies. Fewer studies, used real-time traffic and weather data to examine the severity of an accident given that an accident has occurred. Even fewer studies investigate both accident probability and severity (Xu et al., 2013b). In that study, it is suggested that congested traffic leads to less severe accidents because of the lower driving speeds. Possibly having the same explanation, accidents in adverse weather tends to be less severe as well.

Christoforou et al. (2010) focused on real-time traffic volume and speed data and applied a random parameters ordered pro-bit model. Vehicle occupants' severity levels of accidents on the A4–A86 junction in the Paris region were examined. It is implied that traffic volume does not differ among observations (it has fixed effects and not random effects) and that the high average traffic volume (measured per lane and over 6 min in vehicles) results in less severe accidents. The authors support this finding by stating that drivers are more likely to develop high speeds under free flow and thus involved in more severe accidents. Furthermore, average speed under very high traffic conditions (more than 1120 vehicles/lane/h) was found to increase severity. Some of the limitations of the study were that traffic data were collected over 6 min and the absence of quantitative real-time weather variables. However, Golob et al. (2008) found that traffic conditions had only a marginal influence on accident severity. Recently, Yu and Abdel-Aty (2014a) applied Hierarchical Bayesian probit models and found that large variation of speed and low visibility prior to an accident were some key risk factors which increase accident severity. Furthermore, low visibility in winter results in more severe accidents, while it seems to have no effect for the dry season model.

Jung et al. (2010) focused primarily on weather conditions prior to an accident and examined single-vehicle accident severity in Wisconsin interstate highways. In addition, some real-time parameters were also considered. The advantage of that study is the use of very detailed microscopic real-time weather variables such as15-min rainfall intensity, water film depth, wind speed and temperature. It is suggested that increased rainfall intensity 15 min prior to the accident is 1.787 times more likely to result to the highest level of severity, but on the contrary, wind speed decreases the level of severity.

2.2.4.3 Accident frequency

The number of accident frequency studies in international literature were relatively few as well (Yu and Abdel-Aty, 2013a; Yu et al., 2013). Yu et al. (2013) analysed one year accident data using geometrical, real-time traffic and weather data. Single-vehicle and multi-vehicle accidents were considered for both snow and dry season. Concerning traffic characteristics, the logarithm of vehicle miles travelled had a positive effect indicating that exposure results in high number of accidents. Concerning weather characteristics, the study's results are intuitive and suggest that high precipitation and low visibility increase accident occurrence. However, it is interesting to note that single-vehicle accidents seem to be more influenced by weather characteristics, while multi-vehicle accidents are more correlated with traffic characteristics. An issue of accident frequency studies is that they analyse one year data. The segmentation method that is followed, assumes that if more than one accident occurs in a given segment, the traffic parameters are the average of accidents.

Very few accident frequency studies had a specific focus on weather characteristics (Usman et al., 2010, 2012). For instance, Usman et al. (2012) examined various factors such as precipitation intensity, visibility, air temperature and wind speed. Low visibility, poor road surface conditions, high winds and low temperatures are associated with increased number of accidents. On the other hand, other studies found that air temperature and precipitation do not influence accident frequency (Usman et al., 2010). Overall, these two studies indicate that low visibility has a consistent positive relationship with road accidents increasing accident frequency.

2.2.5 Summary

Taking into consideration the increasing availability of real-time traffic data and stimulated by the importance of proactive safety management, this section attempted to provide a review of the effect of traffic and weather characteristics on road safety, identify the gaps and discuss the needs for further research. Despite the existence of generally mixed evidence on the effect of traffic parameters, a few patterns can be observed. For instance, traffic flow seems to have a non-linear relationship with accident rates, even though some studies suggest linear relationship with accidents. On the other hand, increased speed limits have found to have a straightforward positive relationship with accident occurrence. Regarding weather effects, the effect of precipitation is quite consistent and leads generally to increased accident frequency but does not seem to have a consistent effect on severity. The impact of other weather parameters on safety, such as visibility, wind speed and temperature is not found straightforward so far. The increasing use of real-time data not only makes easier to identify the safety impact of traffic and weather characteristics, but most importantly makes possible the identification of their combined effect. The more systematic use of these real-time data may address several of the research gaps identified by this review.

2.3 A review of Powered-Two-Wheeler behaviour and safety

2.3.1 Introduction

Existing literature regarding PTW safety and behaviour is still not very extensive compared to other road users' (passenger cars, etc.) and lacks from joint consideration of the various contributory factors. Consequently, there is a clear need for further research with respect to PTWs. Taking into account that existing studies reviewing PTW risk factors are very few, this review attempts to provide a broad overview of PTW behaviour and safety. To achieve this aim, the review included both behavioural aspects and interactions of motorised traffic with PTWs as well as injury risk. To present a complete overview of influential factors, a great variety of identified risk factors were examined in the review.

Moreover, in contrast to previous related reviews (e.g. Vlahogianni et al., 2012a) which mainly focus on identifying the critical PTW risk factors, this chapter also provides an overview of data collection methods and the main analysis methods in order to outline the research gaps and propose future research directions regarding PTW safety.

The literature review focused on the most recent and quantitatively substantiated research results in the field of PTW behaviour and safety. A comprehensive search of major databases such as Science Direct and Scopus was carried out. Keywords for the search included motorcycle, moped, PTW, safety, accident, accident, behaviour, risk, accident, severity, frequency, rates, and so on. Both macroscopic analyses (which are associated with accident frequency, accident rates and accident severity) and microscopic analyses (which examine PTW rider behaviour and interaction with other motorised traffic) are examined and discussed in this section. Furthermore, an effort was made to focus on data and analysis methodologies which are more often found in PTW accident analysis.

The structure of the section is as follows: first, the type of data and the collection methods are presented followed by the methods of analysis. Then an overview of contributory factors of PTW accidents is illustrated. Contributory factors are identified as those risk factors which have been found to be examined in international literature. Particularly, subsection 2.3.4.1 “Rider Behaviour” aims to review the factors that influence risky behaviours such as speeding, violations, and so on. The research gaps and suggestions for future research are discussed in section 2.4.

It is noted that at the end of the thesis lies the Appendix B, where there are four tables (Tables B.1-B.4) which summarise the data, methods of analysis and the main findings of the literature. The review of existing studies includes four aspects of PTW safety: PTW rider behaviour, PTW interaction with other road users, PTW accident frequency/rates and PTW accident severity. In this section, PTW rider behaviour concerns the actions of the rider during driving (risky or not) as well as attitudes, perception and other behavioural and psychological characteristics, such as sensation seeking, aggressiveness, and so on.

Particularly, interaction of PTWs with other road users could be entirely incorporated in the examination of behaviour, but it was examined separately in order to emphasise its importance. Of course, this is not always possible, since sometimes it is not clear to discriminate interaction and deal with it separately from behaviour or accidents, because it is involved in some behavioural as well as some accident analysis studies.
It is noted that accident severity reflects the outcome of the accident, and depends on the way it is measured in the respective studies. For example, a number of studies examine the severity of an accident classified as the most injured person involved, while other studies examine the injury severity of occupants.

Potential risk factors for road traffic injuries have been proposed by Haddon (1980) and have been divided to four categories: (1) factors influencing exposure to risk, such as economic, demographic factors and mixture of high speed motorised traffic with vulnerable road users, (2) factors influencing accident involvement such as speeding, alcohol or drugs consumption, being a young male, defects in road design, and so on, (3) factors influencing accident severity, such as not using seatbelts or helmets, presence of roadside objects and, finally, (4) factors influencing post-accident outcome of injuries such as delay in detecting accident, delay in transport of injured persons to a health facility, lack of appropriate hospital care, leakage of hazardous materials and presence of fire resulting from collision. More factors have been investigated regarding PTW studies, for example, attitudes, hazard perception, errors and violations.

Potentially respective data could be extracted from statistical files, questionnaires and simulators (mainly for attitudes and behavioural aspects), police records as well as hospital records. Especially for critical pre-accident phase information, data such as braking and handling are extremely difficult to be massively collected, with an exception of a limited number of studies using naturalistic data.

2.3.2 Data

The type of data and the collection methods are very critical for every researcher wishing to analyse PTW safety and depend heavily on the specific aspect of PTW safety that is analysed. This is better illustrated in Subsections 2.3.2.1, 2.3.2.2, 2.3.2.3 and 2.3.2.4 (PTW rider behaviour, Interaction with other motorised traffic, Accident frequency/rates and Accident severity). It is noted that the other two parts of this section (sections 2.3.3 and 2.3.4) are structured similarly.

2.3.2.1 Rider behaviour

Speeding, sensation seeking, aggression, perceived risk, errors, violations and attitudes towards road safety are considered to be very important issues. The primary methods of data collection that have been used so far depend heavily on stated behaviour and attitudes of riders. They mainly involve interviews and questionnaires in order to capture the attitudes and perception of riders (Broughton et al., 2009; Chen, 2009; Cheng and Ng, 2010; Chung and Wong, 2012; Elliott, Baughan and Sexton, 2007; Haque, Chin, and Lim, 2010; Steg and van Brussel, 2009; Wong, Chung and Huang, 2010), and online surveys as well (Ozkan et al., 2012).

Setting experiments is another way of observing and grasping behavioural characteristics of riders via an environment that attempts to simulate reality. They mainly involve neuropsychological tests, video-based tests, conspicuity tests eye tracking (Cheng et al., 2011; DiStasi et al., 2011; Gershon et al., 2012; Rosenbloom et al., 2011) and simulators (Di Stasi et al., 2009; Filtness et al., 2013; Hosking, Liu and Bayly, 2010; Liu et al., 2009). These methods have been used in order to measure and observe specific actions or choices of participants in an experimental environment under hypothetical situations.

Lastly, direct roadside observations which offer the possibility to capture the real behaviour have not been excessively preferred with the exception of a few studies (Walton and Buchanan, 2012; Woodcock, 2007). As a first remark, aside from roadside observations, experiments with naturalistic data that capture the behaviour of riders in real situations are rare. It is obvious that an evident limitation regarding data exists. Only recently, some studies used naturalistic data (Vlahogianni et al., 2012b; Walker, Stanton and Salmon, 2011).

2.3.2.2 PTW interaction with other motorised traffic

PTW interaction with other motorised traffic takes place in many situations, for example, in overtaking, manoeuvring, braking and in approaching intersections. The importance of interaction is highlighted by the fact that two vehicle collisions involve PTW and car interactions and although many PTW accidents are single-vehicle accidents, there is evidence that a number of those accidents might have occurred in order to avoid a collision with a vehicle (Preusser, Williams and Ulmer, 1995).

A common type of interaction takes place in right-of way accidents, which involve motorcycle conspicuity and motorist's speed/distance judgement (Pai, 2009). The studies that investigated conspicuity and gap acceptance used questionnaires and performed experiments similar to the studies mentioned in the previous section (Cavallo and Pinto, 2012; Crundall et al., 2008; Gould, Poulter, Helman and Wann, 2012; Horswill et al., 2005; Ragot-Court et al., 2012) or traffic accident databases (Clarke et al., 2007; Pai et al., 2009; Radin Umar et al., 1996; Thomson, 1979; Williams and Hoffmann, 1979). Crundall et al., (2012) investigated why car drivers failed to give way to motorcycles at T-junctions using a series of short video clips and then questionnaires. Data were collected from 74 participants, namely 25 novice drivers, 25 experienced drivers and 24 dual drivers. Shahar et al., (2011) implemented more integrated methods of data collection, such as combinations of questionnaires, use of simulator and video clips. Clabaux et al. (2012) aimed to analyse 'looked-but-failed-to-see' accidents with in-depth accident data involving 44 cases.

Summing up, the fact that the real-time interactions that take place before and during an accident cannot be captured by simulators and, consequently, there is need for naturalistic data that reflect the real and not the stated riding behaviour.

2.3.2.3 Accident frequency/rates

Investigation of the risk of accident occurrence concerning PTWs, as expressed through accident frequency and accident rates is primarily based on historical statistics. The main source of data concern safety administrations, organisations (Branas and Knudson, 2001; Houston and Richardson, 2008; Law et al., 2009; Paulozzi, 2005; Paulozzi et al., 2007; Preusser et al., 1995; Supramaniam et al., 1984; Teoh, 2011; Teoh and Campbell, 2010), safety departments (Schneider et al., 2012) and direct police records (Haque et al., 2012; Haque et al., 2009, 2010; Houston, 2007; Kyrychenko and McCartt, 2006; Morris, 2006; Moskal et al., 2012).

Data from hospitals (Ichikawa et al., 2003; Nakahara et al., 2005) or surveys (Harrison and Christie, 2005) were not used so often. On the other hand, more detailed data collection was used in Wanvik's study (2009), which was based on an interactive Internet database containing a huge amount of data concerning injury accidents and property damage accidents in the Netherlands during 1987-2006.

It is noted, however, that most of the previous data sources (administrations, organisations, etc.) extract their data from police records. Consequently, police records constitute the main source of data. The quality and potential underreporting have to be taken into serious account while analysing that kind of data. Various pre-accident variables such as manoeuvring, braking, acceleration, etc. would give very useful information but is difficult to be acquired as has been very correctly stated in other studies (Lord and Mannering, 2010).

2.3.2.4 Accident severity

The literature indicates that there are many measurements of injury severity. The most common measurements are the Abbreviated Injury Scale, the international classification of diseases (ICD), the five-level KABCO (K [fatal], A [incapacitating injury], B [non incapacitating injury], C [possible injury], O [property damage only]) scale and the injury severity score (ISS) (Lin and Kraus, 2008). The philosophy behind data collection in studies investigating severity of accidents involving PTWs is in a way similar to those studies addressing accident frequency/rates and accident characteristics. A number of studies used national accident databases (De Lapparent, 2006; Pai and Saleh, 2008a, 2008b; Yannis et al., 2005; Zambon and Hasselberg, 2007), police databases (Albalate and Fernandez-Villadangos, 2010; Quddus et al., 2002; Savolainen and Mannering, 2007; Shankar and Mannering, 1996) and data from national statistical institutes or safety departments (Montella et al., 2012; Shaheed et al., 2011). Again, the police records are very often the sole source which provides the data to the organisations. National or local accident databases have been structured on police data. A combination of different approaches can be found in some studies. For example, Majdzadeh et al. (2008) investigated which factors contribute to injuries among car drivers and motorcyclists in Iran by using interviews in hospitals and also police records. A similar approach was followed by Langley et al. (2000).

Overall, underreporting, quantity and quality of available data remain a concern when examining both accident severity and frequency, and have not yet been fully addressed as mentioned previously. For example, Lin and Kraus (2008) argue that motorcycle injuries consistently are less likely to be reported by the police when compared with injuries to other motor vehicle occupants. There is also need for a harmonised collection of data on fatalities, hospital discharges and external causes of injuries (Kisser et al., 2009). Finally, similar to accident frequencies, the collection of pre-accident driving data would be essential but is yet very difficult to achieve in a large-scale extent.

2.3.3 Methods of analysis

2.3.3.1 Rider behaviour

Behaviour of riders in relation to safety was usually addressed using relatively straightforward statistical methods. Descriptive statistics were carried out in a great number of studies as a preliminary analysis to illustrate the patterns of the data or the results. Factor analysis is the most common used method, mainly because the extracted data depend primarily on questionnaires and interviews. It aims to identify factors that explain the patterns of correlations within a set of independent variables or for data reduction purposes. In this case, it is attempted to identify a small number of factors that explain most of the variance that is observed in a much larger number of variables. Many types of factor analysis (exploratory, confirmatory and principal components) have been identified in the literature either as the sole analytical tool (Chen and Chen, 2011; Cheng and Ng, 2010; Ozkan et al., 2012; Ulleberg and Rundmo, 2003) or together with another statistical method (Chorlton et al., 2012; Elliott et al., 2007; Rutter and Quinne, 1996; Schwebel et al., 2006). For example, Goldenbeld et al., (2004) applied factor analysis and linear regression in order to investigate the short- and long-term effects of moped rider training.

Steg and van Brussel (2009) used factor analysis to analyse various attitudes and perceptions towards speeding and also a logistic regression in order to investigate accident involvement. Wong et al. (2010) conducted exploratory factor analysis and a structural equation model based on the derived factors. A slightly different approach was applied by Chen (2009), who explored the relationships between personality factors and attitudes towards traffic safety and risky behaviours among young motorcyclists in Taiwan by means of structural equation modelling.

On the other hand, cluster analysis is an exploratory classification tool and attempts to reveal natural groupings (or clusters) within a data-set. This method was identified very often in PTW literature as well (Brandau et al., 2011; Chung and Wong, 2012; Woodcock, 2007). Moreover, Chang and Yeh (2007) carried out a two-step cluster analysis to classify risky behaviours and a logistic regression analysis to investigate the effect of age, gender and risky behaviours on accident involvement.

Regression analysis and discrete choice analysis have been frequently implemented by a number of researchers when the dependent variable is continuous or categorical, respectively. More specifically, hierarchical multiple regression (Elliott, 2010), separate hierarchical regression analyses (Tunnicliff et al., 2012), logistic regression (Liu et al., 2009; Rathinam et al., 2007) and multinomial regression (Mannering and Grodsky, 1995) have been used in the literature. In addition, survival analysis (for example, the Cox regression) which aims to predict the time to event occurrence was also used in some situations (Yeh and Chang, 2009).

Finally, some other methods have been used, such as correlations among some independent variables and behaviour/perception (Cheng et al., 2011), analysis of variance (ANOVA) (Di Stasi et al., 2009; 2011; Rosenbloom et al., 2011), log-linear models (Haque et al., 2010) and non-parametric tests including chi square tests (Broughton et al., 2009; Dandona et al., 2006; Maestracci, et al., 2012).

2.3.3.2 PTW interaction with other motorised traffic

The basic analytical tools when investigating PTW interaction with other motorised traffic are the traditional statistical methods, and they have many similarities with those investigating PTW behaviour. It is noted though that the analysis methods are heavily dependent upon the kind of data. For example, factor analysis is preferred when analysing questionnaires and was chosen by some studies (Crundall et al., 2008; Horswill and Helman, 2003).

Discrete choice methods have been widely applied. A binary logistic regression analysis and a Cox proportional hazard regression model were selected by Li, Doong, Huang, Lai, and Jeng (2009) in order to investigate the survival hazards of road environment factors between motor vehicles and motorcycles. Mixed logit analysis was also found to have been performed (Pai et al., 2009), and also linear or log-linear models (Gershon et al., 2012; Haque et al., 2012).

More straightforward statistical methods, such as chi square tests and ANOVA were also found in literature (Clarke et al., 2007; Crundall et al., 2008; Shahar et al., 2010). Shahar et al. (2011) carried out ANOVA and regression analysis to analyse attitude change of car drivers towards motorcyclists.

In general, statistical analyses are the predominant tool for analysing interactions of motorised users. However, an entirely different tool that is totally absent in PTW literature is game theory. According to this approach road users are players, they follow strategies and finally make decisions according to the pay-offs. This approach may be fruitful for analysing such interactions and need further research (Elvik, 2012).

2.3.3.3 Accident frequency/rates

Examining the frequency of accidents is not always the best way to measure the risk of an accident and road safety. Alternatively, rates of accidents or fatalities per defined unit describe better this phenomenon. For example, Houston and Richardson (2008) chosen three measures to normalise fatalities by risk exposure; namely, motorcyclist fatalities per 10,000 registered motorcycles, number of these fatalities per 100,000 residents and number of fatalities per 10 billion vehicle miles travelled. However, many studies found during the literature review dealt with accident (or fatality) frequency and rates in general, and do not examine motorcycles exclusively.

Count-data modelling is a common method to deal with accident frequency because of the fact that the number of accidents (or fatalities) is a non-negative integer and as a result ordinary least squares regression is not appropriate. Typical models such as Poisson regression models have been found in some studies (Houston, 2007; Teoh and Campbell, 2010). Poisson models are basic models and easy to estimate but cannot handle under-or-over dispersion (Lord and Mannering, 2010).

Abdul Manan and Varhelyi (2012) analysed motorcycle fatal accident data in Malaysia, in terms of frequency and patterns by type of various parameters such as location, area, road, time, accident type, gender and age. A similar approach was followed by Oluwadiya et al. (2009). Lin et al., (2003) investigated the relationship of the risk of a motorcycle accident to the potential risk factors by applying an Andersen Gill multiplicative intensity model (Andersen and Gill, 1982), which is a generalisation of the Cox proportional hazard model. Factors that increase motorcycle rider risk compared to car driver risk were examined by means of advanced logistic models by Keall and Newstead (2012). A fixed effects negative binomial regression analysis was carried out by Law et al. (2009), aiming to investigate the factors which are associated with the relationship between motorcycle deaths and economic growth.

A study carried out by Hyatt, Griffin, Rue, and McGwin (2009) illustrated time-series analysis using autoregressive integrated moving average (ARIMA) models to estimate the association between regular-grade gasoline price, injury and mortality rates. Studies aimed to estimate fatality rates followed a different approach. In those cases, fatality rates were estimated by linear regression (French et al., 2009; Houston and Richardson, 2008; Supramaniam et al., 1984) or generalised linear regression models (Harnen et al., 2003; Morris, 2006). However, discrete choice models such as logistic regression was also used in some cases. Haque et al. (2009) selected logistic regression in order to differentiate between at-fault and not-at-fault accidents of motorcyclists.

Other statistical tests such as non-parametric tests were also present in literature as they are simple and straightforward methods of statistical analysis (Branas and Knudson, 2001; Daniello and Gabler, 2011; Harrison and Christie, 2005; Kasantikul, et al., 2005; Mayrose, 2008; Ouellet and Kasantikul, 2006; Paulozzi, 2005; Teoh, 2011; Xuequn et.al., 2011).

International literature indicates that the primary analytical tools were traditional statistical methods. However, more advanced statistical methods were found. For example, Haque et al. (2010) developed hierarchical Bayesian models to investigate motorcycle accidents at four legged and T- signalised intersections in Singapore. This promising method has not been extensively used in analysing safety data with some exceptions (Ahmed et al., 2011; Yu et al., 2013).

Some alternative methods, such as artificial neural networks (ANN) have not been applied to motorcycle accident data yet. However, Chang (2005) conducted and compared negative binomial regression to ANN to analyse freeway accident frequencies. The author concluded that ANN is a well-promising alternative method to analyse accident frequency as it does not require any pre-defined underlying relationship between dependent and independent variables, and is efficient especially when dealing with prediction and classification problems.

2.3.3.4 Accident severity

The dependent variable of severity generally consists of two or more discrete categories, e.g. fatal/non-fatal or no injury, possible injury, evident injury, disabling injury and fatality. The basic factor that defines the selected method is the nature of the dependent variable. When the dependent variable has two categories, logistic regression is the most common approach. Although the nature of accident severity is ordinal, recognising or denying this nature leads to different statistical approach. For example, ordered logit or probit models are appropriate if accident severity levels are considered ordinal. On the other hand, multinomial or nested logit models are appropriate when accident severity levels are considered unordered or nominal.

Literature indicates that various analysis approaches have been selected by researchers to analyse PTW accident severity data, such as ordered logit or probit models (Albalate and Fernandez-Villadangos, 2010; Pai and Saleh, 2008a, 2008b; Quddus et al., 2002), multinomial models (Shankar and Mannering, 1996) and binary logistic regression models (Keng, 2005; Majdzadeh et al., 2008; Pai, 2009; Zambon and Hasselberg, 2006, 2007). For example, Savolainen and Mannering (2007) applied a nested logit and a standard multinomial logit model in order to analyse motorcyclists' injury severities in single- and multi-vehicle accidents at Indiana. Various ordered response logit models were selected by Rifaat, Tay, and De Barros (2012) in order to investigate the severity of motorcycle accidents in Calgary.

Other analytical methods have also been carried out when a different approach to examine PTW accident severity was selected, such as log-linear models (Haque et al., 2012) and classification trees (Montella et al., 2012). Loglinear analysis was the chosen method followed by Yannis et al. (2005) in order to examine the first- and second-order effects of accident severity, driver age and two-wheeler engine size. De Lapparent (2006) conducted empirical Bayesian analysis for accident severity of motorcyclists (material damages only, slight injury, severe injury and fatal injury) in large French urban areas. More specifically, this approach used an empirical Bayesian method based on the multinomial Dirichlet model.

To conclude, it is observed that traditional statistical methods dominate the field of PTW accident severity. Other computational intelligent methods such as ANN that were also discussed previously have not been applied so far in PTW severity. However, such methods have been applied to analyse severity of accidents or occupant severity (Abdel-Aty and Abdelwahab, 2004; Chimba and Sando, 2009; Delenet al., 2006).

2.3.4 Contributory factors

It is noted that the purpose of the Subsection 2.3.4.1 is to demonstrate the factors which influence behaviour and also predict risky behaviours such as speeding and violations.

2.3.4.1 Rider behaviour

Riding is a complicated task that requires a lot of attention and personal skills. Perceptions and attitudes of PTW riders are considered important because they may reflect their real riding behaviour. Moreover, attitudes towards traffic safety are directly related to risky riding behaviour (Chen, 2009). Ulleberg and Rundmo (2003) designed a questionnaire and attempted to measure risk perception and attitudes towards traffic safety and self-reported risk taking behaviour of adolescents in Norway. Moreover, risky intentions could be predicted by attitudes and sensation seeking (Tunnicliff et al., 2012).

The fact that riding a motorcycle is a dangerous activity may have different effects on the riders' behaviour. For example, risk-seeking individuals may be attracted to this activity (Mannering and Grodsky, 1995). On the other hand, their hazard perception capability is highly likely to be better than that of car drivers (Horswill and Helman, 2003; Rosenbloom et al., 2011). However, this result is not always supported (Maestracci et al., 2012) and some recent studies have shown that this difference in hazard perception capability may be moderated by driving experience (Crundall et al., 2012). Hazard perception is related to risky behaviours, and PTW riders are aware of the risks but believe that those risks are overcome by their skills and experience (Musselwhite et al., 2012). Aside from hazard perception, the decision of a rider to show risky behaviour could also be affected by the perception that this risky behaviour could be detected and by the likelihood of receiving punishment (Rathinam et al., 2007).

A factor that significantly affects riding skills, riding behaviour and also hazard perception is riding under the influence of alcohol (Creaser et al., 2009; Hosking et al., 2010). The effect of alcohol seems to significantly increase the odds of severe and fatal injuries regardless of socio-demographic attributes (Vaez and Laflamme, 2005).

Training and experience seem to also have significant effect on riding behaviour. Training can lead to an improvement in the riding skills of first-time riders, reducing the number of accidents (Di Stasi et al., 2011). Liu et al. (2009) found that novice riders were overconfident about their abilities and they perceived hazards in a less appropriate manner than experienced ones. Other studies have shown that experienced riders respond faster to hazards than inexperienced ones (Hosking et al., 2010). However, skills acquired through training may not remain in the long run (Goldenbeld et al., 2004). Self-assessment tests that have been performed in car drivers have revealed interesting results (De Craen et al., 2011) and need to be performed also to riders.

Age and gender are factors which distinguish heterogeneous rider groups in terms of decision making and influence their risky riding behaviour (Chung and Wong, 2012). Being young and male is associated with risky behaviours (Lin et al., 2003; Mannering and Grodsky, 1995). It is interesting that the three primary attributes of young motorcyclists seem to be sensation seeking, amiability and impatience (Wong et al., 2010). Rutter and Quine (1996) state that age and more specifically youth plays a more significant role than inexperience. Moreover, young riders do not seem to wear protective equipment (De Rome et al., 2011).

Age and gender are also related with errors and violations and more specifically, young and male riders were more likely to disobey traffic rules (Chang and Yeh, 2007). Young riders were also more likely to be unaware or neglect potential risk. Dandona et al. (2006) interviewed PTW riders above 16 years old at petrol filling stations in India and found that about half of the riders committed at least one (assessed) violation during the last three months.

Errors and violations are affected by some other parameters as well. Schwebel et al. (2006) argue that sensation seeking seemed to be the most significant predictor of self-reported riding violations compared to the other parameters (anger/hostility, personality, etc.). A common violation and risky behaviour is considered to be excessive speeding and it is probable that it is affected by many factors and their interactions. The PTW riders' attitudes towards speeding are likely to predict actual behaviour. Psychological flow variables such as perceived enjoyment and concentration seem to positively affect motorcyclists' speeding behaviour (Chen and Chen, 2011). In the same study, it is argued that personal factors (e.g. personality traits, experience and gender) reflect differences in motorcyclist speeding behaviour.

Furthermore, Rathinam et al. (2007) found that young riders were driving faster when they were angry than when they were in any other mood. Steg and van Brussel (2009) argued that moped riders were more likely to speed and disobey speed limits when they have a positive attitude towards speeding, and when they think that other road users expect them to speed. These results were consistent with Elliott (2010) who attempted to investigate which cognitive factors affect motorcyclists' intentions to speed. Chorlton et al. (2012) attempted to predict motorcyclists' intention to ride above the speed limit and also at inappropriate speeds. Some interesting findings were that speeding on motorways would allow riders to beat the traffic and also feel exhilarated.

Finally, according to several studies (Cheng and Ng, 2010; Haque et al., 2010; Rosenbloom et al., 2011), previous involvement in accidents (e.g. past history of accidents) was related to aggressive behaviour (and also sensation seeking) of riders.

2.3.4.2 PTW interaction with other motorised traffic

PTW riders are more vulnerable than car drivers and when car drivers deviate from expected and proper behaviour, they constitute a potential risk to PTW riders (Ragot-Court et al., 2012). It is interesting that car drivers who also hold a motorcycle licence are less responsible for car-motorcycle accidents than those who hold only car driving licence (Magazzu et al., 2006). Moreover, Shahar et al. (2010) revealed that dual drivers responded better to hazards at junctions and also performed better than either experienced or novice drivers.

The attitudes of car drivers towards motorcyclists may influence the interactions between them but have not been extensively investigated. Most empathy towards motorcyclists stems from male drivers who are or know motorcyclists (Musselwhite et al., 2012). Crundall et al. (2008) carried out a survey in order to investigate the car drivers' attitudes towards motorcyclists and some interesting findings suggested that car drivers with an amount of experience between 2 and 10 years expressed the most negative views. Shahar et al. (2011) attempted to deploy a strategy in order to reduce the negative attitudes of car drivers towards motorcyclists.

A number of accidents occur because of the fact that car drivers did not detect the motorcyclist or because the car driver detected the motorcyclist but failed to judge correctly the speed/distance of the oncoming motorcycle (Haque et al., 2012; Thomson, 1979; Williams and Hoffmann, 1979). The fact that motorcycles have relatively small size makes their detection by car drivers more difficult. In regard with night driving, it seems that car drivers are more accurate in judging the speed of cars than motorcycles (Gould et al., 2012). On the other hand, the use of the daytime running lights in cars deteriorates conspicuity (Cavallo and Pinto, 2012). Gershon et al. (2012) concluded that the ability to detect a PTW is affected by a handful of visual factors that have a relation to the PTW, its rider, the driving environment and the car driver's level of awareness. It is important to note that 'looked-but-failed to-see' accidents are over-represented in intersections and constitute a significant contributory factor to PTW accidents (Clabaux et al., 2012; Clarke et al., 2007; Crundall et al., 2012).

2.3.4.3 Accident frequency/rates

A variety of accident-related factors have been identified during literature review. The road environment such as road type, road geometry and roadside installations have been found to have an influence on PTW accident occurrence (Harnen et al., 2003; Kasantikul et al., 2005; Wanvik, 2009). For example, Haque et al. (2010) argue that the number of lanes at four-legged signalised intersections significantly increases motorcycle accidents. In the same study, it is stated that motorcycle accidents increase in high-speed roadways. In the USA, almost 80% of PTW accidents occurred in urban or suburban areas.

Haque et al. (2009) examined fault among motorcyclists involved in accidents and indicate that several geometrical and environmental factors were responsible for non-at fault accidents, for example, wet road surfaces, single-lane roads and median lanes of multi-lane roads.

Aside from the road characteristics, another category of risk factors is vehicle characteristics such as aerodynamic behaviour and engine size. As indicated in a study made by Teoh and Campbell (2010), rider death rates for super sport motorcycles were four times higher than those for standard motorcycles. On the other hand, some vehicle technological characteristics such as anti-lock braking system (ABS) or autonomous emergency braking may be beneficial for the safety of riders (Savino et al., 2013a; Savino 2013b; Teoh, 2011).

Exposure is also a critical factor (Haque et al., 2010; Keall and Newstead, 2012). Harrison and Christie (2005) stated that the rate of accident involvement per kilometre travelled decreases as current riding exposure rises. Lin et al. (2003) state that some exposure factors (number of riding days, average riding distance) were found to increase the risk of being involved in an accident.

The effect of protective equipment such as helmets on reducing fatality rates of riders has been addressed in numerous studies (Branas and Knudson, 2001; Dee, 2009; French et al., 2009; Kyrychenko and McCartt, 2006; Mayrose, 2008; Moskal et al., 2012; Ouellet and Kasantikul, 2006). Helmet laws generally enhance PTW safety

(Kyrychenko and McCartt, 2006; Morris, 2006). Houston (2007) argues that universal helmet laws result in less fatality rates among young motorcyclists. On the other hand, the widespread use of non-standard helmets in low- and middle- income countries may limit the potential gains of helmet use programmes (Ackaah et al., 2013).

However, helmet use seems to have no relation with the risk of being involved in an accident (Lin et al., 2003). De Rome et al. (2011) found no association between riding unprotected and other risk-taking factors. Behavioural characteristics constitute another category of factors correlated with accident involvement. It is noted that factors such as age and experience also have an effect on accident occurrence. Those factors have been discussed in previous sections and also how they predict risky behaviour. Alcohol consumption is a very important behavioural factor related to increased PTW risk and high number of accidents (Ahlm et al., 2009; Huang and Lai, 2011; Lin et al., 2003; Kasantikul et al., 2005; Preusser et al., 1995; Teoh, 2011).

The interaction of behavioural factors seems to increase the risk of accidents. Moskal et al. (2012) indicated that being male, exceeding the legal alcohol limit and travelling leisure trips are related with increased risk. Bjørnskau et al. (2012) attempted to relate rider characteristics, behaviour and accident risk in Norway. The authors conclude that 'the combination of low age, low experience, risky behaviour and "unsafe" attitudes seems to be a particular potent risk factor for Norwegian motorcyclists'. Keall and Newstead (2012) compared the risk between car and motorcycle riders, and found that especially young riders or riders who live in more urbanised settings were exposed to more risk.

The interaction of behavioural characteristics with other factors as well was identified as critical for PTW safety. Schneider et al. (2012) concluded that younger motorcyclists are more likely to be at-fault in an accident, as are riders who are under the influence of alcohol, riding without insurance or not wearing helmet. Oluwadiya et al. (2009) found that the risky behaviour among motorcycle riders interacting with chaotic traffic and road design faults was responsible for the majority of the motorcycle accidents in Nigeria.

Economic indicators seem to be associated with PTW deaths and rates (Law et al., 2009). For instance, Hyatt et al. (2009) observed that whilst the number of injuries and fatalities in motorcycle-related multi-vehicle accidents rise, gasoline price in the USA rises, and rates remained in a great extend stable. Furthermore, Houston (2007) and Houston and Richardson (2008) identified income per capita and registered motorcycles per capita as contributory factors. Lastly, two factors that have not been investigated in a large extent are the weather and traffic characteristics. The effect of those factors on PTW safety has not been deeply explored. Various studies have addressed the effect of weather on vehicle accidents and rates but literature regarding PTW accidents and some weather effects is limited (Branas and Knudson, 2001; Houston and Richardson, 2008; Xuequn et al., 2011). On the other hand, the association between traffic volumes and speed with PTW accidents has not been explored and need further research.

2.3.4.4 Accident severity

Critical factors which affect PTW accident severity are in a way similar to those which affect accident frequency and rates, for example, road infrastructure characteristics, vehicle characteristics, behavioural and environmental factors. Nevertheless, accident severity is influenced in a different way. Consequently, strategies that target accident occurrence are different from strategies targeting mitigation of accident severity.

Road characteristics and roadside installations have been identified as factors that increase PTW accident severity (Albalate and Fernandez-Villadangos, 2010; Rifaat et al., 2012). For example, accidents at curves are associated with fatal injuries (Montella et al., 2012). Daniello and Gabler (2011) indicate that PTW collisions with guardrail are more likely to result in a fatal accident than collisions where the rider hits the ground.

Road Restraint Systems constitute a very essential part of modern road infrastructure and are among the most important life-saving devices available to public authorities and road operators. Not only can they save lives but also significantly reduce the accident related health care cost. Nikiforiadis (2009) provide an excellent review of road restraint systems.

Another typical risk factor is a right-of-way violation. Pai (2009) argued that accident severity is increased when a travelling-straight motorcycle on the main road accidents with a car coming from the minor road and intends to turn right. Moreover, this situation deteriorates particularly at stop-/yield-controlled junctions. Another study (Pai and Saleh, 2008a) indicated that in right-of-way violations, the more severe injuries appeared when stop, give-way signs and markings controlled the junction.

Low visibility is related with increased injury severity of motorcycles (Savolainen and Mannering, 2007). Lighting also affects PTW severity. When motorcycle accidents occurred during the night when the lighting conditions were poor, they resulted in higher injury severity (De Lapparent, 2006; Pai and Saleh, 2007).

Some behavioural factors such as speeding, alcohol consumption and non-helmet use are associated with more severe injuries (Albalate and Fernandez-Villadangos, 2010; Nakahara et al., 2005; Savolainen and Mannering, 2007; Shankar and Mannering, 1996; Zambon and Hasselberg, 2007). High speed causes more difficulty in manoeuvres than passenger cars due to complex dynamics (Elliott et al. 2007; Horswill and Helman, 2003), the stopping distances rise and also results in more severe accidents due to the high amount of kinetic energy. Non-helmet use constitutes a very important factor as it keeps the rider’s neck and head unprotected. The effect of this factor has been highly addressed in many studies (Gabella et al., 1995; Keng, 2005). Moreover, the type of collision also has an effect on PTW accident severity (De Lapparent, 2006; Pai and Saleh, 2008a, 2008b; Shaheed et al., 2011).

Riders' age is associated with increased accident severity (Gabella et al., 1995; Nakahara et al., 2005; Pai, 2009; Savolainen and Mannering, 2007; Yannis et al., 2005). A study (Donate-Lopez et al., 2010) stressed that each one-year increase in age is related with a 3% increase in the risk of death. De Lapparent (2006) stated that women riders between 30 and 50 years old driving motorcycles with high engine size are the most exposed to risk of injury.

In most studies, the interaction of factors seems to affect PTW accident severity rather than a single type of factor. Quddus et al. (2002) identified the most influential factors that result in higher motorcycle accident severity and they appear to be collisions with stationary objects, collisions with pedestrians and the motorcycle engine capacity. Other studies also support this evidence that motorcycle engine size is a contributory factor to accident severity (De Lapparent, 2006; Langley et al., 2000; Pai, 2009; Yannis et al., 2005). Pai and Saleh (2008b) investigated the factors that increase motorcyclists' accident severity by various accident types at T-junctions in the UK and revealed a very interesting number of factors, such as being male and elderly, increased engine size, early morning riding, type of season (spring and summer), fine weather conditions, insufficient lighting, non-built-up areas and collisions with heavy vehicles. Savolainen and Mannering (2007) showed that increasing motorcyclist age, collision type, roadway characteristics, alcohol consumption, non-helmet use and unsafe speed were statistically significant. Similar to the PTW accident frequency, the effect of weather on PTW accident severity has not been excessively explored.

Finally, it is worth mentioning that the socio-economic factors seem to affect accident severity. More specifically, it was found that young motorcycle riders in lower socio-economic groups have higher odds for both minor and severe injuries than those in the highest socioeconomic group (Zambon and Hasselberg, 2006). To conclude, it is obvious that only a few studies investigated the effect of some factors such as real-time speed, weather and traffic on PTW severity. Christoforou, Cohen, and Karlaftis (2010) investigated the association between accident severity and traffic characteristics collected real time during the time the accident has occurred. Quddus, Wang and Ison (2009) used real-time data from the time of the accident and found that as the traffic flow increases the accident severity decreases. There is a clear lack of studies examining the effect of real-time traffic and weather characteristics.

2.3.5 Summary

Powered-two-wheelers constitute a very vulnerable type of road users. The notable increase in their share in traffic and the high risk of severe accident occurrence raise the need for further research. However, current research on PTW safety is not as extensive as for other road users (passenger cars, etc.). A critical review of research on Power-Two-Wheeler behaviour and safety was provided with regard to data collection and methods of analysis. The main contributory factors were examined separately and in conjunction with the other factors. Both macroscopic analyses (accident frequency, accident rates and severity) and microscopic analyses (PTW rider behaviour, interaction with other motorised traffic) were examined and discussed in this section.

2.4 Research gaps and suggestions for further research

2.4.1 Effect of traffic and weather characteristics on road safety

Despite the efforts so far, a number of the various factors affecting road safety still need further investigation and understanding. Traffic and weather characteristics belong to these factors. The literature review presented in this chapter, has showed that most of traffic parameters were found to have mixed effects on road safety and thus need further research. Concerning traffic flow, it seems to be related non-linearly with accident rates, while some studies imply linear correlation with the number of accidents. On the other side, the effect of traffic flow and congestion on accident severity is not clear as far less studies have been examined this phenomenon. Particularly congestion is not extensively investigated and was mostly measured by proxies as correctly stated by Wang et al. (2013a). As a consequence, further research is needed to overcome past limitations and past contradictory findings. It is worth mentioning that traffic density and occupancy have received less attention from researchers than traffic flow. One possible explanation may be that researchers preferred to use traffic parameters like flow instead of other parameters (density for instance), which cannot be directly measured but estimated from fundamental traffic relationships.

Similarly, earlier studies also investigated the effect of speed limits in order to indirectly measure the effect of changes in speed. The effect of speed may be theoretically easy to estimate, however, literature review showed that this effect is not always straightforward. Speed limits have found to influence accident in a more straightforward way, but it has to be stressed, that most of related studies were carried out in the U.S. Consequently, more research regarding European countries or less developed countries could provide more insight. The influence of speed variation although implied to be more hazardous than speed itself, has not excessively been investigated in the past, possibly because these type of data were difficult to be obtained and examined. The implementation of real-time strategies and access to high resolution microscopic traffic data is expected to provide more insight to this problem.

Regarding weather characteristics, the wide majority of literature investigated the impact of precipitation usually aggregated monthly or daily. In general, it seems to have a consistent effect on safety but some recent studies carried out in Mediterranean areas present contradictory results. Other parameters such as temperature and wind speed have not sufficiently been examined while others (low visibility) have a more consistent effect. The recent increasing implementation of traffic and weather characteristics measured at real-time is opening new possibilities in order to assess freeway safety and also to provide insight on past contradictory findings. The most common related risk factors include the average, standard deviation and coefficient of variation of speed and flow aggregated at different upstream and downstream detector locations with respect to the accident location. Weather risk factors include rainfall and low visibility, having more consistent effects than past studies.

An obvious limitation is that the wide majority of studies focus on freeways or urban expressways, while rural and urban roads have not been examined yet. Investigation of these types of areas are more challenging since they are more complex in terms of traffic but also more interesting in terms of safety. Shew et al. (2013) mention that it is possible that a model can be applied in multiple freeways due to similar relationships between traffic and accident risk. Testing the transferability of this methodology and results to urban roads or between urban roads is yet to be investigated. It has to be noted though, that data collection is more difficult in these cases, since management of rural and urban arterials is definitively more complex than managing freeways.

Moreover, there is no specific focus for vulnerable road users such as power-two wheelers, cyclists or pedestrians with some exceptions. Especially power-two wheelers, due to their increased interaction with other heavier motorized traffic may be affected by changes in traffic states and speed variations. Moreover, due to complex dynamics of motorcycles, wet pavement conditions are high likely to be a major contributory factor. It has to be noted though, that the riders may not use their moped or motorcycle under adverse weather conditions. Ouellet et al. (2002) argue that weather conditions are rarely a contributory factor, after an in-depth investigation of 1082 motorcycle accidents in Thailand. As a result, more research is definitively needed to measure the effect of weather conditions especially on power-two wheeler safety.

Concerning methodology, the type of analysis is defined by the type of the dependent variable which is of interest and is straightforward. For instance, accident frequency is analysed by means of count regression techniques (Poisson or negative binomial models) as expected, although recent studies use more advanced models. Logit models dominate the area of accident severity. When examining traffic factors, it can be seen that while early studies were relatively simple in terms of analysis (power functions or simple linear models), there has a sharp increase in the complexity of recently applied statistical models.

Concerning the weather effects, time series analyses are also performed in order to correlate daily or monthly weather data and accidents (Karlaftis and Yannis, 2010; Bergel-Hayat et al., 2013).The majority of recent literature focused on predicting real-time accident probability but only a handful of studies examined accident frequency or accident severity. Accident probability was usually estimated by the use of matched-case control models proved to have adequately predictive power. Nevertheless, it would be interesting to apply another fruitful methodology such as the rare events logit model, which is capable of analysing rare-events data. In such cases, the binary dependent variable has dozens to thousands of times fewer events (such as wars, earthquakes) than non-events. For more information, the reader is encouraged to read King and Zeng (2001b).

As mentioned earlier, accident frequency and severity studies with real-time data are not so many yet, and consequently, more research effort is needed. It is noted that accident frequency studies which are more disaggregate examined homogenous segments, usually consider one year of data. It is challenging to apply a methodology to include more years of data or at least analyse each year separately and compare the findings to understand the seasonal evolution. Real-time traffic variables were extracted for each time slice before the accident and the importance of each variable is measured to assess short-term traffic safety. A slightly more macroscopic approach may be useful by incorporating time series of traffic parameters and examine their impact on safety in a more dynamic way.

Real-time weather characteristics have been explored adequately and have more consistent results. Nevertheless, regional differences among countries such as the U.S. and other European countries should be explored as well. Especially Mediterranean countries, where good conditions are prevailing and adverse weather is rarer, specific attention should be devoted to under-stand potential differences in weather. Since weather greatly varies and could be very different even in nearby areas, the use of interpolation methods may seem essential in order to analyse quantitative weather parameters in more detail.

As mentioned earlier, weather effects are having a more consistent effect on safety than traffic. Despite that, there are still some issues that need to be tackled, such as the negative correlation of adverse weather and road safety of some recent studies as well as the insignificance of lag effect of rain in Mediterranean countries such as Greece. As a result, further research is needed in order to confirm the lag effect of rain in such countries. Further research is needed in order to understand some issues such as transitions from good to adverse weather conditions and vice versa, and acquire a more clear picture of the weather effects on accident severity. It is also important that many weather parameters are correlated such as low temperatures and snow or rainfall, so they must be treated with specific care. Moreover, since weather is a local phenomenon, and there is possible to observe significant variation in weather phenomena in some areas, interpolation methods (such as Thiessen polygons and inverse distance weighting) need to be applied since there were not widely used so far in literature.

A final remark concerns need to further understand the combined effect of weather and traffic. The impact of these factors need to be examined carefully when they are co-considered, since they are likely to be negatively correlated (Keay and Simmonds, 2005) as stated previously. It is imperative to understand the effects of weather and traffic clearly and draw robust conclusions. The application of real-time weather data may be proved very beneficial as discussed in this chapter. Concluding, research in the field of traffic and weather impact on road safety is of critical importance, it is constantly rising and the exploitation of real-time traffic and weather data might provide answers to several open questions. The exploitation of these large amounts of data through appropriate advanced real-time analyses could result in highly effective pro-active traffic safety management systems with great potential for safety improvement. On that purpose, research in the field of traffic and weather impact on road safety with the use of real time data should be intensified.

2.4.2 PTW behaviour and safety

First of all, data basically stem from questionnaires, police records and sometimes experiments. Questionnaires are not the best solution to measure human behaviour, simply because the stated behaviour can be different from the real behaviour. In that context, experiments seem to gain more attention in PTW literature, especially those experiments which try to measure physical attributes such as eye tracking that indirectly express human behaviour. The increased use of enhanced simulators is definitely a step towards the right direction. However, there is need for new naturalistic data which may enable the monitoring of participants' observed behaviour in high detail in real-time situations such as seconds before or during an incident or an accident. PTW interaction with other motorised traffic has been dealt with a similar approach. The use of naturalistic data will also supply researchers with information about observed real-time interactions between PTW riders and car drivers in several situations such as overtaking, braking, evasive actions, giving way, and so on. Furthermore, simulators that would enable various road users to interact in pre-selected scenarios could be helpful to observe these interactions between them.

As far as PTW accidents are concerned, relying mostly on police data records is not always the most appropriate way to analyse PTW accidents either in terms of accident frequency or severity. The reason is that many critical variables may be missing and cannot be measured at post-accident phases, for example traffic volumes, traffic density, speed, pre-accident manoeuvres, overtaking, lane changing, braking/accelerating, etc. This is also mentioned in Lord and Mannering (2010). In addition, police reports may overestimate accident severity remarkably (Tsui et al., 2009). Consequently, there is need for a detailed observation of road sites constantly and for a detailed accident recording system. In this case, naturalistic data which involve accidents may be a good alternative in order to solve these issues. Another approach, of course, is the modelling of near missing accidents (or incidents). A near missing accident or an incident that happen today, could be an accident in the future, so modelling near misses and incidents could provide interesting information. This approach is implemented widely in aviation safety, where the small number of accidents prohibits the classic methodological approach.

In terms of methodology, it is observed that the large majority of studies in regard with every aspect of PTW safety are based upon classic statistical methods with some few exceptions which are mentioned earlier such as neural networks. However, rider behaviour (and, of course, human behaviour) is a very complex variable which is expressed via various actions while riding but even before riding (e.g. alcohol consumption). Thus, modelling human behaviour is a very challenging task and may need a different approach. For example, in transportation systems, legal and social constraints, laws and restrictions may define behaviour, in a sense that human actions and system evolution can be predicted and outcomes may be determined by system dynamics (Frazier and Kockelman, 2004). Then chaos theory may apply to model human behaviour and also to (accident generating) system outcomes. For example, do PTW accidents occur randomly? Do they occur in a designed way that could be called 'deterministic'? In the second case, knowing the initial conditions will lead to accurate forecast. Moreover, traffic may be chaotic and its effect on PTW accidents has not been investigated so far. In any case, chaos applications with regard to PTW accidents seem to be an interesting and promising approach, especially when non-linear relationships seem to exist.

When it comes to interaction, there is not much difference from behaviour in terms of methodology used so far. However, there is one point that has not been investigated sufficiently yet. In every aspect of daily life, humans have to make decisions. Some people even compare real life to chess, where players constantly make decisions and immediately face the effect of each decision. Interactions between road users take place when overtaking, lane changing, braking, giving way, approaching an intersection, and so on. Many studies have focused on gap acceptance, conspicuity and errors but no studies have explored the potential strategy that each individual uses in the transport network and its effect on accidents. In that context, every road user becomes the ‘player’ and every decision has its ‘pay-off’. Authors considered a game theoretical approach to behaviour and interaction as a fruitful direction for further research. To define a game, it is required to identify the players, their alternative strategies and their objectives.

Several questions need an answer, such as what strategy is used by car drivers when confronting PTWs, whether riders have a strategy, whether they make the optimum decision, how their strategy evolves over time and finally how they adapt to critical situations and whether the various measures make riders to adapt their strategic interaction with other motorised traffic and pedestrians. It is also important to mention that the car drivers’ attitudes and perceptions towards motorcyclists are important to be further investigated given that a large proportion of motorcycle accidents results from errors made by other vehicles (Tunnincliff et al., 2012). Finally, another interesting category of games (aside from games between road users) is that which present the interaction between road users and authorities. An authority could be, for example, the police. Each player has different strategies and aims. This could also be very interesting if further explored as Bjørnskau and Elvik (1992) argue. In that study, a game theoretic model was developed in order to explore if traffic law enforcement can permanently reduce accidents. It is stated that the linkage between accident rates, violation rates and police enforcement, examined via a game theoretic model, has not tested so far.

Computational intelligent methods, such as ANN have not been used widely for PTW safety analyses. Neural networks allow non-linear relationships between dependent and independent variables and require no prior assumptions. Nonetheless, the interpretability of their results is not straightforward as in traditional statistical methods of analysis which offer more interpretable results but have some drawbacks (for example, they require prior assumptions to be valid and cannot handle some problems). Although ANN can overcome many problems of statistical methods, they can be time-consuming. A more detailed review of neural networks in transportation research is presented in Karlaftis and Vlahogianni (2011). Other advanced techniques such as Support Vector Machines and Cusp Catastrophe models which may also assume non-linearity are not widely used especially in PTW safety.

Other statistical methods that have been used in accident frequency analysis are Bayesian models with promising results (Ahmed et al., 2011; Chin and Quddus, 2003; Yu et al., 2013). These response models for count data are specified hierarchically in several layers and Marcov Chain Monte Carlo techniques are required in order to obtain inference (Tunaru, 2002). Consequently, it is concluded that other more intelligent computational methods are applied in PTW safety and test their results in comparison with previous research methods.

A limitation identified in PTW literature is the lack of detailed real-time traffic and weather variables. In the large majority of cases, weather variables at the time of the accident are expressed as ‘good weather’, ‘rainy’, ‘snow’, etc., and not in a quantitative scale. PTW riding involves more direct exposure to weather conditions such as temperature, visibility, wind direction/speed, rainfall intensity, and so on. The real-time quantitative weather effects are entirely new in analysing accident frequency, with only a few studies found to have incorporated such data. On the other hand, some studies attempted to relate traffic congestion and accident occurrences (Park and Abdel-Aty, 2011; Quddus et al., 2009).

In most such studies, the annual average daily traffic (AADT) was used. No real-time traffic data were found to have been used to analyse PTW accident occurrences and fatalities. However, some researchers attempted to incorporate such kind of data to investigate accident occurrence and severity. For example, Yu et al. (2013) analysed freeway accident occurrences using real-time weather and traffic data prior to the accident. Vlahogianni et al., (2012c) attempted to model the effect of such variables on the risk of secondary incidents. Very few studies found to have incorporated real-time traffic data to model accident severity (Christoforou et al., 2010; Quddus et al., 2009). All these studies seem to regard such data as promising tools for researchers. As a result, their effect on specific PTW accident frequencies and severity is a good direction for further research. Moreover, it is possible that weather conditions may affect rider behaviour and even traffic conditions (volumes, speed, etc.).

Although it seems that PTW riders prefer to ride in good weather conditions, accidents in adverse weather conditions do occur. However, traffic volumes are not the same in good and adverse weather, and it would be interesting to study the ratio of accidents to the number of PTWs. Lastly, the effect of some important pre-accident variables, which are mentioned earlier in this section, raises the need for further research. Last but not the least, the relationship between exposure and risk has been mentioned (Harrison and Christie, 2005). Of course, this pre-supposes large and highly detailed databases.

Chapter 3 Methodological approach

3.1 General methodological framework

The purpose of Chapter 3 is to serve as an introduction to the other chapters which present the data collection and processing and the results of the analyses that were conducted. Consequently, this chapter is a demonstration of the overall methodological approach that will lead to the accomplishment of the aim of the thesis as defined in Chapter 1: the integrated multi-faceted statistical analysis of accident probability and severity applying advanced statistical and machine learning methods and exploiting high resolution traffic and weather data with specific emphasis on the analysis of Powered-Two-Wheelers. This chapter aims to provide an introduction to the methodological approach followed and therefore the statistical analysis methodologies are not described in this chapter. Detailed descriptions of each statistical methodology are given in the next chapters where each statistical methodology is applied.

The next figure (Figure 3-1) provides the basic graphical illustration of the general methodological framework that is followed to achieve the aims of the thesis. Initially, data for urban roads are collected, processed and analysed. The first step is a preliminary analysis by applying machine learning techniques on the raw 5-min time series data, in order to acquire a first insight on the phenomena of accident probability and accident type of Powered-Two-Wheelers. Afterwards, accident probability and severity are examined and analysed. Various classification techniques are applied (cluster analysis, random forests and so on) on the hourly aggregated data. Then, an advanced approach is followed, where a series of logistic regression models such as finite mixture, mixed effects and Bayesian models are applied. The methodological toolbox is finalized with a non-traditional approach. This approach utilizes cusp catastrophe theory models which aim specifically to explain accident probability and severity and detect potential non-linearity when modelling these complex phenomena.

