



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

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 **mentioned 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.

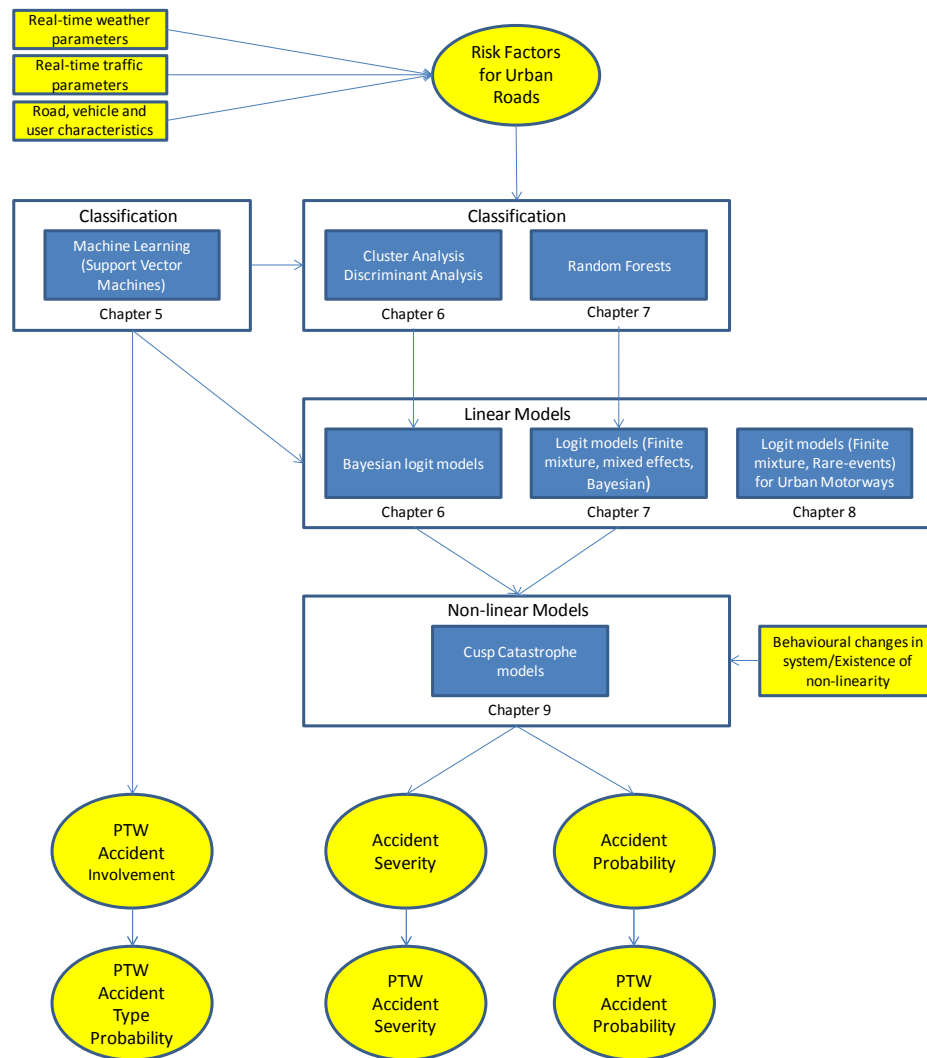


Figure 1: Overview of the methodological framework.

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.