DISS. ETH NO. 25382

Improving Driver Safety through the Identification, Prediction, and Warning of Traffic Accident Hotspots

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES OF ETH ZURICH

(Dr. sc. ETH Zurich)

Presented by:
Benjamin William RYDER

Master of Engineering in Computer Science, from Imperial College London

Born on 05.10.1988
Citizen of United Kingdom

Accepted on the recommendation of:
Prof. Dr. Elgar FLEISCH
Prof. Dr. Andreas HERRMANN
Ass. Prof. Dr. Felix WORTMANN

2018

"Start where you are. Use what you have. Do what you can."

Arthur Ashe

Abstract

Across the globe, injuries sustained from traffic accidents are the eighth leading cause of mortality, and with the number of annual deaths steadily rising to over 1.25 million, now account for 2.5 % of total worldwide fatalities. This growing issue is not limited to the low and middle-income regions of the world, as the frequency and severity of traffic accidents has also been increasing in developed countries over the last decades. For example, between 2014 and 2015 the amount of traffic fatalities in the United States rose sharply by 7.2 %. Through analysing the patterns and locations of traffic accidents, road authorities can identify dangerous sections of the road network and prioritise these locations for infrastructure improvement, helping to prevent these tragic events from occurring. However, traffic accident analysis is traditionally based on historic crash data and is restrictive in many ways, typically suffering from issues including small sample sizes, underreporting of traffic accidents, and data scarcity. Furthermore, in the limited number of countries where it is available, historical accident data is often only provided on a deferred basis and analyses can be severely out-of-date.

Naturalistic driving data, available from the advanced sensors and technology embedded in connected, semi-, and fully-autonomous vehicles, potentially offers both road safety researchers and practitioners a new and dynamic source of variables for analysis. The technology in these vehicles can be leveraged to detect accidents and 'near miss incidents', or critical driving events, such as heavy braking and evasive manoeuvres, and reliably predict locations with a high likelihood of traffic accidents. Both researchers and industry players alike see the promise of this data to combat the existing challenges of accident analyses. For example, real-time assessment of the locations of these events could aid road authorities in monitoring existing accident hotspots, as well as identifying new and developing areas of high accident exposure, offering various possibilities to intervene before incidents occur there. Yet, despite the great potential in identifying the locations of traffic accident hotspots with insights from these vehicles, to date, there is limited empirical evidence on whether the perilousness of locations can be accurately predicted through naturalistic driving data.

Furthermore, with these insights and the rise of increasingly connected and intelligent vehicles, as well as the emergence of smartphone turn-by-turn navigation applications, various safety-focused innovations become a possibility, such as providing safe-routing services and in-vehicle warnings of potential accident hotspots.

Whereas safe-routing will attempt to avoid an accident hotspot entirely, encountering these dangerous locations will always remain a possibility. Consequently, identifying ways of effectively reducing the frequency and severity of traffic accidents at these known locations remains of the utmost importance. Latest studies provide promising evidence that in-vehicle warning systems can have significant positive effects on driving behaviour and collision avoidance, and while the potential of these systems to improve driver safety are undisputed, the vast majority of studies have focused on simulation setups or controlled field experiments. Moreover, the benefit of real world location analytics on accident hotspots as a data source for in-vehicle warnings has widely not been investigated.

In order to address the aforementioned research gaps, a comprehensive intervention system was developed as part of the research at hand and deployed in a realistic field setting. This system both collected naturalistic driving data and provided warnings to drivers based on location analytics applied to a national historical accident dataset, composed of over 266 000 accidents. As such, this thesis depicts the design and field evaluation of an in-vehicle system, that for the first time bridges the gap between real world location analytics, accident hotspot warnings, and a naturalistic driving setting. The presented system was deployed in an 18 week nationwide field study of 72 professional drivers, covering over 690 000 km, and collected high frequency sensor data from the CAN Bus of each of the vehicles.

Ultimately, by going beyond existing research and exploring driver behaviour in a naturalistic driving setting, this thesis demonstrates that in-vehicle warnings of accident hotspots had a significant improvement on driver safety over time. First evidence is additionally provided that an individual's personality plays a key role in the effectiveness of such in-vehicle warning systems. However, in contrast to the promising results of existing lab experiments, an immediate effect on driver behaviour was not observed, further highlighting the importance of conducting experimental research in a realistic field setting.

This thesis additionally identifies the potential of driving data to reliably predict the locations of accident hotspots, assessed through a nationwide spatial regression to determine Crash Frequency across the majority of the Swiss road network. The results demonstrate a proportional relationship between Crash Frequency, and heavy braking events and trip frequency measurements from the field study fleet, along with additional explanatory variables for urban and highway locations. These insights provide initial indications that companies, organisations, and other players in the automotive industry with access to a fleet of connected, semi-, or fully-autonomous vehicles can determine existing and newly arising locations of high accident probability. Such a data-powered approach to road safety both enables the possibility for road authorities to intervene before traffic accidents occur at emerging dangerous locations, and empowers new safety-focused automotive services, such as the in-vehicle warnings that have been shown in this work to encourage safer driving through accident hotspots.

Zusammenfassung

Weltweit sind Verletzungen durch Verkehrsunfälle die achthäufigste Todesursache. Zudem stieg die Zahl der jährlichen Todesfälle stetig auf zuletzt über 1.25 Millionen, was circa 2.5 % der weltweiten Todesfälle entspricht. Dieses wachsende Problem beschränkt sich nicht nur auf die Regionen mit niedrigem und mittlerem Einkommen, da die Häufigkeit und Schwere von Verkehrsunfällen in den letzten Jahrzehnten auch in den Industrieländern zugenommen hat. So stieg die Zahl der Verkehrstoten in den USA zwischen 2014 und 2015 deutlich um 7.2 %. Durch die Analyse der Muster und Orte von Verkehrsunfällen können Straßenverkehrsbehörden gefährliche Abschnitte des Strassennetzes identifizieren, diese Orte für die Verbesserung der Infrastruktur priorisieren und so dazu beitragen, diese tragischen Ereignisse zu verhindern. Allerdings basiert die Analyse von Unfallschwerpunkten traditionell auf historischen Unfalldaten und ist in vielerlei Hinsicht beschränkt. So leidet sie in der Regel unter Problemen wie kleinen Stichprobengrößen, der unvollständigen Erfassung von Verkehrsunfällen und mangelnden Daten. Darüber hinaus werden historische Unfalldaten in der ohnehin begrenzten Anzahl von Ländern, in denen sie verfügbar sind, oft nur mit Verzögerung zur Verfügung gestellt, wodurch entsprechende Analysen nur schwer auf dem neusten Stand zu halten sind.

Reale Fahrdaten, die über hochentwickelte Sensoren und Technologien in vernetzten, halb- und vollautonomen Fahrzeugen erfasst werden, stellen im Bereich Verkehrssicherheit für die Forschung wie auch für die Praxis eine neue und dynamische Datenquelle für unterschiedliche Analysen dar. Die Technologie in diesen Fahrzeugen kann eingesetzt werden, um Unfälle und kritische Fahrereignisse bzw. 'Beinahe-Unfälle', wie z.B. schwere Brems- und Ausweichmanöver, zu erkennen und Orte mit einer erhöhten Unfallwahrscheinlichkeit zuverlässig vorherzusagen. Sowohl die Forschung als auch die Industrie sehen in diesen Daten das Potenzial, den bestehenden Herausforderungen der Unfallanalyse zu begegnen. Beispielsweise könnte die Echtzeit-Bewertung von Gefahrenstellen Straßenverkehrsbehörden helfen, bestehende Unfallschwerpunkte zu überwachen. Darüber hinaus könnten auch Orte mit steigendem Unfallpotential identifiziert werden. Diese Informationen bieten verschiedene Möglichkeiten zum Eingreifen, bevor es zu tatsächlichen Unfällen kommt. Dennoch gibt es trotz des großen Potenzials der Unfallschwerpunkteidentifizierung bisher nur begrenzt empirische Hinweise darauf, ob die Gefährlichkeit von Standorten durch kritische Fahrereignisse genau vorhergesagt werden kann.

Darüber hinaus werden mit diesen Erkenntnissen und dem Aufkommen von

zunehmend vernetzten und intelligenten Fahrzeugen sowie dem Aufkommen von Smartphone-Turn-by-Turn-Navigationsanwendungen verschiedene sicherheitsrelevante Innovationen möglich, wie beispielsweise die Bereitstellung von Safe-Routing-Diensten und die Warnung vor möglichen Unfallschwerpunkten im Fahrzeug. Während Safe-Routing versucht, einen Unfallschwerpunkt vollständig zu vermeiden, bleibt die Begegnung mit Gefahrenstellen immer eine Möglichkeit. Daher ist es nach wie vor von größter Bedeutung, Wege zu finden, um die Häufigkeit und Schwere von Verkehrsunfällen an diesen bekannten Orten wirksam zu reduzieren. Neueste Studien belegen, dass sich Warnsysteme im Fahrzeug deutlich positiv auf das Fahrverhalten und die Kollisionsvermeidung auswirken können. Und während das Potenzial dieser Systeme zur Verbesserung der Fahrsicherheit unumstritten ist, konzentrierte sich die überwiegende Mehrheit der Studien auf Simulationsaufbauten und kontrollierte Feldstudien. Ob sich die Ergebnisse dieser Studien jedoch auch auf die reale Welt übertragen lassen, blieb bisher weitgehend unerforscht.

Um die genannten Forschungslücken zu schließen, wurde im Rahmen der vorliegenden Thesis ein umfassendes Interventionssystem entwickelt und in einem realistischen Umfeld eingesetzt. Dieses System sammelte reale Fahrdaten und zeigte Warnungen für die Fahrer auf der Grundlage von Standortanalysen an, die auf einen nationalen, historischen Unfalldatensatz mit über 266 000 Unfällen angewandt wurden. In dieser Arbeit wird das Design und die Feldevaluation eines Fahrzeugsystems dargestellt, das zum ersten Mal die Lücke zwischen realer Standortanalyse, Unfall-Hotspot-Warnungen und einer realen Fahrsituation schließt. Das vorgestellte System wurde in einer 18-wöchigen landesweiten Feldstudie mit 72 Berufskraftfahrern auf über 690 000 km eingesetzt und sammelte hochfrequente Sensordaten aus dem CAN-Bus der einzelnen Fahrzeuge.

Zusammenfassend zeigt diese Arbeit, dass Warnungen vor Unfallschwerpunkten im Fahrzeug die Sicherheit des Fahrers im Laufe der Zeit deutlich verbessert haben. Sie geht dabei über die bestehenden Forschungsarbeiten hinaus, indem sie Fahrerverhalten in einem naturalistischen Setting untersucht. Zusätzlich wird gezeigt, dass die individuelle Persönlichkeit eine Schlüsselrolle für die Wirksamkeit solcher Warnsysteme im Fahrzeug spielt. Im Gegensatz zu den vielversprechenden Ergebnissen bestehender Laborexperimente wurde jedoch keine unmittelbare Auswirkung auf das Fahrerverhalten beobachtet, was die Bedeutung der Übertragung experimenteller Forschung auf einen realistischen Kontext unterstreicht.

Diese Arbeit identifiziert darüber hinaus das Potenzial von Fahrdaten zur zuverlässigen Vorhersage von Unfallschwerpunkten, basierend auf einer landesweiten räumlichen Regression zur Bestimmung der Unfallfrequenz und deckt dabei den Grossteil des Schweizer Strassennetzes ab. Die Ergebnisse zeigen einen proportionalen Zusammenhang zwischen Unfallfrequenz, starken Bremsvorgängen und Fahrfrequenzmessungen aus der Feldstudienflotte sowie zusätzlichen erklärenden Variablen für Stadt- und Autobahnstandorte. Die Erkenntnisse liefern erste Hinweise, dass Unternehmen, Organisationen und andere Akteure der Automobilindustrie mit Zugang zu einer Flotte von vernetzten, halb- oder vollautonomen Fahrzeugen bestehende und neu entstehende Unfallschwerpunkte identifizieren können. Der vorgestellte datenbasierte Ansatz für die Verkehrssicherheit ermöglicht Straßenverkehrsbehörden einzugreifen, bevor sich Verkehrsunfälle an aufkommenden Gefahrenstellen ereignen. Er ermöglicht außerdem neue, auf Sicherheit ausgerichtete Dienste, wie beispielsweise Warnungen im Fahrzeug, die, wie diese Arbeit zeigt, eine sicherere Fahrt durch Unfallschwerpunkte fördern.

Previous Publications

Parts of this thesis have already been published previously by myself and colleagues as scientific articles in peer-reviewed journals or in conference proceedings. While I am the first author of all of these documents and hereby declare that the majority of the content that has been integrated into this thesis has been written by myself, other co-authors have contributed to these documents with their reviews, changes, suggestions and edits. As a result, some sections of this thesis correspond literally to work previously published by me or bear strong similarities. Thus, the following publications are included in parts, or in an extended version, throughout this thesis:

- Benjamin Ryder, Andre Dahlinger, Bernhard Gahr, Peter Zundritsch, Felix Wortmann, and Elgar Fleisch (2018). "Spatial prediction of traffic accidents with critical driving events – insights from a nationwide field study". In: *Transportation Research Part A: Policy and Practice*. URL: `https://www.sciencedirect.com/science/article/pii/S0965856417310145`
- Benjamin Ryder, Bernhard Gahr, Philipp Egolf, André Dahlinger, and Felix Wortmann (2017). "Preventing traffic accidents with in-vehicle decision support systems – the impact of accident hotspot warnings on driver behaviour". In: *Decision Support Systems* 99, pp. 64–74. URL: `http://www.sciencedirect.com/science/article/pii/S0167923617300829`
- Benjamin Ryder and Felix Wortmann (2017). "Autonomously detecting and classifying traffic accident hotspots". In: *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, pp. 365–370. URL: `https://dl.acm.org/citation.cfm?id=3123199`
- Benjamin Ryder, Bernhard Gahr, and André Dahlinger (2016). "An in-vehicle information system providing accident hotspot warnings". In: *Proceedings of the 24th European Conference on Information Systems (ECIS)*. AIS. URL: `http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1000&context=ecis2016_prototypes`

Furthermore, the following publications were part of my Ph.D. research, but are outside the scope of this thesis:

- Bernhard Gahr, Benjamin Ryder, André Dahlinger, and Felix Wortmann (2018b). "Driver identification via brake pedal signals a replication and advancement of existing techniques (forthcoming)". In: *Proceedings of the 21st*

IEEE International Conference on Intelligent Transportation Systems. IEEE

- Bernhard Gahr, Benjamin Ryder, André Dahlinger, and Felix Wortmann (2018a). "A crowd sensing approach to video classification of traffic accident hotspots". In: *Proceedings of the 14th International Conference on Machine Learning and Data Mining*. Springer. URL: https://link.springer.com/chapter/10.1007/978-3-319-96133-0_14
- André Dahlinger, Felix Wortmann, Benjamin Ryder, and Bernhard Gahr (2018). "The impact of abstract vs. concrete feedback design on behavior - insights from a large eco-driving field experiment". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, p. 379. URL: https://dl.acm.org/citation.cfm?id=3173953
- Arne Meeuw, Sandro Schopfer, Benjamin Ryder, and Felix Wortmann (2018). "Lokalpower: enabling local energy markets with user-driven engagement". In: *Extended abstracts of the 2018 chi conference on human factors in computing systems*. ACM. URL: https://dl.acm.org/citation.cfm?id=3188610
- André Dahlinger, Felix Wortmann, Verena Tiefenbeck, Benjamin Ryder, and Bernhard Gahr (2017). "Feldexperiment zur wirksamkeit von konkretem vs. abstraktem eco-driving feedback". In: *Proceedings of the 13th International Wirtschaftsinformatik Conference*. AIS. URL: https://wi2017.ch/images/wi2017-0292.pdf
- Remo Manuel Frey, Benjamin Ryder, Klaus Fuchs, and Alexander Ilic (2016). "Universal food allergy number". In: *Proceedings of the 6th International Conference on the Internet of Things*. ACM, pp. 157–158. URL: https://dl.acm.org/citation.cfm?id=2998462
- André Dahlinger, Benjamin Ryder, and Felix Wortmann (2015). "Car as a sensor - paying people for providing their car data". In: *Proceedings of the 5th International Conference on Internet of Things*. IEEE. URL: http://www.iot-conference.org/iot2015/wp-content/uploads/2015/11/IoT2015_PS07_1570213122.pdf
- Benjamin Ryder, André Dahlinger, and Felix Wortmann (2015). "Leveraging controller area network data for predicting vehicle position during GPS outage". In: *Proceedings of the 5th International Conference on Internet of Things*. IEEE. URL: http://www.iot-conference.org/iot2015/wp-content/uploads/2015/11/IoT2015_PS05_1570213088.pdf

Acknowledgements

This thesis is result of my time at the Chair of Information Management at ETH Zürich, where I had the privilege to be part of the Bosch IoT Lab, a collaboration between the University of St. Gallen, ETH Zürich, and the Bosch Group. The valuable discussions, collaborations, and expertise of many teams and individuals within Bosch certainly helped to shape the work presented here, and I would like to start by thanking the Bosch Group for supporting me during my Ph.D. studies.

Above all, I would like to thank Prof. Dr. Elgar Fleisch for giving me the opportunity to explore the topic of data-powered road safety and for his supervision over the years. I am not only grateful for the insights and feedback that were provided on the project, which were always encouraging and thought-provoking, but also for the entrepreneurial guidance and the unique working and research environment that I found both refreshing and energising. In addition, I would like to thank Prof. Dr. Andreas Herrmann for his willingness to co-supervise my research, for supporting my work during the final phase of my thesis, and for the inspiration I gained from his recent book on autonomous driving.

Furthermore, I feel that it is not possible to fully express my gratitude to Ass. Prof. Felix Wortmann, Scientific Director of the Bosch IoT Lab, in just a few short lines here. Aside from the day-to-day guidance, constant enthusiasm, and motivation to push the project as far as possible together, I have greatly valued the career advice and publishing direction he has provided throughout the years. Both Prof. Dr. Markus Weinberger and Timo Gessmann also deserve a special mention and thanks for their support in directing the lab, encouraging our research, promoting our work within Bosch, and connecting us to industry partners. I would additionally like to thank Elisabeth Vetsch-Keller from the University of St. Gallen, and Monica Heinz and Judith Holzheimer of ETH Zürich, for their outstanding organisational support and assistance in times of need.

I believe that it is also important to highlight that the work presented in this thesis is the result of a group effort. Over my years at the Bosch IoT Lab I have had the pleasure to work alongside André Dahlinger and Bernhard Gahr in the Connected Car team, two exceptional colleagues without whom none of what is presented here would have been possible. Among many other things, I would like to thank them both for the unwavering dedication they showed to the project, the excitement and passion they bought with them, and most importantly the good times we shared. I am incredibly grateful to have been part of this team and eagerly look forward to our future endeavours together.

I am also thankful for the opportunity I had during my Ph.D. studies to supervise two excellent Master thesis students, Peter Zundritsch, who investigated the relationship between driving data and traffic accidents, and Philipp Egolf, who deservedly won the ETH Medal for his work on accident hotspot analysis.

This work would not have been possible without the contribution of various research partners. In particular I would like to thank the management team and drivers of Touring Club Suisse (TCS), whose commitment, support, and patience enabled the naturalistic driving field study to be a success. Additionally, I would like to thank the Swiss Federal Roads Office (FEDRO) Statistics For Road Accidents department, who kindly provided access to the traffic accident dataset which was foundational to my research.

Finally, the everyday effort of any venture is always eased when it is shared with, and supported by, good friends and colleagues. Over the years at the Bosch IoT Lab I have had the pleasure to enjoy the excellent company of Cotizo Sima, Arne Meeuw, Dominik Bilgeri, Shu Liu, Mathieu Chanson, Andreas Bogner, Dominic Wörner, Thomas von Bomhard, Paul Rigger, Marcus Köhler, Stefanie Turber, and Kristina Flüchter. I would also like to thank all of the incredible and passionate people across the Chair of Information Management at ETH Zürich, and the Institute of Technology Management at the University of St. Gallen, for the amazing times we have already enjoyed together, and those that I am sure we will share in the future.

Last but not least, although my family and friends from home often feel far away, they have remained an ever steadfast presence throughout my life. I also consider myself incredibly lucky to have enjoyed the last years of my studies with Chloé Chong, who has always brought happiness, laughter, and inspiration into every day we shared together. I can only thank them all for their enduring love, support, and guidance, no matter how far away I have travelled over the years, and dedicate this thesis to them.

Contents

Contents (detailed)

List of Figures

List of Tables

Glossary

ADT	**A**verage **D**aily **T**raffic
CAN	**C**ontroller **A**rea **N**etwork
CDE	**C**ritical **D**riving **E**vent
$\mathbf{CF_{Pit}}$	Number of accidents within location *i* from population *P* in timeframe *t*
$\mathbf{CR_{Pit}}$	Number of accidents per 100-million vehicle transits of population *P* in timeframe *t* for location *i*
DBSCAN	**D**ensity-**B**ased **S**patial **C**lustering of **A**pplication with **N**oise
FEDRO	Swiss **Fed**eral **R**oads **O**ffice
GDP	**G**ross **D**omestic **P**roduct
GPS	**G**lobal **P**ositioning **S**ystem
IMU	**I**nertial **M**easurement **U**nit
$\mathbf{JF_{Sit}}$	Number of high jerk events within location *i* from sample (fleet) *S* in timeframe *t*
$\mathbf{JR_{Sit}}$	Number of high jerk events per vehicle transit of sample (fleet) *S* in timeframe *t* for location *i*
OBD-II	**O**n-**B**oard **D**iagnostics
SARTC	**S**wiss **A**utomatic **R**oad **T**raffic **C**ounts
TCS	**T**ouring **C**lub **S**uisse
$\mathbf{TF_{Pit}}$	Number of transits of location *i* of population *P* in timeframe *t*
$\mathbf{TF_{Sit}}$	Number of transits of location *i* from sample (fleet) *S* in timeframe *t*
WHO	**W**orld **H**ealth **O**rganization

Chapter 1

Introduction

"Road traffic injuries are a major but neglected public health challenge that requires concerted efforts for effective and sustainable prevention. Of all the systems with which people have to deal every day, road traffic systems are the most complex and the most dangerous."

World report on road traffic injury prevention,
World Health Organization & World Bank, 2004

1.1 Context and Motivation

In 2004 the World Health Organization (WHO) and the World Bank jointly issued the first major report on road traffic injury prevention, claiming that the existing level of road traffic injury is unacceptable and a largely avoidable issue (World Health Organization and World Bank, 2004). Among other findings, the report concluded that unsafe road traffic systems seriously harm global public health and development, and projected that without fresh commitments to road safety there would be a 65 % increase in traffic fatalities by the year 2025, reaching almost 2 million annual deaths. During 2013 and 2015 the WHO issued new reports on the global status of road safety, and while the previous projections of a dramatic increase in traffic accidents appear to have been averted, the number of road traffic fatalities has continued to steadily increase (World Health Organization, 2013a; World Health Organization, 2015). In fact, the Global Health Estimates report indicates that road traffic crashes are the leading cause of death among young people, i.e. those aged between 15 and 29 years, and the 8th leading cause among all age groups (World Health Organization, 2018). While road injuries have long been the primary cause of death by injury, this new data now suggests that they account for 2.5 % of total worldwide fatalities. As such, it seems clear that the call for urgent action in the WHO global status reports on road safety is more relevant than ever if the 2015 commitment of the United

Nations to halve the global number of deaths and injuries from road traffic crashes by 2020 is to be reached (United Nations, 2015).

Beside the growing humanitarian concerns of so many fatalities worldwide, approximately 50 million people each year suffer from non-fatal injuries due to traffic accidents (World Health Organization, 2015). In countries struggling with other development needs this is an even greater issue, since the scale of the problem created by traffic accidents is rarely inline with the typical investment in road safety. When combined, the economic impact of the consequences related to traffic accident injuries, such as the cost of medical treatment or supporting a family member who cannot work (Jacobs, Aeron-Thomas, and Astrop, 2000; Prinja et al., 2015), and the damages to vehicles and infrastructure, create a significant burden on a country's health, insurance, and legal systems. Overall, the worldwide economic costs of traffic accidents are estimated to be 3% of the global Gross Domestic Product (GDP) (McMahon and Dahdah, 2008), and data suggests that in certain countries these economic losses can reach up 5% of the GDP.

As such, the importance of appropriately tackling the growing issue surrounding traffic accidents is being recognised on a global level. The previously mentioned global status reports on road safety from the WHO present information from over 180 countries, accounting for almost 99% of the world's population (World Health Organization, 2013a; World Health Organization, 2015). On the recommendations of the WHO, progress is being made towards improving both vehicle safety and road traffic legislation. For example, in 2013, just 28 of the countries had detailed safety laws covering fundamental road risk factors (World Health Organization, 2013a). In 2015, a further 17 countries had updated at least one of their laws with recommendations related to these key components of driver safety. Yet despite this progress, the recent data highlights that more needs to be done to prevent traffic accidents from occurring (World Health Organization, 2015), Thus, there are various governmental initiatives and research studies targeted towards reducing the amount of both fatal and non-fatal traffic accidents on our roads. For example, after a sharp increase in traffic fatalities in the United States of 7.2% between 2014 and 2015, the White House and the United States Department of Transportation issued a joint call to action (U.S. Department of Transportation, 2016). In a step to help combat this rising epidemic, the Department of Transportation released an open dataset that contained detailed information about each of the traffic incidents of that year. The two institutions additionally encouraged the continuous research of transportation scholars and data scientists into the different approaches and insights that could help road authorities improve the situation.

Through analysing the patterns and locations of traffic accidents, one of the responsibilities of governmental road authorities is to identify dangerous sections of the road network and prioritise these locations for infrastructure improvement, helping to prevent these tragic events from occurring. However, traditional traffic accident analysis is typically based on historic crash data, such as the dataset issued by the Department of Transportation, and is restrictive in many ways, typically suffering from issues including small sample sizes, underreporting of traffic accidents, and data scarcity (Mannering, Kilareski, and Washburn, 2007). Furthermore, in the limited number of countries where it is available, historical accident data is often only provided on a deferred basis and analyses can be severely out-of-date.

A promising solution to the described drawbacks lies in the post-processing and aggregation of driving data available from the advanced sensors and technology embedded in connected, semi-, and fully-autonomous vehicles. Naturalistic driving data collected from these modern vehicles potentially offers both road authorities and researchers a new and dynamic source of variables for analysis (Mannering and Bhat, 2014). The increasingly connected vehicles on our road networks are enabled by a wealth of technology capabilities, including network connectivity, high precision sensor data from the vehicle's Controller Area Network (CAN) Bus and GPS capabilities, and those with semi-autonomous features often include additional high definition cameras and LIDAR systems (Mannering, Kilareski, and Washburn, 2007). This technology can be leveraged to detect accidents and 'near miss incidents', such as heavy braking and evasive manoeuvres, otherwise known as critical driving events. Both researchers and industry players alike see the potential in the data from these events to help combat the rise in traffic accidents by reliably predicting locations with a high likelihood of collisions occurring. As such, the modern advancements in data quantity and quality has focused many research endeavours toward analysing critical driving situations and near-accidents.

The locations of these incidents could identify areas of high accident exposure, offering automotive manufacturers, insurers, and other industry players, a unique opportunity to reduce traffic accidents through the adoption of safety-focused services and business models. Moreover, this approach holds promise not only for the automotive and insurance industry, but also for policy makers in the long-standing field of road safety, where understanding the new capabilities and the reliance of findings from recent advances in automotive technology will be vital for determining suitable future traffic safety approaches and strategies. By adopting such a data-powered strategy to road safety there is the potential to reduce traffic accidents, and the unrealised opportunity to save lives around the world. Yet despite this, to date, there is limited empirical evidence on whether the likelihood of traffic accidents occurring at specific locations can be accurately identified through driving data and critical driving events.

In addition, driving data insights enable various safety-focused innovations for the automotive and insurance industry, further powered by the rise of increasingly connected vehicles and the emergence of smartphone turn-by-turn navigation applications. For example, in-vehicle warnings of potential accident hotspots could encourage insurance customers to drive more cautiously at locations with high risk exposure, and customised navigation services may offer safe-route options that avoid accident hotspots entirely, effectively enhancing 'pay-how-you-drive' policies with 'pay-where-you-drive' incentives. Yet even if safe-routing technology becomes widely adopted, it is unlikely that encountering accident hotspots can be circumvented in all driving situations. As such, identifying ways of effectively reducing the frequency and severity of traffic accidents at these known locations remains of the utmost importance. Among the wide ranging topics and approaches of improving traffic safety, there is a growing field that has been investigating how in-vehicle warning systems can encourage drivers to adapt their behaviour and drive safer.

Typically, these in-vehicle systems aim to prevent a collision with an upcoming vehicle or pedestrian by providing interventions to drivers, and latest studies demonstrate promising evidence that these systems can indeed have significant positive effects (Kazazi, Winkler, and Vollrath, 2015; Tey et al., 2014; Werneke and Vollrath, 2013). However, the vast majority of studies have focused on either simulation (Naujoks and Totzke, 2014; Seeliger et al., 2014) or controlled lab experiments (Ruscio, Ciceri, and Biassoni, 2015; Zhang, Suto, and Fujiwara, 2009). As such, the existing research falls short of both bringing an in-vehicle warning system into a field study setting, and utilising real world location analytics on traffic accident data to generate in-vehicle warnings. Going one step further, the benefit of detecting dangerous locations from data gathered by connected vehicles, and conceivably using these locations as a source for in-vehicle warnings, has widely not been addressed in this growing domain.

1.2 Objective and Approach

When considering the great potential of in-vehicle warning systems to reduce traffic accidents, along with the increasing availability of connected vehicle driving data, there is a gap in both research and practice regarding the real world assessment of such systems, and the potential of utilising vehicle data to identify hazardous locations on the road network. This leads to the following two research topics that this thesis sets out to investigate: The first, do in-vehicle warnings of accident hotspots, identified through real world location data, have a positive impact on driver behaviour in a realistic driving setting. The second, to what extent can the perilousness of our road networks be predicted on the basis of naturalistic driving data collected from connected vehicles. In the following section these two individual topics are outlined in more detail, and an overview of the thesis objectives and the approaches taken to address these research subjects is presented.

1.2.1 Improving Driver Safety

Many modern vehicles are equipped with sensors that can give guidance while the driver is parking, and through audio notifications prevent minor impacts with the infrastructure surrounding the vehicle. In a similar way, various research studies are geared towards investigating how in-vehicle warning systems can encourage drivers to adapt their behaviour on the road when necessary and improve traffic safety. The interventions explored in prior work have ranged from imminent collision warnings that require immediate evasive manoeuvres from participants, to general advisory warnings of upcoming road hazards that suggest that a more cautious driving style should be adopted (Kazazi, Winkler, and Vollrath, 2015; Tey et al., 2014; Werneke and Vollrath, 2013; Zhang, Suto, and Fujiwara, 2009). These studies provide promising evidence that all of these systems can indeed have significant effects on improving driving behaviour.

The vast majority of research regarding the positive impact these warnings can have on empowering drivers to avoid upcoming collisions and dangerous situations have predominantly focused on controlled setups in lab-based environments. For example, simulation studies have been extensively used to assess the effects of novel warning features and offer researchers a variety of advantages, such as repayable driving scenarios that can be tested on multiple subjects, the ability to monitor participants and their reactions, as well as the option to safely explore more critical circumstances involving collisions (Naujoks and Neukum, 2014a; Seeliger et al., 2014). A second common experimental setup is to make use of controlled field studies, where participants receive in-vehicle warnings while either driving in a location where there are no other road users, i.e. a race track, or following a pre-defined route on the road network as dictated by a researcher who is also in the vehicle (Ruscio, Ciceri, and Biassoni, 2015; Zhang, Suto, and Fujiwara, 2009). In these setups, more critical collision scenarios cannot be explored, but deeper insights may be developed into how participants behave in more realistic driving situations than simulation experiments. However the variables that can be monitored are more restricted and the locations are limited to set routes. While the overall potential of in-vehicle warning systems to improve driver safety are undisputed, it is vital to validate these promising lab-based results, and gain insights into how these interventions impact behaviour in a more naturalistic driving setup. Furthermore, in this field, the benefit of utilising real world location analytics to identify traffic accident hotspots, and use these as a source to generate in-vehicle warnings has not been previously discussed. As such, the first research objective of this thesis is to go beyond existing research, and depict the design and field evaluation of an in-vehicle warning system in order to answer the following research question:

RQ 1 Do warnings of upcoming dangerous locations have a positive effect on driving behaviour?

In order to answer this question, and in contrast to other in-vehicle studies, location analytics were applied to a national historical accident dataset, composed of over 266 000 accidents, and a retrofit in-vehicle warning system was developed to provide accident hotspot interventions to drivers. This system was tested outside of the simulation environment as part of an 18 week country-wide field test of professional drivers in Switzerland. During a dedicated experimental phase the drivers were split into control and warning intervention groups, and over a period of four weeks, with a total of 170 000 km driven, the impact of the system on driver behaviour was assessed with real-time sensor data collected from the vehicles.

1.2.2 Predicting Accident Hotspot Locations

Historical accident datasets, such as the one utilised to generate the accident hotspot warnings, are only collected and distributed in a limited number of countries, typically on a deferred basis. For example, in Switzerland such data is only made available once per year. As such, traditional traffic accident analysis based on historic crash data can be restrictive and severely out of date, additionally suffering from issues including under and over-dispersion, small sample size, and underreporting of traffic accidents (Mannering, Kilareski, and Washburn, 2007). Aside from this, traffic frequency measurements, e.g. average daily traffic (ADT), are vital to accurately calculate exposure measures, such as Crash Rate. However, these traffic frequency variables are expensive to measure and hence tend to mostly suffer from a very low spatial resolution or cover only small fractions of a road network.

A promising solution to the described drawbacks lies in the post-processing and aggregation of data available from the advanced sensors of highly connected and increasingly autonomous vehicles. Naturalistic driving data offers both researchers and practitioners "the potential [to] greatly expand the scope of statistical modelling and the inferences that can be drawn" when compared to the restrictive analysis possible with sparse data collected after an accident has occurred (Mannering and Bhat, 2014). As such, the recent advancements in data quantity and quality from modern vehicles has focused many research endeavours toward analysing critical driving situations and near-accidents. Many problems associated with crash data can be overcome by identifying the 'crash potential' of road sections, and conventional methods can be augmented by insights provided from vehicle data (Guo et al., 2010a; Guo et al., 2010b; Klauer et al., 2006). Moreover, estimations of traffic frequency measurements, such as ADT, can be generated from the frequency and trip details of vehicles travelling through these locations, transmitted via modern telematics, and used to estimate accident exposure measures.

Both researchers and industry players alike see potential in analysing the data collected from the sensors embedded in modern vehicles to detect critical driving events, such as heavy braking and evasive manoeuvres, and help improve road safety. The conditions leading to, and the locations of these events would offer both road safety researchers and practitioners a new and dynamic source of variables for assessment, overcoming many of the issues with traditional traffic accident analysis. Furthermore, these insights could identify areas of high accident exposure, offering automotive manufacturers and insurers a unique opportunity to reduce traffic accidents through the adoption of safety-focused business models, such as providing safe-routing services and in-vehicle warnings. As such, if automotive insurers and manufacturers wish to accurately measure and encourage safer driving and reduce the number of traffic accidents, then the ability to identify locations on the road network that carry a high risk of accident occurrence, so called 'blackspots', 'sites with promise', or 'hotspots' (Cheng and Washington, 2005), is of utmost value. While there is research showing that the situational factors of crashes and near-misses are strongly related, there is limited empirical data on whether potentially dangerous locations can be reliably identified through analysis of driving data and critical driving events (Pande et al., 2017). Thus, the second research objective of this thesis is to collect naturalistic driving data and evaluate the relationship between these events and historically dangerous areas of the road network, and and answer the following research question:

RQ 2 Is it possible to predict the locations of traffic accident hotspots on the basis of driving data and critical driving events?

To address this question, the full 18-week naturalistic driving field study dataset was considered, where the 72 vehicles that took part generated 690 000 km of high-frequency driving data across Switzerland. The locations and frequency of traffic accidents that occurred over the national road network during the field study were compared to the sensor data collected from the vehicles, which was processed to identify the locations of critical driving events. Explanatory variables associated with the likelihood of traffic accidents occurring were additionally considered, including the amount of vehicle traffic through the area, and whether or not the incidents were on the highway and in an urban area.