This methodological toolbox aims to assist in the identification of the broader picture behind the complex phenomena that take place and more specifically to result to the understanding of accident risk, having specific emphasis and focus on PTWs. For that purpose, different definitions of accident severity were required to be used and then further explored, as noted in Chapter 1.

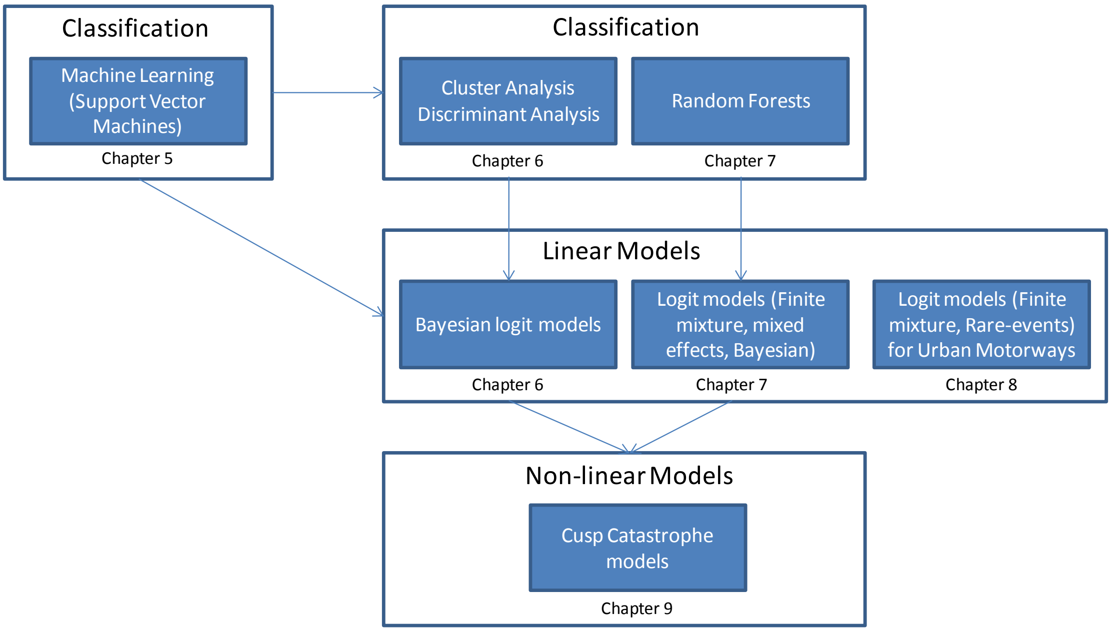


Figure 3-1: Illustration of the general methodological framework.

3.2 Graphical concept

Section 3.2 is dedicated to a graphical representation of the concept behind this thesis. Figure 3-2 is therefore a detailed illustration of the concept and is one step further compared to the general methodological framework presented in the section 3.1.

Firstly, the relevant accident, traffic and weather data are collected and processed, so as to acquire the required databases. The aim is to identify potential risk factors among high resolution traffic (flow, occupancy, speed) and weather (temperature, rainfall, humidity, wind speed, wind direction) parameters, as well as other characteristics (road, vehicle and user characteristics).

The data are extensively explored by utilizing a series of advanced statistical methods. By applying a series of advanced logit models, such as mixed effects and finite mixture models, accident probability, accident severity and occupant injury severity are investigated. Classification techniques, such as random forests, are used for data screening, in a popular technique, where the relevant importance of each variable is assessed before using these variables in the logit models. Using that technique, despite the existence of a large number of candidate predictors, only the most important variables are easily identified. Cluster analysis techniques (finite mixture cluster analysis) are also applied in order to evaluate the impact of traffic states on accident severity and probability.

The next step is to focus solely on Powered-Two-Wheelers. In that context, PTW accident probability and PTW accident type probability are predicted through a combined approach, where time series data (original and wavelet transformed) are used together with advanced machine learning techniques, such as Support Vector Machines (SVM).

It is noted that analytical models for Chapters 5, 6, 7 and 9 are developed by exploiting data extracted from urban roads, while occupant injury severity and PTW occupant injury severity (Chapter 8) are exploited only for urban motorways. Furthermore, accident probability in urban motorways is explored by following a different approach, namely the rare-events logit model. More details about the urban motorway data preparation and the modelling approach are provided in Chapter 8. As mentioned earlier in section 3.1, the methodological approach is completed by applying a series of cusp catastrophe models to investigate accident probability and severity through a different approach, which examines the presence of non-linearity and is described in detail in Chapter 9 of the thesis. There is also specific focus on Powered-Two-Wheelers.

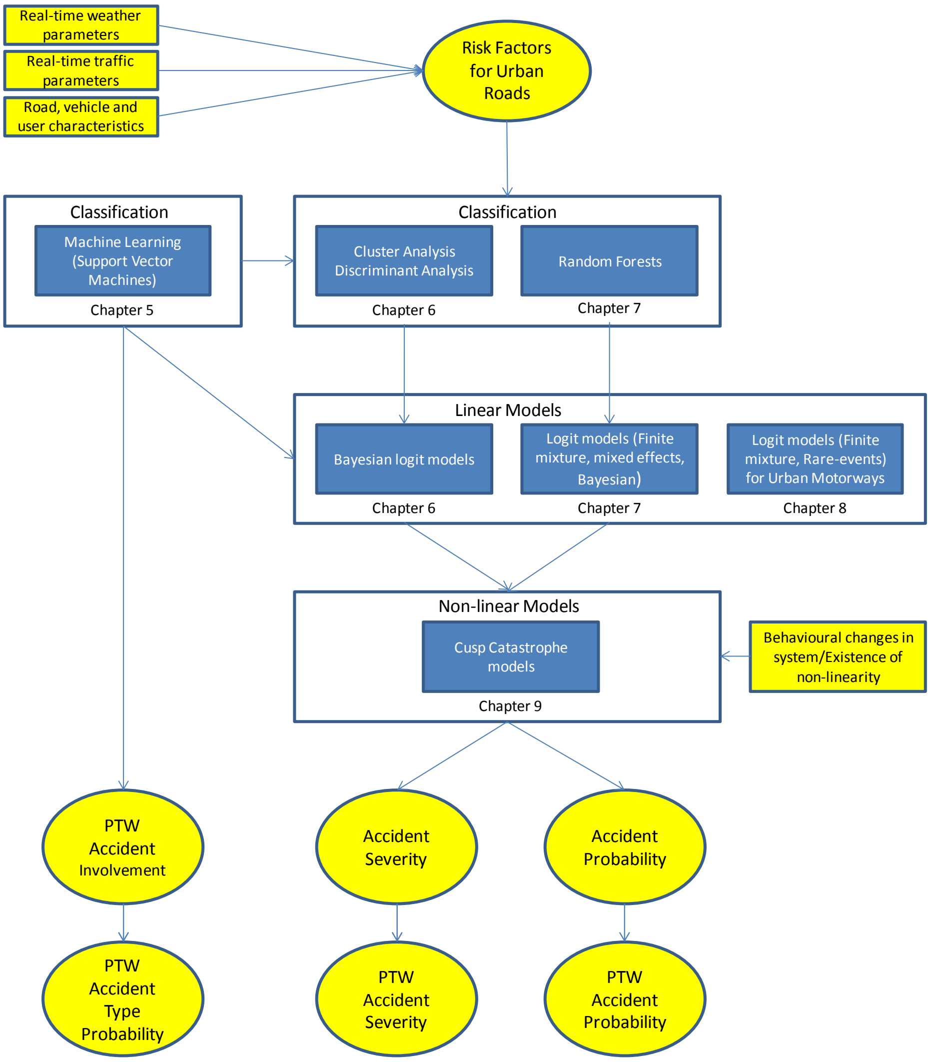


Figure 3-2: Illustration of the graphical concept of the thesis.

3.3 Overview of methodological structure

The methodological structure of the thesis is presented on Table 3-1. The title of each chapter is provided along with the statistical methodology and the road type considered in each analysis (urban road or urban motorway). The table also presents the dependent variables considered in each chapter (e.g. accident probability, accident severity, accident type) and also whether the chapter has a specific dedication to Powered-Two-Wheelers or not. Lastly, the statistical methodology as well as the aim of each chapter are provided (e.g. prediction or explanation of phenomenon).

Chapter	Title	Methodology	Road type		Accident probability		Accident severity		Objective	
			Urban Road	Urban Motorway	All	PTW	All	PTW	Prediction	Phenomenon explanation
5	Time series and machine learning analysis to predict PTW accident involvement and accident type	Time Series/Wavelets, SVM	√			√			√	
6	Modelling the effect of traffic states on road safety in urban roads	Finite mixture cluster, Discriminant analysis, Bayesian logit	√		√	√	√	√		√
7	Multi-statistical investigation of road safety in urban roads	Random Forests, Finite mixture logit, Mixed effects logit, Bayesian logit	√		√	√	√			√
8	Occupant injury severity and accident probability in urban motorways	Finite mixture logit, Rare-events logit, Bayesian logit		√	√	√	√	√		√
9	Cusp catastrophe theory to model accident severity in urban roads	Cusp catastrophe	√		√	√	√	√		√

Table 3-1: General methodological structure.

As shown in Table 3-1, chapter 5 is dedicated to PTWs by utilizing high resolution traffic data from urban roads. A combined approach exploiting time series data and advanced machine learning techniques (Support Vector Machines), is presented so that PTW accident involvement (accident probability) and PTW accident type are predicted. This is one of the first (if not the first) attempts to use this methodology for such research purposes.

In Chapter 6, classification methods such as finite mixture clustering and finite mixture discriminant analysis are applied to classify traffic conditions and then Bayesian logit models are applied to explore the impact of traffic states on accident severity and probability. The aim of this chapter is the explanation of this phenomenon (for the first time on urban roads). Separate models for PTWs are also developed.

Chapter 7, goes one step further and examines the effect of individual traffic and weather parameters. Random forests are applied in order to classify variables according to their relative importance and then the significant variables are entered in logit models in order to investigate accident probability, accident severity and PTW accident probability.

Chapter 8 has a two-fold aim. Firstly, to investigate occupant and PTW occupant injury severity in urban motorways by applying finite mixture logit models. Secondly, to investigate the mechanism of accident and PTW accident occurrence. The stimulation for the research questions of chapter 8 is the fact that there is no specific focus on PTW safety in freeways, as the extensive literature review indicated in Chapter 2.

Lastly, chapter 9 aims to examine the potential presence of non-linearity when analysing accident probability and in urban roads. Cusp catastrophe models were applied for this reason, and therefore this chapter presents a first attempt to apply these complex models. Specific focus on PTWs is being given as well.

This thesis utilizes high resolution traffic and weather data mainly from urban roads which have never been exploited in the past and need examination as stated in literature (Wang et al., 2013a). Another obvious research gap identified during the extensive literature review (Chapter 2 of the thesis) and is attempted to be investigated in this thesis, is the understanding of the influence of high resolution traffic and weather risk factors on Powered-Two-Wheelers. Since PTWs are examined both in urban roads and in urban motorways, this is an important step towards the understanding of PTW safety.

A concluding remark is that, the extensive literature review showed that in terms of methodology, some advanced analyses such as time-series were not considered so far for such research aims. On the other hand, other combined approaches used in this thesis, have been applied in general (e.g. Support Vector Machines) or in freeways (Random Forests), but their performance has not been tested in data from urban roads. Lastly, although much research has been carried out recently in order to investigate and predict accident occurrence, urban roads have not been considered and alternative methodologies have not been considered. It is worth mentioning, that this thesis applies a series of rare-events logit models to estimate the influence of traffic parameters on accident occurrence; an approach used maybe for the first time in transportation data analysis. A detailed description is provided in Chapter 8 of the thesis.

Chapter 4 Data collection and processing

4.1 Introduction

Empirical data have been collected for the period 2006-2011 to investigate the relationship between traffic, weather and other characteristics and road accidents. The roads chosen are the Kifisias and Mesogeion avenues in Athens, Greece, mainly due to the fact that they had very similar characteristics. Secondary, Attica Tollway ("Attiki Odos") was also chosen to be investigated separately.

Kifisias avenue has a total length is about 20 km, beginning 4 km northeast of downtown Athens and ending by the municipal boundary of Nea Erythraia north of Kifisia. The total amount of lanes is three, up to Kifisia, then two through Kifissia, before it turns to a one lane (per direction) road for the rest of its length. The avenue begins at the intersection of Alexandras and Mesogeion Avenues. The avenue has a bus lane for a significant section of its length, close to its start.

Figure 4-1 illustrates the Kifisias Avenue. The blue arrows indicate the beginning and the end of the study area. It is noted that Kifisias avenue was not considered on its entirety because only the section between the two blue arrows is covered by loop detectors (please see section 4.2.2).

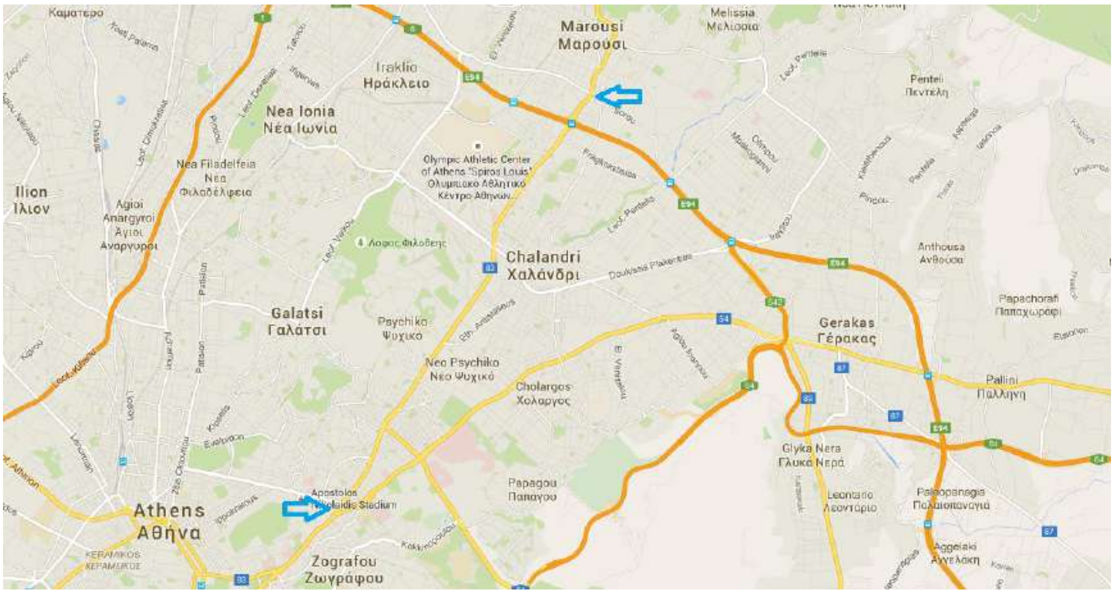


Figure 4-1: The Kifisias avenue.

Mesogeion avenue is also a main road in Athens and its eastern suburbs. The total length is approximately 8 km. Mesogeion avenue also intersects with Michalakopoulou Street, Katechaki avenue and Perikleous avenue. Figure 2 depicts the Mesogeion Avenue, where the arrows show the beginning and the end of the study area.

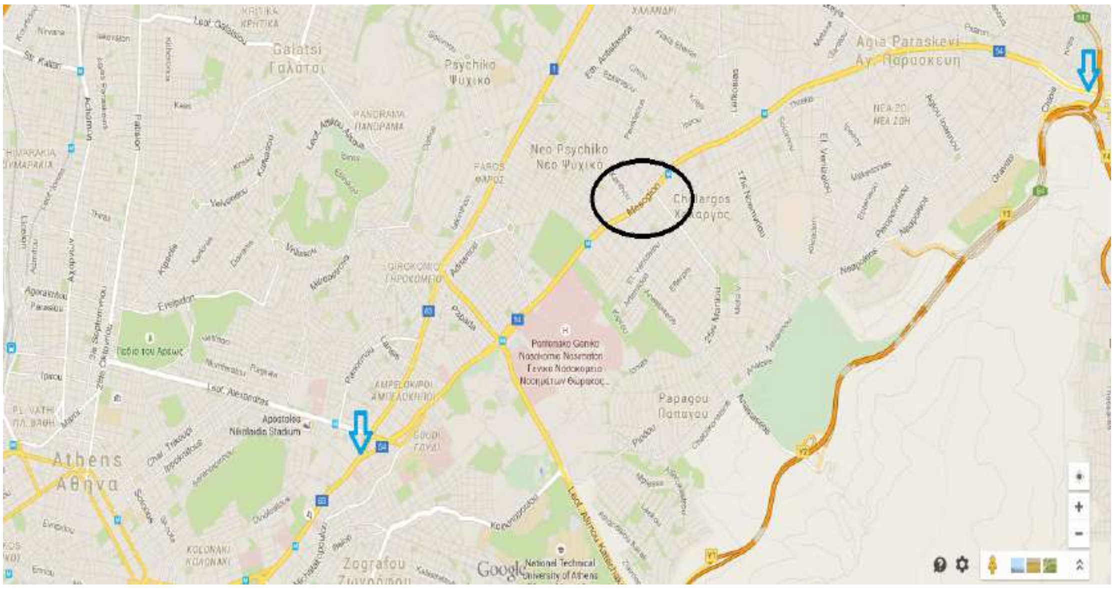


Figure 4-2: The Mesogeion avenue.

Attica Tollway is a modern urban motorway extending along 65.2 km and one of the so-called large projects, most of them satisfying the requirements of Trans European Transport Networks (Papaioannou and Peleka, 2006). Entry to the freeway is through 39 toll plazas with 195 toll lanes. It constitutes the ring road of the greater metropolitan area of Athens and the backbone of the road network of the whole Attica Prefecture. It is essentially a closed toll motorway, within a metropolitan capital, where the problem of traffic congestion is acute. Being a closed motorway, it has controlled access points and consists of two sections, which are perpendicular to one another: The Elefsina - Stavros - Spata A/P motorway (ESSM), extending along approximately 52 km, and the Imittos Western Peripheral Motorway (IWPM), extending along approximately 13 km. Attica Tollway also connects Athens with the International Airport “Eleutherios Venizelos”. Figure 4-3 shows the blue line that indicates the Attika Tollway.

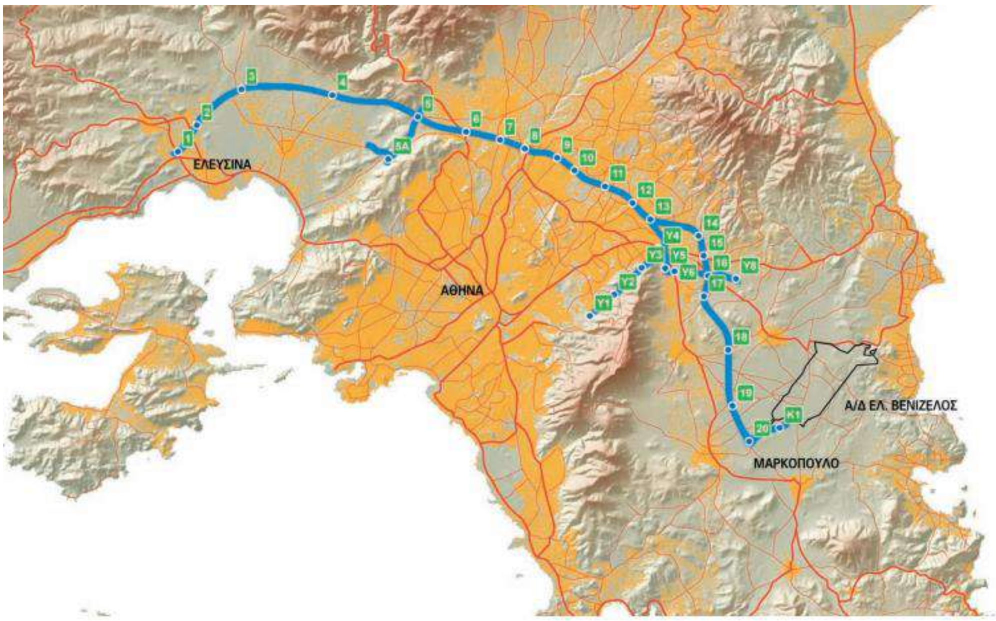


Figure 4-3: Map of Attika Tollway.

4.2 Data in urban roads

4.2.1 Accident data

The required accident data for Kifisias and Mesogeion avenues were collected from the Greek accident database SANTRA, which is provided by the Department of Transportation Planning and Engineering of the National Technical University of Athens. A 6-year period was considered for the analyses of the present thesis, namely 2006-2011.

SANTRA consists of all accident data in Greece, which are filled in high detail by the Police immediately after the occurrence of an accident. Afterwards, data are codified by the Hellenic Statistical Authority (ELSTAT) were each variable gets specific values. This method for national data collection and classification guarantees data quality and also data availability to researchers. SANTRA is a very useful tool for road safety research in Greece, however specific attention is required because such data are confidential and must be utilized only for a scientific and research purpose.

In order to extract accident data the following procedure must be followed: Firstly, the user has to select the version of the SANTRA software use and then to select one of the following objects for data extraction (accident, vehicle or person). Then the user has to specify the place of interest (e.g. a national road) and the desired time period. Lastly, other parameters of interest can be extracted in order to build the required dataset in Microsoft Access based on the queries. The extracted dataset can be easily be transferred in an Excel file. Concerning the urban roads of Kifisias and Mesogeion, the object of interest was the “accident”. Consequently, the severity of the accident (classified as the most severely injured person involved as defined in Chapter 1) is investigated for these two urban roads.

Table 4.1 that follows illustrates the variables extracted from SANTRA as well as their possible values (before the coding took place). In order to have a sufficient and representative sample size 6 years were considered (2006-2011). Lastly, another query in SANTRA had to be performed (the object of interest this time was the “person”), in order to identify if a PTW was involved in the accident.

Variable	Potential values
Year	Factor
Month	Factor
Day	Factor
Time	Numeric
Street number	Numeric
Accident Severity	Fatal
	Severe
	Slight
Median	Yes
	No
Illumination	Day
	Night
	Dusk
Accident Type (collision type)	Off road/Fixed object/Other
	Head-on
	Rear-end
	Side
	Sideswipe
Number of lanes per direction	Numeric
Pavement condition	Good
	Wet
Weather	Good
	Adverse
Road curvature	Straight line
	Curve
Intersection	Yes
	No
Direction	To the centre
	From the centre
Traffic control	Traffic lights
	Other
Number of killed persons involved	Integer
Number of severely injured persons involved	Integer
Number of slightly injured persons involved	Integer

Table 4-1: Variables extracted from SANTRA database regarding Kifisias and Mesogeion avenues.

4.2.2 Traffic data

Traffic data were extracted from the Traffic Management Centre (TMC) of Athens, which operates on a daily basis from July 2004 covering various major arterials in the city of Athens. Traffic Management Centre of Athens consists of 550 loop detectors, 217 cameras, 24 variable message signs and controls 850 junctions in Athens area. It is noted that the Traffic Management Centre does not keep picture files and does not report driving violations. As a consequence, traffic data are confidential and only a small sample can be published. Figure 4-4 shows the interior of TMC while Figure 4-5 shows a simple loop detector.

Figure 4-4: Traffic Management Centre in Athens.

Figure 4-5: Simple loop detector.

Initially, TMC collects traffic occupancy (measured in %) by calculating the percentage of time that a detector is occupied by a vehicle, as well as the traffic flow by simply counting the vehicles passing through a control point. Furthermore, the measurement quality (Quality), classified as High or Low is provided. Traffic flow is measured every 90sec but is measured in vehicles per hour. Moreover, mean-time speed (measured in km/h) is measured either by using image processing or indirectly via loop detectors. When loop detectors are used, speed is calculated through the next equation:

$$v = [(m+l)/10] * [(q*3600/t)/k] \quad \text{(Eq. 4-1)}$$

where,

v: is the mean-time speed (in km/h), m: the length of detector (m), l: average vehicle length (in m), q: current traffic flow (in vehicles), t: time (sec), k: current occupancy (%).

In order to acquire the complete traffic dataset a close collaboration with the TMC was needed. After the initial accident data extraction that was described in section 4.2.1, it was crucial to match the time and location of the accidents of Kifisias and Mesogeion avenues with the respective loop detectors. It was a long process where each accident was assigned to the closest upstream and downstream detectors and then the recorded traffic data had to be extracted from hard disks. It is noted that traffic data for bus lanes are also provided by the TMC, however, bus lanes are not considered in the thesis. It is also worth mentioning, that there were no loop detectors after Agiou Konstantinou intersection in Marousi, and consequently Kifisias avenue was not investigated in its entirety as was shown previously in Figure 4-1.

Having known the time and location for each accident, a 24-hour time series of traffic flow, occupancy and speed (in 5-minutes intervals ending at the time of the accident) were initially extracted. For example, if an accident occurred in Kifisias Avenue on Wednesday 12 August 2009 at 13:00 and is recorded from the loop detector MS259, traffic data from Tuesday 13:00 to Wednesday 13:00 are extracted from loop detector MS259 measured in 5-min intervals. However, a difficulty arose when an accident occurred before 03:00h. Then another excel file from the respective loop detector for the previous day had to be extracted. Section 4.2.4 provides a discussion on the data issues and the actions that had to be taken to overcome these data problems.

It is very important to note, that since accident probability is one of the main aims of the thesis, in order to investigate accident probability random non-accident cases had to be extracted too. This methodology has been widely used in literature (Abdel-Aty and Pande, 2005; Abdel-Aty et al., 2007; Ahmed and Abdel-Aty, 2012; Yu and Abdel-Aty, 2013a). Non-accident cases in this thesis were extracted according to the following assumption: for each accident, two non-accident cases were considered and traffic data for the same location for one week before and one week after the accident occurrence were also extracted. For example, an accident occurred in Kifisias Avenue on Wednesday 12 August 2009 at 13:00 and is recorded from the loop detector MS259, 24 hours of traffic data have been collected covering the period from Tuesday 4 August 2009 13:00h to Wednesday 19 August 2009 for the period 13:00h. It is obvious that for non-accident cases only traffic and weather information could be extracted.

Table 4-2 shows an example of data extraction. An accident occurred in Kifisias Avenue on February 1st at 21:00 and the loop detector MS315 is assigned as the closest downstream detector.

Time	Occupancy [%]	Counting [Veh/h]	Velocity [km/h]	Quality
1/2/06 18:00	14.75	3,386.25	46.75	High
1/2/06 18:05	16.33	3,625.00	45.00	High
1/2/06 18:10	16.33	3,584.33	44.67	High
1/2/06 18:15	16.75	3,620.75	44.25	High
1/2/06 18:20	16.00	3,618.00	45.00	High
1/2/06 18:25	15.67	3,635.00	46.00	High
1/2/06 18:30	15.50	3,750.75	47.50	High
1/2/06 18:35	17.67	3,918.00	43.00	High
1/2/06 18:40	22.33	4,007.00	37.33	High
1/2/06 18:45	28.00	3,826.75	30.25	High
1/2/06 18:50	31.33	3,556.67	26.00	High
1/2/06 18:55	27.67	3,439.00	28.33	High
1/2/06 19:00	18.75	3,330.75	40.75	High
1/2/06 19:05	16.33	3,379.67	45.67	High
1/2/06 19:10	14.67	3,514.67	47.67	High
1/2/06 19:15	13.75	3,535.00	49.00	High
1/2/06 19:20	14.67	3,597.33	47.67	High
1/2/06 19:25	15.00	3,574.33	46.67	High
1/2/06 19:30	14.75	3,535.75	48.00	High
1/2/06 19:35	15.67	3,772.00	48.00	High
1/2/06 19:40	15.33	3,809.00	46.67	High
1/2/06 19:45	15.25	3,800.00	48.00	High
1/2/06 19:50	14.67	3,674.67	49.67	High
1/2/06 19:55	14.33	3,529.67	49.33	High
1/2/06 20:00	15.25	3,595.50	48.25	High
1/2/06 20:05	15.00	3,659.33	48.67	High
1/2/06 20:10	15.33	3,755.00	48.33	High
1/2/06 20:15	14.50	3,625.25	49.25	High
1/2/06 20:20	13.67	3,475.67	49.33	High
1/2/06 20:25	13.67	3,427.00	49.33	High
1/2/06 20:30	13.50	3,511.50	50.00	High
1/2/06 20:35	13.00	3,392.00	51.00	High
1/2/06 20:40	14.00	3,341.33	50.33	High
1/2/06 20:45	13.00	3,339.75	50.25	High
1/2/06 20:50	11.67	3,262.00	52.67	High
1/2/06 20:55	12.00	3,192.67	52.00	High
1/2/06 21:00	12.00	3,108.50	52.50	High

Table 4-2: Time series of counting (flow), speed and occupancy for the loop detector MS315 for February 1st 2006.

4.2.3 Weather data

The weather data were collected from the website created by the Hydrological Observatory of Athens (HOA). The website address is www.hoa.ntua.gr. The Hydrological information provided from this website is a service of the Hydrological Observatory of Athens, operated by the National Technical University of Athens. It provides access to an open-access single database and users are able to locate hydro-meteorological stations on the map, download historic time series and use advanced applications for a statistical analysis of selected hydrological data. HOA offers various weather charts based on historic time series observed in the Greater Athens Area. More specifically:

a) Station charts: Time series derived from single meteorological stations.
b) Variable charts: Diagrams of single variables based on observations from selected meteorological stations.
c) Surface charts: Surface charts of the Greater Athens Area for selected variables.

NTUA's hydrological monitoring network operates since 2005. It consists of more than 10 stations located in the greater Athens area, measuring environmental parameters of hydro-meteorological interest, such as rainfall, temperature, relative humidity, evaporation, air pressure, solar radiation, sunshine duration, wind direction and velocity. Figure 4-6 illustrates the network of meteorological stations in the Greater Athens area monitored by NTUA.

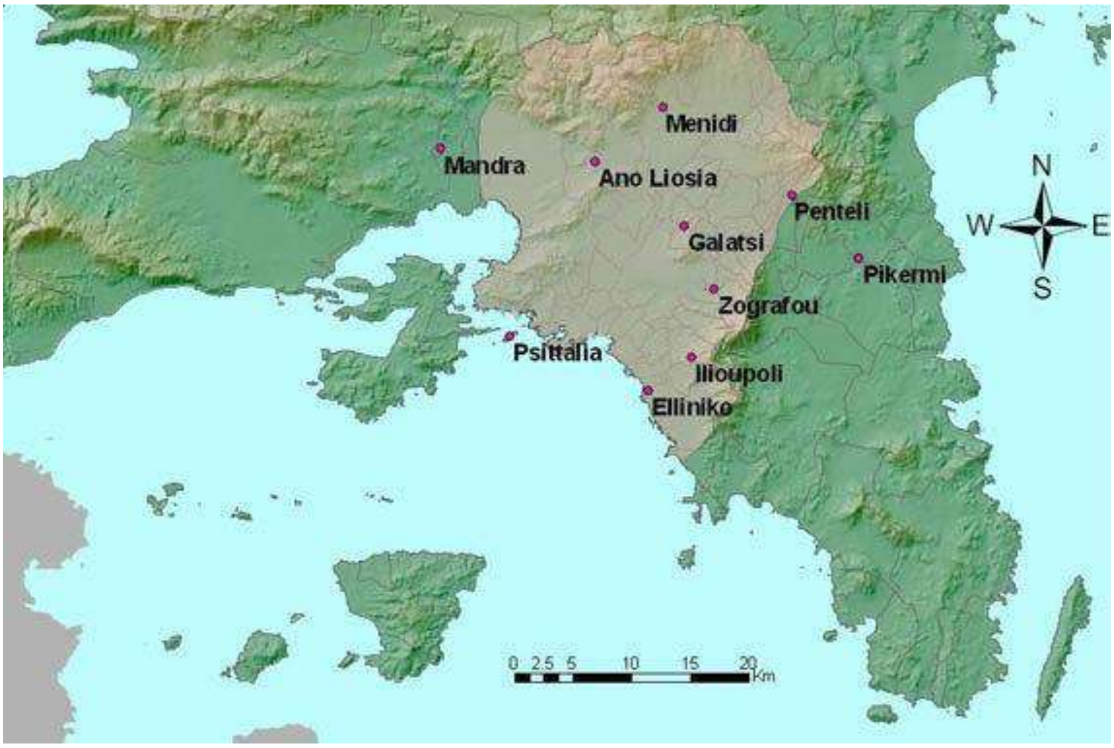


Figure 4-6: Network of meteorological stations in the Greater Athens area controlled by NTUA.

The network has been developed and maintained by the Laboratory of Hydrology and Water Resources, led by Prof. Maria Mimikou and is supported by the Computer Centre of the National Technical University of Athens. Each station is equipped with a data logger, which records, in 10-min intervals, the measurements of all sensors installed in the specific station and, by means of mobile telephony, transmits the data to a central database (HOA, 2014).

In order to acquire weather data for Kifisias avenue, the Galatsi station was chosen as the closest meteorological station. Weather data for Mesogeion Avenue were collected from Zografou station. For those two stations, a time-series in 10-min intervals in a .txt file format was extracted, one for each weather parameter, continuously covering the period 2006-2011. The weather variables considered in this thesis are shown in Table 4-3.

Variable	Unit of measurement
Temperature	^oC
Humidity	%
Rainfall	mm
Wind Speed	m/sec
Wind Direction	degrees (0-360)
Solar Radiation	W/m^2

Table 4-3: Weather parameters considered.

Especially for Wind Direction, zero degrees indicate a north wind, 90 degrees indicate an east wind, 180 degrees a south wind and so on. Solar Radiation which expresses sunlight, is measured in W/m^2 (watts per square metre). In general, practitioners in the business of solar energy may use the unit watt-hour per square metre (Wh/m^2). If this energy is divided by the recording time in hours, it is then a density of power called irradiance, expressed in watts per square metre (W/m^2). Accidents have been assigned to the nearest weather stations as mentioned previously. Based on the reported accident times, the closest weather records 12 hours prior to the accident times were of interest. The next section 4.2.4 describes the steps that were followed so as to acquire and process those specific weather data.

4.2.4 Data processing

This section describes the processing of the accident, the traffic and the weather datasets, the variables' coding and finally presents some descriptive statistics. The data collection that was described in sections 4.2.1, 4.2.2 and 4.2.3 resulted in the creation of three datasets: one accident dataset, one traffic dataset and one weather dataset. The next step was to deeply process the three datasets in order to extract the required parameters and finally create one unified dataset for accident severity and one for accident probability. In order to have an easier data processing, the year, month and day of each observation were re-coded in one column (e.g. 2/12/2007).

Firstly, the accident datasets of Mesogeion and Kifisias were unified in one unique urban accident dataset. For accident severity, each observation in the dataset had to be a record of each accident, the corresponding traffic and weather conditions, and various external factors. For accident probability, for each accident observation of the dataset, there were 2 observations for non-accidents for the same location. It is obvious that for accident probability only traffic and weather parameters were considered. Accidents were the time or the exact location of occurrence was not known were deleted from the dataset.

First of all, the massive traffic data that were extracted from the TMC needed a long processing. Three excel files for each accident were produced and in order to create a whole traffic database several programming tasks had to be undertaken. It is important to note, that in many cases, the time of the accident had to be recalculated in order to be as precise as possible.

The first test was the visual inspection of the traffic time series of traffic variables, so that sudden drops in speeds would reveal the precise time of the accident. When this was not feasible, a time lag of 20 minutes was used, simply meaning that the time of the accident was recalculated by subtracting 20 minutes. This time lag was considered in order to avoid the impact of the crash itself on the traffic variables and also to compensate for any potential 'inaccuracies' in the exact time of the accident. Similar time lags have been applied in other real-time data analyses. For example, Christoforou et al. (2010), used a 12-minute time lag. Quddus et al. (2009), used a similar mesoscopic approach for M25 motorway outside London and used a 30-min time lag. Since the approach of the thesis is more mesoscopic this time lag is considered adequate. More specifically, the raw 5-min traffic data were aggregated into 1-hour intervals (1h), in order to obtain maxima, averages, standard deviations, medians and coefficients of variation for 1h, 2h and 3h prior to the recalculated time of the accident. Such intervals may be too large to capture short-term variations; however, data averaged on shorter intervals were not available. Several authors (Abdel-Aty and Pande, 2005; Oh et al., 2000; Pande and Abdel-Aty, 2006) have used 5-minute intervals to perform similar analyses, however the thesis has a focus on a more mesoscopic approach which has never been followed using high resolution traffic and weather data. It is worth mentioning, that before aggregating, traffic flow was divided by the number of lanes in order to be consistent across Kifisias and Mesogeion road segments.

The traffic data quality was thoroughly checked. In cases of loop detector malfunction or in cases of low measurement quality according to the TMC records, the next closest detector was used to extract data. There were rare cases when loop detectors suffered from problems that might have resulted in unreasonable values for speed, volume, and occupancy. Such unrealistic values (e.g. occupancy>100%, speed>200 km/h or speed>0 along with flow=0) were discarded from the database. Accidents with traffic data unavailability were also discarded. The same procedure has been followed when a loop detector was installed in the place of interest after the time of an accident (e.g.an accident occurred in 8/2/2006 but the closest relevant loop detector was installed there later in 2007).

Regarding the weather data, the weather records for each meteorological station covered the whole period from 2006-2011. For that reason, each accident case had to be assigned to the closest meteorological station and then the relevant weather data had to be extracted. Using advanced programming in Excel, the 10-min raw data were aggregated in order to obtain maxima, averages and standard deviations, for 1-hour, 2-hours, 6-hours and 12-hours prior to the recalculated time of the accident occurrence. Regarding rainfall, the sum of rainfall has also been calculated for 1h, 2h, 6h and 12h prior to the time of the accident.

The non-accident cases (used when modelling accident probability), had been treated by following the exact same procedure as for traffic data. For example, an accident occurred in Kifisias Avenue on Wednesday 12 August 2009 at 13:00, 12 hours of weather data have been extracted from Galatsi Station, for Wednesday 5 August 2009 and Wednesday 19 August 2009 for the period 01:00-13:00. In rare cases, when missing weather data were observed the accident case was removed. However, when only a few 10-min time intervals were missing, the aggregations were performed by considered the rest of 10-min time intervals. The final urban dataset included 527 accidents and 1054 non-accident cases (1581 total cases). PTWs were involved in 326 of those accidents (61.9% of the accidents). The next four tables (4-4, 4-5, 4-6 and 4-7) illustrate the list of variables included along with the description, the coding and descriptive statistics.

Variable	Type	Description	Descriptive Statistics	
Accident Severity	Dummy	Fatal/Severe=1	57	10.8%
		Slight=0	470	89.2%
PTW accident involvement	Dummy	Yes=1	326	61.9%
		No=0	201	38.1%
Median	Dummy	Yes=1	458	86.9%
		No=0	69	13.1%
Illumination	Dummy	Day=1	332	63.0%
		Night/Dusk=0	195	37.0%
Accident Type (collision type)	Dummy	Off road/Fixed object/Other	226	42.9%
		Head-on	43	8.2%
		Rear-end	91	17.3%
		Side	67	12.7%
		Sideswipe	100	19.0%
Road curvature	Dummy	Straight line=1	497	94.3%
		Curve=0	30	5.7%
Intersection	Dummy	Yes=1	174	33.0%
		No=0	353	67.0%
Direction	Dummy	To the centre=1	213	40.4%
		From the centre=0	314	59.6%
Traffic control	Dummy	Traffic lights=1	169	32.1%
		No/Other=0	358	67.9%
Weather	Dummy	Good=1	489	92.8%
		Adverse=0	38	7.2%
Pavement condition	Dummy	Good=1	486	92.2%
		Wet=0	41	7.8%

Table 4-4: List of accident related variables for urban roads.

The next table (4-5) provides an illustration of the traffic variables downstream and upstream of the accident location and their descriptive statistics as well (minimum, maximum, mean, median). It is noted that volume (flow) was divided per lane, while speed and occupancy were across lanes. Table 4-6, shows the description of the weather variables and some descriptive statistics as well (minimum, maximum, mean, median).

Variable	Type	Description	Descriptive Statistics			
			Min	Median	Mean	Max
Q_avg_1h_down	Continuous	1h average flow per lane downstream	25.84	871.61	826.13	1795.47
Q_stdev_1h_down	Continuous	1h st.deviation of flow per lane downstream	4.87	59.08	76.10	477.08
Q_median_1h_down	Continuous	1h median of flow per lane downstream	0.00	877.90	830.10	1786.50
Q_cv_1h_down	Continuous	1h coefficient of variation of flow per lane downstream	0.02	0.08	0.11	0.64
V_avg_1h_down	Continuous	1h average speed downstream	5.69	48.85	48.32	119.39
V_stdev_1h_down	Continuous	1h st.deviation of speed downstream	0.00	2.78	4.21	26.76
V_cv_1h_down	Continuous	1h coefficient of variation of speed downstream	0.00	0.06	0.12	1.08
Occ_avg_1h_down	Continuous	1h average occupancy downstream	0.02	12.13	14.46	51.00
Occ_stdev_1h_down	Continuous	1h st.deviation of occupancy downstream	0.00	1.65	3.03	21.70
Occ_cv_1h_down	Continuous	1h coefficient of variation of occupancy downstream	0.00	0.16	0.22	3.61
Q_avg_2h_down	Continuous	2h average flow per lane downstream	32.29	869.86	827.22	2321.42
Q_stdev_2h_down	Continuous	2h st.deviation of flow per lane downstream	3.26	55.60	69.19	482.43
Q_median_2h_down	Continuous	2h median of flow per lane downstream	32.44	875.00	826.86	2293.75
Q_cv_2h_down	Continuous	2h coefficient of variation of flow per lane downstream	0.01	0.07	0.10	0.85
V_avg_2h_down	Continuous	2h average speed downstream	4.65	49.69	48.96	130.91
V_stdev_2h_down	Continuous	2h st.deviation of speed downstream	0.00	2.49	3.91	30.35
V_cv_2h_down	Continuous	2h coefficient of variation of speed downstream	0.00	0.06	0.11	0.78
Occ_avg_2h_down	Continuous	2h average occupancy downstream	0.28	11.74	14.05	52.50
Occ_stdev_2h_down	Continuous	2h st.deviation of occupancy downstream	0.00	1.47	2.68	21.97
Occ_cv_2h_down	Continuous	2h coefficient of variation of occupancy downstream	0.00	0.15	0.21	1.34
Q_avg_3h_down	Continuous	3h average flow per lane downstream	31.62	857.26	806.89	2588.40
Q_stdev_3h_down	Continuous	3h st.deviation of flow per lane downstream	5.25	57.22	70.54	313.29
Q_median_3h_down	Continuous	3h median of flow per lane downstream	16.84	862.04	808.26	2565.33
Q_cv_3h_down	Continuous	3h coefficient of variation of flow per lane downstream	0.01	0.08	0.11	0.88
V_avg_3h_down	Continuous	3h average speed downstream	5.42	49.75	49.17	142.22
V_stdev_3h_down	Continuous	3h st.deviation of speed downstream	0.00	2.45	3.73	23.05
V_cv_3h_down	Continuous	3h coefficient of variation of speed downstream	0.00	0.05	0.11	0.78
Occ_avg_3h_down	Continuous	3h average occupancy downstream	0.00	11.84	13.81	57.37
Occ_stdev_3h_down	Continuous	3h st.deviation of occupancy downstream	0.00	1.43	2.77	20.65
Occ_cv_3h_down	Continuous	3h coefficient of variation of occupancy downstream	0.00	0.15	0.22	1.61
Q_avg_1h_up	Continuous	1h average flow per lane upstream	52.82	835.64	793.66	1848.91
Q_stdev_1h_up	Continuous	1h st.deviation of flow per lane upstream	10.58	81.81	270.17	1165.50
Q_median_1h_up	Continuous	1h median of flow per lane upstream	13.50	681.63	602.28	1889.13
Q_cv_1h_up	Continuous	1h coefficient of variation of flow per lane upstream	0.02	0.08	0.11	0.58
V_avg_1h_up	Continuous	1h average speed upstream	4.50	45.90	45.52	104.60
V_stdev_1h_up	Continuous	1h st.deviation of speed upstream	0.00	3.06	5.13	28.08
V_cv_1h_up	Continuous	1h coefficient of variation of speed upstream	0.00	0.08	0.15	0.89
Occ_avg_1h_up	Continuous	1h average occupancy upstream	0.15	12.94	15.74	57.52
Occ_stdev_1h_up	Continuous	1h st.deviation of occupancy upstream	0.00	2.06	3.96	30.84
Occ_cv_1h_up	Continuous	1h coefficient of variation of occupancy upstream	0.00	0.17	0.25	1.76
Q_avg_2h_up	Continuous	2h average flow per lane upstream	50.35	836.47	794.83	1772.58
Q_stdev_2h_up	Continuous	2h st.deviation of flow per lane upstream	7.76	56.26	68.55	652.99
Q_median_2h_up	Continuous	2h median of flow per lane upstream	0.00	837.30	794.20	1834.50
Q_cv_2h_up	Continuous	2h coefficient of variation of flow per lane upstream	0.02	0.07	0.10	1.04
V_avg_2h_up	Continuous	2h average speed upstream	5.80	48.14	47.22	116.67
V_stdev_2h_up	Continuous	2h st.deviation of speed upstream	0.00	2.63	4.04	31.49
V_cv_2h_up	Continuous	2h coefficient of variation of speed upstream	0.00	0.06	0.12	0.79
Occ_avg_2h_up	Continuous	2h average occupancy upstream	0.00	11.77	14.62	57.62
Occ_stdev_2h_up	Continuous	2h st.deviation of occupancy upstream	0.00	1.63	2.93	20.35
Occ_cv_2h_up	Continuous	2h coefficient of variation of occupancy upstream	0.00	0.16	0.22	3.61
Q_avg_3h_up	Continuous	3h average flow per lane upstream	28.00	814.50	771.30	1873.90
Q_stdev_3h_up	Continuous	3h st.deviation of flow per lane upstream	6.36	55.09	67.09	381.20
Q_median_3h_up	Continuous	3h median of flow per lane upstream	0.00	835.00	815.20	2308.70
Q_cv_3h_up	Continuous	3h coefficient of variation of flow per lane upstream	0.00	0.08	0.11	0.46
V_avg_3h_up	Continuous	3h average speed upstream	5.88	48.42	47.66	104.51
V_stdev_3h_up	Continuous	3h st.deviation of speed upstream	0.00	2.51	3.84	28.96
V_cv_3h_up	Continuous	3h coefficient of variation of speed upstream	0.00	0.05	0.11	0.75
Occ_avg_3h_up	Continuous	3h average occupancy upstream	0.00	11.54	14.23	55.74
Occ_stdev_3h_up	Continuous	3h st.deviation of occupancy upstream	0.00	1.53	2.90	20.81
Occ_cv_3h_up	Continuous	3h coefficient of variation of occupancy upstream	0.00	0.16	0.22	2.47

Table 4-5: Descriptive statistics of traffic related variables upstream and downstream of the accident location.

Variable	Type	Description	Descriptive Statistics			
			Min	Median	Mean	Max
T_1h_max	continuous	1h maximum temperature	-1.70	17.85	18.72	42.17
T_1h_avg	continuous	1h average temperature	-2.14	17.23	18.20	41.99
T_1h_stdev	continuous	1h st.deviation of temperature	0.02	0.30	0.39	2.84
T_2h_max	continuous	2h maximum temperature	-1.70	18.18	19.04	43.13
T_2h_avg	continuous	2h average temperature	-2.18	17.23	18.09	42.40
T_2h_stdev	continuous	2h st.deviation of temperature	0.03	0.51	0.64	3.91
T_6h_max	continuous	6h maximum temperature	-0.77	18.91	19.90	43.13
T_6h_avg	continuous	6h average temperature	-1.93	16.77	17.55	41.41
T_6h_stdev	continuous	6h st.deviation of temperature	0.04	1.26	1.50	5.27
T_12h_max	continuous	12h maximum temperature	3.08	20.21	20.80	43.13
T_12h_avg	continuous	12h average temperature	-0.07	16.20	16.93	35.38
T_12h_stdev	continuous	12h st.deviation of temperature	0.18	2.18	2.33	7.70
Hum_1h_max	continuous	1h maximum humidity	12.49	58.96	58.27	97.50
Hum_1h_avg	continuous	1h average humidity	12.20	55.37	55.84	97.20
Hum_1h_stdev	continuous	1h st.deviation of humidity	0.06	1.37	1.82	13.12
Hum_2h_max	continuous	2h maximum humidity	12.87	61.73	60.68	98.50
Hum_2h_avg	continuous	2h average humidity	11.62	55.58	56.17	97.55
Hum_2h_stdev	continuous	2h st.deviation of humidity	0.14	2.25	2.92	20.33
Hum_6h_max	continuous	6h maximum humidity	13.37	70.50	68.46	98.50
Hum_6h_avg	continuous	6h average humidity	11.16	58.20	57.98	97.62
Hum_6h_stdev	continuous	6h st.deviation of humidity	0.51	5.21	6.26	22.26
Hum_12h_max	continuous	12h maximum humidity	25.03	78.05	74.86	99.60
Hum_12h_avg	continuous	12h average humidity	13.55	61.49	60.19	97.29
Hum_12h_stdev	continuous	12h st.deviation of humidity	0.57	8.61	9.40	27.58
Rain_1h_sum	continuous	1h sum of rainfall	0.00	0.00	0.05	6.60
Rain_1h_st.dev	continuous	1h st.deviation of rainfall	0.00	0.00	0.01	1.20
Rain_2h_sum	continuous	2h sum of rainfall	0.00	0.00	0.15	23.60
Rain_2h_st.dev	continuous	2h sum of rainfall	0.00	0.00	0.02	1.84
Rain_6h_sum	continuous	6h sum of rainfall	0.00	0.00	0.33	35.60
Rain_6h_st.dev	continuous	6h st.deviation of rainfall	0.00	0.00	0.02	1.90
Rain_12h_sum	continuous	12h sum of rainfall	0.00	0.00	0.53	52.80
Rain_12h_st.dev	continuous	12h st.deviation of rainfall	0.00	0.00	0.02	1.67
W.Sp_1h_max	continuous	1h maximum wind speed	0.00	2.32	2.72	9.67
W.Sp_1h_avg	continuous	1h average wind speed	0.00	1.75	2.16	7.95
W.Sp_1h_stdev	continuous	1h st.deviation of wind speed	0.00	0.34	0.39	1.39
W.Sp_2h_max	continuous	2h maximum wind speed	0.00	2.49	2.93	9.67
W.Sp_2h_avg	continuous	2h average wind speed	0.00	1.67	2.14	8.06
W.Sp_2h_stdev	continuous	2h st.deviation of wind speed	0.00	0.40	0.46	1.63
W.Sp_6h_max	continuous	6h maximum wind speed	0.01	3.03	3.33	9.67
W.Sp_6h_avg	continuous	6h average wind speed	0.00	1.68	2.02	7.57
W.Sp_6h_stdev	continuous	6h st.deviation of wind speed	0.00	0.58	0.64	2.02
W.Sp_12h_max	continuous	12h maximum wind speed	0.41	3.38	3.71	9.67
W.Sp_12h_avg	continuous	12h average wind speed	0.02	1.47	1.85	6.68
W.Sp_12h_stdev	continuous	12h st.deviation of wind speed	0.06	0.75	0.84	2.43
W.Dir_1h_avg	continuous	1h average wind direction	0.00	141.00	128.36	351.67
W.Dir_2h_avg	continuous	2h average wind direction	0.00	138.77	128.31	328.52
W.Dir_6h_avg	continuous	6h average wind direction	0.43	129.80	125.19	303.27
W.Dir_12h_avg	continuous	12h average wind direction	9.89	128.74	122.50	285.90
Sol_1h_max	continuous	1h maximum solar radiation	0.00	264.75	354.07	1100.00
Sol_1h_avg	continuous	1h average solar radiation	0.00	164.38	284.92	1007.29
Sol_2h_max	continuous	2h maximum solar radiation	0.00	362.50	396.92	1146.00
Sol_2h_avg	continuous	2h average solar radiation	0.00	177.04	283.00	1023.15
Sol_6h_max	continuous	6h maximum solar radiation	0.00	555.60	517.50	1168.00
Sol_6h_avg	continuous	6h average solar radiation	0.00	176.25	255.60	896.69
Sol_12h_max	continuous	12h maximum solar radiation	0.00	654.90	621.00	1168.00
Sol_12h_avg	continuous	12h average solar radiation	0.00	158.01	203.30	678.19

Table 4-6: Descriptive statistics of weather related variables in urban roads.

4.3 Data in urban motorways

4.3.1 Accident data

The required accident data for Attica Tollway were extracted from the Greek accident database SANTRA provided by the Department of Transportation Planning and Engineering of the National Technical University of Athens, by following the exact same procedure that was followed in section 4.2.1 for Kifisias and Mesogeion avenues. Similarly, a 6-year period was considered as well (years 2006-2011). The unit of analysis in Attica Tollway was any vehicle occupant involved in an accident (rider, driver or passenger) resulting in at least one person being slightly injured. Consequently, a single accident would correspond to various observations that are equal to the number of all injured persons that were involved in the accident. It is noted than only basic freeway segments (BFS) were considered and not ramp areas. Table 4-7 that follows illustrates the variables of interest that were extracted from SANTRA database as well as their possible values (before coding took place).

Variable	Potential values
Year	Factor
Month	Factor
Day	Factor
Time	Numeric
Vehicle Occupant Severity	Fatal
	Severe
	Slight
Vehicle type	Car
	PTW
	Other
Illumination	Day
	Night
	Dusk
Accident Type (collision type)	Off road/Fixed object/Other
	Head-on
	Rear-end
	Side
	Sideswipe
Engine size	Numeric
Age of occupant	Numeric
Gender	male
	female
Nationality of occupant	Greek
	Other
Number of lanes per direction	Numeric
Road curvature	Straight line
	Curve
Number of killed persons involved	Integer
Number of severely injured persons involved	Integer
Number of slightly injured persons involved	Integer

Table 4-7: Variables extracted from SANTRA database for Attica Tollway.

4.3.2 Traffic parameters

Traffic data for the Attica Tollway were extracted after a close collaboration with the Traffic Management and Motorway Maintenance. It is located in Paiania and operates on a 24-hour basis. The Traffic Management and Motorway Maintenance Department contains one (1) 80 inches monitor, forty (40) 21 inches monitors and 7 workstations with access to motorway systems.The main purposes of the Traffic Management and Motorway Maintenance Department are, among others: Traffic control and monitoring; Management of emergency incidents and planned activities; Inspection, maintenance and repairs, as may be required in order to preserve the good condition of the motorway; Maintenance of operation and maintenance vehicles and facilities; Routine works, such as cleaning the motorway.

Inductive loops (sensors) are placed every 500 meters inside the asphalt pavement of the open sections of the motorway and every 60 meters inside tunnels, providing information regarding the volume, speed and density of traffic. These sensors enable the prompt detection of any problems causing disruption to the smooth flow of traffic and automatically activates intervention procedures to deal with the problem. Aside from the extensive loop detector system, traffic monitoring and management are conducted through a series of Closed Circuit TV cameras (CCTV), Variable Message Signs, Variable Speed Limit Signs and Over Height Vehicle Detectors.

The data that were collected for Attica Tollway consisted of the following traffic parameters: flow, speed, occupancy and truck proportion. The raw data that stem from each single loop detector are calculated directly every 20 sec. Then data are aggregated in 5 minutes. Concerning flow, the 5-min data are calculated simply by adding the values of the 20 sec raw data. Regarding speed and occupancy the average values are obtained. After the initial accident data extraction that was described in section 4.3.1, it was then aimed to match the time and location of the accidents of Attica Tollway with the respective loop detectors. Each accident was assigned to the closest upstream detector. Aside of the traffic parameters of interest, the number of lanes are provided and also the quality of measurements expressed in % observed. In cases of low percentages of observed values (lower than 90%) or in cases of loop detector malfunction the next closest upstream detector was used to extract data.

For each accident, a 3-hour time series of traffic flow, occupancy, speed and truck proportion (in 5-minutes intervals ending at 5 minutes before the time of the accident) were extracted. This approach has been followed because the analysis of Attica Tollway data was more microscopic than that of urban roads data analysis (see section 4.3.4). This time lag was used to avoid the impact of the crash itself on the traffic variables and also to compensate for any ‘inaccuracy’ in the exact time of the accident. For example, if an accident occurred at 21:00h, the traffic data considered were obtained from the 17:55-20:55h period. Similar techniques have been applied in other real-time data analyses (Abdel-Aty et al., 2007; Christoforou et al., 2010).

Table 4-8 shows an example of data extraction involving the time series of flow and speed. An accident occurred in Attica Tollway on December 1st 2011 at 18:00 and traffic data are extracted by the closest upstream detector (the date is in mm/dd/yyyy format). Consequently, the 3-hour time series of flow, speed, occupancy and truck proportion ending at 5-min prior to the time of the accident (14:55-17:55) are considered. Each excel file could contain information for only two traffic parameters (e.g. speed and flow), and therefore two separate excel files for each accident had to be created.

5 Minutes	Flow (Veh/5 Minutes)	Speed (kph)	# Lane Points	% Observed
12/1/2011 14:55	81	87.2	2	100
12/1/2011 15:00	70	89.8	2	100
12/1/2011 15:05	78	90	2	100
12/1/2011 15:10	67	89.7	2	100
12/1/2011 15:15	78	91.5	2	100
12/1/2011 15:20	77	91.9	2	100
12/1/2011 15:25	76	92.2	2	100
12/1/2011 15:30	83	89.2	2	100
12/1/2011 15:35	63	89.9	2	100
12/1/2011 15:40	87	89.8	2	100
12/1/2011 15:45	92	90	2	100
12/1/2011 15:50	79	91.2	2	100
12/1/2011 15:55	78	92.2	2	100
12/1/2011 16:00	81	85.8	2	100
12/1/2011 16:05	77	88.9	2	100
12/1/2011 16:10	80	90.5	2	100
12/1/2011 16:15	94	87.4	2	100
12/1/2011 16:20	69	89.5	2	100
12/1/2011 16:25	52	89.9	2	100
12/1/2011 16:30	87	87.4	2	100
12/1/2011 16:35	70	88.5	2	100
12/1/2011 16:40	77	90	2	100
12/1/2011 16:45	76	90.4	2	100
12/1/2011 16:50	80	89.7	2	100
12/1/2011 16:55	71	90	2	100
12/1/2011 17:00	85	90.5	2	100
12/1/2011 17:05	84	89.4	2	100
12/1/2011 17:10	78	87.4	2	100
12/1/2011 17:15	77	89.9	2	100
12/1/2011 17:20	77	90.5	2	100
12/1/2011 17:25	59	90.1	2	100
12/1/2011 17:30	78	89.9	2	100
12/1/2011 17:35	80	90.6	2	100
12/1/2011 17:40	71	80.9	2	100
12/1/2011 17:45	61	68.6	2	100
12/1/2011 17:50	54	64.5	2	100
12/1/2011 17:55	63	68.5	2	100
12/1/2011 18:00	59	70.3	2	100
12/1/2011 18:05	69	69.3	2	100
12/1/2011 18:10	66	68.1	2	100
12/1/2011 18:15	78	78.4	2	100
12/1/2011 18:20	54	92.2	2	100
12/1/2011 18:25	53	86.8	2	100
12/1/2011 18:30	66	89.9	2	100
12/1/2011 18:35	79	89.5	2	100
12/1/2011 18:40	59	89.8	2	100
12/1/2011 18:45	55	87.8	2	100
12/1/2011 18:50	59	90	2	100
12/1/2011 18:55	63	89	2	100
12/1/2011 19:00	53	86	2	100
12/1/2011 19:05	59	87.7	2	100
12/1/2011 19:10	56	91.5	2	100
12/1/2011 19:15	41	93.4	2	100
12/1/2011 19:20	52	89	2	100
12/1/2011 19:25	38	89.9	2	100
12/1/2011 19:30	59	91.9	2	100
12/1/2011 19:35	48	90.3	2	100
12/1/2011 19:40	61	86.9	2	100
12/1/2011 19:45	45	90.1	2	100
12/1/2011 19:50	48	91.8	2	100
12/1/2011 19:55	46	88.2	2	100

Table 4-8: Time series of flow and speed for December 1st 2011.

Accident probability modelling for urban motorway data were analysed following a different approach than that in urban roads (a detailed modelling description is provided in chapter 8). In that context, in order to apply the rare-events logit model, there was no need to consider non-accident cases, and hence the complete traffic time series in 1-hour intervals from two random loop detectors were collected for the whole period 2008-2011 in order to be analysed.

4.3.3 Weather parameters

Collection of weather data in the urban motorway of Attica Tollway were collected by the same source as earlier for the weather data for urban roads (www.hoa.ntua.gr). According to the location of occurrence, each accident was assigned to the nearest meteorological station. The following meteorological stations were considered: Ano Liosia, Galatsi, Mandra, Penteli, Pikermi and Zografou. For each of those stations, a time-series in 10-min intervals in a .txt file format was extracted, one for each weather parameter. The weather variables considered for Attica Tollway are the same as in urban roads: Temperature (in ^{o}C), Humidity (in %), Rainfall (in mm), Wind Speed (in m/sec), Wind Direction (in degrees ranging from 0 to 360) and lastly Solar radiation (in W/m^2). Similarly to urban roads analysis, only the 12-hour time series of weather variables prior to the accident were of interest. The next section 4.3.4 describes the steps that were followed so as to acquire and process those specific weather data.

4.3.4 Data processing

As in section 4.2.4, this section also describes the processing of the accident, the traffic and the weather datasets, the variables' coding and finally presents some descriptive statistics. The data collection that was described in sections 4.3.1, 4.3.2 and 4.3.3 resulted in the creation of three datasets: one accident dataset, one traffic dataset and one weather dataset. The next step was to deeply process the three datasets in order to extract the required parameters and finally create one unified dataset for accident severity and one for accident probability. In order to have an easier data processing, the year, month and day of each observation were re-coded in one column (e.g. 2/12/2007). In contrast to urban roads, in motorways the unit of analysis was the occupant, and therefore each row in the dataset corresponds to an injured individual who has been involved in an accident. Accidents were the time or the exact location of occurrence was not known were deleted from the dataset.

The traffic data extracted from the Traffic Management and Motorway Maintenance of Attica Tollway needed a deep processing. Two excel files for each accident were produced (each containing two traffic variables of interest) and in order to create a whole traffic database several programming tasks had to be undertaken. Before aggregating, traffic flow was divided by the number of lanes in order to be consistent across road segments. The raw 5-min traffic data were further aggregated in 15min, 30min and 60min in order to obtain maxima, averages, standard deviations, medians and coefficients of variation prior to the accident. Therefore, the analysis approach for Attica Tollway was more microscopic than in urban roads.

The traffic data quality had to be thoroughly checked. In cases of loop detector malfunction or in cases of low measurement quality according to the TMC records, the next closest detector was used to extract data. There were rare cases when loop detectors suffered from problems that might have resulted in unreasonable values for speed, volume, and occupancy. Such values (e.g. occupancy>100%, speed>200 km/h or speed>0 along with flow=0) were discarded from the database. Accidents with traffic data unavailability were also discarded.

Regarding the weather data, the weather records for each meteorological station covered the whole period from 2006-2011. For that reason, each accident case had to be assigned to the closest meteorological station and then the relevant weather data had to be extracted. For example, an accident occurred in Attica Tollway on December 1st 2011 at 18:00. Consequently, the 12-hour weather time series ending at 5-min prior to the time of the accident (05:55-17:55) are considered. Using advanced programming in Excel, the 10-min raw data were aggregated in order to obtain maxima, averages and standard deviations, for 1-hour, 2-hours, 6-hours and 12-hours prior to the time of the accident occurrence. Regarding rainfall, the total amount of rain has also been calculated for 1h, 2h, 6h and 12h prior to the time of the accident. In extremely rare cases, when missing data were observed the accident case was removed. However, when only a few observations of 10-min time intervals were missing, the aggregations were performed by considered the rest of 10-min time intervals. The descriptive statistics of weather variables are shown on Table 4-12.

The final motorway dataset included 285 accidents (387 injured persons). Tables 4-9 4-11 and 4-12, illustrate the list of variables included along with the description, the coding and descriptive statistics. More specifically, tables 4-11 and 4-12 provide an illustration of the traffic variables upstream of the accident location and also their descriptive statistics (minimum, maximum, mean, median etc.). It is noted that traffic flow was divided per lane, while speed and occupancy were across lanes. Furthermore, flow for Attica Tollway was measured vehicles/5 minutes. Table 4-9 shows that severity is recoded in order to have two levels, namely fatal-severe injury (KSI) and slight injury (SI). Moreover, night and dusk categories of Illumination were unified as well.

It is noticed that accident probability for Attica Tollway required another dataset, because the collection of non-accident cases was entirely different than that of urban roads. In order to model accident probability for Attical Tollway a different approach was followed. As mentioned in section 4.3.2, in order to proceed in the accident probability modeling, the complete traffic time series measured in 1-hour intervals from 2008-2011 were considered. More specifically, traffic and accident data from three random loop detectors were extracted and unified. The final dataset, consists of 17 accident cases (occurred nearby these three loop detectors) as well as 91118 non-accident cases. Traffic variables were measured in 1-h intervals (flow, speed, occupancy and truck proportion). The flow is the total number of vehicles in 1 hour and is measured in vehicles per hour. On the other hand, speed, occupancy and truck proportion are the averaged values of the 5-minutes intervals, which were automatically aggregated in 1-hour intervals.

Accident occurrence was defined as binary variable taking the values of 0 (non-accident) and 1 (accident). Therefore, in each 1-h time interval there is the information if an accident has occurred or not. In order to avoid the post-accident traffic conditions where low mean speeds may prevail due to the accident itself, the traffic time series before accident cases had to be checked. The aim was to identify potential sudden fall in mean speeds which would lead to erroneous estimates of the effect of traffic variables, because of the decrease in average mean speed, due to the effect of the accident itself. In such cases, the accident is assigned to the previous 1-hour time interval. Table 4-10, presents a small sample of the data of one loop detector before uniting the relevant datasets.

Variable	Type	Description	Descriptive statistics	
Severity	Dummy	Killed/Severely injured=0	344	88.89%
		Severe=1	43	11.11%
Vehicle type	Dummy	Car	205	52.97%
		PTW	161	41.60%
		Other	21	5.43%
Illumination	Dummy	Day=1	283	73.13%
		Night/Dusk=0	104	26.87%
Accident Type (collision type)	Dummy	Off road/Fixed object/Other	116	29.97%
		Rear-end	157	40.57%
		Side	43	11.11%
		Sideswipe	71	18.35%
Engine size	Continuous	Numeric in cc	mean=1167.553	st.dev= 1303.694
Age of occupant	Continuous	Numeric in years	mean=37.7804	st.dev=14.648
Gender	Dummy	Male=1	283	73.13%
		Female=0	104	26.87%
Nationality of occupant	Dummy	Greek=1	357	92.25%
		Other=0	30	7.75%
Road curvature	Dummy	Straight line=1	313	80.88%
		Curve=0	74	19.12%
Weather	Dummy	Good=1	367	94.83%
		Adverse=0	20	5.17%

Table 4-9: List of accident related variables for Attica Tollway.

TIME	Accident	Flow	Speed	Occupancy	Truck Prop.
1/1/2008 0:00	0	1414	110.7	0.02	4.2
1/1/2008 1:00	0	2800	102.7	0.04	4.6
1/1/2008 2:00	0	2201	103.8	0.03	5.5
1/1/2008 3:00	0	1443	108.1	0.02	6.5
1/1/2008 4:00	0	1134	109	0.02	4.1
1/1/2008 5:00	0	1015	112.1	0.01	3.8
1/1/2008 6:00	0	950	114	0.01	3.3
1/1/2008 7:00	0	725	114.7	0.01	2.8
1/1/2008 8:00	0	569	115.9	0.01	5.3
1/1/2008 9:00	0	561	115.8	0.01	5
1/1/2008 10:00	0	794	116.9	0.01	2.4
1/1/2008 11:00	0	1316	116.3	0.02	4.8
1/1/2008 12:00	0	2021	114.5	0.03	5.7
1/1/2008 13:00	0	2723	114.6	0.04	4
1/1/2008 14:00	0	2419	116.6	0.03	4.1
1/1/2008 15:00	0	1662	118.3	0.02	4.3
1/1/2008 16:00	0	1868	116.8	0.03	4.1
1/1/2008 17:00	0	2212	113	0.03	4.7
1/1/2008 18:00	0	2726	108.2	0.04	4.4
1/1/2008 19:00	0	2555	110.6	0.04	4.8
1/1/2008 20:00	0	2446	110	0.04	4.9
1/1/2008 21:00	0	2061	108.7	0.03	5.2
1/1/2008 22:00	0	1510	110.3	0.02	5
1/1/2008 23:00	0	1247	112.8	0.02	5.1
2/1/2008 0:00	0	666	115.9	0.01	4.4

Table 4-10: Example of dataset for accident probability in Attica Tollway.

Variable	Type	Description	Descriptive Statistics			
			Min	Median	Mean	Max
Q5min.prior_up	continuous	5min cumulative flow	4.00	162.00	189.50	583.00
Q_sum_15m_up	continuous	15min cumulative flow	6.00	167.00	189.10	629.00
Q_avg_15m_up	continuous	15min average flow	2.00	55.66	63.05	209.67
Q_stdev_15m_up	continuous	15min st.dev of flow	0.19	4.06	5.48	45.57
Q_median_15m_up	continuous	15min median flow	1.50	54.71	63.34	201.00
Q_cv_15m_up	continuous	15min coefficient of variation of flow	0.01	0.08	0.10	0.54
Q_sum_30m_up	continuous	30min cumulative flow	13.00	321.10	380.70	1253.00
Q_avg_30m_up	continuous	30min average flow	2.17	53.52	63.45	208.83
Q_stdev_30m_up	continuous	30min st.dev of flow	0.60	5.05	6.29	44.24
Q_median_30m_up	continuous	30min median flow	1.75	53.71	63.48	210.00
Q_cv_30m_up	continuous	30min coefficient of variation of flow	0.02	0.10	0.13	1.60
Q_sum_1h_up	continuous	1h cumulative flow	29.00	627.80	756.90	2470.00
Q_avg_1h_up	continuous	1h average flow	2.42	52.32	63.08	205.83
Q_stdev_1h_up	continuous	1h st.dev of flow	0.65	5.83	7.05	36.24
Q_median_1h_up	continuous	1h median flow	2.00	53.00	63.02	200.50
Q_cv_1h_up	continuous	1h coefficient of variation of flow	0.03	0.11	0.14	1.38
V5min.prior_up	continuous	5min speed	8.40	104.40	96.96	125.90
V_avg_15m_up	continuous	15min average speed	8.27	104.98	99.30	123.47
V_stdev_15m_up	continuous	15min st.dev of speed	0.06	2.21	4.18	54.91
V_cv_15m_up	continuous	15min coefficient of variation of speed	0.00	0.02	0.05	0.71
V_avg_30m_up	continuous	30min average speed	10.82	105.33	100.13	123.75
V_stdev_30m_up	continuous	30min st.dev of speed	0.44	2.58	4.20	38.95
V_cv_30m_up	continuous	30min coefficient of variation of speed	0.00	0.02	0.05	0.53
V_avg_1h_up	continuous	1h average speed	28.40	105.28	100.97	124.19
V_stdev_1h_up	continuous	1h st.dev of speed	0.47	2.76	4.44	32.58
V_cv_1h_up	continuous	1h coefficient of variation of speed	0.00	0.03	0.05	0.78
Occ5min.prior_up	continuous	5min occupancy	0.00	0.03	0.05	0.20
Occ_avg_15m_up	continuous	15min average occupancy	0.00	0.03	0.05	0.20
Occ_stdev_15m_up	continuous	15min st.dev of occupancy	0.00	0.00	0.00	0.06
Occ_cv_15m_up	continuous	15min coefficient of variation of occupancy	0.01	0.09	0.12	0.76
Occ_avg_30m_up	continuous	30min average occupancy	0.00	0.03	0.05	0.19
Occ_stdev_30m_up	continuous	30min st.dev of occupancy	0.00	0.00	0.00	0.05
Occ_cv_30m_up	continuous	30min coefficient of variation of occupancy	0.02	0.11	0.14	1.53
Occ_avg_1h_up	continuous	1h average occupancy	0.00	0.03	0.04	0.17
Occ_stdev_1h_up	continuous	1h st.dev of occupancy	0.00	0.00	0.01	0.05
Occ_cv_1h_up	continuous	1h coefficient of variation of occupancy	0.03	0.13	0.16	1.29
Tr.Prop5min.prior_up	continuous	5min truck proportion	0.00	4.05	5.30	25.20
Tr.Prop_avg_15m_up	continuous	15min average truck proportion	0.00	4.09	5.12	27.30
Tr.Prop_stdev_15m_up	continuous	15min st.dev of truck proportion	0.00	1.25	1.68	17.67
Tr.Prop_cv_15m_up	continuous	15min coefficient of variation of truck proportion	0.00	0.28	0.43	1.73
Tr.Prop_avg_30m_up	continuous	30min average truck proportion	0.00	4.40	5.14	25.58
Tr.Prop_stdev_30m_up	continuous	30min st.dev of truck proportion	0.00	1.43	1.92	14.00
Tr.Prop_cv_30m_up	continuous	30min coefficient of variation of truck proportion	0.00	0.36	0.58	2.45
Tr.Prop_avg_1h_up	continuous	1h average truck proportion	0.00	4.55	5.30	25.53
Tr.Prop_stdev_1h_up	continuous	1h st.dev of truck proportion	0.00	1.70	2.27	17.03
Tr.Prop_cv_1h_up	continuous	1h coefficient of variation of truck proportion	0.00	0.36	0.64	3.46

Table 4-11: Descriptive statistics of traffic related variables upstream of the accident location.

			Descriptive Statistics			
Variable	**Type**	**Description**	**Min**	**Median**	**Mean**	**Max**
T_1h_max	continuous	1h maximum temperature	0.00	19.83	20.23	38.88
T_1h_avg	continuous	1h average temperature	-0.86	19.48	19.74	38.46
T_1h_stdev	continuous	1h st.deviation of temperature	0.01	0.31	0.40	2.06
T_2h_max	continuous	2h maximum temperature	0.51	20.32	20.61	38.88
T_2h_avg	continuous	2h average temperature	-0.84	19.79	19.66	38.47
T_2h_stdev	continuous	2h st.deviation of temperature	0.03	0.52	0.62	5.87
T_6h_max	continuous	6h maximum temperature	0.51	21.54	21.37	38.88
T_6h_avg	continuous	6h average temperature	-0.76	18.83	19.29	37.50
T_6h_stdev	continuous	6h st.deviation of temperature	0.07	1.10	1.29	6.04
T_12h_max	continuous	12h maximum temperature	0.51	22.17	22.10	38.88
T_12h_avg	continuous	12h average temperature	-0.58	18.84	18.68	35.25
T_12h_stdev	continuous	12h st.deviation of temperature	0.17	1.85	2.04	5.37
Hum_1h_max	continuous	1h maximum humidity	15.20	56.31	56.03	99.60
Hum_1h_avg	continuous	1h average humidity	14.06	52.81	53.82	99.59
Hum_1h_stdev	continuous	1h st.deviation of humidity	0.04	1.27	1.74	9.94
Hum_2h_max	continuous	2h maximum humidity	15.20	57.37	57.65	99.60
Hum_2h_avg	continuous	2h average humidity	13.69	54.25	53.83	99.59
Hum_2h_stdev	continuous	2h st.deviation of humidity	0.03	1.97	2.55	12.37
Hum_6h_max	continuous	6h maximum humidity	17.81	63.77	63.29	99.90
Hum_6h_avg	continuous	6h average humidity	15.19	53.08	54.64	99.54
Hum_6h_stdev	continuous	6h st.deviation of humidity	0.06	3.93	4.93	17.82
Hum_12h_max	continuous	12h maximum humidity	24.88	69.01	68.56	99.90
Hum_12h_avg	continuous	12h average humidity	17.85	53.84	56.26	99.18
Hum_12h_stdev	continuous	12h st.deviation of humidity	0.06	6.77	7.33	19.29
Rain_1h_sum	continuous	1h sum of rainfall	0.00	0.00	0.08	16.40
Rain_1h_st.dev	continuous	1h st.deviation of rainfall	0.00	0.00	0.01	3.34
Rain_2h_sum	continuous	2h sum of rainfall	0.00	0.00	0.10	17.00
Rain_2h_st.dev	continuous	2h sum of rainfall	0.00	0.00	0.01	2.64
Rain_6h_sum	continuous	6h sum of rainfall	0.00	0.00	0.19	23.60
Rain_6h_st.dev	continuous	6h st.deviation of rainfall	0.00	0.00	0.01	1.65
Rain_12h_sum	continuous	12h sum of rainfall	0.00	0.00	0.56	55.40
Rain_12h_st.dev	continuous	12h st.deviation of rainfall	0.00	0.00	0.03	2.28
W.Sp_1h_max	continuous	1h maximum wind speed	0.37	3.22	3.54	11.94
W.Sp_1h_avg	continuous	1h average wind speed	0.09	2.64	2.90	10.72
W.Sp_1h_stdev	continuous	1h st.deviation of wind speed	0.08	0.38	0.45	3.40
W.Sp_2h_max	continuous	2h maximum wind speed	0.45	3.53	3.76	11.94
W.Sp_2h_avg	continuous	2h average wind speed	0.10	2.55	2.88	9.63
W.Sp_2h_stdev	continuous	2h st.deviation of wind speed	0.11	0.46	0.52	2.88
W.Sp_6h_max	continuous	6h maximum wind speed	0.45	3.93	4.27	12.26
W.Sp_6h_avg	continuous	6h average wind speed	0.06	2.39	2.78	9.88
W.Sp_6h_stdev	continuous	6h st.deviation of wind speed	0.11	0.70	0.74	2.68
W.Sp_12h_max	continuous	12h maximum wind speed	0.49	4.50	4.72	12.26
W.Sp_12h_avg	continuous	12h average wind speed	0.06	2.36	2.63	8.86
W.Sp_12h_stdev	continuous	12h st.deviation of wind speed	0.12	0.89	0.97	2.92
W.Dir_1h_avg	continuous	1h average wind direction	6.34	171.22	159.71	347.97
W.Dir_2h_avg	continuous	2h average wind direction	8.83	168.53	159.96	348.18
W.Dir_6h_avg	continuous	6h average wind direction	7.42	161.51	157.43	339.09
W.Dir_12h_avg	continuous	12h average wind direction	14.42	161.61	159.19	331.83
Sol_1h_max	continuous	1h maximum solar radiation	0.00	301.10	389.27	1112.00
Sol_1h_avg	continuous	1h average solar radiation	0.00	191.15	321.96	1056.43
Sol_2h_max	continuous	2h maximum solar radiation	0.00	338.00	432.06	1115.00
Sol_2h_avg	continuous	2h average solar radiation	0.00	152.28	318.32	997.15
Sol_6h_max	continuous	6h maximum solar radiation	0.00	578.60	548.10	1117.00
Sol_6h_avg	continuous	6h average solar radiation	0.00	192.59	290.97	971.00
Sol_12h_max	continuous	12h maximum solar radiation	0.00	690.90	647.70	1226.00
Sol_12h_avg	continuous	12h average solar radiation	0.00	183.01	228.21	698.26

Table 4-12: Descriptive statistics of weather related variables for Attica Tollway.

Chapter 5 Time series and machine learning analysis to predict PTW accident involvement and accident type

5.1 Introduction

Predicting and explaining accident probability with high resolution traffic data has been a continuously researched topic in the last years as shown from the literature review conducted. However, there is no specific focus on Powered-Two-Wheelers. Moreover, urban roads were not considered so far. In terms of methodology, there is a clear gap, because of the fact that time series were not utilized in predicting road accidents so far. Moreover, as Li et al. (2008) suggest, the assessment of SVM models' performance when only traffic flow is considered, should be implemented.

This chapter aims to add to the current knowledge by utilizing time series data and by exploiting them by applying advanced machine learning techniques to analyse and predict PTW accident involvement and PTW accident type in urban roads. The term "predicting PTW accident involvement", refers to the prediction of whether a PTW is involved or not in an accident that has occurred (as it was defined in Chapter 1).

5.2 Methodology

In order to predict PTW involvement and PTW accident type a combined methodological approach was followed. More specifically, a time series classification was performed by applying SVMs (Support Vector Machines), which is a very powerful and advanced machine learning technique. Firstly, the SVMs were applied by utilizing the original time series data and secondly by utilizing the DWT (Discrete Wavelet Transform) transformed data. Lastly, the results are compared.

5.2.1 Time series approach

5.2.1.1 Original time series data

Time series classification is used when it is desired to build a classification model based on labelled time series and then use the model to predict the label of unlabelled time series. Classification of unlabelled time series to existing classes is a further traditional data mining task. By "labelled time series", it means that a training dataset with correctly classified observation is used, and then the built models are used to predict the labels of a test dataset (Kleist, 2015).

It is possible to extract new features from time series in order to potentially improve the performance of classification models. There are various such techniques for feature extraction such as the Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), Piecewise Aggregate Approximation (PAA), Perpetually Important Points (PIP), Piecewise Linear Representation and Symbolic Representation (Zhao, 2012).

In this approach, the original time series data are used, namely data which have been sampled at equispaced points in time, without applying any techniques for feature extraction.

5.2.1.2 Wavelet transform

The Wavelet transform provides a multi-resolution representation using wavelets. In this chapter, a Discrete Wavelet Transform (DWT) is used to extract features from time series and then build a classification model (Burrus et al., 1998). The time series and its transform can be considered to be two representations of the same mathematical entity.

The very name wavelet originates from the requirement that they should integrate to zero, “waving" above and below the x-axis (Vidakovic and Mueller, 1991). A DWT is any wavelet transform for which the wavelets are discretely sampled and is an orthornormal transform. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information (location in time).

McLeod et al., (2012) provide a very good description of Discrete Wavelet Transform. A time series of dyadic length is considered z_t,t= 1,......n, where n = 2^J. The discrete wavelet transformation (DWT) decomposes the time series into J wavelet coefficients vectors, $W_{i,j}$= 0,......,J-1 each of length n_j= 2^{J-j}, j= 1,......,J plus a scaling coefficient V_J. Each wavelet coefficient is constructed as a difference of two weighted averages each of length λ^j= 2^{j-1}. Similarly to Discrete Fourier Transformation, the DWT provides an orthonormal decomposition, W= WZ, where W′= (W'_1,…,W'_{J-1}, V'_{J-1}), Z= (z_1,......,z_n)′.

There are two functions that play a primary role in wavelet analysis, the scaling function (father wavelet) and the wavelet (mother wavelet). The simplest wavelet analysis is based on Haar scaling function, (Haar Wavelet Transform), (Struzik and Siebes, 1999). The Haar wavelet is a sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. The Haar sequence is recognised as the first known wavelet basis and it was proposed in 1909 by Alfréd Haar (Haar, 1910).

The Haar scaling function φ(x) is defined as:

$$\varphi(x) = \begin{cases} 1, & 0 \le x < 1 \\ 0, & otherwise \end{cases} \qquad \text{(Eq. 5-1)}$$

The Haar Wavelet's mother function is then defined as $\psi(x)= \varphi(2x)-\varphi(2x-1)$

$$\psi(x) = \begin{cases} 1, & 0 \leq x < 1/2 \\ -1, & 1/2 \leq x < 1 \\ 0, & otherwise. \end{cases} \quad \text{(Eq. 5-2)}$$

Figure 5-1 that follows, illustrates an example of a graphical representation of a Haar Wavelet Transform, which is the simplest DWT (Zhao, 2012). Figure 5-2 shows another graphical example of a Haar Wavelet (Vidakovic and Mueller, 1991).

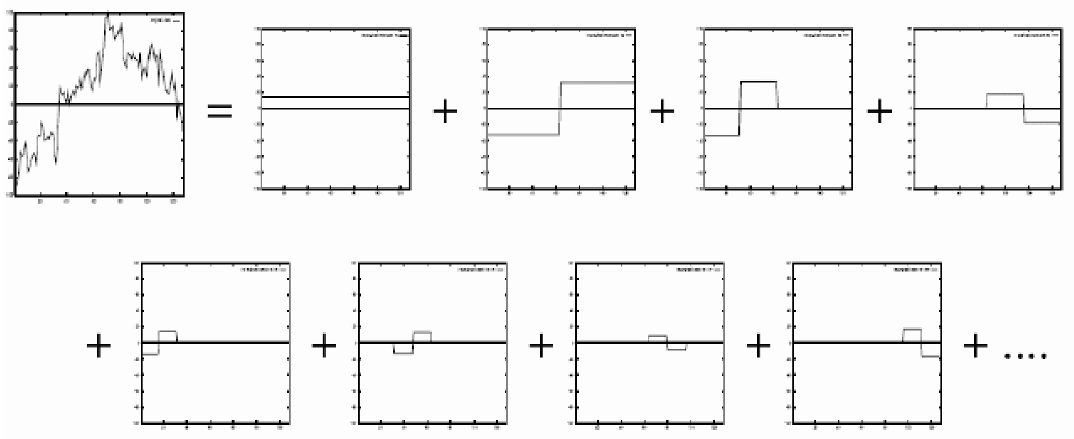


Figure 5-1: A graphical representation of a simple Haar Wavelet Transform.

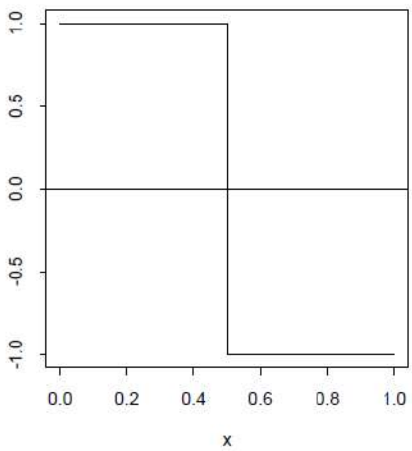


Figure 5-2: A Haar Wavelet.

5.2.2 Support vector machines

Traditional statistical modelling has been widely used for transportation data analysis. However, such approach contains some limitations, for example modelling assumptions that may not always be true. Non-parametric and artificial intelligent methods could then be applied to overcome such limitations.

Support Vector Machines (SVMs), constitute a relatively new modelling technique, which is useful for classification problems (Keckman, 2005). In transportation science, the studies having used SVMs are relatively rare (Li et al., 2008; Li et al., 2012), especially in real-time crash risk evaluation (Yu and Abdel-Aty, 2013b and 2014b).

SVMs have originated from statistical learning theory (Vapnik, 1998), and have been developed by Cortes and Vapnik (1995) mainly for binary classification. Basically, when building a SVM model, the aim is the optimal separating hyperplane between two classes by maximizing the margin between the classes' closest points (Meyer, 2001). Therefore, different classes are separated by the hyperplane:

$$\langle w, \Phi(x)\rangle + b = 0 \qquad \text{(Eq. 5-3)}$$

which corresponds to the decision function

$$f(x) = sign(\langle \Phi(x_i), w\rangle + b) \qquad \text{(Eq. 5-4)}$$

The points lying on the boundaries are the support vectors, while the middle of the margin is the optimum separating hyperplane. Figure 5-3 provides a graphical illustration of a linear separable example of SVMs (Meyer, 2001).

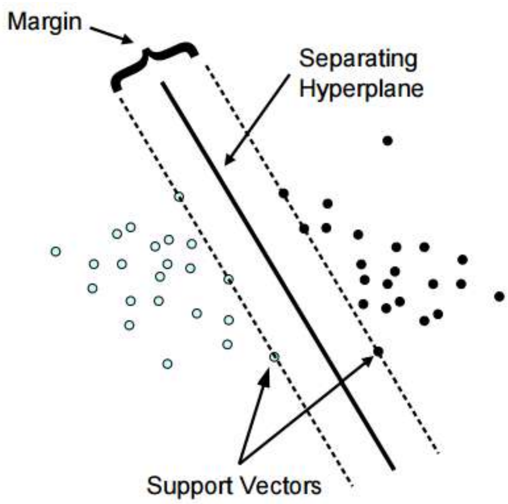


Figure 5-3: Graphical illustration of SVM classification (linear separable example).

In addition, SVMs can be enhanced to tackle nonlinear classification problems, regression and outlier detection. The major limitation of SVMs is that the models cannot be directly used to identify the relationships between the dependent and the independent variables. Therefore, SVMs can be considered as a "black box" technique. The reader is encouraged to refer to Karatzoglou et al. (2006), where a very detailed description of SVM model formulation is provided for R language.

From all the available kernel-based algorithms (kernels) (e.g. linear, polynomial, Gaussian Radial-basis function and sigmoid), the Gaussian Radial-basis function kernel was considered in this thesis (Karatzoglou et al., 2005):

Radial-basis function kernel (RBF):

$K(x_i, x_j) = \exp(-\gamma)||x_i, x_j||^2), \gamma > 0$ (Eq. 5-5)

where, γ is the kernel parameter.

Moreover, with the Gaussian Radial-basis kernel function, the SVM model has two parameters (C, γ) which need to be determined. The cost parameter C controls the penalty for misclassifying a training point and consequently the complexity of the prediction function (Karatzoglou et al., 2006). A high cost value C, will result in a complex prediction function in order to misclassify as few training cases as possible. On the other hand, a low cost parameter C, results in simpler prediction functions. Thus, this type of SVM model is called C-SVM (Karatzoglou et al., 2006).

Karatzoglou et al. (2006), provide the primal form of the bound constraint C-SVM formulation:

minimize $t(w, \xi) = (\frac{1}{2})||w||^2 + (\frac{1}{2})\beta^2 + (\frac{C}{m}) \sum_{i=1}^{m} \xi_i$

subject to $y_i(\langle \Phi(x_i), w \rangle + b) \geq 1 - \xi_i$ (Eq. 5-6)

where, i= 1,……m

and $\xi_i \geq 0$, where i= 1,……m.

The dual form of the bound constraint C-SVM formulation (Karatzoglou et al., 2006) is:

maximize $W(\alpha) = \sum_{i=1}^{m} a_i - \frac{1}{2} \sum_{i,j=1}^{m} a_i a_j (y_i y_j + k(x_i, x_j))$

subject to $0 \leq a_i \leq \frac{C}{m}$, where i= 1,......m (Eq. 5-7)

and $\sum_{i=1}^{m} a_i y_i = 0$.

5.3 Data preparation

In order to apply the SVM models, a number of different datasets had to be prepared. Having known the time and location for each accident, the 3-hour time series of traffic flow, occupancy and speed (in 5-minutes intervals ending at the calculated time of the accident) from the closest upstream and closest downstream loop detector were utilized. Thus, each dataset contains one of the following time series: traffic flow upstream, traffic flow downstream, speed upstream, speed downstream, occupancy upstream, occupancy downstream. Then, by following the Discrete Wavelet Transform (DWT) procedure described in section 5.2.1.2, all datasets are then transformed. Consequently, for each dataset containing the original time series data a duplicate dataset is created with the transformed time series data.

Then for each dataset, the SVM models were applied in order to predict:

a) PTW accident involvement given an accident occurrence
b) PTW accident type given an accident occurrence.

In order to enhance the classification performance of SVMs, PTW accident type was classified as a binary outcome, namely single and multi-vehicle accidents, transforming the classification problem to a two-category classification. This approach was followed, because literature indicates that typical multi-classification problems have been very commonly observed when methods such as SVMs, Artificial Neural Networks (ANN) or classification trees are applied (Li et al., 2012; Delen et al., 2006). For example, Li et al. (2012), developed SVMs to model injury severity and the SVM model ignored severity categories with small proportions (namely fatal and incapacitating injuries) to improve the overall classification accuracy.

The final datasets consist of 527 accidents where the PTWs were involved in 326 of them (61.9%). Regarding PTW accident type, PTWs were involved in 107 single-vehicle accidents (32.8%) and 219 multi-vehicle accidents (67.1%).

5.4 Results

The DWT procedure was conducted with the use of the package *wavelets* (Aldrich, 2010) in R software (R, 2010), and all SVM models were developed through the package *e1071* (Dimitriadou et al., 2005) in R software. It is noted that when building a SVM model, a training and a testing set have to be defined. For each dependent variable; PTW accident involvement and PTW accident type, respectively.

When building a SVM model, a training and a testing dataset have to be defined. The models are calibrated on the training set and then are used for prediction of the dependent variable on the testing set. Two different training and testing sets have to be prepared in order to further compare the results and to reduce the bias that arises when the accident database is randomly separated as Li et al. (2008) propose. In this thesis, a 10-fold cross validation technique was applied on each dataset, in order to have a measure of the overall classification performance of the SVM models. Generally, in k-fold cross-validation, the original sample is randomly divided into k equal sized subsamples. Of the k subsamples, a single subsample is used as the validation dataset for testing the prediction performance of the model while the remaining k−1 subsamples are used as training data to calibrate the model. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data (Kohavi, 1995). Consequently, 10 subsamples were created. The total accuracy of SVM models was calculated with this simple equation:

$$\text{Accuracy} = \frac{number\ of\ correctly\ predicted\ cases}{total\ number\ of\ cases} \qquad \text{(Eq. 5-8)}$$

Moreover, different parameters of C and γ were tested each time in order to improve the performance of SVMs. The need for an investigation of SVMs performance by applying different values of C and γ was emphasized by Yu and Abdel-Aty (2013b), being another step further made by this thesis.

5.4.1 PTW accident involvement prediction

In general, the SVM models of this thesis showed relatively good classification accuracy compared to other similar studies (Li et al., 2012). However, it is noted that this is the first attempt to incorporate real-time traffic time series in SVMs. The development of the models showed that by modifying the two parameters, C and γ accordingly, the classification accuracy can be substantially influenced. A detailed presentation of SVM models' performance follows. Table 5-1 illustrates the total classification accuracy of SVM models regarding PTW accident involvement. In general, the original time series approach produced better results than the DWT approach. Overall, in regard with the parameters C and γ, the models performed better when the Cost parameter C had taken the value of 10 or 100, while the parameter γ values ranged from 0.001 to 0.01.

Total SVM performance	Speed downstream	Speed upstream	Flow downstream	Flow upstream	Occupancy downstream	Occupancy upstream
Original Time series						
10-fold cross validation mean accurracy %	62.17%	65.64%	61.90%	64.15%	60.72%	65.38%
(C, γ)	(100, 0.002)	(10, 0.01)	(100, 0.001)	(100, 0.001)	(100, 0.0001)	(100, 0.001)
DWT Time series						
10-fold cross validation mean accurracy %	60.65%	58.38%	60.82%	61.39%	59.20%	61.54%
(C, γ)	(10, 0.001)	(100, 0.01)	(100, 0.002)	(100, 0.0001)	(100, 0.01)	(100, 0.0001)

Table 5-1: Total classification accuracy of SVM models to predict PTW accident involvement.

When the original time series are considered, the best performance is consistently higher than 60%. The best accuracy is achieved when the speed upstream (65.64%) and occupancy upstream (65.38%) are considered. On the other hand, the time series of occupancy downstream of the accident location, were the worse predictors of PTW accident involvement (60.72%). When the DWT time series are considered, the total classification accuracy was generally around 60%. Speed upstream and occupancy downstream had the lowest accuracies (58.38% and 59.20% respectively).

The lower difference between original and DWT time series accuracy, was observed in speed downstream (1.52%) and occupancy downstream (1.52%). On the contrary, the highest difference was observed in speed upstream time series (7.26%). In all aforementioned situations, the DWT performed worse. The lower prediction performance of the transformed time series in predicting PTW accident involvement, may imply that there may be not necessary to extract features from time series but use the original time series instead. Alternatively, a different transformation could be applied as mentioned earlier in section 5.2.1.1.

5.4.2 PTW accident type prediction

Table 5-2 illustrates the total classification accuracy of SVM models regarding PTW accident type. As a first remark, neither of the two approaches (original or DWT time series) clearly outperformed the other, as the classification accuracies were similar. In regard with the parameters C and γ, the models performed better when the Cost parameter C had taken the value of 10 or 100, while the parameter γ values ranged from 0.000001 to 0.01.

Total SVM performance	Speed downstream	Speed upstream	Flow downstream	Flow upstream	Occupancy downstream	Occupancy upstream
Original Time series						
10-fold cross validation mean accurracy %	65.20%	66.76%	65.64%	66.57%	66.56%	66.67%
(C, γ)	(100, 0.01)	(100, 0.0001)	(10, 0.01)	(100, 0.0001)	(100, 0.0001)	(100, 0.0001)
Wavelet transformation						
10-fold cross validation mean accurracy %	63.63%	66.46%	66.26%	66.67%	65.31%	65.77%
(C, γ)	(100, 0.01)	(10, 0.001)	(100, 0.0001)	(10, 0.000001)	(10, 0.001)	(10, 0.001)

Table 5-2: Total classification accuracy of SVM models using to predict PTW accident type.

Overall, the models perform better in accident type prediction than in PTW accident involvement. When the original time series are considered, the best performance is consistently higher than about 65%. The best accuracy is achieved when the speed upstream (66.74%) is utilized. On the other hand, the utilization of time series of traffic flow downstream of the accident location, provided the lowest prediction of PTW accident involvement (65.64%). When the DWT time series are considered, the total classification accuracy was generally around 63-66%. Speed downstream had the lowest accuracy (63.63%), whilst flow upstream had the best accuracy (66.67%).

Moreover, when predicting PTW accident type, it is observed that both approaches provided similar results. Original time series performed better that DWT, when speed downstream and occupancy (both in upstream and downstream) are utilized. On the other hand, DWT performed slightly better for flow downstream. For speed upstream and for flow upstream, the SVM models showed similar results for original and DWT time series. Therefore, it may be suggested that both approaches are appropriately adequate for predicting PTW accident type.

The next two figures (Figure 5-4 and 5-5), demonstrate the relative performance of the SVM models when predicting PTW accident involvement and PTW accident type, by utilizing original and DWT time series.

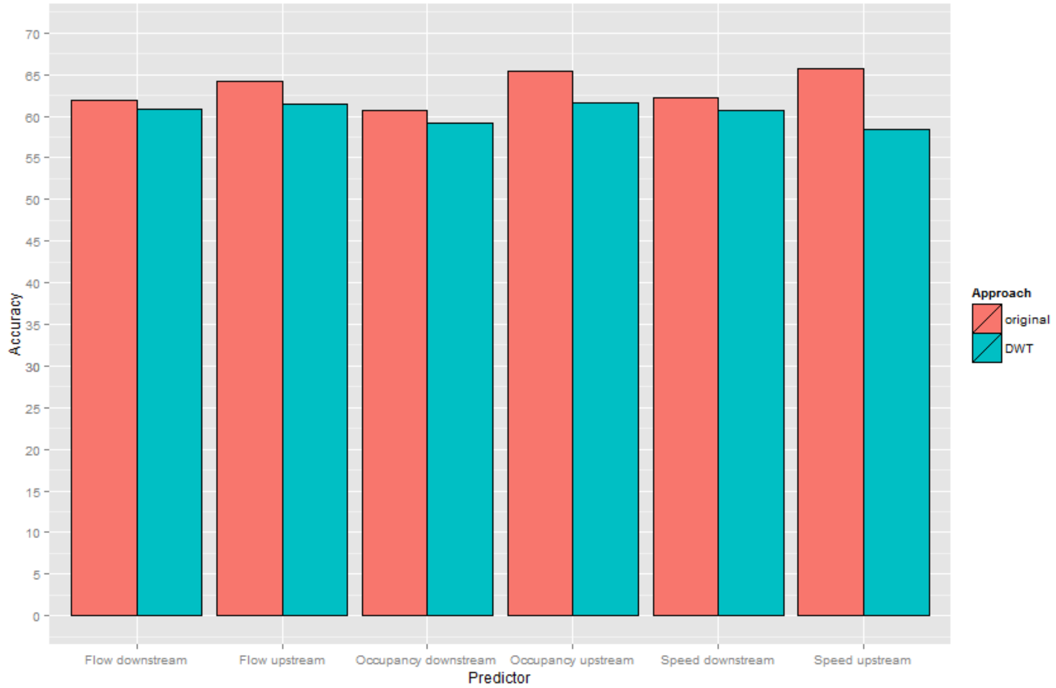


Figure 5-4: Graphical illustration of SVM models performance when analysing PTW accident involvement.

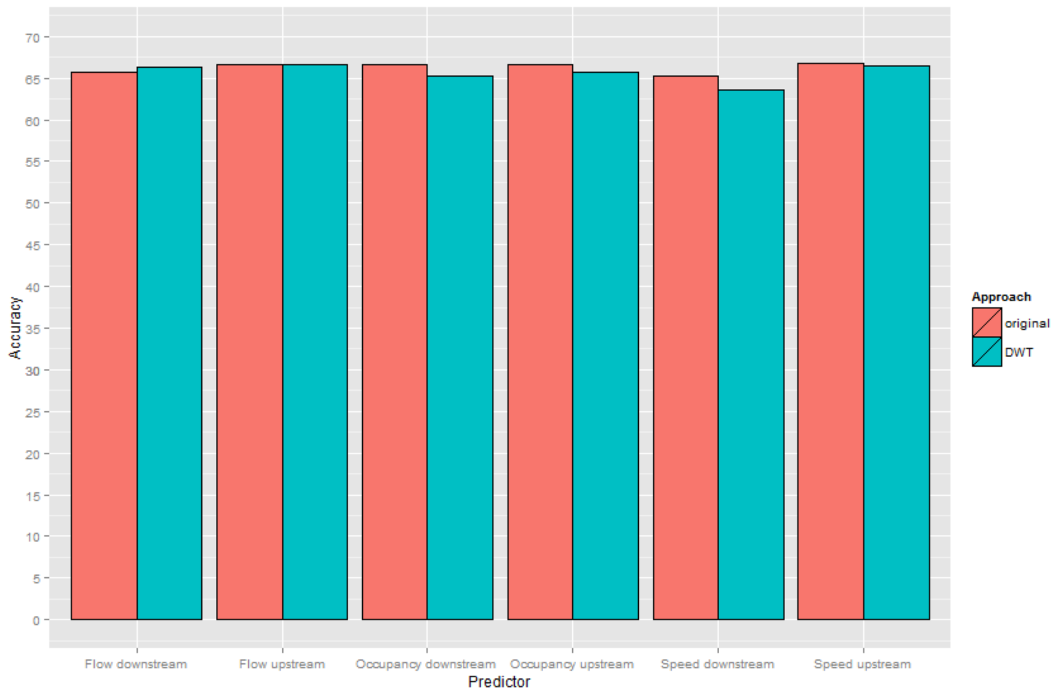


Figure 5-5: Graphical illustration of SVM models performance when analysing PTW accident type.

5.5 Summary

This chapter, has presented the prediction results from the Support Vector Machine (SVM) models, which were applied on the original time series and on the Discrete Wavelet Transformed (DWT) time series of flow, speed and occupancy upstream and downstream of the accident location. This methodological approach, was a first attempt to incorporate time-series data when analysing road safety with real-time traffic data. Moreover, the opportunity of applying of a relatively new and scientifically strong classification technique such as Support Vector Machines, in road safety with real-time traffic data was only recently been explored.

The prediction of whether a PTW is involved in an accident, given that the accident has occurred, is two-category classification problem. Overall, the original time series data performed better than the DWT time series. However, in a few cases, the difference between original time series and DWT time series was very low. Consequently, original time series are preferred for predicting PTW accident involvement.

This combined approach proposed by this thesis, is considered promising when investigating the accident type of Powered-Two-Wheelers (PTW) that occur in urban roads. The dependent variable (single and multi-vehicle accident) is considered as a two-category classification problem as well. In this case, both original time series and DWT time series performed well. Neither of the two approaches clearly outperformed the other. Therefore, both type of time series could be utilized and produce good prediction.

In general, it is observed that this combined approach provides better results when accident type is aimed to be predicted. Moreover, when original time series are utilized, upstream speed had consistently the best accuracy both for PTW accident involvement and PTW accident type.

Summing up, this methodological approach showed promising results, having produced a number of adequately high correct classification percentages in some cases. The major conclusion of this chapter is that the combination of SVM models and time-series data can be used for road safety purposes especially by utilizing high resolution traffic data. Clearly, this direction has to be exploited further. It is interesting though, that despite the fact that Zhao (2012) suggests to extract features from time-series when perform time series data-mining, the performance of SVMs on the DWT data did not generally outperformed original time series. This means that the transformation applied in the thesis is not always necessary, but alternatively, a different transformation could be applied, such as Singular Value Decomposition (SVD), Discrete Fourier Transform, Piecewise Aggregate Approximation, Perpetually Important Points and so on.

Chapter 6 Modelling the effect of traffic states on safety in urban roads

6.1 Introduction

During the past decade, increased attention has been given to build relationships between real-time traffic characteristics and risk of accident occurrence in freeways. The impact of traffic states on traffic safety has only relatively recently been explored. For example, Abdel-Aty et al. (2005), divided freeway traffic flow in high and low speed states and then examined severity and mechanism of multi-vehicle accident occurrence under these two different states, finding different results. Moreover, Xu et al. (2012), demonstrate the need to divide traffic in states and explore their effect on safety due to the fact that different traffic states may have different influence on the risk of an accident.

This approach has only recently been explored in freeways, but urban roads are not considered yet. Ensuring safety in major urban roads holds high priority, consequently, the primary objective of this chapter is to divide urban traffic flow into different states and to investigate their effect on accident probability and severity. Furthermore, PTW accident probability and severity are also explored for the first time by using this approach. It is noted that the analysis did not focus on prediction but on explaining accident severity and accident probability. In the present thesis, the term "PTW accident severity" is defined as the severity of the accident (classified by the most severely injured person) when a PTW is involved in the accident. PTW accident probability, is the likelihood that a PTW is involved in an accident given that this accident has already occurred. In this chapter, accident probability in general is defined as the probability of an accident occurrence by using a sample of non-accident cases as described in Chapter 4 of the thesis.

The findings of this chapter demonstrate that urban traffic can be divided into different states by using average traffic occupancy and its standard deviation, measured by nearby upstream and downstream loop detectors. Moreover, the hazardous traffic conditions were identified and discussed.

6.2 Methodology

The statistical methods that applied to achieve the aims of this chapter, are described in the following subsections. Expectation Maximization clustering (EM) (or Finite Mixture) was used to classify traffic into different states. In addition, Bayesian logistic regression models were applied in order to correlate traffic states with traffic safety.

6.2.1 Finite mixture cluster analysis

In order to divide traffic in states, a Gaussian finite mixture model-Based clustering (covariance parameterization and number of cluster selected via the Bayesian Information Criterion) was followed. The models were fitted by Expectation-Maximization algorithm. Fraley et al., (2012) and Fraley and Raftery (2002), provide a detailed description of Normal Mixture Modeling. All following equations appear in Fraley et al. (2012).

A normal or Gaussian mixture model is assumed:

$$\prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k \varphi_k(x_i|\mu_k, \Sigma_k), \quad \text{(Eq. 6-1)}$$

where x are the data, G is the number of components, τ_k is the probability that a case belongs to the k^{th} component ($\tau_k \geq 0$; $\Sigma_k^{-1}\tau_k = 1$) and

$$\varphi_k(x|\mu_k, \Sigma_k) = (2\pi)^{-\frac{p}{2}}|\Sigma_k|^{-\frac{1}{2}}exp\left\{-\frac{1}{2}(x_i - \mu_k)^T\Sigma_k^{-1}(x_i - \mu_k)\right\}. \quad \text{(Eq. 6-2)}$$

The cluster is ellipsoidal, centered at the means μ_k. Their other geometrical features are determined by Σ_k. Banfield and Raftery (1993), suggested that each covariance matrix parameterized by eigenvalue decomposition takes the following form:

$$\Sigma_k = \lambda_k D_k A_k D_k^T, \quad \text{(Eq. 6-3)}$$

where λ_k is a scalar, D_k is the orthogonal matrix of eigenvectors, A_k is a diagonal matrix where all the elements are proportional to Σ_k. It is suggested that characteristics of distributions such as orientation, volume and shape, are estimated by the data, and can either vary between clusters or remain the same for all clusters (Banfield and Raftery, 1993; Fraley et al., 2012: Celeux and Govaert, 1995).

According to Fraley et al. (2012), the distribution for Expectation Maximization (EM) algorithm for multidimensional data, can be Spherical, Diagonal or Ellipsoidal. The Volumes and the Shapes of clusters can be equal or variable. The combination of these characteristics, defines each model (namely the covariance matrix Σ_k). For more information, the reader is encouraged to read the report by Fraley et al. (2012).

The best model is determined according to the BIC (Bayesian Information Criterion) as initially proposed by Schwarz (1978). The BIC is basically the maximized log-likelihood but in order to avoid overfitting it includes a penalty term for the number of parameters in the model. The optimum number of clusters and the best model are defined by the value of BIC. A larger value of BIC indicates stronger evidence for the best model and number of clusters (Fraley and Raftery, 2002).

6.2.2 Finite mixture discriminant analysis

In general, the cluster analysis provides the optimum number of clusters as well as their centres. However, as corrected stated by Xu et al. (2012), new observations cannot be directly assigned to the defined traffic states. In that case a discriminant analysis is needed to be conducted. According to Johnson and Wichern (1998), discriminant analysis allocates new cases to the pre-defined cluster groups. Fraley and Raftery (2002), state that in discriminant analysis (or supervised classification), known classifications (training set) are used to classify others (testing set).

Several discriminant analysis methods exists. In this thesis, a Discriminant analysis through Eigenvalue Decomposition was applied. More specifically, the procedure of applying a Gaussian finite mixture modelling for discriminant analysis, where each known group (class) is modelled by a single Gaussian term with the same covariance structure among classes is named as Eigenvalue Decomposition Discriminant Analysis (EDDA) by Bensmail and Celeux (1996). When the model is a normal mixture fitted by model-based clustering, the procedure is known as mclustDA (Fraley and Raftery 2002).

In the followed approach, a separate mean vector for each class is calculated, but with the same ellipsoidal covariance matrix, which basically is the same with linear discriminant analysis.

6.2.3 Bayesian logit

In this chapter several Bayesian logistic regression models were developed to estimate the effect of traffic states on accident severity and probability with focus on PTWs. The classical statistical approach (also called frequentist approach) is different than the Bayesian approach. The general philosophy behind Bayesian approach, is that the prior distributions for each parameter are defined and then the data are used to update beliefs about the behaviour of parameters. Moreover, the updated probability of the parameters are used and the posterior credible intervals are produced. The correct interpretation is that a parameter of interest lies within the credible interval with 95% probability. In that context, instead of a t-test, each parameter is statistically significant if the 95% credible interval (2.5%-97.5%) of the beta coefficient does not contain zero (Lunn et al., 2012). As stated by some studies (Ahmed et al., 2012b), the Bayesian inference can effectively treat overfitting problems.

Bayesian inference for logistic regression follows the usual procedure for all Bayesian analysis. More specifically, a prior distribution for all unknown parameters has to be formed, then the likelihood function of the data has to be defined and lastly, the Bayes theorem has to be applied so as to find the posterior distribution of all parameters.

The likelihood function for Bayesian logistic regression is the same as in the frequentist inference. More specifically,

$$likelihood_i = \pi(x_i)^{y_i}(1-\pi(x_i))^{(1-y_i)} \quad \text{(Eq. 6-4)}$$

where $\pi(x_i)$ is the probability of the event for the i^{th}subject which has covariate vector x_i. The term y_i is the response variable which has the outcomes y=1 (occurrence of event) or y=0 (absence of event). The logistic regression equation is:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \quad \text{(Eq. 6-5)}$$

where β_0 is the intercept, β_i is a coefficient for the explanatory variable x_i. In addition, similarly to the frequentist approach, taking the exp(β) provides the odds ratio for one unit change of that parameter.

Any prior distribution can be used for unknown parameters, however, it is usually preferable to use the so-called "vague" or "non-informative" priors if little is known about the coefficient values (Lunn et al., 2012). A non-informative prior could be for example a normal distribution with zero mean and very large variance, for example its form could be: $\beta_j \sim normal(0, 100^2)$. Another popular non-informative prior could be for example a uniform distribution with large boundaries a and b, e.g. $\beta_i \sim uniform(-100, 100)$.