1.3 Outline

The remainder of this thesis is structured as follows: In the opening chapter, the background and related work on the subjects of traffic accident analysis, driver safety and in-vehicle warnings, and the potential that access to driving data brings, are outlined. Following this, foundational materials and methods utilised throughout the work are presented, highlighting the datasets that support the research, as well as describing both the system that was developed and the field study that was conducted in order to answer the introduced research objectives. Subsequently, the first of the research objectives is explored, and describes the generation of the accident hotspot warnings and the impact that the in-vehicle system had on driver behaviour during the naturalistic driving experiment. The second research objective is investigated in the penultimate chapter, which focuses on the relationship between the perilousness of locations across the road network, and the frequency of heavy braking events at these sites. Finally, the thesis is concluded with a summary of the key findings, the implications for both research and practice, and an outlook on the opportunities that the technology and analyses presented bring to the future of road safety, along with the work that should be conducted to advance this topic in the future.

Chapter 2

Background and Related Work

In order to support the presented research objectives, it is important to consider the existing work and research background of the following three key topics: traffic accident analysis, driver safety and in-vehicle warnings, and the potential that driving data has to support these fields. This chapter outlines the existing fundamental research on each of these subjects, and an additional discussion of the related 'State of the Art' work can be found in Chapters 4 and 5 respectively.

2.1 Road Traffic Accident Analysis

2.1.1 The Evolution of Accident Analysis

Through analysing the many attributes of traffic accidents that occur, such as the type of vehicles that were involved, the speed being travelled, and the condition of the driver, both researchers and road authorities seek to understand the root causes of these tragic incidents in order to better inform policy and practice and to improve road safety. In particular, the locations and patterns of traffic accidents on our road networks have been extensively researched over the past sixty years, with various methods developed to assess the need for, and impact of, road improvements (Hauer, 1996). The most common approaches have historically been non-spatial techniques, considering traffic accidents which occurred on locations defined by the underlying road structure. The so-called Crash Frequency method (Deacon, Zegeer, and Deen, 1974) is probably the most fundamental identification technique of this type. In this approach, the number of accidents that occurred during a specified period determines a road segment's perilousness. Estimations for Crash Frequency typically utilise count data models at selected locations (Anastasopoulos and Mannering, 2009; Bhat et al., 2014). Moreover, other examples of this approach have examined, on the basis of multivariate analysis, how the frequency of highway accidents is impacted by roadway geometries (e.g. horizontal and vertical alignments), weather, and seasonal effects (Shankar, Mannering, and Barfield, 1995). In a separate study, the influence of road segment length was explored with regard to accident frequency (Thomas, 1996), and the impact of horizontal curvature and an auxiliary lane has also been recently investigated (Pande et al., 2017).

The Crash Rate method is similar to the concept of Crash Frequency, but provides a measure of accident exposure of vehicles as it takes traffic volume into account (Hauer and Persaud, 1984). However, there are some evident drawbacks of both the Crash Frequency and Crash Rate methods, such as not considering random fluctuations of the number of accidents and the general lack of traffic volume data for generating the Crash Rate (Yu et al., 2014). As such, over time researchers have developed and utilised statistical models in the analysis of hazardous road sections (Miaou, 1994; Oppe, 1979). Probably the most prominent and applied identification technique using a statistical model is the Empirical Bayesian method (Hauer, 1997; Hauer et al., 2002). From a statistical perspective, it is argued that the Empirical Bayesian method can outperform both Crash Rate and Crash Frequency (Montella, 2010; Yu et al., 2014). Yet despite this, both Crash Frequency and Crash Rate are still commonly in use today, with their popularity stemming from the ease of implementation and interpretation.

The most common models for predicting accident data, that is count data, are negative binomial and Poisson regression models and their variants (Gianfranco, Soddu, and Fadda, 2017; Greibe, 2003; Lord and Mannering, 2010; Miaou, 1994; Quddus, 2008). Following this, negative binomial and hierarchical Bayesian models have been utilised to investigate the prominently discussed topic of the relationship between speed and accidents, where it was found that, when controlling for traffic volume and road geometry, average speed is not associated with accident rates whereas speed variations can be positively associated (Quddus, 2013). The effects of land use on accident rates has also been modelled, with a uniform grid made up of 0.259km^2 cells, resulting in statistically significant results using negative binomial models to predict the number of accidents (Kim, Brunner, and Yamashita, 2006). Finally, classification analysis has been applied to create a learning model based on attributes such as road, weather and traffic conditions, and social factors (Park, Kim, and Ha, 2016).

Alternative approaches to identify the perilousness of hazardous locations include measuring injury severity and cost estimates for crashes, where various statistical methods have been investigated, such as the ordered logit and ordered probit models (Kockelman and Kweon, 2002; O'Donnell and Connor, 1996). These models are suitable when the dependent variable has multiple possible outcomes and a natural ordering, for example, injury severity can be encoded by "no injury (0), minor injury (1), severe injury (2), and fatal injury sustained by driver (3)" (Kockelman and Kweon, 2002). Throughout the literature, locations identified with higher values for risk measures, e.g. Crash Frequency, Crash Rate and Crash Severity, are regarded to be more hazardous than locations with lower values. There are, however, common issues in many of these studies, which are associated with the difficulties in using crash data for the analysis of traffic safety (Lord and Mannering, 2010).

2.1.2 Spatial Accident Hotspot Analysis

The classic techniques outlined so far mostly neglect the spatial aspects and patterns of accidents, and focus on capturing details of the traffic accidents which are mapped onto predefined sections of road (Flahaut et al., 2003; Whitelegg, 1987). In contrast to these traditional approaches, recent analyses have considered the actual locations of individual accidents that occurred, regardless of the underlying road structure, utilising techniques that can capture the spatial relationship between these incidents. For example, spatial autocorrelation is a property found across geographic space, according to which random variables at certain distances from each other are either "more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations" (Legendre, 1993). Thus, observations are dependent on their spatial location, and data points can be aggregated to areas such as those used in census tracts or other zoning schemes (Quddus, 2008), grid based representations, or divided into regular or irregular polygons (Wang and Kockelman, 2013). As such, models for accident analysis have been proposed to account for this spatial relationship. For example, traditional count data models have been compared to spatial lag and error models, as well as spatial Bayesian hierarchical models (Quddus, 2008). These new models address spatial heterogeneity in geographical data and have been determined to perform well across the different model types.

Furthermore, with the increasing appearance of Geographic Information Systems and the larger availability of precise, geo-coded data, as well as digital maps, researchers have started to use spatial data analysis methods for identifying accident hotspots (Anderson, 2009; Flahaut et al., 2003; Okabe, Satoh, and Sugihara, 2009).

This follows the theory that the concentration of individual accidents at certain locations is called forth by a set of common causes – implying a spatial dependence of the accidents (Anderson, 2009; Flahaut et al., 2003). Underlying causes for such a concentration may include weather effects, infrastructure challenges, or traffic conditions (Geurts, Thomas, and Wets, 2005; Montella, 2010; Xie and Yan, 2008). The most commonly used spatial accident hotspot identification approaches are either the K-means clustering technique, spatial autocorrelation, or the Kernel Density Estimation method. In particular, the Kernel Density Estimation method has been extensively researched (Erdogan et al., 2008; Pulugurtha, Krishnakumar, and Nambisan, 2007; Xie and Yan, 2008), and in general it is argued that it outperforms other hotspot identification methods, such as the spatial autocorrelation, Crash Frequency or Crash Rate, and might perform equally well as the Empirical Bayesian method (Flahaut et al., 2003; Yu et al., 2014). In an example using this approach, accident hotspots on highways in Turkey were identified and explored with two different methods of Kernel Density Estimation analysis and repeatability analysis (Erdogan et al., 2008). The authors in this study additionally presented a Geographic Information System that was used as a management tool for accident analysis and the detection of hotspots with statistical methods. Furthermore, in another study, both a Kernel Density Estimation and a K-means clustering approach were used to profile road accident hotspots (Anderson, 2009).

However, more recently researchers have started to use the data mining clustering technique 'Density-Based Spatial Clustering of Application with Noise' (DBSCAN), to identify road accident hotspots (Szénási, 2015; Szénási and Csiba, 2014; Szénási and Jankó, 2016). DBSCAN is a density-based algorithm which classifies elements into clusters in such way that inside a cluster, the density of elements is higher compared to the outside of the cluster (Ester et al., 1996). Therefore, it can efficiently identify members of arbitrarily shaped clusters, as well as noise, and remain robust against outliers (Khan et al., 2014). After receiving significant attention in both theory and practice, the manuscript presenting the DBSCAN algorithm was awarded the '2014 SIGKDD test of time award' for the important impact it has had on the data mining research community (SIGKDD, 2014). DBSCAN has additionally been attributed with some key advantages when compared to other clustering techniques. Among these are the attributes that it does not require a predefined number of clusters as an input to the algorithm, and the two parameters, the minimum number of points to form a cluster and the distance that a point is considered a neighbour to another, can be set by a domain expert.

2.2 Driver Safety and In-Vehicle Warnings

2.2.1 Fundamental Approaches to Driver Safety

Supported by the ongoing worldwide research and analysis to reduce the number of traffic accidents on our roads, and the risks associated with these, a multitude of approaches to improve driver safety have arisen. In general, these approaches can be split into three categories: governmental laws and the associated enforcement that punishes dangerous driving behaviour, improvements to vehicles that aim to prevent or reduce the impact of traffic accidents, and infrastructure improvements that hope to reduce the likelihood of accidents occurring or limit the injuries that might be sustained.

One of the most critical factors in road safety are the laws that countries strongly and sustainably enforce in order to improve road user behaviour. Reductions in road traffic crashes, injuries, and fatalities can all be achieved through legislative change and effective road safety programmes, such as those targeting speeding in urban areas or drinking and driving, as seen in a variety of countries around the world (Peden et al., 2004; World Health Organization, 2013b). In a similar way, regulatory requirements have additionally led to the development of safer and more advanced vehicles, especially in higher income countries. Consumer demand has led to the introduction of initially expensive safety features to vehicles, which have then become cheaper over time, and some even made mandatory requirements for new vehicles after demonstrating a contribution to improving road safety. Perhaps the most successful example of this is electronic stability control, a technology that aims to prevent the loss of vehicle control and swerving in the case of oversteering or understeering, which has proven so successful in preventing traffic accidents that it has become a required feature for all vehicles through UN regulation (Global New Car Assessment Programme, 2015). Through this regulation, which both encourages and requires the adoption of this safety feature, studies have shown that there has been a reduction in serious and fatal injuries that would have been sustained from various types of traffic accident (Erke, 2008; Lie et al., 2006).

Moreover, with respect to the improving technology within vehicles themselves, the advent of autonomous vehicles brings the promise of a new era of traffic safety, where the frequency of road accidents can be drastically reduced, and potentially eliminated entirely (Fagnant and Kockelman, 2015). However, even in the most advanced markets, it will take decades to make this vision a reality. For example, recent predictions indicate that it is unlikely that the majority of the light-duty vehicle fleet in the U.S. will be capable of full self-driving automation by the year 2045 (Bansal and Kockelman, 2017). As such, semi-autonomous vehicles will become the dominating paradigm in the coming years, and traffic accidents from the manual driving of these vehicles will persist as a key issue (Albright et al., 2016). Although recent advancements in areas such as autonomous braking hope to reduce the number of fatalities from traffic accidents in the near future (Fildes et al., 2015), the road infrastructure itself remains an important road safety factor.

The extensive research on the locations of road traffic accidents, and identifying areas with a higher risk of collisions, is often motivated by determining suitable locations and priorities for road infrastructure improvement (Hauer, 1996). These infrastructure changes are typically in the form of either significantly altering the road structure, e.g. converting a dangerous junction into a roundabout, or adding new road sign infrastructure to convey safety information to drivers and improve awareness. Perhaps an unexpected example of the importance of altering the existing road structure can be seen in typical urban road networks that have often had insufficient attention paid to the integration of footpaths and cycle paths during their planning and development stages (World Health Organization, 2015). The rise in health and sustainability programs in many countries, encouraging walking and cycling within cities, has led to a mix of road users where fast-moving traffic shares the road with pedestrians and cyclists. Despite the worthy intentions of these programs, these circumstances potentially increase the risk of traffic accidents as the non-vehicle participants have to negotiate various dangerous situations on the road. Thankfully, studies have shown that the modification of the existing road structure, for example, the addition of cycle paths to urban roads, can reduce cyclist injuries and fatalities by 35 % (Peden et al., 2004).

Road sign infrastructure has long been associated with providing directional guidance to drivers and in enhancing traffic safety, with many countries adopting simple and standardised warning signs to aid driver understanding and adherence to traffic regulation. The road sign infrastructure on road networks that impact road safety can be grouped into the following distinct types that have evolved as a result of improving technology. Traditional traffic signs are almost always static, and most commonly take the form of physical road signs, such as stop signs and speed limits. Temporary physical signs are used in situations where drivers need to be notified of potential danger for a limited time, examples of this include signs during road works or current icy conditions. Variable-message traffic signs have recently become more prevalent on highways, and can provide a variety of dynamic information, such as changeable speed limits, and warnings of adverse weather condition warnings

FIGURE 2.1: Examples of existing road sign infrastructure. From left to right: traditional static sign, temporary static sign, variable-message sign, vehicle activated sign. All images courtesy of www.wikipedia.org

and upcoming traffic jams (Emmerink et al., 1996). An extension of these variable-message signs includes dynamic, or smart, signs that are equipped with sensors so that warnings are shown only when an approaching vehicle is displaying certain behaviour, the most common instance being exceeding the speed limit (Gregory et al., 2016). Visual examples of each of these types of warning can be found in Figure 2.1.

Overall, various studies have demonstrated that a significant improvement in driver safety can be achieved through warnings of upcoming hazardous road features, such as sharp corners and icy conditions, thus mitigating the risk of vehicle accidents (Carson and Mannering, 2001; Persaud et al., 1997). However, the cost of installing and maintaining all of these types of infrastructure can be high, especially for developing countries (International Road Federation, 2006). This has motivated the development of smartphone applications and in-vehicle systems that can provide warnings of hazardous situations in a more cost-effective and dynamic way. An interesting example of this approach has been in place in Iceland since 2014, where hazardous conditions, such as volcanic behaviour, are alerted to every phone in the area via SMS, effectively serving as a 'virtual warning' (Iceland Review, 2014).

2.2.2 In-Vehicle Warnings

The promise of in-vehicle warning systems to improve driving safety has generated a substantial body of research (An and Harris, 1996; Hirst and Graham, 1997), and a positive impact can be seen when they are compared with traditional warning approaches in relation to driver behaviour and accident frequency, e.g. in the context of railway crossings (Tey et al., 2014). For example, common conventional warning devices, such as the passive stop sign, were compared by researchers to active variations, i.e. flashing lights and a half bloom-barrier with flashing lights. The results showed that, on average, driver responses to passive and static warnings were poor in comparison to active warnings (Tey, Ferreira, and Wallace, 2011). In a later follow-up study, rumble strips and in-vehicle audio warnings were compared to the previous active and passive warnings at the railway level crossings (Tey et al., 2014). Results indicated that both of the novel warning devices produced much higher levels of driver compliance than the existing conventional warnings, demonstrating the positive impact in-vehicle warning systems can have when compared to conventional approaches.

Furthermore, various simulator based studies have shown that in-vehicle warning systems can have a positive effect on driving behaviour. In one example, early warning signals displayed while approaching an intersection showed a positive shift towards participants driving safer (Werneke and Vollrath, 2013). Those drivers that were shown the intervention adapted their driving behaviour by turning with a lower velocity after waiting longer at the intersection, and so avoided collisions. Visual warnings have also had a positive effect on drivers braking reaction time, for both older and younger participants (Kazazi, Winkler, and Vollrath, 2015). The largest improvement was seen in critical situations, where collisions were successfully avoided due to the warnings.

Outside of simulation experiments, a few controlled field studies have investigated the impact of in-vehicle warnings on driver safety. The influence of warning expectancy and automation complacency on real-life emergency braking has been investigated (Ruscio, Ciceri, and Biassoni, 2015). In particular, reliable warnings quickened the decision making process and misleading warnings generated automation complacency, slowing visual search for hazard detection. Additionally, specific spatially located hazards have been investigated with regard to the effect of in-vehicle warning systems (Zhang, Suto, and Fujiwara, 2009). In the study, the hazardous area tested was an intersection near an arch-shaped bridge, where traffic accidents had often occurred due to poor visibility, and the effects of different combinations of audio and visual warnings provided to a driver were investigated. It was demonstrated that information about the cause of accidents was more effective than information on road infrastructure in helping drivers to avoid dangerous driving situations.

2.3 Potential of Driving Data

In many of the studies highlighted regarding the positive effects on in-vehicle warnings, driver behaviour was either manually observed by the researchers or operationalised with variables extracted from the simulation environment, such as velocity and braking reaction time (Kazazi, Winkler, and Vollrath, 2015; Werneke and Vollrath, 2013). With the modern developments in data quantity and quality from the advanced sensors in smartphones and connected, semi-, and fully-autonomous vehicles many research endeavours have additionally started to investigate insights that can be gained from driving data from these sources (Guo et al., 2010a; Guo et al., 2010b; Klauer et al., 2006). Furthermore, access to such naturalistic driving data offers both road safety researchers and practitioners a new and dynamic source of data for analyses, which can be used independently from, or to augment, the typically restrictive road traffic accident datasets (Mannering and Bhat, 2014).

One of the simplest, but also most powerful, of these technologies is the ease of accessing localisation data from the Global Positioning System (GPS), technology that is now built into almost every smartphone and modern vehicle, while driving. Not only does this enable accurate navigation services, but when the data is transmitted via modern telematics, estimations of traffic frequency measurements, such as average daily traffic, can be generated from the frequency and trip details of vehicles travelling across the road network. In fact, even without GPS capability and internet connectivity, the infrastructure powering mobile phones has been shown to be capable of providing accurate traffic flow information just from the data generated when a phone changes location, and then swaps from being serviced by one phone mast to another (Swisscom, 2017).

Due to the overall wide and increasing availability of smartphones, many recent research efforts have identified the capability for these devices to be used as a source of driving data. As well as GPS connectivity, modern smartphones are typically equipped with an inertial measurement unit (IMU) to provide three-dimensional acceleration values, as well as a camera and microphone that can also be utilised for detection of various driving related activities, for example, stop-and-go or fluid traffic conditions, and road quality (Mohan, Padmanabhan, and Ramjee, 2008). In other studies, road roughness conditions and features, such as potholes and manhole covers, were detected using a mounted smartphone's accelerometer and microphone data (Mednis, Elsts, and Selavo, 2012), and sensor fusion utilised to identify driving style, differentiating between aggressive and non-aggressive driving manoeuvres (Johnson and Trivedi, 2011). Finally, gyroscope and magnetometer sensor data can also be obtained in certain smartphones and have been considered, in conjunction with the accelerometer, for detecting risky and safe driving behaviour, i.e. by comparing the similarity of an identified event to predefined risky and safe event patterns (Eren et al., 2012).

However, a limitation of these approaches is that the smartphone must typically be mounted in the vehicle to obtain accurate results (Fazeen et al., 2012). If the smartphone is moved while driving then false positives for events and road features will almost certainly be introduced into any system. An option to overcome this limitation is to add an additional mounted IMU to a vehicle and use the same signal processing techniques as applied to smartphone accelerometer data (Tin Leung et al., 2011). These two different sensor approaches have been compared for the detection of driving acceleration, braking and steering events (Paefgen et al., 2012). While the authors found correlations between the smartphone and IMU-based events, differences were also reported between the statistical distributions of generated event counts, primarily due to variations in how the smartphone was mounted and positioned inside the vehicle.

An alternative approach for the collection of driving data is to make use of increasingly available access to the sensors in the vehicle itself. Since access to standardised On-Board Diagnostics (OBD-II) data is mandatory for all vehicles manufactured or sold in the USA from 1996, many of these parameters have been widely used in research. This data gives insights into features such as vehicle velocity and engine speed, and has been used to detect hazardous driving behaviour (Imkamon et al., 2008). As such, the automotive insurance market has harnessed these insights through telematics solutions and smartphone applications in order to measure 'pay-as-you-drive' trip distances, and reward safer driving for 'pay-how-you-drive' policies by assessing dangerous driving patterns, such as speeding, swerving, and heavy braking (AXA Winterthur, 2016; Metromile, 2018). From a research perspective, OBD-II data has been utilised in a study that predicted passenger ratings of driving behaviour by assigning current hazardousness to a range of values between driving safely and driving dangerously (Castignani, Frank, and Engel, 2013). The authors developed classification models of these levels by combining OBD-II sensor data, 3-axis accelerometer data, and fuzzy logic, a technique suitable when a binary distinction is not appropriate. In other work, the behaviour of different driver groups in the US State of Georgia was studied by collecting data from vehicle OBD-II systems, as well as GPS location data (Jun, Ogle, and Guensler, 2007), and it was noted that clusters of hard deceleration events were co-located with clusters of historical accident data.

Yet OBD data is a small subset of the data that is potentially available within a vehicle. The Controller Area Network (CAN) Bus holds unstandardised information on the inner-workings and communications between the subsystems of a vehicle. Deeper insights into a vehicle's operation are available on the CAN Bus of a variety of vehicles, and as such, have been used in recent research studies of driver behaviour. For example, one proposed platform, based on a feature extraction algorithm and a fuzzy system, identified current driving conditions and provided energy consumption feedback to the driver (Araújo et al., 2012). Feedback for the system was based on GPS measurements from a smartphone combined with CAN Bus sensor readings, including vehicle speed, acceleration, throttle signal, fuel consumption, and engine speed. Other studies have also made use of the CAN Bus of several vehicles to identify fuel efficient driving behaviour (D'Agostino et al., 2015; Ferreira, Almeida, and Silva, 2015), and characteristics of aggressive and calm driving have been previously identified (Karaduman et al., 2013).

Chapter 3

Foundational Materials and Methods

Addressing the research objectives of this thesis, presented in Section 1.2, requires several key components that are described in the following chapter. The first research objective, determining what impact in-vehicle warnings, utilising real world location data to generate interventions, have on driver behaviour, requires knowledge of places that are dangerous on the road network, and a system that can provide such warnings, as well as measure driver behaviour. The second research objective, investigating to what extent the perilousness of our road networks can be predicted on the basis of critical driving events, also requires the insights into where traffic accidents have historically occurred, along with data regarding the traffic flow of the road network, and driving data collected in a naturalistic setting. As such, this chapter initially outlines the two supporting datasets utilised in this thesis, the first featuring fundamental information on the traffic accidents occurring within Switzerland since 2011, and the second providing traffic frequency measurement at specific locations across the country. Following this, the development and features of an in-vehicle system that in parallel collects high frequency driving data for analyses and provides accident hotspot warning interventions to drivers is then presented. Finally, the chapter concludes by detailing the field study that was conducted to answer the introduced research objectives utilising the in-vehicle system, the demographics of the participants, and the experimental design that was applied.

3.1 Supporting Datasets

Two fundamental datasets are presented in this initial section that the subsequent development and analysis of the thesis build upon. The first, a dataset containing contextual information on traffic accidents occurring in Switzerland is the foundation of generating the accident hotspot warnings presented in Chapter 4, and determining the perilousness of locations in Chapter 5. The second, a dataset providing traffic flow measurements for a variety of locations across the Swiss road network is utilised in the analysis of Chapter 5.

3.1.1 FEDRO Traffic Accident Dataset

In order to facilitate the investigation into data-powered road safety of both Chapters 4 and 5, and determine whether in-vehicle warnings of accident hotspots can help improve driver behaviour, the Swiss Federal Roads Office (FEDRO) Statistics For Road Accidents provided access to a historical accident dataset. Since 2011, FEDRO has collected detailed information regarding every Swiss road accident for which the police were called, building up an extensive accident record. The data collection procedure of FEDRO is comprised of the following steps: Police officers are called to traffic accidents and have to fill out a report for each incident. This report includes information regarding the involved persons and vehicles, details about the location of the accident, the leading cause, and the circumstances and course of events leading to the collision. Once the report has been filled out, it is then sent to the FEDRO, which reviews and controls the data to ensure integrity and credibility. Appendix A provides an example of the report template of the FEDRO (in German). This data is made available for research purposes on a yearly basis, and so the initial dataset received and utilised for generating in-vehicle warnings was composed of over 266 000 geo-located accident records, which occurred in Switzerland between the years 2011 to 2015. Following this, two additional datasets were later received covering the years 2016 and 2017. Combined, these datasets provide details on over 377 000 traffic accidents, and, as seen in Figure 3.1, the number of accidents occurring in Switzerland follows the global trend of remaining at an unacceptably high level year on year, with an average of almost 54 000 annual reported accidents.

The provided dataset includes a multitude of features related to each accident, such as the reason for the incident occurring, the number of people and vehicles involved, and the severity of the suffered injuries. From this data, and as shown in Figure 3.2, it can be seen that, at least in the traffic accidents reported to the police, roughly 80 % of the participants are thankfully uninjured. However, an average of 16.1 % of the participants suffer minor injuries, dropping from 18 805 to 17 759 between 2011 and 2017, a change of 5.6 %. This trend is also seen in the number serious

FIGURE 3.1: Chart showing the number of traffic accidents within Switzerland

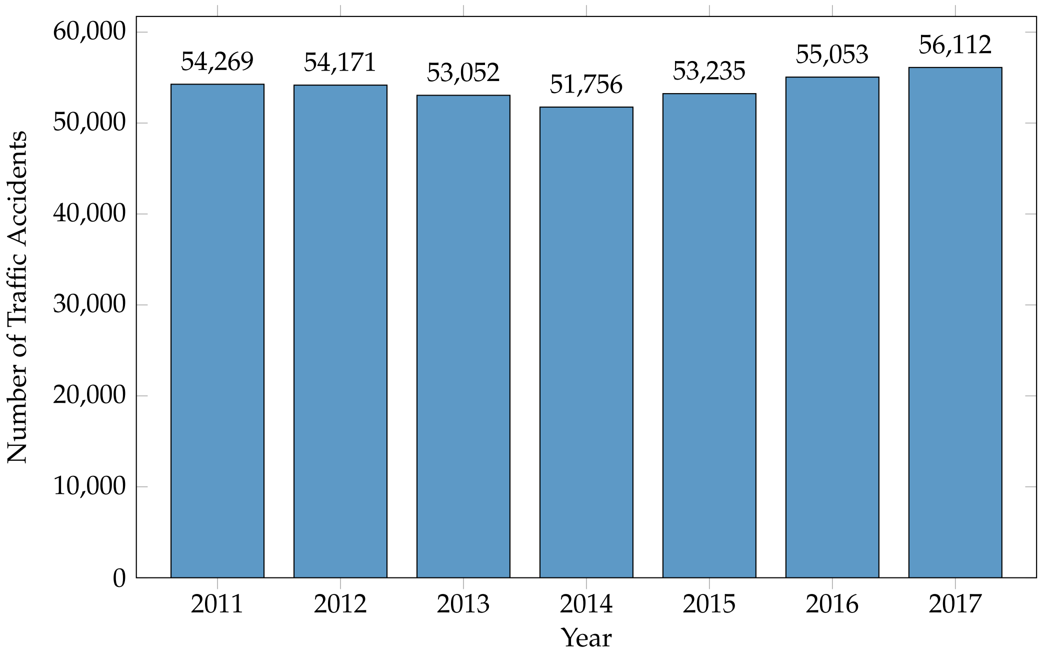


FIGURE 3.2: Chart showing the number of people involved in traffic accidents, by degree of injury, within Switzerland

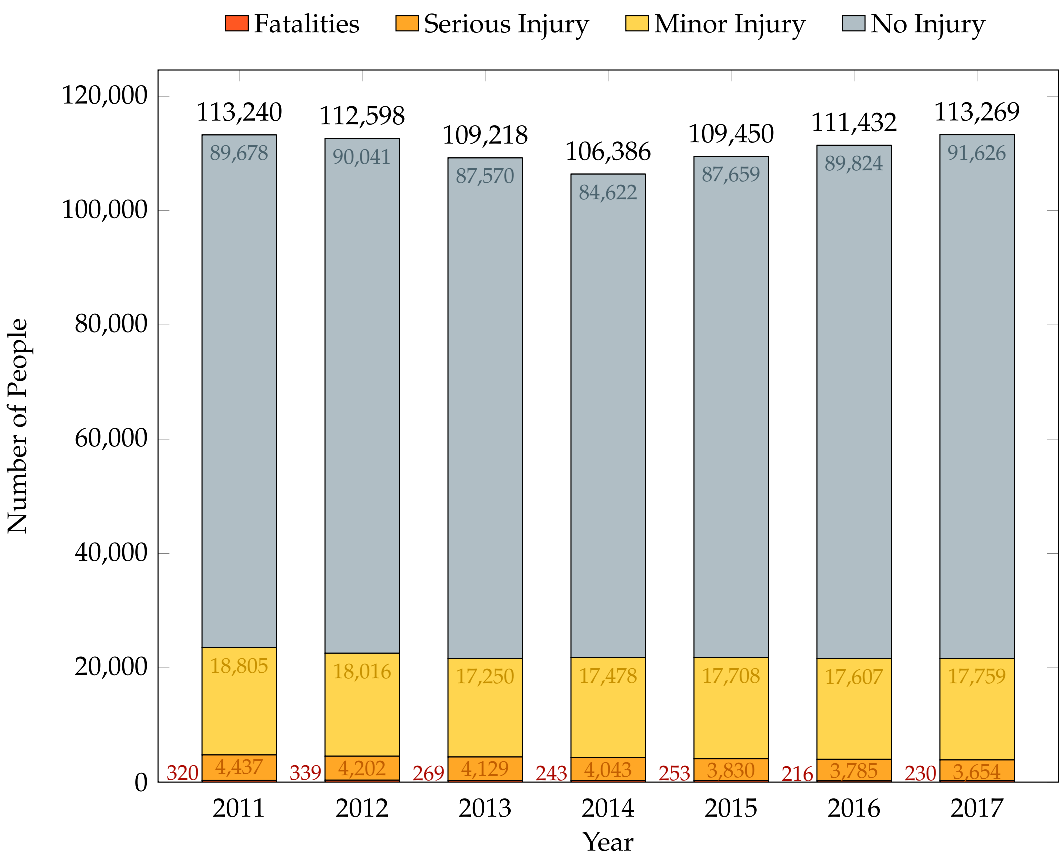
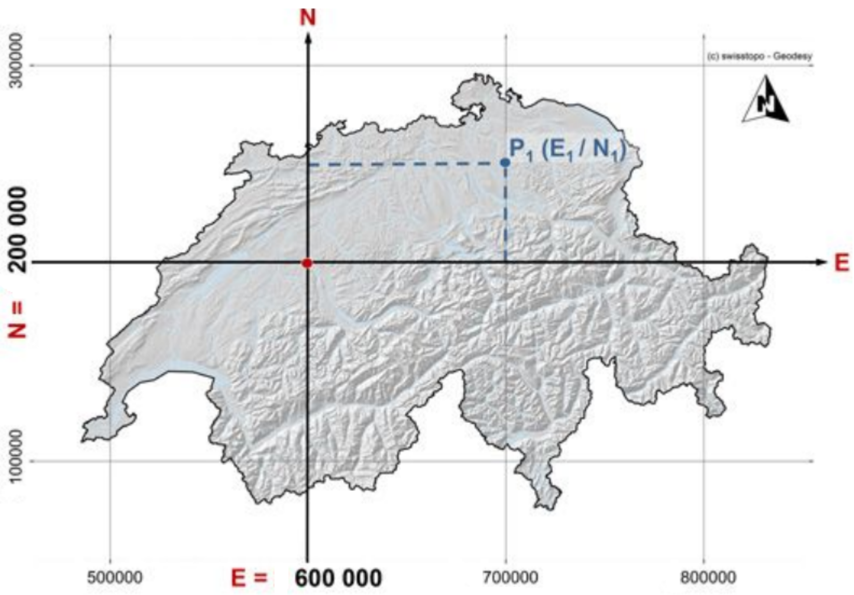


FIGURE 3.3: Switzerland's national coordinate system - Swiss coordinates LV03 (Federal Office of Topography, 2016)



injuries incurred due to traffic accidents, where approximately 3.6 % of those involved are affected, which has dropped from 4437 to 3654 (17.7 %) over the years. Finally, the average number fatalities from traffic accidents is 0.2 % of the total people involved over a year, falling from 320 to 230 deaths, a dramatic change in seven years of 28.13 %. Yet, while the proportion of people suffering minor to fatal injuries from traffic accidents in Switzerland has been dropping, significantly more work is needed in order to reach the United Nations 2015 commitment of halving the number of injuries and deaths from traffic accidents by 2020.

Aside from the descriptive features of each accident, the FEDRO traffic accident dataset also includes the geo-location information of where each incident occurred. As FEDRO is a federal institution of Switzerland, the location data of the accidents is provided in the so-called LV03 Swiss coordinate system. This coordinate system is typically the standard used by federal authorities to collect and store geo-locations, and represents the Earth in a projected form. The point of origin of the projection is in Bern, which is labeled as $E = 600\,000$ m (east) and $N = 200\,000$ m (north), and is illustrated in Figure 3.3 demonstrating the LV03 coordinate frame. The main advantage of using the Swiss coordinate system is that, unlike systems that represent the Earth as a sphere, i.e. the Global coordinates (WGS84), distance calculations between two points are simpler and can use Euclidean distance.