It is noted that in the vast majority of cases when choosing a normal distribution as a prior distribution, the precision is considered. The precision is defined as $\frac{1}{\sigma^2}$, where σ^2 is the variance. Therefore, the distribution of the aforementioned example is transformed to $\beta_j \sim normal(0, 0.0001)$. Lastly, the posterior distribution is derived if the prior distribution over all parameters is multiplied by the full likelihood function. Thus,

$$prior \times likelihood = posterior \quad \text{(Eq. 6-6)}$$

6.3 Data preparation

In this chapter, accident and traffic data from the Kifisias and Mesogeion avenues were used to explore the effect of traffic states on accident probability and severity. Thus, only urban roads were considered as similar studies explored freeways (Xu et al., 2012). Furthermore, Powered-Two-Wheelers (PTWs) were considered as well. The final urban dataset for this chapter, included 527 accidents and 1581 non-accident cases for Kifisias and Mesogeion from 2006 to 2011. PTWs were involved in 326 of those accidents (61.9% of the accidents). Accident severity had two levels, namely fatal-severe injury (KSI) and slight injury (SI). There were 57 severe accidents (KSI) and 470 slight accidents.

For the needs of the analysis, a new variable was also created, namely accident occurrence, which was also a binary variable (1 for accident occurrence, 0 for non-accident occurrence). More specifically, when exploring accident probability, for each accident observation of the dataset, there were 2 observations for non-accidents for the same time and location, one week before and one week after the accident occurrence. For example, if an accident has occurred in Kifisias Avenue on

Wednesday 12 August 2009 at 13:00 and is recorded from the loop detector MS259, 3 hours data have also been collected for Wednesday 5 August 2009 and Wednesday 19 August 2009 for the period 10:00-13:00.

Traffic data from the nearest both downstream and upstream loop detectors were considered. As stated in Chapter 4 of the thesis, the 5-min traffic data were further aggregated to 1-hour level to obtain averages, standard deviations and so on, prior to an accident occurrence. It was anticipated that the 60-min traffic data before accident occurrence would cover the hazardous traffic conditions leading to accident occurrence and consequently only traffic data 1-hour prior to accident occurrence were initially considered. The same time slice (1-hour prior to an accident) was considered for non-accident cases as well.

The accident data were combined with the traffic data to build the required dataset for the statistical analysis. The thesis followed the approach of Xu et al. (2012), who used the traffic occupancy of both upstream and downstream loop detectors to classify traffic states. This approach was extended by considering also the standard deviation of occupancy and by developing more advanced statistical models. It is noted that weather data were not considered in the proposed thesis approach.

6.4 Results

To achieve the aims of the chapter, by utilizing average and standard deviation of occupancy, a finite mixture cluster analysis took place to identify traffic states, followed by a finite mixture discriminant analysis to assign traffic to traffic states. All required EM cluster and discriminant analyses were conducted via the package *mclust* (Fraley and Raftery, 1999 and 2003) in R software. Lastly, the effects of traffic states on accident probability and severity by applying Bayesian logit models was conducted.

6.4.1 Finite mixture cluster analysis

The finite mixture cluster analysis was conducted by using upstream and downstream occupancy measurements (average and standard deviation). Firstly, the accident cases were explored and then the non-accident cases. Therefore, two (2) clustering and discriminant models were developed.

6.4.1.1 Accident cases

The average and standard deviation of occupancy were assigned to the dataset of the 527 accidents for the two urban roads. The finite mixture cluster analysis results revealed five clusters. This was the optimal number of clusters as determined by the BIC criterion. Moreover, the finite mixture models showed the optimal covariance matrix Σ_k. More specifically, it has the form $\Sigma_k = \lambda_k D_k A_k D_k^T$ having an Ellipsoidal distribution, with a varying volume, shape and orientation between clusters (abbreviated VVV, please see Fraley et al., 2012). The optimum number of clusters was determined to be five. Table 6-1 that follows, illustrates the percentage of cases in each cluster, as well as the mean value (clustering centre) for each cluster. The description of the five produced clusters is provided below.

Traffic states	State 1	State 2	State 3	State 4	State 5
Percentage of total cases (%)	26%	25.24%	12.71%	23.72%	12.33%
Average Occupancy upstream (%)	11.28	4.92	21.1	26.53	24.28
Average Occupancy downstream (%)	11.49	6.01	30.16	12.03	27.06
St.deviation of Occupancy upstream (%)	1.66	0.75	8.52	8.17	4.38
St.deviation of Occupancy downstream (%)	2.45	0.84	6.29	1.94	7.88

Table 6-1: Clustering centres for different traffic states (accident cases).

<u>Traffic state 1:</u> 26% of accident cases belong to traffic state 1. Traffic state 1 is characterized by quite identical medium occupancy values in both upstream and downstream loop detectors. More specifically, the clustering centre for average occupancy upstream is 11.28%, while for downstream the respective mean is 11.49%. The fluctuations in occupancy are similar as well. The level of service in that state is considered as high.

<u>Traffic state 2:</u> 25.24% of accident cases belong to traffic state 2. This state is characterized by quite homogenous low occupancy across the two loop detectors (4.29% and 6.01%). The standard deviations of occupancy upstream and downstream are homogenous and low as well. The level of service in this state is considerably high.

<u>Traffic state 3:</u> 12.71% of cases are assigned to cluster 3. The main characteristic of this cluster, is the great difference observed in occupancy between upstream and downstream loop detectors. More specifically, the clustering centre for average occupancy downstream is 30.16%, whilst the respective centre upstream is 21.1%. These high values of occupancy indicate potential traffic congestion or to be more specific, transition from congestion to even higher. Another interesting attribute of this cluster, is that the standard deviation of occupancy upstream is higher (8.52%) than downstream (6.29%).

<u>Traffic state 4:</u> This traffic states consists of 23.72% of accident cases. The main difference from traffic state 4 is the opposite traffic conditions. In other words, there is a transition from congestion (26.53% average occupancy upstream) to free flow (12.03% average occupancy downstream). The clustering centre for standard deviation of occupancy is greatly higher upstream (8.17%) compared to downstream (only 1.94%).

<u>Traffic state 5:</u> Cases in this clusters are characterized by homogenous and high occupancy across upstream and downstream loop detectors (24.28% and 27.06% respectively). The cluster's centre for standard deviation of occupancy downstream is slightly higher than upstream. Only 12.33% of accidents belong to this cluster.

6.4.1.2 Accident and non-accident cases

The average and standard deviation of occupancy were assigned to the total dataset of the 1581 accident and non-accident cases (527 accidents and 1054 non-accidents) for the two urban roads. The finite mixture cluster analysis results revealed nine clusters. This was the optimal number of clusters as determined by the BIC criterion. Similarly, the finite mixture models showed the optimal covariance matrix Σ_k. More specifically, it has the form $\Sigma_k = \lambda_k D_k A_k D_k^T$ having an Ellipsoidal distribution, with a varying volume, shape and orientation between clusters (abbreviated VVV). The optimum number of clusters was determined to be five. The horizontal axis shows the number of clusters (components), while the vertical axis shows the respective values of BIC. Table 6-2, presents the percentage of cases in each cluster, as well as the mean value (clustering centre) for each cluster for non-accident cases. The nine clusters are summarized below:

Traffic states	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
Percentage of cases (%)	12.02%	21.38%	17.43%	11.30%	10.28%	1.90%	6.45%	9.49%	9.74%
Average Occupancy upstream (%)	10.81	4.35	10.52	23.52	18.89	30.82	28.59	12.87	29.74
Average Occupancy downstream (%)	14.15	5.45	7.95	11.24	14.89	6.07	34.78	27.9	24.53
St.deviation of Occupancy upstream (%)	1.32	0.67	1.45	3.75	7.86	11.81	10.41	1.57	5.67
St.deviation of Occupancy downstream (%)	2.82	0.76	1.03	1.18	3.99	1.44	9.55	6.69	4.15

Table 6-2: Clustering centres for different traffic states (non-accident cases).

Traffic state 1: A relatively low percentage of total cases (12.02%) is assigned in cluster 1 (traffic state 1). In traffic state 1, there is a difference between upstream and downstream average occupancy (10.81% and 14.15% respectively). Both traffic conditions are generally in a free flow state. The standard deviation is low in both loop detectors, however a difference is observed.

Traffic state 2: Almost 1/5 of cases (21.38%) were classified in traffic state 2. In this state, especially low and homogenous values of occupancy measurements can be observed (both lower than 5.5%). The standard deviation of occupancies is also very low.

Traffic state 3: 17.43% of total cases were assigned to cluster 3. The traffic characteristics of this traffic state, are quite opposite to these of traffic state 1. This state presents a situation where there is a decrease in occupancy from upstream to downstream. However the occupancy variation is quite low and similar.

Traffic state 4: 11.30% of total cases belong to cluster 4. As shown in table 6-2, the clustering centre for upstream loop detector is 23.52%, while the respective centre for downstream detector is 11.24%. Thus, there is a great difference in traffic occupancy between upstream and downstream loop detectors, implying a transition from congestion to much better traffic conditions. A difference in occupancy variation can also be observed, where the upstream detector has the higher.

Traffic state 5: The traffic conditions in this state could be characterized as opposite to these of traffic state 1. In this state, both occupancies are relatively low, however, occupancy upstream is higher (almost 19%) than downstream (almost 15%). It is interesting, that the upstream occupancy faces a relatively high variation, having a high clustering centre for standard deviation (7.86%), while the standard deviation downstream is almost 4%. About 10% of cases belong to this traffic state.

Traffic state 6: Despite consisting only of 1.90% of cases, this traffic state might have some of the most interesting characteristics. There is a very high difference in occupancy between upstream (30.82%) and downstream (only 6.07%) loop detectors, meaning the existence of a transition of very high congestion to a very high level of service. The clustering centre for upstream standard deviation of occupancy is also higher.

Traffic state 7: Traffic state 7 consists of less than 7% of total cases. Both occupancy measured from downstream and upstream are considered very high (34.78% and 28.59%) indicating very low level of service and congestion. The clustering centres for standard deviations can be considered homogenous as well (9.55% for downstream and 10.41% for upstream).

Traffic state 8: 9.49% of the sample is assigned to traffic state. The traffic characteristics of this cluster, are opposite to that of cluster 4 (traffic state 4). As shown in table 6-2, the clustering centres of average occupancy and standard deviation of occupancy upstream are 12.87% and 1.57% respectively, while those at downstream loop detector are 27.9% and 6.69% respectively. Therefore, a transition from a good level of service to high congestion can be observed.

Traffic state 9: The attributes of this traffic state are opposite to those of traffic state 7. Both occupancy measurements of upstream and downstream loop detectors are high (29.74% and 24.53% respectively), indicating traffic congestion, however the clustering centres for standard deviation of occupancies at this state are significantly lower than these observed in traffic state 7, indicating more stable conditions.

6.4.2 Finite mixture discriminant analysis

The cluster analysis revealed groups of traffic states, firstly for the accident cases and then for the total cases (accident and non-accident cases). However, new observations cannot be directly assigned to previously pre-defined states. For that reason a discriminant analysis. Each dataset was randomly divided into a training and testing set. Each training set accounted for the 80% of each sample, while each testing set accounted for the rest 20% of each sample. The training sets were used for calibrating the models, while the testing sets were used to validate the models and test the accuracy for identifying traffic states and assigning new observations to the traffic states.

6.4.2.1 Accident cases

The finite mixture discriminant analysis was firstly performed on the accident cases. The accuracy of predicted traffic state memberships for the 20% testing set is presented on Table 6-3. The 80.21% of validation cases can be correctly classified. The lowest classification accuracy was observed for traffic state 5 (50%), however all other traffic states give significantly higher classification accuracies. Especially classification accuracy for traffic states 1, 3 and 4 was substantially higher, reaching 82.14%, 100% and 89.4% respectively.

Traffic states	Predicted group membership using discriminant analysis				
	State 1 (%)	State 2 (%)	State 3 (%)	State 4 (%)	State 5 (%)
State 1	**82.14%**	17.86%	0.00%	0.00%	0.00%
State 2	25.00%	**75.00%**	0.00%	0.00%	0.00%
State 3	0.00%	0.00%	**100.00%**	0.00%	0.00%
State 4	10.53%	0.00%	0.00%	**89.47%**	0.00%
State 5	25.00%	0.00%	16.67%	8.33%	**50.00%**

Table 6-3: Validation results of discriminant analysis (accident cases).

6.4.2.2 Accident and non-accident cases

The finite mixture discriminant analysis on the total cases (accident and non-accident) performed slightly worse and could classify the 76.58% of total validation cases. The accuracy of predicted traffic state memberships for the 20% testing set is presented on Table 6-4. The lowest classification accuracy was observed for traffic state 3 (60.71%), while some substantially higher accuracies were observed, for example for traffic states 2 (91.04%), 6 (100%) and 8 (92.31%).

Traffic states	Predicted group membership using discriminant analysis								
	State 1 (%)	State 2 (%)	State 3 (%)	State 4 (%)	State 5 (%)	State 6 (%)	State 7 (%)	State 8 (%)	State 9 (%)
State 1	**69.23%**	0.00%	30.77%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State 2	0.00%	**91.04%**	8.96%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State 3	1.79%	37.50%	**60.71%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State 4	0.00%	0.00%	20.59%	**76.47%**	2.94%	0.00%	0.00%	0.00%	0.00%
State 5	7.14%	0.00%	10.71%	0.00%	**67.86%**	7.14%	0.00%	7.14%	0.00%
State 6	0.00%	0.00%	0.00%	0.00%	0.00%	**100.00%**	0.00%	0.00%	0.00%
State 7	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**76.47%**	11.76%	11.76%
State 8	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**92.31%**	0.00%
State 9	0.00%	0.00%	0.00%	7.14%	10.71%	0.00%	3.57%	7.14%	**71.43%**

Table 6-4: Validation results of discriminant analysis (accident and non-accident cases).

6.4.3 Bayesian logit models

All the required Bayesian logit models were developed via WinBUGS (Bayesian inference Using Gibbs Sampling) software (Lunn et al., 2000). The WinBUGS is a free and open-source software which is popular for analysing complex statistical models using Markov chain Monte Carlo (MCMC) methods. It applies the Gibbs sampling (Geman and Geman 1984) and the Metropolis algorithm (Metropolis et al., 1953) in order to generate a Markov chain by sampling from full conditional distributions. The DIC, a Bayesian generalization of AIC, is used as a measure of model fit (Spiegelhalter et al., 2003). In general, when comparing models, the lower DIC indicates a better model.

6.4.3.1 Effect of traffic states on accident severity

Using finite mixture cluster analysis, traffic data in accident cases were separated into 5 different traffic states on the basis of average and standard deviation of occupancy upstream and downstream of the accident location. A Bayesian logistic regression model was developed to examine the relationship between traffic states and accident severity. Other variables such as traffic flow and speed were not included in the model because of the potential correlation with the traffic states, following the approach of previous studies (Xu et al., 2012).

The priors for the constant and for the independent variables were all "vague" (non-informative), assuming to follow a normal distribution with zero mean and very low precision. The prior for the constant was $alpha \sim dnorm(0, 0.0001)$. All categories of the independent variable "traffic state", were following the exact same non-informative normal distribution, e.g. for traffic state 2, $beta1 \sim dnorm(0, 0.0001)$. The first 1,000 samples were discarded as adaptation and burn-in. Three chains and 5,000 more samples were used to ensure convergence. Aside from visual inspection of the chains, the Monte Carlo (MC) errors (i.e. the Monte Carlo standard error of the mean values) were also monitored. According to Spiegelhalter et al. (2003), MC errors less than 0.05 indicate that convergence may have been achieved. In the model all MC errors were very low (less than 0.005) indicating convergence.

Table 6-5 summarizes the findings of the Bayesian logistic regression model for accident severity, and provides the estimates of beta coefficients, the standard deviation and the 95% credible interval (2.5%-97.5%) and the odds ratios (OR). Only statistical significant parameters are illustrated on the table. The value of the DIC of the model was 331.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-2.589	0.357	0.075	-3.353	-1.936
traffic state 1 (ref)	-	-	-	-	-
traffic state 2	1.377	0.419	3.963	0.5839	2.242
DIC	331				

Table 6-5: Significant parameters estimates, credible intervals and odds ratios for accident severity model.

Traffic state 1 was used as a reference category and only traffic state 2 was found to be significant because the 95% credible interval of the beta coefficient does not contain zero (0.5839-2.242). Other traffic states, do not provide a credibly nonzero predictiveness for accident severity, because zero was well among the credible interval for the posterior. The positive sign of the mean value of the parameter of traffic state 2, means that traffic state 2 is associated with higher accident severity risk than traffic state 1. More specifically, the odds ratio of 3.963 indicates, that the odds of an accident being severe or fatal in traffic state 2 is almost 4 times higher than for traffic state 1. Both traffic conditions in upstream and downstream loop detectors have a similar and low variation in occupancy. Since traffic state 1 indicates more congested traffic conditions, less congestion is associated with higher severity levels. This finding might be considered consistent with findings of similar studies in the past (Martinm 2002; Quddus et al., 2009; Christoforou et al., 2010), verifying the assumption that under less congestion, drivers tend to drive at higher speeds and therefore more severe accidents may occur.

6.4.3.2 Effect of traffic states on accident probability

Using finite mixture cluster analysis, traffic data in both accident and non-accident cases were separated into 9 different traffic states on the basis of average and standard deviation of occupancy upstream and downstream of the accident location. A Bayesian logistic regression model was developed to examine the relationship between traffic states and accident probability.

The priors for the constant and for the independent variables were all "vague" (non-informative), assuming to follow a normal distribution with zero mean and very low precision. The prior for the constant was $alpha \sim dnorm(0, 0.0001)$. All categories of the independent variable "traffic state", were following the exact same non-informative normal distribution, e.g. for traffic state 2, $beta1 \sim dnorm(0, 0.0001)$. The first 5,000 samples were discarded as adaptation and burn-in. Three chains and 20,000 more samples were used to ensure convergence. In addition, monitoring of the MC errors was performed as previously.

Table 6-6 summarizes the findings of the Bayesian logit model for accident probability and provides the estimates of beta coefficients, the standard deviation and the 95% credible interval CI (2.5%-97.5%) and the odds ratios (OR). Only statistical significant parameters are illustrated on the table. The DIC of the model was 1821.37.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-0.7932	0.1643	0.452	-1.121	-0.4779
traffic state 1 (ref)	-	-	-	-	-
traffic state 5	0.5862	0.2334	1.797	0.1323	1.047
traffic state 6	1.183	0.4332	3.264	0.3482	2.047
DIC	1821.37				

Table 6-6: Significant parameters estimates, credible intervals and odds ratios for accident probability model.

Traffic state 1 was the reference category for this model as well. The mean value of coefficient on traffic state 5, provides a credibly nonzero predictiveness for accident probability, because the 95% credible interval of the beta coefficient does not contain zero (b=0.5862, CI=0.1323-1.047). Furthermore, traffic state 6 was also found to be significant (b=1.183, CI=(0.3482-2.047)). All other traffic states, were not considered significant, because zero was well among the credible interval for the posterior distributions. The positive signs of the mean values of the parameters of traffic states 5 and 6, indicate that these traffic states are associated with higher accident risk than traffic state 1, and thus being considered being hazardous traffic states.

The odds ratio for traffic state 5 is 1.797, meaning that the odds for an accident being severe/fatal for traffic state 5 is almost twice than for traffic state 1. Traffic conditions at this state present a situation quite opposite of that of traffic state 1. In traffic state 5 the occupancy faces a fall (from 18.89% upstream to 14.89% downstream), while in traffic 1 a small increase is present (10.81% to 14.15%), and also the values of occupancy are relatively lower. However, traffic state 6 was found to be associated with even higher accident risk (OR=3.264). These results show that the transition from very high occupancy (30.82%) to very low occupancy (6.07%), is associated with high accident probability. It is also worth noticing, that at this state, a very significant change in standard deviation of occupancy from upstream (11.81%) to downstream (1.44%) is observed, indicating high transition and correlation with risk of accident occurrence.

6.4.3.3 Effect of traffic states on PTW accident probability

The relationship between traffic states and PTW accident probability, was examined through the application of Bayesian logit models. The approach was the same as in previous models.

The priors for the constant and for the independent variables were all "vague" (non-informative), assuming to follow a normal distribution with zero mean and very low precision. The prior for the constant was $alpha \sim dnorm(0, 0.0001)$. All categories of the independent variable "traffic state", were following the exact same non-informative normal distribution, e.g. for traffic state 2, $beta1 \sim dnorm(0, 0.0001)$. The first 1,000 samples were discarded as adaptation and burn-in. Three chains and 20,000 more samples were used to ensure convergence.

Table 6-7 summarizes the findings of the Bayesian logit model for PTW accident probability, and provides the posterior mean, the standard deviation and the 95% credible interval CI (2.5%-97.5%) and the odds ratios (OR). Only statistical significant parameters are illustrated on the table. The DIC of the model was 625.51.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	0.53	0.1864	1.699	0.1679	0.902
traffic state 1 (ref)	-	-	-	-	-
traffic state 2	-0.6467	0.2613	0.524	-1.165	-0.138
traffic state 3	0.8263	0.3671	2.285	0.1274	1.564
DIC	625.51				

Table 6-7: Significant parameters estimates, credible intervals and odds ratios for PTW accident probability model.

Traffic state 1 was the reference category for this model as well. Traffic states 2 (b=-0.6467, CI=(-1.165-0.138)) and 3 (b=0.8263, CI=(0.1274-1.564)) were found to be significantly associated with PTW accident probability. However, traffic state 2 had a negative posterior mean value, meaning that at this traffic state (lower occupancy) the risk of PTW accident involvement is lower. The positive sign of the posterior mean for traffic state 3, shows that a consistent very high occupancy in upstream (21.1%) and downstream (30.16%) loop detectors, and high variation in occupancy upstream (8.52%) is associated with high probability of accidents with a PTW (given the accident occurs). All other traffic states, were not considered significant, because zero was well among the credible interval for the posterior distributions of the beta coefficients.

6.4.3.4 Effect of traffic states on PTW accident severity

Following the same approach as previously, a Bayesian logit model was developed to examine the relationship between traffic states and PTW accident severity. The priors for the constant and for the independent variables were all "vague" (non-informative), assuming to follow a normal distribution with zero mean and very low precision. The prior for the constant was $alpha \sim dnorm(0, 0.0001)$. All categories of the independent variable "traffic state", were following the exact same non-informative normal distribution, e.g. for traffic state 2, $beta1 \sim dnorm(0, 0.0001)$. The first 1,000 samples were discarded as adaptation and burn-in. Three chains and 5,000 more samples were used to ensure convergence. Table 6-8 illustrates the findings of the Bayesian model for PTW accident severity, and provides the estimates of posterior mean, the standard deviation, the odds ratios and the 95% credible interval (2.5%-97.5%). The DIC of the model was 176.52.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-2.561	0.4546	0.077	-3.545	-1.753
traffic state 1 (ref)	-	-	-	-	-
traffic state 2	1.095	0.5702	2.989	0.01806	2.257
DIC	176.52				

Table 6-8: Significant parameters estimates, credible intervals and odds ratios for PTW accident severity model.

The summary of the results of PTW accident severity (accidents where a PTW is involved), were found to be similar to the findings of section 6.4.3.1, when accident severity was analysed. Traffic state 1 was set as the reference category for the independent variable, and only traffic state 2 was found to be significant because the 95% credible interval does not contain zero (0.01806-2.257). The other traffic states were not found to be statistically significant. The positive sign of the mean value of the posterior parameter of traffic state 2, means that traffic state 2 is associated with higher accident PTW severity risk than traffic state 1 (OR=2.989).

6.5 Summary

This chapter of the thesis investigated the influence of traffic states on road safety on urban roads by using high resolution traffic data. If not the first, then it is one of the first implementations of such approach and modelling that also considers urban roads. More specifically, finite mixture cluster analysis was performed on the basis of traffic average and standard deviation of occupancy measured at the two nearby loop detectors, in order to classify urban traffic conditions into different states (5 clusters for accident cases, 9 clusters for accident and non-accident cases). Then a finite mixture discriminant analysis was conducted to identify and predict traffic state membership, by using a validation dataset (20% of each sample), given the presence of that real-time urban traffic occupancy data. Lastly, Bayesian logit models were applied to unveil the influence of different traffic conditions (states) on accident severity and accident probability, having an emphasis on PTWs as well.

The findings of this section demonstrates that this approach is promising when applied on urban roads. When analysing accident severity, traffic state 2 was found to have the greatest influence, correlating high accident severity with lower occupancy levels and therefore less congestion. Concerning the probability of accident occurrence, higher occupancy was found to lead to an increased risk of accident occurrence. Consequently, it can be concluded that higher occupancy and congestion contribute to higher accident occurrence but also result to less severe accidents.

The traffic condition that was identified as the most hazardous for accident occurrence was the traffic state when the transition from very high to very low occupancy took place. This finding is consistent with the results of Hossain and Muromachi (2013), who argue that a fast moving uncongested downstream when being followed by slow moving and congested upstream may be more risky. One possible reason for this finding may be the fact that drivers may compensate for travel time loss and consequently accelerate (Hossain and Muromachi, 2013).

The findings of the statistical modelling of PTW accident severity, were very similar to that of accident severity, suggesting that uncongested traffic is correlated with more severe accidents. It was also found that PTWs are more likely to be involved in accidents when the occupancy is very high (congestion). On the contrary, in very low levels of occupancy (traffic state 2), PTWs are less likely to be involved in accidents, indicating that congestion is correlated with increased probability of PTW involvement in accidents. Consequently, traffic state 2 is associated with higher accident severity for both accidents and accidents with PTWs, given that an accident has occurred, but is correlated with lower PTW accident probability.

Chapter 7 Multi-statistical investigation of road safety in urban roads

7.1 Introduction

The effective treatment of road accidents and thus the enhancement of road safety is a major concern to societies due to the losses in human lives and the economic and social costs. Tremendous efforts have been dedicated by transportation researchers and practitioners to improve road safety. Recently, real-time data were utilized when analysing road safety in freeways. An increasing number of studies (Oh et al., 2001; Lee et al., 2003; Zheng et al., 2010; Abdel-Aty et al., 2012; Xu et al., 2013b) have developed freeway accident risk models in order to model accident probability (or crash likelihood as frequently stated in literature) and accident severity (Christoforou et al., 2010; Yu and Abdel-Aty, 2014a, 2014b).

However, although freeway safety has been extensively explored, only freeways and urban expressways have been considered. Very little research has been done on urban roads and arterials so far (Yannis et al., 2014). Moreover, there is no specific focus on Powered-Two-Wheelers (PTWs) which constitute a vulnerable type of road users and therefore the investigation of the impact of traffic parameters could be of great importance. Taking also into account the speeding and the manoeuvring capabilities of PTWs, investigation of PTW safety by incorporating traffic conditions would be of particular interest. These gaps in knowledge have been discussed in the detailed literature review in Chapter 2 as well.

Consequently, the aim of this chapter is to analyse traffic safety in urban roads by incorporating high resolution traffic and weather data. More specifically, it is aimed to explore accident probability, accident severity and PTW accident probability by applying a series of advanced statistical models. Firstly, data mining techniques such as the Random Forests (RF) are applied to rank variable importance and consequently provide a first insight on the significant variables. Then, having acquired information from the Random Forest models, a series of advanced logit models such as finite mixture models, mixed effects models and Bayesian logistic models are applied.

The results of Chapter 7 aim to provide an insight on accident probability and severity on urban roads and add to the current knowledge, by including high resolution traffic and weather data and by considering PTWs as well.

7.2 Methodology

The statistical methods that applied to achieve the aims of this chapter, are described in the following subsections. The first task was to develop random forest model in order to rank variable importance, which would allow the use of the most vital variables in the modelling procedures that follow. The final step was to develop advanced logit models by using the information acquired by the data mining achieved through the Random Forests.

7.2.1 Classification techniques

7.2.1.1 Random forests

Random Forests (RF) have been frequently used in traffic safety studies (Abdel-Aty and Haleem, 2011; Ahmed and Abdel-Aty, 2012; Yu and Abdel-Aty, 2014b). All these studies aimed to rank the variable importance and thus select the appropriate variables before applying other statistical models.

A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1,\ldots\}$, where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x (Breiman, 2001). Breiman (2000), argue that random forest models are different from other classification and regression tree models, in a sense that RF models are capable of providing unbiased error estimates and does not require a cross-validation dataset. Another advantage of RF models is accuracy and stability (e.g. if data are slightly changed the individual trees may change but the forest is relatively stable since it is a combination of many trees). Strobl and Zeileis, (2008) argue that in order to construct a random forest, a number of bootstrap samples from original sample have to be drawn, and then a classification tree to each bootstrap sample has to be fitted (number of trees).

The variable importance as unveiled from the RF models are extremely helpful to define which variables are significant. However, the magnitude of the effect and the sign of each variable are not identified. It should be noted that variable importance should be interpreted as a relative ranking of predictors, since the absolute values of the importance scores should not be interpreted or compared over different studies (Strobl, et al., 2009).

7.2.2 Logit models

A series of logit models were applied in this chapter to explore accident severity, accident probability and PTW accident probability. In order to account for the unobserved heterogeneity as suggested by literature (Savolainen et al., 2011) accident severity was explored through finite mixture and mixed effects logistic regression, while Bayesian logistic regression models were applied for accident probability and PTW accident probability. However, accident probability was also explored with finite mixture logistic regression.

7.2.2.1 Finite mixture logit

This logit modelling approach is an extension of the binary logit model based on a finite mixture approach in which the unobserved heterogeneity is accounted for via latent classes. The finite mixture logit analysis is also referred as Latent Class analysis.

The Finite Mixture Logit (FML) model can be considered as an extension of the standard binary logit model which also includes a latent class model that captures the effect of unobserved variables on the binary outcome variable. The response variable is Y and takes the values y = 0 (non-event) and y = 1 (event). Holm et al. (2009) suggest the FML model with J (j = 1,..., J) latent classes as:

$$P(Y=1|x) = \sum_{j=1}^{j=J} P(Y=1|x, \Xi=\varepsilon_j)P(\Xi=\varepsilon_j) = \sum_{j=1}^{j=J} \frac{\exp(\alpha+\beta x+\varepsilon_j)P(\Xi=\varepsilon_j)}{1+\exp(\alpha+\beta x+\varepsilon_j)}, \quad \text{(Eq. 7-1)}$$

where α is a constant term, x is a vector of independent variables, β is a corresponding row vector of regression coefficients, ε_j is the effect of the j'th latent class on the probability of observing Y =1, and P(Ξ=ε_j) is the proportion of the population that belongs to the j'th latent class. According to Holm et al. (2008), the log-likelihood function lnL for n independent observations can be defined as:

$$lnL = \sum_{i=1}^{i=n} lnP(Y=1|\,x_i) + (1-y_i)\ln(1-P(Y=1|x_i)) \quad \text{(Eq.7-2)}$$

where,

$$P(Y=1|x_i) = pP_{0i} + (1-p)P_{ie}, \quad \text{(Eq.7-3)}$$

and

$$P_{0i} = \frac{\exp(a+bx_i)}{1+\exp(a+bx_i)} \quad \text{(Eq.7-4)}$$

$$P_{ie} = \frac{\exp(a+bx_i+\varepsilon)}{1+\exp(a+bx_i+\varepsilon)} \quad \text{(Eq. 7-5)}$$

where P(Ξ = 0) = p and P(Ξ = e) = $1-p$.

In finite mixture formulation, parameter heterogeneity across observations is explored with a discrete distribution or set of classes (Greene and Hensher, 2003; Shaheed and Gkritza, 2014). Finite mixture analysis divides the sample into distinct classes with homogenous attributes. As Greene and Hensher, (2003) correctly state, an important issue is the determination of the number of classes. Roeder et al. (1999) suggest the implementation of the Bayesian information criterion (BIC):

$$BIC(model) = lnL + \frac{(model\ size)lnN}{N} \quad \text{(Eq.7-6)}$$

where lnL is the log-likelihood for the sample.

7.2.2.2 Mixed effects logit

In this chapter, mixed effects logit models were also applied to explore accident severity. Mixed effects logit model enable the incorporation of both "fixed effects" and "random effects". Gelman (2005), provided clear definition (Gelman, 2005, page 21): *"We define effects (or coefficients) in a multilevel model as constant if they are identical for all groups in a population and varying if they are allowed to differ from group to group".*

Generalized linear models (GLMs) are different from linear models if a link function between the predicted mean μ and the linear predictor u is specified:

$$u = a + \beta_1 X_1 + \cdots + \beta_n X_n, \qquad \text{(Eq.7-7)}$$

and also an error distribution function is specified. The error distribution function specifies the probability distribution around the predicted mean and can be normal, Poisson or Binomial. In multi-level generalized linear models a stochastic component is added to the linear predictor:

$$\mathrm{l}(\mu) = \mathrm{u} = a + \beta_1 X_1 + \cdots + \beta_n X_n + b_o + b_1 X_1 + \cdots + b_{\mathrm{m}} \mathrm{Z}_{\mathrm{m}}, \qquad \text{(Eq.7-8)}$$

where the random effects vector $\vec{b}$ is assumed to be normally distributed with a mean of zero and variance-covariance matrix Σ. In mixed effects logistic regression model, the stochastic component is assigned to the binomial error function:

$$P(y;\mu) = \binom{n}{yn} \mu^{yn} (1-\mu)^{(1-y)n} \qquad \text{(Eq. 7-9)}$$

and the logit function applies:

$$u = log \frac{\mu}{1-\mu} \qquad \text{(Eq. 7-10)}$$

and

$$\mu = \frac{e^u}{e^u + 1}. \qquad \text{(Eq. 7-11)}$$

7.2.2.3 Bayesian logit

In this chapter, Bayesian logistic regression models were developed to estimate the effect traffic, weather and other characteristics on accident probability and PTW accident probability. The theoretical background of Bayesian logit models was demonstrated previously in the thesis in Chapter 6 (section 6.2.3). The procedure was the same, and "non-informative" priors were used for the needs of the modelling of this chapter as well.

The DIC as the Bayesian generalization of AIC (Akaike information criterion), is a measure of model fit and the effective number of parameters. Spiegelhalter et al.

(2003), argue that differences of more than 10 might definitely rule out the model with higher DIC.

7.3 Data preparation

In this chapter, accident, traffic and weather data from the Kifisias and Mesogeion avenues were used to explore accident probability and severity. In addition, Powered-Two-Wheelers (PTWs) were considered as well. The final urban dataset for this chapter, included 527 accidents and 1581 non-accident cases for Kifisias and Mesogeion avenues from 2006 to 2011. Accident severity had two levels, namely fatal-severe injury (KSI) and slight injury (SI). A percentage of 10.8% of accidents were classified as severe (KSI), while 89.2% were classified as slight (SI). PTWs were involved in 326 of those accidents (61.9% of the accidents). Consequently, this new variable is a binary (1 for PTW accident involvement, 0 when no PTWs are involved in an accident) and expresses whether a PTW is involved in an accident-given the accident has already occurred.

For the needs of the analysis, a new variable was also created, namely accident occurrence, which was also a binary variable (1 for accident occurrence, 0 for non-accident occurrence). More specifically, when exploring accident probability, for each accident observation of the dataset, there were 2 observations for non-accidents for the same time and location, one week before and one week after the accident occurrence, resulting in 527 accident cases and 1054 non-accident cases (1581 total cases).

Traffic data from the closest upstream loop detector were considered. As stated in Chapter 4 of the thesis, the 5-min traffic data were further aggregated to 1-hour level to obtain averages, standard deviations and so on, prior to an accident occurrence. It was anticipated that the 60-min traffic data before accident occurrence would cover the hazardous traffic conditions leading to accident occurrence, and consequently only the traffic data 1-hour prior to accident occurrence were initially considered. The same time slice (1-hour prior to an accident) was considered for non-accident cases as well.

Weather records for each meteorological station covered the whole period from 2006-2011. For that reason, each accident case had to be assigned to the closest meteorological station and then the relevant weather data had to be extracted. Using advanced programming in Excel, the 10-min raw data were aggregated in order to obtain maxima, averages and standard deviations, in the time-slice of 1-hour prior to the time of the accident occurrence. Regarding rainfall, the sum and standard deviation of rainfall has also been calculated for 1h, 2h, 6h and 12h prior to the time of the accident. Moreover, the *weather* variable (good/adverse weather) and the pavement condition (good/wet) as originally coded in SANTRA database system were also considered.

The non-accident cases (used for modelling accident probability), were treated by following the exact same procedure as for traffic data. For example, an accident occurred in Kifisias Avenue on Wednesday 12 August 2009 at 13:00, 12 hours of weather data have been extracted from Galatsi Station, for Wednesday 5 August 2009 and Wednesday 19 August 2009 for the period 01:00-13:00. Lastly, the accident data were combined with the traffic and weather data in order to construct the required dataset for the statistical analysis.

7.4 Results

In this thesis all the RF analyses were conducted via the *party* package in R (Hothorn et al., 2010). This package can also handle correlated predictor variables (Strobl and Zeileis, 2008). Each respective dataset was randomly split into a training (80%) and a testing set (20%). The RF models were applied on the training set. According to Strobl et al. (2009a), insignificant variables vary randomly around zero. Additionally, a correlation matrix of the variables entered in the final models has been checked to avoid multicollinearity problems. The finite mixture logit analysis was conducted via the *flexmix* package in R (Gruen and Leisch, 2008), while the mixed effects logistic regression was performed in *lme4* package (Bates et al., 2015).

7.4.1 Accident severity

By using Random Forests, the variable importance ranking was unveiled by entering mtry = 7 meaning that seven variables were randomly sampled as candidates for each split. As Strobl et al. (2009b) suggest, the number of *mtry* should equal to the square root of the number of candidate variables. For that reason in this analyses a large number of *mtry* was chosen. Furthermore, totally 1000 trees were constructed since a suitably large number of trees will guarantee more stable and robust results.

Figure 7-1 shows the final results of variable importance rankings. Predictors to the right of dashed vertical line are identified to be significant. This vertical line on the plot is set at the value of the lowest important predictor. It can be drawn from the figure that the 1-hour average flow upstream (Q_avg_1h_up), accident type (Acc.type), the 1-hour coefficient of variation of flow upstream (Q_cv_1h_up), the 1-h average speed upstream (V_avg_1h_up) as well as the 1-hour coefficient of variation of speed upstream (V_cv_1h_up) are the most important variables. Hence, these variables had to be further explored as inputs in the modelling steps. However, it was obvious that some of these variables were highly correlated, consequently the correlation matrix had to be constructed before entering the logit models. For example, average speed and coefficient of variation of speed cannot simultaneously be considered since they are correlated (r=-0.59). The same is for average flow and coefficient of variation of flow (r=-0.47).

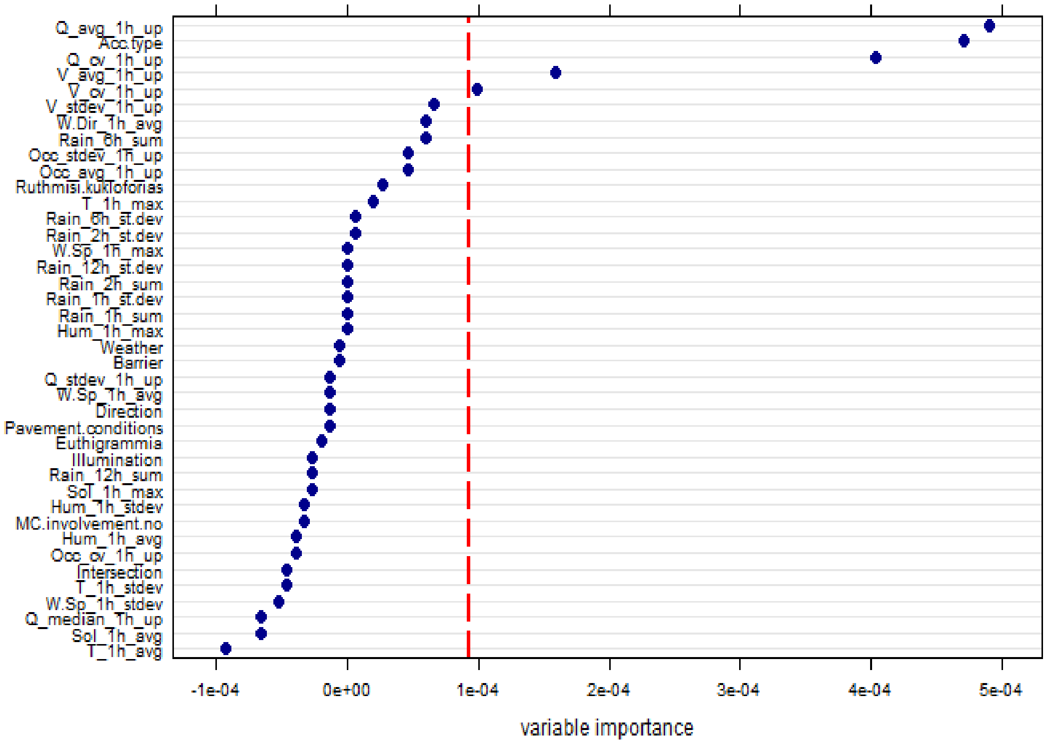


Figure 7-1: Accident severity variable importance provided by random forest.

The most significant variables as identified by the Random Forest model, were tested firstly on the finite mixture logit model and then on the mixed effects logit model. The finite mixture analysis revealed that two distinct classes with homogenous attributes were found significant: latent class 1 with probability 47.27% and latent class 2 with probability 52.73%. These are the probabilities for each accident case to belong in each latent class respectively. The optimum number of latent classes was determined according to the BIC criterion, with lower BIC indicating best models.

Table 7-1 shows the finite mixture logit estimation results. A total of four variables were found to be significant (accident type, coefficient of variation of flow, coefficient of variation of speed and involvement of PTW), the likelihood of the final model was -151.328 and the McFadden R-square was considered adequate with a value of 0.197, since it is suggested that values between 0.2 and 0.4 are of very good fit. Moreover, the likelihood ratio test between the null and the full model is considered significant. Despite the fact the RF model did not figure potential involvement of a PTW as a significant predictor, its effect was tested. It is noted that correlation of V_cv_1h_up (coefficient of variation of flow) and Q_cv_1h_up (coefficient of variation of speed) was low (r=0.31) and therefore they could be simultaneously considered in the model. Only accident type was set as fixed across latent classes, while the constant, the coefficient of variation of flow, the coefficient of variation of speed, the involvement of PTW were set as random and free to vary across the two latent classes, in order to account for unobserved heterogeneity.

Variables	Latent Class 1			Latent Class 2		
	Mean	t-statistic	p-value	Mean	t-statistic	p-value
Constant (random)	-1.823	-1.904	0.057	-0.402	-0.330	0.741
Acc.type0 (reference cat.)	-	-	-	-	-	-
Acc.type1 (fixed)	-1.099	-1.539	0.124	-1.099	-1.539	0.124
Acc.type2 (fixed)	-1.483	-2.246	0.025	-1.483	-2.246	0.025
Acc.type3 (fixed)	-16.942	-0.014	0.988	-16.942	-0.014	0.988
Acc.type4 (fixed)	-1.152	-2.264	0.023	-1.152	-2.264	0.023
Q_cv_1h_up (random)	10.936	1.977	0.048	-5.142	-0.619	0.536
MC.involvement.no (random)	-1.359	-1.733	0.083	1.289	1.312	0.190
V_cv_1h_up (random)	-0.387	-0.165	0.869	-32.089	-2.090	0.037
Log-likelihood at zero	-188.400					
Final Log-likelihood	-151.328					
Likelihood ratio test	74.144					
McFadden R^2	0.197					

Table 7-1: Summary of the finite mixture logit model for accident severity.

The results showed that aside from accident type which has a consistent effect as it was set to be fixed across classes, the other variables either have opposite signs across the two classes, or are significant only in one of them, suggesting that there is heterogeneity between the two classes.

Consequently, elasticity analysis (marginal effects) is suggested in order to gain a better understanding of the effect of variables that are not fixed across latent classes (Shaheed and Gritza, 2014; Xie et al., 2012; Cerwik et al., 2014). The elasticities were calculated for each class using the same method as that of conventional logit models (Train, 2003). For continuous variables, the elasticity shows the change in probability for 1% change in the continuous variable. The equations for the elasticity are not duplicated since they have been presented in various studies (Shankar and Mannering, 1996; Ulfarsson and Mannering, 2004). For discrete variables, the pseudo-elasticity expresses the simple difference of the two probabilities, with and without the variable (Shaheed and Gritza, 2014). The final elasticity of a variable for the finite mixture model, is the summation of the marginal effects for each class weighted by their posterior latent class probabilities. Table 7-2, demonstrates the elasticities and the pseudo-elasticities for the estimated model.

Variable	e
Acctype2	-0.656
Acctype4	-0.572
V_cv_1h_up	-0.379
Q_cv_1h_up	0.427
MC.involvement.no	-0.282

Table 7-2: Average elasticity and pseudo-elasticity of each variable for the finite mixture model for accident severity.

The type of accident (Acc.type) has a consistent effect among the two classes. Rear-end (Acc.type2) and sideswipe (Acc.type4) have negative beta coefficients, meaning that these collisions lead to less severe severity outcomes than collisions with fixed object or run-off-road collisions (Acc.type0). On the other hand, head-on collisions (Acc.type1) and side collisions (Acc.type3) were not found to be statistically significant. The overall results of the effect of accident type on accident severity inside urban roads, are consistent with Theofilatos et al. (2012), who explored accident severity inside and outside urban areas in Greece.

Regarding the influence of the other variables, the results between the two classes generally differ. First of all only in latent class 1 is the constant statistically significant. The impact of the coefficient of variation of flow (Q_cv_1h_up) is positive on class 1, indicating that increased flow variations lead to more severe accidents, but has no effect on class 2. The positive correlation of flow variations and severity may be consistent with past studies (Yu and Abdel-Aty, 2014a and 2014b), however, the insignificance in class 2 indicates presence of heterogeneity. To solve this problem, the real effect of the coefficient of variation of flow is achieved with the interpretation of the elasticity (0.427), suggesting a positive correlation.

On the other hand, speed variation is not significant on class 1, but it is on class 2. The most interesting fact is that the beta coefficient of the coefficient of variation of speed (V_cv_1h_up), has a negative sign, implying that in class 2 traffic variations as expressed by coefficient of variation of speed, are more likely to lead to less severe accidents. However, it is not statistically significant for observations in class 1. Therefore, the final effect is given by the elasticity (-0.379). One explanation for the negative effect, could be the potential existence of traffic oscillations in low speeds, resulting in occurrence of non-severe crashes. As such, it is interesting that variations in traffic seem to have both positive and negative correlation with severity.

Lastly, involvement of PTWs in accidents has a negative beta coefficient for class 1 but it is not statistically significant for class 2. The negative elasticity suggest a negative effect on the probability of severe/fatal accidents. Therefore, PTWs are more likely to be involved in non-severe accidents. This may seem counterintuitive, but the fact that the effect is significant only for cases of class 1, indicates the presence of heterogeneity. This variable is significant for 90% level of confidence only, consequently the negative effect cannot be supported with confidence. That is probably the reason for not having been detected significant in the Random Forest model.

Summing up, the results of the finite mixture logit analysis has shown that the impact of a number of exogenous factors differ in sign and significance, and consequently there is a need for segmentation of accidents (Shaheed et al., 2014).

Next, based upon the random forest model, a mixed effects logit model was developed to model accident severity. The results of the mixed model are summarized on Table 7-3. The model diagnostics showed an adequate fit but worse than the finite mixture model. The only fixed effect defined for the model was standard deviation of occupancy. The negative beta coefficient of the standard deviation of occupancy means that as occupancy variations increase, there is increased probability of having non-severe accidents. This finding may not be consistent with past studies, however, this contradictory finding might be attributed to the fact that the study area consists of urban roads which frequently face more severe congestion than freeways. Moreover, the existence of traffic lights may cause variations in traffic flow. As a result, occupancy variations as expressed in the model, indicate stop-and-go conditions with probably low speeds and consequently lower severities.

After many tests, two random effects were defined. Firstly, the variability among accident types was tested in terms of the coefficient of variation of flow, simply assessing if the effect of the coefficient of variation of flow differs among accident type. Then the variability of PTW accidents (variable MC.involvement.no) was tested, in terms of the constant, by simply assigning a different constant. Thus, accident with PTWs have a different constant than accident without a PTW.

The first random effect (coefficient of variation of flow per accident type). The variance of the first random effect (Q_cv_1h_up|Acc.type) has the largest share of the variance (5.553). The second random effect (constant|MC.involvement.no) has less variability, with a variance of 2.3. Furthermore, ANOVA tests were performed and shown that both random effects were significant with respective p-values significantly lower than 0.05 ($0.0001337<0.05$ and $0.000000022<0.05$ respectively), and as consequence their involvement in the final model did make sense.

Fixed Effects	Mean	t-statistic	p-value
Occ_stdev_1h_up	-0.07867	-2.074	0.038
Random Effects			
Groups	Variable	Variance	
Acc.type	Q_cv_1h_up	5.553	
MC.involvement.no	Constant	2.3	
Log-likelihood at zero		-188.370	
Final Log-likelihood		-166.700	
Likelihood ratio test		43.340	
McFadden R^2		0.115	

Table 7-3: Summary of the mixed effects logit model for accident severity.

Table 7-4 illustrates the estimates of the coefficients of the random effects. When examining the coefficients of the random effects, it is observed that the effect of the constant on the variable MC.involvement.no is always negative. This means that despite individual variation, there is consistency.

Concerning the effect of the coefficient of variation of flow, higher variability is observed, as it was expected from the high variance value. The two positive coefficients (6.781 and 3.015), mean that high variations in flow increase severity when run-off-road and collisions with fixed objects occur (Acc.type0) and when rear-end crashes occur (Acc.type2). In all other accident types (side, sideswipes, head-on collisions), flow variation is more likely to reduce accident severity. The existence of heterogeneity is very probable, since in the previous finite mixture logit the impact of traffic variation was not consistent.

Coefficients of Random Effects	
	Q_cv_1h_up
Acc.type0	6.781
Acc.type1	-1.126
Acc.type2	3.015
Acc.type3	-6.110
Acc.type4	-0.146
	Constant
MC.involvement.no (0)	-2.007
MC.involvement.no (1)	-2.486

Table 7-4: Coefficients of the random effects.

7.4.2 Accident probability

For accident probability, the variable importance ranking was explored by using mtry = 6 meaning that seven variables were randomly sampled as candidates for each split. In addition, 1000 trees were constructed.

Figure 7-2 demonstrates the variable importance rankings. Predictors to the right of dashed vertical line are identified to be significant. As previously, the vertical line on the plot is set at the value of the lowest important predictor.

It can be drawn from the figure that the 1-hour coefficient of variation of flow upstream (Q_cv_1h_up), the 1-h standard deviation of occupancy upstream (Occ_stdev_1h_up), the 1-h standard deviation of speed upstream (V_stdev_1h_up), the 1-hour coefficient of variation of speed upstream (V_cv_1h_up) and lastly the 1-hour coefficient of variation of occupancy upstream (Occ_cv_1h_up) are identified as the most important variables. Hence, these variables had to be further explored as inputs in the modelling steps of the Bayesian logit model.

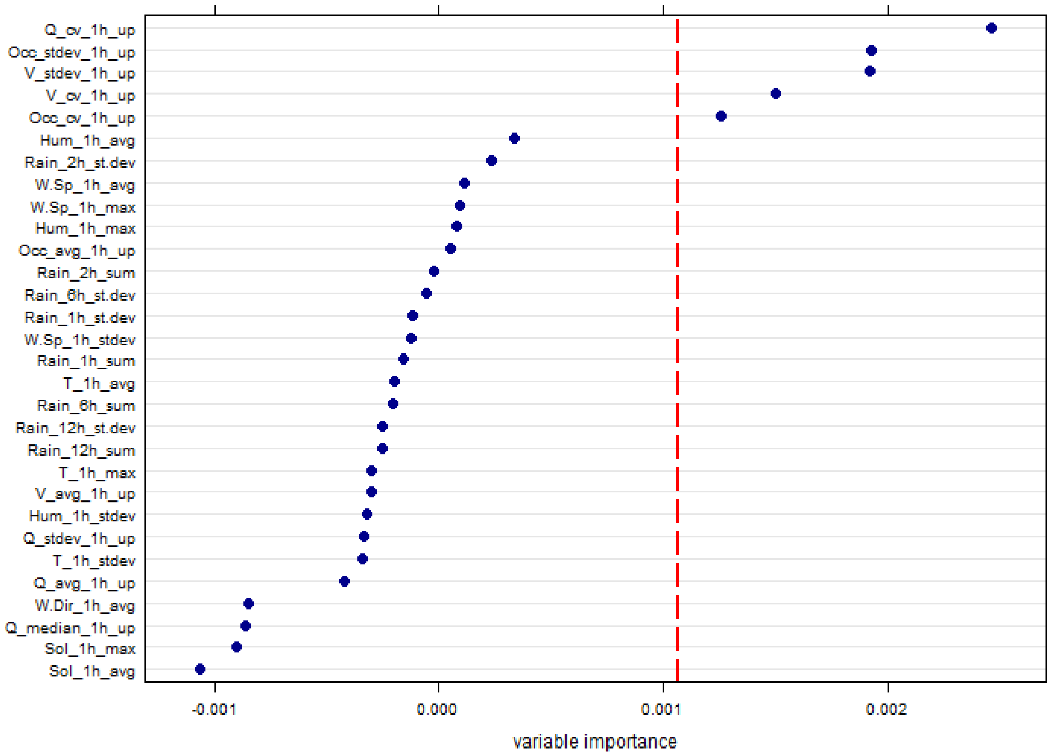


Figure 7-2: Accident probability variable importance provided by random forest.

An important remark concerning the Random Forest models so far, is the consistently high importance of some traffic variables, while weather variables do not seem to explain accident severity and accident probability.

Since most of the traffic variables are high likely to be correlated to each other, several attempts had to be conducted. For example, the standard deviation of occupancy was found to be highly correlated with the standard deviation of speed (r=0.85) and with the coefficient of variation of speed (r=0.92). As a result, specific attention was needed when constructing the logit models.

The next table (7-5) summarized the findings of the final Bayesian logistic regression model. All the prior distributions of the parameters were non-informative, following a normal distribution with zero mean and precision of 0.0001, namely ~dnorm(0, 0.0001). The first 5,000 iterations were discarded and used as burn-in and 3 chains of 20,000 iterations were set up. Regarding the final traffic variables of the Bayes logit model, the standard deviation of occupancy was not correlated with the coefficient of variation of flow (r=0.29), neither with the logarithm of the coefficient of variation of flow (r=0.31).

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant (n.s.)	-0.328	0.2183	-	-0.7425	0.1074
Occ_stdev_1h_up	0.055	0.014	1.057	0.028	0.082
log(Q_cv_1h_up)	0.217	0.081	1.242	0.069	0.380
DIC	1980.64				

Table 7-5: Summary of the Bayesian logit model for accident probability.

The constant in the model were not significant as the 95% credible interval of the beta coefficient of the constant contain zero (-0.7425-0.1074). However, this model had the lowest DIC value (1980.64). Moreover, the DIC was substantially lower than the null model as well. The positive beta coefficient (b=0.055), shows that as the variation in occupancy, expressed by the standard deviation of occupancy increases, the risk of accident occurrence increases as well. The Odds Ratio for the beta coefficient of standard deviation of occupancy, indicates that for 1 unit increase in the standard deviation of occupancy upstream, the odds for accident occurrence are 1.057 times higher than before (5.7% increase in the odds of accident).

Moreover, the model indicates that the logarithm of the coefficient variation of flow increases the risk of an accident as well, as it has a positive beta coefficient (0.217). The Odds Ratio of the beta coefficient (1.242), means that for 1 unit increase in the logarithm of the coefficient variation of flow upstream, there is a 24.2% increase in the odds of an accident occurrence.

These findings are consistent with other studies, which suggest that variations in occupancy have an impact on accident occurrence (Yu and Abdel-Aty, 2013a; Yu et al., 2013; Xu et al., 2013a). In addition, a few studies suggest that speed variations increase risk (Ahmed et al., 2012b). For example, Hassan and Abdel-Aty (2013) found that the coefficient of variation of speed upstream was a main risk factor under clear weather. Consequently, it can be concluded that large variation of occupancy and volume, indicating turbulent and stop-and-go traffic scenarios have a consistent influence on accident occurrence on urban roads as well, as they tend to increase accident risk.