The Swiss coordinate system LV03 can be transformed into the more commonly used

WGS84 projection by applying a set of derived formulas (Federal Office of Topography, 2016). The accuracy of these approximate formulas is at least $0.12''$ (arcseconds) in longitude, $0.08''$ in latitude and 0.5 metres in height everywhere in Switzerland. As such, the coordinate transformation of the accident data was achieved via the following approximation approach:

1. Transformation of the projection coordinates (east (y)/ north (x)) into the civil coordinate frame (where Bern = (0,0)) and into units of [1000 km]:

$$y' = \frac{(y - 600000)}{1000000} \tag{3.1}$$

$$x' = \frac{(x - 200000)}{1000000} \tag{3.2}$$

2. Calculation of longitude (λ) and latitude (ϕ) [$10\,000''$]:

$$\lambda' = 2.6779094 + 4.728982y' + 0.791484y'x' + 0.1306y'x'^2 - 0.0436y'^3 \tag{3.3}$$

$$\phi' = 16.9023892 + 3.238272x' - 0.270978y'^2 - 0.002528x'^2 - 0.0447y'^2x' - 0.0140x'^3 \tag{3.4}$$

3. Transformation to units of [$^\circ$]:

$$\lambda = \lambda' \frac{100}{36} \tag{3.5}$$

$$\phi = \phi' \frac{100}{36} \tag{3.6}$$

Once transformed into GPS latitude and longitude coordinates, the locations of the traffic accidents across the country can be visualised. Figures 3.4 to 3.6 all present heatmap representations of the initial 2011 to 2015 traffic accident datasets provided by FEDRO, where *green* locations indicate a lower frequency of accidents, and *red* a higher frequency. As one would perhaps expect, Figure 3.4 shows that the cities and urban areas in Switzerland have a higher concentration of traffic accidents, typically explained by a higher number of vehicles and other road participants, e.g. pedestrians, in these areas. Additionally, main transportation routes, for example highways, can be easily identified linking these urban locations due to a higher frequency of accidents than the surrounding rural areas. Figure 3.5 provides a more detailed examination of one of the cities, Zurich, where the main transportation routes can again be seen, as well as a higher concentration of accidents when compared to the nearby suburban and rural locations. Furthermore, certain areas inside the city demonstrate

FIGURE 3.4: Swiss FEDRO accident data - Nationwide heatmap
Green represents low, and red indicates high accident frequency

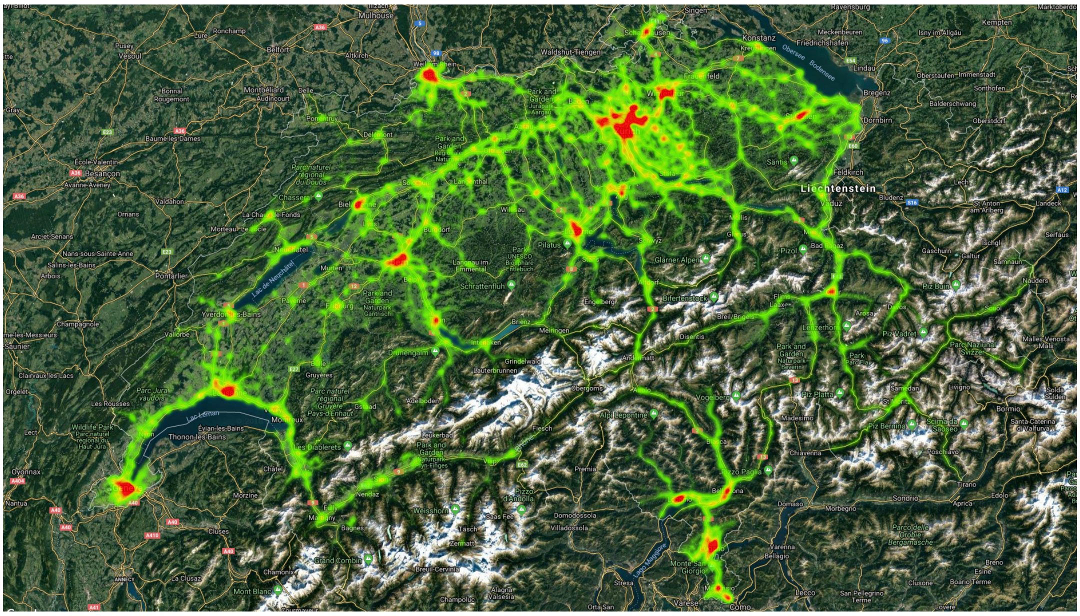


FIGURE 3.5: Swiss FEDRO accident data - Zurich heatmap
Green represents low, and red indicates high accident frequency

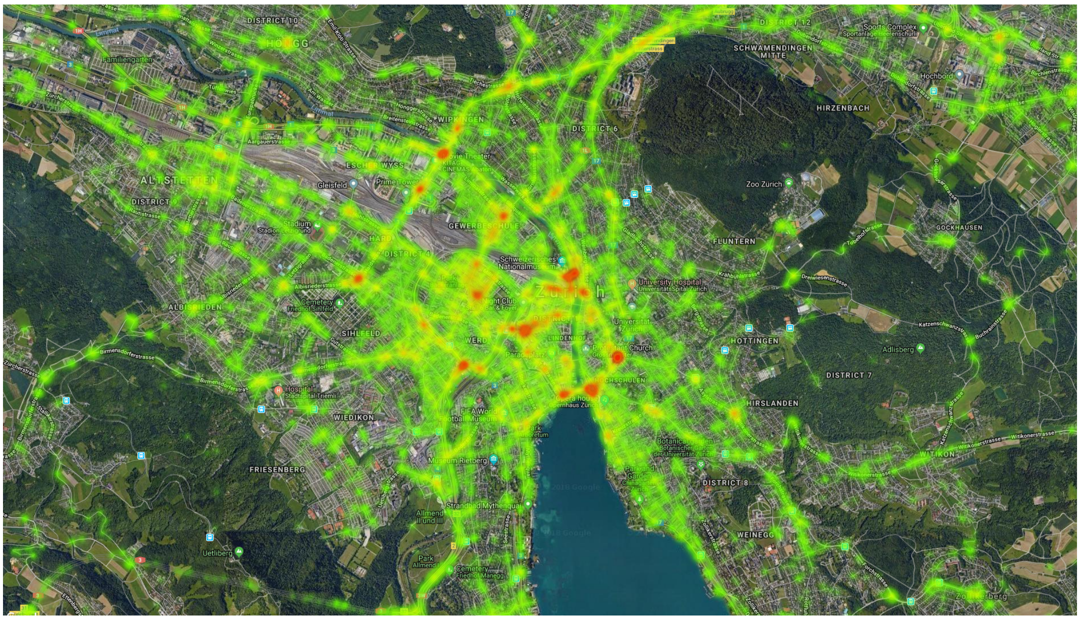
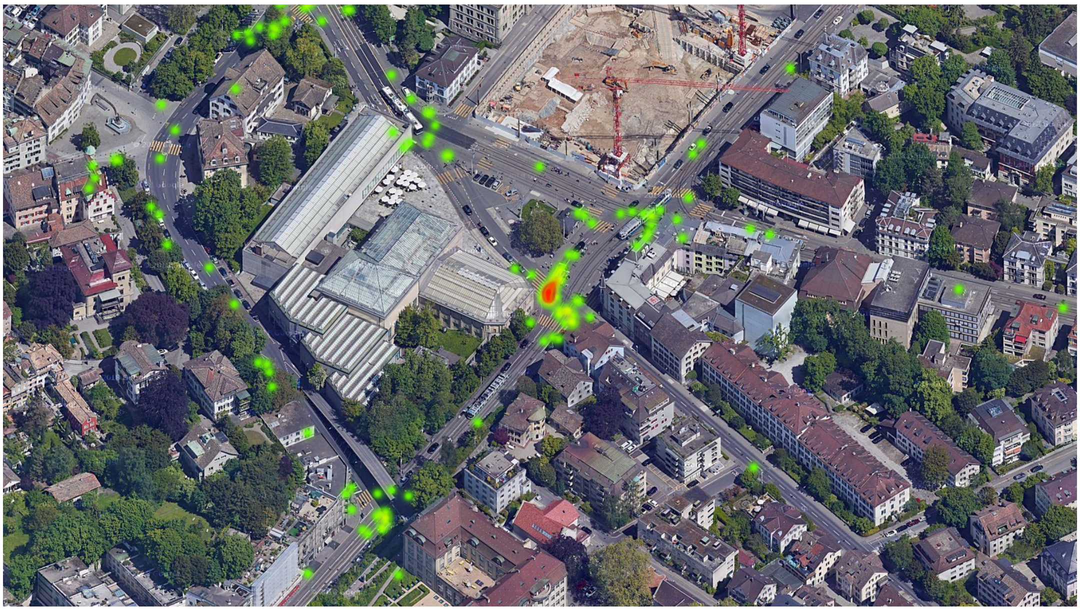


FIGURE 3.6: Swiss FEDRO accident data - Zurich Kunsthaus heatmap
Green represents low, and red indicates high accident frequency



a significantly higher amount of traffic accidents than even the surrounding urban locations.

Figure 3.6 shows an example of one of these locations in greater detail, here the intersection and crossing in the centre of the image has a much higher frequency than the nearby road infrastructure. For some reason, this location is clearly more dangerous than other intersections and crossings in the same area. Looking in more detail in the map based image, it can be seen that this particular location is a combination of pedestrian crossings, a traffic light controlled multi-lane intersection, and a tram line. With just a visual inspection it is difficult to identify a clear cause explaining why these traffic accidents are occurring. However, when considering the features provided in the FEDRO dataset for the traffic accidents that contribute to this dangerous location, a clearer idea of what the underlying cause of these accidents can start to be constructed.

Each traffic accident in the FEDRO dataset is composed of several sets of features, which can be grouped into three fundamental categories:

- **'What'** was involved in the accident, e.g. cars, motorcycles, pedestrians
- **'Why'** did the accident occur, e.g. speeding, aquaplaning, distraction
- **'Where'** did the accident happen, e.g. at a roundabout, intersection, pedestrian crosswalk

TABLE 3.1: Potential attributes assigned to each traffic accident in the FEDRO dataset

What Objects involved	**Why** Reasons for accident	**Where** Surrounding road infrastructure
Agricultural vehicle	Aquaplaning	Bus/Tram stop
Bicycle (Moped, E-Bike)	Careless backing up	Crosswalk
Bus	Collision with animals (pet, wild animals)	Curve
Coach	Condition / Intention of the driver	Cycle path
Lorry/Juggernaut	Cutting corners	Entry/ exit parking lot / property
Motorcycle	Disregard right of way	Intersection
Passenger car	Disregard signalisation	Junction
Pedestrian	Disregard traffic light	Lay-by (Rest stop)
Train	Illegal crossing	Parking lot
Tram	Inattentiveness / Distraction	Parking space
	Lack of consideration while lane changing	Protection island
	Lack of space between cars	Railroad crossing without barrier
	Not adapted speed (e.g. curve)	Railroad crossing with barrier
	Not adapted speed (e.g. road condition)	Roundabout
	Sun dazzling	Square
	Technical defects of the vehicle	Straight road
		Traffic abatement
		Tunnel

Table 3.1 provides an overview of these categories and the features that each traffic accident might be labelled with[1], and Figure 3.7 presents the two pages of the FEDRO report form (in German) associated with these categories, with the related attributes highlighted in *blue*.

If the features provided in the FEDRO dataset for the traffic accidents contributing to the previously discussed dangerous location in Zurich are considered, more than half of the accidents are reported with the 'why' main cause being "lack of consideration while lane changing" of the vehicles involved, and so a clearer picture can be built of what the underlying cause is of these accidents. Hotspots such as this, where there is a dominant reason for the traffic accidents occurring, strongly motivate research investigating whether in-vehicle warnings of these dangerous locations can improve driver safety.

3.1.2 FEDRO Traffic Frequency Dataset

Aside from the attributes of accidents previously described, such as the location, objects involved, and the underlying causes, traffic accident analysis has traditionally considered additional variables related to the road infrastructure itself in order to

[1] For brevity, the full set and subsets of all possible objects, reasons, and locations have been summarised in this table. For the complete list of attributes and combinations, please see the FEDRO report forms (in German) in Appendix A.

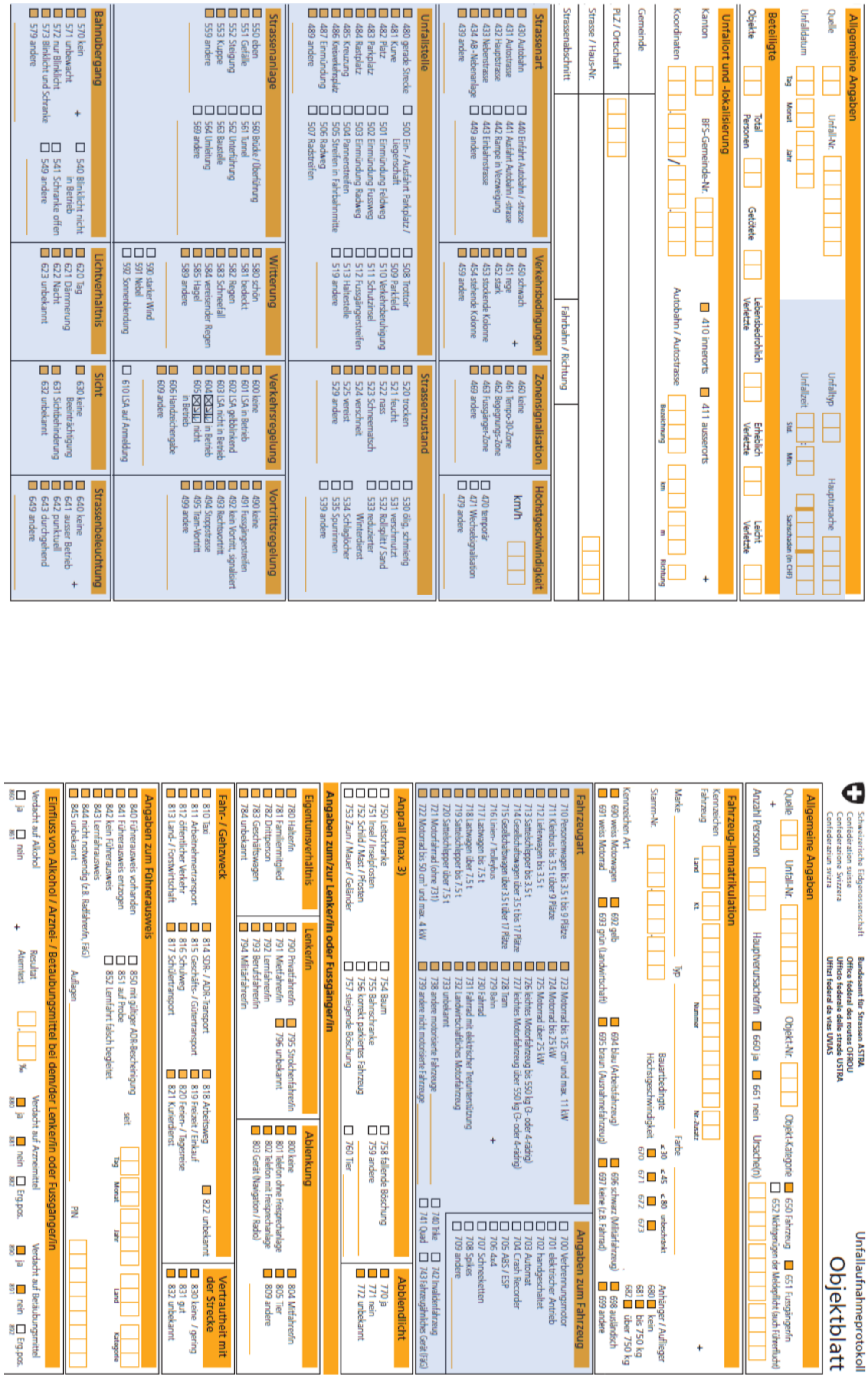


FIGURE 3.7: FEDRO Police Accident Report Form (in German) - Pages 1 and 3, with the attributes related to 'What', 'Why', and 'Where' highlight in *blue*

FIGURE 3.8: Examples of Swiss automatic traffic counting stations (Swiss Federal Road Office, 2013)

(A) An automatic traffic counting station (B) Integrated road surface induction loops

investigate safety related topics. For example, the rate of traffic flow, i.e. the number of cars travelling on a particular stretch of road, has long been associated with traffic accidents (Hauer and Persaud, 1984). The argument for this is that with a fixed probability of a traffic accident occurring, sections of road with higher traffic flow will see a higher number of traffic accidents per year. Therefore, to assess a specific location's Crash Rate or risk exposure based on the number of traffic accidents, it is important to account for traffic frequency. In order to incorporate this into potential analysis, traffic data was obtained from FEDRO and the Swiss Automatic Road Traffic Counts (SARTC) department. These datasets are freely available online, and are powered by a network of permanent automatic traffic counting stations on the country's most important thoroughfares. As shown in Figure 3.8, these automated stations detect and count vehicles via connected induction loops which are integrated into the road surface. In total, 466 such stations have been in operation as of May in 2013 in Switzerland, and of these, 424 were still operational for at least part of 2016 . However, as is the case in many countries, in Switzerland there is only partial location coverage of such automatic counting stations.

The ultimate goal of SARTC is to record both the direction and time of traffic movements, and in this case the average daily traffic (ADT) and average weekday traffic (AWT) volumes are considered the two most important criteria for evaluation. For ADT, the figure is calculated by SARTC on every day of the year for 24 hour traffic volumes, and AWT from traffic volumes on all weekdays (Monday to Friday) except public holidays. These statistics are only recorded for roads when they are open, so some data is not collected during winter when certain roads are forced to close. Annual and monthly results, between January 2002 and August 2017 at the time of writing, for ADT and AWT are available online (Swiss Federal Road Office, 2017),

FIGURE 3.9: Locations of Switzerland's automatic traffic counting stations (Swiss Federal Road Office, 2013)

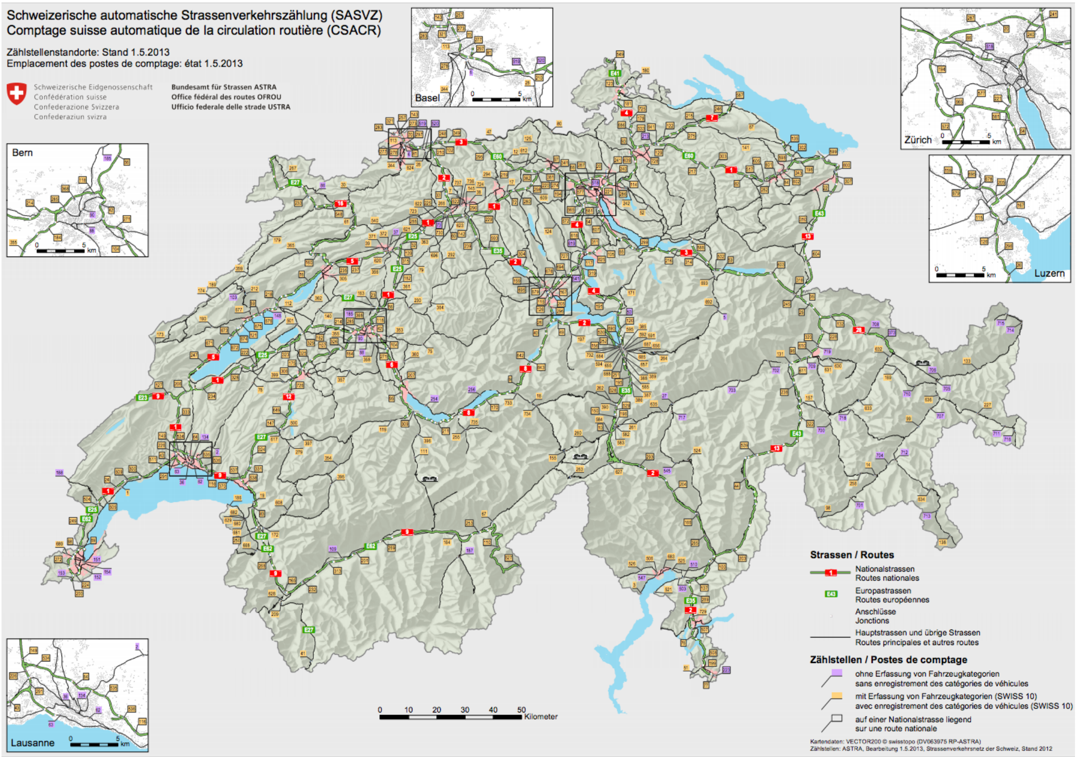
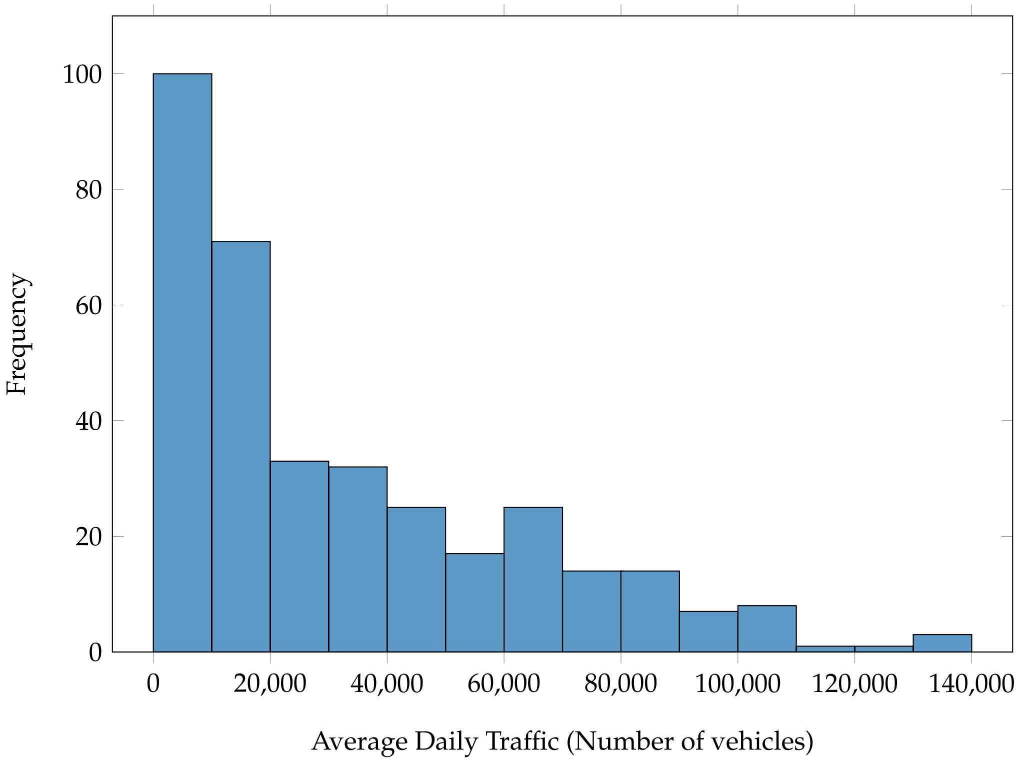


FIGURE 3.10: Histogram showing the distribution of Average Daily Traffic from 2016 across the automatic counting stations in Switzerland



along with a separate dataset containing the locations of the counting stations (Swiss Federal Road Office, 2013). As with the FEDRO traffic accident dataset outlined in Section 3.1.1, the locations of the counting stations are represented in the LV03 Swiss coordinate system, and presented in Figure 3.9. Thus, the same formulas from Section 3.1.1 can be applied to transform the Swiss coordinate system LV03 into the WGS84 projection (Federal Office of Topography, 2016).

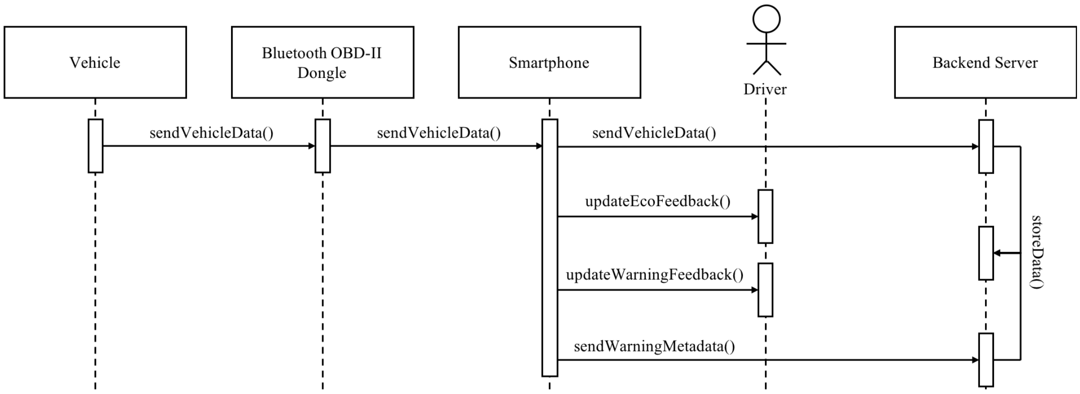
Figure 3.10 provides a histogram showing the distribution of the yearly ADT across the Swiss road network, as available in the FEDRO ADT dataset. In this case, the ADT measurements for each month of the year 2016 were averaged to generate the yearly ADT for each counting station. A total of 351 of the 424 operational counting stations in 2016 collected measurements for each of the 12 months, the remaining 73 stations were not included in the summary.

3.2 System Description

The outlined datasets are used throughout the thesis in order to facilitate the investigation into the impact of in-vehicle warnings, and the relationship between critical driving events, such as heavy braking, with the locations of accident hotspots. In order to achieve these goals, an in-vehicle system was developed that could display accident hotspot warnings to drivers, as well as collect driving data from the vehicle itself to investigate driving behaviour and road safety. This system was developed through multiple build and evaluation stages, typical of iterative development, following the design science research paradigm (Gregor and Hevner, 2013; Peffers et al., 2007; Von Alan et al., 2004). The prototype was developed in conjunction with regular drivers over the course of a year, comprising of surveys and an initial field test of eight drivers for approximately two weeks, where approximately 5000 km were driven.

The subsequent in-vehicle system was comprised of three core components: a bluetooth OBD-II dongle to access data from the vehicle, a smartphone application that acts as the user-interface for the driver, and a backend server infrastructure to collect, store, and process the driving data. First, the driving data from the vehicle is collected through the OBD-II dongle that was configured to access CAN Bus data of specific vehicles, this data is passed to the smartphone in the vehicle via a Bluetooth connection. Second, a smartphone application receives the CAN Bus data from

FIGURE 3.11: UML sequence diagram showing an overview of the information flow of the system.



the OBD-II dongle, and transmits this in real-time to the backend server for analysis, along with metadata, such as the GPS location associated with each measurement. For the majority of the time while driving, the user interface of the smartphone application displays a variety of eco-driving feedback for the driver, and the differences between the types of feedback has been investigated in a supporting study (Dahlinger, 2018; Dahlinger et al., 2018). The smartphone application additionally determines if and when the accident hotspot warning interventions should be displayed to the driver, changing the application screen when necessary from the eco-driving feedback when not in a hazardous area, to a contextual warning when an accident hotspot is being approached. Finally, the backend server architecture provides an updatable reference of the locations of historic accident hotspots which can be queried by the smartphone application, along with processing and storing the transmitted vehicle data for post-processing and analyses. The remainder of this section provides a more detailed description of each of these three components, and Figure 3.11 provides a UML sequence diagram showing an overview of the system.

3.2.1 Vehicle Data Collection

When access to the CAN Bus of a vehicle is not restricted, then the messages transmitted from the vehicle's internal sensors can be obtained. As such, vehicle data, representative of that available in modern connected and semi-autonomous vehicles, was collected via the retrofit system by accessing the CAN Bus of the vehicles via an OBD-II dongle. The dongle was configured to interpret the CAN messages on compatible vehicles, and was paired via Bluetooth with a smartphone in the vehicle. The signals and messages accessed are unstandardised and vary between makes and models of a vehicle, and the dongle was configured to transmit these at a maximum rate of 30 Hz per measurement. Based on the prior research into driving related insights, the following measurements were identified and collected for the purpose of measuring driver behaviour and potentially identifying hazardous areas:

- Engine Speed [rpm]
- Individual Wheel Speed [km/h]
- Vehicle Acceleration [m/s^2]
- Throttle Pedal Position [%]
- Brake Pedal Position [%]
- Steering Wheel Angle [°]
- Longitudinal Acceleration [m/s^2]
- Lateral Acceleration [m/s^2]
- Yaw Rate [°/s]
- Antilock Braking System [on/off]
- Electronic Stability Control [on/off]
- Traction Control System [on/off]
- Wheel Slip Status [on/off]
- Outside Temperature [°C]
- Windshield Wiper [on/off]
- Headlight Setting [on/off]

Additionally, the vehicle's fuel consumption measurements (Litres per hour) were collected and used to provide eco-driving feedback to the driver in the dedicated smartphone application. Finally, the driving data and CAN Bus signals received by the smartphone application were streamed in real-time via the phone's internet connection to a backend server, augmented with the smartphone's own GPS location, timestamp, accelerometer data, and metadata regarding the trip status and general usage information.

3.2.2 Smartphone Application

The visual user interface of the Android smartphone application was separated into two components, non-warning feedback and warning feedback. In order to encourage use of the system, a variety of non-warning information was capable of being provided to the user when they were not travelling through accident hotspots. Of this non-warning feedback, there were two types of eco-driving feedback available within the application. The first is concrete feedback that extends the vehicles dashboard and provides real-time fuel consumption information. The second is abstract feedback, where a tree grows and shrinks based on the average fuel consumption of the current trip. Figure 3.12 shows these two types of eco-driving feedback. Further, an additional 'g-radar' screen was also available, showing the longitudinal and lateral acceleration values that the vehicle was currently experiencing.

Prior driving studies have demonstrated that, in general, brief glances away from the forward roadway, for the purpose of scanning the driving environment and the

FIGURE 3.12: Non-warning feedback provided by the information system

(A) Concrete Eco-Driving Feedback (B) Abstract Eco-Driving Feedback

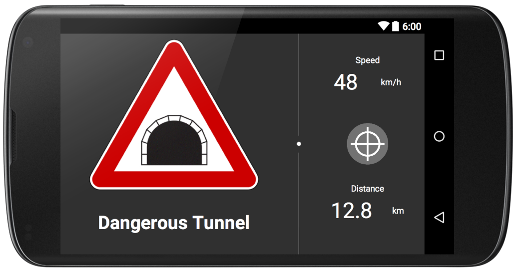
in-vehicle instrument cluster, are safe and decrease crash and near crash risk (Klauer et al., 2006). In addition, simulation studies that provided early warning signals to drivers approaching a hazardous location reported a positive shift towards safer behaviour, where participants receiving the intervention drove at a lower velocity after waiting longer at the intersection, and so avoided collisions (Werneke and Vollrath, 2013). In order to promote this safe driving behaviour, and encourage driver alertness through potentially hazardous areas, warning feedback is displayed via the smartphone application based on contextual information of the accident hotspot. This contextual information is based on the 'What', 'Where', and 'Why' attributes of the traffic accidents that together form the accident hotspot, and the method for generating this is described in greater detail in Section 4.3.1. As an example, a 'Caution Pedestrians' warning is provided when the majority of accidents making up the hotspot are reported to have included pedestrians. These interventions are provided to the drivers when they approach an accident hotspot, and the non-warning feedback is replaced by the relevant warning for that location, as illustrated in Figure 3.13.

Naturalistic driving studies typically observe driving behaviour of participants through cameras and other recording devices, and previous work has found that a short glance from the forward roadway for a simple task only increases marginally, if at all, the risk of a crash or near crash while driving (Klauer et al., 2006). However, the importance of non-intrusive interactions with the driver are also highlighted. Previous in-vehicle warning studies have shown that audio-based warnings may be as effective as both audio and visual warning information combined (Zhang, Suto, and Fujiwara, 2009). However, since it would have been difficult to detect and control for whether the driver had manually disabled the volume of the smartphone, audio warnings were not included in the experimental version of the application. Additionally, tactile warnings, e.g. vibrations delivered through the vehicle seat, throttle pedal, or seat belt, have previously been provided to drivers (Meng and

FIGURE 3.13: Two examples of the in-vehicle warning intervention shown to drivers approaching an accident hotspot

(A) Pedestrian Crosswalk Warning

(B) Dangerous Tunnel Warning



Spence, 2015), however at the time of development it was not possible to deliver tactile feedback through smartphone applications[2]. As such, only visual warnings were implemented as part the in-vehicle system considered in the research at hand. In order to reduce driver inattention due to the accident hotspot warning functionality, and convey actionable information to the driver, the visual warnings were implemented following the guidelines of previous studies (Cao et al., 2010).

3.2.3 Backend Infrastructure and Data Collection

In order to process and store the high volume and frequency of data being transmitted from each vehicle and smartphone, a backend server infrastructure was developed utilising the latest Big Data technology (Manyika et al., 2011). Apache Kafka, Storm, and Cassandra were used to process the vehicle data messages in real-time and store them for post-processing (Ranjan, 2014). Furthermore, in order to enable the updating of existing, and the addition of new, accident hotspot warnings if required, the locations and details of the each hotspot were stored on the server, and queried and loaded onto the smartphone whenever the vehicle was started.

Aside from sensor measurements from the vehicle and smartphone, additional data was generated and collected for each trip, such as the start and end location, along with notifications and metadata of when accident hotspots were encountered and warnings shown to the driver. All data was initially stored on the smartphone, and streamed to the backend infrastructure in real-time when a suitable internet connection was available. So as to not lose any data in the case of weak internet signal, a set

[2] Although it is worth noting that recent technology advancements have potentially enabled the option for delivering such tactile feedback to drivers, by incorporating the warning interventions into a vibrating smartwatch notification.

of transfer buffers and databases were implemented in the smartphone application in order to store information until the connection was restored.

3.3 Touring Club Suisse Field Study

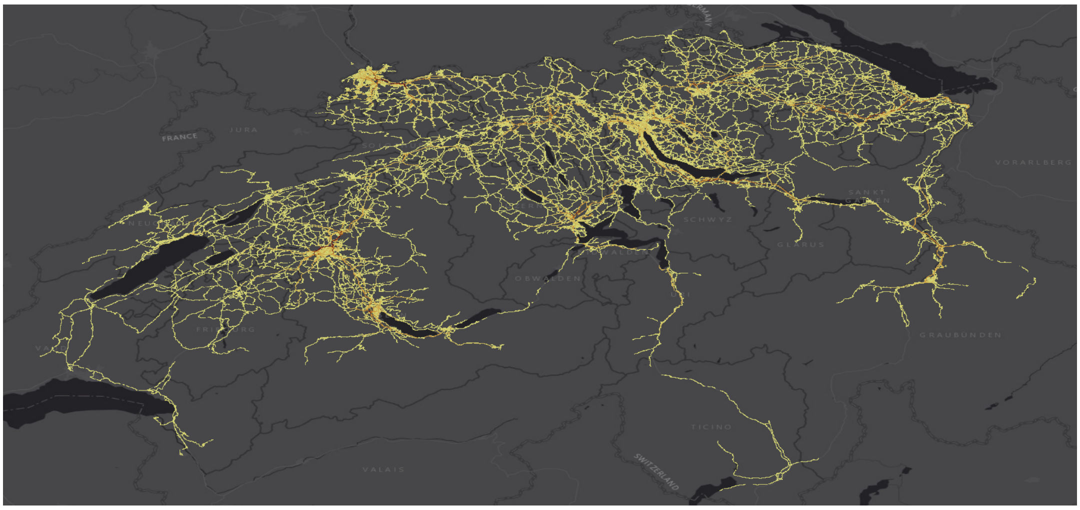
In order to answer the previously outlined research questions, and in contrast to the state-of-the-art in vehicle warning studies that focus on simulations and controlled experiments, the outlined in-vehicle warning system was tested in a country-wide field test of professional drivers. The following section provides an overview of this field study, along with details of the participant demographics, and the experimental design that was applied.

3.3.1 Field Study Overview

Professional drivers were recruited to participate in an 18 week field study from a fleet of road assistance patrollers, in cooperation with their employer, Touring Club Suisse (TCS), a Swiss road assistance club. Participation was voluntary and anonymous, and limited to German speaking drivers who drove cars compatible with the prototype system for extracting CAN Bus data from the vehicle, as described in Section 3.2.1. Out of the 92 potential patrollers that were invited via email to sign up for the study, 72 agreed to participate. As such, each of these drivers worked for the same roadside assistance company across a variety of locations, and all drove a Chevrolet Captiva of similar make and model.

During the 18 week period these drivers drove for approximately four hours, with an average of 144km travelled, per driver per day. Drivers followed their usual day-to-day routines, and did not undertake any specific driving tasks as part of the naturalistic driving field study. In total, this resulted in over 690 000 km of driving data collected using the system across the majority of the Swiss road network. Figure 3.14 shows the routes in Switzerland which were travelled by the participants during the study. Since the location data acquisition relied on the GPS position of the smartphone, in areas where there was no GPS signal, such as in tunnels, no location data was captured.

FIGURE 3.14: Locations in Switzerland where naturalistic field study driving data was collected over the 18 week period



3.3.2 Study Participants

Of the 72 recruited participants, 66 provided demographic details, such as age and gender, and completed a set of questionnaires regarding their personality and subjective driving style. Each of the participants providing this information was offered a voucher worth 100 Swiss francs as an incentive and compensation. All but one of the participants providing demographic data were male, and ranged from 21 to 64 years of age, with a mean age of 39.2 and a median of 37.

Existing research into the impact of driver personality on driving behaviour shows that various Big Five traits can be linked to subjective driving styles (Taubman-Ben-Ari and Yehiel, 2012). A previous study has shown that there are correlations between 'reckless' and 'angry' driving styles and high levels of Extraversion, and low Agreeableness and Conscientiousness. Additionally, high levels for Agreeableness, Conscientiousness, and Openness were correlated to the 'careful' driving style. Finally, the 'anxious' driving style was linked to high Neuroticism. Since a driver's personality may impact their behaviour (Taubman-Ben-Ari and Yehiel, 2012), the personality of the participants was measured in order to control for these factors when assessing the effect of the eco-driving and warning interventions. Therefore, the Big-Five-Inventory-10 (BFI-10) questionnaire (Rammstedt and John, 2007), a short version of the well-established Big-Five-Inventory (BFI) (John and Srivastava, 1999), was provided to the drivers. The BFI-10 consists of 10 items to cover the five personality factors, Agreeableness, Consciousness, Extraversion, Neuroticism and Openness, each with two items accordingly and measured on a Likert scale from 1 (very low) to 5 (very high). Psychometric properties do not reach the

TABLE 3.2: Descriptive statistics of the Big Five personality traits of the participants, measured on a Likert scale from 1 (very low) to 5 (very high)

Personality Trait	Mean	Standard Deviation	Median	Minimum	Maximum
Agreeableness	2.85	0.65	3.00	1.00	4.50
Conscientiousness	3.51	0.70	3.50	2.00	5.00
Extraversion	2.91	0.76	3.00	1.50	4.50
Neuroticism	3.50	0.65	3.50	2.00	5.00
Openness	2.94	0.56	3.00	1.50	4.50

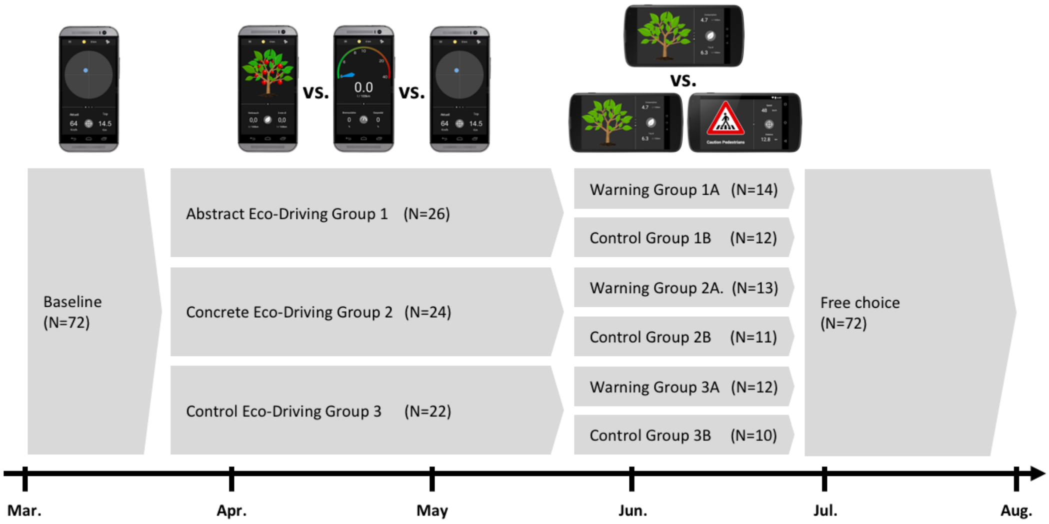
quality of the original BFI, but deliver sufficient values. The short version of the questionnaire was chosen due to restrictions on participant's time and to avoid attrition. Table 3.2 provides an overview of the distribution of each of the Big Five personality traits among the drivers. Additionally, the subjective driving style of the drivers was assessed using the multidimensional driving style inventory (Taubman-Ben-Ari, Mikulincer, and Gillath, 2004). This inventory measures driving style on a 7 point Likert scale across eight dimensions: 'angry', 'anxious', 'careful', 'dissociative', 'distress-reduction', 'high-velocity', 'patient', and 'risky'.

3.3.3 Experimental Design

The field study was conducted over a course of 18 weeks, and consisted of a 2 week baseline and on-boarding phase, a 12 week experimental phase during which two intervention experiments were performed, and a final 4 week 'free driving' phase where drivers could choose between the experimental interventions. The first experiment explored the impact of eco-driving feedback on driver behaviour, where, after the 2 week baseline and on-boarding phase, participants were randomly allocated into one of three groups and shown either a control intervention, concrete eco-driving feedback, or abstract eco-driving feedback[3]. After another 8 weeks, these three experimental groups of drivers were randomly split a second time, between a control group and a warning intervention group. This second group allocation was undertaken to control for the effect that the eco-driving intervention had on the group driving behaviour, and the eco-driving feedback of each driver remained the same during the two experimental phases. Randomisation checks were conducted at both of these group allocation stages, and indicated that there were no significant group differences with respect to drivers' demographic and subjective driving style

[3] For a detailed analysis of this phase, please see the two supporting studies from Dahlinger et al., 2018, and Dahlinger, 2018.

FIGURE 3.15: Overview of the field study and experimental setup, adapted from Dahlinger, 2018



survey responses. Overall, the warning experimental phase lasted for 4 weeks, after which all drivers had the warning intervention enabled and were given free choice between the three styles of eco-driving feedback for the final 4 weeks of the field study. Figure 3.15 provides an overview of this experimental setup, and the group sizes at each stage of the randomly assigned allocations.

Chapter 4

The Impact of Accident Hotspot Warnings on Driver Behaviour

The previous chapters have so far outlined the background work on driver safety and in-vehicle warnings, and presented the development of a prototype system that could in parallel display accident hotspot warnings to drivers and collect data from the car itself. The naturalistic field study setup has additionally been described, that enabled the collection of vehicle data and the analyses of the following chapter, which investigate the impact that the accident hotspot warnings had on the drivers. This chapter opens by revisiting the motivation of providing such warnings, and going into more detail on the latest studies that the analyses build upon. The specific materials and methods that were utilised to identify traffic accident hotspots from the historical accident data, and the generation of contextual warnings from these, are then described. Finally, the results of the intervention on the drivers' behaviour are presented, and the chapter closes with a discussion of the findings and the conclusions that can be drawn from the work.

4.1 Context and Motivation

According to the World Health Organization (WHO) road traffic accidents are now the eighth leading cause of death globally, with the number of road traffic fatalities steadily increasing since 2001 to over 1.25 million people each year (World Health Organization, 2015). An example of the growing risk can be seen in the United States, where, according to the National Highway Traffic Safety Administration (NHTSA), the number of deaths from traffic accidents in 2015 rose by 7 % from the year before, up to 35 092 fatalities (NHTSA, 2007). Aside from the humanitarian concerns of so many injuries and mortalities, the worldwide economic costs caused by the impact of traffic accidents are estimated to account for a loss of approximately 3 % of the global GDP (McMahon and Dahdah, 2008). As such, the United States Department of Transport, along with the White House, issued a call to action, encouraging the continuous research into different approaches that can help to reduce the number of traffic accidents, both fatal and non-fatal (U.S. Department of Transportation, 2016).

In light of the depicted challenges, a huge variety of research endeavours and systems have emerged to help tackle the problem of traffic accidents on our road networks. For example, due to the data requirements and the complexity of urban planning and transportation problems, there has been a growing interest in the use of new methods and tools to analyse the strategic planning (Coutinho-Rodrigues et al., 1997; Ülengin et al., 2007), the multi-vehicle tactical (Santos, Coutinho-Rodrigues, and Antunes, 2011) and the individual vehicle operational levels (Keenan, 1998; Ray, 2007). In particular, spatial systems have been shown to play a vital role in this domain, enabling a variety of analytics on the existing challenges of road infrastructure. An instance of such as a system was presented in an accident hotspot evaluation study that took place in India, where inadequate development of transport networks led to traffic congestions and accidents (Prasannakumar et al., 2011). In the study, geo-information technology was used to help examine the location and distribution of hotspots, highlighting the influence of spatial and temporal factors in their formation.

Furthermore, at the individual vehicle operational level, various research studies are geared towards how in-vehicle systems can encourage drivers to adapt their driving behaviour when necessary. The latest studies in this field provide promising evidence that these systems can indeed have significant positive effects on driving behaviour and collision avoidance (Kazazi, Winkler, and Vollrath, 2015; Tey et al., 2014; Werneke and Vollrath, 2013). Moreover, the benefits of these systems can be delivered to vehicles either directly through their own data connections and displays, or through existing mobile or standalone satellite navigation systems (Zhang, Suto, and Fujiwara, 2009).

4.1.1 Research Objectives

While the potential of in-vehicle systems is undisputed, the vast majority of studies have focused on simulation experiments (Naujoks and Neukum, 2014a; Seeliger et al., 2014) and controlled field studies (Ruscio, Ciceri, and Biassoni, 2015; Zhang, Suto, and Fujiwara, 2009), typically providing warnings to drivers to prevent a collision with an upcoming vehicle or pedestrian. In this field, the existing research falls short of both bringing an in-vehicle warning system into a field study setting, and utilising real world location analytics on traffic accident data as a source for generating in-vehicle warnings. As such, this chapter of the thesis goes beyond existing studies and depicts the design and field evaluation of an in-vehicle warning system that aims to improve driver safety at known dangerous locations. Thus investigating the first research objective of this thesis, determining whether warnings of upcoming dangerous locations have a positive effect on driving behaviour, as discussed in Section 1.2.1.

The existing studies that measure the impact of in-vehicle warnings through simulation environments and controlled field studies typically make use of a variety of features to assess the effect of warnings on driver behaviour. The effectiveness of these interventions are typically assessed using variables that are unique to a more controlled experimental setting, and examples include measuring either the number of collisions incurred or braking reaction time (Kazazi, Winkler, and Vollrath, 2015) Since the research objective is the field evaluation of a system that warns users of historically hazardous locations, rather than specific upcoming objects, many of the typical validation variables in this field are unavailable for analysis. For example, 'minimum-time-to-collision', the time left for a participant to avoid a collision with another object (Naujoks and Totzke, 2014), is unsuitable as the distance to other objects is unknown and there may be no need for a change in the driver's behaviour. Therefore, a more applicable dependent variable to measure the effect of the warnings are the number of potentially dangerous braking events incurred by the driver, which are a key result of decision making driving behaviour through hazardous locations. With this dependent variable, and the approach of applying real world location analytics on traffic accident data, the research objective can be operationalised to generate the first concrete research question of this thesis:

RQ 1a To what extent do warning interventions of upcoming dangerous locations, i.e. accident hotspots, impact the braking behaviour of drivers in a naturalistic driving setting?

As the warnings are designed to encourage awareness of historically dangerous areas, drivers that receive the interventions should be both more alert and effective at planning for situations ahead (Zhang, Suto, and Fujiwara, 2009). Therefore, the hypothesis of the research at hand is that when the warnings are shown to drivers approaching these accident hotspots then dangerous braking events are less likely to occur at these locations.

4.1.2 Research Implications and Structure

In order to assess the impact of the system, and in contrast to other in-vehicle studies, location analytics were applied to a national historical accident dataset, composed of over 266 000 accidents, and a complete in-vehicle warning system providing warning interventions to drivers deployed to the field. This system was tested outside of the simulation environment in a country-wide field test of 72 professional drivers in order to assess the impact of the system on driver behaviour and safety.

The supporting description of the prototype system, and the national traffic accident dataset utilised to generate the in-vehicle warnings have been previously discussed in Chapter 3. As such, the remainder of this chapter initially focuses on the state-of-the-art research with regard to in-vehicle systems that provide warnings to drivers, and proceeds to present the methods regarding the identification of the accident hotspots and the classification techniques to generate the contextual warning interventions. Subsequently, the data collection from the naturalistic driving experimental phase is presented, along with the generation of the 'heavy braking' dependent variable. Finally, the chapter concludes with an evaluation of the system with regard to its effect on safe driving behaviour, and a discussion of the results and implications of the research.

Overall, the topic at hand of driver safety and in-vehicle warning systems is highly relevant to both policy makers and industry players, such as vehicle manufactures and insurances. Numerous hardware-based vehicle safety systems have become mandatory in various countries throughout the last decades, for example, air-bags and electronic stability programs. Similarly, policy makers could consider promoting data-powered in-vehicle warning systems that encourage safer driver behaviour. Eventually, such warning intervention setups that have proven to prevent accidents could also be enforced by corresponding regulation. The automotive industry might also recognise that data-powered prevention services are positioned to be an effective means to address the distinct safety needs of consumers and form a basis for sustainable competitive differentiation.

4.2 State of the Art

4.2.1 Simulation Studies

Recent research has indicated that active warning signs, e.g. flashing lights, and in-vehicle warning devices both produced higher levels of driver compliance than the existing conventional roadside warnings, demonstrating the positive impact in-vehicle warning systems can have when compared to traditional road safety approaches (Tey et al., 2014). This overall positive effect is supported by various other simulation studies that demonstrate how in-vehicle warning systems, such as early warning signals displayed while approaching an intersection, can encourage safer driving behaviour (Werneke and Vollrath, 2013). Participants in the study that received the interventions were able to avoid collisions by waiting for a longer time at the intersection and turning with a lower velocity, thus adapting their behaviour to better reflect the driving situation. In particular, it has been reported that in-vehicle warnings led to a significant reduction in collisions during simulated critical situations, and that both older and younger drivers demonstrated improved braking reaction time during non-critical scenarios (Kazazi, Winkler, and Vollrath, 2015).

In other simulation studies, the effects on driving behaviour from advisory warnings were found to be strongly dependent on warning time, with earlier warnings more effective than late warnings (Naujoks and Neukum, 2014a). The interventions were more greatly appreciated by drivers when given earlier, even though in critical situations shorter warning times were still effective (Naujoks and Neukum, 2014b). In situations where there is low visibility of potential hazardous situations, the frequency of critical situations was reduced when early advisory warnings were provided, especially in surprising or unexpected situations (Naujoks et al., 2015). With regard to the types of warning that can be provided, contextual warnings had limited importance to the behaviour of the driver, but users rated the system much higher due to them (Naujoks and Neukum, 2014a). Additionally, it has been shown that in less critical situations a contextual caution warning sign is more suitable than a stop sign warning (Kazazi, Winkler, and Vollrath, 2015).

Furthermore, the recent developments in heads-up displays and the visualisation technology built into the dashboard of modern vehicles have started to enable even more innovative approaches to in-vehicle warnings. For example, simulation studies have started to explore how augmented reality systems can provide even greater improvements to driver behaviour than other in-vehicle approaches (Schwarz and Fastenmeier, 2017; Schwarz and Fastenmeier, 2018). When the researchers compared a control condition to augmented reality warnings with spatial referencing of a hazard, it was reported that drivers' reactions consistently improved, as well as their ability to identify unnecessary warnings. Participants additionally reported that they subjectively preferred scaling animations as part of the augmented reality setup, where warnings grew in size as the hazardous location was approached, however there was no impact on drivers' objective behaviour (Schwarz and Fastenmeier, 2017). Additionally, when measuring participants' gaze and braking reaction times, passing speeds, and collision rates, visual warnings with contextual information on the upcoming hazard were reported to have advantages over the three other warning designs (Schwarz and Fastenmeier, 2018).

4.2.2 Controlled Field Experiments

Several controlled field studies have investigated the impact of in-vehicle warnings on driver safety by going beyond the simulation environment and into more realistic driving situations. For example, the impact on real-life emergency braking has been investigated in a study with regard to automation complacency and warning expectancy (Ruscio, Ciceri, and Biassoni, 2015). The study involved 30 participants that drove a customised vehicle, equipped with a variety of measuring devices and cameras, around a small controlled track with no other road participants, where a simple collision warning system was tested, along with misleading warnings and unexpected events for the driver. Aside from the task of stopping as quickly as possible when the warnings were shown, the researchers tested automation complacency by, unexpectedly for the driver and without warning from the system, throwing a foam rubber cube into the path of the vehicle on a straight section of track before the end of the experimental task. The results of the study highlight that reliable warnings for drivers quickened the overall decision making process for emergency braking, and show that misleading warnings of upcoming collisions generated automation complacency, slowing the visual search for unexpected hazards[1].

Moreover, in a separate study an existing spatially located hazard on the Japanese road network was considered and the effects of visual and audio warning combinations on driver behaviour were investigated (Zhang, Suto, and Fujiwara, 2009). Poor visibility had previously led to traffic accidents occurring at a downhill road section of the intersection near an arch-shaped bridge, and this hazardous area was the source of warnings provided to drivers in the study as they approached the location. In the experiment that was conducted, 14 students drove with an observer on the identified highway, and were presented with either an audio warning, or a combination of an audio and heads-up display visual warning. Additionally, two types

[1] However, it is worth noting that the reported results on automation complacency should be further validated in a different setting, as when asked about the surprise appearance of the foam cube, several drivers reported that they either "had not seen the obstacle", "did not have time to realise there was an obstacle to be avoided", or "were not sure whether they had to brake or not, as they were not sure [whether] it was part of the experiment".

of intervention were evaluated, the first was static information, that warned only of the approaching traffic light infrastructure, the second was dynamic information, triggered manually by the researcher in the car, and warning of a vehicle ahead that was stopping. The authors demonstrated that the drivers were better able to avoid dangerous situations when the warnings included information about the causes of accidents, i.e. the dynamic information on stopping vehicles, rather than on the road infrastructure itself. The results also indicate that the audio voice-based interventions may be as effective in improving driver awareness as the combination of both the voice and heads-up display warnings.

4.3 Materials and Methods

In order to go beyond these existing studies, and provide meaningful warnings of traffic accident hotspots to drivers at a national level, two key prerequisite goals must be achieved. The first is the spatial identification of these dangerous areas, and the second is the classification of why each of the identified locations is dangerous. As such, the following section initially opens by outlining the data-mining and rule-based approaches that were utilised to tackle these two goals respectively, and the FEDRO traffic accident dataset that contained details of over 266 000 incidents occurring in Switzerland between 2011 to 2015. Subsequently, the logic and conditions of how and when the accident hotspot warnings were displayed to the drivers are outlined. The section then closes with a description of the experimental phase of the field study where the interventions were tested, along with the data and post-processing applied to generate the variables for analysis.

4.3.1 Accident Hotspot Identification and Classification

The literature review revealed multiple approaches to spatially identify accident hotspots, however, not all of them were compatible with the available dataset. The Empirical Bayesian approach, which is commonly applied by governmental institutions, is well known to produce good results. However, it is very sensitive to the quality of the estimation function and requires detailed information about risk variables, such as traffic volumes or road parameters, e.g. curve radii (Deublein et al., 2015). Since the provided dataset of the FEDRO consisted only of rough traffic flow estimates and lacked other road parameters, Empirical Bayesian was discarded as a suitable approach. Nonetheless, other promising analysis methods, such as spatial autocorrelation or Kernel Density Estimation, are appropriate for identifying locations where many accidents occurred. Meanwhile, current research suggests that spatial clustering techniques can achieve similar results, and have the advantage of being simpler to interpret and perform much more efficiently on large datasets (Szénási, 2015). As it is not trivial to identify the hotspot boundaries that Kernel Density Estimation generates, and therefore, identify the accidents which contribute toward a hotspot being formed, DBSCAN was selected as a natural density based clustering technique that clearly identifies observations contributing to a cluster.

DBSCAN classifies elements into clusters in such a way that inside a cluster the density of elements is higher compared to the outside of the cluster, and elements that are not part of any group are considered as noise (Ester et al., 1996). As such, identified clusters can be considered hotspots with a significantly higher density of accidents compared to other areas. Noise elements represent "random" accidents, which have no, or very little, spatial dependencies to other crashes. DBSCAN's performance in identifying clusters is very sensitive to the distance between points that are considered to be part of the same cluster (ε), and the minimum number of points which must be within ε to form together a cluster ($MinPts$). There exists no optimal choice of these parameters, and domain expertise is suggested to identify optimum values based on the intentions of the analysis. If ε is too small, only accidents occurring in very close proximity to each other will be considered as hotspots, and if too large, hotspots can grow in size and cover parallel roads. Likewise, if $MinPts$ is too high, only the most severe clusters are identified, and if too low, many small and "random" hotspots are found.

Therefore, the following practical approaches were considered when applying DBSCAN on the FEDRO dataset. The value of $MinPts$ was discussed with experts from one of the largest automotive clubs in Europe and hence defined with the following heuristic: to call a specific location a hotspot, on average, more than two accidents per year had to occur at that location. As a result, it was decided that $MinPts$ cannot be smaller than ten as the hotspots were formed out of a dataset covering accidents over five years. Finally, ε was fine-tuned by a visual inspection of a selection of accident hotspots, this was achieved using a map-based tool, which was purpose built for the research at hand, and enabled the validation the DBCAN parameters and provided statistical overviews of the clusters. The selection of hotspots that were considered in this review stage contained both examples which were closely connected to certain road infrastructure, and others were more spatially distributed. It was found that with $\varepsilon = 15m$ DBSCAN produced results where individual clusters did not span multiple roads. These parameters can be loosely defined by the following natural definition: For an observation to be included in an accident hotspot, at least ten accidents must have occurred within 15 meters of that location over five years. With these parameter settings for ε and $MinPts$, a total of 1608 unique accident hotspots were found in Switzerland from over 266 000 geo-located accident records.

With the accident hotspots across Switzerland identified, the second goal of this subsection is to generate contextual reasons for why each of the identified locations are considered dangerous, with the intention of providing drivers with meaningful warning feedback whenever they are approaching the accident hotspots, rather than a generic warning (Kazazi, Winkler, and Vollrath, 2015). Following the guidelines from the NHTSA, drivers should be provided with warning messages in the form of signs and non-critical supporting text (Campbell et al., 2007). The feedback information varies depending on the available contextual information derived from the spatially identified accident hotspots. The assumption is that drivers can directly and quickly relate the warning sign to the upcoming dangerous location. Additionally, the warning text should provide further non-time critical information, e.g. the predominant cause of the accident hotspot.

The contextual information of each accident hotspot was derived based on the corresponding accident protocols of the FEDRO. In these reports, police officers recorded all related accident information and determined, besides other details, the leading cause and type of the accident. In order to not overwhelm the drivers with too detailed or complicated warnings, a simple categorisation algorithm was developed. The detailed contextual information of each accident was summarised into three main categories: “What”, “Why”, and “Where”. “What” refers to the type of objects which were involved in the accident, e.g. cars, cyclists or pedestrians. “Why” refers to the predominant cause and type of the accident, e.g. disregarding right of way, speeding or swerving. Lastly, “Where” refers to the location information about the predominant type of road infrastructure at which the accident happened, e.g. at a crossroad intersection, roundabout or traffic light. In other words, where possible, information was captured about what objects were involved in the accident, why it happened and where it occurred. Furthermore, previous simulation studies have shown that warnings making use of contextual objects and directions are preferred by users (Naujoks and Neukum, 2014a). Therefore, when generating the warning, the preference of information primarily shown was ranked in the following order: “What”, “Why” and “Where”.

As such, the warning intervention was generated through a ranked majority-voting of the categorisation statistics of each accident hotspot (Lam and Suen, 1997). In order to capture accurate contextual information of a hotspot, more than 50 % of the accidents involved had to share the same predominant contextual detail information. Otherwise, a general warning sign and message was shown. Algorithm 1 outlines the classical DBSCAN pseudocode (Ester et al., 1996), which was modified so that once a hotspot was identified it was assigned a contextual warning type through the classification pseudocode also provided. In the majority of cases the official road warning signs of Switzerland were matched to the generated warnings and were used in the intervention. This was to prevent any potential confusion about the meaning of the shown warning messages. However, in some cases the creation of new symbols was inevitable. In total, six new signs were generated, following NHTSA standards (Campbell et al., 2007). This classification approach of the 1608 detected hotspots led to a total number of 20 different warning signs, and 36 unique combinations of sign and text. Figure 4.1 shows a selection of four different types of the accident hotspots that were detected using DBSCAN and this classification approach. The full list of sign and text combinations, in both English and German, can be found in Appendix B, and the top ten most commonly encountered combinations can be found in Chapter 4 in Table 4.1.

In order to empower a more cautious driving style, the warnings built into the application should be shown at a point in time that a driver has the ability to adapt their behaviour while approaching an accident hotspot. Not only this, but once the hotspot has been passed the warning should disappear as it is no longer providing useful information to the driver. These two conditions come with an added constraint, in that the warnings should not be shown too early, which could potentially create confusion for the driver, as well as not shown too late and prevent the driver from altering his or her behaviour. With these goals in mind, two approaches to determine when to present the warnings to the driver were explored, developed and tested.

The first of these approaches was a point-to-point map-matching algorithm, making use of freely available OpenStreetMap data as the digital data input (OpenStreetMap, 2016). OpenStreetMap provides digital map data in the following format, GPS points making up a road network are referred to as 'nodes' with a unique ID, and individual sections of roads, known as 'ways', are a collection of these nodes. Additionally, sections of road that intersect can be encoded using this approach through two or more ways sharing the same node. The implemented point-to-point map-matching technique found the closest way to a provided GPS coordinate, and then returned the closest node on the identified way. Using this method, the centre

Algorithm 1 Modified DBSCAN and Hotspot Classification Pseudocode

accidentHotspotDBSCAN(D, ε, MinPts)
  hotspot H = 0
  for each accident A in dataset D
    if A is visited **then continue** next accident
    mark A as visited
    Neighbours = all accidents within ε-radius around A
    if sizeOf(Neighbours) $<$ MinPts **then mark** A as NOISE
    else
      H = next hotspot
      expandHotspot(A, Neighbours, H, ε, MinPts)
      assignHotspotContext(H,D)

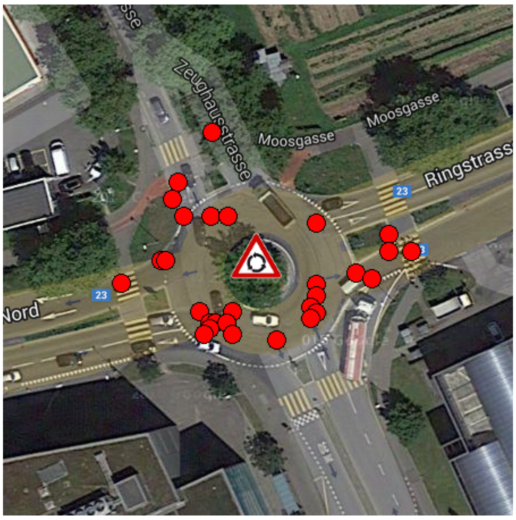
expandHotspot(A, Neighbours, H, ε, MinPts)
  add A to hotspot H
  for each accident A′ in Neighbours
    if A′ is not visited **then**
      mark A′ as visited
      Neighbours′ = all accidents within ε-radius around A′
      if sizeOf(Neighbours′) $>=$ MinPts **then**
        Neighbours = Neighbours union Neighbours′
    if A′ is not yet member of any hotspot **then add** A′ to hotspot H

assignHotspotContext(H, D)
  associativeArray WhatTypeCount =
    for each whatType T in dataset D **initialise** with 0
  associativeArray WhyTypeCount =
    for each whyType T in dataset D **initialise** with 0
  associativeArray WhereTypeCount =
    for each whereType T in dataset D **initialise** with 0

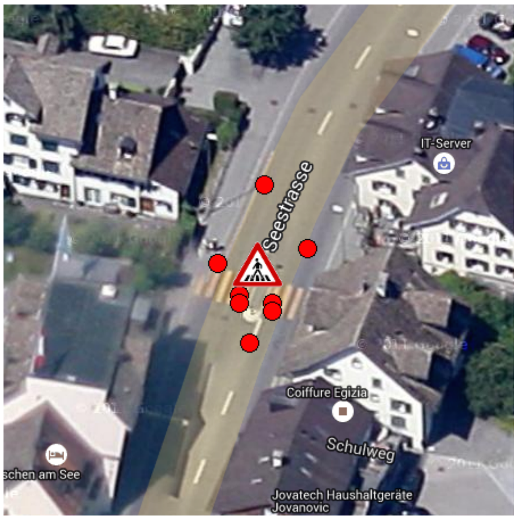
  for each accident A in hotspot H
    increment WhatTypeCount(whatType(A))
    increment WhyTypeCount(whyType(A))
    increment WhereTypeCount(whereType(A))

  if maxCount(WhatTypeCount) $>$ 50% sizeOf(H) **then**
    set warning W of hotspot H to maxType(WhatTypeCount)
  else if maxCount(WhyTypeCount) $>$ 50% sizeOf(H) **then**
    set warning W of hotspot H to maxType(WhyTypeCount)
  else if maxCount(WhereTypeCount) $>$ 50% sizeOf(H) **then**
    set warning W of hotspot H to maxType(WhereTypeCount)
  else set warning W of hotspot H to ‘General’

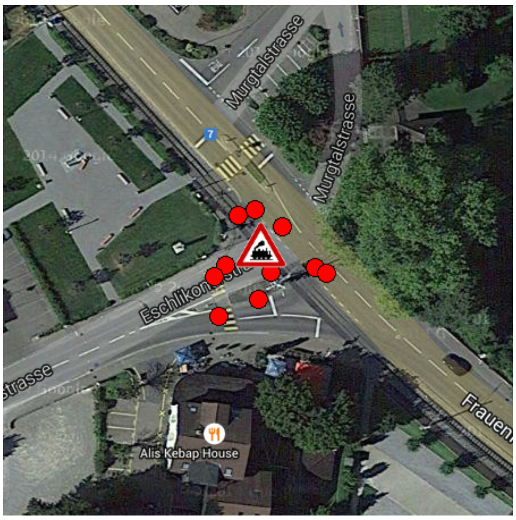
FIGURE 4.1: Selection of accident hotspots identified through DB-SCAN, accidents contributing to the hotspot are shown in red



(A) Roundabout Hotspot



(B) Pedestrian Crossing Hotspot



(C) Train Hotspot

(D) Rear-end Hotspot

of each accident hotspot can be mapped to a representation of the road network, with each hotspot assigned a unique node ID, which is then the basis for providing warnings for specific sections of road. The same point-to-point map-matching technique is additionally suitable for identifying a vehicle's current road segment in real-time, and then determining if an accident hotspot is being approached and if a warning should be provided to the driver. For example, a background process can find the closest way to vehicle's current GPS coordinates, and then return any warnings associated with nodes which are encompassed in this way.

The second, and ultimately more robust approach, was to develop a 'warning cone' to determine whether a vehicle was approaching an accident hotspot. This cone effectively projects a digital circular slice ahead of the vehicle, which adapts in size and shape depending on the speed and bearing of the vehicle, and is updated with each new GPS measurement. The logic behind this approach is that if an accident hotspot falls within this cone, then the vehicle is on a trajectory towards the hotspot and a set of rules can be applied to determine whether or not to show the associated warning. This warning cone is demonstrated in Figure 4.2, and alters its shape and size based on the following driving parameters:

- The cone length L [m], or radius of the circular sector from vehicle to outer edge, is determined by the speed S [km/h] of the vehicle, and the desired warning time T [s] by applying the equation:

$$L = T \cdot \frac{S}{3.6} \tag{4.1}$$

- The cone angle θ [°] is determined by a fixed cone width W [m], and the cone length L [m], and determined by the following formula:

$$\theta = \frac{\sin\left(\frac{W}{L}\right) \cdot 180}{\pi} \tag{4.2}$$

- The steering angle offset γ [°] is calculated with the current steering wheel angle β [°] and the maximum steering wheel angle α [°], and is subtracted from the vehicle's current bearing in order to take into account the expected direction that the vehicle will continue in. The steering angle offset is defined by:

$$\gamma = \operatorname{sgn}\beta \cdot \log|\beta| \cdot 2\pi \cdot \frac{360}{\alpha} \tag{4.3}$$

Extensive user testing of this approach led to the warning cone parameters being set to $T = 15[s]$ and $W = 90[m]$, along with the addition of a circular area surrounding

FIGURE 4.2: Accident Hotspot Warning Cone Example Sequence
Red Google Maps marker indicates centre of accident hotspot

(A) vehicle speed: 122 km/h
cone length: 507 m
cone angle: 14°
steering angle offset: −11°
hotspot in cone: *FALSE*
GPS measurements in cone: *N/A*
warning shown: *FALSE*

(B) vehicle speed: 103 km/h
cone length: 429 m
cone angle: 14°
steering angle offset: 7°
hotspot in cone: *FALSE*
GPS measurements in cone: *N/A*
warning shown: *FALSE*

(C) vehicle speed: 77 km/h
cone length: 321 m
cone angle: 14°
steering angle offset: 20°
hotspot in cone: *FALSE*
GPS measurements in cone: *N/A*
warning shown: *FALSE*

(D) vehicle speed: 58 km/h,
cone length: 244 m,
cone angle: 15°
steering angle offset: −21°,
hotspot in cone: *FALSE*,
GPS measurements in cone: *N/A*
warning shown: *FALSE*

(E) vehicle speed: 52 km/h
cone length: 216 m
cone angle: 16°
steering angle offset: −22°
hotspot in cone: *TRUE*
GPS measurements in cone: *1*
warning shown: *FALSE*

(F) vehicle speed: 51 km/h
cone length: 213 m
cone angle: 17°
steering angle offset: −17°
hotspot in cone: *TRUE*
GPS measurements in cone: *2*
warning shown: *TRUE*

the vehicle with a radius of 50 m. Further, warnings were displayed only when the GPS centre of a hotspot was within this cone or radius for at least two GPS measurements from the vehicle, which added confidence that the hotspot was about to be encountered. These warnings would remain shown to the driver until the accident hotspot had been passed, i.e. when the centre of the hotspot no longer falls within the cone or the radius surrounding the vehicle, and always remained visible for at least 6 s. With these parameters, warnings were reliably shown in a good time frame as a driver approached an accident hotspot.

When comparing the two approaches, significant user-experience difficulties were encountered using the map-matching technique. The first of these was that it did not take into account the direction of travel, and so warnings would remain in view for longer than desired once the accident hotspot had been passed, and it was challenging to ensure the warnings were provided with enough time for the drivers to alter their behaviour. The second disadvantage was that the implemented setup was limited to performing the real-time map-matching process on a backend server, and so became reliant on a constant internet connection, this occasionally led to a delay in receiving the warnings when the connection was poor. Although a map-matching approach could have been developed to run locally on the smartphone, due to resource constraints it was decided that the cone provided both a better user-experience and more reliable warning results. However, it is worth noting that both of these approaches suffered weaker performance in situations where there was poor GPS signal or the road network featured multilevel sections, i.e. a bridge crossing over a highway, as in the prototype stage neither could distinguish between road levels.

In order to help address the multilevel issue, a simple rule based approach was added to the smartphone implementation to ensure sensible warnings were shown to the driver given the driving situation. As the maximum speed limit in Switzerland for non-highway roads is 80 km/h, the vehicle was determined to be driving on a highway when travelling over 90 km/h, and so infrastructure and object warnings not associated with highways could be disregarded. These warning locations included: bus stops, crossroads, parking, roundabouts, and traffic lights. Additionally, warnings for pedestrians and pedestrian crossings, rail crossings, trains, and trams were also not shown in highway driving situations.

FIGURE 4.3: Two examples of the in-vehicle warning intervention shown to drivers approaching an accident hotspot

(A) Pedestrian Crosswalk Warning (B) Dangerous Tunnel Warning

4.3.2 Accident Hotspot Warning Generation

Finally, regarding visual warnings, the NHTSA released a guideline for the design of crash warning devices (Campbell et al., 2007). The project reflects a review of the human factors associated with the implementation of such warning system interfaces, and the lessons learned were then developed into guidelines for interface design. The highest efficiency was achieved by the choice of a discrete display, providing binary on-off information, and symbol or icon based information. Additionally, it was reported that the alphanumeric display type led to poor results and is commented with "Only appropriate for non-time-critical complex information". Based on these design suggestions, the warning sign is primarily displayed with the additional non-time-critical warning text below. Visual warnings are displayed on the smartphone application as the driver approaches an accident hotspot, and remain until the area surrounding the hotspot is passed. As earlier warnings are more effective and greater appreciated by drivers than late warnings, the warnings were shown up to 15 seconds before a driver encountered an accident hotspot (Naujoks and Neukum, 2014a; Naujoks and Neukum, 2014b). Figure 4.3 shows examples of the in-vehicle warning intervention provided to the drivers when approaching an accident hotspot.

4.3.3 Field Study Description

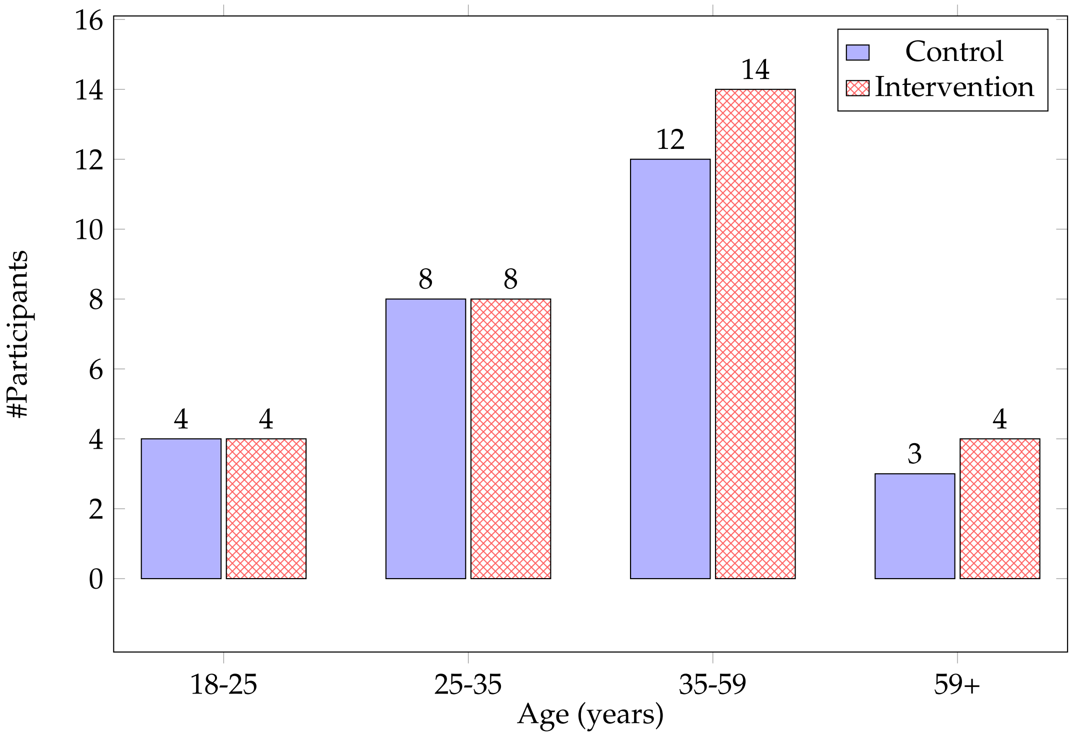
The impact of the in-vehicle warning system, presented in Section 3.2, and the described accident hotspot warnings on users' decision making behaviour while driving was assessed as part of the 18 week field study of professional drivers, as outlined in Chapter 3. Each of the drivers worked for the same firm across a variety of locations, and drove a company issue Chevrolet Captiva, all of similar make and model. After the 8 week eco-driving study the participants were randomly split between a control group ($N = 33$) and a warning intervention group ($N = 39$). As the eco-driving intervention of each participant remained the same during the warning experimental phase, this second random group allocation controlled for the effect that the eco-driving intervention had on the driving behaviour.

Of the 72 recruited participants, 6 did not provide important demographic details, such as age and gender. Randomisation checks on the remaining 66 drivers indicated that there were no significant group differences with respect to drivers' demographic and subjective driving style survey responses, as detailed in Section 3.3. A further 9 of the 66 drivers were excluded from the subsequent analyses as they drove significantly less during the four week experimental period than the other participants, i.e. because of planned vacation. This resulted in roughly even split of the remaining 57 drivers suitable for analyses between the control group ($N = 27$) and intervention group ($N = 30$). During the four week period, over 170 000 km were driven using the system, with an average of 144 km travelled per driver per day.