7.4.3 PTW accident probability

When modelling PTW Accident Probability, the variable importance ranking was explored by constructing 1000 trees and using mtry = 7. Figure 7-3 illustrates the variable importance rankings. Predictors to the right of dashed vertical line are identified to be potentially significant. However, another vertical line was set in order to highlight the very highly important predictors (to the right of the blue line).

As previously, the vertical line on the plot is set at the value of the lowest important predictor. Similar to the RF model for accident severity, the type of accident was identified as a considerably highly important predictor as well. Furthermore, the 1-h average speed upstream (V_avg_1h_up), the 1-h average occupancy upstream

(Occ_avg_1h_up), 1-h average flow upstream (Q_avg_1h_up) and lastly the 1-h median flow upstream (Q_median_1h_up) were identified as significant as well. As in previous analyses, the variables had to be checked once more for potential correlations.

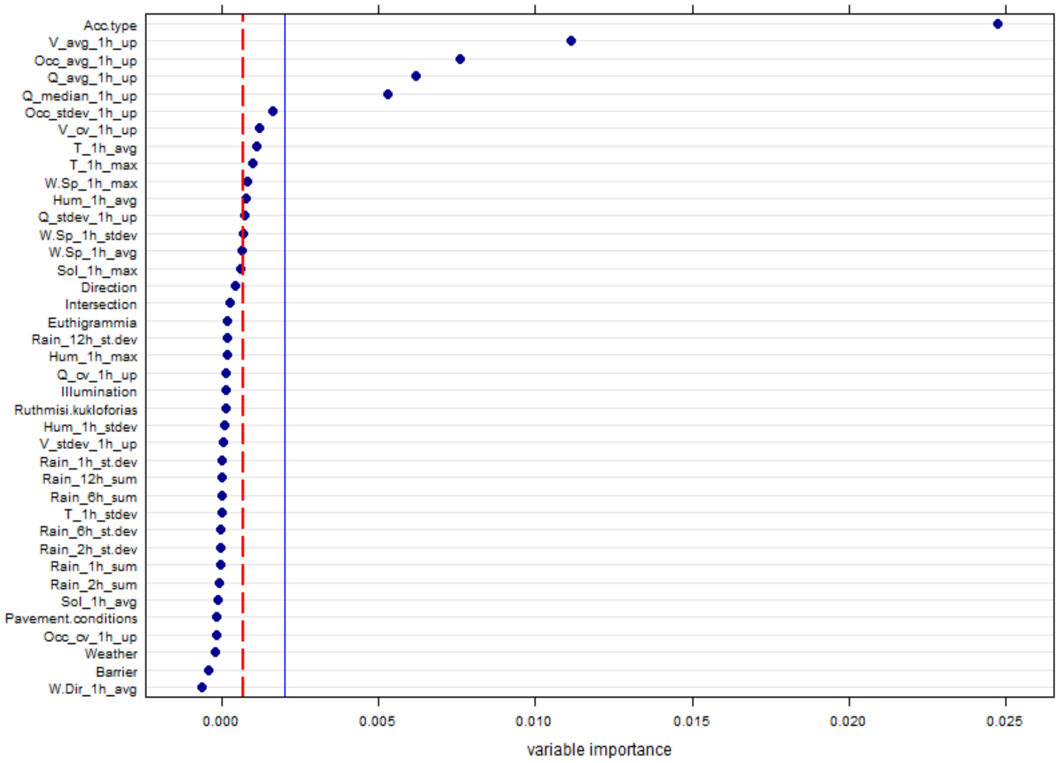


Figure 7-3: PTW accident probability variable importance provided by random forest.

Since most of the traffic variables were highly correlated to each other, several attempts had to be conducted. The median flow although indicated by the RF model as significant, it was found to be correlated with the average flow (r=0.62) and thus removed from the model. Moreover, only the coefficient of variation of speed was retained to the model since it was correlated with standard deviation of occupancy (r=0.91).

The next table (7-6) summarizes the findings of the final Bayesian logistic regression model. This was the model with the lowest DIC value (640.05) and thus having the best fit. All prior distributions of the parameters were chosen to be non-informative, following a normal distribution with zero mean and a very low precision of 0.0001, namely $\sim \mathrm{dnorm}(0, 0.0001)$. The first 5,000 iterations were discarded and used as burn-in and 3 chains of 20,000 iterations were set up. Regarding the two traffic variables in the model, the correlation was very low (r=0.16).

The beta coefficient of average flow has a positive sign, indicating that when flow increases, there is an increase in the probability of a PTW to have been involved in an accident (given the accident occurs), simply meaning that in more congested traffic conditions, PTWs are more likely to be involved in accidents. One explanation could be the existence of increased interaction with other motorized traffic and the potential need for manoeuvres under these dense traffic conditions.

Another interesting finding was that the coefficient of variation of speed (basically expressing variations in mean time speed) increase the probability of an accident involving a PTW. In literature, speed variations have been linked to high risk of accidents on freeways (Zheng et al., 2010; Xu et al., 2013b; Ahmed et al., 2012b; Hassan and Abdel-Aty, 2013). The finding of the Bayesian model suggests that large variations in speed, have an influence on PTW accidents in urban roads as well. It is important to comment on the high odds ratio (3.487) of the coefficient of variation of speed variable, meaning that 1 unit increase results in 3.487 higher odds of an accident involving a PTW than that before the increase.

Lastly, it was found that accident type was strongly associated with PTW accidents. It was found that PTWs are more associated with head-on collisions (Acc.type1), side (Acc.type3) and sideswipe collisions (Acc.type4). The beta coefficient of Acc.type2 (rear-end collisions) include zero meaning non-significance. The reference category was set us an accident with a fixed object or run-off road collisions (Acc.type0). What is interesting is that by interpreting the odds ratios, PTWs are more likely to be involved in head-on, side and sideswipe collisions rather than a collision with a fixed-object or to run-off-road (7.737, 7.546 and 2.652 times more likely respectively). This finding means that PTWs are more vulnerable and thus are more affected by interactions with other motorized traffic, as they are more likely to be involved in multi-vehicle accidents than in single vehicle accidents.

In order to further support this finding, a table of descriptive statistics was constructed. Table (7-7) shows that 107 PTW accidents are single-vehicle, while 219 PTW accidents are multi-vehicle. It is also shown that that the proportion of PTWs is significantly higher in head-on, side and sideswipe collisions, while in off-road and fixed object collisions the PTWs and other vehicles are almost equally involved.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-1.321	0.295	0.267	-1.905	-0.740
Q_avg_1h_up	0.001	0.000	1.001	0.001	0.002
V_cv_1h_up	1.249	0.603	3.487	0.077	2.441
Acc.type0 (ref)	-	-	-	-	-
Acc.type1	2.046	0.479	7.737	1.165	3.053
Acc.type2	0.435	0.260	-	-0.071	0.948
Acc.type3	2.021	0.417	7.546	1.244	2.888
Acc.type4	0.976	0.263	2.652	0.478	1.497
DIC	640.095				

Table 7-6: Summary of the Bayesian logit model for PTW accident probability.

Accident	Accident type				
	Off-road/fixed object	Head-on	Rear-end	Side	Sideswipe
no PTW	119	6	37	8	31
PTW	107	37	54	59	69

Table 7-7: Descriptive statistics of accident type in relation to PTW accident involvement.

7.5 Summary

The aim of this chapter of the thesis was to investigate road safety and more specifically accident severity and accident probability, having taken into account Powered-Two-Wheelers. For that reason, high resolution traffic and weather data as well as other accident characteristics from urban roads were exploited. Due to the existence of high number of candidate independent variables, a famous data-mining technique was utilized, that of Random Forests in order to rank significant variables according to their relative importance. Next, the most important variables as highlighted by the RF models, were mainly applied to the logit models. In general, the models showed adequate statistical fit and unveiled useful information regarding the severity and the likelihood of accidents in urban roads.

In general, the Random Forest technique which is frequently applied in transportation safety research, revealed a consistent significant effect of traffic variables and especially of traffic variation expressed as standard deviation and coefficient of variation. On the other hand, weather variables were not highlighted as significant, and this non-significance was consistent, whether severity or probability was investigated.

Concerning accident severity, the finite mixture model showed better fit than the mixed effects model and is proved to be more appropriate because the latent classes are defined by the model (through the BIC criterion) and therefore the researcher does not have to specify the random effects as in mixed effects model. Through this approach, the issue of unobserved heterogeneity is accounted for. The results indicate a generally mixed influence of traffic variations on severity, although international literature states that traffic variations increase severity. For example, the results revealed a consistent but negative effect of the standard deviation of occupancy, whilst the coefficient variation of flow had a mixed effect depending on the accident type, indicating the presence of heterogeneity. Moreover, weather did not found to have a direct influence on severity, being consistent with Christoforou et al., (2010).

On the other hand, the results of the Bayesian logistic regression revealed more consistent findings. Traffic variation were found to significantly influence accident probability. Summing up, the findings of the Bayesian logistic model for the thesis, could be considered consistent with past literature concerning accident probability, indicating that variations observed in traffic characteristics increase accident risk.

Lastly, PTW accident probability was found to be positively influenced by variations in traffic flow. Moreover, PTWs are more likely to be involved in multi-vehicle accidents. However, rear-end accidents might probably occur in high congestion or in near signalized intersections and seem to affect more other types of motorized traffic, such as cars. The fact that PTWs seem to be prone to be involved in such types of accidents when combined with fluctuations in traffic conditions, implies that in urban road roads, PTWs are particularly affected by the interaction with other motorized traffic.

Chapter 8 Occupant injury severity and accident probability in urban motorways

8.1 Introduction

Understanding the various factors that affect accident severity is of particular concern to decision makers and researchers. Recently, the incorporation of real-time (high resolution) traffic and weather data has proven to be a very fruitful approach when analysing accident severity. Until now, the vast majority of relevant research concerns motorways and urban expressways. A major limitation is the fact, that there is no specific focus on vulnerable road users such as Powered-Two-Wheelers (PTWs). Due to the low weight and the high manoeuvre capabilities of PTWs, and considering the lack of protection of riders, it is of high importance to understand the influence of traffic conditions on PTWs safety.

This chapter aims to analyse accident severity and probability in the urban motorway of Attica Tollway ("Attiki Odos") by using high resolution traffic and weather data. Furthermore, there is specific focus on PTWs, as PTW probability and severity are examined separately. Accident severity is expressed through occupant injury severity, while PTW severity is expressed through injury severity of PTWs occupants. It is observed that these definitions of severity are different that those used so far in this thesis. It is noted though, that PTW accident probability followed the very same definition as in earlier chapters of the thesis.

Severity is explored through finite mixture logit models. These models are capable of accounting for unobserved heterogeneity, which has been recognized as a critical issue when analysing accident severity (Savolainen et al., 2011). This happens because some of the factors affecting severity are not observable. Mannering and Bhat (2014), correctly argue, that when analysing motorcycle severity, heterogeneity is major issue as it may occur from various unobserved factors (motorcycle speed, riders' physical condition etc.), and consequently resulting in biased estimates.

The literature indicates, that the main approach for analysing accident probability (also referred as crash likelihood) is the case-control method, where a number of non-accident cases are collected under appropriate assumptions. However, in reality, accident are rare events. Consequently, a different approach might be more appropriate. For that reason, instead of applying the traditional approach (as applied in Chapter 7), a number of rare-events logit models are applied. A complete description is provided in the following relevant subsections.

The results of this chapter attempt to contribute to the understanding of accident probability and severity in urban motorways, by having a special consideration of PTWs for the first time in motorways and also by developing novel models such as the rare-events logit for the first time in safety evaluation of motorways.

8.2 Methodology

Accident severity is explored through finite mixture logit models which as stated earlier, can account for the unobserved heterogeneity. Furthermore, accident probability is explored through rare-events logit models, being the first time applied in transportation safety.

8.2.1 Finite mixture logit

Random parameters logit and finite mixture logit models accommodate heterogeneity by allowing parameters to vary across observations. More specifically, the purpose of the finite mixture logit model is to divide the sample into homogenous groups (latent classes). Moreover, the finite mixture model does not require to assume a distribution relating the way in which parameters vary across observations. A disadvantage of the finite mixture model is the fact that variation within observations in the same class are not allowed (Mannering and Bhat, 2014). The theoretical background of the finite mixture model was earlier presented in Chapter 7 (section 7.2.2.1).

8.2.2 Bayesian logit

In this chapter, Bayesian logistic regression models were developed to estimate the effect traffic, weather and other characteristics on PTW accident probability. The theoretical background of Bayesian logit models was demonstrated previously in the thesis in Chapter 6 (section 6.2.3). The procedure was the same, and "non-informative" priors were used for the needs of the modelling of this chapter as well. The well-known DIC was used in this analysis to assess the model fit.

8.2.3 Rare-events logit

Most of significant events in reality are rare events. They occur very rarely-that is we have dozens to thousands of times fewer events (e.g. wars, volcano explosions) than non-events. King and Zeng (2001a), identified two major causes for problems when analysing such kind of data. Firstly, traditional statistical procedures underestimate the probability of rare events, and second, the inefficient data-collection strategies. In general, serious problems arise due to the fact that maximum likelihood estimation of the logistic model suffers from small-sample bias, with the degree of bias being strongly dependent on the number of cases in the less frequent of the two categories of the dependent variable y. For example, even with a sample size of 100,000 cases, if there are only 20 events in the sample, substantial bias exists. Consequently, scholars cannot rely on logit coefficients. To solve these problems, King and Zeng (2001a, b), proposed an adapted version of the logistic regression, the so-called rare-events logistic regression.

This approach, applies a number of corrections. The first correction concerns data collection. The authors suggest to perform a case-control sampling design, based on stratified sampling. Thus, it is recommended to include all events and a random selection of non-events. Then, in order to account for the biased estimation of constant term due to the case-control design, a prior correction has to be applied to the constant term. Then the next equation applies:

$$\alpha_0 = \hat{\alpha} - \ln[\left(\frac{1-\tau}{\tau}\right) * \left(\frac{1-\bar{y}}{\bar{y}}\right)], \quad \text{(Eq. 8-1)}$$

where α_0 is the new corrected constant, $\hat{\alpha}$ is the uncorrected constant, τ is the proportion of events in the population and $\bar{y}$ is the proportion of events in the sample. Another method proposed by for correction is the "weighting" method, which was not used in this thesis and thus not described in this section. Moreover, the underestimation of the probabilities when using the corrected intercept α_0 need a similar correction. For that reason, a correction factor C_i is added to the estimated probability p_i. If we assume the corrected logit form based on the corrected constant term:

$$logitp_i = \ln\left(\frac{1-p_i}{p_i}\right) = a_0 + \sum \hat{\beta}_i x_i, \quad \text{(Eq. 8-2)}$$

then,

$$p_i' = p_i + C_i \quad \text{(Eq. 8-3)}$$

where C_i is calculated according to King and Zeng (2001b):

$$C_i = (0.5 - p_i)p_i(1 - p_i)x_0 V(\beta)x_0', \quad \text{(Eq. 8-4)}$$

where p_i is the probability of an event estimated using the corrected estimated coefficient a_0, x_0 is the $1 * (m + 1)$ vector of values for each independent variable, V(β) is the variance-covariance matrix, and lastly x_0' is the x_0 transposed.

In this thesis, the rare-events model is applied where accidents are treated as rare events, and thus the term "event" corresponds to an occurrence of an accident.

8.3 Data Preparation

In this chapter, the required accident, traffic and weather data were extracted from Attica Tollway as described in Chapter 4 (section 4.3). Two entirely different datasets were prepared for the needs of the analysis. Only basic motorway segments (BFS) were considered and not ramp areas. Accident severity, which has two levels, namely fatal-severe injury (KSI) and slight injury (SI), is explored in the first dataset. PTW accident probability is explored in the first dataset as well. The second dataset is utilized to model accident probability.

The first motorway dataset for this chapter, includes 387 cases, from 2006 to 2011. The unit of analysis was any vehicle occupant involved in an accident (rider, driver or passenger) resulting in at least one person being slightly injured. Therefore, each one of the 387 cases in the dataset is a record of the severity level sustained by each vehicle occupant involved in the accident. Therefore, a single accident would correspond to various observations that are equal to the number of all injured persons involved in the accident. "PTW occupant injury severity" refers to the injury severity levels of every injured person involved in the accident who was a PTW occupant (rider or passenger). A percentage of 11.11% of occupants sustained severe injuries or were killed (KSI), while 88.89% of occupants slightly injured (SI). The number of Powered-Two-Wheeler riders was 161 (41.6% of total injured persons). 23 PTW riders (14.29%) were killed or severely injured (KSI) and 138 (85.71%) were slightly injured (SI). A subset of PTW occupants was also created to explore PTW occupant severity separately.

In order to explore the probability of PTW accident involvement, a subset had to be created. More specifically, the accident cases had to be defined (and not the injured persons). Therefore, each row of this subset corresponds to an accident, resulting in 285 accident cases, where a new binary variable was defined, namely "PTW accident involvement". This variable takes two possible values 1 if a PTW was involved in an accident and 0 if no PTWs were involved. It is interesting that PTWs are involved in almost half of the total accident cases (49.5%), while the respective percentage of PTWs in the examined urban roads of the thesis corresponds to 61.9% of the total respective accident cases. The high percentage of PTWs indicates the need to examine the factors that influence the PTW accident probability in motorways as well.

The consideration of relevant traffic data was similar to that of previous chapters. For each accident, a 3-hour time series of traffic flow, occupancy, speed and truck proportion (in 5-minutes intervals ending at 5 minutes before the time of the accident) were extracted. This approach has been followed since, the analysis of Attica Tollway data was more microscopic than that of urban roads (please see section 4.3.4). This time lag was used to avoid the impact of the crash itself on the traffic variables and also to compensate for any 'inaccuracy' in the exact time of the accident. For example, if an accident occurred at 21:00, the traffic data considered were obtained from the 17:55-20:55 period. Similar techniques have been applied in other real-time data analyses (Abdel-Aty et al., 2007; Christoforou et al., 2010). The raw 5-min traffic data before the time of the accident occurrence (considering the 5-min time lag that was described earlier), were extracted from the closest upstream loop detector and then were further aggregated in 15min, 20min, 30min and 60min in order to obtain maxima, averages, standard deviations, medians and coefficients of variation. Therefore, the analysis approach for Attica Tollway was more microscopic than in urban roads. Traffic flow was divided by the number of lanes in order to be consistent across road segments.

Regarding the weather data, the weather records for each meteorological station covered the whole period from 2006-2011. For that reason, each accident case had to be assigned to the closest meteorological station and then the relevant weather data had to be extracted. The 10-min raw data were aggregated in order to obtain maxima, averages and standard deviations, for 1-hour, 2-hours, 6-hours and 12-hours prior to the time of the accident occurrence (considering the 5-min time lag). Regarding rainfall, the total amount of rain has also been calculated for 1h, 2h, 6h and 12h prior to the time of the accident after considering the 5min time-lag that was described earlier. In rare cases, when missing data were observed the accident case was removed. It is noted, that weather parameters were not considered for accident probability modelling.

The second dataset consists of accidents and complete traffic time series measured in 1-hour intervals from 2008-2011 in three random loop detectors in BFS areas with the same number of lanes. The data of the three loop detectors were unified in one combined dataset. This motorway dataset, includes 17 accident cases (occurred nearby these three loop detectors) and 91118 non-accident cases. This dataset was utilized in order to proceed in the accident probability modelling, via the rare-events models. Accident occurrence was defined as binary variable taking the values of 0 (non-accident) and 1 (accident). Therefore, in each 1-h time interval there is the information if an accident has occurred or not.

Traffic variables extracted from the Traffic Management System of Attica Tollway, were measured in 1-h intervals (flow, speed, occupancy and truck proportion). The flow is the total number of vehicles in 1 hour and is measured in vehicles per hour. On the other hand, speed, occupancy and truck proportion are the averaged values of the 5-minutes intervals, which were automatically aggregated into 1-hour intervals. In order to avoid the post-accident traffic conditions where low mean speeds may prevail due to the accident itself, the traffic time series before accident cases had to be checked. The aim was to identify potential sudden fall in mean speeds which would lead to erroneous estimates of the effect of traffic variables, because of the decrease of low mean speeds that could take place due to the effect of the accident itself. In this case, the accident is assigned to the previous 1-hour time interval.

8.4 Results

Sections 8.4.1 and 8.4.2 present the extracted results for occupant and PTW occupant injury severity respectively, while accident probability results are presented in section 8.4.3. Severity models were developed via the *flexmix* package in R (Gruen and Leisch, 2008). Accident probability analyses were performed through the package *zelig* (Imai et al., 2007, 2008) in R software. It is also noted that this procedure offers the option to correct the coefficients β in order to account for the rare events bias.

8.4.1 Occupant Injury Severity

The finite mixture analysis revealed two distinct classes of injured occupants with homogenous attributes; latent class 1 with probability 84.4% and latent class 2 with probability 15.6%. These are the probabilities that each person belongs in each latent class respectively. The optimum number of latent classes was determined according to the BIC criterion, with lower BIC indicating best models.

Table 8-1 summarizes the findings of the finite mixture logit model. The likelihood of the final model was -112.152 and the McFadden R-square was considered adequate with a value of 0.169, since it is suggested that values between 0.2 and 0.4 are of very good fit. Moreover, the likelihood ratio test between the null and the full model is considered significant. Five explanatory variables were found to be significant (accident type, 30min average flow, 30min standard deviation of occupancy, 30min average truck proportion and engine size). The variables of the final model were checked for potential multicollinearity and after having ensured that they were not correlated, it was decided to be retained in the final model. Accident type, average truck proportion and engine size (CC) were set as fixed across latent classes, while the constant, the average flow and the standard deviation of occupancy were found to be random and free to vary across the two latent classes.

Consequently, elasticity analysis (marginal effects) is suggested in order to gain a better understanding of the effect of an variables that differ across latent classes, following the same approach as in Chapter 7 (Shaheed and Gritza, 2014; Xie et al., 2012; Cerwik et al., 2014). The elasticities and the pseudo-elasticities for each variable are presented on Table 8-2.

It is observed that the type of accident was fixed across classes similar to the finite mixture models for urban roads in Chapter 7. It is shown, that side (Acc.type2) and sideswipe collisions (Acc.type3) were found to be statistically significant, while rear-end collisions (Acc.type1) was not significant. The negative signs of the beta coefficients indicate lower injury severities for the aforementioned significant variables than the reference category, which is collision with fixed object/run-off road (Acc.type0). The negative effect of these variables is consistent with Al-Ghamdi (2002) and Theofilatos et al. (2012). However, in Theofilatos et al. (2012), rear-end collisions were found to have a negative effect on accident severity outside urban areas.

The engine size (CC) has a negative sign across latent classes, indicating that occupants of vehicles with small engine are more likely to sustain severe injuries, regardless of the class that they belong, implying a consistent fixed and homogenous effect of engine size. This negative effect may be attributed to the fact, that heavy vehicles such as SUVs and trucks have large engine size offer more security to their occupants.

It is observed that average truck proportion has a fixed negative sign across latent classes. This means that when the proportion of heavy vehicles and trucks is lower, occupants are more likely to be severely or fatally injured. This is probably the first time that this variable is examined, when analysing severity with real-time traffic data, and as such the understanding of the impact on occupant injury severity needs further investigation. In order to understand this effect, the effect of traffic flow needs to be co-examined. Average traffic flow has a negative sign of the beta coefficient and was found significant for class 2, but was not significant for class 1. The negative elasticity value (-0.059), suggests that as average flow increases, there is a reduction in the probability of a severe/fatal severity outcome. This means that high average traffic volumes result in less severe accidents, which was also found in Christoforou et al. (2010). Consequently, less congestion leads to more severe accidents and in that context, the presence of low number of heavy vehicles may worsen the situation, as more space for manoeuvres is left. Moreover, considering that off-road collisions and collisions with fixed objects were found to result in more severe injuries, one might conclude that injury severities in Attica Tollway may not be a matter of interaction among road users.

The standard deviation of occupancy was found to have mixed effects on severity, depending on the latent class, meaning that there is variation in the effect. The positive sign in the first latent class, shows that large variation of occupancy has higher probably of resulting in severe and fatal injuries of occupants. In general, traffic variations are linked to severe accidents (Yu and Abdel-Aty, 2014b), but in this is not supported for latent class 2, where a negative sign of the beta coefficient is observed. Consequently, the final effect is based on the elasticity, where the positive elasticity value (0.086), suggests that as occupancy variation increases, there is an increase in the probability of a severe/fatal severity outcome.

Variables	Latent Class 1			Latent Class 2		
	Mean	t-statistic	p-value	Mean	t-statistic	p-value
Constant (random)	-1.720	-0.961	0.337	36.153	2.573	0.010
Acc.type0 (reference cat.)	-	-	-	-	-	-
Acc.type1 (fixed)	-2.338	-1.352	0.176	-2.338	-1.352	0.176
Acc.type2 (fixed)	-26.779	-2.262	0.024	-26.779	-2.262	0.024
Acc.type3 (fixed)	-7.196	-2.429	0.024	-7.196	-2.429	0.024
CC (fixed)	-0.002	-1.755	0.079	-0.002	-1.755	0.079
Tr.Prop_avg_30m_up (fixed)	-1.038	-2.429	0.015	-1.038	-2.429	0.015
Q_avg_30m_up (random)	0.027	1.362	0.173	-0.236	-2.354	0.019
Occ_stdev_30m_up (random)	154.060	2.597	0.009	-1629	-2.134	0.033
Log-likelihood at zero			-134.998			
Final Log-likelihood			-112.152			
Likelihood ratio test			45.692			
McFadden R^2			0.169			

Table 8-1: Summary of the finite mixture logit model for occupant injury severity.

Variable	e
Acc.type2	-0.843
Acc.type3	-0.933
CC	-0.558
Tr.Prop_avg_30m_up	-0.639
Q_avg_30m_up	-0.059
Occ_stdev_30m_up	0.086

Table 8-2: Average elasticity and pseudo-elasticity of each variable for the finite mixture model for occupant injury severity.

8.4.2 PTW Occupant Injury Severity

The finite mixture analysis revealed two distinct classes of injured PTW occupants with homogenous attributes. PTW occupants have a probability of 80.7% to be assigned in latent class 1 and a probability of 19.3% to be assigned in latent class 2. The optimum number of latent classes was determined according to the BIC criterion, with lower BIC indicating best models, following the same approach as in section 8.4.1.

Table 8-3 summarizes the findings of the finite mixture logit model for PTW occupant injury severity. The final likelihood of the final model was -53.922 and the McFadden R-square was considered relatively adequate with a value of 0.363. Values between 0.2 and 0.4 suggest a very good fit. Moreover, the likelihood ratio test between the null and the full model is considered highly significant.

Three explanatory variables were found to be significant (30min average flow, 30min coefficient of variation of speed and engine size). The variables of the final model were checked for potential multicollinearity and after having ensured that they were not correlated, it was decided to be retained in the final model. Only engine size (CC) was set as fixed across latent classes, while the constant, 30min average flow, 30min coefficient of variation of speed were better explaining the phenomenon if set as be random and free to vary across the two latent classes. The true effect of class-varying variables will be revealed with the elasticity analysis (Table 8-4).

Average traffic flow has a negative sign of the beta coefficient and was found significant for latent class 1, but was not significant for latent class 2. Concerning latent class 1, high average traffic volumes result in less severe injuries to PTW occupants, probably because of the lower speeds that prevail in more congested traffic conditions. It is observed, that the impact of traffic flow is not fixed across observations of the two latent class, but the overall effect is negative, having an elasticity of -0.224.

Another similar result between the two models was found, regarding the impact of traffic variations. Coefficient of variation of speed was found to have mixed effect on injury severity. Large variations in speed increase the likelihood of severe or fatal injuries for latent class 1, while they seem to have the opposite effect on severity for latent class 2. However, the elasticity analysis showed, that if the coefficient of variation of speed increases, this results in an increase in the probability of severe/fatal injury outcome for the PTW occupant. The positive correlation of variations in speed and severity is also found in other studies (Yu and Abdel-Aty, 2014a and b).

Lastly, the effect of engine size was found to be the same as in the previous analysis for Attica Tollway. The negative sign indicates that occupants are more likely to be severely injured as engine size increases. For example, motorcycles are more likely to be involved in less severe accidents than mopeds. This may be attributed to the fact that riders or large motorcycles are probably more cautious when riding and may have better riding capabilities. Furthermore, another trend is observed. The mean value of engine size of slightly injured occupants was found to be 642.842cc, while the respective mean value of severely or fatally injured occupants was 510.41cc.

Variables	Latent Class 1			Latent Class 2		
	Mean	t-statistic	p-value	Mean	t-statistic	p-value
Constant (random)	12.420	1.796	0.072	20.440	1.946	0.052
CC (fixed)	-0.009	-1.934	0.053	-0.009	-1.934	0.053
V_cv_30m_up (random)	16.812	1.106	0.037	-516.820	-2.081	0.037
Q_avg_30m_up (random)	-0.446	-1.647	0.099	-0.020	-0.602	0.547
Log-likelihood at zero	-64.212					
Final Log-likelihood	-40.927					
Likelihood ratio test	46.570					
McFadden R^2	0.363					

Table 8-3: Summary of the finite mixture logit model for PTW occupant injury severity.

Variable	e
CC	-0.523
V_cv_30m_up	0.036
Q_avg_30m_up	-0.224

Table 8-4: Average elasticity and pseudo-elasticity of each variable for the finite mixture model for occupant injury severity.

8.4.3 PTW Accident Probability

The relationship between traffic and weather parameters and PTW accident probability, was examined through the application of Bayesian logit models as previously in the PhD thesis. The followed methodological approach was the same as in previous models. The priors for the constant and for the candidate independent variables were all "vague" (non-informative), assuming to follow a normal distribution with zero mean and very low precision. The prior for the constant was alpha~dnorm(0,0.0001). All candidate independent variables were following a non-informative normal distribution, e.g. beta~dnorm(0,0.0001). The first 5,000 samples were discarded as adaptation and burn-in. Three chains and 20,000 more samples were used to ensure convergence. The Monte Carlo (MC) errors (i.e. the Monte Carlo standard error of the mean values) were monitored. According to Spiegelhalter et al. (2003), MC errors less than 0.05 indicate that convergence may have been achieved. In the model all MC errors were very low (less than 0.005) indicating convergence.

Tables 8-5 and 8-6, summarize the findings of the Bayesian logit model for PTW accident probability, and provides the posterior mean, the standard deviation and the 95% credible interval CI (2.5%-97.5%) and the odds ratios (OR). Only statistical significant parameters are illustrated on the tables.

Linear model	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-0.8357	0.2355	-	-1.2970	-0.3816
Q_avg_30m_up	0.0130	0.0033	1.0131	0.0068	0.0194
DIC	376.926				

Table 8-5: Significant parameters estimates, credible intervals and odds ratios for PTW accident probability linear model.

Non-linear model	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-1.6950	0.3958	-	-2.5090	-0.9607
Q_avg_30m_up	0.0435	0.0114	1.0445	0.0223	0.0666
Q_avg_30m_up^2	-0.0002	0.0001	0.9998	-0.0003	-0.0001
DIC	367.84				

Table 8-6: Significant parameters estimates, credible intervals and odds ratios for PTW accident probability non-linear model.

Only the 30-min average traffic flow was found to be statistically significant and is interesting that a quadratic relationship was revealed as well. More specifically, two models were developed, one linear and one non-linear. The equations are provided below:

$$\mathrm{U1} = -0.8357 + 0.013 * \mathrm{Q\_avg\_30m\_up} \qquad \text{(Eq. 8-5)}$$

$$\mathrm{U2} = -1.695 + 0.0435 * \mathrm{Q\_avg\_30m\_up} - 0.0002 * \mathrm{Q\_avg\_30m\_up}^2 \qquad \text{(Eq. 8-6)}$$

In the first model the average flow has a positive relationship with PTW accident involvement, suggesting that as traffic flow increases, PTWs are more likely to be involved in accidents. However, in the second model a quadratic term of the flow was found to be significant, implying a non-linear relationship between flow and PTW accident probability. The DIC of the non-linear model is lower, suggesting that this model is preferable over the linear model. Therefore, a quadratic relationship between PTW accident probability and average flow is more likely to exist.

The next figures illustrate a graphical representation of the relationship between average flow and the utility function, as well as the probability of PTW accident involvement. More specifically, Figures 8-1 and 8-2 regard the linear model, while Figures 8-3 and 8-4 demonstrate the non-linear model.

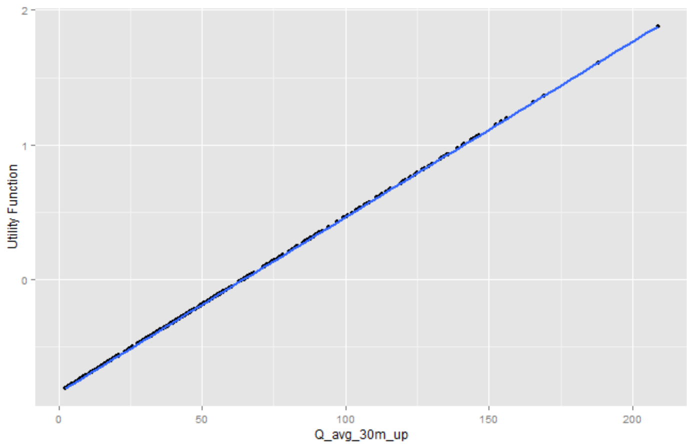


Figure 8-1: Diagram of the relationship between the average flow upstream and the utility function of the linear PTW accident probability model.

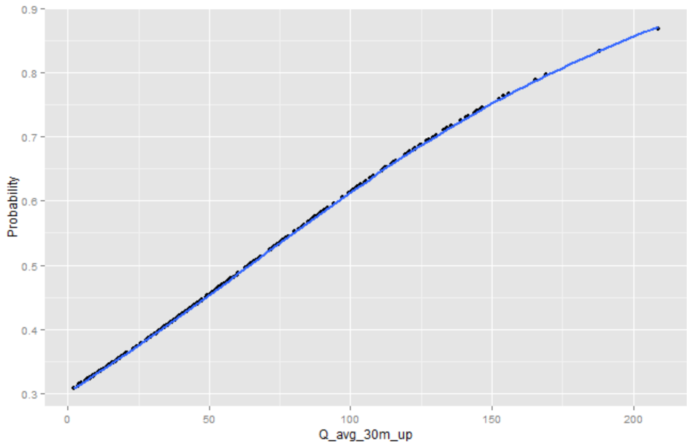


Figure 8-2: Diagram of the relationship between the average flow upstream and the probability of PTW accident involvement (linear model).

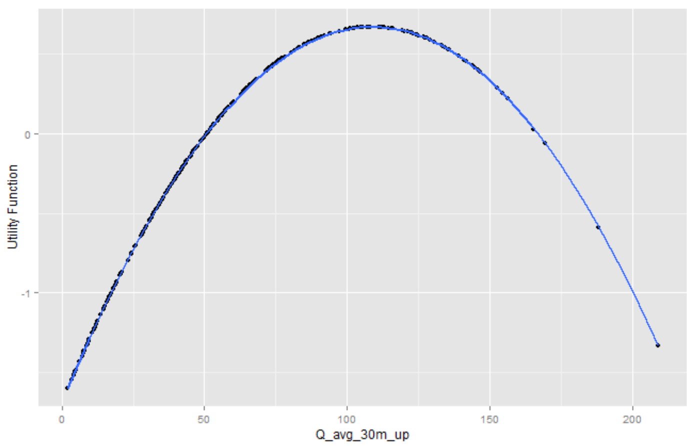


Figure 8-3: Diagram of the relationship between the average flow upstream and the utility function of the non-linear PTW accident probability model.

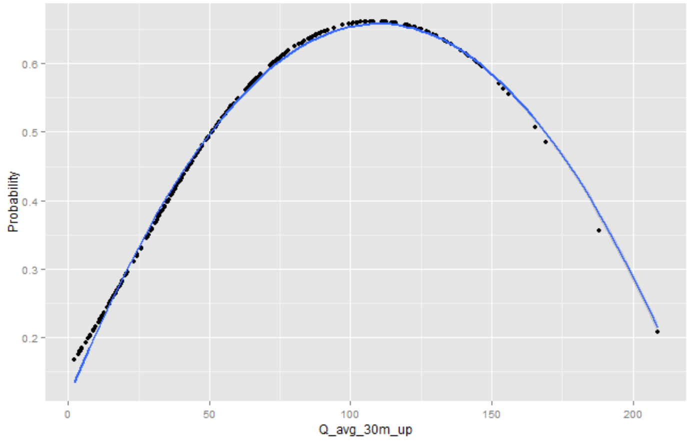


Figure 8-4: Diagram of the relationship between the average flow upstream and the probability of PTW accident involvement (non-linear model).

Figure 8-4, implies that an inverse U-shape relationship exists, meaning that as average flow increases the probability of PTW accident involvement increases until it reaches a maximum and then it starts to decrease.

8.4.4 Accident Probability

This is the first time that accident probability in urban motorways was explored with the application of the rare-events logistic regression. Therefore, the model results presented in this thesis, are a first trial and is attempted to observe whether this methodological approach creates promising results and thus may be potentially considered fruitful.

For the stratified sampling, a proportion of 1:10 for the ratio of events (accidents) to non-events (non-accidents) was used in each sample. The main drawback of the rare-events logistic regression is the dependency of results on the stratified sampling. As a result, three trials were conducted. As suggested by King and Zeng (2001a and b) all accident cases were retained in each sample. Therefore, in each trial, there were 17 accident cases and 170 non-accident cases.

It was not possible to include all explanatory variables in the models, due to serious multicollinearity problems among traffic variables. Therefore, several tests had to be performed in order to find the best combination of independent variables. In order to illustrate the model, but also to highlight which variables are consistently significant, non-significant variables are also included in the final models.

Table 8-7 presents the results of the rare-events logit models for the tree trials, each of them having a different sample of non-accident cases. The results include the logistic coefficients β, the standard error of β, the z-test, the p-value as well as the odds ratios for the significant explanatory variables (not for the constant). Although the previous tables in the thesis did not include the standard error, this statistic is presented in this table, in order to further compare the models of the tree trials. All the models include the "prior correction", where τ is the proportion of events in the population

$(17/91118 = 0.00019)$ and $\bar{y}$ is the proportion of events in the sample $(17/170 = 0.1)$.

The fit of the three models is reasonable. The values of McFadden-R^2 may be considered adequate, since it is suggested that values between 0.2 and 0.4 indicate a very good fit. Furthermore, the change in the log-likelihood is significant in all three models. It is interesting though, that all three models showed very similar fit, either in terms of McFadden-R^2 or AIC. The values of AIC are low indicating good fit and similar, ranging from 106.6 to 106.9.

Trial 1	β	S.E.	z value	p-value	Odds Ratio
Constant	26.4158	11.3706	2.3232	0.0212	-
Truck.Prop.	-0.0394	0.1072	-0.3684	0.7129	-
log(Speed)	-7.4700	2.4369	-3.0653	0.0025	0.0006
Log-likelihood at zero		-113.9			
Final log-likelihood		-100.9			
Likelihood ratio test		26.0			
AIC		106.9			
McFadden R^2		0.1141			

Trial 2	β	S.E.	z value	p-value	Odds Ratio
Constant	33.2999	14.3741	2.3117	0.0216	-
Truck.Prop.	0.0157	0.0981	0.1597	0.8733	-
log(Speed)	-9.0004	3.0874	-2.9152	0.0039	0.0001
Log-likelihood at zero		-113.9			
Final log-likelihood		-100.6			
Likelihood ratio test		26.6			
AIC		106.6			
McFadden R^2		0.1168			

Trial 3	β	S.E.	z value	p-value	Odds Ratio
Constant	29.8363	12.6321	2.3619	0.0192	-
Truck.Prop.	-0.0444	0.0964	-0.4600	0.6460	-
log(Speed)	-8.2035	2.7063	-3.0311	0.0028	0.0003
Log-likelihood at zero		-113.9			
Final log-likelihood		-100.8			
Likelihood ratio test		26.2			
AIC		106.8			
McFadden R^2		0.1150			

Table 8-7: Summary of rare-events logistic regression for three trials.

The three models showed a consistent negative effect of the logarithm of average speed, while average truck proportion was not found to affect accident occurrence. Moreover, the constant was significant in all three models having a positive sign. The negative sign of the beta coefficient of logarithm of average speed may seem counterintuitive, however this finding is consistent with Ahmed et al., (2012b), who found that low average speeds increase on accident occurrence on freeways under clear weather. Therefore, considering the prevalence of good weather conditions in the Greater Athens Area, this negative effect of low speeds on accident occurrence may be considered consistent with the aforementioned study. Moreover, this finding may indicate that accidents in Attica Tollway are more likely to occur in more dense traffic conditions with lower mean speeds.

8.5 Summary

The aim of this chapter of the thesis was to investigate road safety in motorways by utilizing high resolution traffic and weather data and other accident attributes. Firstly, occupant injury severity was explored. PTW occupant injury severity was modelled separately, as the thesis has a focus on PTWs. Then, accident probability was explored.

By applying finite mixture logit models, it was found that a number of traffic parameters such as truck proportion, average flow and standard deviation of occupancy, as well as other risk factors, such as accident type and engine size, significantly affect the severity outcome of vehicle occupants. Moreover, this model accounted for the heterogeneity among two distinct groups of observations. For example, the impact of average flow and standard deviation of occupancy was not fixed across the two produced latent classes and diverse results were produced. However, the elasticity analysis revealed a negative effect of average flow and a positive effect of occupancy variation. On the other hand, high percentage of trucks consistently reduce severity levels of involved injured occupants. Collisions with fixed objects or run-off road collisions as well low engine size vehicles are associated with higher severity levels.

When PTW occupants are examined separately, the model produced slightly different results. For example, accident type was not found to have an impact on PTW occupant severity. On the other hand, variations in traffic (measured by the coefficient of variations in speed) and average flow have mixed effects on severity as well, depending on the latent class in which each observation belongs, but the elasticity analysis revealed a positive influence. Lastly, engine size was negatively correlated with severity levels.

PTW accident probability was explored through Bayesian logistic regression models. It was found the only statistically significant variable is the average 30-min flow upstream of the accident location. Two models were developed: one linear and one non-linear. The fit of the non-linear was better indicating that a quadratic relationship exists, namely an inverse U-shape.

Accident probability was explored by developing rare-events logit models. The method of stratified sampling was followed and three models were developed. The

main risk factor for accident occurrence was found to be the logarithm of average speed, as lower speeds are more likely to result in accident occurrence.
It is interesting, that weather parameters were not found to significantly affect injury severity of occupants in motorways. The insignificance of weather parameters in urban motorways and in urban roads, may be attributed to the fact that weather parameters may not be linearly related with road safety indicators such as severity and probability.

This chapter contributes on the current knowledge, by having a specific consideration of PTW occupants in analysing severity in urban motorways and also by developing finite mixture models combined with high resolution traffic and weather data. Moreover, the application of the rare-events is considered fruitful. From a methodological point of view, the application of finite mixture models not only accounts for the unobserved heterogeneity but also proved capable of providing an understanding of the factors affecting occupant injury severity. In addition, the rare-events logit model was applied in road safety for the first time. The results are considered promising, since the risk factors are identified. However, the beta coefficients were not found to be totally identical among the models, maybe due to the dependence of rare-events logistic on stratified sampling. In order to overcome this limitation and to improve similar models more efforts are needed. Moreover, it would be interesting to examine the impact of weather parameters by applying such logit models.

Chapter 9 Cusp catastrophe theory to model road safety in urban roads

9.1 Introduction

Reducing the severity of road accidents has been primary emphasis of researchers, policy makers and motor-vehicle industry. The literature review conducted in this thesis, has showed that most of existing studies aimed to predict accident probability according to short-term traffic parameters to develop proactive management strategies. Accident severity has received less attention from researchers since, there are fewer studies which used real-time traffic and weather data to examine accident severity (Christoforou et al., 2010; Jung et al, 2010; Xu et al., 2013b; Yu and Abdel-Aty 2014 and 2014b). Savolainen et al. (2011), provided a comprehensive review and assessment of methodological alternatives of statistical models used to analyse severities. The authors argue that the major drawback of alternative methods applied, such as data mining techniques and other artificial intelligent methods (e.g. neural networks) do not possess interpretive abilities of the classical statistical models.

On the other hand, accident probability has gained increased attention from researchers in the last decade and strong efforts have been made to predict or explain the accident occurrence phenomenon mainly in freeways. As indicated earlier in the thesis, the main tool to analyse accident probability in freeways is the development of classical statistical methods (e.g. logistic models, classification trees) and the control-case approach. A clear research gap was identified to be the lack of urban data and the lack of dedication to vulnerable road users such as Powered-Two-Wheelers.

An approach which is different from the classical statistical methodology and has been applied in traffic flow theory (Dendrinos, 1978; Navin 1986; Papacharalampous and Vlahogianni, 2014) but very rarely in transportation safety (Park and Abdel-Aty, 2011) is the catastrophe theory. This method is entirely different from the existing classical statistical methodology, as catastrophe theory searches for the existence of non-linearity in the system and explains sudden transition between states of a dynamic system due to small changes in the input parameters. More details about this methodology is presented in subsection 9.2.1 that follows next.

Summing up, the aim of this chapter is to investigate the feasibility to model severity and probability in urban roads by applying the cusp catastrophe model. When plausible, the cusp catastrophe model is compared to the classic linear regression model. This modelling approach can be considered as first trial and the results of this chapter aim to provide a first insight on accident severity and probability on urban roads by including high resolution traffic and weather data and by considering PTWs as well.

To conclude, the results of the analysis indicate the existence of strong nonlinearity in the safety system. Moreover, the cusp catastrophe models are proved to be useful alternative for explaining the nonlinear relationships that potentially exist between road safety indicators and independent predictors.

9.2 Methodology

The methodology that was followed to achieve the aims of this chapter, is described in the following subsections. The main analytical tool was the cusp catastrophe theory. The theoretical background described in this subsection of the thesis, has a focus on the cusp catastrophe, however, a more brief description of the catastrophe theory in general is also provided. For a very detailed description about the cusp catastrophe, the reader is encouraged to refer to Grasman et al. (2009). Most of the following equations appearing in the subsection are provided in Grasman et al. (2009). Lastly, the fit of the linear regression models was also tested and compared with the cusp models.

9.2.1 Catastrophe theory

Catastrophe theory examines the qualitative changes in the behaviour of systems when the control factors that influence their behavioural state face smooth and gradual changes (Poston and Stewart, 1978). In other words, the catastrophe theory assumes the existence of a dynamic system and explains the sudden transition between the system states, when small changes in the parameters of the system (known as α and β) take place. The term “catastrophe” may be confusing, as it has nothing to do with the consequences of the event. In mathematical sciences, the term catastrophe implies a nonlinear transition from one state to another.

Catastrophe theory became popular in the 1970’s and since then its applications range from economics to psychology. However, this approach had a few major drawbacks. The major reason of criticism stems from the qualitative methodology used in the aforementioned applications, due to the fact that catastrophe theory concerned deterministic dynamical systems (Grasman et al., 2009). Another issue is the ad hoc nature of the selection of the variables that would be used as control factors. Rosser (2007), comprehensively summarizes the critiques of catastrophe theory and the reader may refer to this study. Consequently, there was a deep need to make a stochastic formulation in the catastrophe theory. Indeed, several stochastic formulations have been found along with statistical methods, so that the quantitative comparison of catastrophe models with data is enabled (Cobb et al., 1983; Guastelo 1982; Wagenmakers et al., 2005).

9.2.1.1 Cusp catastrophe

Of the seven elementary types of catastrophe models, perhaps the most popular and easy is the cusp catastrophe. The cusp catastrophe model is capable of handling complex linear and nonlinear relationships simultaneously, by applying a high-order probability density function (Zeeman, 1976). This density function has the advantage to involve sudden behavior jumps and transitions.

The notations used in Barunik and Vosvrda (2009), Grasman et al. (2009) and Park and Abdel-Aty (2011), are followed in this chapter. A deterministic dynamical system is considered:

$$\frac{\partial y}{\partial t} = -\frac{\partial V(y;\alpha,\beta)}{\partial y} \quad \text{(Eq. 9-1)}$$

where y represents the state variable (can be considered as the dependent variable) and α, β are the two control parameters that determine the behaviour of the system. The canonical form of the cusp catastrophe function is:

$$-V(y;\alpha,\beta) = \alpha y + \frac{1}{2}\beta y^3 - \frac{1}{4}\beta y^4 \quad \text{(Eq. 9-2)}$$

This system moves towards equilibrium and will reach one when:

$$-\frac{\partial V(y;\alpha,\beta)}{\partial y} = 0 = \alpha + \beta y - y^3 \quad \text{(Eq. 9-3)}$$

There is one solution to this equation if $\delta > 0$, and three solutions if $\delta < 0$. The term δ is also called Cardan's discriminant and is defined as

$$\delta = 27\alpha - 4\beta^3 \quad \text{(Eq. 9-4)}$$

The set of values α and β for which $\delta = 0$, determines the bifurcation set. Grasman et al. (2009) and Barunik and Vosvrda (2009), have well-explained the cusp equilibrium surface for nonlinear deterministic systems. Statistically speaking, the cusp equilibrium surface may be considered as a response surface, where depending on the values of α and β, its height predicts the value of the dependent variable y. Moreover, the dependent variable y cannot be necessarily observed (and thus being an observed quantity), but it is rather a canonical variable depending on a number of measured dependent variables.

In that context, the control variables α and β are canonical as well and depend on a number of actual measured independent variables. Figure 9-1 illustrates a cusp surface as presented in Grasman et al. (2009). The bifurcation set as defined earlier is depicted on the figure (the crosshatch grey shaped region on the bottom of the figure).

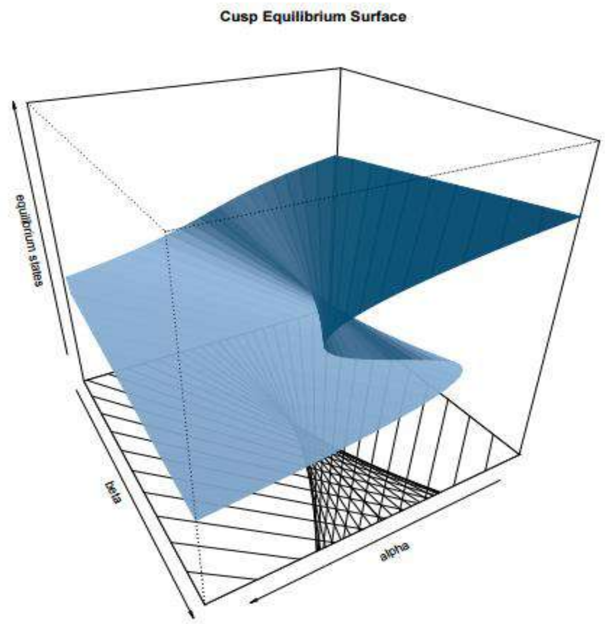


Figure 9-1: Example of cusp surface.

A number of qualitative behaviours of the cusp model were derived by Gilmore (1983). These characteristics are called catastrophe flags. These characteristics are of major importance because the existence of some (or all) of them indicates a strong presence of a good fit to the data and therefore evidence is gathered for the presence of cusp catastrophe in the system. Some of them are sudden jumps in the value of the canonical state variables, hysteresis and multi-modality.

As stated earlier, catastrophe models are applied in deterministic systems. As a consequence, these models cannot be directly applied in stochastic environments. For that reason, a stochastic catastrophe theory was proposed (Cobb and Ragade, 1978; Cobb, 1980; Cobb and Watson, 1980; Cobb and Zacks, 1985) by adding a white noise Wiener process, namely $dW(t)$ to the initial Equation 9-1. Therefore, Eq. 9-1, is transformed to a stochastic differential equation:

$$dY = \frac{\partial V(Y;\alpha,\beta)}{\partial Y} dt + dW(t) \qquad \text{(Eq. 9-5)}$$

This stochastic differential equation is affiliated with a probability density that describes the allocation of system states at any moment in time. It can be expressed as follows:

$$f(y) = \frac{\sigma}{\psi^2} exp[\frac{\alpha(y-\lambda)+\frac{1}{2}\beta(y-\lambda)^2-\frac{1}{4}(y-\lambda)^4}{\sigma^2}] \qquad \text{(Eq. 9-6)}$$

where ψ is the normalizing constant and λ determines the origin of scale of the state variable. The variable β is the bifurcation factor and α is the asymmetry factor. The asymmetry factor governs how close the system is to a sudden discontinuous change of events, while the bifurcation factor governs how large a change will take place. As stated earlier, the variables y, α and β are canonical. Now let's assume a set of measured dependent variables $Y_1, Y_2, \dots, Y_n$, then:

$$y = w_0 + w_1 Y_1 + w_2 Y_2 + \cdots + w_n Y_n. \qquad \text{(Eq. 9-7)}$$

Similarly, if a set of measured independent variables $X_1, X_2, \dots, X_n$ is considered the control factors α and β can be estimated as:

$$\alpha = \alpha_0 + a_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n, \qquad \text{(Eq. 9-8)}$$

and

$$\beta = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n. \qquad \text{(Eq. 9-9)}$$

In order to assess the fit of the cusp model a set of diagnostic tools have been suggested such as the pseudo-R^2 (Cobb, 1998), the well-known AIC (Akaike, 1974) and BIC (Schwarz, 1978). It is noted that the pseudo-R^2 can become negative. In order to further evaluate the cusp model fit, Cobb (1998), suggests more diagnostics. For example, each one of the coefficients $w_1, w_{2,} \dots w_n$ should be statistically significant (except w_0) as well at least one of the a's or the b's. Moreover, at least 10% of the pairs (a_i, β_i) should lie inside the bifurcation region. One alternative diagnostic measure according to a number of studies such as this of Hartelman (1997) and van der Maas et al. (2003), is the comparison of the cusp model with a nonlinear logistic model:

$$y = \frac{1}{1+\exp(-\frac{\alpha}{\beta})} + \varepsilon \qquad \text{(Eq. 9-10)}$$

where the parameters (x, y, z) were defined previously in Equations 9-7, 9-8 and 9-9, while ε is the random disturbance. The nonlinear logistic model has the ability to model the sudden changes in the response variable y in a way "similar" to the sudden transition in the cusp model.

9.2.1.2 Linear regression models

As mentioned earlier, the model fit of the cusp model is compared to the fit of the simple linear regression model. However, in cases when the dependent variable is discrete (e.g. accident probability) the comparison with the linear model does not make sense. In this case, the cusp model is compared only to the nonlinear logistic model that was defined previously in equation 9-10. Since the linear regression was not the main analytical tool for this chapter of the thesis, the theoretical background is very briefly presented in this subsection. For more information, the reader is encourage to refer to Washington et al. (2011). The simplest form of linear regression is:

$$Y_i = \beta_o + \beta_1 X_{1i} + \varepsilon_i \qquad \text{(Eq. 9-11)}$$

where, Y_i is the continuous dependent variable, β_0 is the constant, β_1 is the beta coefficient for the value x_i of independent variable X for observation i, and ε_i is the disturbance term (Washington et al., 2011). One of the major linear regression assumptions is that the disturbances are independent of X and have a value of zero. Moreover, they are identically distributed as normal. In notation these assumptions require:

$$\varepsilon_i \approx rnorm(0,\, \sigma^2) \qquad \text{(Eq. 9-12)}$$

The significance of the candidate variables is usually tested via the t-statistic for 95% level of confidence. The main goodness-of-fit statistic for the linear regression models is the coefficient of determination R^2, ranging from 0 to 1, with 1 meaning perfect fit as all of the variance is explained by the model. When R^2 equals to 0, there is no association between x and Y.

9.3 Data preparation

In this chapter, accident, traffic and weather data from the Kifisias and Mesogeion avenues were used to model accident severity. In addition, Powered-Two-Wheelers (PTWs) were considered as well. The final urban dataset for this chapter, included 527 accident cases for Kifisias and Mesogeion avenues from 2006 to 2011.

Accident severity had two levels, namely fatal-severe injury (KSI) and slight injury (SI). A percentage of 10.8% of accidents were classified as severe (KSI), while 89.2% were classified as slight (SI). PTWs were involved in 326 of those accidents (61.9% of the accidents). A percentage of 7.66% of such accidents were severe/fatal (KSI) while the vast majority (301 accidents), were classified as slight (SI). As stated in Chapter 1, the term “PTW accident severity” is defined very similarly as “accident severity”. The core difference is that it is measured only when a PTW is involved in the accident.

The accident occurrence variable, is a binary variable (1 for accident occurrence, 0 for non-accident occurrence). More specifically, when exploring accident probability, for each accident observation of the dataset, there were 2 observations for non-accidents for the same time and location, one week before and one week after the accident occurrence, resulting in 527 accident cases and 1054 non-accident cases (1581 total cases). The approach was described on Chapter 4. PTW accident probability was explored as well by using the same definition as in previous chapters of the thesis.

In order to explore accident severity in more depth, accident severity was decided to be re-defined and re-coded. More specifically, the dependent variable S*everity*, was re-coded as a continuous variable. Three definitions of severity were used and explored for both accident severity and PTW accident severity. The first definition of severity was the classic definition used in previous chapters (SI or KSI) and is notated as *Severity_1*. Secondly, severity was defined as a percentage of the total severely or killed persons involved in each accident by the total number of persons involved in an accident (Severity_2). The last definition of severity is the total number of persons involved in an accident divided by the total number of vehicles involved in an accident (Severity_3). The two latter definitions of severity are better illustrated throughout the next equations (Eq.8-13 to Eq.8-14):

$$\text{Severity\_2} = \frac{Number\ of\ severely\ injured\ and\ killed}{Total\ number\ of\ persons\ invovled} \qquad \text{(Eq. 9-13)}$$

$$\text{Severity\_3} = \frac{Total\ number\ of\ persons\ invovled}{Total\ number\ of\ vehicles\ involved} \qquad \text{(Eq. 9-14)}$$

The next two tables show the descriptive statistics for the definitions of severity regarding for both the total accidents and those accidents involving a PTW.

Dependent variable	Mean	St.deviation
KSI/total persons	0.0882	0.2644
total persons/total vehicles	0.8864	0.5177

Table 9-1: Descriptive statistics for the dependent variables for accident severity.

Dependent variable	Mean	St.deviation
KSI/total persons	0.0583	0.2108
total persons/total vehicles	0.8413	0.5172

Table 9-2: Descriptive statistics for the dependent variables for PTW accident severity.

As in previous chapters of the thesis, the traffic data from the closest upstream loop detector were considered. As stated in Chapter 4 of the thesis, the 5-min traffic data were further aggregated to 1-hour level to obtain averages, standard deviations and so on, prior to an accident occurrence. It was anticipated that the 60-min traffic data before accident occurrence would cover the hazardous traffic conditions, consequently only the traffic data 1-hour prior to accident occurrence were initially considered.

Weather data for this chapter were the same as in Chapters 6 and 7, and concern the whole period from 2006 to 2011. Each accident case had to be assigned to the closest meteorological station and then the relevant weather data had to be extracted. Then the 10-min raw data were aggregated over hour in order to obtain maxima, averages and standard deviations, in the time-slice of 1-hour, 2-hours, 6-hours and 12-hours prior to the time of the accident occurrence.

9.4 Results

In this thesis, a series of cusp catastrophe models are developed using high resolution traffic and weather data, to model the micro-level safety (accident severity and accident probability) in urban roads. The dependent variable is related to the state variable y, and the traffic and weather parameters as well as other accident variables, constitute the control factors α and β. However, there is no a priori determination of which parameter (traffic, weather, other) would be assigned to each control factor. This happens because there is a lack of objective criteria to determine whether a predictor variable should be classified as an asymmetry or as a bifurcation variable (Cobb, 1981; Grasman et al., 2009). As a consequence, several attempts had to be conducted in order to develop the best possible models.

In this thesis, several cusp models were developed with one measured dependent variable each time, therefore the equation 9-7 is simplified to:

$$y = w_0 + w_1 Y_1. \qquad \text{(Eq. 9-15)}$$

The intention was to investigate the potential existence of non-linearity in the system and also the potential transition of the “accident severity state” from a lower accident severity state (safe state) to a higher accident severity state (unsafe state), or vice versa, through the changes of the various traffic, weather and other accident predictors. The same was attempted for accident probability where two safe states were defined namely the unsafe state (accident occurrence) and safe state (no accident occurrence).

Due the special nature of the cusp modeling approach, the issues and limitations discussed in Park and Abdel-Aty (2011) probably concerns this approach as well, and as such the results must be interpreted very carefully. All the required statistical analyses were developed via the package *cusp* (Grasman et al., 2009) in the R software (R Development Core Team, 2015). This add-on package applies the stochastic method of Cobb et al., 1983).

Consequently, eight cusp catastrophe models were developed; six models regard accident and PTW accident severity, and two models regard accident and PTW accident probability.

9.4.1 Accident severity

Accident severity was explored by considering three severity definitions. The results presented on Table 9-3 confirm the promising applicability of the cusp catastrophe, but the findings are not consistent among severity definitions. However, there is a promising evidence of presence of cusp at least in the first two situations, namely for model 1 and for model 2. The significant variables are highlighted in bold.

Model		Assymetry α				Bifurcation β				Severity y		
		Variable	Coefficient	p-value		Variable	Coefficient	p-value		Variable	Coefficient	p-value
Model 1	a0	**Constant**	-0.2899	0.0000	b0	**Constant**	5.7148	0.0000	w0	**Constant**	-2.3233	0.0000
	a1	**V_cv_1h_up**	-0.6741	0.0110	b1	log(Q_avg_1h_up)	-0.1096	0.3319	w1	**Severity_1**	4.7048	0.0000
	a2	**Acc.type1**	-0.3146	0.0424								
	a3	**Acc.type2**	-0.2468	0.0194								
	a4	**Acc.type3**	-1.2944	0.0012								
	a5	**Acc.type4**	-0.2851	0.0071								
pseudo-R^2	0.9874											
AIC	-270.2550											
BIC	-227.5830											
Linear model R^2	-											
Logistic model R^2	0.0749											
Model 2	a0	Constant	-0.1527	0.1335	b0	**Constant**	4.9843	0.0000	w0	**Constant**	-2.3389	0.0000
	a1	**Q_avg_1h_up**	-0.0006	0.0000	b1	T_1h_avg	0.0102	0.1992	w1	**Severity_2**	4.8234	0.0000
	a2	**T_1h_avg**	0.2549	0.0135	b2	**MC.involvement.no**	-0.2843	0.0372				
	a3	Acc.type1	-0.2566	0.1638								
	a4	**Acc.type2**	-0.2937	0.0175								
	a5	**Acc.type3**	-1.2182	0.0039								
	a6	**Acc.type4**	-0.4165	0.0027								
	a7	**MC.involvement.no**	-0.1481	0.0994								
	a8	**W.Dir_1h_avg**	0.0009	0.0552								
pseudo-R^2	0.8675											
AIC	96.6890											
BIC	156.4296											
Linear model R^2	0.0938											
Logistic model R^2	0.1576											
Model 3	a0	**Constant**	-0.8214	0.0000	b0	**Constant**	1.0920	0.0000	w0	**Constant**	-2.6870	0.0000
	a1	**Acc.type1**	-0.9527	0.0000	b1	**Q_avg_1h_up**	0.0003	0.0000	w1	**Severity_3**	1.4460	0.0000
	a2	**Acc.type2**	-2.7280	0.0000	b2	**Rain_12h_sum**	0.0162	0.0000				
	a3	**Acc.type3**	-3.6840	0.0000								
	a4	**Acc.type4**	-3.0340	0.0000								
	a5	**Sol_1h_max**	-0.0006	0.0000								
	a6											
pseudo-R^2	0.2282											
AIC	1060.4348											
BIC	1107.2481											
Linear model R^2	0.2798											
Logistic model R^2	0.2904											

Table 9-3: Summary results of the cusp models for accident severity.

In the first model (for *Severity_1*), the traditional definition of accident severity that was considered throughout the thesis was utilized. The cusp model has a surprisingly high value of pseudo-R^2 (0.9874). However, this diagnostic is not a panacea as it does not guarantee the presence of cusp and therefore more evidence is needed. It is interesting though that the value of the logistic curve value of R^2 is significantly lower (0.0749) showing that the cusp model is superior to the nonlinear logistic model. The values of the AIC and BIC of the cusp model are also found to be considerably low, indicating a very good fit (-270.2550 and -227.5830), but more graphical diagnostics were conducted.

Figure 9-2 visualizes the 2D projection of the cusp catastrophe surface where the x-axis is the α parameter and the y-axis is the β parameter. Each dot is a single case and its size varies according to the observed bivariate density of the control factors' values at the location of the point. More specifically, the colour of the dots varies depending on the value of the response (state) variable y, with higher values being associated with more intense purple, while lower values with more intense green.

It is observed that the 100% of cases fall inside the V-shaped curve which is the bifurcation area (instability area), meaning that all cases are in a very vulnerable condition where the current (low) severity state could be easily turned into a high severity state. Strictly speaking, the cases that lie inside the bifurcation areas are the cases where a sudden or dramatic change in severity level can occur when there is a small change in α parameter. Therefore, there is a high possibility of the presence of a dynamic nonlinear system. The 3D projection of the cusp catastrophe surface is also illustrated (Figure 9-3).

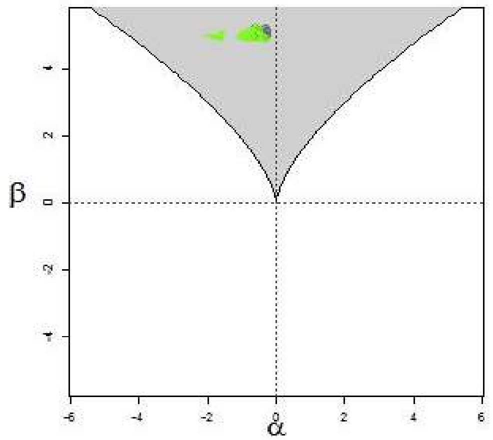


Figure 9-2: 2D projection of cusp surface for accident severity model 1.

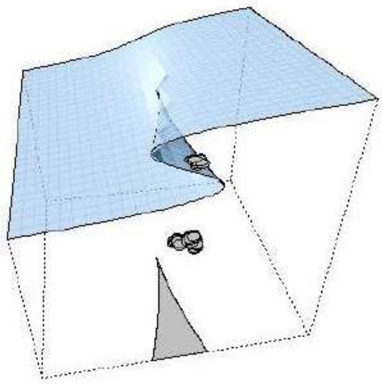


Figure 9-3: 3D cusp catastrophe surface for accident severity model 1.

One more prerequisite for the confirmation of a good fit of the cusp model is the significance of parameter w1 and at least one of a and b. The model shows several significant variables and provides evidence for the existence of nonlinearity. Following Equation 9-7 and examining Table 9-3, the dependent variable of the first model (Model 1) is defined as a linear equation:

$$y = -2.3233 + 4.7048 * Severity\_1. \quad \text{(Eq. 9-16)}$$

The asymmetry factor α is now defined as:

$$a = -0.2899 - 0.6741 * \mathrm{V\_cv\_1h\_up} - 0.3146 * Acc.type1 - 0.2468 * Acc.type2 - 1.2944 * Acc.type3 - 0.2851 * Acc.type4. \quad \text{(Eq.9-17)}$$

It is noted that treatment of discrete variables (e.g. Acc.type) is similar to the classic statistical methods where there is a reference category. In the previous equation (Eq. 9-17), the first category of accident type (collision with fixed object/run off road) was used as a reference category, similar to the previous models at the other chapters of the thesis. The bifurcation factor β can be also extracted easily from Table 9-3. The logarithm of the coefficient of variation of flow (log(Q_avg_1h_up)) is not significant having a p-value of 0.3319, therefore only the constant defines it:

$$\beta = 5.7148. \qquad \text{(Eq. 9-18)}$$

It is shown that the type of accident and the coefficient of variation of speed (V_cv_1h_up) strongly determine the asymmetry factor α. Therefore, small changes in the α factor as determined by these variables can dramatically influence severity level.

Model 2 considers severity as a percentage of the total severely or killed persons involved in each accident by the total number of persons involved in an accident (*Severity_2*). The pseudo-R^2 is high (0.8675), while the values of R^2 of the linear and logistic models are inferior (0.0938 and 0.1576 respectively). The graphical illustrations of model 2 were found to be similar to model 1 (Figure 9-4).

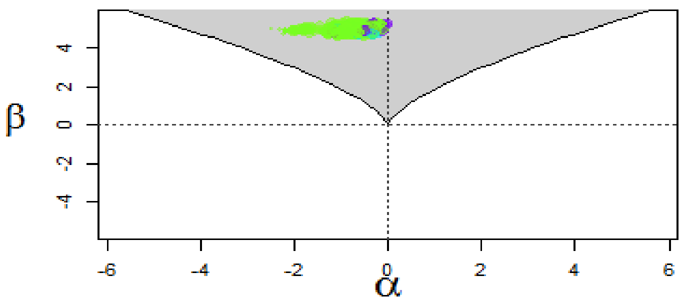


Figure 9-4: 2D projection of cusp surface for accident severity model 2.

The design matrix of this model is constructed similar to model 1, simply by applying the equations 9-7 to 9-9. Table 9-3 shows that the type of accident (Acc.type), the average flow (Q_avg_1h_up) and average wind direction (W.Dir_1h_avg) strongly determine the asymmetry factor α. Moreover, the involvement of a Powered-Two-Wheeler (MC.involvement.no) has a moderate impact as it was found significant for 90% (p-value= 0.0994). Concerning accident type, head-on collisions do not have an impact (p-value=0.1638). The bifurcation factor β is influenced only by the involvement of a PTW (MC.involvement.no), as the 1-h average temperature (T_1h_avg) was not significant (0.1992). It is interesting to notice that the potential involvement of a PTW variable was found to be influence both the asymmetry and the bifurcation factor.

The last definition of severity that was explored, was the total number of persons involved in an accident divided by the total number of vehicles involved (Severity_3). The statistical fit of this cusp model was found not superior to the linear and logistic model. The R^2 of the cusp model is relatively low (0.2282) compared to that of the linear (0.2798) and the logistic (0.2904), implying that the latter models can describe the severity phenomenon slightly better than the cusp. Nevertheless, a number of traffic and weather parameters were found to be statistically significant (average flow, solar radiation, rainfall) and accident type as well. The graphical representation of the cusp model sheds more light (Figure 9-5).

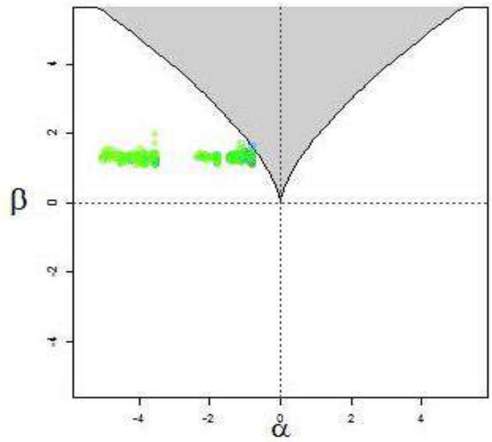


Figure 9-5: 2D cusp catastrophe surface for accident severity model 3.

One can observe that the vast majority of cases lie outside the bifurcation area. Although a small number of cases are placed inside or very close to the instability area, the system cannot be considered as dynamic, taking also into account the superior fit of the linear and logistic model that was found earlier. Therefore it is very possible that the linearity of the system is preserved even if the model shows that a number of cases lie inside the instability area.

Concerning speed variation, this is an important finding, because this parameter can be used as an important tool to improve safety in urban arterials. On the other hand, the type of accident seems to plays a vital role in road accident severity as found in Chapter 7 of the thesis, however a nonlinear relationship was found to exist, implying that the type of accident strongly determines accident severity. The involvement of a PTW plays a role in determining the dynamics of accident severity, having an impact on both asymmetry and bifurcation factors. Thus, it might be said that accident severity might be strongly affected by the potential involvement of a PTW through a nonlinear relationship.

Lastly, the weather parameters that were found to have an impact are the average temperature and the average wind direction. The linear effect of wind is not extensively explored in literature as the related studies exploring this effect are relatively few (Baker and Reynolds, 1992; Levine et al., 1995; Hermans et al., 2006; Brijs et al., 2008). Baker and Reynolds (1992) found during heavy storms in the UK almost 50% of accidents are overturning accidents and 66% of accidents involved high trucks implying a risk. The other studies do not support the hypothesis of significant effect of wind on accidents. For that reason, the effect of wind direction need to be further investigated in future studies. Concerning the average temperature, the association with injury accidents was explored in literature (Hermans et al., 2006; Scott, 1986). It is important though that non-linear relationships were found to exist by Brijs et al. (2008).

Summing up, when modelling accident severity, there were cases when the evidence of existing dynamic system cannot be guaranteed, but in general the application of cusp catastrophe models seems fruitful and very promising.

9.4.2 PTW accident severity

The cusp models of this sections are explored by considering the very same severity definitions as previously in accident severity. The severity notations are still the same. For example model 1 of PTW accident severity uses the same severity definition as model 1 for accident severity and is notated as *Severity_1*. The same applies for models 2 and 3. The summary results of the cusp models for PTW accident severity are presented on Table 9-4. The promising applicability of the cusp catastrophe is observed when PTW accident severity is modelled, having consistent results with the accident severity models of section 9.4.1.

Model 1 has a considerably high value of pseudo-R^2 (0.9954) and very good values both of AIC (-113.6183) and BIC (-71.9624). On the contrary, the logistic fit is significantly inferior to the cusp model, having a low R^2 (0.1107). The graphical illustrations strongly support the evidence that the system is nonlinear. Figure 9-6 depicts the 2D cusp surface for this model showing that the complete number of cases fall within the critical instability area. It is also observed that Figure 9-2 and Figure 9-6 are very similar (accident severity and PTW accident severity).

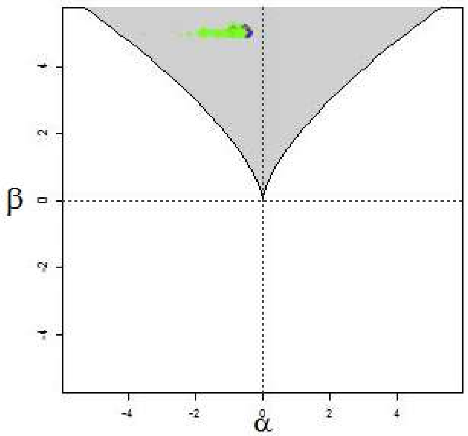


Figure 9-6: 2D projection of cusp surface for PTW accident severity model 1.

In general, accident severity model 1 has lower AIC and BIC values and had a slight better fit than PTW accident severity model 1. Concerning the significant variables, it was found that accident type and rainfall have an impact on the asymmetry factor α. More specifically, Acc.type 2 (rear-end collisions) and Acc.type3 (side collisions) were found to be significant. The high significance of the type of collision is consistent with the findings of the previous section regarding accident severity. The total 12-hour precipitation was found less influential but still it was significant for 90% level (p-value= 0.0892). This may imply that small changes in the long-term precipitation levels during the day may dramatically either increase or decrease the severity level of accidents with PTWs in urban roads. The bifurcation factor β is influenced only by the constant b0, as the logarithm of the 1-hour average flow was not significant.

The second model for PTW accident severity (Model 2) was found to have a high value of pseudo-R^2 (0.8041) and very good values both of AIC (-113.6183) and BIC (-71.9624). The linear and the logistic models were found to be inferior to the cusp model, having low R^2 values (0.0983 and 0.0585 respectively). The visual diagnostic confirms the nonlinearity of the system (Figure 9-7).

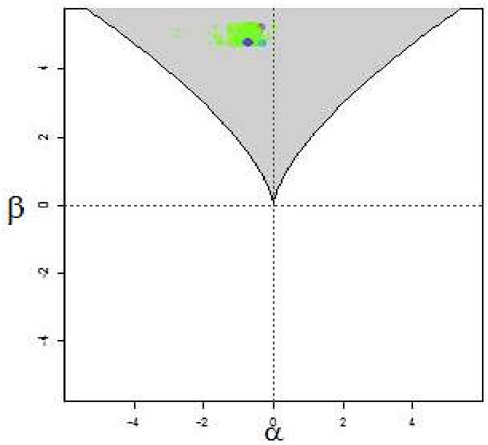


Figure 9-7: 2D projection of cusp surface for PTW accident severity model 2.

When accident severity and PTW severity are compared, accident severity model 2 has lower AIC and BIC values and show a slight better fit than PTW accident severity model 2. In PTW accident severity model 2 (Severity_2). Concerning the significant variables, it was found that the logarithm of the 1-hour average flow (Q_1h_avg_up) and total rainfall have an impact on the asymmetry factor α. Rainfall was found to be less influential with generally at 90% level.

On the contrary, the significance of average at a 95% demonstrates the important role of traffic flow on determining the level of accident severity with PTWs. Accident type is the only independent variable that has an impact on the bifurcation factor β. Acc.type 2 (rear-end collisions) and Acc.type3 (side collisions) were found to be significant.

Lastly, model 3 (Severity_3) deserves a discussion. In spite of the fact that all previous models for PTW accident severity were slightly worse than the accident severity models, PTW accident severity model 3, performs better than the respective model 3 of accident severity, mainly due to the much lower values of AIC and BIC. As a first remark, it is more likely to support the evidence for cusp for PTW accident severity than accident severity. Moreover, the values of the R^2 of the cusp is more or less the same with the linear and logistic model. This of course does not guarantees the presence of cusp because it implies that both three candidate models (cusp, linear, logistic) do not substantially differ.

It is worth noticing though, that almost all predictors are statistically significant. This may justify the applicability of the cusp catastrophe model. More specifically, all accident types and the 1-hour maximum solar radiation were found to be significantly affecting asymmetry factor α, while the 1-hour average flow determines the bifurcation factor β. The visual display of the model fit gives further insight (Figure 9-8). Although a number cases lie inside or very close to the bifurcation area, the presence of cusp cannot be easily confirmed, however there is stronger evidence for cusp presence in this model than the respective model 3 of accident severity.

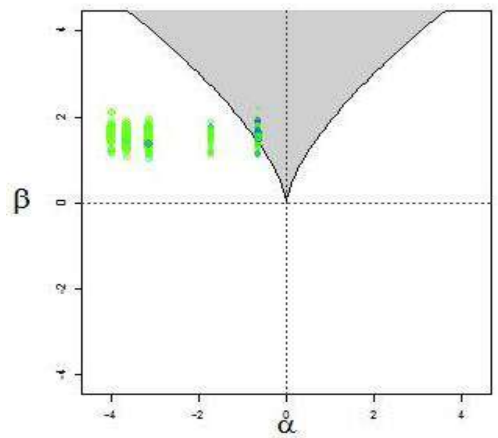


Figure 9-8: 2D projection of cusp surface for PTW accident severity model 3.

Summing up, the presence of cusp and thus the nonlinearity of the safety system is not always preserved. The first two models of PTW accident severity show a strong evidence of nonlinearity, with several traffic and weather variables playing a consistently important role in determining the asymmetry and the bifurcation factors. On the other hand, the only significant predictor among the other accident variables is the type of accident. Therefore, smooth changes in these parameters significantly affect the stability of the safety system, causing transitions from the existing low severity level to higher severity levels. Concerning the third model, the presence of cusp is not as strongly supported as in the previous two models, but the possibility for the presence of such nonlinear dynamics cannot be excluded.

Model		Assymetry α				Bifurcation β				Severity y		
		Variable	Coefficient	p-value		Variable	Coefficient	p-value			Coefficient	p-value
Model 1	a0	**Constant**	-0.4610	0.0026	b0	**Constant**	5.5490	0.0000	w0	**Constant**	-2.3390	0.0000
	a1	T_1h_avg	0.0019	0.7909	b1	log(Q_avg_1h_up)	-0.0835	0.6314	w1	**Severity_1**	4.7722	0.0000
	a2	**Rain_12h_sum**	-0.1787	0.0892								
	a3	Acc.type1	-0.1650	0.3120								
	a4	**Acc.type2**	-0.4432	0.0330								
	a5	Acc.type3	-1.3155	0.0037								
	a6	Acc.type4	-0.2030	0.1300								
pseudo-R^2	0.9954											
AIC	-113.6183											
BIC	-71.9624											
Linear model R^2	-											
Logistic model R^2	0.1107											
Model 2	a0	**Constant**	1.1122	0.0650	b0	**Constant**	4.7636	0.0000	w0	**Constant**	-2.3328	0.0000
	a1	**log(Q_avg_1h_up)**	-0.2729	0.0036	b1	Acc.type1	0.3304	0.2158	w1	**Severity_2**	4.9400	0.0000
	a2	**Rain_12h_sum**	-0.1949	0.0827	b2	**Acc.type2**	0.4835	0.0405				
					b3	**Acc.type3**	0.3808	0.0992				
					b4	Acc.type4	0.2014	0.3541				
pseudo-R^2	0.8041											
AIC	164.003											
BIC	201.749											
Linear model R^2	0.0983											
Logistic model R^2	0.0585											
Model 3	a0	**Constant**	-0.6272	0.0000	b0	**Constant**	0.9996	0.0000	w0	**Constant**	-2.7342	0.0000
	a1	**Acc.type1**	-1.1054	0.0000	b1	**Q_avg_1h_up**	0.0007	0.0000	w1	**Severity_3**	1.5773	0.0000
	a2	**Acc.type2**	-2.5298	0.0000	b2	T_1h_avg	-0.0004	0.4500				
	a3	**Acc.type3**	-3.3955	0.0000								
	a4	**Acc.type4**	-3.0374	0.0000								
pseudo-R^2	0.3479											
AIC	607.4205											
BIC	645.166											
Linear model R^2	0.3317											
Logistic model R^2	0.3539											

Table 9-4: Summary results of the cusp models for PTW accident severity.

9.4.3 Accident probability

One cusp model was developed for accident probability. Table 9-5 illustrates the summary results of the cusp model. The findings of the model are considered promising as the value of the pseudo-R^2 of the cusp model is considered very highly satisfactory (0.9258), and is superior to the respective value of the logistic model (0.0136) which cannot adequately describe the phenomenon. Moreover, all a, b and w parameters were found statistically significant and thus the requirement for significance is met.

Model		Assymetry α				Bifurcation β				Accident Occurrence y		
		Variable	Coefficient	p-value		Variable	Coefficient	p-value			Coefficient	p-value
	a0	**Constant**	-0.2840	0.0000	b0	**Constant**	-	-	w0	**Constant**	-1.996	0.0000
	a1	**V_cv_1h_up**	0.5587	0.0000	b1	**Sol_1h_max**	0.0004	0.0000	w1	**Accident Occurrence**	3.978	0.0000
	a2	**Sol_1h_max**	0.0004	0.0002	b2	**Q_avg_1h_up**	0.0044	0.0000				
pseudo-R^2	0.9258											
AIC	1626.5610											
BIC	1658.7100											
Linear model R^2	-											
Logistic model R^2	0.0136											

Table 9-5: Summary results of the cusp models for accident probability.

Furthermore, visual diagnostics are applied to confirm the nonlinear dynamics of accident and non-accident cases. The visual display of the model cusp surface shows that the vast majority of observations lie inside the bifurcation area (Figure 9-9), implying the presence of cusp, as the minimum percentage for confirmation of cusp acceptance is required to be only 10%. These findings provide evidence supporting the nonlinear nature of accident occurrence in urban roads. Figure 9-10 displays a 3D projection of the cusp catastrophe surface for this model. It is noticed that the dots represents the cases and the V-shaped grey area represents the bifurcation area (instability area).

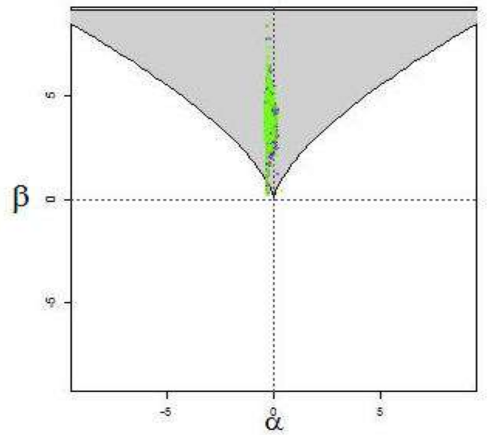


Figure 9-9: 2D projection of cusp surface for accident probability model (accident occurrence).

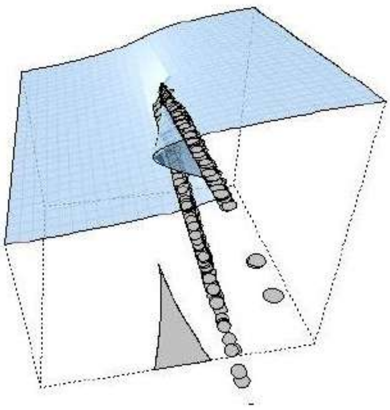


Figure 9-10: 3D projection of cusp surface for accident probability model (accident occurrence).

The design matrix of the cusp model can be now constructed, on the basis of the results of Table 9-5. The response variable y takes the simplest possible form, as only one dependent variable was considered, namely accident occurrence, and is the following:

$$y = -1.996 + 3.978 * Accident\ Occurrence. \quad \text{(Eq. 9-19)}$$

The asymmetry factor α is defined as a linear function:

$$a = -0.2840 + 0.5587 * \mathrm{V\_cv\_1h\_up} + 0.0004 * Sol\_1h\_max. \quad \text{(Eq. 9-20)}$$

Likewise, the bifurcation variable β is defined as another linear equation of the predictor variables:

$$\beta = 0.0004 * Sol\_1h\_\max + 0.0044 * \mathrm{Q\_avg\_}1h\_up. \quad \text{(Eq. 9-21)}$$

The significance of the vast majority of variables supports the existence of nonlinear relationships. It appears that the asymmetry factor is affected by 1-hour coefficient of variation of speed and by 1-hour maximum solar radiation, whilst the bifurcation factor β is affected by 1-hour average flow as well as 1-hour maximum solar radiation. It is interesting to notice the lack of a constant in the bifurcation factor.

It is important that traffic flow and variations in speed were found to have a strong effect on the safety level. This constitutes an interesting finding due to the fact that, if these traffic parameters are controlled road safety in urban arterials will be improved. This will in turn be very useful for policy makers. Consequently, further research is needed towards that direction, in order to gain a deep understanding of the way in which gradual changes in these parameters determine the occurrence of accidents.

Another interesting remark is that solar radiation, which is an indirect measure of weather condition, is used in both factors, thus having high influence on the stability of the system. Note that the1-h maximum solar radiation was found to be positively correlated with 1-hour average (0.496) and 1- hour maximum temperature (0.502) and negatively correlated with the 1-h precipitation (-0.87) could be another indicator. The problem is that the cusp catastrophe model is not as explanatory as a traditional linear model. However, it can be drawn that small changes in the maximum values of solar radiation may cause a sudden change in the probability of an accident occurrence. It may be not entirely appropriate, but it will be aimed to link maximum solar radiation with heat, therefore small changes in maximum solar radiations will be translated as small changes in the heat level. One explanation for the effect of this parameter on accident occurrence could be the fact that driving is a demanding task that requires increased attention, with some studies having showed that heat is associated with driver distraction and errors (Daanen et al. 2003; Mackie and O'Hanlon 1977).

Summing up, it can be concluded that small changes in factors α and β caused by the aforementioned traffic and weather variables, such as average flow, speed variation and weather conditions (expressed indirectly via solar radiation) can cause a sudden transition from the safe state (non-accident occurrence) to the unsafe state (accident occurrence) and vice versa, due to the fact that the dynamic safety system seems to be vulnerable and unstable.

9.4.4 PTW accident probability

As indicated earlier, PTW accident probability was set in a different manner than accident probability. In the thesis, PTW accident probability is explored when an accident has already occurred and a PTW was involved. For that reason it can be also defined as *"PTW accident involvement"*. The summary of the cusp catastrophe model that was developed for PTW accident probability is demonstrated on Table 9-6.

Model		Assymetry α			Bifurcation β				PTW Accident invovlement y			
		Variable	Coefficient	p-value		Variable	Coefficient	p-value			Coefficient	p-value
	a0	**Constant**	-0.7148	0.0000	b0	Constant	-	-	w0	**Constant**	-1.1511	0.0000
	a1	**Q_avg_1h_up**	0.0009	0.0000	b1	**V_cv_1h_up**	4.3153	0.0000	w1	**PTW Accident Involvement**	2.4605	0.0000
	a2	**Acc.type1**	1.4027	0.0000								
	a3	**Acc.type2**	0.2717	0.0368								
	a4	**Acc.type3**	1.3054	0.0000								
	a5	**Acc.type4**	0.6211	0.0000								
pseudo-R^2	0.333											
AIC	1052.6164											
BIC	1091.0120											
Linear model R^2	-											
Logistic model R^2	0.1165											

Table 9-6: Summary results of the cusp models for PTW accident probability.

At first glance, the cusp model seems to be superior to the logistic model, due to the higher value of R^2 (0.333). The routine of constructing visual diagnostic graphs applies in this section as well, in order to gather evidence for the support or the rejection of the nonlinearity in the system. Figure 9-11, demonstrates the cusp surface, where the V-shaped grey area is the bifurcation area (instability area), and the dots representing the accident cases. One remark is that a considerable number of cases lie inside or very close to the bifurcation area. These cases are instable and can easily be changed from an accident without a PTW to an accident with a PTW, while cases outside the bifurcation area are more stable. Another remark is a strong difference observed on the size and the colour of the dots. Each dot is a single case and its size varies according to the observed bivariate density of the control factors' values at the location of the point. As stated earlier in the chapter, the colour of the dots depends on the value of the response (state) variable y (higher values are associated with more intense purple, lower values with more intense green).

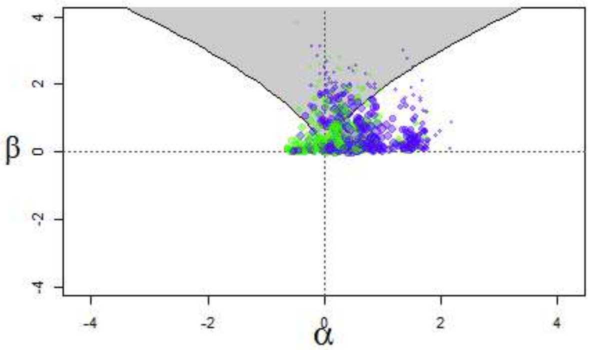


Figure 9-11: 2D projection of cusp surface for PTW accident probability model (PTW accident involvement).

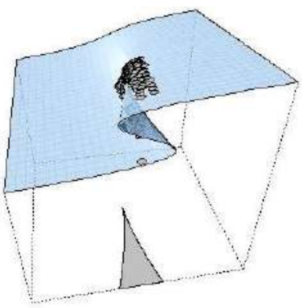


Figure 9-12: 3D projection of cusp surface for accident probability model (accident occurrence).

Therefore, following the same approach as earlier in the chapter, the equations are designed:

$$y = -1.1511 + 2.4605 * PTW\ Accident\ Involvement. \qquad \text{(Eq. 9-22)}$$

The asymmetry factor α is defined as:

$$a = -0.7148 + 0.0009 * Q\_avg\_1h\_up + 1.4027 * Acc.type1 + 0.2717 * Acc.type2 + 1.3054 * Acc.type3 + 0.6211 * Acc.type4. \qquad \text{(Eq. 9-23)}$$

Then the bifurcation variable β is defined as:

$$\beta = 4.3153 * V\_cv\_1h\_up. \qquad \text{(Eq. 9-24)}$$

The significance of the vast majority of variables supports the existence of nonlinear relationships as suggested by the guidelines of Cobb (1998) and Hartelman (1997). The asymmetry factor α is affected by 1-hour average flow and by the type of accident. On the other hand the bifurcation factor β is affected only by 1-hour of coefficient of variation of speed. It is interesting to notice the lack of a constant in the bifurcation factor in a way similar to accident probability results (see section 9.4.3).

The results indicate that cusp model is able to describe the data and give some insight on the underlying complex relationships that determine the probability that a PTW is involved in accident. Nonetheless, the model fit is not excellent, but the assumption that accident mechanism regarding PTW accident involvement is nonlinear cannot be rejected.

9.5 Summary

This chapter has presented the analysis of road safety in urban roads by applying sophisticated cusp models. More specific, accident severity, Powered-Two-Wheeler accident severity (that is accidents with a PTW), accident probability (accident occurrence) and lastly PTW accident probability (PTW accident involvement). The aim was to examine the assumption that safety of the system as expressed by the aforementioned dependent variables, could be considered as a nonlinear dynamic system, where the transitions from safe to unsafe conditions and vice versa, can happen due to smooth or small changes to some control factors. Traffic, weather and traditional accident information were considered as potentially critical to the construction of the control factors.

The results of accident severity models justify the potential applicability of the cusp catastrophe. In general, variations in speed, average flow upstream, accident type and a number of weather parameters such as average temperature, total rainfall, solar radiation and wind direction, were found to have a potential effect on the system dynamics. However, findings do not always confirm the strong presence of a dynamic system. The same applies when PTW accident severity is considered, where in some cases, linear models are proved equally capable of describing the underlying phenomenon. One can conclude that in such cases the linearity of the safety system is preserved. Nonetheless, PTW accident severity shows slightly stronger evidence for existence of nonlinearities than accident severity.

On the other hand, accident occurrence (i.e. accident probability) could be potentially considered as a dynamic phenomenon, since the findings gives evidence to support the nonlinear nature of accident occurrence in urban roads. This implies that sudden jumps from accident to non-accident situations (and vice versa) can be easily happen due to changes in flow, speed variations and maximum solar radiation.

Lastly, the cusp model seems capable of describing the probability that a PTW is involved in accident, when parameters like average flow, speed variations and accident type are considered. It is noted though, that the model fit is inferior to the previous models of the chapter. Therefore, examining the accident mechanism (involving a PTW) as a nonlinear dynamic system must be done carefully. For that reason, more research efforts have to be done towards that direction, in order to gather more evidence.

An interesting note on the findings of this chapter, concerns the statistically significant effect of weather variables in the majority of the cusp catastrophe models. This raise the need for more research, because it implies that high resolution weather variables have an impact on road safety of urban roads, only if nonlinear relationships are considered. In that context, most of the respective results of Chapters 7 and 8 are inconsistent, in regard with the effect of weather (see for example accident probability models).

The obtained results of this chapter confirm in general that road safety in urban roads could be treated as nonlinear dynamic systems, when high resolution traffic and weather traffic data are considered. Moreover, some other accident characteristics such as the type of accident, consistently influence the system dynamics. In other words, the findings indicate that the dynamic change in urban road safety levels expressed by accident severity and accident probability is very likely to be nonlinear in nature.

Unlike the traditional linear modelling approach, the results indicate the possible existence of a catastrophic influence of relatively short-term changes in traffic and weather factors on the system, as sudden changes between different states of the system take place. As a consequence, this theory could be proved to be a useful tool for developing indicators of a catastrophe, although the actual points at which the catastrophic changes occur cannot be easily predicted. Although there is definitely much room for additional research, this chapter clearly demonstrates the possibility of using high resolution traffic and weather data to estimate accident severity and probability, through the development of an advanced stochastic differential equation (i.e., cusp catastrophe model).

One last remark, is that the results of this chapter have to be treated with care, as the statistically satisfactory fit of the majority of the proposed models of this chapter, by no means gives us definitive evidence for the presence of dynamical phase transition. In that context, if more research is done towards this direction, the prediction and the qualitative assessment of the catastrophe points would definitely have an outstanding contribution to road safety, due to the enhancement of the proactive safety management system.

Chapter 10 Critical synthesis of findings

10.1 Introduction

This PhD thesis explored accident probability and severity in urban roads and urban motorways with linear as well as non-linear models. Each chapter demonstrated the main methodological results as well the main findings, regarding the issues of accident probability and severity. This chapter serves as supplement to the previous chapters. Therefore the aim of this chapter is twofold. Firstly, a critical comparison of the analysis models is demonstrated. Secondly, a critical synthesis of the results regarding accident probability and severity is provided, where the main findings of accident probability and severity models are summarized and critically discussed.

10.2 Critical comparison of analysis models

The methodological contribution of this PhD thesis is the introduction of a variety of statistical methodologies to identify the influence of various independent variables on traffic safety. In terms of statistical and econometric models, popular models were initially considered and employed. Subsequently, a number of models (Support Vector Machines, finite mixture logit, cusp catastrophe) were used for the first time when high resolution data in urban roads are exploited, while some other or for the first time in road safety in general (rare-events logit). The aim of the various statistical models that were developed was either predict or explain the accident probability and severity phenomena.

When **accident and PTW accident severity** are examined, popular models employed by road safety modellers were considered in this thesis. To account for the unobserved heterogeneity, finite mixture logit (or latent class logit) models as well as mixed effects logit models were applied. One exception is Chapter 6, where a fixed effects logistic model with full Bayesian inference was applied, because the main objective of that chapter was to focus on the effect of traffic states only. In general, the finite mixture models showed better goodness-of-fit when compared with the mixed effects models. However, a drawback of the mixed effects models is that the researcher needs to have a prior knowledge about the random effects, otherwise several tests should be carried out. On the other hand, the finite mixture models automatically divides the population into heterogeneous groups. The effect of the independent variables can be fixed across the groups or can vary. One minor drawback of the finite mixture model is that heterogeneity among observations of the same group cannot be fully accounted for.

For the case of **accident and PTW probability** analyses, several categorical outcome models were considered, including both the frequentist and the Bayesian approach. The Bayesian logit models which were applied are considered to be superior to the classic logistic models as is suggested in literature. The constantly increasing popularity of Bayesian methods is also considered. The classical logistic regression treats the parameters of the models as fixed and the data are used solely to best estimate the unknown values of the parameters. On the other hand, the Bayesian approach treats the parameters as random variables. In addition, the beliefs about the variables are updated as data are entered into the model. Moreover, Bayesian inference can effectively avoid over fitting problems that arise when the sample size is low and the number of variables is large. The drawback of Bayesian inference is that prior distributions for the various parameters have to be determined, however, as stated in previous chapter of the PhD thesis, this was overcame with the use of non-informative prior distributions.

A frequentist approach was applied to model accident probability in motorways, mainly because of the data opportunities offered by the Attica Tollway. More specifically, this approach followed in Chapter 8 to model accident probability, namely the **rare-events logit model, is used for the first time in road safety**. The rare-events logistic regression is used when the number of events (e.g. accidents) are very few in relation to the non-events (non-accidents). As discussed earlier in Chapter 8, traditional statistical procedures are not appropriate to model rare events. One major problem that arises is the small-sample bias. The rare-events logit which is a modified version of the logit model, applies a number of corrections to the parameter estimates.

Moreover, a core difference when modelling of accident probability, is the kind of utilized data that were considered in each accident probability model. The rare-events model used various samples extracted from **massive non-accident data collected from urban motorways**, through the process of stratified endogenous sampling as suggested in literature (King and Zeng, 2001a and b). In this context three candidate models were applied, because of the dependence of the rare-events logit on the stratifying sample. One the other hand, the previous methods to model accident probability in Chapters 6 and 7 were constructed on the basis of different data collection strategies, namely, on the traditional data collection strategy as used in recent literature, which is less efficient than the rare-events logit model, mainly because of the absence of corrections that exist in the rare-events logit.

A remark should be added in relation to the PTW accident probability modelling techniques used in this PhD thesis. The thesis has shown the applicability and the promising results of a combined modelling technique, which is proposed by Zhao (2012). This technique uses original or transformed labelled time series data as input to advanced models such as k-NN or Support Vector Machine models so as to predict the label of unlabelled time series. In this thesis, a number of Support Vector Machines (SVMs) were developed. The main advantage of SVMs is that can provide very good predictability and can consider nonlinear relationships as well. The main problem that occurs in such techniques (e.g. in neural networks), is the limited interpretability of the effect of individual independent variables, as these techniques are not used for identifying and quantifying relationships but only for prediction purposes. In order to overcome this limitation, the thesis used SVMs on the basis of time series data and raised the question whether time series data can contribute to the proactive safety management. **This methodological approach, was a first attempt to incorporate time-series data when analysing road safety with real-time traffic data.** Moreover, the opportunity of applying of a relatively new and scientifically strong classification technique such as Support Vector Machines, in road safety with real-time traffic data was only recently been explored.

It is noted, that a similar technique, namely, the Random Forests (RFs) are mainly used for prediction, but can also be used for data mining purposes. In this thesis, RFs were used in order to give a first insight on the potentially significant independent variables (Chapter 7). When a great number of candidate variables exists, **the RFs can show the relevant importance of variables** and therefore provide the researcher a first idea as to what variables should be selected for further consideration. Similarly to SVMs, the type of correlation (positive or negative) is not unveiled and consequently, these type of models are better off when used for preliminary analysis (and of course for prediction).

Lastly, notable attention should be given to the cusp catastrophe models, which were proposed as an alternative to the traditional statistical and other computational methods (e.g. SVMs) used throughout the thesis. In most cases, **cusp catastrophe models proved to describe accident severity and probability very accurately,** having better goodness of fit in comparison to linear models. A similarity observed between cusp catastrophe and computationally intelligent methods such as neural networks, is the difficulty that arises when attempting to interpret the effect of independent variables.

The cusp catastrophe considers a dynamic safety system and reveals existence of potential nonlinearities also whether small changes in control factors, may result in sudden transitions from safe to unsafe conditions and vice versa. A drawback of cusp catastrophe models is that they are deterministic in nature, and as such, may not be always appropriate to describe phenomena that are subject to random variations. Nevertheless, stochastic formulations have been proposed in literature to overcome this limitation. Moreover, there is no prior knowledge towards which variables should be included in the asymmetry and in the bifurcation factor. On the basis of this limitation, several tests were performed and the most appropriate variables which determine each control factor are identified.

Concluding, **cusp catastrophe may pave the ground for further introducing nonlinear relationships when modelling road safety, such as chaos theory**. However, even if the results of the cusp models may be considered very promising and fruitful, one has to be cautious before entirely relying on such approaches. The main reason is that a complex system is non-deterministic. On the contrary, a chaotic system remains deterministic. Consequently, in order to further develop cusp catastrophe (even after the stochastic formulations that were carried out in this thesis) and chaos theory modelling a strong step should be taken: the consideration of the safety system as chaotic and not as a complex system.

10.3 Critical synthesis of accident probability findings

The effect of traffic states was firstly explored in urban roads (Chapter 6), and the results revealed a positive association between occupancy and risk of accident occurrence. Therefore, if the respective severity results are considered, **it can be concluded that higher occupancy and congestion contribute to higher accident occurrence but to less severe accidents**. Since traffic states were constructed on the basis of both upstream and downstream detectors, the most hazardous situation was when transitions from very high to very low occupancy levels took place. This finding is consistent with the findings of Hossain and Muromachi (2013).

In addition**, application of other methodologies applied in the thesis, have produced consistent results regarding accident probability**. For example, Bayesian logistic regression demonstrated that variations in occupancy and flow increase significantly the risk of an accident as Chapter 7 indicates. These findings are consistent with previous studies, which suggest that traffic variations a positive impact on accident occurrence (Yu and Abdel-Aty, 2013a; Yu et al., 2013; Xu et al., 2013a; Hassan and Abdel-Aty, 2012b; Ahmed et al., 2012b). As stated in Elvik (2006), the increased complexity of such traffic situations may increase the probability of accident occurrence. Moreover, a link between congestion and aggressive behaviour is identified in literature (Hennessy and Wiesenthal, 1997; Shinar, 1998). There is also empirical evidence that stop and go actions that take place in congested environments have an association with increased accident probability of (Hanbali and Fornal, 1997).

It is interesting that accident occurrence (that is accident probability) is likely to be considered as a dynamic system**,** since the findings of Chapter 9 provide evidence to support the nonlinear nature of accident occurrence in urban roads. To be more specific, **sudden jumps from accidents to non-accidents (and vice versa) can take place due to small changes in flow, speed variation and maximum solar radiation**. It is noted that weather effects do not seem to affect accident occurrence through linear relationships but have an effect only when cusp models are used.

Lastly, another contribution of this thesis is the rare-events logit model, which was applied on the motorway dataset to explore accident probability. The method of stratified sampling was followed and three models were developed, indicating that the main risk factor for accident occurrence was the logarithm of average speed, with **lower mean speeds being more likely to result in accident occurrence.**

Summing up, although a few studies indicate that driving in congestion increases drivers' stress and leads to more aggressive behaviour (Hennessy and Wiesenthal, 1997), the thesis showed that **accident occurrence is more a matter of low congestion or variations in traffic flow.**

10.4 Critical synthesis of accident severity findings

Based on the results, a number of findings regarding accident severity were unveiled. Linear models (Chapters 6, 7 and 8) as well as non-linear models (Chapters 9) were used. It is noted that different definitions of severity were used, in order to have a more **broad understanding of the phenomenon**.

Firstly, **the effect of traffic states** on the basis of average and standard deviation of occupancy measured from downstream and upstream loop detectors is considered, indicating that **low occupancy levels are associated with higher accident severity**. Therefore a negative association between congestion and severity can be assumed to exist.

When the impact of individual traffic and weather parameters is considered, it was found that traffic parameters have mixed effects on severity. For example, the results of the mixed effects logit model, revealed a negative effect of the standard deviation of occupancy. One explanation for the negative effect of traffic, could be the potential existence of traffic oscillations in low speeds and thus occurrence of non-severe crashes. On the other hand, the coefficient of variation of flow had a mixed effect depending on the accident type, indicating the presence of heterogeneity. The finite mixture model revealed different effect of speed and flow variations depending on the latent class. It is therefore concluded that **the effect of variations in traffic parameters on accident severity varies across different road accidents in urban environments**, although in general, traffic variations are correlated with severe accidents (Yu and Abdel-Aty, 2014b).

On the contrary, **when motorway environments are explored**, the finite mixture model initially showed mixed effects of variations in occupancy on severity. However, the elasticity analysis showed positive association. In addition, average truck proportion and average flow were negatively associated with severe accidents, something that was indicated in urban roads as well. **Therefore, it can be concluded that more congestion leads to less severe accidents.** Accident type has a consistent significant effect on accident severity, both inside and outside urban environments, with collisions with fixed objects and run-off-road collisions being associated with more severe accidents. The hazardous nature of this type of collisions, was highlighted also by Mintsis and Pitsiava-Latinopoulou (1990), who examined rural roads in Greece. The findings of this PhD thesis suggest that such collisions increase risk of severe injuries in urban roads as well.

Regarding weather parameters, **no direct association between weather parameters and severity seems to exist**, both inside and outside urban environments, as was also found in Christoforou et al. (2010).

When cusp catastrophe models are applied in urban roads, the results justify well their usefulness. **Variations in speed, average flow upstream, accident type and a number of weather parameters such as average temperature, total rainfall, solar**

radiation and wind direction, were found to have a potential effect on the system dynamics.

Concluding, it is observed that some parameters have a consistent effect (accident type, low congestion), or mixed effects (traffic variations) on accident severity. It is interesting that **weather parameters were found statistically significant only when nonlinear relationships are used** to explore accident severity, being non-significant in all other linear models.

10.5 Critical synthesis of PTW accident probability findings

The thesis also investigated the probability that a PTW is involved in accident. Chapter 5 demonstrated how **time series traffic data can be used as input to SVM** models following Zhao (2012). Time series of traffic parameters, such as flow, speed and occupancy which were extracted from the closest upstream and downstream loop detectors were considered separately. This combined approach suggested in this thesis is considered promising when investigating the involvement of Powered-Two-Wheelers (PTW) in accidents that occur in urban roads. In this case, both original time series and Discrete Wavelet Transformed (DWT) time series performed relatively well, however **the original time series produced better results than transformed time series**. This implies that extracting features from time series may not be necessary. As Chapter 5 is dedicated to PTWs, it is noted that it was aimed to predict **PTW accident type** as well, **both original and DWT time series data resulted in good total classification accuracies** and performed better than in PTW accident involvement. In that case both original and transformed time series produced very similar results.

For the traffic states modelling approach, **it was found that PTWs are more likely to be involved in accidents when the occupancy is very high** (congestion). On the contrary, in very low levels of occupancy PTWs are less likely to be involved in accidents. This implies that that congestion is correlated with increased probability of PTW involvement in accidents.

Considering the findings of Chapter 7, it can be argued **that in urban environments, PTWs might be strongly affected by interaction with other motorized traffic**. Thus PTW accident occurrence seems to be a matter of behavioural interaction with other motorized traffic. This may be attributed to the fact that high flow variations and multi-vehicle collisions (except for rear-end collisions), were found to have a strong association with accidents involving a Powered-Two-Wheeler.

When PTW accident probability in urban motorways is investigated, some interesting underlying relationships were unveiled. Firstly, **a linear positive relationship between traffic flow and PTW accident probability** was found to exist. However, further analysis showed that non-linear relationships fit the data better. More specifically, **a polynomial (quadratic) relationship** was revealed.

Lastly, the cusp model seems capable of describing the probability that a PTW is involved in accident, when parameters like **average flow, speed variations and accident type are considered to govern the two control parameters**. Since the performance of cusp model is slightly worse compared to the rest of the cusp models of the thesis, it is suggested that the consideration of the accident mechanism (involving a PTW) as a nonlinear dynamic system must be done with care and be further justified. Consequently, it be concluded that in such cases, despite the existence of nonlinearities, the linearity of the safety system might be still preserved and further research is needed.

10.6 Critical synthesis of PTW accident severity findings

When traffic states are considered, the findings of the models of PTW accident severity were very similar to that of accident severity, **suggesting that uncongested traffic is correlated with more severe accidents.**

Findings from the cusp catastrophe models regarding PTW accident severity in urban roads, indicated that linear models are proved equally capable of describing this underlying phenomenon. The first two models of PTW accident severity show a strong evidence of nonlinearity, with several traffic and weather variables playing a consistently important role in determining the asymmetry and the bifurcation factors. On the other hand, the only significant predictor among the other traditional accident variables is the type of accident. Therefore, smooth changes in these parameters significantly affect the stability of the safety system, causing transitions from the existing low severity level to higher severity levels. Concerning the third model, the presence of cusp is not as strongly supported as in the previous two models, but the possibility for the presence of such nonlinear dynamics cannot be excluded. **Overall, PTW accident severity shows slightly stronger evidence for existence of nonlinearities** than accident severity.

In the motorway dataset, the models produced slightly different results for occupants of PTWs. For example, accident type was not found to have an impact on PTW occupant severity. **On the other hand, variations in traffic (measured by the coefficient of variations in speed) and average flow have mixed effects on severity** as well, depending on the latent class in which each observation belongs, but the elasticity analysis revealed a positive influence. Lastly, **engine size was negatively correlated with severity levels**.

It is interesting, that **weather parameters were not found to significantly affect injury severity of vehicle occupants in urban motorways**. The insignificance of weather parameters in motorways and in urban roads, may be attributed to the fact that weather parameters may not be linearly related with road safety indicators such as severity and probability. It is expected that complex non-linear relations exist as discussed in Chapter 8.

Chapter 11 Conclusion

11.1 Overview of the research

The effective treatment of road accidents and the improvement of road safety level is a major concern to societies due to the losses in human lives and the economic and social cost. Tremendous efforts have been dedicated by transportation researchers and practitioners to improve road safety. Recently, high resolution real-time traffic and weather data started to be used when analysing road safety in freeways. Regardless of modelling techniques, a major gap is that **very limited research has been conducted so far for urban roads**. Moreover, there is **no specific focus on Powered-Two-Wheelers (PTWs)**, which constitute a vulnerable type of road users and are affected by the interaction with other motorized traffic. Taking also into account the speeding and the manoeuvring capabilities of PTWs, investigation of PTW safety by incorporating traffic conditions would be of particular interest. It should be noted, that **an integrated methodology** was needed in order to understand accident probability and severity, due to the complex nature of these phenomena.

In that context, **the main research question of the thesis was whether and how traffic and weather parameters affect accident probability and severity in urban roads and urban motorways.** The thesis objectives were achieved through the utilization of high resolution traffic and weather data collected on a real-time-basis in order to conduct **a multi-faceted statistical exploration of accident probability and severity.**

To this end, **a number of main research activities** were carried out:

1) A literature review of relevant research.
2) Data collection and processing.
3) Statistical analysis of accident probability in urban roads and motorways.
4) Statistical analysis of accident severity in urban roads and motorways.
5) Consideration of PTWs in the aforementioned statistical analyses.