Of the 57 participants, all were male and ranged from 21 to 64 years of age, and Figure 4.4 shows the distribution of driver age between the control and intervention groups, using the same categories as previous accident analysis studies (Paefgen, Staake, and Fleisch, 2014; Paefgen, Staake, and Thiesse, 2013). The majority of drivers (45.61 %) fall between the ages of 35 to 59 years, with a mean of 40.3 and a median of 39 years of age. Furthermore, existing research into the impact of a driver's personality on driving behaviour linked various Big Five traits to four identified driving styles (John and Srivastava, 1999; Taubman-Ben-Ari and Yehiel, 2012). Previous studies found that the 'anxious' driving style was linked to high Neuroticism, and high levels for Agreeableness, Conscientiousness and Openness were correlated to the 'careful' driving style. Additionally, the reported results found correlations between high levels of Extraversion, and low levels of Agreeableness and Conscientiousness, to the 'reckless' and 'angry' driving styles. As such, the personality of participants was measured at the beginning of the study with the Big-Five-Inventory-10 (BFI-10) questionnaire, which consists of 10 items to cover the five personality factors, each with two items respectively, and is measured on a Likert scale from 1 to 5 (Rammstedt and John, 2007).

As previously described, the existing studies measuring the impact of in-vehicle warnings through simulation environments and controlled field studies made use of a variety of features to assess the effect of warnings on driver behaviour which are unsuitable in a naturalistic setting. These include variables such as the time left for a participant to avoid a collision with another object, counting the number of

FIGURE 4.4: Chart showing distribution of age amongst the drivers in the control and intervention groups



collisions, or measuring braking reaction time (Kazazi, Winkler, and Vollrath, 2015; Naujoks and Totzke, 2014). As such, the dependent variable in this study is the effect of the warnings on potentially dangerous braking events incurred by the driver, a key result of decision making driving behaviour. Chapter 2 presented various methods of capturing insights from driving data, including accessing smartphone accelerometer data (Johnson and Trivedi, 2011) and on-board diagnostics (OBD-II) standardised data (Imkamon et al., 2008). However, data available on the Controller Area Network (CAN) Bus of vehicles can give deeper insights into a vehicle's operation, for example, by directly accessing the values of the built in sensors of the vehicle while driving, characteristics of aggressive and calm driving can be identified (Karaduman et al., 2013).

As such, braking behaviour is captured from the vehicle's longitudinal acceleration sensor, accessed on the CAN Bus of each of the field study vehicles via an OBD-II Bluetooth dongle and transmitted to the smartphone in the vehicle. The current driving speed was additionally collected, and calculated from averaging the individual speeds of each of the four wheels. Braking events have previously been categorised as Low Danger, Dangerous and High Danger levels, based on thresholds of deceleration values (Bergasa et al., 2014). Low Danger events are those where vehicle deceleration was between $1.0\,\mathrm{m/s^2}$ to $2.0\,\mathrm{m/s^2}$, Dangerous events between $2.0\,\mathrm{m/s^2}$ to $4.0\,\mathrm{m/s^2}$ and High Danger events as greater than $4.0\,\mathrm{m/s^2}$. As the system provides warnings in historically hazardous locations, Low Danger braking behaviour may be unavoidable in many situations, and a large portion of events are expected to fall into this category. Therefore, only the deceleration events over $2.0\,\mathrm{m/s^2}$ are considered, capturing Dangerous and High Danger level braking events that are above the level that occupants feel comfortable experiencing as passengers (Hermann, Brenner, and Stadler, 2018). When reviewing the brake events that occurred as drivers crossed each of the accident hotspots during the experimental phase of the field study, in only 0.28 % of the cases were more than one dangerous braking event observed. Thus, a binary measure was applied to generate the dependent variable, i.e. whether or not one or more dangerous events were experienced while the hotspot was encountered.

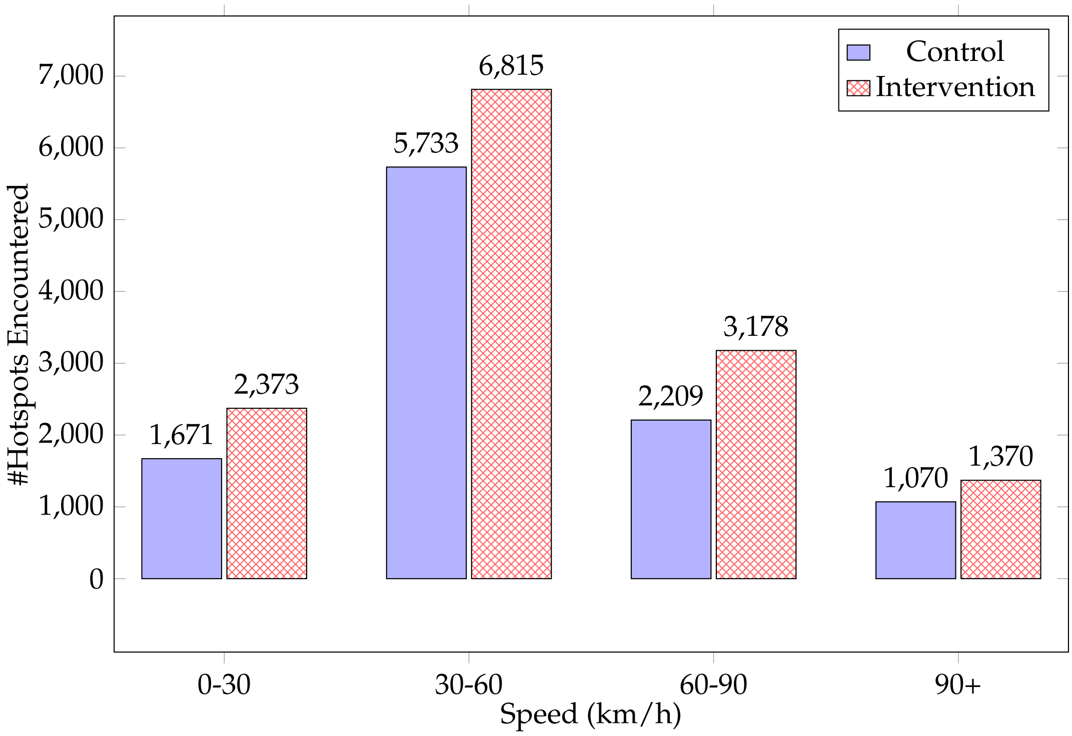
The evaluation was conducted through collecting vehicle sensor data during times that the warning intervention was shown to the driver. In the case of the control group, data were collected while the warning would have been shown, i.e. when drivers crossed an identified hotspot but no warning was shown. Finally, erroneous observations of encountered hotspots were cleansed from the dataset in certain situations, i.e. where there were issues with the sensors in the vehicle and data were not collected. This led to a total of 24 419 observations of encountered hotspots; 10 683 in the control group where no intervention was provided, and 13 736 in the intervention group where the location based accident hotspot warnings were shown. Table 4.1 shows the top ten most commonly encountered warning interventions across the study. The table additionally shows the number of hotspots encountered, as well as occurrences of one or more heavy braking events at the locations, for both the control and intervention groups.

Along with the sensor data, various other variables were collected which have been shown to have an effect on the likelihood of a traffic accident occurring (Paefgen, Staake, and Fleisch, 2014; Paefgen, Staake, and Thiesse, 2013). These values include the time of day, the day of the week and the speed that the vehicle was travelling when the hotspot was encountered. For comparison, these variables are categorised into bands on the basis of previous studies (Paefgen, Staake, and Fleisch, 2014; Paefgen, Staake, and Thiesse, 2013). As shown in Figure 4.5, the majority of hotspots were encountered travelling between 30 km/h to 60 km/h, with similar distributions between the control and intervention group. Figure 4.6 additionally shows

TABLE 4.1: The top ten most commonly encountered hotspot types, for both the control and intervention groups, along with the frequency they were encountered by drivers, and occurrences of one or more heavy braking events at the locations

Hotspot Warning		Total		Control Group		Intervention Group	
Sign	**Text**	**Count**	**Events**	**Count**	**Events**	**Count**	**Events**
	Disregarding Right of Way	4541	777	2135	384	2406	393
	Dangerous Crossroad	4378	783	2022	354	2356	429
	Rear-end Collisions	3532	375	1407	138	2125	237
	Disregarding Traffic Light	2037	284	951	120	1086	164
	Control Speed	1927	87	586	36	1341	51
	Caution Dangerous Area	1471	112	803	60	668	52
	Swerving Accidents	964	108	317	35	647	73
	Disregarding Right of Way	931	200	438	101	493	99
	Dangerous Roundabout	880	201	407	95	473	106
	Caution Cyclists	689	112	342	59	347	53
OTHER		3069	349	1275	150	1794	199
TOTAL		24419	3388	10683	1532	13736	1856

FIGURE 4.5: Distribution of speed and number of hotspots encountered for the control and intervention groups

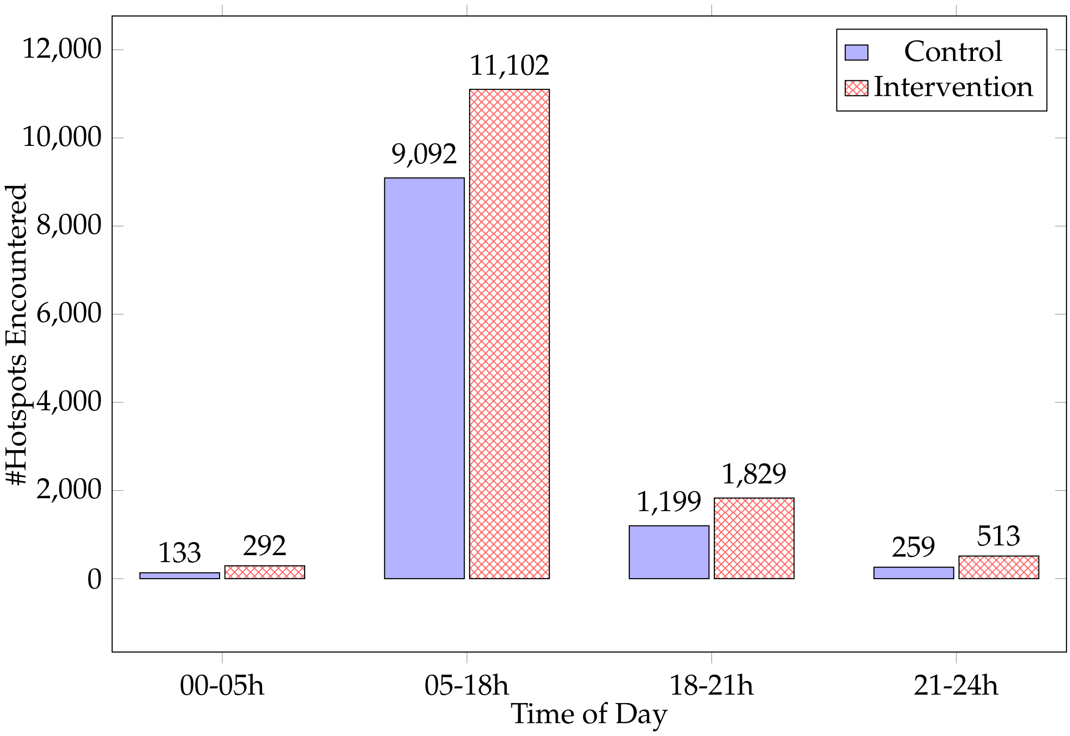


that the drivers were typically operating, and so crossing the accident hotspots, between the 'working hours' of 5 am and 6 pm, and that both the control and intervention groups again follow a similar distribution. In addition, an incremental count was collected for each driver in the intervention group every time they were shown the warning intervention for each specific accident hotspot, effectively generating a 'number of warnings shown' variable for each driver per hotspot. For the control group, this count variable was therefore always zero.

4.4 Results

In order to account for the impact of the individual drivers among the control and intervention groups, multilevel mixed-effects logistic regression was utilised (Snijders, 2011), and the dependent variable was a binary measure of whether one or more dangerous braking events occurred while each accident hotspot was encountered. Seven models were iteratively developed to test the impact that the in-vehicle accident hotspot warnings had on the braking behaviour of the drivers, and these regression results are shown in Table 4.2. In the remainder of this section, each of these models and the insights that they present are discussed.

FIGURE 4.6: Distribution of time of day and number of hotspots encountered for the control and intervention groups



Firstly, in Model (1) the regression was run with only the independent variable 'warning' capturing whether or not the warning intervention was provided to the driver, and thus the difference between the control and intervention groups. Here a significant impact is not observed of the warnings provided to the drivers on their braking behaviour. This indicates that when only comparing the behaviour between the control group and the intervention group, the occurrence of a warning had no significant impact on driver safety. Thus, it was not possible to confirm the immediate positive effect of warnings seen across many lab studies.

Exploring this further in Model (2), an additional independent variable 'number of warnings' was added to the previous analysis. This variable describes the number of times a driver in the intervention group had been shown the warning for a specific accident hotspot. This way the learning effect that repeated warnings of the same area had on a driver can be explored. Instead of linear effects of the number of warnings experienced, the impact of repeated warnings shown is expected to decrease with each additional warning experienced. Thus, in line with existing studies (Paefgen, Staake, and Fleisch, 2014; Paefgen, Staake, and Thiesse, 2013), the time effects of the variable are explored using a logarithmic transformation. In this model it can be observed that the number of times a warning has been shown has a significant effect on the likelihood of dangerous braking events occurring. This indicates that the more times a driver is exposed to the same warning in a hazardous area then the more cautious that driver operates, and a dangerous braking event is less likely to occur. Overall this is a positive result, and shows that the accident hotspot warning intervention has a significant learning effect on driving behaviour over time, but not an immediate short-term one.

Various models were analysed to test whether the significance of the learning effect of the in-vehicle warnings remained stable. Additional independent variables capturing the speed that the vehicle was travelling when the warning was shown, the time of day, day of week, driver age, and each of the Big Five personality traits of the driver were considered individually. When incorporating the speed as a predictor to generate Model (3), the immediate warning effect continues to be insignificant, and the learning effect remains. Additionally there are significant variations in the likelihood of a heavy braking event based on the speed that an accident hotspot was approached at. The stability of the learning effect was also tested with temporal variables in Model (4), which have historically been found to influence the rate of crashes (Paefgen, Staake, and Fleisch, 2014). Both time of day and day of week categorical independent variables were incorporated into the regression model. The short-term warning continues to be insignificant, and the learning effect remains at the same level as seen in Model (2). Overall, in contrast to previous studies, no time of day category was more or less dangerous at a significant level when compared to the '00 - 05 h' category. The only significant temporal effect was comparing the days of the week, Monday compared to Tuesday, where on Tuesday it was found to be more likely to incur a dangerous braking event.

Various studies have shown the effect of a driver's age and personality factors on driving related behaviour (Taubman-Ben-Ari and Yehiel, 2012). In order to incorporate this into the model, the age categories and personality information of the drivers were added in Model (5). As with the other models, the short-term effect of the warnings remained insignificant and the learning effect remained significant at a similar level to Models (2) and (4). Although investigating the effect of driver personality on driving behaviour is not the primary aim of this thesis, a significant effect of the Agreeableness trait is observed to reduce the likelihood of a dangerous braking event. This seems to confirm findings from a previous study (Taubman-Ben-Ari and Yehiel, 2012), where low levels of Agreeableness are correlated to 'reckless' and 'angry' driving styles, and high levels correlate to 'careful' driving behaviour.

Each of the additional independent variables discussed in Models (3), (4) and (5) were merged into the combined Model (6). The insignificant immediate effect of the

TABLE 4.2: Binary Logistic Regression Odds Ratio and Significance. Dependent Variable: Occurrence of a Dangerous or High Danger Braking Event. N = 24,419

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Warning	0.950	1.115	1.081	1.124	0.997	0.976	1.021
Number of Warnings		0.892***	0.922**	0.892***	0.891***	0.921**	1.276
Speed (vs. 0-30 km/h)							
30-60 km/h			2.008***			2.035***	2.031***
60-90 km/h			1.611***			1.632***	1.632***
90+ km/h			0.472***			0.477***	0.479***
Time of Day (vs. 00-05 h)							
05-18 h				1.038		1.054	1.049
18-21 h				0.874		0.876	0.876
21-24 h				0.768		0.759	0.765
Day of Week (vs. Monday)							
Tuesday				1.183*		1.212*	1.219*
Wednesday				1.078		1.102	1.108
Thursday				1.067		1.094	1.101
Friday				1.107		1.126	1.128
Saturday				0.959		0.970	0.976
Sunday				1.127		1.143	1.148
Driver Age (vs. 18-25)							
25-35					0.781	0.718	0.734
35-59					0.803	0.717	0.714
59+					0.895	0.813	0.830
Driver Personality							
Agreeableness					0.913*	0.913*	0.954
Conscentiousness					1.038	1.034	1.023
Extraversion					1.001	0.996	1.006
Neurotiscism					1.046	1.049	1.037
Openness					0.974	0.986	0.985
Driver Age Interactions (vs. 18-25)							
Number of Warnings × 25-35							0.930
Number of Warnings × 35-59							1.000
Number of Warnings × 59+							0.978
Driver Personality Interactions							
Number of Warnings × Agreeableness							0.944*
Number of Warnings × Conscentiousness							1.013
Number of Warnings × Extraversion							0.988
Number of Warnings × Neurotiscism							1.005
Number of Warnings × Openness							0.994
Constant	0.164***	0.164***	0.103***	0.151***	0.266**	0.158***	0.128***
L1 error	0.354***	0.350***	0.338***	0.353***	0.324***	0.314***	0.317***

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

intervention and the learning effect continues to remain stable across all the models, showing a reduction in the likelihood of a dangerous braking event. The other significant effects discussed in the previous models also remain in the fully merged model, and no new features become significant.

Finally, existing evidence suggests that, aside from the previously discussed impact on general driving behaviour, an individual's characteristics, such as age and personality, are important variables which might affect the generalisability of the findings, for example, in the form of moderators (Sprotles and Kendall, 1986; Taubman-Ben-Ari and Yehiel, 2012; Tiefenbeck et al., 2016). Thus, Model (6) was enhanced with interactions between the observed learning effect and age, as well as personality. Specifically, the interactions between the 'number of warnings' and driver age and personality variables were added to generate Model (7). Here it can be seen that neither the learning effect nor the effect of Agreeableness remain significant. However, the interaction between the two is. This interaction indicates that the learning effect previously identified is dependent on an individual's level of Agreeableness, where only those with reasonable levels improve their driving behaviour due to the warnings provided by the system.

4.5 Discussion

In summary, the analyses presented on data from a large naturalistic field study demonstrate that in-vehicle warnings of accident hotspots can have a significant improvement on driver behaviour over time. However, neither an immediate positive nor negative effect of the warnings on dangerous braking behaviour of the drivers was observed. When investigating generalisability on the basis of interactions, results indicate that the learning effect requires an adequate level of driver Agreeableness. The Agreeableness personality trait is linked to characteristics such as cooperation and social harmony (John and Srivastava, 1999). Hence, drivers who do not accept advice from an in-vehicle system (lack of willingness to cooperate), or do not care or reflect that they might harm others (lack of social harmony), might not benefit from such a system. Thus, in order to improve driver safety, future work should investigate the key determinants of, and how to best facilitate, the learning effects of such in-vehicle warning systems (Bokhove and Drijvers, 2012; Brendryen and Kraft, 2008). Finally, although the impact of vehicle speed on the likelihood of heavy braking events is not a focus of this thesis, a significant effect is observed in the analyses, which is similar to previous work that considers the influence of velocity on the exposure-accident relationship (Paefgen, Staake, and Fleisch, 2014). In line with these findings, the literature generally associates higher velocities with a greater risk of accident involvement (Aarts and Van Schagen, 2006). This is primarily due to larger stopping distances and reduced manoeuvrability at higher speeds.

The presented work has implications both for researchers and practitioners. From a research perspective, the learning effect observed is well known with regard to digital interventions. Similar long-term effects are seen in other domains, such as health and education (Bokhove and Drijvers, 2012; Brendryen and Kraft, 2008), where significant effects are reported the more often an intervention was triggered. The results further confirm the importance of measuring personality traits when researching interventions with any kind of system that supports decision making behaviour. Personality traits have long been recognised as a strong predictor of human decision-making outcomes (Sprotles and Kendall, 1986). However, research on real-time feedback interventions have only recently considered the impact of personality as a key factor in human behaviour (Tiefenbeck et al., 2016). Additionally, the results emphasise the importance of field research. The large effects that are often reported from very controlled settings have to be verified under real-world conditions to ensure generalisability.

On a more general note, the interaction of Agreeableness and the learning effect that can be seen in the results trigger a call for action towards the personalisation of support systems in general, as the effectiveness of digital interventions vary according to an individual's characteristics (Tiefenbeck et al., 2016). However, measuring personality traits is inconvenient and often perceived as intrusive by the user (Tourangeau and Yan, 2007). As such, one can either seek to identify the user's personality unobtrusively, or rely on the stable effect of the warnings over time and on the consumers' self-selection.

Importantly, the results of this study should be seen in the light of its limitations. While there are many thresholds for heavy braking values suggested in the related research, as well as other methods to identify and measure safe driving behaviour, in this early stage just one threshold was considered in order to generate the dependent variable. Furthermore, the described system makes use of historical accident data from a national dataset, restricting the adaptation of this approach to regions with similar sources of information. However, there is increasing work in both research and practice to utilise naturalistic driving data for traffic accident analysis. As such, Chapter 5 of this thesis discusses and investigates the potential of identifying such accident hotspots from near-miss events detected through connected vehicles.

Finally, this research is geared towards the development and validation of an innovative artefact. In accordance with this goal and in conformance with latest discussions in the scientific community (Von Alan et al., 2004), the chapter does not focus on theory. Future research should cover theoretical models of human behaviour to further increase generalisability of the findings.

4.6 Conclusions

Overall, in-vehicle systems can encourage drivers to adapt their driving behaviour when necessary, and have therefore been the focus of various research endeavours. Latest studies provide promising evidence that these systems can indeed have significant positive effects on driving behaviour and collision avoidance. Going beyond the existing research, a complete in-vehicle warning system was designed and implemented, which provided accident hotspot warning interventions to drivers based on location analytics applied to a national historical accident dataset. The system was tested with 57 drivers in a field test covering over 170 000 km. As such, this thesis is among the first to bring research on in-vehicle systems and warnings for drivers into the field in a realistic experimental setting.

Ultimately, the results show that in-vehicle warnings of accident hotspots can have a significant improvement on driver braking behaviour over time. Thereby demonstrating that such systems can play a fruitful role in the field of connected vehicles, a domain which has traditionally not been a core focus of the decision support and information systems research fields. In addition, the positive intervention effects are bound to drivers' Agreeableness, i.e. drivers have to be willing to "listen" to the in-vehicle system. Hence, future research should carefully reflect the role and impact of subjects' personality. Moreover, there appears to be potential for fields such as design science research to develop and validate effective strategies that help to overcome technology adoption challenges, which are based on a lack of Agreeableness (Peffers et al., 2007). Finally, in contrast to existing lab experiments with very promising results, an immediate effect of warnings on driver behaviour was not able to be confirmed, and thus demonstrates the importance of building innovative artefacts and conducting experimental research in a realistic field setting. As such, there remains a strong need for further field experiments with high resolution vehicle data in order to determine whether the impressive results of existing lab-based studies can deliver in diverse field situations.

In practice, safety-focused services such as the one presented could particularly benefit the automotive insurance industry, where the ever-increasing digitisation of the physical world has, until recently, primarily been viewed as an advanced concept with limited short- and mid-term impact (Roland Berger, 2015). Hence, many insurers have previously adopted a cautious attitude to this new technology paradigm. Meanwhile, early-adopters have been disrupting the way insurers traditionally conduct business by demonstrating how vehicle sensors, along with smartphones and wearable technology, can proactively engage policyholders and help improve risk assessment in loss prevention (Koenig et al., 2016). In the insurance industry, as advice-led customer interactions become both more frequent and in real-time, a shift is starting as companies pivot more toward active loss prevention (Reifel et al., 2014). As such, many insurers are moving away from purely 'reactive' business models and incorporating 'preventative' measures into their products in order to help their customers and cut long-term costs. Some of the first companies to adopt this approach were insurers in the health domain, which have introduced preventive actions into their portfolios, such as offering customers incentives to engage in healthy behaviour (CSS Insurance, 2017; Sanitas, 2017). Likewise, in the automotive insurance market, safer driving can be both encouraged with an in-vehicle warning system such as the one described, and measured through vehicle sensors and 'pay-how-you-drive' smartphone applications that assess dangerous driving patterns.

Chapter 5

Spatial Prediction of Traffic Accidents with Heavy Braking Events

The previous chapters have so far presented the motivation for, and background work on traffic accident analysis, and the potential that driving data brings to this long standing field. The prototype system that was developed has also been outlined, and its capability to access high-frequency naturalistic driving data has been discussed, along with the naturalistic field study setup that enabled the collection of vehicle data and the analyses featured in the following chapter. While the results of the previous study have shown that in-vehicle warnings of accident hotspots have a significant improvement on driver behaviour over time, a key limitation is the difficulty in adapting the approach to regions that either do not collect or do not provide historical accident data. As such, this chapter opens by revisiting the motivation of spatially predicting the potential locations of traffic accidents with driving data, and provides more information on the state-of-the-art work that the analyses are structured upon. The materials and methods that are specific to these analyses are subsequently detailed, and the results presented. Finally, a discussion of the results and the conclusions that can be drawn from the work bring the chapter to a close.

5.1 Context and Motivation

The advent of fully-autonomous vehicles brings the promise of drastically reducing the frequency of road traffic accidents, and potentially eliminating them entirely, ushering in a new era of traffic safety (Fagnant and Kockelman, 2015). However, it may take decades to make this vision a reality in even some of the most advanced markets. For example, recent predictions indicate that by the year 2045 it is unlikely that the majority of the light-duty vehicle fleet in the U.S. will be capable of full self-driving automation (Bansal and Kockelman, 2017). As such, connected and semi-autonomous vehicles will become the dominating paradigm in the coming years, and traffic accidents from the manual driving of these vehicles will persist as a key issue (Albright et al., 2016). As these vehicles become increasingly common, a shift will be expected away from the traditional model-based approaches of accident prediction and 'rough proxies' of exposure (Sheehan et al., 2017), to a crowd-sourcing approach using the real-time collection and analysis of data from these vehicles. Modern connected vehicles are enabled by a wealth of technology capabilities, including network connectivity, high precision sensor data from the vehicle's Controller Area Network (CAN) Bus and GPS capabilities, and their semi-autonomous counterparts with additional high definition cameras and LIDAR systems (Mannering, Kilareski, and Washburn, 2007). These technologies can be leveraged to detect accidents and 'near miss incidents', such as heavy braking and evasive manoeuvres, otherwise known as critical driving events (CDEs).

The data from these incidents could be employed to determine existing and emerging locations of high accident probability and enable more proactive business models. For example, in the automotive insurance industry usage-based incentives for taking 'safe routes' could be provided and pay-how-you-drive insurance plans could be optimised (Sheehan et al., 2017). Furthermore, with the knowledge that a potentially dangerous location is ahead, a semi-autonomous vehicle might drive in a more cautious mode to reduce risk or hand over control to the driver to transfer insurance liability. Furthermore, recent research studies, including the previous chapter, of in-vehicle warning systems have shown that drivers themselves can be encouraged to adapt their driving behaviour at potentially dangerous locations (Kazazi, Winkler, and Vollrath, 2015; Tey, Ferreira, and Wallace, 2011; Werneke and Vollrath, 2013). Ultimately, companies with access to such insights can collaborate with road authorities to improve the road infrastructure that contribute to dangerous locations, and manufactures to better understand vehicle capabilities and advance safety-focused offerings (Sheehan et al., 2017).

As such, if players in the automotive industry wish to accurately measure and encourage safer driving and progress further toward more loss prevention business models, then the ability to identify locations on the road network that carry a high risk of accident occurrence, so called 'blackspots', 'sites with promise', or 'hotspots' (Cheng and Washington, 2005), is of utmost value. In this regard, there are two common measures of a roads perilousness that have been extensively researched. The first is Crash Frequency, the actual number of accidents that have occurred on a road section during a specified period of time (Deacon, Zegeer, and Deen, 1974). The second is Crash Rate, a measure of accident exposure risk of vehicles on a road segment, which is typically estimated by normalising the Crash Frequency by the traffic volume, e.g. average daily traffic (ADT) (Hauer and Persaud, 1984). While there is research showing that the situational factors of crashes and near-misses are strongly related, there is limited empirical data on whether the Crash Frequency and Crash Rate of potentially dangerous locations can be reliably identified through analysis of CDEs (Pande et al., 2017).

5.1.1 Research Objectives

Currently, low levels of connected, semi-, and fully-autonomous vehicle adoption make the large-scale collection of CDE data a challenge. However, preliminary analysis is made possible through the installation of the retrofit in-vehicle system previously presented, that collects data analogous to that found in these more modern vehicles. As a result, this chapter of the thesis goes beyond existing research and provides early results from the nationwide naturalistic driving field study of 72 cars covering over 690 000 km in Switzerland. Thus investigating the second research objective of this thesis, determining whether it possible to predict the locations of traffic accident hotspots on the basis of driving data and critical driving events, as discussed in Section 1.2.2. These vehicles were equipped with the system that collected sensor information from each car's CAN Bus in order to access data synonymous with that available in more advanced vehicles. The study presented in this chapter can be split into two parts. In a first step, an ideal setting is analysed for a sub-region of Switzerland, where data for both Crash Frequency and the volume of vehicle traffic, i.e. ADT, were available. This enabled the Crash Rate of these locations to be determined, and the operationalisation and investigation of the second concrete research question of this thesis:

RQ 2a To what extent can the Crash Rate of a location be predicted by CDE information from that location, assuming that Crash Frequency and ADT data are available?

In practice, however, regions where both Crash Frequency and ADT are available are exceedingly rare. While government bodies are now regularly collecting police-recorded Crash Frequency data across whole road networks, accurate measures for traffic frequency remain only partially available and come most often from counting stations. Since traffic frequency is necessary to calculate exposure measures such as Crash Rate, traffic safety analysis are commonly limited to utilising Crash Frequency as a dependent variable for all road segments in this sparse setting. However, to reliably estimate the Crash Frequency of a location on the basis of CDEs, the relationship between traffic frequency of the utilised vehicle fleet and total traffic frequency (of the overall vehicle population) must be known for each location. This dilemma can be addressed by fulfilling a key assumption: The fleet should be 'representative' of the overall population of vehicles.

Therefore, the final stage of the analyses is built upon the assumption that a 'representative' fleet of vehicles is available for data capture. More specifically, the assumption is that there is a well-defined relationship between traffic frequency of the overall population (TF_{Pit}) and traffic frequency of the sample, i.e. field study fleet (TF_{Sit}), which is independent of location i for a given timeframe t of the analysis. From this the relationship between fleet-generated CDEs and the Crash Frequency variable can be tested, and the final research question operationalised as:

RQ 2b To what extent can the Crash Frequency of a location be predicted by CDE information from that location, assuming we have sparse data coverage (only Crash Frequency information is available) and a well-defined relationship between TF_{Sit} and TF_{Pit}, i.e. a 'representative' fleet?

To address this question, the partially available TF_{Pit} data is first leveraged to test to what extent the field study fleet utilised for the analysis fulfils the 'representative' fleet assumption. The analysis then concludes by estimating Crash Frequency on the basis of CDEs for the majority of Switzerland, where TF_{Pit} was unavailable, utilising spatial regression.

5.1.2 Research Implications and Structure

The remainder of this chapter is structured as follows. In the next section, the state-of-the-art research is outlined and the data collection process and the field study setting revisited. Next, the results of the study are presented with regard to the relationship between Crash Rate and CDEs. This is followed by an exploration of fleet data from the field study to explain Crash Frequency, and finally demonstrate the relationship between Crash Frequency and CDEs through nationwide spatial regression, for situations in practice where traffic frequency data is unavailable. The final sections discuss the limitations of the research and how they could be addressed in future work, and the chapter closes with the general conclusions.

Overall, this research has implications for companies, organisations, and other players in the automotive industry both today and in the near future. The presented approach empowers those with access to data from connected, semi-, and fully-autonomous vehicles to rapidly identify areas of high accident exposure and to improve their management of customer risk. Furthermore, the importance of such a fleet satisfying the 'representative' fleet assumption is demonstrated, and how this enables analysis where traffic frequency measurements are unavailable. Moreover, this information may have potential not only for industry practitioners, but also for policy makers in this long-standing field, where understanding the new capabilities and the reliance of findings from recent automotive technology advances will be vital for determining suitable traffic safety approaches and strategies.

5.2 State of the Art

5.2.1 The Naturalistic Driving Approach

In the last years, several researchers have built upon the idea that driving behaviour displayed in avoiding a crash is similar to that of a crash event itself, and hence, if detected, can be considered in the identification of crash potential. In prior work, a naturalistic driving study of 100 cars equipped with cameras, radar, and OBD-II sensors was conducted, and analyses established that accident 'near-misses' can act as a suitable safety metric surrogate for rarely occurring crashes (Dingus et al., 2006). This foundational study was the first to equip predominantly privately owned vehicles with unobtrusive recording devices in order to collect naturalistic driving data at a large-scale. The drivers taking part in the study were given no specific driving instructions, and unlike previously discussed controlled field studies, no researcher was present in the vehicle with the driver. The resulting dataset reportedly contains primarily examples of typical driving behaviour, along with a select number of instances where the participants demonstrate drowsiness, judgment error, risk taking and aggressive driving behaviour, and traffic violations, along with many other actions that would be considered unsafe.

From these extreme cases, a 'near-miss' accident has been described operationally as "any circumstance that requires a rapid, evasive manoeuvre by the participant vehicle, or any other vehicle, pedestrian, cyclist, or animal, to avoid a crash" (Dingus et al., 2006). Furthermore, a rapid, evasive manoeuvre is further defined as "steering, braking, accelerating, or any combination of control inputs that approaches the limits of the vehicle capabilities" (Guo et al., 2010a; Guo et al., 2010b). Overall, the identified 'near-misses' that occurred in the study were determined to have a very strong correlation with actual crash events. It was found that the reaction to events displayed by drivers was typically much lower for crashes than for near-crashes, suggesting that often the level of awareness of the driver is the deciding factor between a collision and a near-miss. As such, the study and the related analyses demonstrate how naturalistic driving data offers both road safety researchers and practitioners a source of variables for analysis that are not limited to rarely occurring traffic accidents. Whereas the participants in the so called '100-Car Naturalistic Driving Study' were recorded with cameras in order to operationalise dangerous driving behaviour and establish the connection between traffic accidents and near-miss incidents, the behaviour itself can be measured using the advanced sensors embedded in connected, semi-, and fully-autonomous vehicles.

5.2.2 Driving Data and Traffic Accidents

The technology in these modern vehicles can be potentially be leveraged to detect accidents and critical driving events, such as heavy braking and evasive manoeuvres, and used to predict locations with a high likelihood of traffic accidents. Both researchers and industry players alike see the promise of this data to combat the existing challenges of accident analyses, which include issues surrounding small sample sizes, underreporting of traffic accidents, and data scarcity (Mannering, Kilareski, and Washburn, 2007). For example, the behaviour of different driver groups in the US State of Georgia was studied by collecting data from vehicle OBD-II systems, as well as GPS location data, and it was noted that clusters of hard deceleration events were co-located with clusters of historical accident data (Jun, Ogle, and Guensler, 2007). Furthermore, in very recent work, a small naturalistic driving study of 33 university staff members, with an average of 10 days of data collected per participant, modelled accident frequency on US Highway 101 using only data gathered from a GPS logger (Pande et al., 2017). In the study, a total of 39 segments were analysed in a negative binomial model by considering variables such as the percentage of high jerk (the rate of change of acceleration) events, the slope of the road segment, and the average daily traffic. As a result, the authors found a promising relationship between the collected jerk events and the frequency of traffic accidents, and call for additional research on a larger scale in order to validate and generalise their findings.