The **first research activity** included an extensive literature review, investigating the research topics examined: the effect of traffic and weather characteristics on road safety and afterwards the critical parameters of PTW behaviour and safety. More specifically, a systematic review of the effect of traffic and weather characteristics on road safety was conducted firstly, having a specific focus on recent studies featuring high resolution traffic and weather data. Then studies related to rider behaviour, PTW interaction with other motorized traffic, accident frequency, accident rates and accident severity were examined. The extensive literature review led to the identification of the research gaps and open research questions.

The **second research activity** concerns the data collection and processing. Empirical data have been collected for the period 2006-2011 to investigate the relationship between traffic, weather and other characteristics and road accidents. The road axes chosen were the Kifisias and Mesogeion avenues in Athens, Greece, mainly due to the fact that they had very similar characteristics. Secondary, Attica Tollway ("Attiki Odos") was also chosen to be investigated separately.

The required **accident data** were collected from the Greek accident database SANTRA provided by the Department of Transportation Planning and Engineering of the National Technical University of Athens, based on data collected by the Police and coded by the Hellenic Statistical Authority. A 6-year period was considered, from 2006 to 2011. **Traffic data** were extracted from the Traffic Management Centre (TMC) of Athens for Kifisias and Mesogeion avenues, and from the Traffic Management Centre of Attica Tollway for the urban motorway. **Weather** data were collected from the Hydrological Observatory of Athens (HOA), operated by the Laboratory of Hydrology and Water Resources Management of the National Technical University of Athens.

Data collection led to the **data processing**. In this step, data quality was ensured (e.g. false values of traffic measurements were removed). Concerning the **accident cases**, the raw 5-min traffic and the 10-min weather data were further aggregated into 1-hour intervals in order to obtain averages, standard deviations and so on, following a more **mesoscopic analysis approach**. For accident probability examination purposes, data from **non-accident cases** were also collected, following the usual procedure described in international literature (Abdel-Aty and Pande, 2005; Abdel-Aty et al., 2007; Ahmed and Abdel-Aty, 2012; Yu and Abdel-Aty, 2013a).

In order to achieve the aims of the thesis through the **aforementioned research activities (third, fourth and fifth activity)**, a set of statistical analyses were carried out:

f) combined utilization of time series data and machine learning techniques (Support Vector Machine models) to predict PTW accident involvement and PTW accident type (Chapter 5),
g) finite mixture cluster analysis to identify traffic states and then explore the effect of traffic states on accident probability, accident severity, PTW accident severity and PTW accident involvement (Chapter 6),
h) investigation of the effect of individual traffic and weather parameters on accident probability and severity, by applying Random Forests (to detect potential significant variables) and then by applying finite mixture and Bayesian logit models (Chapter 7),
i) development of finite mixture, Bayesian and rare-events logit models to explore the factors affecting accident probability and severity in Attica Tollway (Chapter 8) and
j) investigation of the feasibility of applying the cusp catastrophe theory to estimate accident probability and accident severity in urban roads (Chapter 9).

The overview of the methodological framework is demonstrated on Figure 11-1.

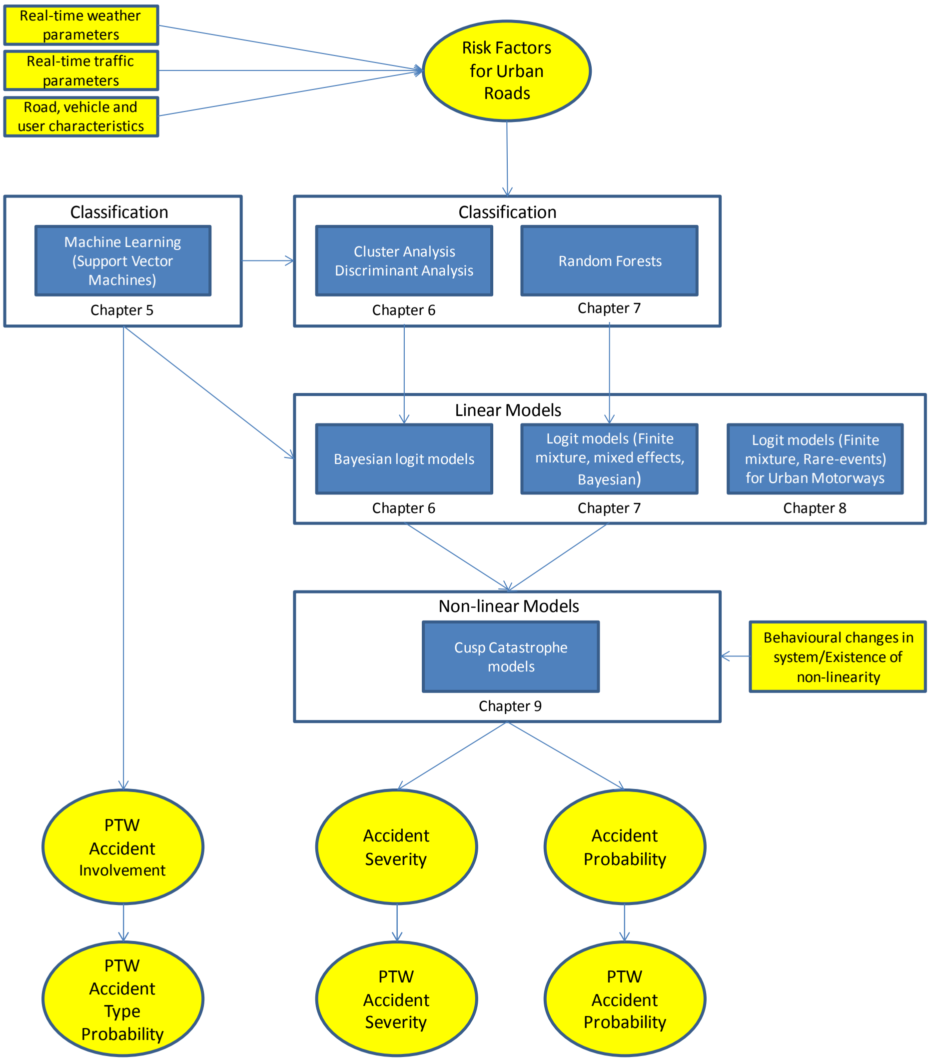


Figure 11-1: Overview of the methodological framework.

Since this PhD thesis deals with accident probability and accident severity having specific focus on Powered-Two-Wheelers. For that reason, separate PTW severity and probability models were developed. Different models and datasets were used throughout the thesis to achieve the aim of the research.

Chapter 5 **(machine learning)** is dedicated solely on PTWs, where PTW accident probability and PTW accident type are examined, with the use of a combined approach (Zhao, 2012). In this approach machine learning models such as Support Vector Machines were developed by utilizing original and transformed time series data.

Chapter 6 **(cluster analysis, discriminant analysis, Bayesian logit models)**, aimed to create meaningful groups of observations according to traffic characteristics (traffic states) on the basis of occupancy measurements. Then, the effect of group membership (traffic states) on accident probability and severity of all vehicle users as well as specifically PTW users was explored.

Chapter 7 **(random forests, finite mixture and Bayesian logit models)** as well as Chapter 8 **(finite mixture logit, Bayesian and rare-events logit models)** examine the effect of individual independent variables on accident probability and severity in urban roads and urban motorways respectively. Random forest models were first applied in order to acquire a preliminary analysis where potential significant variables are indicated. Afterwards, the significant variables were used as input to the finite mixture and Bayesian logit models in order to quantify the effect of traffic and weather variables on accident probability and severity. An innovative model, namely the rare-events logit model, was applied on urban motorways so as to explore accident probability when accidents are treated as rare events.

Lastly, in Chapter 9 **(cusp catastrophe)** advanced nonlinear cusp catastrophe models were developed. These advanced type of models consider the safety system as a dynamic system, explore the presence of nonlinear relationships and examine whether small changes significant variables can cause sudden and dramatic changes in the level of safety.

11.2 Conclusions

The present PhD thesis resulted in a number of original scientific contributions which are presented at the following sections. Section 11.2.1 demonstrates the main methodological contributions and conclusions, whilst the key research findings are presented on section 11.2.2. The original scientific contributions are the following:

i. Utilization of high resolution traffic and weather data in urban roads.
ii. Specific research focus on Powered-Two Wheelers in urban roads and motorways.
iii. Development of an integrated multifaceted approach to model accident probability and severity.
iv. Introducing advanced methods of analysis in exploring high resolution traffic and weather data and in road safety.
v. Investigating the existence of non-linear relationships when analysing accident probability and severity.

11.2.1 Methodological conclusions and contributions

Road safety is a complex scientific field and consequently the identification of relevant risk factors and the way which they influence road safety can be proved to be a very challenging and time consuming task. In this context, only a combination of different accident data modelling approaches can potentially provide a comprehensive explanatory picture of the underlying mechanisms.

The utilization of high resolution traffic and weather data in urban roads and the simultaneous co-consideration of Powered-Two-Wheelers, covered several gaps of knowledge, as indicated by the extensive literature review that was conducted. Due to the fact that the large majority of similar studies considered freeway data, it was needed to investigate urban environments. This is considered of great importance, since traffic and safety dynamics in freeway and urban environments are very different. For example, in urban environments the road users are more vulnerable to interactions with other motorized traffic and of course the presence of intersections plays a critical role. The focus on PTWs in such studies is essential, because they are very vulnerable to interaction especially in urban environments. Therefore, the effect of flow conditions on PTW safety had to be investigated.

This thesis proposed **an innovative approach to investigate accident probability and severity in urban roads and motorways with the use of differently oriented advanced modelling approaches** in order to acquire the larger picture of the accident severity and probability phenomena. For that reason, several probability and severity definitions were used. It was aimed to acquire the larger picture of accident severity and probability phenomena. Various data sources (e.g. real-time traffic data, real-time weather data and traditional accident data) have been obtained, processed and utilized. Although the core part of the thesis concerned urban roads, analyses of urban motorway data were also performed to complement the research design.

It is noted, that **some of the methods were applied for the first time** when such data are utilized (finite mixture logit, cusp catastrophe) or for the first time in road safety (rare-events logit model). Moreover, the time series data mining techniques through the combined application of original and transformed time series with Support Vector Machines, provided promising results and should be expanded in more relevant studies.

The mesoscopic accident analysis approach of this thesis, has a lot to contribute to the better understanding of the road accident phenomenon. To be more specific, this approach possesses some advantages over both macroscopic methods and the real-time microscopic data analysis, mainly because of the following reasons: 1) it enables the provision of sufficient time that allow authorities to develop a proactive safety management system without losing information of critical variables caused by large time interval measurement and 2) provides much more information than aggregate measures of traffic parameters (e.g. hourly traffic or annual average daily traffic). Modelling accident probability and severity in urban environments is a highly complex procedure and that fact should always be taken into serious consideration when analytical models are developed. The various statistical models that were developed can either predict or explain the accident probability and severity phenomena.

A methodological remark that derived from the analyses carried out in the thesis, **is the general superiority in terms of goodness of fit, of the non-linear modelling techniques** when accident severity and probability are examined. Although a number of linear models achieved to adequately describe the aforementioned phenomena, the cusp catastrophe models were proved to be considerably promising and fruitful. In a sense, these results may be considered as a first trial and a first step towards the incorporation of chaos theory in accident research. While one cannot definitely say that these methods outperform the traditional statistical analysis methods, it is without doubt that new directions are opened. It is very promising that the application of non-linear cusp catastrophe models, produced new original results and it is therefore suggested that more research should be conducted towards that direction in order to accurately predict the points of catastrophe.

A remark is worth of discussion. This concerns a number of similarities between cusp catastrophe and chaos theory, such as the presence of strong nonlinear relationships, the fact that control factors govern the system and the potential tremendous impact that small changes in control factors have on the system. Consequently, **these results may be used as a first step towards the incorporation of chaos theory in accident research**. While one cannot definitely say that these methods outperform the traditional statistical analysis methods, it is without doubt that new directions are opened.

Lastly, it is **suggested that an advanced multi-faceted statistical analysis of accident probability and severity exploiting high resolution traffic and weather data, can be proved as a very useful tool** for accident and injury causation analysis, but also for support of real-time road safety decision making.

11.2.2 Key research findings

This PhD thesis aimed to unveil the influence of high resolution traffic and weather parameters on accident probability and severity. Despite the emphasis given on such kind of data, traditional accident information was also used to enrich the interpretability of models. Overall, the findings of the thesis suggest that **high resolution traffic and weather data are capable of opening new dimensions in road accident analysis in urban roads and motorways**. In addition, the combination of traffic and weather data leads to a clearer picture of the road accident phenomenon, in terms of both probability and severity.

The multi-faceted statistical analysis conducted in the thesis has revealed a **consistent and strong influence of traffic parameters on accident probability and severity**. This finding suggests that similar accident studies and investigations should always consider and incorporate the traffic conditions before the occurrence of an accident. If effective real-time measures are implemented then accident probability and accident severity will be reduced. Nevertheless, the **statistical significance of some other specific accident attributes**, such as accident type, suggests that more data should be utilized, as they provide useful and important information.

It is without doubt that accident probability and severity and are two entirely distinct phenomena, which were found to be influenced by a number of common and non-common parameters. Each phenomenon was found to have different characteristics for different types of vehicles (passenger cars, Powered-Two-Wheelers) involved in the accident, but this does not always happen. For example, **the effect of traffic states on overall accident severity was found to be very similar to PTW accident**

severity (i.e. accidents were a PTW is involved). However, the importance of separate models for PTWs was justified in the rest of the chapters.

In general, it was found that **traffic parameters have mixed effects on accident severity**. For example, speed and flow variations had different effect on accident severity, depending on the latent class to which the accident was assigned (see table 7.1). Similar findings were revealed when accident severity and PTW accident severity were explored in the urban motorway.

When urban roads are analysed, accident occurrence with PTWs could be a matter of the behavioural interaction of PTWs with other motorized traffic, rather than PTW errors. This may be attributed to the fact that high fluctuations in traffic flow and multi-vehicle collisions (except for rear-end collisions), were found to have a strong association with accidents involving a PTW. On the other hand, PTWs are less likely to be involved in single-vehicle accidents. However, in urban motorways, PTW accident involvement was found to be correlated only with traffic flow and not with accident type. More specifically, **a non-linear relationship between traffic flow and accident with PTWs was observed in urban motorways**.

It is interesting that **weather parameters were not found statistically significant when linear relationships are considered.** This trend was observed regardless of the analysis method, the dependent variable of interest (i.e. severity or probability) or the area type (i.e. urban or motorway). However, the cusp catastrophe models indicated a strong significant effect of a number of weather parameters on the asymmetry and/or bifurcation factors, which determine the transition of safe to unsafe regimes and vice versa.

Lastly, the development of cusp catastrophe models implied that it is likely that **even small traffic and weather changes may have a critical impact on safe and unsafe traffic conditions in urban roads**. This regards not only overall accident probability and accident severity, but PTWs as well.

For example, severe accidents could be very easily turned into slight accidents in the future (and vice versa). Therefore, it should be further investigated if traditionally linear relationships are not appropriate in investigating accident probability and accident severity. When following this approach, the assumption that a dynamic system exists is required.

11.3 Future Research

It is recognised that the conducted thesis has a number of limitations that need further investigation. Similarly, some suggestions for further research are also demonstrated.

Concerning data issues, it would be interesting to **distinguish future analysis among lanes and among separated traffic regimes (peak and off-peak)**. Furthermore, loop detector traffic data were provided in 5-min intervals and then were aggregated in 1-hour intervals. Therefore, **more microscopic data** could be considered especially for road safety exploration in urban environments. For example, a number of analyses that were performed in this thesis, such as the combined approach of time series traffic data and the Support Vector Machines, could be performed when more microscopic data are used, i.e. 30sec traffic data.

Regarding road infrastructure and weather, if the **skid resistance and the precise pavement condition at the site and at the time of an accident are known**, then wet weather accidents could be further explained, especially with the utilization of real-time traffic and weather data. Consequently, more research is suggested to be conducted towards that direction.

It should be mentioned that the potential **spatial and temporal transferability** of the results should be tested in order to examine accident severity and probability in different major urban roads in different time periods. It is suggested that more recent data are collected in order to investigate the effect of traffic conditions under economic recession in Greece, since externalities of accidents during recession were are expected to be lower (Sotiriou et al., 2015).

Since the thesis examined accident probability and accident severity, it is suggested that **accident frequency should be explored in urban roads**, with both mesoscopic and microscopic traffic and weather data. In addition, rural areas should also be monitored and explored as suggested also by Wang et al. (2013a).

The fact that accident probability and severity are multidimensional and complex issues, raises the **need to incorporate human factors** in order to become more familiar with the parameters which eventually cause accidents. Besides, human factors have to be considered in many aspects of road safety (Kanellaidis 1996; Kanellaidis and Vardaki, 2010).

In terms of methodology, **the promising results of the rare-events logit models, raise the need to further enhance this approach by considering more accident data and by incorporating weather parameters**. The application of such approach on data from urban roads is suggested. Moreover, it is very important that rare-events modelling provides the capability of conducting accident probability analyses in sites or segments with limited number of accidents, because in such cases traditional logistic models are not appropriate.

Furthermore, it would be interesting to **apply different transformations on the time series**, such as Singular Value Decomposition (SVD), Discrete Fourier Transform, Piecewise Aggregate Approximation, Perpetually Important Points and so on. Aside from time series, the SVM models could be further applied in order to investigate if traditional traffic and weather variables in urban roads predict severity and probability.

Lastly, an interesting methodological remark concerns the fact that non-linear accident modelling is explaining far more accurately accident severity and probability in terms of goodness of fit. It is very promising, that the application of non-linear models such as the cusp catastrophe models, produced fruitful results and it is therefore suggested that more research should be conducted towards that direction in order to accurately predict the points of catastrophe. In addition, **the results of the cusp catastrophe models may be used as a first step towards the incorporation of chaos theory in accident research**.

List of references

Aare M., Von Holst, H. (2003). Injuries from motorcycle and moped crashes in Sweden from 1987-1999. Injury Control and Safety Promotion 10, 131-138.

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

Abdel-Aty, M., Abdelwahab, H. (2004). Modelling rear-end collisions including the role of driver’s visibility and light truck vehicles using a nested logit structure. Accident Analysis and Prevention 36(3), 447-456.

Abdel-Aty, M., Ekrama, A.-A., Huang, H., Choic, K. (2011). A study on crashes related to visibility obstruction due to fog and smoke. Accident Analysis and Prevention 43, 1730– 1737.

Abdel-Aty, M., Haleem, K., (2011). Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accident Analysis and Prevention 43, 461–470.

Abdel-Aty, M., Hassan, H.M., Ahmed, M., Al-Ghamdi, A.S. (2012). Real-time prediction of visibility related crashes. Transportation Research Part C 24, 288–298.

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

Abdel-Aty, M., Pande, A., Lee, C., Gayah, V., Dos Santos, C. (2007). Crash risk assessment using intelligent transportation systems data and real-time intervention strategies to improve safety on freeways. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 11(3), 107-120.

Abdel-Aty, M., Uddin, N., Pande, A. (2005). Split models for predicting multi-vehicle crashes during high-speed and low-speed operating conditions on freeways. Transportation Research Record 1908, 51-58.

Abdul Manan, M.M., Varhelyi, A. (2012). Motorcycle fatalities in Malaysia. IATSS Research 36(1), 30-39.

Ackaah, W., Afukaar, F., Agyemang, W., Anh, T.T., Hejar, A. R., Abdul, G., Yu, J. (2013). The use of non-standard motorcycle helmets in low- and middle-income countries: A multicentre study. Injury Prevention 19, 158-163.

Ahlm, K., Björnstig, U., Öström, M. (2009). Alcohol and drugs in fatally and non-fatally injured motor vehicle drivers in northern Sweden. Accident Analysis and Prevention 41, 129-136.

Ahmed, M., Abdel-Aty, M. (2012). The viability of using Automatic Vehicle Identification data for real-time crash prediction. IEEE Transactions of intelligent transportation systems 13(2), 459-468.

Ahmed, M., Abdel-Aty, M. (2013). Data fusion framework for real-time risk assessment on freeways. Transportation Research Part C 26, 203–213.

Ahmed, M., Abdel-Aty, M., Yu, R. (2012a). A bayesian updating approach for real-time safety evaluation using AVI data. Transportation Research Record 2280, 60–67.

Ahmed, M., Abdel-Aty, M., Yu, R. (2012b). Assessment of the interaction between crash occurrence, mountainous freeway geometry, real-time weather and AVI traffic data. Proceedings of the 91st Annual meeting of the Transportation Research Board, January 22-26, Washington, D.C.

Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B. (2011). Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis and Prevention 43, 1581-1589.

Aguero-Valverde, J., Jovanis, P.P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention 38, 618-625.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19. 716–723.

Albalate, D., Fernandez-Villadangos, L. (2010). Motorcycle injury severity in Barcelona: The role of vehicle type and congestion. Traffic Injury Prevention 11(6), 623-631.

Aldrich, E. (2010) Wavelets: A package of functions for computing wavelet filters, wavelet transforms and multiresolution analyses. Retrieved from: http://cran.rproject.org/web/packages/wavelets/index.html.

Al-Ghamdi, A.S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. Accident Analysis and Prevention 34, 729–741.

Al-Ghamdi, A.S. (2007). Experimental evaluation of fog warning system. Accident Analysis and Prevention 39, 1065–1072.

Aljanahi, A.A.M., Rhodes, A.H., Metcalfe, A.V. (1999). Speed, speed limits and road traffic accidents under free flow conditions. Accident Analysis and Prevention 31(1–2), 161–168.

Anastasopoulos, P.C., Mannering, F.L. (2009). A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41, 153-159.

Andersen, P.K., Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. Annual Statistics 10, 1100-1120.

Andreescu, M.-P., Frost, D.B. (1998). Weather and traffic accidents in Montreal, Canada. Climate Research 9, 225-230.

Andrey, J., Yagar, S. (1993). A temporal analysis of rain-related crash risk. Accident Analysis and Prevention 25(4), 465-472.

Antoniou, C., Yannis, G., Katsohis, D. (2013). Impact of meteorological factors on the number of injury accidents. Proceedings of the 13th World Conference on Transportation Research, COPPE - Federal University of Rio de Janeiro at Rio de Janeiro, Brazil, 15-18 July.

Baker, C.J., Reynolds, S. (1992). Wind-induced accidents of road vehicles. Accident Analysis and Prevention 24(6), 559-575.

Banfield, J.D., Raftery, A.E. (1993). Model-based Gaussian and non-Gaussian clustering. Biometrics 49, 803-821.

Barunik, J., Vosvrda, M. (2009). Can a stochastic cusp catastrophe model explain stock market crashes? Journal of Economic Dynamics and Control 33(10), 1824-1836.

Baruya, B., (1998). MASTER: Speed-accident relationship on European roads. Working Paper R 1.1.3, Deliverable D7, Technical Research Centre of Finland VTT, Espoo.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8. Retrieved from: http://CRAN.R-project.org/package=lme4.

Belmont, D.M., Forbes, T.W. (1953). Effect of average speed and volume on motor vehicle accidents on two-lane tangents. Proceedings of Highway Research Board 383-395.

Bensmail, H., Celeux, G. (1996). Regularized Gaussian Discriminant Analysis Through Eigenvalue Decomposition. Journal of the American Statistical Association 91, 1743-1748.

Bergel-Hayat, R., Debbarh, M., Antoniou, C., Yannis, G. (2013). Explaining the road accident risk: weather effects. Accident Analysis and Prevention 60, 456-465.

Bergel-Hayat, R., Depire, A. (2004). Climate, road traffic and road risk – an aggregate approach. Proceedings of the 10th World Conference on Transport Research (CD-ROM), World Conference on Transport Research Society, Instabul, Turkey.

Bijleveld, F., Churchill, T. (2009). The influence of weather conditions on road safety: An assessment of the effect of precipitation and temperature. Report R-2009-9, SWOV Institute for Road Safety Research, Netherlands.

Bjørnskau, T., Elvik, R. (1992). Can road traffic law enforcement permanently reduce the number of accidents? Accident Analysis and Prevention 24(5), 507-520.

Bjørnskau T., Nævestad, T.O., Akhtar, J. (2012). Traffic safety among motorcyclists in Norway: A study of subgroups and risk factors. Accident Analysis and Prevention 49 50-57.

Branas, C.C., Knudson, M.M. (2001). Helmet laws and motorcycle rider death rates. Accident Analysis and Prevention 33, 641-648.

Brandau, H., Daghofer, F., Hofmann, M., Spitzer, P. (2011). Personality subtypes of young moped drivers, their relationship to risk-taking behaviour and involvement in road crashes in an Austrian sample. Accident Analysis and Prevention 43, 1713-1719.

Breiman, L. (2001). Random Forests. Machine Learning 45, 5-32.

Breiman, L. (2000). Some infinity theory for predictor ensembles. Technical Report 579, Department of Statistics, University of California, Berkeley, US.

Brijs T., Karlis D., Wets, G. (2008). Studying the effect of weather conditions on daily crash counts using a discrete time-series model. Accident Analysis and Prevention 40, 1180–1190.

Brodsky, H., Hakkert, A.S. (1983). Highway accident rates and rural travel densities. Accident Analysis and Prevention 15(1), 73–84.

Brodsky, H., Hakkert, A.S. (1988). Risk of a road accident in rainy weather. Accident Analysis and Prevention 20(3), 161-176.

Broughton, P.S, Fuller, R., Stradling, S., Gormley, M., Kinnear, N., O'dolan, C., Hannigan, B. (2009). Conditions for speeding behaviour: A comparison of car drivers and powered two wheeled riders. Transportation Research Part F 12, 417-427.

Brown, D.B., Maghsoodloo, S., 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, 125-139.

Burrus, C.S., Gopinath, R.A., Guo, H. (1998). Introduction to wavelets and wavelet transforms: A Primer. Prentice-Hall, Inc.

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

Cavallo V., Pinto, M. (2012). Are car daytime running lights detrimental to motorcycle conspicuity? Accident Analysis and Prevention 49, 78-85.

Ceder, A. (1982). Relationships between road accidents and hourly traffic flow- II. Probabilistic Approach. Accident Analysis and Prevention 14(1), 35-44.

Ceder, A., Livneh, M. (1978). Further evaluation of the relationships between road accidents and average daily traffic. Accident Analysis and Prevention 10, 95-109.

Ceder, A., Livneh, M. (1982). Relationships between road accidents and hourly traffic flow-I. Analyses and Interpretation. Accident Analysis and Prevention 14(1), 19-34.

Celeux, G., Govaert, G. (1995). Gaussian parsimonious clustering models. Pattern Recognition 28, 781-793.

Cerwik, D.M., Gkritza, K., Shaheed, M.S., Hans, Z. (2014). A comparison of the mixed logit and latent class methods for crash severity analysis. Analytic Methods in Accident Research 3-4, 11-27.

Chang, H.L., Yeh, T.H. (2007). Motorcyclist accident involvement by age, gender, and risky behaviours in Taipei, Taiwan. Transportation Research Part F 10, 109-122.

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

Chang, L.Y., Chen, W.-C. (2005). Data mining of tree-based models to analyze freeway accident frequency. Journal of Safety Research 36, 365-375.

Chang, L.-Y., Wang, H.W. (2006). Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis and Prevention 38, 1019–1027.

Chen, C.F. (2009). Personality safety attitudes and risky driving behaviours - evidence from young Taiwanese motorcyclists. Accident Analysis and Prevention 41, 963-968.

Chen, C.F., Chen, C.W. (2011). Speeding for fun? Exploring the speeding behaviour of riders of heavy motorcycles using the theory of planned behaviour and psychological flow theory. Accident Analysis and Prevention 43, 983-990.

Cheng, A.S.K., Ng, T.C.K. (2010). Development of a Chinese motorcycle rider driving violation questionnaire. Accident Analysis and Prevention 42, 1250-1256.

Cheng, A.S.K., Ng, T.C.K, Lee, H.C. (2011). A comparison of the hazard perception ability of accident-involved and accident-free motorcycle riders. Accident Analysis and Prevention 43, 1464-1471.

Chimba, D., Sando, T. (2009). Neuromorphic prediction of highway injury severity. Advances in Transportation Studies 19(1), 17-26.

Chin, H.C., Quddus, M.A. (2003). Applying a random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis and Prevention 35, 253-259.

Chorlton, C., Conner, M., Jamson, S. (2012). Identifying the psychological determinants of risky riding: An application of an extended theory of planned behaviour. Accident Analysis and Prevention 49, 142-153.

Christoforou, Z., Cohen, S., Karlaftis, M. (2010). Vehicle occupant injury severity on highways: An empirical investigation. Accident Analysis and Prevention 42, 1606-1620.

Chung, Y.S., Wong, J.T. (2012). Beyond general behavioural theories: Structural discrepancy in young motorcyclist's risky driving behaviour and its policy implications. Accident Analysis and Prevention 49, 165-176.

Clabaux, N., Brenac, T., Perrin, C., Magnin, J., Canu, B., Van Elslande, P. (2012). Motorcyclists' speed and "looked-but failed- to-see" accidents. Accident Analysis and Prevention 49, 73-77.

Clarke, D.D., Ward, P., Bartle, C., Truman, W. (2007). The role of motorcyclist and other driver behaviour in two types of serious accident in the UK. Accident Analysis and Prevention 39, 974-981.

Cobb, L. (1980). Estimation theory for the cusp catastrophe model. Proceedings of the Section on Survey Research Methods, 772-776.

Cobb, L. (1981). Parameter estimation for the cusp catastrophe model. Behavioral Science 26, 75–78.

Cobb, L. (1998). An introduction to the cusp surface analysis. Technical Report, Aetheling Consultants, Louisville, CO, USA. Retrieved from: http://www.aetheling.com/models/cusp/Intro.htm.

Cobb, L., Koppstein, P., Chen, N. (1983). Estimation and moment recursion Relations for multimodal distributions of the exponential family. Journal of the American Statistical Association 78(381), 124-130.

Cobb, L., Ragade, R.K. (1978). Applications of Catastrophe Theory in the Behavioural and Life Sciences. Behavioral Science 23, 291-419.

Cobb, L., Watson, B. (1980). Statistical catastrophe theory: an overview. Mathematical Modelling 1(4), 311-317.

Cobb, L., Zacks, S. (1985). Applications of the catastrophe theory for statistical modeling in the biosciences. Journal of the American Statistical Association 80(392), 793-802.

Colgate, M.G., Tanner, J.C. (1967). Accidents at rural three-way junctions. RRL Report LR87. Crowthorne, U.K.: Transport and Road Research Laboratory.

Cortes, C., Vapnik, V. (1995). Support-vector network. Machine Learning 20, 1-25.

Creaser, J.I., Ward, N.J., Rakauskas, M.E., Shankwitz, C., Boer, E.R. (2009). Effects of alcohol impairment on motorcycle riding skills. Accident Analysis and Prevention 41 906-913.

Crundall, D., Bibby, P., Clarke, D., Ward, P., Bartle, C. (2008). Car drivers' attitudes towards motorcyclists: A survey. Accident Analysis and Prevention 40, 983-993.

Crundall, D., Crundall, E., Clarke, D., Shahar, A. (2012). Why do car drivers fail to give way to motorcycles at t-junctions? Accident Analysis and Prevention 44, 88- 96.

Crundall, D., Humphrey, K., Clarke, D. (2008). Perception and appraisal of approaching motorcycles at junctions. Transportation Research Part F 11, 159-167.

Daanen, H.A., van de Vliert, E., Huang, X. (2003). Driving performance in cold, warm, and thermoneutral environments. Applied Ergonomics 34(6), 597-602.

Dandona, R., Kumar, G.A., Dandona, L. (2006). Risky behaviour of drivers of motorized two wheeled vehicles in India. Journal of Safety Research 37, 149-158.

Daniello, A.L., Gabler, H.C. (2011). Fatality risk in motorcycle collisions with roadside objects in the United States. Accident Analysis and Prevention 43, 1167-1170.

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H., Brookhuis, K.A. (2011). Do young novice drivers overestimate their driving skills more than experienced drivers? Different methods lead to different conclusions. Accident Analysis and Prevention 43, 1660-1665.

De Lapparent, M. (2006). Empirical Bayesian analysis of accident severity for motorcyclists in large French urban areas. Accident Analysis and Prevention 38, 260-268.

De Pauw, E., Daniels, S., Thierie, M., Brijs, T. (2014). Safety effects of reducing the speed limit from 90 km/h to 70 km/h. Accident Analysis and Prevention 62, 426-431.

De Rome, L., Ivers, R., Haworth, N., Heritier, S., Du, W., Fitzharris, M. (2011). Novice riders and the predictors of riding without motorcycle protective clothing. Accident Analysis and Prevention 43, 1095-1103.

Dee, T.S. (2009). Motorcycle helmets and traffic safety. Journal of Health Economics 28, 398-412.

Delen, D., Sharda, R., Bessonov, M. (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis and Prevention 38(3), 434-444.

Dendrinos, D. (1978). Operating speeds and volume to capacity ratios: the observed relationship and the fold catastrophe. Transportation Research 12(3), 191-194.

Di Stasi, L.L., Álvarez -Valbuena, V., Cañas, J.J., Maldonado, A., Catena, A., Antolí, A., Candido, A. (2009). Risk behaviour and mental workload: Multimodal assessment techniques applied to motorbike riding simulation. Transportation Research Part F 12, 361-370.

Di Stasi, L.L., Contreras, D., Candido, A., Cañas, J.J., Catena, A. (2011). Behavioural and eye-movement measures to track improvements in driving skills of vulnerable road users: First-time motorcycle riders. Transportation Research Part F 14, 26-35.

Dickerson, A., Peirson, J., Vickerman, R. (2000). Road Accidents and Traffic Flows: An Econometric Investigation. Economica, London School of Economics and Political Science, 67(265), 101-21.

Donate-Lopez, C., Espigares-Rodríguez, E., Jimenez-Moleon, J. J., Luna-del-Castillo, J.D, Bueno-Cavanillas, A., Claret, P.L. (2010). The association of age, sex and helmet use with the risk of death for occupants of two-wheeled motor vehicles involved in traffic crashes in Spain. Accident Analysis and Prevention 42, 297-306.

Edwards, J.B. (1996). Weather related road accidents in England and Wales: a spatial analysis. Journal of Transport Geography 4(3), 201-212.

Edwards, J.B. (1999). Speed adjustment of motorway commuter traffic to inclement weather. Transportation Research Part F 2, 1-14.

Eisenberg, D. (2004). The mixed effects of precipitation on traffic crashes. Accident Analysis and Prevention 36, 637–647.

Eisenberg, D., Warner, K.E. (2005). Effects of Snowfalls on Motor Vehicle Collisions, Injuries, and Fatalities. American Journal of Public Health 95(1), 120-124.

Elliott, M.A. (2010). Predicting motorcyclists' intentions to speed: Effects of selected cognitions from the theory of planned behaviour, self-identity and social identity. Accident Analysis and Prevention 42, 718-725.

Elliott, M.A., Baughan, C.J., Sexton, B.F. (2007). Errors and violations in relation to motorcyclists' crash risk. Accident Analysis and Prevention 39, 491-499.

Elvik, R. (2006). Laws of accident causation. Accident Analysis and Prevention 38(4), 742-747.

Elvik, R. (2012). A review of game theoretic models of road user behaviour. Presented at the 25th ICTCT Workshop, November 8-9, 2012, Hasselt University, Hasselt, Belgium.

Elvik, R., Christensen, P., Amundsen, A. (2004). Speed and Road Accidents: an Evaluation of the Power Model. TOI report 740. Oslo: The Institute of Transport Economics.

European Road Safety Observatory, ERSO. (2011). Traffic safety basic facts “motorcycles and mopeds”. European Commission, Traffic Safety Basic Facts on Motorways, European Commission, Directorate General for Transport.

European Road Safety Observatory, ERSO. (2015a). Traffic Safety Basic Facts on Urban Areas. European Commission, Directorate General for Transport, June 2015.

European Road Safety Observatory, ERSO. (2015b). European Commission, Traffic Safety Basic Facts on Motorways, European Commission, Directorate General for Transport, June 2015.

European Transport Safety Council, ETSC. (2013). Back on track to reach the EU 2020 Road Safety Target? 7th Road Safety PIN Report.

El-Basyouny, K., Kwon, D.-W. (2012). Assessing time and weather effects on collision frequency by severity in Edmonton using multivariate safety performance functions. Proceedings of the 91st Annual meeting of the Transportation Research Board, January 22-26, Washington, D.C.

Filtness, A.J., Rudin-Brown, A.J., Mulvihill, C.M., Lenné, M. G. (2013). Impairment of simulated motorcycle riding performance under low dose alcohol. Accident Analysis and Prevention 50, 608-615.

Fraley, C., Murphy, T.B., Scrucca, L. (2012). mclust version 4 for R: Normal Mixture modeling for model-based clustering, classification, and density estimation. Technical Report No.597, Department of Statistics, University of Washington, USA.

Fraley, C., Raftery, A.E. (2002). Model-based clustering, discriminant analysis and density estimation. Journal of the American Statistical Association 97, 611-631.

Fraley, C., Raftery, A.E. (1999). MCLUST: Software for model-based cluster analysis. Journal of Classification 16, 297-306.

Fraley, C., Raftery, A.E. (1999). Enhanced software for model-based clustering, density estimation, and discriminant analysis: MCLUST. Journal of Classification 20, 263-286.

Frazier, C., Kockelman, K.M. (2004). Chaos theory and transportation systems: An instructive example. In Proceedings of the 83rd annual meeting of the Transportation Research Board, January 11-15 2004, Washington DC.

French, M.T., Gumus, G., Homer, J.F. (2009). Public policies and motorcycle safety. Journal of Health Economics 28, 831-838.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., Thomsen, L.K. (1995). Measuring the contribution of randomness, exposure, weather and daylight to the variation in road accidents. Accident Analysis and Prevention 27(1), 1-20.

Gabella, B., Reiner, K.L., Hoffman, R.E., Cook, M., Stallones, L. (1995). Relationship of helmet use and head injuries among motorcycle crash victims in El Paso County, Colorado, 1989-1990. Accident Analysis and Prevention 27(3), 363-369.

Gandhi, P.M., Colucci, B., and Gandhi, S.P. (1991). Polishing of Aggregates and Wet-Weather Accident Rates for Flexible Pavements. Transportation Research Record 1300, 71-79.

Garber, N.J., Gadirau, R. (1988). Speed variance and its influence on accidents. AAA Foundation for Traffic Safety, Washington, DC.

Garber, N.J., Subramanyan, S. (2002). Incorporating crash risk in selecting congestion-mitigation strategies: Hampton Roads area (Virginia) case study. Transportation Research Record 1746, 1–5.

Gelman, A. (2005). Analysis of variance - why it is more important than ever. The Annals of Statistics 33(1), 1- 53.

Geman, S, Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Gershon, P., Ben-Asher, N., Shinar, D. (2012). Attention and search conspicuity of motorcycles as a function of their visual context. Accident Analysis and Prevention 44(1), 97-103.

Gilmore, R. (1983). Catastrophe Theory for Scientists and Engineers. Dover, New York, NY, USA.

Goldenbeld, C., Twisk, D., De Craen S. (2004). Short and long term effects of moped rider training: A field experiment. Transportation Research Part F 7, 1-16.

Golias, J. (1992). Establishing relationships between accidents and flows at urban priority road junctions. Accident Analysis and Prevention 24(6), 689- 694.

Golob, T.F., Recker, W.W. (2002). Relationships among Urban Freeway Accidents, Traffic Flow, Weather and Lighting Conditions. Institute of Transportation Studies Report UCI-ITS-WP-02-1. University of California, Irvine, 2002.

Golob, T.F., Recker, W.W. (2003). Relationships Among Urban Freeway Accidents, Traffic Flow, Weather, and Lighting Conditions. Journal of Transportation Engineering 129(4), 342-353.

Golob, T.F., Recker, W.W., Alvarez, V.M. (2004). Freeway safety as a function of traffic flow. Accident Analysis and Prevention 36, 933–946.

Golob, T.F., Recker, W.W., Pavlis, Y. (2008). Probabilistic models of freeway safety performance using traffic flow data as predictors. Safety Science 46, 1306-1333.

Gould, M., Poulter, D., Helman, R.S., Wann, J.P. (2012). Errors in judging the approach rate of motorcycles in night-time conditions and the effect of an improved lighting configuration. Accident Analysis and Prevention 45, 432-437.

Graham, D.J., Glaister, S. (2003). Spatial variation in road pedestrian casualties: The role of urban scale, density and land-use mix. Urban Studies 40(8), 1591–1607.

Grasman, R.P.P.P., van der Maas, H.L.J., Wagenmakers, E-J. (2009). Fitting the cusp catastrophe in R: A cusp Package Primer. Journal of Statistical Software 32(8), 1-28. Retrieved from: http://www.jstatsoft.org/v32/i08/.

Greene, W.H., Hensher, D.A. (2003). A latent class model for discrete choice analysis: contrasts with the mixed logit. Transportation Research Part B 37(8), 681-698.

Gruen, B., Leisch, F. (2008). FlexMix Version 2: Finite mixtures with concomitant variables and varying and constant parameters Journal of Statistical Software 28(4), 1-35.

Guastello, S.J. (1982). Moderator regression and the cusp catastrophe-Application of 2-stage personnel selection, training, therapy and policy evaluation. Behavioral Science 27(3), 259-272.

Gwynn D.W. (1967). Relationship of accident rates and accident involvements with hourly volume. Traffic Quarterly 21(3), 407-418.

Haar, A. (1910). Zur zheorie der orthogonalen funktionensysteme. Mathematische Annalen 69(3), 331–371.

Haddon, W. Jr. (1980). Advances in the epidemiology of injuries as a basis for public policy. Public Health Report 95 411-421.

Haghighi-Talab, D. (1973). An investigation into the relationship between rainfall and road accident frequencies in two cities. Accident Analysis and Prevention 5, 343-349.

Hanbali, R., Fornal, C. (1997). Methodology for evaluating effectiveness of traffic responsive systems on intersection congestion and traffic safety. Transportation Research Record 1603(1) 137-149.

Haque, M.M., Chin, H.C., Debnath, A.K. (2012). An investigation on multi-vehicle motorcycle crashes using log-linear models. Safety Science 50, 352-362.

Haque, M.M., Chin, H.C., Huang, H. (2009). Modelling fault among motorcyclists involved in crashes. Accident Analysis and Prevention 41, 327-335.

Haque, M.M., Chin, H.C., Huang, H. (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis and Prevention 42, 203-212.

Haque, M.M., Chin, H.C., Lim, B.C. (2010). Effects of impulsive sensation seeking, aggression and risk-taking behaviours on the vulnerability of motorcyclists. Asian Transportation Studies 1(2), 165-180.

Harnen, S., Wong, S.V., Radin Umar, R.S., Wan Hashim, W.I. (2003). Motorcycle crash prediction model for non-signalized intersections. IATSS Research, 27(2), 58-65.

Harrison, W.A., Christie, R. (2005). Exposure survey of motorcyclists in New South Wales. Accident Analysis and Prevention 37, 441-451.

Hartelman, P.A.I. (1997). Stochastic Catastrophe Theory. Ph.D. thesis, University of Amsterdam, Amsterdam, the Netherlands.

Hassan, H.M., Abdel-Aty, M.A. (2013). Predicting reduced visibility related crashes on freeways using real-time traffic flow data. Journal of Safety Research 45, 29–36.

Hellenic Statistical Authority, ELSTAT. (2012). Retrieved from: http://www.statistics.gr/portal/page/portal/ESYE.

Hennessy, D.A., Wiesenthal, D.L. (1997). The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours. Ergonomics, 40(3), 348-361.

Hermans, E., Brijs, T., Stiers, T., Offermans, C. (2006). The Impact of weather conditions on road safety investigated on an hourly basis. Proceedings of the 85th Annual meeting of the Transportation Research Board, January 22-26, Washington, D.C.

Hiselius L.W. (2004). Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows. Accident Analysis and Prevention 36 (2004) 985–992.

Holm, A., Jaeger, M.M., Pedersen, M. 2009. Unobserved heterogeneity in the binary logit model with cross-sectional data and short panels: A finite mixture approach. Working paper no. 2009-04, Centre for Applied Microeconometrics, University of Copenhagen.

Hoogendoorn, S.P. (2007). Traffic flow theory and simulation. Lecture notes course vk4821. Transportation and Traffic Engineering Section. Delft University of Technology, Delft.

Horswill, M.S., Helman, S. (2003). A behavioural comparison between motorcyclists and a matched group of non-motorcycling car drivers: Factors influencing accident risk. Accident Analysis and Prevention 35, 589-597.

Horswill, M.S., Helman, S., Ardiles, P., Wann, J.P. (2005). Motorcycle accident risk could be inflated by a time to arrival illusion. Optometry and Vision Science 82(8), 740-746.

Hosking, S.G., Liu, C.C., Bayly, M. (2010). The visual search patterns and hazard responses of experienced and inexperienced motorcycle riders. Accident Analysis and Prevention 42, 196-202.

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

Hothorn, T., Hornik, K., Strobl, C., Zeileis, A. (2010). Party: A laboratory for recursive partytioning. Retrieved from: http://cran.r-project.org/web/packages/party/.

Houston, D.J. (2007). Are helmet laws protecting young motorcyclists? Journal of Safety Research 38, 329-336.

Houston, D.J., Richardson, L.E. (2008). Motorcyclist fatality rates and mandatory helmet-use laws. Accident Analysis and Prevention 40, 200-208.

Huang, W.S., Lai, C.H. (2011). Survival risk factors for fatal injured car and motorcycle drivers in single alcohol-related and alcohol-unrelated vehicle crashes. Journal of Safety Research 42, 93-99.

Hyatt, E., Griffin, R., Rue, L.W.III, McGwin, G. Jr. (2009). The association between price of regular-grade gasoline and injury and mortality rates among occupants involved in motorcycle- and automobile-related motor vehicle collisions. Accident Analysis and Prevention 41, 1075-1079.

Hydrological Observatory of Athens, HOA. (2014). Retrieved from: www.hoa.ntua.gr.

Iamtrakul, P., Tanaboriboon, Y., Hokao, K. (2003). Analysis of motorcycle accidents in developing countries: A case study of Khon Kaen, Thailand. Journal of the Eastern Asia Society for Transportation Studies 5, 147-162.

Ichikawa, M., Chadbunchachai, W., Marui, E. (2003). Effect of the helmet act for motorcyclists in Thailand. Accident Analysis and Prevention 35, 183-189.

Imai, K., King, G., Lau O. (2007). Zelig: Everyone's statistical software. Retrieved from: http://GKing.harvard.edu/zelig.

Imai, K., King, G., Lau O. (2008). Toward a common framework for statistical analysis and development. Journal of Computational and Graphical Statistics 17(4), 892-913.

International Traffic Safety Data and Analysis Group (IRTAD). (2011). Road Safety Annual Report. Retrieved from: http://internationaltransportforum.org/irtadpublic/pdf/11IrtadReport.pdf.

Ivan, J.N., Wang, C., Bernando, N.R. (2000). Explaining two-lane highway crash rates using land use and hourly exposure. Accident Analysis and Prevention 32, 787-795.

Jamson, S., Chorlton, K. (2009). The changing nature of motorcycling: Patterns of use and rider characteristics. Transportation Research Part F 12, 335-346.

Jellie, J. H. (2003). A study of Factors Affecting Skid Resistance Characteristics. Unpublished Thesis (DPhil), University of Ulster, Jordanstown, Northern Ireland.

Johansson, P. (1996). Speed limitation and motorway casualties: a time series count data regression approach. Accident Analysis and Prevention 28(1), 73-87.

Johnson, R.A., Wichern, D.W. (1998). Applied Multivariate Statistical Analysis. Prentice-Hall International, Inc., Upper Saddle River, NJ.

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

Jung S., Qin X., Noyce D.A. (2010). Rainfall effect on single-vehicle crash severities using polychotomous response models. Accident Analysis and Prevention 42, 213-224.

Kanellaidis, G. (1996). Human Factors in Highway Geometric Design. ASCE Journal of Transportation Engineering, 122(1), 59-66.

Kanellaidis G., Vardaki, S. (2010). Integrating safety into Highway Geometric Design. Proceedings of the 4th International Symposium on Highway Geometric Design, Valencia, June 2010.

Karatzoglou, A., Meyer, D., Hornik, K. (2006). Support Vector Machines in R. Journal of Statistical Software 15(9), 1-28.

Karatzoglou, A., Smola, A., Hornik, K., Zelies, A. (2005). Kernlab-Kernel Methods. R package, Version 0.6-2. Retrieved from: http://CRAN.R-project.org/.

Karlaftis, M., Vlahogianni, E. (2011). Statistical methods versus neural networks in Transportation Research: Differences, similarities and some insights. Transportation Research Part C 19, 387-399.

Karlaftis M., Yannis, G. (2010). Weather effects on daily traffic accidents and fatalities: A time series count data approach. Proceedings of the 89th Annual meeting of the Transportation Research Board, January 10-14, Washington, D.C.

Kasantikul, V., Ouellet, J.V., Smith, T., Sirathranont, J., Panichabhongse, V. (2005). The role of alcohol in Thailand motorcycle crashes. Accident Analysis and Prevention 37, 357-366.

Keay, K., Simmonds, I. (2005). The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. Accident Analysis and Prevention 37, 109–124.

Keay, K., Simmonds, I. (2006). Road accidents and rainfall in a large Australian city. Accident Analysis and Prevention 37, 109–124.

Keckman, V. (2005). Support vector machines-an introduction. In: Wang, L. (eds). Support Vector Machines: Theory and Applications, 1-48. Springer-Verlag, Berlin, Heidelberg, New York.

Khattak, A.J., P. Kantor, F.M. Council. (1998). Role of adverse weather in key crash types on limited-access: Roadways implications for advanced weather systems. In Transportation Research Record 1621, 10-15.

Kim, J.-K., Kim, S., Ulfarsson, G.F., Porrello, L.A. (2007). Bicyclist injury severities in bicycle–motor vehicle accidents. Accident Analysis and Prevention 39, 238-251.

King, G., Zeng, L. (2001a). Explaining rare events in international relations. International Organization 55(3), 693-715.

King, G., Zeng, L. (2001b). Logistic regression in rare events data. Political Analysis 9(2), 137-163.

Kleinbaum, D.G., Klein, M. (2010). Introduction to logistic regression, in: Logistic regression, Statistics for biology and health. Springer New York, 1-39.

Kleist, C. (2015). Time series data mining methods: A review. Master Thesis, Humboldt-Universitat zu Berlin, School of Business and Economics.

Kockelman, K.M., Ma, J. (2007). Freeway speeds and speed variations preceding crashes, within and across lanes. Journal of Transportation Research Forum 46(1), 43-61.

Koetse, M.J., Rietveld, P. (2009). The impact of climate change and weather on transport: An overview of empirical findings. Transportation Research Part D 14, 205–221.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th international joint conference on Artificial intelligence 2, 1137-1143.

Kononov, J., Bailey, B., Allery, B. (2008). Relationships between safety and both congestion and number of lanes on urban freeways. Transportation Research Record 2083, 26-39.

Kononov, J., Durso, C., Reeves, D., Allery, B.K. (2012). Relationship between traffic density, speed, and safety and its implications for setting variable speed limits on Freeways. Transportation Research Record 2280, 1-9.

Kuttesch, J.S. (2004). Quantifying the Relationship between Skid Resistance and Wet Weather Accidents for Virginia Data. Master thesis. Faculty of Virginia Polytechnic Institute and State University, Blacksburg, Virginia, U.S.

Langley, J., Mullin, B., Jackson, R., Norton, R. (2000). Motorcycle engine size and risk of moderate to fatal injury from a motorcycle crash. Accident Analysis and Prevention 32, 659-663.

Lassare, S. (1986). The introduction of variables "traffic volume", "speed" and "belt-wearing" into a predictive model of the severity of accidents. Accident Analysis and Prevention 18(2), 129-134.

Lave, C.A. (1985). Speeding co-ordination and the 55-miles/h limit. American Economic Review 75, 1159-1164.

Law, T.H., Noland, R.B., Evans, A.W. (2009). Factors associated with the relationship between motorcycle deaths and economic growth. Accident Analysis and Prevention 41, 234-240.

Lee, C., Hellinga, B., Saccomanno, F. (2003). Real-time crash prediction model for the application to crash prevention in freeway traffic. Proceedings of the 82nd Annual meeting of the Transportation Research Board, January 12- 16, Washington, D.C.

Leong, W.H.J. (1973). Relationship between accidents and traffic volumes at urban intersections. Journal of Australian Road Research Board, 5(3), 72- 90.

Leutzbach W. (1966). Traffic accidents and traffic flow on the Frankfurt- Mannheim motorway-investigation of the relationship between traffic accidents and traffic flow (in German). Hessisches Ministerium fur Wirtschaft und Verkehr, March 1966.

Levine, N., Kim, K.K., Nitz, L.H. (1995). Daily fluctuations in Honolulu motor vehicle crashes. Accident Analysis and Prevention 27(6), 785-796.

Li, M.D., Doong, J.L., Huang, W.S., Lai, C.H., Jeng, M.C. (2009). Survival hazards of road environment factors between motor-vehicles and motorcycles. Accident Analysis and Prevention 41, 938-947.

Li, Z., Liu, P., Wang, W., Xu, C. (2012). Using support vector machine models for crash injury severity analysis. Accident Analysis and Prevention 45, 478-486.

Li, X., Lord, D., Z., Y., Xie, Y. (2008). Predicting motor vehicle crashes using Support Vector Machine models. Accident Analysis and Prevention 40, 1611-1618.

Lin, M.R., Chang, S.H., Pai, L., Keyl, P.M. (2003). A longitudinal study of risk factors for motorcycle crashes among junior college students in Taiwan. Accident Analysis and Prevention 35, 243-252.

Lin, M.R., Kraus, J.F. (2008). Methodological issues in motorcycle injury epidemiology. Accident Analysis and Prevention 40, 1653-1660.

Lin, M.R., Kraus, J.F. (2009). A review of factors and patterns of motorcycle injuries. Accident Analysis and Prevention 41, 710-722.

Liu, C.C., Hosking, S.G., Lenné, M.G. (2009). Hazard perception abilities of experienced and novice motorcyclists: An interactive simulator experiment. Transportation Research Part F 12, 325-334.

Lord, D., Manar, A. Vizioli, A. (2005). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. Accident Analysis and Prevention 37, 185–199.

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

Lunn, D.J., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D. (2012). The BUGS Book: A practical introduction to Bayesian Analysis, 1st ed., Boca Raton, FL., Chapman and Hall/CRC.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10, 325-337.

Mackie, R.R., O'Hanlon, J.F. (1977). A study of the combined effect of extended driving and heat stress on driver arousal and performance. In: Mackie, R.R. Vigilance: Theory, Operational Performance and Physiological Correlates. New York, USA: Plenum Press: 537–558.

Maestracci, M., Prochasson, F., Geffroy, A., Peccoud, F. (2012). Powered two-wheelers road accidents and their risk perception in dense urban areas: Case of Paris. Accident Analysis and Prevention 49, 114-123.

Magazzu, D., Comelli, M., Marinoni, A. (2006). Are car drivers holding a motorcycle license less responsible for motorcycle - car crash occurrence? A non-parametric approach. Accident Analysis and Prevention 38, 365-370.

Majdzadeh, R., Khalagi, K., Naraghi, K., Motevalian, A., Eshraghian, M.R. (2008). Determinants of traffic injuries in drivers and motorcyclists involved in an accident. Accident Analysis and Prevention 40, 17–23.

Majdzadeh, R., Khalagi, K., Naraghi, K., Motevalian, A., Eshraghian, M.R. (2008). Determinants of traffic injuries in drivers and motorcyclists involved in an accident. Accident Analysis and Prevention 40, 17-23.

Malyshkina N.V., Mannering, F.L. (2009). Markov switching multinomial logit model: An application to accident-injury severities. Accident Analysis and Prevention 41, 829–838.

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

Mannering, F.L., Grodsky, L.L. (1995). Statistical analysis of motorcyclists' perceived accident risk. Accident Analysis and Prevention 27(1), 21-31.

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

Mayrose, J. (2008). The effects of a mandatory motorcycle helmet law on helmet use and injury patterns among motorcyclist fatalities. Journal of Safety Research 39, 429-432.