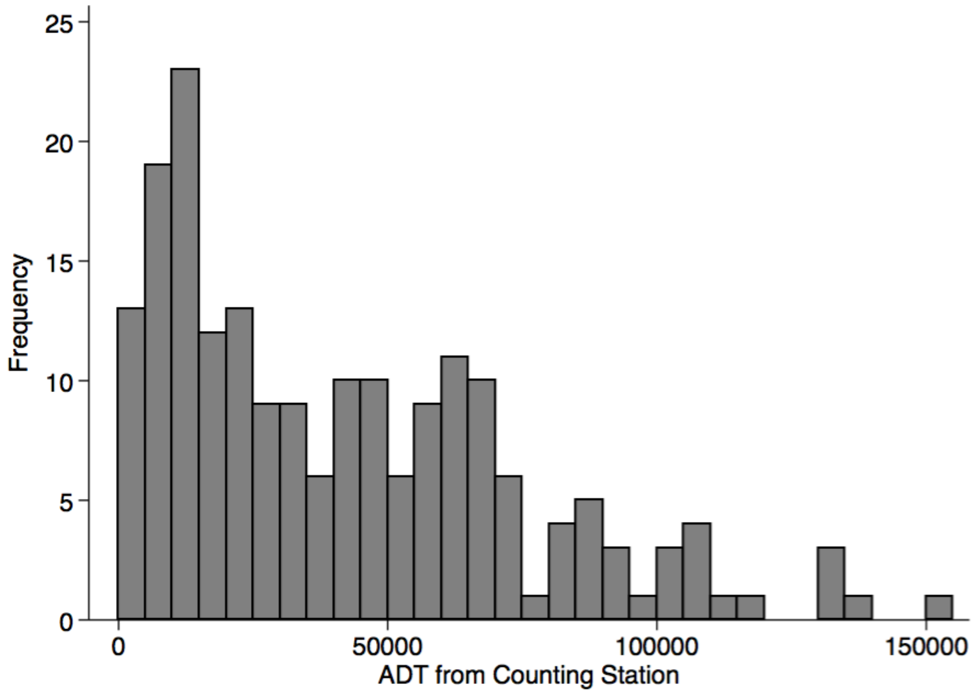
In summary, by extending existing work to detect high jerk events with a large enough fleet of vehicles and validating this approach with existing traffic accident analysis techniques, it becomes theoretically possible to identify crash potential and dangerous locations before an accident occurs. While traditional vehicles require retrofitting to deliver such insights, the growing number and popularity of increasingly connected vehicles offer a unique opportunity to detect the locations of real-time CDEs. These events can act as a predictor for rarely occurring traffic accidents, and hence address issues of data scarcity and time delay (Guo et al., 2010a; Guo et al., 2010b; Klauer et al., 2006). A prominent example of how this could be applied in practice addresses traffic queues forming on highways, where many lives are lost each year from drivers approaching these unforeseen queues too quickly (Li, Chung, and Cassidy, 2013). By detecting such hazardous situations early enough and in real-time, loss prevention services can be provided, such as triggering in-vehicle warnings to encourage drivers to approach with caution, and vehicles with autonomous driving features can take risk reducing measures, such as adapting their speed gradually rather than abruptly (Herrmann, Brenner, and Stadler, 2018).

5.3 Materials and Methods

5.3.1 FEDRO Average Daily Traffic Dataset

Traditionally, the amount of traffic flow, i.e. the number of cars travelling on a particular stretch of road, has long been associated with traffic accidents (Hauer and Persaud, 1984). The argument for this is that sections of road with higher traffic flow will see a higher number of traffic accidents per year, assuming there is a fixed probability of a traffic accident occurring. Therefore, in order to assess a specific location's Crash Rate or risk exposure based on the number of traffic accidents, it is important to account for traffic frequency. In order to incorporate this into the analysis, traffic data was obtained from the Swiss Federal Roads Office (FEDRO) Statistics For Road Accidents. This dataset is comprised of the average number of vehicles per day passing a variety of counting stations across the Swiss road network, as outlined in Section 3.1. However, as is the case in many countries, in Switzerland there is only partial location coverage of ADT counting stations. This data was filtered to cover the same 18 week time period as the full field study, as presented in Section 3.3. As such, the final set of observations were constructed from locations where measurements were available for this period, resulting in 194 counting stations and ADT measurements. The distribution of the ADT counts at these locations is shown in Figure 5.1.

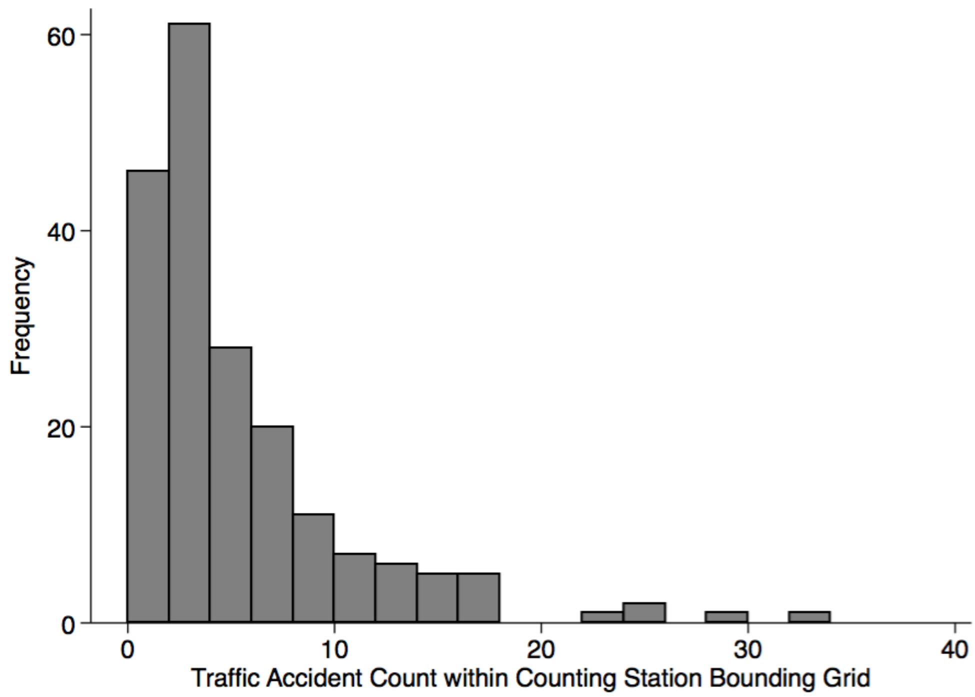
FIGURE 5.1: Histogram showing the distribution of ADT from counting stations during the 18-week field study, with a bin size of 5,000



5.3.2 FEDRO Traffic Accident Dataset

Traffic accident data, as initially presented in Section 3.1, was obtained from FEDRO in order to generate the dependent variables for the analyses. This dataset contained GPS locations, as well as contextual information on the causes, on over 266 000 traffic accidents which occurred in Switzerland between 2011 and 2017. Since accident locations can change over time due to improved road infrastructure, time-relevant Crash Rate and Crash Frequency dependent variables were generated by first sampling only the traffic accidents which occurred during the 18 week field study from the overall dataset, a total of 25 493 accidents across the country. Secondly, a naive grid-count approach was applied to count the traffic accidents which occurred within a 1 km^2 grid of each of the 194 traffic counting stations from Section 3.1, where the centre of the grid was each traffic frequency counting station, similar to a previous traffic accident study (Yang and Kim, 2003). Over the 18 week period, this came to a total of 972 traffic accidents, with an arithmetic mean of 5.01 and a geometric mean of 3.29 accidents per counting station location. Figure 5.2 shows the distribution of traffic accidents within the bounding grid of the counting stations.

FIGURE 5.2: Histogram showing the distribution of traffic accidents within counting station bounding grids during the 18-week field study, with a bin size of 2



5.3.3 Field Study Dataset

Following a recent study, high-jerk events were considered for classifying near-miss events and CDEs (Pande et al., 2017). For the purpose of this analysis the jerk event behaviour was calculated from the vehicle's longitudinal acceleration sensor, which was sampled at 10 Hz from the CAN Bus of each of the field study vehicles. Post-processing of the acceleration values enabled the generation of jerk events for each participant of the field study. These values were calculated for each measurement time step based on the following equation:

$$j = da/dt \tag{5.1}$$

where da is the change in acceleration [m/s^2], and dt is the change in time [s]. The thresholds suggested for identifying near-miss jerk events significantly vary between research projects (Aichinger et al., 2016). These thresholds range from very strong events between $-9.9\,\text{m/s}^3$ to $-12.6\,\text{m/s}^3$ (Bagdadi and Várhelyi, 2011), which occur very infrequently in the fleet of professional drivers, with less than 600 events in the 18 week period, to very low events between $-0.15\,\text{m/s}^3$ to $-0.61\,\text{m/s}^3$

(Pande et al., 2017), which were triggered for almost every braking event that occurred in the study. As such, a threshold value of $-2\,\mathrm{m/s^3}$ was chosen to classify an event as high jerk. These jerk events were then limited to situations where the vehicles were decelerating, resulting in a dataset of roughly 912 000 geo-located jerk events over the course of the field study. Therefore, deceleration jerk events were calculated using the pseudocode shown in Algorithm 2, where successive jerk events were counted as one single event.

Algorithm 2 Pseudocode for deceleration jerk event generation

```
DecelerationJerkEventGeneration(A, J)
    set successiveJerkEvent = false
    set jerkEventList = []
    for each value A′ in acceleration values list A:
        if (A′ < 0 m/s²):
            // vehicle is decelerating
            J′ = jerk values list J[position of A′ in A]
            if (J′ < −2 m/s³):
                // vehicle jerk rate is under threshold
                if (successiveJerkEvent == false):
                    add J′ to jerkEventList
                    set successiveJerkEvent = true
            else if (J′ > 0 m/s³):
                set successiveJerkEvent = false
        else
            // vehicle is accelerating
            set successiveJerkEvent = false
    return jerkEventList
```

5.3.4 Variable Definitions

On the basis of the average daily traffic and traffic accident datasets previously outlined, the following definitions are put forward:

- Let CF_{Pit} (Crash Frequency) be the number of accidents within location i from population P in timeframe t
- Let TF_{Pit} (Traffic Frequency) be the number of transits of location i of population P in timeframe t – in our case the average ADT over the period of the field study multiplied by the length of the study

A standardised measure of roadway safety, and long-standing alternative to analysing Crash Frequency, is to consider the exposure risk of locations, such as Crash Rate. In research and practice, Crash Rate is often reported and analysed as number of accidents per 1 million or even 100 million vehicle miles travelled. There are two fundamental reasons for the context-specific scaling of Crash Rates. Firstly, very small decimal numbers, as well as very high numbers, are hard to communicate. Secondly, the scaling of Crash Rate enables the application of well-proven techniques, such as established count regressions. Based on the scaling of a previous study (Anastasopoulos et al., 2012), $\mathrm{CR}_{\mathrm{Pit}}$ is defined as:

$$CR_{\mathrm{Pit}} = CF_{\mathrm{Pit}} / TF_{\mathrm{Pit}} \cdot 10^8 \tag{5.2}$$

The third dataset is the connected vehicle fleet dataset from the field study. In accordance to the definitions above, the following definitions are put forward on the basis of fleet data:

- Let $\mathrm{JF}_{\mathrm{Sit}}$ (Jerk Frequency) be the number of high jerk events within location i from sample (fleet) S in timeframe t
- Let $\mathrm{TF}_{\mathrm{Sit}}$ (Trip Frequency) be the number of transits of location i from sample (fleet) S in timeframe t – in our case the number of trips through that location during the period of the field study

Finally, Jerk Rate, $\mathrm{JR}_{\mathrm{Sit}}$, is defined as the following:

$$JR_{\mathrm{Sit}} = JF_{\mathrm{Sit}} / TF_{\mathrm{Sit}} \tag{5.3}$$

5.4 Results

The results of four sequential sets of analysis are presented in the following section, with the underlying theme of investigating the link between CDEs and traffic accidents. The analyses of Sections 5.4.1, 5.4.2, and 5.4.3 primarily consider locations where ADT measurements from FEDRO are available. In Section 5.4.4 the approach is extended to the majority of the Swiss road network, where the population traffic frequency measurements are unavailable. As such, this section proceeds by addressing the following problems:

- Section 5.4.1 examines an optimal scenario for determining accident exposure on the basis of driving data, in that we consider the subset of locations with available ADT measurements. This enables an investigation into the relationship between the Crash Rate of the population and the Jerk Rate of vehicles at these sites, and further motivates the subsequent sections.

- Section 5.4.2 follows by investigating the assumption that the Trip Frequency of the field study fleet is representative of the overall population traffic frequency, i.e. ADT, through these locations. By satisfying this, the more practical situation where traffic frequency data is unavailable can start to be modelled.
- Section 5.4.3 then continues by utilising the assumption from Section 5.4.2 and testing of the relationship between the Crash Frequency of the population, and the Jerk Rate and Trip Frequency of the field study fleet at these locations.
- Section 5.4.4 concludes the analyses by demonstrating the same relationship from Section 5.4.3 through nationwide spatial regressions covering the majority of the Swiss road network. The robustness of this model is additionally explored by iteratively including explanatory variables well-known to be associated with impacting the likelihood of traffic accidents occurring.

5.4.1 Crash Rate and Jerk Rate

In a first step, the analysis focuses on an ideal setting, where Crash Frequency as well as traffic frequency are available for all locations. More specifically, locations were defined using a naive grid-count approach, i.e. traffic accidents and traffic volumes were determined within a 1 km^2 grid of each of the 194 traffic counting stations. Since the dependent variable in the initial analysis is CR_{Pit}, i.e. the expected number of traffic accidents of the grid per 100 million transits, a count variable is effectively being analysed as a measure of exposure. There are indications of over-dispersion of the variable, as there are large differences between the mean (183.3) and the variance (123 340.2). In addition, when considering the JR_{Sit} independent variable, it is also observed to follow a Poisson distribution. As such, this is transformed via the commonly used natural log function to normalise the variable (Quddus, 2008), and negative binomial regression utilised, thus assuming that:

$$CR_{\text{Pit}} = e^{a1} \cdot JR_{\text{Sit}}{}^{b1} \tag{5.4}$$

Regressions were run for the whole set of counting stations ($\text{TF}_{\text{Sit}} > 0$), and additional models generated where counting stations were excluded based on a minimum value of TF_{Sit}. As shown in Table 5.1, the correlation between Crash Rate and

TABLE 5.1: Negative binomial regression results for the Crash Rate (CR_{Pit}) dependent variable and the independent variable Jerk Rate (JR_{Sit}), sampled by number of trips through counting station bounding grids (TF_{Sit})

	Dependent Variable: CR_{Pit}					
	Model 1.1 $TF_{Sit} > 0$	**Model 1.2** $TF_{Sit} > 5$	**Model 1.3** $TF_{Sit} > 10$	**Model 1.4** $TF_{Sit} > 15$	**Model 1.5** $TF_{Sit} > 20$	**Model 1.6** $TF_{Sit} > 25$
Independent Variables						
ln (JR_{Sit})	0.293***	0.338***	0.354***	0.362***	0.385***	0.383***
	(3.86)	(4.43)	(4.56)	(4.40)	(4.74)	(4.66)
Constant	20.03***	20.09***	20.12***	20.14***	20.16***	20.16***
	(273.63)	(266.12)	(260.01)	(251.74)	(252.18)	(252.45)
ln alpha						
Constant	-0.326***	-0.352***	-0.377***	-0.375***	-0.393***	-0.396***
	(-4.22)	(-4.30)	(-4.39)	(-4.29)	(-4.43)	(-4.44)
Observations	**194**	**172**	**159**	**152**	**149**	**147**
McFadden's Pseudo R^2	**0.079**	**0.108**	**0.121**	**0.120**	**0.137**	**0.133**
χ^2	**14.90**	**19.66**	**20.82**	**19.36**	**22.48**	**21.71**

t statistics in parentheses
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

Jerk Rate is highly significant in all models and has a similar coefficient, ranging between 0.293 to 0.383. In addition, the over-dispersion constant, i.e. $ln\ alpha$, is significant for all models, confirming the suitability of negative binomial regression. The best fitting model is where counting stations were excluded with 20 or less trips (TF_{Sit}) through the surrounding bounding grid, with a pseudo R^2 of 0.137[1]. This indicates that there is a minimum number of times an area should be driven through before an assessment should be conducted on the likelihood of traffic accidents occurring based on driving data. However, the model fit drops when limiting with higher number of trips, most likely due to decreased observations in the model.

In order to answer the first concrete research question of the chapter, **RQ 2a**: To what extent can the Crash Rate of a location be predicted by CDE information from that location, assuming that Crash Frequency and ADT data are available, the null hypothesis that CR_{Pit} of a location is not explained by $ln(JR_{Sit})$ is tested. Through testing the χ^2 values for each of the generated models this null hypothesis is rejected, with $p < 0.001$ for all models. Since negative binomial regression utilises the log-link, and JR_{Sit} is natural log transformed, the interpretation of the best-fitting model

1

allows us to conclude that with a 10 % increase in the Jerk Rate there would be a 3.7 % ($1.10^{0.385} = 1.037$) increase in the Crash Rate within the surrounding area.

However, for practitioners and researchers alike, while CF_{Pit} can often be obtained, such ideal data coverage for TF_{Pit}, e.g. ADT, is rarely available to generate CR_{Pit}. It is therefore important to consider more common situations in traffic analysis where only the Crash Frequency can be leveraged as a dependent variable for the whole road network. Therefore, while Crash Rates have been used in the first part of this analysis, the remainder of the analyses focus on Crash Frequency, additionally motivated by previous work investigating traffic accident exposure (Hauer, 1995).

5.4.2 Representative Fleet Assumption

The second stage of the analysis is built upon the following assumption to determine Crash Frequency: That there has to be a well-defined relationship between traffic frequency of the population (TF_{Pit}) and traffic frequency of the fleet sample (TF_{Sit}), which is independent of location i for a specific timeframe t, i.e. a 'representative' fleet. A roadside assistance fleet, as employed in the field study, could be considered representative, since they are called to locations where typical drivers have broken down or require assistance. Thus, in order to test the second research question of this chapter, regarding the relationship between fleet-generated events and the Crash Frequency variable, the representative assumption on the fleet dataset at hand should first be assessed. This assumption can be tested by considering the relationship between TF_{Pit} and TF_{Sit}. Since TF_{Pit} for the set of stations is generated by multiplying the ADT of the location by the length of the field study, a count variable is being considered which follows a Poisson distribution. Additionally, by reviewing the mean (41 964) and variance (1.10×10^9) of TF_{Pit}, it can be seen that the variance is much larger, indicating that the variable is over-dispersed and that negative binomial regression is a suitable technique for analysis.

When considering TF_{Sit}, the traffic frequency of the sample of fleet drivers, a Poisson distribution of the data is also observed, and thus the commonly used normalisation technique of natural logarithm transformation is applied to the variable (Quddus, 2008). As in the previous section, there is the additional assumption that there is a trade-off with the sample, where there is a minimum number of times an area should be driven through for a reliable assessment on the basis of driving data. As such, by removing grids with less trips from the fleet of drivers (TF_{Sit}) the fit of the model should increase, until a certain point where too many data points have been removed. The results of these regressions are shown in Table 5.2.

TABLE 5.2: Negative binomial regression results for the Population Traffic Frequency ($\mathrm{TF_{Pit}}$) dependent variable and the independent variable Fleet Traffic Frequency ($\mathrm{TF_{Sit}}$) sampled by number of trips through counting station bounding grid ($\mathrm{TF_{Sit}}$)

	Dependent Variable: $\mathrm{TF_{Pit}}$					
	Model 2.1 $\mathrm{TF_{Sit}} > 0$	**Model 2.2** $\mathrm{TF_{Sit}} > 5$	**Model 2.3** $\mathrm{TF_{Sit}} > 10$	**Model 2.4** $\mathrm{TF_{Sit}} > 15$	**Model 2.5** $\mathrm{TF_{Sit}} > 20$	**Model 2.6** $\mathrm{TF_{Sit}} > 25$
Independent Variables						
ln ($\mathrm{TF_{Sit}}$)	0.203***	0.269***	0.342***	0.425***	0.424***	0.428***
	(5.78)	(5.30)	(7.19)	(10.14)	(9.28)	(9.04)
Constant	14.49***	14.13***	13.72***	13.24***	13.24***	13.22***
	(74.18)	(49.44)	(51.08)	(56.77)	(51.51)	(49.50)
ln alpha						
Constant	-0.482***	-0.642***	-0.740***	-0.808***	-0.792***	-0.792***
	(-4.97)	(-5.30)	(-6.25)	(-6.46)	(-6.33)	(-6.25)
Observations	**194**	**172**	**159**	**152**	**149**	**147**
McFadden's Pseudo R^2	**0.204**	**0.241**	**0.279**	**0.327**	**0.309**	**0.302**
χ^2	**33.43**	**28.12**	**51.77**	**102.8**	**86.11**	**81.72**

t statistics in parentheses
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

The results show that the natural logarithm of $\mathrm{TF_{Sit}}$ has a highly significant coefficient with $\mathrm{TF_{Pit}}$. Through testing the null hypothesis that $\mathrm{TF_{Pit}}$ of a location is not explained by $ln(TF_{\mathrm{Sit}})$ of the field study fleet, it can be determined whether the overall population traffic volume can be explained by fleet traffic volume on the basis of the assumed relationship:

$$TF_{\mathrm{Pit}} = e^{a2} \cdot {TF_{\mathrm{Sit}}}^{b2} \tag{5.5}$$

Through testing the χ^2 values for each of the generated models, this null hypothesis is rejected, with $p < 0.001$ for all models. The best fitting model was where counting stations were excluded with 15 or less trips through the surrounding bounding grid and with a pseudo R^2 of 0.327. Finally, Figure 5.3 shows the distribution of the natural logarithm transformed $\mathrm{TF_{Sit}}$ independent variable, and Figure 5.4 the relationship between $\mathrm{TF_{Pit}}$ and $\mathrm{TF_{Sit}}$. Both figures feature the full dataset, and a vertical cut-off demonstrating the data used in the negative binomial regression model with the best fit, i.e. only including grids with greater than 15 trips.

FIGURE 5.3: Histogram showing log-normal distribution of field study fleet trips through counting station bounding grids, with a bin size of 0.4. Vertical line indicates 15 trips cut-off

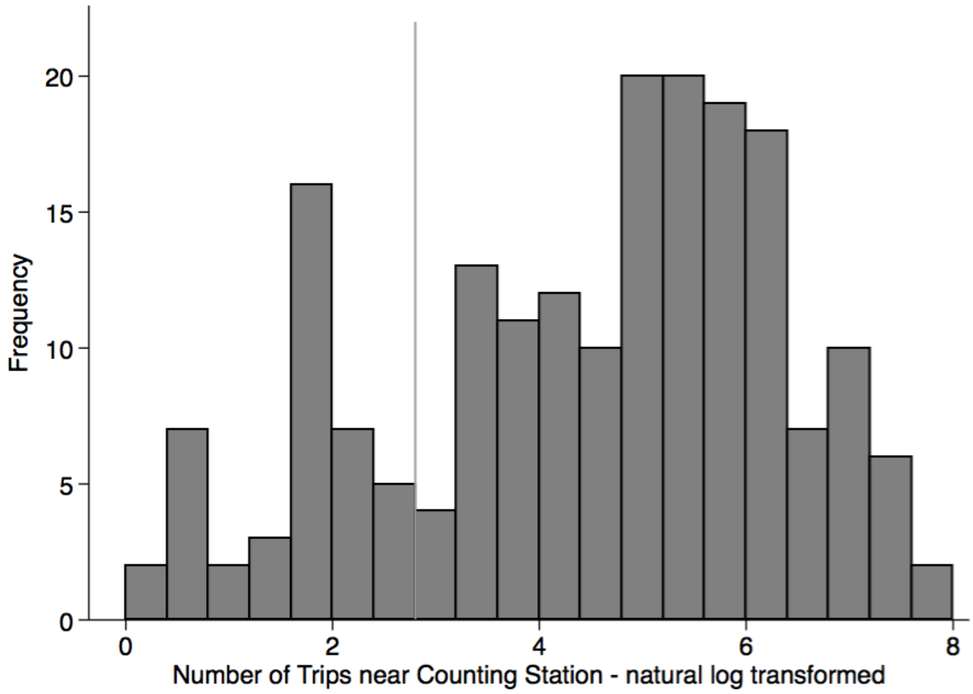
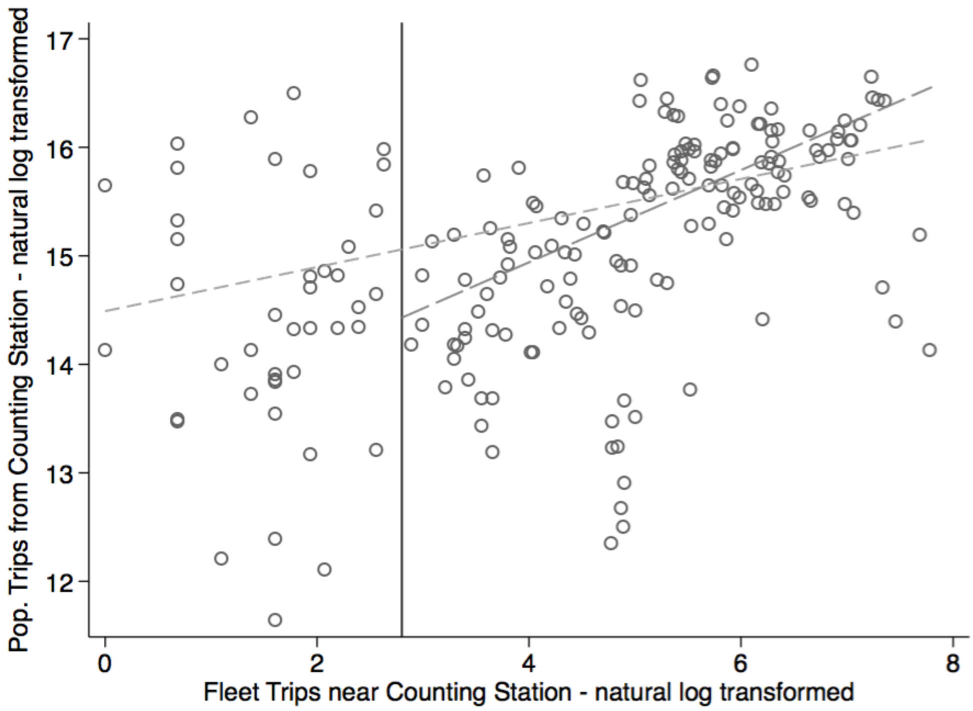


FIGURE 5.4: Scatter plot showing relationship between TF_{Pit} and TF_{Sit}. Dashed line is Model (2.1). Long dashed line is Model (2.4). Vertical line indicates 15 trips cut-off



5.4.3 Crash Frequency and Jerk Frequency

Building upon the evidence that we have a well-defined relationship between $\mathrm{TF}_{\mathrm{Sit}}$ and $\mathrm{TF}_{\mathrm{Pit}}$, the Crash Frequency of the locations can be estimated based on the Jerk Rate and Trip Frequency of the sample, substituting $\mathrm{TF}_{\mathrm{Pit}}$ with $\mathrm{TF}_{\mathrm{Sit}}$ and thereby eliminating the need for ADT information. More specifically, based on the previous definitions and assumptions:

$$\text{(from Equation 5.2)} \qquad CR_{\mathrm{Pit}} = CF_{\mathrm{Pit}}/TF_{\mathrm{Pit}} \cdot 10^8$$

$$\text{(from Equation 5.4)} \qquad CR_{\mathrm{Pit}} = e^{a1} \cdot JR_{\mathrm{Sit}}{}^{b1}$$

$$\text{(from Equation 5.5)} \qquad TF_{\mathrm{Pit}} = e^{a2} \cdot TF_{\mathrm{Sit}}{}^{b2}$$

it can be concluded that:

$$CF_{\mathrm{Pit}}/TF_{\mathrm{Pit}} \cdot 10^8 = e^{a1} \cdot JR_{\mathrm{Sit}}{}^{b1} \tag{5.6}$$

$$CF_{\mathrm{Pit}} = e^{a1-ln(10^8)} \cdot JR_{\mathrm{Sit}}{}^{b1} \cdot TF_{\mathrm{Pit}} \tag{5.7}$$

$$CF_{\mathrm{Pit}} = e^{a1-ln(10^8)} \cdot JR_{\mathrm{Sit}}{}^{b1} \cdot e^{a2} \cdot TF_{\mathrm{Sit}}{}^{b2} \tag{5.8}$$

$$CF_{\mathrm{Pit}} = e^{a1+a2-ln(10^8)} \cdot JR_{\mathrm{Sit}}{}^{b1} \cdot TF_{\mathrm{Sit}}{}^{b2} \tag{5.9}$$

Thus, through a Poisson or negative binomial regression with Crash Frequency ($\mathrm{CF}_{\mathrm{Pit}}$) as the dependent variable, one can derive the coefficients $b1$ as well as $b2$. If $a2$ is available from a previous analysis determining the relationship between $\mathrm{TF}_{\mathrm{Pit}}$ with $\mathrm{TF}_{\mathrm{Sit}}$, i.e. through the analysis in Section 5.4.2, $a1$ can also be determined. With estimates for $a1$ and $b1$, $\mathrm{CR}_{\mathrm{Pit}}$ can ultimately also be predicted on the basis of:

$$\text{(from Equation 5.2)} \qquad CR_{\mathrm{Pit}} = e^{a1} \cdot JR_{\mathrm{Sit}}{}^{b1}$$

Considering the $\mathrm{CF}_{\mathrm{Pit}}$ dependent variable, a Poisson distribution continues to be observed, where the variance (28.20) is higher than the mean (5.01), indicating overdispersion of the variable and thus negative binomial regression should be applied. Results of the regressions, sampled by number of trips ($\mathrm{TF}_{\mathrm{Sit}}$) are shown in Table 5.3. When compared to the $\mathrm{CR}_{\mathrm{Pit}}$ set of models, improved model fit is observed, with pseudo R^2 ranging from 0.281 to 0.336. In addition, the same pattern seen in the previous regressions of Section 5.4.2 can be seen, where model fit improved when grids with less than 15 trips were excluded from the analysis. This reaffirms that

TABLE 5.3: Negative binomial regression results for the Crash Frequency (CF_{Pit}) dependent variable and the independent variables Jerk Rate (JR_{Sit}) and Fleet Trip Frequency (TF_{Sit}), sampled by number of trips through counting station bounding grid (TF_{Sit})

	Dependent Variable: CF_{Pit}					
	Model 3.1 $TF_{Sit} > 0$	**Model 3.2** $TF_{Sit} > 5$	**Model 3.3** $TF_{Sit} > 10$	**Model 3.4** $TF_{Sit} > 15$	**Model 3.5** $TF_{Sit} > 20$	**Model 3.6** $TF_{Sit} > 25$
Independent Variables						
ln (JR_{Sit})	0.330***	0.367***	0.405***	0.371**	0.380***	0.384***
	(4.88)	(5.18)	(5.31)	(4.92)	(5.02)	(4.93)
ln (TF_{Sit})	0.232***	0.259***	0.314***	0.378***	0.367***	0.363***
	(6.43)	(5.20)	(6.15)	(7.95)	(7.54)	(7.30)
Constant	0.479**	0.333	0.0171	-0.366	-0.302	-0.276
	(2.61)	(1.25)	(0.06)	(-1.49)	(-1.19)	(-1.06)
ln alpha						
Constant	-0.929***	-0.905***	-0.960***	-1.020***	-1.012***	-1.004***
	(-6.38)	(-5.83)	(-5.84)	(-5.82)	(-5.74)	(-5.68)
Observations	**194**	**172**	**159**	**152**	**149**	**147**
McFadden's Pseudo R^2	**0.281**	**0.275**	**0.303**	**0.336**	**0.331**	**0.322**
χ^2	**56.77**	**49.49**	**62.09**	**73.27**	**69.24**	**64.32**

t statistics in parentheses
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

there should be a minimum number of fleet transits of an area before analyses are conducted on the basis of driving data.

With the outlined results, the second concrete research question of this chapter can now be addressed, **RQ 2b**: To what extent can the Crash Frequency of a location be explained by CDE information from that location, assuming we have sparse data coverage (only Crash Frequency information is available) and a well-defined relationship between TF_{Sit} and TF_{Pit}, i.e. a 'representative' fleet. This is achieved by separately testing two null hypotheses, i.e. that CF_{Pit} of a location is not explained by either $ln(JR_{Sit})$ or by $ln(TF_{Sit})$ of our field study fleet. Through independently testing the χ^2 values for each variable of the generated models, both null hypotheses are rejected, with $p < 0.001$ for both variables and all models. Regarding the $b1$ coefficient, which gives the proportional increase in both Crash Rate, and Crash Frequency in the models, it can be observed that in the CF_{Pit} regressions it falls within a similar range (0.330 to 0.405) to the coefficient in the CR_{Pit} set of models (0.293 to 0.383). In addition, the $b2$ coefficient can be seen to also fall in a similar range (0.232 to 0.378) to the coefficient in the TF_{Pit} models (0.203 to 0.428). From this it can be concluded that, with a fleet which can be demonstrated to be representative of the population, one can estimate Crash Frequency, and from the coefficient make an approximation for the Crash Rate exposure measure.

5.4.4 Crash Frequency Spatial Regression

In practice, such an analysis would not be run on a small subset of grids, but over the whole road network. In order to provide insights into a nationwide approach, this problem is now considered for the whole field study dataset, covering the majority of the Swiss road network. Following the same approach as the first part of the analysis, the country was divided into regular square grid cells of $1\,\mathrm{km}^2$ in order to perform the analysis (Kim, Brunner, and Yamashita, 2006). Crash Frequency was generated within each grid, along with the Jerk Frequency, and Traffic Frequency of the fleet, allowing the generation of new values for the variables CF_{Pit}, JR_{Sit}, and TF_{Sit}. Based on the results of the previous sections, grids with a value for TF_{Sit} less than 15 were excluded from the analysis, resulting in a total of $4197\,\mathrm{km}^2$ of the country's road network being considered. Of the 25 493 traffic accidents which occurred during the 18 week period, 15 450 were included in the grids covered by the field study and that fulfilled the minimum number of transits by the utilised fleet.

Urban areas, as well as highways, often show very specific accident patterns (Kim, Brunner, and Yamashita, 2006; Pande et al., 2017). These variables can be obtained for locations through map-matching services and open data sources. Therefore, in order to include these as explanatory variables in the regression, each grid was further enriched with two binary variables, one for 'Urban' or 'Rural' and another for 'Highway' or 'Non-highway' through a professional map-matching service provider. The dataset was split roughly in half between urban and rural areas, as shown in Figure 5.5, and Figure 5.6 shows the distribution of highway and non-highway roads. As mentioned above, significant differences are expected between the areas and road-types due to the difference in both population and traffic frequency.

Spatial autocorrelation is where measurements are dependent on their location and surroundings, i.e. not independent and identically distributed. Thus, it can be defined as a property found across geographic space, where variables are either more similar or less similar at certain distances from each other than would be expected for randomly associated pairs of observations (Legendre, 1993). This spatial autocorrelation can be tested with various indicators, where the most commonly used measure is the Moran's I statistic (Anselin and Rey, 2014; Bivand et al., 2008). Moran's

FIGURE 5.5: Distribution of Urban (red) and Rural (yellow) locations in Switzerland where naturalistic field study driving data was collected

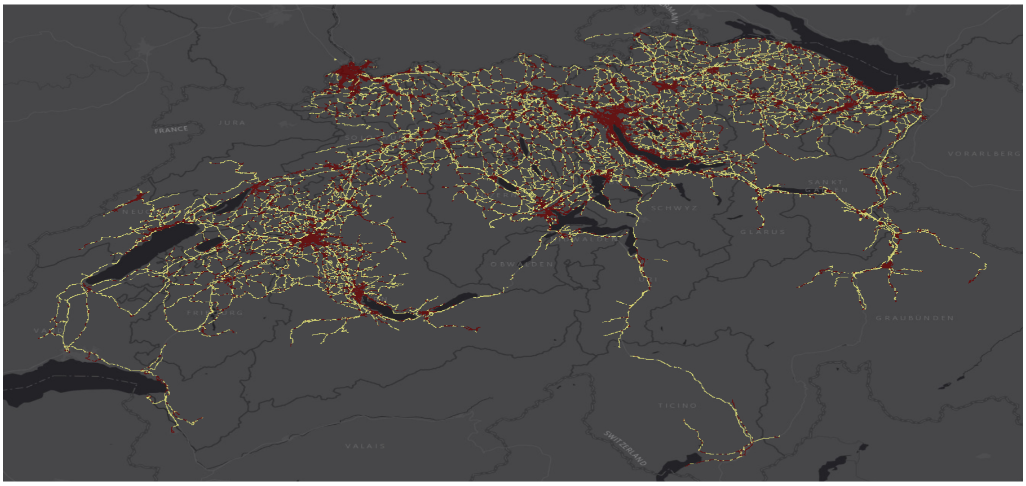
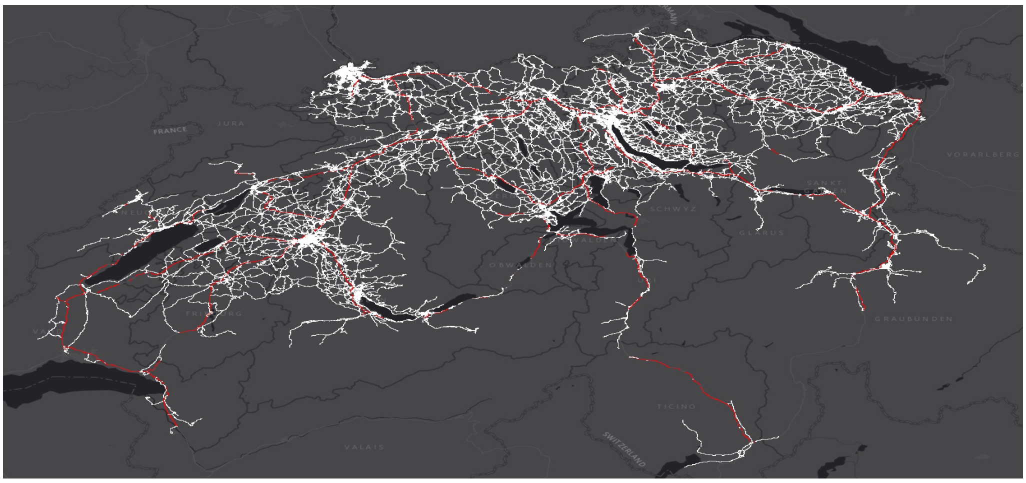


FIGURE 5.6: Distribution of Highway (red) and Other (white) locations in Switzerland where naturalistic field study driving data was collected



I can be interpreted as a regression coefficient for autocorrelation, where the value falls between -1 and 1. A value of 1 indicates a strong positive spatial autocorrelation, -1 indicates a strong negative spatial autocorrelation and 0 means that observations are random, and thus not spatially auto-correlated (Anselin, 1993). For the dataset and a grid size of $1\,\mathrm{km}^2$, the Moran's I value is 0.426 ($p = 0.000$), hence a statistically significant spatial autocorrelation is found which should be accounted for in the model.

Another important concept for statistical tests is homo- and heteroscedasticity. Homoscedasticity is the assumption of constant error variance (Anselin and Rey, 2014). That is, the variances around the regression line are equal for all values of the independent variable. If the distribution of values is heteroskedastic, the variance of the independent variable cannot be assumed to be constant. Due to the spatial distribution of accidents one can expect heteroscedasticity of accidents and vehicle data along different road sections. To address the problem of heteroscedasticity, spatial regression models can include local and global spatial relationships in order to mitigate or reduce the effect. To utilise these spatial effects, spatial weights are used (Anselin and Rey, 2014; Bivand et al., 2008). For the grid-based dataset, contiguity weights are the most logical choice, where only bordering cells are expected to have a direct effect on the dependent variable. Other techniques, such as threshold and inverse distance weights, are impractical as they use different distances for diagonal than horizontally or vertically adjacent cells due to the calculation using the centre point of the square. Contiguity weights also allow higher order weights for cells that are further away than the cell directly adjacent.

Lagrange Multiplier tests showed that the dataset is susceptible to both spatial lag and spatial error, where tests are highly significant for both lag and error models for the general and the robust specification. Therefore, the spatial combo model is used as it includes both spatial error and spatial lag. The dependent variable, $\mathrm{CF}_{\mathrm{Pit}}$, is transformed with the natural logarithm, to replicate the log-link relationship from the earlier negative binomial regressions. For this analysis four models, presented in Table 5.4, are iteratively developed in order to test the robustness of the relationships when controlling for factors that traditionally explain large parts of the variance in Crash Frequency, specifically Urban and Highway environments. The first of these, Model 4.1, uses just the natural logarithm transformed $\mathrm{JR}_{\mathrm{Sit}}$ and $\mathrm{TF}_{\mathrm{Sit}}$ as dependent variables, extending the results from the previous sections to a country-wide setting and accounting for spatial autocorrelation. In Models 4.2 and 4.3 the binary variables for Highway and Urban locations are respectively added to the regression. Finally, both of these binary variables are included to generate the combined Model 4.4.

TABLE 5.4: Spatial combo regression results for the Crash Frequency (CF_{Pit}) dependent variable and iteratively added independent variables, with grids limited to those with a fleet Trip Frequency greater than 15 (TF_{Sit})

	Dependent Variable: ln (CF_{Pit})			
	Initial Model 4.1	**Urban Model 4.2**	**Highway Model 4.3**	**Full Model 4.4**
Independent Variables				
ln (JR_{Sit})	0.3176***	0.2948***	0.3224***	0.3043***
	(35.7135)	(39.8947)	(34.1915)	(31.2257)
ln (TF_{Sit})	0.1060***	0.1304***	0.0857***	0.0766***
	(7.6174)	(12.2520)	(5.1154)	(4.6636)
Urban		0.1417***		0.1651***
		(6.8906)		(7.0837)
Highway			0.0292	0.1250***
			(4.9491)	(3.6352)
Constant	-0.8754***	-0.9776***	-0.8694***	-0.9376***
	(-13.0789)	(-15.4511)	(-14.0448)	(-17.3417)
Spatial Lag	0.0618	0.0840*	0.1183***	0.2022***
	(1.4026)	(6.8906)	(3.0585)	(6.0833)
Spatial Error	0.6330***	0.5994***	0.5559***	0.4296***
	(25.9786)	(24.0536)	(21.4148)	(14.6869)
Observations	**4197**	**4197**	**4197**	**4197**
Spatial Pseudo R^2	**0.3899**	**0.4017**	**0.3913**	**0.4083**

z statistics in parentheses
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

Considering Model 4.1 initially, as seen in the negative binomial regressions from the previous sections, highly significant coefficients for both $ln(JR_{\text{Sit}})$ and $ln(TF_{\text{Sit}})$ are reported. Here the coefficient of $ln(JR_{\text{Sit}})$ falls within a similar range as the results from the previous section, and so with an increase of 10 % in Jerk Rate from the fleet at a location we would expect an increase of 3.1 % ($1.10^{0.3176} = 1.031$) in the Crash Frequency at that location. On the other hand, while the number of trips the fleet makes through the location also shows a significant proportional increase in the Crash Frequency, the coefficient is lower than in the previous models. In this case, the coefficient can be interpreted such that a 10 % increase in trips leads to a 1.0 % ($1.10^{0.1060} = 1.010$) increase in the number of accidents. Here the importance of incorporating spatial factors in such an analysis is potentially witnessed, since the spatial error is highly significant in all of the models. Therefore, there is an indication that the error of an observation affects the errors of its neighbours. As the individual trips of the fleet naturally pass through neighbouring grids, this behaviour can potentially be accounted for in the spatial model.

Model 4.2 sees the significance impact of the Urban binary variable that was included in the regression, and the spatial lag additionally becomes significant. Since spatial lag is the variable which captures the influence of neighbouring observations on the dependent variable, it can be determined that controlling for Urban and Rural locations highlights the similarity of neighbouring observations in the dataset. Model 4.3 adds the binary variable for whether the location is a Highway, here the variable itself does not have a significant effect on the number of crashes within that grid. However, we see that the coefficient and the significance of the spatial lag increase, indicating higher importance of the influence of neighbouring observations on the number of crashes.

Finally, Model 4.4 is the combined approach and is the best-fitting model with a pseudo R^2 of 0.4083. Here the significance of both the binary variables of Urban and Highway can be observed, along with the spatial lag and error for these models remaining significant. With regard to the impact of the binary variables, the results of this model are primarily considered. In including these variables, minor changes can be observed in the coefficients of $ln(JR_{\text{Sit}})$, from 0.3176 to 0.3043, and $ln(TF_{\text{Sit}})$, from 0.1060 to 0.0766, with the significance of both variables remaining. The binary variable coefficients in the model indicate the direct effect on the dependent variable of being in urban and highway areas respectively. Thus, the impact of being in an urban location increases the number of accidents at that location by 16.5 % when compared to rural areas. Moreover, highway locations had an increase of 12.5 % in the Crash Frequency when compared to other non-highway roads.

5.5 Discussion

The topic at hand has several implications for research, policy, and industry. At the core of the research is the hypothesis that the rate of CDEs at a specific location is related to the Crash Rate and Crash Frequency of that location. Thus, an explorative nationwide study was conducted to provide early evidence that the Crash Rate can be modelled by the Jerk Rate of vehicles that represent a connected, semi-, or fully-autonomous fleet. Here, we observe the significance of coefficients that would be expected from the literature and see promising quality metrics with regard to pseudo R^2 and χ^2. In addition, model fit generally increased where grids were only considered for analysis when they had a higher number of transits by the utilised fleet. This demonstrates that with more trips the measure of Jerk Rate becomes more representative of the population Crash Rate. However, in reality, traffic safety analysis is commonly limited to estimating Crash Frequency as a dependent variable, since the necessary traffic frequency data needed to calculate exposure measures, such as Crash Rate, is rarely available for whole road networks.

In order to reliably estimate the Crash Frequency of a location on the basis of Jerk Rate (without the corresponding data for overall population traffic frequency), the relationship between the traffic frequency of the field study fleet and total traffic frequency of the population has to be known. The results indicate that a road assistance fleet, such as the one used in this field study, can address this requirement. More specifically, evidence is provided of a well-defined relationship between the number of trips made by the fleet and the overall traffic frequency measure, ADT. In practice, it may be easier for interested parties to use a professional fleet to model both Crash Rate and Crash Frequency, rather than recruiting and collecting data from typical drivers within the general population. After validating the assumption that the fleet in the field study was representative of the population, models were developed to estimate Crash Frequency, and from the coefficient make an approximation for the Crash Rate exposure measure.

Through these steps, the results contribute to the existing body of research by going beyond the most recent studies that consider individual highway segments (Pande et al., 2017), applying the analysis to a subset of locations with ADT data, and providing model quality measures to help researchers and practitioners validate their approaches. Furthermore, in practice such an analysis would not be run merely on a small subset of grids where traffic frequency information is available, but over the whole road network where only Crash Frequency is typically collected. Therefore, the study at hand, which covers the majority of the Swiss road network, provides first insights into a nationwide approach. The importance of controlling for spatial lag and error effects is also highlighted, and attention drawn to how urban and highway variables, obtainable from map-matching services, impacted Crash Frequency at those locations and improved model fit when included in the analyses. Finally, the models presented in each section form a concrete starting point for automotive insurance companies looking to exploit the growing opportunity for active loss prevention, such as providing in-vehicle warnings and safe-routing services. These services, which build upon an accurate knowledge of dangerous locations on the road network, will be further enabled by the growing popularity and advanced features of the semi- and fully-autonomous vehicles in the coming decades.

Whilst this study demonstrates how traffic accident exposure can be modelled by naturalistic driving data, the results should be seen in the light of their limitations. While there are many jerk threshold values suggested in the related research, as well as other variables to detect CDEs, in this early stage just one threshold of jerk was considered in order to generate the dependent variables. In addition, spatial point data can be analysed in numerous ways, and it is important to mention that improvements in the analysis could be made by considering non-grid based spatial regression techniques. For example, map-matching traffic accidents, trips, and CDEs onto a representation of the road network and dividing this into network segments could be a promising future endeavour (Pande et al., 2017). Moreover, including attributes for these road segments, e.g. road curvature, as seen in other studies, could provide deeper insights into the relationship between CDEs and Crash Frequency.

5.6 Conclusions

Previous research has shown that the situational factors of crashes and near-misses are strongly related, and that naturalistic driving data can provide promising insights for traffic accident analysis, yet thus far there has been limited empirical evidence on whether the Crash Frequency and Crash Rate of locations can be reliably identified through analysis of CDEs. While low adoption levels of connected, semi-, and fully-autonomous vehicles currently make investigating this potential a challenge, preliminary analysis was made possible with the installation of the retrofit system that could access vehicle sensor measurements analogous to those found in these vehicles. Using this setup, jerk events were detected from 72 vehicles equipped with the in-vehicle system presented in Section 3.2, and sensor information collected from each car's CAN Bus, synonymous with data that facilitate the advanced features in modern vehicles. As such, the research at hand presented results from the 18 week field study of these vehicles, which covered over 690 000 km in Switzerland.

Overall, this early-stage research shows a consistent and robust relationship between the critical driving events from a naturalistic driving field study, and both the Crash Rate and Crash Frequency of locations across the Swiss road network. This not only validates the results reported by other studies (Pande et al., 2017), that have typically focused on limited segments of road over a small timeframe, but further demonstrates the potential for traffic safety analysis that could be achieved with the driving data collected from advanced vehicles. Moreover, the relationship between the fleet trip data and the traffic frequency of the general population was investigated, and the results highlight that a road assistance company or similar partner operating a professional fleet can provide valuable road usage insights. Furthermore, there is an indication that there is a minimum number of times an area should be driven through by such a fleet before an assessment conducted on the likelihood of traffic accidents occurring based on driving data.

These findings could be of particular importance to the automotive insurance industry, where there is an ongoing shift of moving away from purely 'reactive' business models toward active loss prevention. In this increasingly data-powered world, the ability to identify high risk locations on the road network could become of utmost value and the cornerstone for future safety-focused innovations and services. Through this, preventative measures can be incorporated into insurance products, such as encouraging customers to adapt their driving behaviour at potentially dangerous locations, or offering safe-routing options that avoid accident hotspots entirely, and enhance 'pay-how-you-drive' policies with 'pay-where-you-drive' incentives. Furthermore, semi-autonomous vehicles approaching these locations might drive more carefully to reduce risk or hand over control to the driver to transfer insurance liability. Finally, with the continued rise of automotive technology improvements, understanding the new practical opportunities of connected vehicles, such as those presented in this work, will become vital for determining suitable traffic safety approaches and strategies, both for insurers and policy makers in this long-standing field of traffic accident prevention.

Chapter 6

General Discussion and Implications

The previous chapters of this thesis have so far presented the motivation for, and background work on the two research topics that have focused on the impact of warnings on driving behaviour and the potential of driving data to predict accident hotspots. The results of these two studies have been presented in the corresponding chapters, along with the foundational materials and methods that enabled the investigations, and the naturalistic driving field study that was conducted. This chapter opens by briefly reintroducing the overall context and motivation of the work. Subsequently, the core results of the two studies are revisited, summarising the key findings of this thesis and outlining the implications for research, practice, and policy making. This is then followed by a discussion on the limitations of the research and how future work can advance the topics at hand, and finally, the chapter brings this thesis to its conclusion and closes with final remarks regarding the work presented.

6.1 Summary and Key Findings

Overall, the topic of road safety and how to improve it has been extensively discussed by academics, road authorities, automotive manufacturers, and governmental bodies alike over the last 60 years. The issues surrounding traffic accidents have been steadily increasing over the previous decades, and the injuries sustained from these incidents can often prove to be fatal and are now the eighth leading cause of death worldwide. This serious problem is not only seen in developing countries with limited investment in road infrastructure, but additionally affects high income countries. For example, the United States saw a sharp increase of 7.2 % in the amount of traffic fatalities between 2014 and 2015, bringing the annual number of lives lost to 35 092 people across the country. Among the many endeavours to reduce both the frequency and severity of accidents, road authorities prioritise dangerous sections of the road network for infrastructure improvement, identified by analysing the patterns and locations of traffic accidents. However, this traffic accident analysis is traditionally based on historic crash data and is restrictive in many ways, typically suffering from issues including small sample sizes, underreporting of traffic accidents, and data scarcity.

A new and dynamic source of naturalistic driving data variables, collected from the advanced sensors and technology embedded in connected, semi-, and fully-autonomous vehicles, potentially offers the opportunity to overcome the existing challenges and traditional limitations of accident analysis. Both road safety researchers and practitioners see the opportunity to reliably predict locations with a high level of traffic accident exposure by leveraging the technology in these vehicles to detect accidents and 'near miss incidents', otherwise known as critical driving events (CDEs). Yet, to date, there is limited empirical evidence on whether the perilousness of locations can be accurately predicted through naturalistic driving data, despite the great potential for improving road safety by identifying the locations of traffic accident hotspots with insights from these vehicles.

Furthermore, various safety-focused innovations become a possibility with the insights of these dangerous locations, for example, automotive manufacturers and insurances can start to offer reliable 'pay-where-you-drive' safe-routing services and incentives, and in situations where hazardous locations cannot be avoided, provide in-vehicle warnings of accident hotspots. While the potential of such in-vehicle warning systems are undisputed in their ability to improve driver safety in critical situations, such as avoiding imminent collisions, the vast majority of studies have focused on simulation and controlled field experiments. Moreover, the benefit of real world location analytics on accident hotspots as a data source for in-vehicle warnings has widely not been investigated.

When considering the promise of in-vehicle warning systems to reduce traffic accidents, along with the increasing availability of connected vehicle driving data, this thesis identified gaps in both research and practice regarding the real world assessment of such systems, and the potential of vehicle data to identify hazardous locations. From this, two research topics were investigated in this thesis: The first, what impact do in-vehicle warnings, utilising real world location data to generate interventions, have on driver behaviour in a realistic driving setting. The second, to what extent can the perilousness of our road networks be predicted on the basis of naturalistic driving data collected from connected vehicles. The results from the sets of analysis tackling these two topics are presented in the following sections respectively, summarising the key findings of the work and their implications for research, practice, and policy making.

6.1.1 Improving Driver Safety

In order to address the first research objective highlighted in Section 1.2.1, and determine whether warnings of upcoming dangerous locations have a positive effect on driving behaviour, a smartphone based in-vehicle warning system was developed and tested as part of an 18 week nationwide field test of professional drivers. In contrast to other in-vehicle studies that focus on simulation and controlled field experiments, typically involving imminent collision scenarios, location analytics were applied to a national historical accident dataset in order to identify accident hotspots on the Swiss road network. The locations and contextual causes for these historically hazardous locations were then used to generate warnings, which were provided through the in-vehicle system to an intervention group of drivers as they approached the specific accident hotspots during a 4 week experimental phase of the field study. During this period, real-time sensor data was collected from the vehicles of both a control and intervention group of drivers, and post-processing detected heavy braking events from this data so that the following research question from Section 4.1.1 could be investigated:

RQ 1a To what extent do warning interventions of upcoming dangerous locations, i.e. accident hotspots, impact the braking behaviour of drivers in a naturalistic driving setting?

In order to answer this research question, and account for the effect of the individual drivers among the control and intervention groups, multilevel mixed-effects logistic regression was applied and seven different models iteratively generated to test the robustness of the results with different independent variables. Based on these analyses the following three key findings, and their implications for research, practice and policy, are highlighted:

In contrast to existing studies, an immediate effect of warnings was not confirmed. Throughout all seven models that were generated, the independent variable that represented the difference between the control and the warning intervention group, i.e. whether a warning was shown or not, did not demonstrate a significant effect on the likelihood of a heavy braking event at the location occurring. Overall, this result offers two insights, the first is that unlike the promising reports from simulation and controlled field experiments, the strong effect of in-vehicle warnings could not be confirmed in our naturalistic driving setting. However, given that the vast majority of other studies typically focus on scenarios involving an upcoming collision, i.e. with another vehicle or pedestrian, invoking an immediate braking or evasive manoeuvre from the driver is typically the main aim of the in-vehicle warnings. In contrast to these studies, the research at hand focused on warning drivers of accident hotspots, locations that historically pose a safety risk and require the driver to be more aware of the driving situation and surrounding traffic in general, but that may not demand a guaranteed action in order to avoid a collision. Nonetheless, the warnings themselves had neither an immediate positive nor negative effect on the driver's braking behaviour at these locations, which leads to a second insight: that the interventions did not appear to distract the drivers and raise the likelihood of heavy braking events occurring. Overall, this is a positive result for the field of in-vehicle warnings, and the lack of a direct effect demonstrates the importance of exploring this topic in a realistic driving setting, along with building innovative artefacts and conducting experimental research. Despite this, field evidence supporting the impressive results reported in simulation and controlled experiments is still weak, and there remains a strong need for further naturalistic driving experiments with high resolution vehicle data to confirm if these systems can deliver in diverse field situations on their "lab promises".

However, in-vehicle warnings of historically dangerous locations had a significant improvement on driver braking behaviour over time. Unlike the direct effect of the accident hotspot warnings on driver behaviour, a positive and significant learning effect of the warnings was observed that remained stable throughout the first six models that were generated. This learning effect implies the following: the more often a driver in the intervention group encountered an accident hotspot and received the associated in-vehicle warning for that location, the less likely he was to have a heavy braking event at that location, effectively driving smoother and more controlled through the hotspot each time. Other studies outside of the traffic safety and accident prevention field, for example those in the health and education domains, have reported similar long-term effects of interventions, where significant results are found the more often an intervention was triggered (Bokhove and Drijvers, 2012; Brendryen and Kraft, 2008).

Both policy makers and industry players alike should find the main implication of these results, that accident hotspot warnings can improve driver behaviour over time, highly relevant to their existing work and approaches to road traffic safety. Numerous countries throughout the last decades have made various hardware-based vehicle safety systems mandatory in new vehicles, for example, air-bags and electronic stability programs. Similarly, policy makers should start to consider promoting software-based and data-powered in-vehicle warning systems that have been shown to encourage safer driver behaviour, since such tools are lightweight, low cost and highly scalable (Brendryen and Kraft, 2008). Hence, they could efficiently complement traditionally more complex and expensive approaches to improve traffic safety, such as altering the existing hazardous road infrastructure and installing and maintaining road warning signs. Ultimately, when such systems have been conclusively proven to prevent traffic accidents across multiple vehicle and driver demographics, and on various different road types, then corresponding regulation could enforce the inclusion of these in new vehicles. Prior to this occurring, consumer demand has historically led to the introduction of novel safety features being included in new vehicles. As such, the automotive industry might recognise that data-driven prevention services are positioned to be an effective means to address the distinct safety needs of consumers and form a basis for sustainable competitive differentiation. For vehicle manufacturers and mobility solution providers the presented accident hotspot warning use-case could extend their portfolio of safety features and connected car services, both of which are increasingly important for car buyers' purchase decisions (Wee et al., 2015).

Finally, there is evidence that an individual's personality plays a key role in the effectiveness of in-vehicle systems that support decision making behaviour. This key finding was identified when investigating the generalisability of the reported learning effect of the warnings on driver behaviour with respect to the individual drivers' characteristics, which was tested by including each driver's age and personality as moderators. When considering the final model that was generated, the significant interaction between the 'number of warnings' shown and the individual's level of Agreeableness indicates that the learning effect previously identified is dependent on this personality trait. As such, only those with reasonable levels of Agreeableness improved their driving behaviour due to the warnings provided by the system. The Agreeableness personality trait is linked to characteristics such as cooperation and social harmony, i.e. drivers have to be willing to "listen" to the in-vehicle system (John and Srivastava, 1999). This result is important for future research, not just in the field of in-vehicle warnings, but when researching interventions with any kind of system that supports decision making behaviour, of which personality is a longstanding predictor for (Sprotles and Kendall, 1986). Yet research on real-time feedback interventions have only recently started to consider the impact of personality as a key factor in human behaviour (Tiefenbeck et al., 2016), and the results of this thesis further confirm the importance of carefully reflecting the role and impact of subjects' Agreeableness in such systems. Thus, in order to improve driver safety, future work should identify the key determinants of the learning effects of such in-vehicle warning systems, and how best to facilitate these moving forward (Bokhove and Drijvers, 2012; Brendryen and Kraft, 2008).

Overall, the in-vehicle warning system that was developed in this thesis incorporates automatically generated accident hotspot interventions as an alternative to human selected locations or up-coming collision warnings, and is the first of its kind among the comprehensively reviewed systems. Moreover, the conducted field study was the first to test such warning interventions in real world conditions across a longitudinal field study, and provides evidence of an improvement on driver braking behaviour over time. Additionally, the importance of carefully reflecting the role and impact of subjects' Agreeableness personality trait when assessing any kind of system that supports decision making behaviour is also highlighted in the results, a finding not just applicable to the field of in-vehicle warnings. Furthermore, due to the integration of accident location reports and details, the proposed system can be easily extended to other parts of the world where similar data are compiled, either at a regional or national level, or where insurances have collected their own information on the locations and causes of traffic accidents. In regions where such reports are not routinely collected, then in-vehicle warning could still be offered if the locations of traffic accident hotspots can be predicted on the basis of driving data, and the results of research into this topic are discussed in the forthcoming section.

6.1.2 Predicting Accident Hotspot Locations

The traditional limitations and challenges associated with analysis of historical accident records can be potentially overcome by considering variables available from naturalistic driving data, which are increasingly available in connected, semi-, and fully-autonomous vehicles from the advanced sensors that enable their modern features. By leveraging the technology in these vehicles to detect critical driving events there is the opportunity to identify areas on the road network with a high level of traffic accident exposure. Thus, the full 18 week field study, along with the FEDRO traffic accident and frequency datasets, was utilised in order address the second research objective highlighted in Section 1.2.2, and determine whether it is possible to predict the locations of traffic accident hotspots on the basis of driving data and critical driving events.

As is the case in many countries, in Switzerland traffic frequency measurements are only available at limited locations across the road network, therefore accurate accident exposure analysis is restricted to these areas. Thus, in order to first investigate the relationship between Crash Rate and naturalistic driving data, both the traffic accident and field study real-time sensor datasets were sampled to only include data collected within 1 km^2 of the 192 traffic counting stations. At each of these locations, the frequency of traffic accidents occurring during the field study was normalised by the average daily traffic (ADT) of the same period to generate the Crash Rate dependent variable. Furthermore, post-processing identified heavy jerk events from the field study vehicles, along with the number of times each location was driven through, in order to investigate the first research question from Section 5.1.1:

RQ 2a To what extent can the Crash Rate of a location be predicted by CDE information from that location, assuming that Crash Frequency and ADT data are available?

To answer this research question, the jerk rate dependent variable was generated by normalising the number of jerk events by the number of trips through each location, and six different models were iteratively generated using negative binomial regression to test the robustness of the results. Each of these successive models increasingly restricted the locations that were modelled based on a minimum number of trips that the field study fleet had to have been determined to have made in the surrounding area. Based on these analyses the following key finding is highlighted:

There is a proportional relationship between the Crash Rate and the Jerk Rate of vehicles driving through an area. Throughout each of the six models that were assessed the coefficient for the Jerk Rate falls within a similar range and is highly significant. Furthermore, the null hypothesis that the Jerk Rate of a location does not explain the Crash Rate is rejected for all models when testing the χ^2 values. Thus, based on the coefficient of the best fitting model, where only locations that had been visited more than 15 times by the field study fleet were considered, the results of this thesis indicate that a 3.7 % increase in the Crash Rate of a location can be expected with a 10 % increase in the Jerk Rate. Additionally, as model fit generally increased when only grids with a higher number of transits by the utilised fleet were considered for analysis, the results also demonstrates that with more trips the measure of Jerk Rate becomes more representative of the population Crash Rate.

For both research and practice, if this relationship is validated by future studies, then heavy jerk events may be determined to be a suitable surrogate for traffic accidents when assessing the impact of new vehicles and safety related systems. In the near future, the presented approach of measuring critical driving events may form the basis to develop procedures for the assessment of the advanced features of semi- and fully-autonomous vehicles. Since an evaluation based on rarely occurring traffic accidents will naturally take a substantial amount of time, access to technology to measure these events could be considered crucial for the approval and insurance of such systems before they are permitted on the road network. Furthermore, the ability to consider the variables collected from naturalistic driving studies as a surrogate for accident risk and Crash Rate may form the basis for differentiating between the capabilities of different vehicles and the software powering their autonomous functions. In the same way that automotive manufacturers allow access to their vehicles' inner sensors to provide fuel efficiency measures, policy makers might consider mandating the sharing of the location and frequency of heavy jerk events incurred while driving. Access to this information would both enable the safety assessment and comparison of these vehicles and their modern systems, along with potentially providing road authorities new variables that may revolutionise traffic accident analysis.

However, one of the aforementioned limitations of traffic accident analysis includes the limited availability of traffic frequency measures, i.e average daily traffic, and the lack of data coverage restricts the generation of the Crash Rate variable across the road network. It is therefore important to consider in accident analysis the more common situation, where only the Crash Frequency can be leveraged for the whole road network as a dependent variable. Nevertheless, reliably assessing Crash Frequency on the basis of driving data comes with its own challenges, and the relationship between the traffic frequency of the utilised vehicle fleet (TF_{Sit}) and the total traffic frequency of the population (TF_{Pit}) should be known for each location. Thus, the assumption that the fleet driving behaviour was representative of that of the global population was tested at the locations where ADT data was available, and led to the second key finding in this section:

The number of trips made by the connected fleet in the field study has a proportional relationship with the population traffic frequency. By generating negative binomial regression models that increasingly limited the considered locations by the number of field study fleet trips, evidence is provided of a well-defined relationship between the number of these trips and the ADT of locations. The results indicate that the representative fleet assumption can be fulfilled by a road assistance fleet, such as the one used in this field study, and that there is a minimum number of trips that should be made through a location before this relationship becomes stable and an assessment of road safety should be made. For research, this offers a new variable to potentially act as a proxy for the rarely available traffic frequency data, and in practical settings, rather than recruiting and collecting data from typical drivers within the general population, it may be easier for interested parties to use or partner with a professional fleet. Once the assumption that the fleet in the field study was representative of the population had been validated, the second research question of Section 5.1.1 could be investigated:

RQ 2b To what extent can the Crash Frequency of a location be predicted by CDE information from that location, assuming we have sparse data coverage (only Crash Frequency information is available) and a well-defined relationship between TF_{Sit} and TF_{Pit}, i.e. a 'representative' fleet?

In order to answer this question, models were developed to estimate Crash Frequency both on the sub-sample of locations where traffic frequency measurements were available, and then across the majority of the Swiss road network utilising spatial regression techniques. These sets of analysis led the final key finding of Chapter 5:

There is a proportional relationship between the Crash Frequency and both the fleet trip frequency and the Jerk Rate of vehicles driving through an area. Following the approach of the prior sets of analysis, these relationships were initially modelled using negative binomial regression. In each of the six models the coefficient of both the jerk rate and trip frequency of the fleet are significant, and the null hypotheses that neither variables explain the Crash Frequency are rejected when testing the χ^2 values. From this, and with vehicles that satisfy the 'representative fleet' assumption, it can be concluded that Crash Frequency can be estimated from naturalistic driving data, and an approximation made for the Crash Rate exposure measure from the coefficient and equations presented in Section 5.4.3. The robustness of these relationships were then tested by a considering the more practical case, where such analyses would typically be conducted over a whole country, rather than on a small subset of grids where traffic frequency information is available. As such, the results of the spatial regression, covering over 4000 km^2 of the Swiss road network, highlight the importance of controlling for spatial lag and error effects, as well as how including explanatory variables for urban and highway locations impacted the Crash Frequency and improved model fit when included in the analyses.

As such, the study at hand provides first insights into a nationwide approach for utilising naturalistic driving data to overcome the challenges faced by traditional traffic accident analysis. The analyses presented form a concrete starting point for both research and practice to consider critical driving events as a surrogate for rarely occurring traffic accidents. However, while it can be demonstrated that there is a proportional relationship between heavy jerk events and the frequency of traffic accidents at the 1 km^2 grid based level, in practice this may not be sufficient to take action on specific infrastructure challenges or to provide the safety-focused services previously discussed. For such use-cases, future research should consider exploring the benefits of using either smaller grids, conceivably down to the level of 100 m in urban areas, or map-matching both the driving data and the traffic accidents to a representation of the road network.

Overall, a thorough search of the relevant literature indicates that this is the first work to incorporate a national level traffic accident and traffic volume dataset covering the same period as a nationwide naturalistic driving field study. Utilising these datasets, the results of this thesis demonstrate that, on a sub-sample of data where average daily traffic measurements were available, the rate of high jerk events of vehicles has a proportional relationship with the Crash Rate of the population. Since traffic frequency measures are expensive to collect and rarely available over the whole road network, it was further shown that the number of trips made by the connected fleet in the field study has a proportional relationship with the population traffic frequency. From this, this thesis further modelled Crash Frequency with the rate of high jerk events and the number of trips made by the fleet. Finally, spatial regression analysis was applied on Crash Frequency across locations covering the majority of the Swiss national road network, and highlights the relationship with the rate of high jerk events and fleet traffic frequency, along with urban and highway explanatory binary variables for these locations.

6.2 Limitations and Research Outlook

The promising findings and implications of this thesis should be considered in the light of the challenges of the research setting, and as discussed in Sections 4.5 and 5.5 respectively, the field study at the core of this thesis and the two set of analysis performed share several limitations. This section revisits the previously mentioned challenges while additionally highlighting other limitations of the thesis, and includes discussions on the future work that should be conducted to address these issues and advance the research topics in the future.

Firstly, an important key weakness of the field study that was conducted and presented in this thesis is the homogeneity of the researched sample. Out of the 72 participants taking part in the field study just one was female, and all were professional drivers. Moreover, when considering the sub-sample of 57 drivers who took part in the experimental phase for assessing the impact of the in-vehicle accident hotspot warning system, all were male. Their profession means that they are typically expected to be more experienced and demonstrate safer driving behaviour than other road users, thus this sample may not be easily generalisable to regular drivers. However, the proficiency of the sample of drivers, implies that the effect of the in-vehicle warning solution may have been dramatically underestimated, as there is potentially greater opportunity to improve driving behaviour for more regular drivers. Furthermore, with regard to the potential of naturalistic driving data to predict accident hotspots, the relationships between the various measures of road perilousness and critical driving events may be more pronounced when collecting data from less proficient drivers. Finally, while the overall sample size of the study was larger than comparable work (Pande et al., 2017; Zhang, Suto, and Fujiwara, 2009), the total number of drivers participating could be increased in future research, and the length of both the 18 week field study and 4 week experimental phase extended. Overall, in order to improve the reliability and generalisability of the results for both research objectives, future studies that wish to develop and validate this work should make use of a larger and more diverse sample of drivers, as well as measure the behaviour of drivers for a longer length of time.

A second common limitation shared by the two studies stems from the lack of a clear and common operationalisation of dangerous or 'near-miss' driving behaviour, for example heavy braking and other critical driving events. While there are many thresholds suggested in the related literature for heavy braking and jerk values, as well as other methods to identify and measure safe driving behaviour, in this early stage just one threshold was considered in order to generate each of the dependent variables. In a similar way, both researchers and practitioners should apply the models presented in this thesis with an element of caution. For example, with regards to the results of the spatial prediction of accident hotspots, transferring and applying these models to other locations outside of Switzerland may not be a trivial task based on the varying driving behaviour displayed across different cultures. In the majority of European countries, the United States, and Canada the road usage culture is typically calmer, steadier, and more respectful than might be experienced on a road in China, many Middle-Eastern countries, or northern Africa (Herrmann, Brenner, and Stadler, 2018). As such, the presented relationship between the locations of heavy braking events and traffic accidents may well be very different in regions of the world displaying a more organic and chaotic approach to the mixing of pedestrian with vehicles than the stop-and-wait culture that can be seen in Switzerland. Thus, in order to gain greater and more accurate insights into road safety there is a strong need to develop clear operationalisations of driver safety variables, and reliably validate these for multiple vehicle and driver demographics, on various different road types across the world.

With regard to the specific studies of this thesis, the investigation into predicting the perilousness of locations through naturalistic driving data in Chapter 5 has several additional limitations that encourage future research on this topic. Firstly, following the methodology of similar studies, the spatial analyses considered a grid-based approach of $1\,\text{km}^2$ unit cells to combine and measure the traffic accident and jerk event frequency across Switzerland (Kim, Brunner, and Yamashita, 2006). This approach comes with many advantages, including the ease of implementation and interpretation of the results, along with well defined strategies to capture the spatial effects, such as the contiguity weights that were utilised in this thesis. However, while the approach potentially provides actionable insights in rural and highway locations, the cell size of $1\,\text{km}^2$ might not create results that can be employed in urban areas where road infrastructure is more densely located. In order to more accurately pinpoint dangerous locations on the basis of driving data, future work could consider either applying smaller cell sizes across the road network, or utilising clustering techniques on the traffic accident locations that fall within specific cells of interest. A second consideration for future research could be analysing the spatial point data in alternative ways, in particular, map-matching traffic accidents, trips, and critical driving events onto a representation of the road network could provide more applicable results, especially for practitioners (Pande et al., 2017).

Furthermore, the work surrounding the in-vehicle accident hotspot warnings and their effect on driver behaviour in Chapter 4 is geared towards the development and validation of an innovative artefact. In accordance with this goal and in conformance with latest discussions in the scientific community (Von Alan et al., 2004), the chapter does not focus on theory, and future research should cover theoretical models of human behaviour to further increase generalisability of the findings. It also important to note that visual warnings of accident hotspots are just one type of feedback than can be provided by such a system. Due to the limitations of using a smartphone-based approach, and the inability to detect if drivers had disabled the audio of the device, only the visual warning was considered in the work presented in this thesis. Future studies should consider extending existing research and investigate the impact on driver behaviour and the preference of participants with regard to warnings delivered through tactile, audio, and visual feedback, and the combination of these, in a naturalistic setting. Furthermore, modern technological advances regarding the capabilities of the dashboard of modern vehicles and heads-up displays enable even more innovative approaches to accident hotspot warnings. Recent studies have shown the benefits of augmented reality systems in simulation experiments, which if validated in a naturalistic setting, could have the potential to improve accident hotspot warnings by better conveying spatial information about the location (Schwarz and Fastenmeier, 2017; Schwarz and Fastenmeier, 2018).