McCarthy, P. (1998). Effect of speed limits on speed distributions and highway safety: a survey of literature in managing speed. In: Review of Current Practice for Setting and Enforcing Speed Limits (No. Special Report 254). Purdue University, Department of Economics, Indiana.

McCarthy, P. (1993). The effect of higher rural interstate speed limits in alcohol related studies. Journal of Health Economics 11, 281-299.

McLeod, A. I., H. Yu and E. Mahdi (2012). Time series analysis with R. In Time Series Analysis: Methods and Applications, Chapter 23 (pp. 661-712) in Handbook in Statistics, Elsevier, Volume 30, Edited by T. S. Rao, S. S. Rao and C. R. Rao.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, H., Teller, E. (1953). Equation of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1092.

Meyer, D. (2001). Support vector machines. R News 1(3), 23-26.

Milton, J.C., Shankar, V.N., Mannering, F.L. (2008). Highway accident severities and the mixed logit model: An exploratory empirical analysis. Accident Analysis and Prevention 40, 260–266.

Mintsis, G., Pitsiava-Latinopoulou, M. (1990). Traffic accidents with fixed roadside obstacles: A study of Greek rural road network. Traffic Engineering and Control, 31(5), 306-311.

Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F. (2012). Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accident Analysis and Prevention 49, 58-72.

Moskal, A., Martin, J.L., Laumon, B. (2012). Risk factors for injury accidents among moped and motorcycle riders. Accident Analysis and Prevention 49, 5-11.

Morris, C.C. (2006). Generalized linear regression analysis of association of universal helmet laws with motorcyclist fatality rates. Accident Analysis and Prevention, 38, 142-147.

Musselwhite, C.B.A., Avineri, E, Susilo, Y.O., Bhattachary, D. (2012). Public attitudes towards motorcyclists' safety: A qualitative study from the United Kingdom. Accident Analysis and Prevention 49, 105-113.

Nakahara, S., Chadbunchachai, W., Ichikawa, M., Tipsuntornsak, N., Wakai, S. (2005). Temporal distribution of motorcyclist injuries and risk of fatalities in relation to age, helmet use, and riding while intoxicated in Khon Kaen, Thailand. Accident Analysis and Prevention 37, 833-842.

National Road Safety Observatory, NRSO. (2009). “Helmet use rate, Greece, 2009”. Retrieved from: http://nrso.ntua.gr/index.php/data/item/252-helmet-use-rate-greece-2009.html.

Navin, F. (1986). Traffic congestion catastrophe. Transportation Planning and Technology 11, 19-25.

Nikiforiadis, A.L. (2009). Vehicle restraint systems in roads. First edition, Gramma, Thessaloniki, GR.

Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety. Bulletin, 221. Lund Institute of Technology, Department of Technology and Society, Traffic Engineering, Lund.

Noland, R.B., Quddus, M.A. (2005). Congestion and safety: A spatial analysis of London. Transportation Research Part A 39, 737–754.

Oluwadiya, K.S., Kolawole, I.K., Adegbehingbed, O.O., Olasindee, A.A., Agodirina, O., Uwaezuoke, S.C. (2009). Motorcycle crash characteristics in Nigeria: Implication for control. Accident Analysis and Prevention 41, 294-298.

Oh, C., Oh, J-S., Ritchie, S.G., Chang, M. (2001). Real-time estimation of freeway accident likelihood. Presented at the Annual Meeting of the Transportation Research Board, 8–12 January 2001, Washington DC.

Ossiander, E.M., Cummings, P. (2002). Freeway speed limits and traffic fatalities in Washington State. Accident Analysis and Prevention 34, 13-18.

Ouellet, J.V., Kasantikul, V. (2006). Motorcycle helmet effect on a per-crash basis in Thailand and the United States. Traffic Injury Prevention 7(1), 49-54.

Ouellet, J.V., Smith, T.A., Thom, D.R., Kasantikul, V. (2002). Environmental contributing factors in Thailand motorcycle crashes. Institut fur Zweiradsicherheit (Hrsg.): Sicherheit Umwelt Zukunft IV.

Ozkan, T., Lajunen, T., Dogruyol, B., Yıldırım, Z., Coymak, A. (2012). Motorcycle accidents, rider behaviour, and psychological models. Accident Analysis and Prevention 49, 124-132.

Pai, C.W. (2009). Motorcyclist injury severity in angle crashes at T-junctions: Identifying significant factors and analyzing what made motorists fail to yield to motorcycles. Safety Science 47, 1097-1106.

Pai, C.W., Hwang, K.P., Saleh, W. (2009). A mixed logit analysis of motorists’ right-of-way violation in motorcycle accidents at priority T-junctions. Accident Analysis and Prevention 41, 565-573.

Pai, C.W., Saleh, W. (2007). An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK. Safety Science, 45, 832-847.

Pai, C.W., Saleh, W. (2008a). Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: Focusing on the effects of driver's failure to yield and junction control measures. Accident Analysis and Prevention 40, 479-486.

Pai, C.W., Saleh, W. (2008b). Modelling motorcyclist injury severity by various crash types at T-junctions in the UK. Safety Science 46, 1234-1247.

Papaioannou, P. (2007). Driver behaviour, dilemma zone and safety effects at urban signalised intersections in Greece. Accident Analysis and Prevention 39, 147-158.

Papaioannou, P., Peleka, M. (2006). Recent Experience on Success and Failure Stories from Funding Large Transportation Projects in Greece. Proceedings of the 1st International Conference on Funding Transportation Infrastructure, Banff, Alberta, Canada, August 2006.

Park, P.Y., Abdel-Aty, M. (2011). A stochastic catastrophe model using two-fluid model parameters to investigate traffic safety on urban arterials. Accident Analysis and Prevention 43, 1267-1278.

Paulozzi, L.J. (2005). The role of sales of new motorcycles in a recent increase in motorcycle mortality rates. Journal of Safety Research 36, 361-364.

Paulozzi, L.J., Ryan, G.W., Espitia-Hardeman, V.E., Xi, Y. (2007). Economic development's effect on road transport related mortality among different types of road users: A cross-sectional international study. Accident Analysis and Prevention 39, 606-617.

Poston, T., Stewart, I. (1996). Catastrophe Theory and Its Applications. London: Pitman.

Preusser, D.F., Williams, A.F., Ulmer, R. (1995). Analysis of fatal motorcycle crashes: Crash typing. Accident Analysis and Prevention 27(6), 845-851.

Quddus, M.A. (2013). Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. Journal of Transportation Safety and Security 5(1), 27-45.

Quddus, M.A., Noland, R.B., Chin, H.C. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research 33, 445-462.

Quddus, M.A., Wang, C., Ison, S.G. (2009). The impact of road traffic congestion on crash severity using ordered response models. Road traffic congestion and crash severity: econometric analysis using ordered response models. Journal of Transportation Engineering 136 (5), 424-435.

Quddus, M.A., Wang, C., Ison, S.G. (2010). Road traffic congestion and crash severity: econometric analysis using ordered response models. Journal of Transportation Engineering 136(5), 424–435.

R Development Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org.

Radin Umar, R.S., Mackay, M.G., Hills, B.L. (1996). Modelling of conspicuity-related motorcycle accidents in Seremban and Shah Alam, Malaysia. Accident Analysis and Prevention 28(3), 325-332.

Ragot-Court, I., Mundutéguy, C., Fournier, J.Y. (2012). Risk and threat factors in prior representations of driving situations among powered two-wheeler riders and car drivers. Accident Analysis and Prevention 49, 96-104.

Rathinam, C., Nair, N., Gupta, A., Joshi, S., Bansal, S. (2007). Self-reported motorcycle riding behaviour among school children in India. Accident Analysis and Prevention 39, 334-339.

Reeder, A.I., Chalmers, D.J., Langley, J.D. (1996). The risky and protective motorcycling opinions and behaviours of young on-road motorcyclists in New Zealand. Social Science and Medicine 42(9), 1297-1311.

Rifaat, M.R.S., Tay, R., De Barros, A. (2012). Severity of motorcycle crashes in Calgary. Accident Analysis and Prevention 49, 44-49.

Rock S.M. (1995). Impact of the 65mph speed limit on accidents, deaths and injuries in Illinois. Accident Analysis and Prevention 27(2), 207-214.

Rogers, M. P., Gargett, T. (1991). Skidding resistance standard for the National Road Network. Highways and Transportation, 38(4).

Rosenbloom, T., Perlman, A., Pereg, A. (2011). Hazard perception of motorcyclists and car drivers. Accident Analysis and Prevention 43, 601-604.

Rosser, J. (2007). The rise and fall of catastrophe theory applications in economics: Was the baby thrown out with the bathwater? Journal of Economic Dynamics and Control 31(10), 3255-3280.

Rutter, D.R., Quine, L. (1996). Age and experience in motorcycling. Accident Analysis and Prevention 28(1), 15-21.

Salt, G.F. (1977). Research on Skid-Resistance at the Transport and Road Research Laboratory (1927-1977). Transportation Research Record 622, 26-38.

Satterthwaite, S.P. (1976). An assessment of seasonal and weather effects on the frequency of road accidents in California. Accident Analysis and Prevention 8, 87-96.

Savino, G., Giovannini, F., Baldanzini, N., Pierini, M., Rizzi, M. (2013a). Assessing the potential benefits of the motorcycle autonomous emergency braking using detailed crash reconstructions. Traffic Injury Prevention, 14(sup1), S40-S49.

Savino, G., Pierini, M., Rizzi, M., Frampton, R. (2013b). Evaluation of an autonomous braking system in real-world PTW crashes. Traffic Injury Prevention, 14(5), 532-543.

Savolainen, P., Mannering, F. (2007). Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. Accident Analysis and Prevention, 39, 955-963.

Savolainen, P., Mannering, F., Lord, D., Quddus, M. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives Accident Analysis and Prevention 43, 1666–1676.

Schneider IV, W.H, Savolainen, P., Van Boxel, D., Beverley, R. (2012). Examination of factors determining fault in two vehicle motorcycle crashes. Accident Analysis and Prevention 45, 669-676.

Schwarz, G. (1978). Estimating the dimension of a model. The annals of Statistics 6, 461-464.

Schwebel, D.C., Severson, J., Ball, K., Rizzo, M. (2006). Individual difference factors in risky driving: The roles of anger/hostility, conscientiousness, and sensation-seeking. Accident Analysis and Prevention 38, 801-810.

Scott, P.P. (1986). Modelling time-series of British road accident data. Accident Analysis and Prevention 18(2), 109-117.

Shahar, A., Clarke, D., Crundall, D. (2011). Applying the motorcyclist's perspective to improve car drivers' attitudes towards motorcyclists. Accident Analysis and Prevention 43, 1743-1750.

Shahar, A., Poulter, D., Clarke, D., Crundall, D. (2010). Motorcyclists' and car drivers' responses to hazards. Transportation Research Part F 13, 243-254.

Shaheed, M.S., Gkritza, K. (2014). A latent class analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research 2, 30-38.

Shaheed, M.S.., Zhang, W., Gkritza, K., Hans, Z. (2011). Differences in motorcycle conspicuity related factors and motorcycle crash severities in daylight and dark conditions. In Proceeding of the Third International Conference on Road Safety and Simulation, September 11-16 2011, Indianapolis, IN.

Shankar, V., Mannering, F. (1996). An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of Safety Research 27(3), 183-194.

Shankar, V., Mannering, F., Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention 27(3), 371-389.

Shefer, D. (1994). Congestion, air pollution, and road fatalities in urban areas. Accident Analysis and Prevention 26(4), 501–509.

Shefer, D., Rietveld, P. (1997). Congestion and safety on highways: towards an analytical model. Urban Studies 34(4), 679-692.

Sherretz, L.A., Farhar, B.C. (1978). An analysis of the relationship between rainfall and the occurrence of traffic accidents. Journal of Applied Meteorology 17, 711-715.

Shew, C., Pande, A., Nuworsoo, C. (2013). Transferability and robustness of real-time freeway crash risk assessment. Journal of Safety Research 46, 83-90.

Shinar, D. and Compton, R., 2004. Aggressive driving: an observational study of driver, vehicle, and situational variables. Accident Analysis & Prevention 36(3), 429-437.

Smeed R.J. (1955). Accident rates. International Road Safety and Traffic Review 3(2), 30-40.

Smith, K. (1982). How seasonal and weather conditions influence road accidents in Glasgow, Scottish Geographical Magazine, 98(2), 103-114.

Solagberu, B.A., Ofoegbu, C.K.P., Nasir, A.A., Ogundipe, O.K., Adekanye, A.O., Abdur-Rahman, L.O. (2006). Motorcycle injuries in a developing country and the vulnerability of riders, passengers, and pedestrians. Injury Prevention, 12(4), 266-268.

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

Sotiriou M., Gavanas N., Pitsiava-Latinopoulou M. (2015). External costs due to congestion, accidents, energy consumption and emissions before and into the economic crisis: Pilot study along selected roadways of Thessaloniki, Greece. Spatium, 33, 33-40.

Spiegelhalter, D., Best, N., Carlin, B., Linde, V. (2003). Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society B 64 (4), 583–616.

Steg, L., Van Brussel, A. (2009). Accidents, aberrant behaviours, and speeding of young moped riders. Transportation Research Part F 12, 503-511.

Strobl, C., Malley, J., Tutz, G. (2009a). An introduction to recursive partitioning: Rational, application, and characteristics of classification and regression trees, Bagging, and Random Forests. Psychological Methods 14(4), 323-348.

Strobl, C., Malley, J., Tutz, G. (2009b). Supplement to ‘An introduction to recursive partitioning: Rational, application, and characteristics of classification and Regression trees, bagging, and random forests. Retrieved from: http://dx.doi.org/10.1037/a0016973.supp.

Strobl, C., Zeileis, A. (2008). Danger: High power! – exploring the statistical properties of a test for random forest variable importance. In Proceedings of the 18th International Conference on Computational Statistics, Porto, Portugal.

Struzik, Z.R., Siebes, A. (1999). The haar wavelet transform in the time series similarity paradigm. In Principles of Data Mining and Knowledge Discovery, pages 12-22. Springer.

Supramaniam, V., Van Belle, G., Sung, J.F.C. (1984). Fatal motorcycle accidents and helmet laws in Peninsular, Malaysia. Accident Analysis and Prevention 16(3), 157-162.

Sze, N.N, Wong, S.C. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accident Analysis and Prevention 39, 1267–1278.

Tanner, J. C. (1953). Accidents at rural three-way junctions. Journal of the Institution of Highway Engineers II(ll):56-67.

Taylor, M.C., Baruya, A., Kennedy, J.V. (2002). The relationship between speed and accidents on rural single-carriageway roads. TRL Report 511. Crowthorne, Berkshire: Transport Research Laboratory.

Taylor, M.C., Lynam, D.A., Baruya, A. (2000). The effects of drivers’ speed on the frequency of road accidents. TRL Report 421. Crowthorne, Berkshire: Transport Research Laboratory.

Teoh, E.R. (2011). Effectiveness of antilock braking systems in reducing motorcycle fatal crash rates. Traffic Injury Prevention 12(2), 169-173.

Teoh, E.R., Campbell, M. (2010). Role of motorcycle type in fatal motorcycle crashes. Journal of Safety Research 41, 507-512.

Theofilatos, A., Graham, D.J., Yannis, G. (2012). Factors affecting accident severity inside and outside urban areas in Greece. Traffic Injury Prevention, Volume 13(5), 458-467.

Thomson, G.A. (1979). The role frontal motorcycle conspicuity has in road accidents. Accident Analysis and Prevention 12, 165-178.

Tiwari, G. (2013). Motorcycle mobility and traffic risk. International Journal of Injury Control and Safety Promotion 20(2), 101-102.

Toroyan, T., Peden, M., Iaych, K., Krug, E. (2013). More action needed to protect vulnerable road users. The Lancet, 381(9871), 977-979.

Train, K. (2003). Discrete Choice Methods with Simulation. Cambridge University Press, New York.

Turner, D.J., Thomas, R. (1986). Motorway accidents: an examination of accident totals, rates and severity and their relationship with traffic flow. Traffic Engineering and Control 27(78), 377-387.

Tsui, K.L., So, F.L., Sze, N.N., Wong, S.C, Leung, T.F. (2009). Misclassification of injury severity among road casualties in police reports. Accident Analysis and Prevention 41, 84-89.

Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. Austrian Journal of Statistics, 31(3), 221-229.

Tunnicliff, D.J., Watson, B.C., White, K.M., Hyde, M.K., Schonfeld, C.C., Wishart, D.E. (2012). Understanding the factors influencing safe and unsafe motorcycle rider intentions. Accident Analysis and Prevention 49, 133-141.

Ulleberg, P., Rundmo, T. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. Safety Science 41, 427-443.

Usman, T., Fu, L., Miranda-Moreno, L.F. (2010). Quantifying safety benefit of winter road maintenance: Accident frequency modeling. Accident Analysis and Prevention 42, 1878-1887.

Usman, T., Fu, L., Miranda-Moreno, L.F. (2012). A disaggregate model for quantifying the safety effects of winter road maintenance activities at an operational level. Accident Analysis and Prevention 48, 368– 378.

Vaez, M., Laflamme, L. (2005). Impaired driving and motor vehicle crashes among Swedish youth: An investigation into drivers' sociodemographic characteristics. Accident Analysis and Prevention 37(4), 605-611.

Van der Maas, H.L.J., Kolstein, R., van der Pligt., J. (2003). Sudden transitions in attitudes. Sociological Methods & Research 23(2), 125–152.

Van Eslande, P., Elvik, R. (2012). Powered two-wheelers within the traffic system. Accident Analysis and Prevention, 49, 1-4.

Vapnik, V. (1998). Statistical learning theory. Wiley, New York, U.S.

Vernon, D.D., Cook, L.J., Peterson, K.J., Dean J.M. (2004). Effect of repeal of the national maximum speed limit law on occurrence of crashes, injury crashes, and fatal crashes on Utah highways. Accident Analysis and Prevention 36, 223–229.

Vidakovic, B, Muller, P (1991). Wavelets for kids, a tutorial introduction. Wavelet Digest, 1991.

Vitaliano, D.F., Held, J. (1991). Road accident external effects: An empirical assessment. Applied Economics 23, 373–378.

Vlahogianni, E., Karlaftis M., Orfanou, E. (2012c). Modelling the effects of weather and traffic on the risk of secondary incidents. Journal of Intelligent Transportation Systems, 16(3), 109-117.

Vlahogianni, E., Yannis, G., Golias. (2012a). Overview of critical risk factors in power-two-wheeler safety. Accident Analysis and Prevention 49, 12-22.

Vlahogianni, E., Yannis, G., Golias, J., Eliou, E. (2012b). Identifying outlying powered-two-wheeler riding behaviours at the emergence of an incident. In Proceedings of the 91st TRB Annual Meeting, January 22-26 2012, Washington DC.

Wagenmakers, E.J., Molenaar P.C.M., Grasman, R.P.P.P., Hartelman, P.A.I., van der Maas, H.L.J. (2005). Transformation invariant stochastic catastrophe theory. Physica D (211), 263-276.

Walker, G.H., Stanton, N.A., Salmon, P.M. (2011). Cognitive compatibility of motorcyclists and car drivers. Accident Analysis and Prevention 43, 878-888.

Walton, D., Buchanan, J. (2012). Motorcycle and scooter speeds approaching urban intersections. Accident Analysis and Prevention 48, 335-340.

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

Wang, C., Quddus, M.A., Ison, S.G. (2013a). The effect of traffic and road characteristics on road safety: A review and future research direction. Safety Science 57, 264-275.

Wang, C., Quddus, M.A., Ison, S.G. (2013b). Spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK. Transportmetrica A: Transport Science 9(2), 124-148.

Wanvik, P.O. (2009). Effects of road lighting: An analysis based on Dutch accident statistics 1987-2006. Accident Analysis and Prevention 41, 123-128.

Washington, S., Karlaftis, M., Mannering, F. (2011). Statistical and econometric methods for transportation data analysis. Second edition, Chapman and Hall/CRC, Boca Raton, FL.

Warren, D.L. (1982). Speed Zoning and Control. In: Synthesis of Safety Research Related to Traffic Control and Roadway Elements, Vol. II. Federal Highway Administration, Washington, DC, No. FHWA-TS-82–233.

Wegman, F., Aarts, L., Bax, C. (2008). Advancing sustainable safety: National road safety outlook for The Netherlands for 2005-2020. Safety Science 46, 323-343.

Williams, M.J., Hoffmann, E.R. (1979). Motorcycle conspicuity and traffic accidents. Accident Analysis and Prevention 11, 209-224.

Wilson, D.J. (2006). An Analysis of the Seasonal and Short-Term Variation of Road Pavement Skid Resistance. Unpublished thesis (DPhil), The University of Auckland, New Zealand.

Wong, J.T., Chung, Y.S., Huang, S.H. (2010). Determinants behind young motorcyclists' risky riding behaviour. Accident Analysis and Prevention 42, 275-281.

Wong, S.C., Sze, N.N., Lo, H.K., Hung, W.T., Loo, B.P.Y. (2005). Would relaxing speed limits aggravate safety? A case study of Hong Kong. Accident Analysis and Prevention 37, 377-388.

Woodcock, K. (2007). Rider errors and amusement ride safety: Observation at three carnival midways. Accident Analysis and Prevention 39, 390-397.

World Health Organization, WHO. (2013). Global status report on road safety 2013. Supporting a decade of action.

Xie, W., Zhao, K., Huynh, N. (2014). Analysis of driver injury severity in rural single-vehicle crashes. Accident Analysis and Prevention 47, 36-44.

Xu, C., Liu, P., Wang, W., Li, Z. (2012). Evaluation of the impacts of traffic states on crash risks on freeways. Accident Analysis and Prevention 47, 162– 171.

Xu, C., Tarko, A.P., Wang, W., Liu, P. (2013b). Predicting crash likelihood and severity on freeways with real-time loop detector data. Accident Analysis and Prevention 57, 30– 39.

Xu, C., Wang, W., Liu, P. (2013a). Identifying crash-prone traffic conditions under different weather on freeways. Journal of Safety Research 46, 135– 144.

Xuequn, Y., Ke, L., Ivers, R., Du, W., Senserrick, T. (2011). Prevalence rates of helmet use among motorcycle riders in a developed region in China. Accident Analysis and Prevention 43, 214-219.

Yamamoto, T., Shankar, V.N. Bivariate ordered response probit models of driver's and passenger's injury severities in collisions with fixed objects. Accident Analysis and Prevention 36, 869–876.

Yannis, G., Golias, J., Papadimitriou, E. (2005). Driver age and vehicle engine size effects on fault and severity in young motorcyclists accidents. Accident Analysis and Prevention 37, 327-333.

Yannis, G., Theofilatos, A., Ziakopoulos, A., Chaziris, A. (2014). Investigation of road accident severity and likelihood in urban areas with real-time traffic data. Traffic Engineering and Control 55(1), 31-35.

Yannis, G., Vlahogianni, E., Golias, J., Saleh, P. (2010). Road infrastructure and safety of powered-two wheelers. In Proceedings of the 12th World Conference on Transport Research, July 11-15 2010, Lisbon.

Yau, K.K.W. (2004). Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accident Analysis and Prevention 36, 333–340.

Yeh, T.H., Chang, H.L. (2009). Age and contributing factors to unlicensed teen motorcycling. Safety Science 47, 125-130.

Yu, R., Abdel-Aty, M., (2013a). Investigating the different characteristics of weekday and weekend crashes. Journal of Safety Research 46, 91–97.

Yu, R., Abdel-Aty, M., (2013b). Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis and Prevention 51, 252-259.

Yu, R., Abdel-Aty, M., (2014a). Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. Accident Analysis and Prevention 62, 161-167.

Yu, R., Abdel-Aty, M., (2014b). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety Science 63, 50-56.

Yu, R., Abdel-Aty, M., Ahmed, M. (2013). Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis and Prevention 50, 371-376.

Zambon, F., Hasselberg, M. (2006). Socioeconomic differences and motorcycle injuries: Age at risk and injury severity among young drivers: A Swedish nationwide cohort study. Accident Analysis and Prevention 38, 1183-1189.

Zambon, F., Hasselberg, M. (2007). Factors affecting the severity of injuries among young motorcyclists - a Swedish nationwide cohort. Traffic Injury Prevention 7(2), 143-149.

Zhao, Y. (2012). R and data mining: Examples and case studies. ISBN 978-0-12-396963-7, December 2012. Academic Press, Elsevier. Retrieved from: http://www.rdatamining.com/docs/RDataMining.pdf.

Zheng, Z., Ahn, S., Monsere, C.M. (2010). Impact of traffic oscillations on freeway crash occurrences. Accident Analysis and Prevention 42, 626– 636.

Zhou, M., Sissiopiku, V. (1997). Relationship between volume-to-capacity ratios and accident rates. Transportation Research Record 1581, 47-52.

Appendices

Appendix A

Author	Year	Country	Study area	Study Period	Dependent variable		Traffic variables considered								Summary of key findings
					Accident risk (frequencies, rates)	Accident Severity	Flow	Density	Occupancy	Speed	Speed variations	Speed limits	Congestion	V/C	
Ceder	1982	Israel	4-lane divided roadway sections.	1967-1975	•		•								For free flow data, total accident rates were related with hourly flow with a U-shaped curve for both single and multi-vehicle accidents. In congestion the multi-vehicle accident rate was found to have an abrupt increase with hourly flow.
Ceder and Livneh	1982	Israel	4-lane road sections.	1967-1975	•		•								Consistent results for single-vehicle accidents: Accident rates are inversely related with hourly flows. For multi-vehicle accidents a positive relationship was sometimes observed.
Brodsky and Hakkert	1983	USA	Urban and rural Interstates and primary highways. Rural secondary highways.	1976-1979	•			•							Moderate positive effect of density on U.S. Primary and Secondary highways, non-significant effect on the Interstate, negative association with fatality rates.
Caliendo et al.	2007	Italy	4-lane median-divided motorways.	1999-2003	•	•	•								Positive effect of AADT on total and severe accidents on curves and tangents. Wet pavements increase the number of severe accidents accidents on curves and tangents more than accidentes with property damage only.
Martin	2002	France	2- and 3-lanes Interurban motorways.	1997-1998	•	•	•								Property-damage accidents and injury rates reach their maximum in light traffic (under 400veh/h).
Dickerson et al.	2000	UK	London area.	1993-1995	•		•								Accident externality is almost zero for low or moderate flows, whist in increases sharply at high traffic flows.
Vitaliano and Held	1991	USA	Urban and rural roads in New York State.	1985	•		•								No significant externalities were found to exist.
Chang	2005	Taiwan	National Freeway 1 in Taiwan.	1997-1998	•		•								Accident frequency increases when the traffic per lane increases.
Milton et al.	2008	USA	Washington State's multilane divided highways.	1990–1994		•	•								Complex interaction between AADT and severity as in almost half of segments there is a positive relationship, while on the other segments there is a negative relationship. For 39% of the segments, increasing snowfall decreases the probability of property damage, for about 61% it increases the probability of property damage.
Anastasopoulos and Mannering	2009	USA	Rural interstate highways in Indiana.	1995-1999	•		•								Positive effect of AADT on accidents. However, because AADT was found to be a random parameters there is some variation among segments.
Zhou and Sissiopiku	1997	USA	A 26 km segment of Interstate I-94 in the Detroit area.	1993-1994	•	•							•	•	U-shaped relationships exist. In congestion, accident rates are high but accidents tend to be less severe.
Noland and Quddus	2005	UK	Greater London metropolitan area.	1999-2001		•							•		Accident severity does not seem to be affected by congestion.
Kononov et al.	2008	USA	Multilane freeways in Colorado, California and Texas.	5-year	•		•						•		Congestion increases the number of accident rates (fatal and injury rates).
Wang et al.	2009	UK	M25 London orbital motorway.	2004-2006	•								•		No or little impact of traffic congestion on accident frequency.
Quddus et al.	2010	UK	M25 London orbital motorway.	2003-2006		•							•		No impact of traffic congestion on accident severity.
Wang et al.	2013b	UK	M25 London orbital motorway.	2003-2007	•								•		Congestion is correlated with more accidents with killed and severely injured occupants but has little impact on slight injury accidents.
Ivan et al.	2000	USA	2-lane highways in Connecticut.	6-year period	•			•						•	The natural log of the segment volume to capacity ratio was found to have a negative relationship with accident rates.
Lord et al.	2005	Canada	Urban and rural freeway segments downtown and outside of Montreal, Quebec.	1994-1998	•		•	•						•	Overall, density and V/C ratio have an inverse U-shaped relationship with the number of accidents. This relationship does not describe multi-vehicle accidents. Separate models for single- and multi-vehicle accidentes should be developed.
Garber and Subramanyan	2002	USA	Interstate highways in Hampton.	1995-1998	•				•						U-shaped relationships exist.
Kononov et al.	2012	USA	Freeways in Colorado.	2001-2006	•		•	•							As density and flow increase, the accident rate remains constant until it reaches a certain critical value (threshold). If this threshold is exceeded, the accident rate rises rapidly
Nilsson	2004	Sweden	At national level.	1997-1999	•					•					Speed and accidents are related with a power-function.
Taylor et al.	2002	UK	Rural single-carriageway roads in England.	1992-1998	•					•					Positive relationship between average speed and accidents.
Quddus	2013	UK	13 motorways and 17 trunk A-class roads around London.	2003-2007	•					•	•				Average speeds are not associated with accidents, while speed variation is positively correlated with accidents.
Aljanahi et al.	1999	UK/Bahrain	Dual carriageways.	1988-1992	•					•	•	•			Mean speed and speed variations are positively associated with accidents. Speed limits reduce accidents.
De Pauw et al.	forthcoming	Belgium	Primary, Secondary and Local roads inside and outside urban areas in Limburg province.	2001-2002	•							•			Lower speed limits enhance road safety by decreasing fatal and serious accidents.
Rock	1995	USA	Rural highways in Illionois.	1983-1991	•							•			Increased speed limits led to an increase in the number of accidents.
Ossiander and Cummings	2002	USA	Rural freeways in Washington state.	1974-1994	•					•	•	•			Increased speed limits led to an increase in the number of fatal accidents.
Wong et al.	2005	Hong Kong	Major urban roadway sections.	1999-2002	•							•			Increased speed limits led to an increase in the number of all type of accidents (fatal, serious, slight).
Brown et al.	1991	USA	Rural interstates in Alabama.	1986-1988	•							•			Total number of accidents are unaffected by the increase in speed limits.
McCarthy	1993	USA	Highways in Indiana.	1981-1989	•							•			Increased speed limits also increased the number of alcoho-related accidents.
Johansson	1996	Sweden	Swedish motorways.	1982-1991	•							•			Reduced speed limits reduced the number of slight accidents (with minor injuries and/or property damage)
Vernon et al.	2004	USA	Urban and rural Interstates, rural non-Interstates, high-speed non-Interstate highways.	1992-1999	•							•			Increased speed limits did not have a major overall effect, but different effects in the various road types
Chang and Chen**	2005	Taiwan	National Freeway 1 in Taiwan.	2001-2002	•		•								Accidents are positively associated with increased ADT and annual precipitation.

**Weather effects are also investigated

Table A-1: Summary of traffic effect on safety.

Author	Year	Country	Study area	Study Period	Dependent variable: Accident risk (frequencies, rates)	Dependent variable: Accident Severity	Weather variables considered: Temperature	Rainfall	Number of rainy days	Sunlight	Dry spell*	Wind speed	Humidity	Low visibility	Snowfall	Summary of key findings
Caliendo et al.**	2007	Italy	4-lane median-divided motorways.	1999-2003	•	•		•								Positive effect of AADT on total and severe accidents on curves and tangents. Wet pavements increase the number of severe crashes, accidents on curves and tangents more than crashes with property damage only.
Milton et al.**	2008	USA	Washington State's multilane divided highways.	1990–1994		•									•	Complex interaction between AADT and severity as in almost half of segments there is a positive relationship, while on the other segments there is a negative relationship.For 39% of the segments, increasing snowfall decreases the probability of property damage, for about 61% it increases the probability of property damage.
Andrey and Jagar	1993	Canada	Cities of Calgary and Edmonton.	1979-1983	•			•								The overall accident risk during rainfall was 70% higher than normal conditions.
Levine et al.	1995	USA	City and County of Honolulu.	1990	•			•				•				Accident risk is substantially increased with rainfall.
Chang and Chen**	2005	Taiwan	National Freeway 1 in Taiwan.	2001-2002	•			•								Accidents are positively associated with increased ADT and annual precipitation.
Aguero-Valverde and Jovanis	2006	USA	Roads in the state of Pennsylvania.	1996-2000	•			•								Positive effect of rainfall in the Negative Binomial model, not significant in Full Bayes model.
Yannis and Karlaftis	2010	Greece	Urban area of Athens.	1985-2005	•		•	•			•					High precipitation (with its lagged effect) and low temperatures reduce accidents.
Bergel-Hayat et al.	forthcoming	France, Netherlands and Greece	Motorways, rural roads or urban roads.	1975-2005	•		•	•							•	Rainfall and temperature are positively correlated with the number injury accidents nationally. Negative correlation between precipitation and the number of injury accidents for the Athens region.
Shankar et al.	1995	USA	Interstate I-90 in Seattle.	1988-1993	•			•	•							Average daily rainfall in the month increases rear-end collisions, while the number of rainy days in the month decrease sideswipe and rear-end collisions but increase fixed object collisions.
Eisenberg	2004	USA	At national level.	1975-2000	•			•			•				•	There is a negative relationship between monthly precipitation and monthly fatal accidents. The opposite relatiosnhip exists for daily data. High risk imposed by precipitation as the time since last precipitation event increases.
Brodsky and Hakkert	1988	Israel and USA	At national level.	1979-1981, 1983-1984	•			•								Rainfall and rain lag increase the number of accidents.
Brijs et al.	2008	Netherlands	3 cities in Netherlands	2001	•		•	•		•	•	•				Rainfall and rainfall intensity increase the number of accidents. Negative non-linear relationship of accidents and temperature. Other weather variables were not significant.
Abdel-Aty et al.	2011	USA	All state roads in Florida.	2003-2007		•								•		Low visibility increases the level of severity.
Baker and Reynolds	1992	UK	At national level.	1990	•							•				Wind increases risk of overturning accidents involving mainly heavy vehicles.
Scott	1986	UK	Built up and non built-up A class roads and other roads,motorways.	1970-1978	•		•	•								Higher rainfall being associated with more accidents and warmer temperatures associated with fewer number of accidents.
Antoniou et al.	2013	Greece	Greater Athens area.	1997-2005	•		•	•								Low temperatures and increased rainfall reduce the number of accidents due to reduced mobility.
Hermans et al.	2006	Netherlands	Primary Dutch road system	2002	•		•		•	•		•	•			Increased wind gusts and longer duration of precipitation are among the most significant variables and they are correlated with increased number of accidents. Increased radiation and sunshine hours seem to have the same effect.
Eisenberg and Warner	2005	USA	At national level.	1975-2000		•									•	Snowy weather resulted in less fatal accidents than non-fatal and property damage accidents.
Keay and Simmonds	2006	Australia	Metropolitan area of Melbourne.	1987-1991, 1992-1996, 1997-2002	•			•			•					Rainfall consistently increases accident risk. The first rain after a dry spell has a large impact on safety especially when the duration of dry spell is long and the rainfall is high.
El-Basyouny and Kwon	2012	Canada	City of Edmonton.	2000-2010	•	•	•	•				•			•	For severe collisions, there was an inverse relationship with mean temperature and a positive relationship with total snowfall and total daily precipitation. For property-damage-only collisions, there was an inverse relationship with mean temperature and a significant positive relationship with total and previous snow fall and total daily precipitation
Andreescu and Frost	1998	Canada	Province of Quebec.	1990-1992	•		•	•							•	Increased snowfalls increase the number of accidents sharply. Rainfall has also a strong positive relationship with accidents.Little effect of air temperature.
Jones et al.	1991	USA	Interstate 5 and State Route 520, Seattle.	1987-1898	•			•						•	•	Very little effect of rainfall on accident frequencies. Wet surface significatly affects accidents.
Haghighi-Talab	1973	UK	12 inner London Boroughs and Huddersfield.	1966-1967	•			•								Rainfall increased accident rates especially during nightfall and in Spring months. Moderate and heavy rainfall have similar effects.
Fridstrøm et al.	1995	Nordic countries	Provinces in Nordic countries.	1975-1987	•			•	•	•					•	Rainfall increases the accidents, but snowfall has the opposite effect. First snowfall in winter is slightly significant and increases risk.
Malyshkina and Mannering	2009	USA	Selected Indiana interstate segments.	1995-1999		•	•	•				•		•	•	Severe and fatal accidents as well as minor accidents increase in adverse weather, but their proportion is approximately the same as in good weather.

*(days from the previous rainfall event-lag effect)

** Traffic effects are also investigated

Table A-2: Summary of weather effect on safety.

Author	Year	Country	Study area	Study Period	Accident probability	Traffic variables considered								Weather variables considered									Summary of key findings
						Flow	Density	Occupancy	Speed	Congestion	Speed variations	Flow variations	Occupancy variations	Temperature	Rainfall	Number of rainy days	Sunlight	Dry spell*	Wind speed	Humidity	Visibility	Snowfall	
Ahmed and Abdel-Aty	2012	USA	78 miles on the expressway network in Orlando	2008	•	•			•		•												Increased odds of an accident when the variation in speed increases and the average speed decreases, at the segment of the crash at 5–10 min before the crash occurrence.
Kockelman and Ma	2007	USA	6 California freeways	1998	•	•		•	•		•												Changes in 30-second speed prior to an accident occurrence were insignificant.
Abdel-Aty and Pande	2005	USA	Interstate-4 in the Orlando metropolitan area, Florida	1999	•	•		•	•		•												The logarithm of coefficients of variation in speed for five minute interval is the major risk factor.
Abdel-Aty et al.	2007	USA	Interstate-4 in the Orlando metropolitan area, Florida	1999-2002	•	•		•	•		•	•	•										Different results were produced for low and moderate-high speed regimes.
Yu and Abdel-Aty	2013	USA	I-70 in Colorado	2006-2011	•	•		•	•		•	•											Weekday crashes are more likely to occur in congestion, the weekend accidents mostly occur under free flow. For weekday crashes, high occupancy and low speed downstream increase the crash probability. The same effect has the downstream congested traffic with high volumes upstream.
Ahmed et al.	2012a	USA	Florida state road 417	2007-2009	•				•		•												High speed variation at the crash segment with decrease in the average speed in the downstream segment increase the risk of having accidents at this location.
Ahmed et al.	2012b	USA	I-70 in Colorado	2007-2009	•				•		•				•						•		For dry season, low average speeds and low visibility increase accident risk. For snow season, low visibility, high precipitation and speed variation increase the probability of accidents.
Xu et al.	2013a	USA	I-880N freeway in California	2008, 2010	•	•		•	•		•	•	•		•						•		In clear weather, the standard deviation of speed at the downstream station and the speed difference between upstream and downstream stations increase risk.The standard deviation of occupancy at downstream and the speed difference between upstream and downstream station increased risk in low visibility weather. The speed difference between upstream and downstream stations, the rainfall intensity, and the linear and quadratic terms of the occupancy at upstream station increased accident under rainy weather.
Abdel-Aty et al.	2012	USA	I-4 and I-95 freeways in Central Florida	2007-2009	•				•		•				•						•		Reduction in average speed and high standard deviation in speeds at the same segment (all 5-10min prior to the crash), increase the risk of visibility related accidents.
Hassan and Abdel-Aty	2013	USA	I-4 and I-95 freeways in Central Florida	2007-2009	•	•		•	•		•				•						•		In low visibility, average speed downstream, average speed upstream and average occupancy downstream (all 5-10min prior to the accident) increased risk. The logarithm of average occupancy downstream and the coefficient of variation of speed upstream (all 5-10min prior to a crash) increased risk in clear weather.
Zheng et al.	2010	USA	I-5 freeway in Portland.	2004-2007	•	•		•	•		•	•	•										Traffic oscillations as expressed by the standard deviation of speed increased the probability of rear-end accidents.
Hossain and Muromachi	2013	Japan	Tokyo urban metropolitan expressways.	2007-2009	•	•		•	•	•													For basic freeway segments high speed upstream and low speed downstream as well as movement from high occupancy locations to a much lower occupancy location, are the major risk factors. Level of congestion difference between downstream and upstream increase risk for almost all ramp areas.
Xu et al.	2012	USA	I-880 freeway in San Francisco Bay Area.	2010	•	•		•	•		•	•	•										Different effect of each of the 5 traffic states.
Xu et al.	2013b	USA	I-880 freeway in San Francisco Bay Area.	2008	•	•	•	•	•	•	•	•	•		•						•		Accident probability is high when the traffic density upstream, the speed variance upstream and/or downstream, volume difference between upstream and downstream station and the occupancy difference between upstream and downstream station are also high.Adverse weather conditions also increase accident probability. Upstream traffic density, downstream traffic volume and adverse weather reduce severity.

*(days from the previous rainfall event-lag effect)

Table A-3: Summary of the effect of real-time traffic and weather on accident probability.

Author	Year	Country	Study area	Study Period	Accident Severity	Traffic variables considered								Weather variables considered									Summary of key findings
						Flow	Density	Occupancy	Speed	Congestion	Speed variations	Flow variations	Occupancy variations	Temperature	Rainfall	Number of rainy days	Sunlight	Dry spell*	Wind speed	Humidity	Visibility	Snowfall	
Xu et al.**	2013b	USA	I-880 freeway in San Francisco Bay Area.	2008	•	•	•	•	•	•	•	•	•		•						•		accident probability is high when the traffic density upstream, the speed variance upstream and/or downstream, volume difference between upstream and downstream station and the occupancy difference between upstream and downstream station are also high.Adverse weather conditions also increase accident probability. Upstream traffic density, downstream traffic volume and adverse weather reduce severity.
Christoforou et al.	2010	France	Highway A4–A86 in Paris	2000-2002, 2006	•	•			•						•							•	High average traffic volume (measured per lane and over 6 minutes in vehicles) reduces severity. Average speed under very high traffic conditions increases severity.
Golob et al.	2008	USA	California highways	2001	•	•		•	•														Marginal effect of traffic conditions on accident severity.
Yu and Abdel-Aty	2014	USA	I-70 freeway in Colorado and in an urban expressway	2007-2009, 2010-2011	•	•			•		•										•		Large variance of speed prior to a accident increases severity. Low visibility increases severity. No influence of AADT.
Jung et al.	2010	USA	Wisconsin interstate highways	2004-2006	•	•		•	•		•	•	•	•	•				•				Increased rainfall intensity 15 minutes prior to the accident is more likely to result to the highest level of severity, but wind speed has the opposite effect.

*(days from the previous rainfall event-lag effect)
** Accident probability is also investigated

Table A-4: Summary of the effect of real-time traffic and weather on accident severity.

Author	Year	Country	Study area	Study Period	Accident frequency	Traffic variables considered								Weather variables considered									Summary of key findings
						Flow	Density	Occupancy	Speed	Congestion	Speed variations	Flow variations	Occupancy variations	Temperature	Rainfall	Number of rainy days	Sunlight	Dry spell*	Wind speed	Humidity	Visibility	Snowfall	
Yu and Abdel-Aty	2013	USA	I-70 in Colorado	2006-2011	•	•		•	•		•	•											Weekday accidents are more likely to occur in congestion, the weekend accidents mostly occur under free flow. For weekday accidents, high occupancy and low speed downstream increase the accident probability. The same effect has the downstream congested traffic with high volumes upstream.
Yu et al.	2013a	USA	I-70 freeway in Colorado	2010-2011	•	•			•					•	•					•			Logarithm of vehicle miles travelled had a positive effect on the number of accidents. High precipitation and low visibility increase accident occurrence.
Usman et al.	2010	Canada	2 highways and 2 patrol routes in Ontario	-	•	•	•		•					•	•				•		•		Reduced visibility increase accident frequency. No significant effect of temperature and precipitation.
Usman et al.	2012	Canada	31 different highway routes across Ontario, Canada.	2000-2006	•	•	•		•					•	•				•		•		Low visibility, poor road surface conditions, high winds and low temperatures are associated with increased number of accidents.

*(days from the previous rainfall event-lag effect)

Table A-5: Summary of the effect of real-time traffic and weather on accident frequency.

APPENDIX B

Author(s)	Year	Data						Method of analysis														Contributory risk factors
		questionnaire	simulator	experiment (virtual environment, video based, field)	direct observation	interview	police records	correlation tests	descriptive statics	ANOVA, ANCOVA, MANCOVA	factor analysis (PCA, exploratory, confirmatory FA)	cluster analysis	chi square test	discrete choice (logistic, multinomial)	structural modeling analysis	survival analysis	reliability analysis	log-linear	regression (hierarchical, generalized, linear, Poisson	in-depth analysis	other	
Cheng et al.	2011	•	•					•		•				•								aggressive driving violations, hazard perceptions
Steg and van Brussel	2009	•									•			•			•					errors, lapses, violations, positive attitudes towards speeding
Yeh and Chang	2009	•														•						age, family motorcycle ownership
Rutter and Quine	1996	•									•		•									age,willingness to break the law and violate the rules
Di Stasi et al.	2011			•						•												training, experience
Chung and Wong	2012	•										•	•									age, gender
Broughton et al.	2009	•											•									type of area, time of day
Wong et al.	2010	•									•				•							sensation seeking, amiability, impatience
Cheng and Ng	2010	•									•					•						past history of accidents
Haque et al.	2010a	•												•				•				past history of accidents, aggression, risk-taking
Elliott et al.	2007	•									•								•			traffic errors, control errors, speed violations, performance of stunts, safety equipment, age, exposure
Mannering and Grodsky	1995	•												•								exposure, riding above the speed limit, passing vehicles on the shoulder or passing between lanes of traffic
Liu et al.	2009		•											•							•	experience
Rosenbloom et al.	2011			•						•												motorcycle licence,past history of accidents
Chorlton et al.	2012	•									•								•			past behaviour, control and behavioural beliefs, attitudes, moral norm, normative belifs, age, self-identity, anticipated regret, training status, engine size
Schwebel et al.	2006	•		•							•								•			sensation seeking, anger-hostility
Özkan	2012	•									•						•					traffic errors, control errors, speed violations, performance of stunts, safety equipment, age, exposure
Chang and Yeh	2007	•										•		•								age, gender, experience, driving skills
Chen	2009	•													•							personality traits, attitudes toward traffic safety
Maestracci et al.	2012	•					•						•									anxiety, conspiquity, ptw lane changing, other users' lane changing, skidding and failure to give way
Elliott	2010	•						•											•			affective attitude, self identity, perceived group norm, group identification, interaction between perceived group norm and group identification
Woodcock	2007				•							•									•	mistakes, violations
Di Stasi et al.	2009	•	•					•		•												risky behaviour, saccadic velocity, subjective mental workload
Dandona et al.	2006					•							•									violations
Rathinam et al.	2007	•												•								experience, learning source of riding, speed, riding frequency and distance, aggressive behaviour, encounter with the police
Chen and Chen	2011	•									•				•							perceived enjoyment, concentration, sensation seeking, experience
Reeder et al.	1996	•											•									licence, convictions, exposure, alcohol
Hosking et al.	2010		•							•												experience, hazard response times
Tunnicliff et al.	2012	•						•											•			theory of planned behavior variables, perceived behavioural control, attitudes, sensation seeking
Brandau et al.	2011	•										•										neuroticism, risking personality, risking driving style, inatention, impulsivity
Creaser et al.	2009			•																	•	alcohol, task demand, time pressure, tolerance
Goldenbeld et al.	2004	•									•						•		•			training, experience
Walton and Buchanan	2012				•				•				•									speed, presence of car at t-intersection, free headway, location

Table B-1: Overview of PTW behaviour in terms of data, method of analysis and contributory risk factors.

Author(s)	Year	Data								Method of analysis										Contributory risk factors
		questionnaire	simulator	experiment (video based, picture-based, field)	police records	hospital records	national accident database	other database	road site observation	correlation tests	chi square	descriptive statics	ANOVA, ANCOVA, MANCOVA	factor analysis (PCA, exploratory, confirmatory FA)	regression (hierarchical, generalized, linear, Poisson etc	discrete choice (logistic, multinomial)	survival analysis	log-linear	in-depth	
Shahar et al.	2011	●											●		●					empathic/perceptual factors, spatial factors
Pai et al.	2009						●									●				type of area, lighting, age, gender
Gershon et al.	2012			●											●					type of area, PTW rider's outfit
Crundall et al.	2008a	●											●	●						gender, experience, negative attitudes, empathic attitudes, awareness of perceptual problems, spatial understanding
Horswill and Helman	2003	●		●					●							●				speed, hazard perception, sensation seeking, attitudes to riding/driving
Li et al.	2009				●	●		●								●	●			type of area, road type, location, speed
Haque et al.	2012				●													●		nightime riding, wet road surface, failure of drivers to notice ptws, failure to judge correctly the speed/distance, stop/waiting vehicles, location, interaction with opposing traffic
Shahar et al.	2010	●	●																	type of vehicle, age, experience, driving/riding skills, mindset
Clarke et al.	2007				●						●	●								age, experience, collision type, ptw type, loss of control, curvature, conspicuity
Crundall et al.	2012			●									●							dual drivers, experience, eye movements
Clabaux et al.	2012						●												●	location, area type
Crundall et al.	2008b	●		●									●							junction, distance, appraisal and perceptual errors
Cavallo and Pinto	2012			●									●							daytime running lights, distance
Gould et al.	2012		●										●							time of day, vehicle type, tri-headlight configuration
Radin Umar et al.	1996				●										●					week of the year, running headlights, fasting in Ramadhan, Balik Kampong culture
Williams and Hoffmann	1979				●							●								collision type, maneuver type, inadequate motorcycle visibililty, conspiquity
Ragot-Court et al.	2012	●									●	●								car driver's behaviour in events/type of driving/detection problems/internal condition
Horswill et al.	2005			●									●							vehicle type, speed, viewing times

Table B-2: Overview of PTW interaction with other motorized traffic in terms of data, method of analysis and contributory risk factors.

Author(s)	Year	Data							Method of analysis									Contributory risk factors
		questionnaire	interview	direct observations	police records	national accident database	other databases	hospital data	descriptive statistic	regression (hierarchical, generalized, linear, Poisson etc.)	discrete choice (logistic, multinomial)	ARIMAmodels	survival analysis	chi square	cluster analysis	other	in-depth	
Lin et al.	2003	•														•		age, past crash history, exposure, risk-taking level, alcohol consumption, traffic violations
Preusser et al.	1995				•	•			•									speed, alcohol, collision type
Haque et al.	2010b				•					•								type of intersection, road type, red light cameras, exposure
Teoh	2011				•	•			•									age, speed, alcohol, antilock brake system
Wanvik	2009					•			•									road lighting, weather conditions, road surface conditions
Harrison and Christie	2005	•													•	•		age, gender, riding patterns, exposure, skills
Paulozzi et al.	2005						•			•								gross national income per capita
Law et al.	2009					•				•								per capita GDP, infant mortality rate, medical care, political factors, helmet laws, motorcycles per capita, changes in road infrastructure and vehicle design
Hyatt et al.	2009					•	•					•						gasoline price, age of motorcycle, age of occupants, gender
Harnen et al.	2003						•			•								intersection, approach traffic flow, approach speed, lane width, number of lanes, shoulder width, land use
Moskal et al.	2012				•						•							age, gender, helmet, alcohol, leisure travel, licence
Teoh and Campell	2010				•	•				•	•							ptw type, speed, alcohol, age, gender
Paulozzi	2005				•				•									motorcycle age
Haque et al.	2009				•					•								time of day, wet road surface, location, road type, number of lanes, speed limit, number of occupants, engine capacity, age
Oluwadiyaa et al.	2009		•						•									risky behavior, chaotic traffic, road design faults
Ahlm et al.	2009							•			•					•		alcohol, pharmaceuticals, drugs, day of the week
Houston	2007				•	•				•								helmet laws, age, speed limit, alcohol limit, alcohol consumption, income per capita, motorcycles per capita
Morris	2006				•	•				•								seasonality, climate measures, helmet laws
Ichikawa et al.	2003							•	•									sex, age and occupational, time of crash, helmet act
Supramaniam et al.	1984				•					•								state, year,helmet laws
Kyrychenko and Mc Cartt	2006				•						•							helmet laws, age, gender
Branas and Knudson	2001				•	•				•								helmet laws, population density and temperature
Ouellet and Kasantikul	2006						•		•					•				helmet use
Dee	2009				•	•										•		helmet laws, population,state, time of day
Houston and Richardson	2008				•	•				•								helmet laws, population/density, speed limit, climate measures, per capita alcohol consumption, income per capita, motorcycles per 1000 people, age, education
Huang and Lai	2011				•	•							•					alcohol, age, crashing in fixed object, time of day, curvature, area type
Mayrose	2008				•				•									helmet laws, helmet use, gender, state, age
Kasantikul et al.	2005				•				•					•				alcohol, location, type of collision, gender, error, collision type, curves
French et al.	2009									•								helmet laws, education, speed limit, administrative license revocation
Keall and Newstead	2012					•					•							area of residence, age, protection aids, exposure, vehicle age
Daniello and Gabler	2011				•				•									collision with roadside fixed objects
Schneider IV et al.	2012				•	•					•							age, alcohol, riding without insurance, helmet, education, risk behaviour
Abdul Manan and Várhelyi	2012				•	•			•									area type, road type, age, gender, licence, collision type, helmet, time of day
Xuequn et al.	2011			•							•							helmet, road type, number of passengers, registration status, gender, weather
Bjørnskau et al.	2012	•									•			•				motorcycle type, age, engine size, speed, experience, risky behaviour, unsafe attitudes

Table B-3: Overview of PTW accident frequency/rates in terms of data, method of analysis and contributory risk factors.

Author(s)	Year	Data						Method of analysis										Contributory risk factors
		police records	national accident database	hospital data	local and state databases	interview	surveillance data	ordered response (logit, pro etc.)	multinomial	nested logit	logistic regression	poisson regression	log-linear	classification trees	bayesian method	chi square	non parametric tests	
Quddus et al.	2002	•						•										nationality, engine size, time of day, headlight, collisions pedestrians and stationary objects
Yannis et al.	2005	•											•					age, engine size
Pai	2009	•	•								•							age, engine size, collision partner, number of vehicles involved, weather, time of day, day of the week, speed limit, right-of-way violation, motorcycle manoeuvre
Majdzadeh et al.	2008	•				•					•						•	fire, cost of damage, collision type, weather conditions, gender, safety equipment
Zambon and Hasselberg	2006	•		•							•							age, socioeconomic status
Rifaat et al.	2012				•			•										street type, manoeuver type, truck, speed, alcohol
Pai and Saleh	2008a	•	•					•										age, gender, manoeuver type, engize size, exposure, control, number of invovled vehicles, type of collision, time of day, day of the week, month, weather, lighting, road type
Pai and Saleh	2008b	•	•					•										age, gender, engine size, collision type, crash partner, exposure, speed limit, control, lighting, weather, overtaking or changing lanes
Savolainen and Mannering	2007	•							•	•								alcohol, helmet, speed, collision type, roadway characteristics
de Lapparent	2006	•													•			gender, age, collision type, weather, helmet, engine size, time of day, intersection
Shankar and Mannering	1996	•							•									environmental factors, rider attributes, roadway conditions, vehicle characteristics
Montella et al.	2012	•	•											•				collision type, pavement conditions, surface level, weather, intersection, road alignment, curvature, area type, road type
Donate-López et al.	2010	•										•						age, gender, helmet use
Shaheed et al.	2011				•				•									crash type, lighting, failure to yield right-of-way
Langley et al.	2000			•		•					•							engine size, licence
Albalate and Fernández-Villadangos	2010	•						•										gender, speed, road width, alcohol, congestion
Zambon and Hasselberg	2007	•	•								•							alcohol, traffic environment, speed limit, type of crash, area type
Keng	2005	•									•							helmet, vehicle type, age, gender, time of day, area type, speed limit, weather, head/neck injuries
Gabella et al.	1995	•					•				•							helmet, age, alcohol, time of day, motorcycle damage, risky behaviour
Nakahara et al.	2005			•												•		helmet, age, alcohol, time of day

Table B-4: Overview of PTW severity in terms of data, method of analysis and contributory risk factors.