Finally, the in-vehicle warnings presented benefited from the detailed situational information recorded in the FEDRO reports, whereas outside of Switzerland similar data collection procedures are rarely standardised across a whole country. Various studies have shown that a general warning sign with limited information can still improve driver behaviour, and so both accident and naturalistic driving hotspots could be warned of in countries without such detailed accident accounts. However, users have previously reported that they generally preferred to receive more contextual warnings (Naujoks and Neukum, 2014a). If insurances, automotive manufacturers, and other industry players wish to offer such a warning service, then automatically and accurately classifying the identified accident hotspot locations without detailed situational accident reports might be a beneficial topic to explore in future research. Methods to generate insights into the underlying challenges of these area could include infrastructure recognition through satellite imagery or LIDAR data of the hotspots, or alternatively 'dash-cam' videos captured by vehicles themselves as they cross these locations (Gahr et al., 2018a).

To summarise, the research presented in this thesis contributes to creating a foundational starting point for a variety of future studies and developments, both in research and in practice, that focus on the large-scale assessment of road safety on the basis of driving data, and improving driver behaviour through in-vehicle warnings. For practice, these insights can facilitate the accident analysis work of road authorities and the implementation of in-vehicle warning services of automotive manufacturers and insurers, along with encouraging the field deployment of systems that can collect naturalistic driving data from fleet vehicles. The insights additionally demonstrate for researchers the importance of conducting longitudinal field assessments to validate the promising findings of 'lab-based' experiments, open up a new set of dynamic variables for traffic accident analysis, and provide a concrete starting point for promising follow-up studies. Furthermore, the in-vehicle system that was developed to in parallel collect naturalistic driving data from vehicles and provide real-time feedback through the smartphone interface to the drivers has proven to be a robust and stable research tool that provided rich naturalistic driving insights. With the growing popularity of smartphone turn-by-turn navigation applications, along with the emergence of driver assistance and insurance ‘driver scoring’ applications, such a retrofit smartphone solution might offer a platform to explore additional questions and theories related to both road safety and improving driver behaviour.

6.3 Conclusion

While the importance of preventing traffic accidents is not a new topic, without the continuous research of transportation scholars, road authorities, and automotive manufacturers and insurers into the different approaches and insights that could improve the situation, these accidents will continue to remain the primary cause of death by injury worldwide. The advent of autonomous vehicles brings the promise of a new era of traffic safety, where ultimately the threat of road accidents may potentially be drastically reduced or eliminated entirely. However, even in the most advanced markets, it may take decades to make this vision a reality, and in the meantime alternative options must be explored into how road safety can be dramatically improved if the steady increase in traffic accidents is to be reversed. With connected and semi-autonomous vehicles becoming increasingly common on our road networks, and the data available from the technology and sensors within these modern vehicles increasing in quality and quantity, traditional model-based approaches of accident prediction and exposure can start to be enhanced with the crowd-sourcing and real-time collection of data from these vehicles. Furthermore, the latest lab-based studies provide promising evidence that in-vehicle warning systems of dangerous locations can have positive effects on collision avoidance and driver safety, yet the benefit of real world location analytics on traffic accident hotspots as a source for these warnings had not been previously considered. As such, this thesis has presented results from the field assessment of in-vehicle warnings generated from the historical locations of traffic accidents and the impact these had on driving behaviour, along with insights into the potential of naturalistic driving data and critical driving events to predict accident hotspots.

Overall, if new data-powered road safety services are to be realised, and the growing opportunity to exploit active loss prevention business models explored, then automotive manufacturers and insurers will need to build upon an accurate knowledge of dangerous locations on the road network. The research at hand demonstrates that this knowledge will be further enabled by the growing popularity and advanced features of connected, semi-, and fully-autonomous vehicles in the coming decades, and the presented approach empowers those with access to data from these vehicles to rapidly identify areas of high accident exposure. Thus, with the increasing collection of driving data from almost every vehicle, automotive manufacturers, phone service providers, and insurance companies are set to become the gate keepers to highly accurate estimations of traffic and safety levels on our roads. Nonetheless, without a collaborative approach to data sharing, especially with regards to traffic frequency estimates, it may be hard to obtain a true picture of real-time road safety from critical driving events and fully realise the benefits that this knowledge can bring. Therefore, when determining suitable traffic safety approaches and strategies, it becomes vital for policy makers to understand the new capabilities and the reliance of findings from recent automotive technology advances, and to consider how the data from these vehicles can be best used for the benefit of all.

Ultimately, the insights of this thesis show that the exploration of naturalistic driving data holds great promise to combat the traditional challenges of traffic accident analysis and identify accident hotspots on our road networks without relying on historical crash data. The results further demonstrate how in-vehicle warnings of accident hotspots can improve driver behaviour over time by encouraging awareness of upcoming road challenges, offering the opportunity of a cost-effective, scalable, and dynamic approach to promoting safer driving on our roads. This data-powered perspective to road safety should encourage the collaboration between road authorities and the many players in the automotive industry. Together, they have the unrealised potential to determine existing and newly arising locations of high accident probability, and the power to intervene through in-vehicle warnings and other methods in order to prevent traffic accidents from occurring.

References

Aarts, Letty and Ingrid Van Schagen (2006). "Driving speed and the risk of road crashes: a review". In: *Accident analysis & prevention* 38.2, pp. 215–224.

Aichinger, Claus et al. (2016). "Using low-cost smartphone sensor data for locating crash risk spots in a road network". In: *Transportation research procedia* 14, pp. 2015–2024.

Albright, J et al. (2016). *Automobile insurance in the era of autonomous vehicles*. Tech. rep. URL: `https://home.kpmg.com/content/dam/kpmg/pdf/2016/05/marketplace-change.pdf`.

An, P Edgar and Chris J Harris (1996). "An intelligent driver warning system for vehicle collision avoidance". In: *Ieee transactions on systems, man, and cybernetics-part a: systems and humans* 26.2, pp. 254–261.

Anastasopoulos, Panagiotis Ch and Fred L Mannering (2009). "A note on modeling vehicle accident frequencies with random-parameters count models". In: *Accident analysis & prevention* 41.1, pp. 153–159.

Anastasopoulos, Panagiotis Ch et al. (2012). "A study of factors affecting highway accident rates using the random-parameters tobit model". In: *Accident analysis & prevention* 45, pp. 628–633.

Anderson, Tessa K. (2009). "Kernel density estimation and K-means clustering to profile road accident hotspots". In: *Accident analysis and prevention* 41.3, pp. 359–364. ISSN: 00014575.

Anselin, Luc (1993). *The moran scatterplot as an ESDA tool to assess local instability in spatial association*. Regional Research Institute, West Virginia University Morgantown, WV.

Anselin, Luc and Sergio Joseph Rey (2014). *Modern spatial econometrics in practice: a guide to GeoDa, GeoDaSpace and PySAL*. GeoDa Press LLC.

Araújo, Rui et al. (2012). "Driving coach: a smartphone application to evaluate driving efficient patterns". In: *Intelligent vehicles symposium (iv), 2012 ieee*. IEEE, pp. 1005–1010.

AXA Winterthur (2016). *AXA drive recorder*. Accessed on 06.06.2016. URL: `https://www.axa-winterthur.ch/de/privatpersonen/angebote/drive-recorder`.

Bagdadi, Omar and András Várhelyi (2011). "Jerky driving—an indicator of accident proneness?" In: *Accident analysis & prevention* 43.4, pp. 1359–1363.

Bansal, Prateek and Kara M Kockelman (2017). "Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies". In: *Transportation research part a: policy and practice* 95, pp. 49–63.

Bergasa, Luis M et al. (2014). "Drivesafe: an app for alerting inattentive drivers and scoring driving behaviors". In: *2014 ieee intelligent vehicles symposium proceedings*. IEEE, pp. 240–245.

Bhat, Chandra R et al. (2014). "A count data model with endogenous covariates: formulation and application to roadway crash frequency at intersections". In: *Analytic methods in accident research* 1, pp. 53–71.

Bivand, Roger S et al. (2008). *Applied spatial data analysis with R*. Vol. 747248717. Springer.

Bokhove, Christian and Paul Drijvers (2012). "Effects of a digital intervention on the development of algebraic expertise". In: *Computers & education* 58.1, pp. 197–208.

Brendryen, Håvar and Pål Kraft (2008). "Happy ending: a randomized controlled trial of a digital multi-media smoking cessation intervention". In: *Addiction* 103.3, pp. 478–484.

Campbell, John L et al. (2007). "Crash warning system interfaces: human factors insights and lessons learned". In: *Nhtsa report# hs* 810, p. 697.

Cao, Yujia et al. (2010). "Local danger warnings for drivers: the effect of modality and level of assistance on driver reaction". In: *Proceedings of the 15th international conference on intelligent user interfaces*. ACM, pp. 239–248.

Carson, Jodi and Fred Mannering (2001). "The effect of ice warning signs on ice-accident frequencies and severities". In: *Accident analysis and prevention* 33.1, pp. 99–109. ISSN: 00014575.

Castignani, German, Raphaël Frank, and Thomas Engel (2013). "An evaluation study of driver profiling fuzzy algorithms using smartphones". In: *2013 21st ieee international conference on network protocols (icnp)*. IEEE, pp. 1–6.

Cheng, Wen and Simon P Washington (2005). "Experimental evaluation of hotspot identification methods". In: *Accident analysis & prevention* 37.5, pp. 870–881.

Coutinho-Rodrigues, João et al. (1997). "Interactive spatial decision-support system for multiobjective hazardous materials location-routing problems". In: *Transportation research record: journal of the transportation research board* 1602, pp. 101–109.

CSS Insurance (2017). *Health Account and Health Account Bonus*. URL: https://www.css.ch/en/home/privatpersonen/krankenkasse/zusatzversicherung/gesundheitskonto.html.

D'Agostino, C. et al. (2015). "Learning-based driving events recognition and its application to digital roads". In: *Intelligent transportation systems, ieee transactions on* 16.4, pp. 2155–2166.

Dahlinger, André (2018). *The impact of abstract vs. concrete feedback information on eco-driving*. Doctoral Thesis, University St. Gallen.

Dahlinger, André, Benjamin Ryder, and Felix Wortmann (2015). "Car as a sensor - paying people for providing their car data". In: *Proceedings of the 5th International Conference on Internet of Things*. IEEE. URL: http://www.iot-conference.org/iot2015/wp-content/uploads/2015/11/IoT2015_PS07_1570213122.pdf.

Dahlinger, André et al. (2017). "Feldexperiment zur wirksamkeit von konkretem vs. abstraktem eco-driving feedback". In: *Proceedings of the 13th International Wirtschaftsinformatik Conference*. AIS. URL: https://wi2017.ch/images/wi2017-0292.pdf.

Dahlinger, André et al. (2018). "The impact of abstract vs. concrete feedback design on behavior - insights from a large eco-driving field experiment". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, p. 379. URL: https://dl.acm.org/citation.cfm?id=3173953.

Deacon, John A, Charles V Zegeer, and Robert C Deen (1974). *Identification of hazardous rural highway locations*. Transportation Research Board.

Deublein, Markus et al. (2015). "A bayesian network model to predict accidents on Swiss highways". In: *Infrastructure asset management* 2.4, pp. 145–158.

Dingus, Thomas A et al. (2006). *The 100-car naturalistic driving study, phase ii-results of the 100-car field experiment*. Tech. rep.

Emmerink, Richard HM et al. (1996). "Variable message signs and radio traffic information: an integrated empirical analysis of drivers' route choice behaviour". In: *Transportation research part a: policy and practice* 30.2, pp. 135–153.

Erdogan, Saffet et al. (2008). "Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar". In: *Accident analysis and prevention* 40.1, pp. 174–181. ISSN: 00014575.

Eren, Haluk et al. (2012). "Estimating driving behavior by a smartphone". In: *Intelligent vehicles symposium (iv), 2012 ieee*. IEEE, pp. 234–239.

Erke, Alena (2008). "Effects of electronic stability control (ESC) on accidents: a review of empirical evidence". In: *Accident analysis & prevention* 40.1, pp. 167–173.

Ester, Martin et al. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise". In: *Kdd*. Vol. 96. 34, pp. 226–231.

Fagnant, Daniel J and Kara Kockelman (2015). "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations". In: *Transportation research part a: policy and practice* 77, pp. 167–181.

Fazeen, Mohamed et al. (2012). "Safe driving using mobile phones". In: *Ieee transactions on intelligent transportation systems* 13.3, pp. 1462–1468.

Federal Office of Topography (2016). *Approximate formulas for the transformation between Swiss projection coordinates and WGS84*. Tech. rep. URL: {https://www.swisstopo.admin.ch/content/swisstopo-internet/en/online/calculation-services/_jcr_content/contentPar/tabs/items/documents_publicatio/tabPar/downloadlist/downloadItems/19_1467104393233.download/ch1903wgs84_e.pdf}.

Ferreira, J.C., J. de Almeida, and A. Rodrigues da Silva (2015). "The impact of driving styles on fuel consumption: a data-warehouse-and-data-mining-based discovery process". In: *Intelligent transportation systems, ieee transactions on* 16.5, pp. 2653–2662.

Fildes, Brian et al. (2015). "Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes". In: *Accident analysis & prevention* 81, pp. 24–29.

Flahaut, Benoıt et al. (2003). "The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach". In: *Accident analysis & prevention* 35.6, pp. 991–1004.

Frey, Remo Manuel et al. (2016). "Universal food allergy number". In: *Proceedings of the 6th International Conference on the Internet of Things*. ACM, pp. 157–158. URL: `https://dl.acm.org/citation.cfm?id=2998462`.

Gahr, Bernhard et al. (2018a). "A crowd sensing approach to video classification of traffic accident hotspots". In: *Proceedings of the 14th International Conference on Machine Learning and Data Mining*. Springer. URL: `https://link.springer.com/chapter/10.1007/978-3-319-96133-0_14`.

Gahr, Bernhard et al. (2018b). "Driver identification via brake pedal signals a replication and advancement of existing techniques (forthcoming)". In: *Proceedings of the 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE.

Geurts, Karolien, Isabelle Thomas, and Geert Wets (2005). "Understanding spatial concentrations of road accidents using frequent item sets". In: *Accident analysis & prevention* 37.4, pp. 787–799.

Gianfranco, Fancello, Stefano Soddu, and Paolo Fadda (2017). "An accident prediction model for urban road networks". In: *Journal of transportation safety & security*, pp. 1–19.

Global New Car Assessment Programme (2015). *Democratising car safety: road map for safer cars 2020*. Tech. rep. URL: `http://www.globalncap.org/wp-content/uploads/2015/04/road-map-2020.pdf`.

Gregor, Shirley and Alan R Hevner (2013). "Positioning and presenting design science research for maximum impact". In: *Mis quarterly* 37.2, pp. 337–356.

Gregory, Bree et al. (2016). "Differential effects of traffic sign stimuli upon speeding in school zones following a traffic light interruption". In: *Accident analysis & prevention* 86, pp. 114–120.

Greibe, Poul (2003). "Accident prediction models for urban roads". In: *Accident analysis & prevention* 35.2, pp. 273–285.

Guo, Feng et al. (2010a). "Evaluating the relationship between near-crashes and crashes: can near-crashes serve as a surrogate safety metric for crashes?" In:

Guo, Feng et al. (2010b). “Near crashes as crash surrogate for naturalistic driving studies”. In: *Transportation research record: journal of the transportation research board* 2147, pp. 66–74.

Hauer, Ezra (1995). “On exposure and accident rate”. In: *Traffic engineering & control* 36.3.

Hauer, Ezra (1996). “Identification of sites with promise”. In: *Transportation research record: journal of the transportation research board* 1542, pp. 54–60.

Hauer, Ezra (1997). *Observational before/after studies in road safety. estimating the effect of highway and traffic engineering measures on road safety*. Elsevier.

Hauer, Ezra and Bhagwant N Persaud (1984). *Problem of identifying hazardous locations using accident data*. Transportation Research Board.

Hauer, Ezra et al. (2002). “Estimating safety by the empirical bayes method: a tutorial”. In: *Transportation research record: journal of the transportation research board*, pp. 126–131.

Herrmann, Andreas, Walter Brenner, and Rupert Stadler (2018). *Autonomous driving: how the driverless revolution will change the world*. Emerald Publishing Limited.

Hirst, Stephen and Robert Graham (1997). “The format and presentation of collision warnings”. In: *Ergonomics and safety of intelligent driver interfaces*, pp. 203–219.

Iceland Review (2014). *Hekla volcano eruption hazard signs put up*. visited on 11/26/2015. URL: `http://icelandreview.com/news/2014/03/19/hekla-volcano-eruption-hazard-signs-put`.

Imkamon, T et al. (2008). “Detection of hazardous driving behavior using fuzzy logic”. In: *Electrical engineering/electronics, computer, telecommunications and information technology, 2008. ecti-con 2008. 5th international conference on*. Vol. 2. IEEE, pp. 657–660.

International Road Federation (2006). *A cost effective means of optimising road safety*. Tech. rep. visited on 24.02.2016. URL: `http://www.irfnet.ch/files-upload/pdf-files/PTCRS_publication.pdf`.

Jacobs, Goff, Amy Aeron-Thomas, and Angela Astrop (2000). “Estimating global road fatalities”. In:

John, Oliver P and Sanjay Srivastava (1999). "The big five trait taxonomy: history, measurement, and theoretical perspectives". In: *Handbook of personality: theory and research* 2.1999, pp. 102–138.

Johnson, Derick A and Mohan M Trivedi (2011). "Driving style recognition using a smartphone as a sensor platform". In: *2011 14th international ieee conference on intelligent transportation systems (itsc)*. IEEE, pp. 1609–1615.

Jun, Jungwook, Jennifer Ogle, and Randall Guensler (2007). "Relationships between crash involvement and temporal-spatial driving behavior activity patterns using GPS instrumented vehicle data". In: *86th annual meeting of the transportation research board, washington, dc*.

Karaduman, O et al. (2013). "An effective variable selection algorithm for aggressive/calm driving detection via CAN bus". In: *Connected vehicles and expo (iccve), 2013 international conference on*. IEEE, pp. 586–591.

Kazazi, Juela, Susann Winkler, and Mark Vollrath (2015). "Accident prevention through visual warnings: how to design warnings in head-up display for older and younger drivers". In: *2015 ieee 18th international conference on intelligent transportation systems*. IEEE, pp. 1028–1034.

Keenan, Peter B (1998). "Spatial decision support systems for vehicle routing". In: *Decision support systems* 22.1, pp. 65–71. ISSN: 01679236.

Khan, Kamran et al. (2014). "DBSCAN: past, present and future". In: *Applications of digital information and web technologies (icadiwt), 2014 fifth international conference on the*. IEEE, pp. 232–238.

Kim, Karl, I Made Brunner, and Eric Y Yamashita (2006). "Influence of land use, population, employment, and economic activity on accidents". In: *Transportation research record* 1953.1, pp. 56–64.

Klauer, Sheila G et al. (2006). *The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car naturalistic driving study data*. Tech. rep.

Kockelman, Kara Maria and Young-Jun Kweon (2002). "Driver injury severity: an application of ordered probit models". In: *Accident analysis & prevention* 34.3, pp. 313–321.

Koenig, Kevin et al. (2016). *The Internet of Things in insurance*. Tech. rep. URL: `http://www.ey.com/Publication/vwLUAssets/EY_-_The_internet_of_things_in_insurance/$FILE/EY-the-internet-of-things-in-insurance.pdf`.

Lam, Louisa and SY Suen (1997). "Application of majority voting to pattern recognition: an analysis of its behavior and performance". In: *Ieee transactions on systems, man, and cybernetics-part a: systems and humans* 27.5, pp. 553–568.

Legendre, Pierre (1993). "Spatial autocorrelation: trouble or new paradigm?" In: *Ecology* 74.6, pp. 1659–1673.

Li, Zhibin, Koohong Chung, and Michael Cassidy (2013). "Collisions in freeway traffic: influence of downstream queues and interim means to address them". In: *Transportation research record: journal of the transportation research board* 2396, pp. 1–9.

Lie, Anders et al. (2006). "The effectiveness of electronic stability control (ESC) in reducing real life crashes and injuries". In: *Traffic injury prevention* 7.1, pp. 38–43.

Lord, Dominique and Fred Mannering (2010). "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives". In: *Transportation research part a: policy and practice* 44.5, pp. 291–305.

Mannering, Fred, Walter Kilareski, and Scott Washburn (2007). *Principles of highway engineering and traffic analysis*. John Wiley & Sons.

Mannering, Fred L and Chandra R Bhat (2014). "Analytic methods in accident research: methodological frontier and future directions". In: *Analytic methods in accident research* 1, pp. 1–22.

Manyika, James et al. (2011). *Big data: the next frontier for innovation, competition, and productivity*. URL: `http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation`.

McMahon, Kate and Said Dahdah (2008). *The true cost of road crashes: valuing life and the cost of a serious injury*. London: iRAP.

Mednis, Artis, Atis Elsts, and Leo Selavo (2012). "Embedded solution for road condition monitoring using vehicular sensor networks". In: *The 6th ieee international conference on application of information and communication technologies (aict'12)*, pp. 1–5.

Meeuw, Arne et al. (2018). "Lokalpower: enabling local energy markets with user-driven engagement". In: *Extended abstracts of the 2018 chi conference on human factors in computing systems*. ACM. URL: `https://dl.acm.org/citation.cfm?id=3188610`.

Meng, Fanxing and Charles Spence (2015). "Tactile warning signals for in-vehicle systems". In: *Accident analysis and prevention* 75, pp. 333–346. ISSN: 00014575.

Metromile (2018). *Metromile app*. Accessed on 01.08.2016. URL: `https://www.metromile.com/technology/`.

Miaou, Shaw-Pin (1994). "The relationship between truck accidents and geometric design of road sections: poisson versus negative binomial regressions". In: *Accident analysis & prevention* 26.4, pp. 471–482.

Mohan, Prashanth, Venkata N Padmanabhan, and Ramachandran Ramjee (2008). "Nericell: rich monitoring of road and traffic conditions using mobile smartphones". In: *Proceedings of the 6th acm conference on embedded network sensor systems*. ACM, pp. 323–336.

Montella, Alfonso (2010). "A comparative analysis of hotspot identification methods". In: *Accident analysis & prevention* 42.2, pp. 571–581.

Naujoks, Frederik and Alexandra Neukum (2014a). "Specificity and timing of advisory warnings based on cooperative perception". In: *Mensch & computer workshopband*, pp. 229–238.

Naujoks, Frederik and Alexandra Neukum (2014b). "Timing of in-vehicle advisory warnings based on cooperative perception". In: *Proceedings of the human factors and ergonomics society europe chapter annual meeting*.

Naujoks, Frederik and Ingo Totzke (2014). "Behavioral adaptation caused by predictive warning systems – the case of congestion tail warnings". In: *Transportation research part f: traffic psychology and behaviour* 26, pp. 49–61.

Naujoks, Frederik et al. (2015). "Effectiveness of advisory warnings based on cooperative perception". In: *Iet intelligent transport systems* 9.6, pp. 606–617.

NHTSA (2007). *Federal Motor Vehicle Safety Standards: Electronic Stability Control Systems; Controls and Displays*. Tech. rep., p. 327.

O'Donnell, CJ and DH Connor (1996). "Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice". In: *Accident analysis & prevention* 28.6, pp. 739–753.

Okabe, Atsuyuki, Toshiaki Satoh, and Kokichi Sugihara (2009). "A kernel density estimation method for networks, its computational method and a GIS-based tool". In: *International journal of geographical information science* 23.1, pp. 7–32.

OpenStreetMap (2016). *Openstreetmap*. Accessed on 26.02.2016. URL: `https://www.openstreetmap.org`.

Oppe, Siem (1979). "The use of multiplicative models for analysis of road safety data". In: *Accident analysis & prevention* 11.2, pp. 101–115.

Paefgen, Johannes, Thorsten Staake, and Elgar Fleisch (2014). "Multivariate exposure modeling of accident risk: insights from pay-as-you-drive insurance data". In: *Transportation research part a: policy and practice* 61, pp. 27–40.

Paefgen, Johannes, Thorsten Staake, and Frédéric Thiesse (2013). "Evaluation and aggregation of pay-as-you-drive insurance rate factors: a classification analysis approach". In: *Decision support systems* 56, pp. 192–201.

Paefgen, Johannes et al. (2012). "Driving behavior analysis with smartphones: insights from a controlled field study". In: *Proceedings of the 11th international conference on mobile and ubiquitous multimedia*. ACM, p. 36.

Pande, Anurag et al. (2017). "A preliminary investigation of the relationships between historical crash and naturalistic driving". In: *Accident analysis & prevention* 101, pp. 107–116.

Park, Seong-hun, Sung-min Kim, and Young-guk Ha (2016). "Highway traffic accident prediction using VDS big data analysis". In: *The journal of supercomputing* 72.7, pp. 2815–2831.

Peden, Margie et al. (2004). *World report on road traffic injury prevention*.

Peffers, Ken et al. (2007). "A design science research methodology for information systems research". In: *Journal of management information systems* 24.3, pp. 45–77.

Persaud, Bhagwant et al. (1997). "Crash reductions related to traffic signal removal in Philadelphia". In: *Accident analysis & prevention* 29.6, pp. 803–810.

Prasannakumar, V. et al. (2011). "Spatio-temporal clustering of road accidents: GIS based analysis and assessment". In: *Procedia - social and behavioral sciences*. Vol. 21, pp. 317–325. ISBN: 1877-0428.

Prinja, Shankar et al. (2015). "Estimation of the economic burden of injury in north India: a prospective cohort study". In: *The lancet* 385, S57.

Pulugurtha, Srinivas S, Vanjeeswaran K Krishnakumar, and Shashi S Nambisan (2007). "New methods to identify and rank high pedestrian crash zones: an illustration". In: *Accident analysis & prevention* 39.4, pp. 800–811.

Quddus, Mohammed (2013). “Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS”. In: *Journal of transportation safety & security* 5.1, pp. 27–45.

Quddus, Mohammed A (2008). “Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data”. In: *Accident analysis & prevention* 40.4, pp. 1486–1497.

Rammstedt, Beatrice and Oliver P John (2007). “Measuring personality in one minute or less: a 10-item short version of the big five inventory in English and German”. In: *Journal of research in personality* 41.1, pp. 203–212.

Ranjan, Rajiv (2014). “Streaming big data processing in datacenter clouds”. In: *Cloud computing, ieee* 1.1, pp. 78–83.

Ray, Julian J. (2007). “A web-based spatial decision support system optimizes routes for oversize/overweight vehicles in Delaware”. In: *Decision support systems* 43.4, pp. 1171–1185. ISSN: 01679236.

Reifel, J et al. (2014). *The internet of things: opportunity for insurers*. Tech. rep. URL: https://www.atkearney.com/documents/10192/5320720/Internet+of+Things+-+Opportunity+for+Insurers.pdf/4654e400-958a-40d5-bb65-1cc7ae64bc72.

Roland Berger (2015). *Internet of Things and Insurance*. Tech. rep. URL: https://www.the-digital-insurer.com/wp-content/uploads/2015/06/537-Roland_Berger_Internet_of_Things_and_insurance_20150513.pdf.

Ruscio, Daniele, Maria Rita Ciceri, and Federica Biassoni (2015). “How does a collision warning system shape driver's brake response time? The influence of expectancy and automation complacency on real-life emergency braking”. In: *Accident analysis and prevention* 77, pp. 72–81. ISSN: 00014575.

Ryder, Benjamin, André Dahlinger, and Felix Wortmann (2015). “Leveraging controller area network data for predicting vehicle position during GPS outage”. In: *Proceedings of the 5th International Conference on Internet of Things*. IEEE. URL: http://www.iot-conference.org/iot2015/wp-content/uploads/2015/11/IoT2015_PS05_1570213088.pdf.

Ryder, Benjamin, Bernhard Gahr, and André Dahlinger (2016). “An in-vehicle information system providing accident hotspot warnings”. In: *Proceedings of the 24th European Conference on Information Systems (ECIS)*. AIS. URL: http://aisel.

aisnet.org/cgi/viewcontent.cgi?article=1000&context=ecis2016_prototypes.

Ryder, Benjamin and Felix Wortmann (2017). "Autonomously detecting and classifying traffic accident hotspots". In: *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, pp. 365–370. URL: https://dl.acm.org/citation.cfm?id=3123199.

Ryder, Benjamin et al. (2017). "Preventing traffic accidents with in-vehicle decision support systems – the impact of accident hotspot warnings on driver behaviour". In: *Decision Support Systems* 99, pp. 64–74. URL: http://www.sciencedirect.com/science/article/pii/S0167923617300829.

Ryder, Benjamin et al. (2018). "Spatial prediction of traffic accidents with critical driving events – insights from a nationwide field study". In: *Transportation Research Part A: Policy and Practice*. URL: https://www.sciencedirect.com/science/article/pii/S0965856417310145.

Sanitas (2017). *Sanitas Active app*. URL: https://www.sanitas.com/en/index/private-customers/bewegung/sanitas_active_app.html.

Santos, Luís, João Coutinho-Rodrigues, and Carlos Henggeler Antunes (2011). "A web spatial decision support system for vehicle routing using Google Maps". In: *Decision support systems* 51.1, pp. 1–9.

Schwarz, Felix and Wolfgang Fastenmeier (2017). "Augmented reality warnings in vehicles: effects of modality and specificity on effectiveness". In: *Accident analysis & prevention* 101, pp. 55–66.

Schwarz, Felix and Wolfgang Fastenmeier (2018). "Visual advisory warnings about hidden dangers: effects of specific symbols and spatial referencing on necessary and unnecessary warnings". In: *Applied ergonomics* 72, pp. 25–36.

Seeliger, F et al. (2014). "Advisory warnings based on cooperative perception". In: *Intelligent vehicles symposium*, pp. 246–252.

Shankar, Venkataraman, Fred Mannering, and Woodrow Barfield (1995). "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies". In: *Accident analysis & prevention* 27.3, pp. 371–389.

Sheehan, Barry et al. (2017). “Semi-autonomous vehicle motor insurance: a bayesian network risk transfer approach”. In: *Transportation research part c: emerging technologies* 82, pp. 124–137.

SIGKDD (2014). *2014 SIGKDD test of time award*. Accessed on 18.04.2016. URL: http://www.kdd.org/News/view/2014-sigkdd-test-of-time-award.

Snijders, Tom AB (2011). *Multilevel analysis*. Springer.

Sprotles, George B and Elizabeth L Kendall (1986). “A methodology for profiling consumers' decision-making styles”. In: *Journal of consumer affairs* 20.2, pp. 267–279.

Swiss Federal Road Office (2013). *Swiss automatic road traffic counts - networks of counting stations*. Accessed on 26.02.2016. URL: https://www.astra.admin.ch/astra/en/home/documentation/traffic-data/data-and-publication/swiss-automatic-road-traffic-counts--sartc-/networks-of-counting-stations.html.

Swiss Federal Road Office (2017). *Swiss automatic road traffic counts - annual and monthly results*. Accessed on 26.02.2016. URL: https://www.astra.admin.ch/astra/en/home/documentation/traffic-data/data-and-publication/swiss-automatic-road-traffic-counts--sartc-/annual-and-monthly-results.html.

Swisscom (2017). *When big data brings digital agility to traffic management*. Accessed on 26.02.2016. URL: https://ict.swisscom.ch/2017/05/when-big-data-brings-digital-agility-to-traffic-management/.

Szénási, Sándor (2015). “GPU implentation of DBSCAN algorithm for searching multiple accident black spots”. In: *15th sgem geoconference on informatics, geoinformatics and remote sensing*.

Szénási, Sándor and Peter Csiba (2014). “Clustering algorithm in order to find accident black spots identified by GPS coordiantes”. In: *14th sgem geoconference on informatics, geoinformatics and remote sensing*. Vol. 1. SGEM2014 Conference Proceedings, ISBN 978-619-7105-10-0/ISSN 1314-2704, June 19-25, 2014, Vol. 1. STEF92 Technology, pp. 505–512.

Szénási, Sándor and Domokos Jankó (2016). “A method to identify black spot candidates in built-up areas”. In: *Journal of transportation safety & security*, pp. 1–25.

Taubman-Ben-Ari, Orit, Mario Mikulincer, and Omri Gillath (2004). "The multidimensional driving style inventory—scale construct and validation". In: *Accident analysis & prevention* 36.3, pp. 323–332.

Taubman-Ben-Ari, Orit and Dalia Yehiel (2012). "Driving styles and their associations with personality and motivation". In: *Accident analysis & prevention* 45, pp. 416–422.

Tey, Li Sian, Luis Ferreira, and Angela Wallace (2011). "Measuring driver responses at railway level crossings". In: *Accident analysis and prevention* 43.6, pp. 2134–2141. ISSN: 00014575.

Tey, Li Sian et al. (2014). "Microsimulation modelling of driver behaviour towards alternative warning devices at railway level crossings". In: *Accident analysis and prevention* 71, pp. 177–182. ISSN: 00014575.

Thomas, Isabelle (1996). "Spatial data aggregation: exploratory analysis of road accidents". In: *Accident analysis & prevention* 28.2, pp. 251–264.

Tiefenbeck, Verena et al. (2016). "Overcoming salience bias: how real-time feedback fosters resource conservation". In: *Management science*, mnsc.2016.2646.

Tin Leung, King et al. (2011). "A review of ground vehicle dynamic state estimations utilising GPS/INS". In: *Vehicle system dynamics* 49.1-2, pp. 29–58.

Tourangeau, Roger and Ting Yan (2007). "Sensitive questions in surveys". In: *Psychological bulletin* 133.5, p. 859.

Ülengin, Füsun et al. (2007). "An integrated transportation decision support system for transportation policy decisions: the case of Turkey". In: *Transportation research part a: policy and practice* 41.1, pp. 80–97.

United Nations (2015). "Transforming our world: the 2030 agenda for sustainable development". In: *Seventieth united nations general assembly, new york* 25. URL: https://sustainabledevelopment.un.org/post2015/transformingourworld/publication.

U.S. Department of Transportation (2016). *Traffic Fatalities Up Sharply in 2015*. URL: https://www.transportation.gov/briefing-room/traffic-fatalities-sharply-2015.

Von Alan, R Hevner et al. (2004). "Design science in information systems research". In: *Mis quarterly* 28.1, pp. 75–105.

Wang, Yiyi and Kara M Kockelman (2013). "A poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods". In: *Accident analysis & prevention* 60, pp. 71–84.

Wee, Dominik et al. (2015). *Competing for the connected customer – perspectives on the opportunities created by car connectivity and automation*. Tech. rep.

Werneke, Julia and Mark Vollrath (2013). "How to present collision warnings at intersections? – a comparison of different approaches". In: *Accident analysis and prevention* 52, pp. 91–99. ISSN: 00014575.

Whitelegg, John (1987). "A geography of road traffic accidents". In: *Transactions of the institute of british geographers*, pp. 161–176.

World Health Organization (2013a). *Global status report on road safety 2013*. Tech. rep. Accessed 15.02.2017, pp. 1–318. URL: http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/.

World Health Organization (2013b). *Strengthening road safety legislation: a practice and resource manual for countries*. World Health Organization.

World Health Organization (2015). *Global status report on road safety 2015*. Tech. rep. Accessed 15.02.2017, pp. 1–318. URL: http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/.

World Health Organization (2018). *Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016*. Tech. rep. Accessed 15.02.2017. URL: http://www.who.int/healthinfo/global_burden_disease/estimates/en/.

World Health Organization and World Bank (2004). *World report on road traffic injury prevention*. Tech. rep. Accessed on 24.02.2016. URL: http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en.

Xie, Zhixiao and Jun Yan (2008). "Kernel density estimation of traffic accidents in a network space". In: *Computers, environment and urban systems* 32.5, pp. 396–406. ISSN: 01989715.

Yang, Bong-Min and Jinhyun Kim (2003). "Road traffic accidents and policy interventions in Korea". In: *Injury control and safety promotion* 10.1-2, pp. 89–94.

Yu, Hao et al. (2014). "Comparative analysis of the spatial analysis methods for hotspot identification". In: *Accident analysis & prevention* 66, pp. 80–88.

Zhang, Junyi, Koji Suto, and Akimasa Fujiwara (2009). "Effects of in-vehicle warning information on drivers' decelerating and accelerating behaviors near an arch-shaped intersection". In: *Accident analysis and prevention* 41.5, pp. 948–958. ISSN: 00014575.

Appendix A

FEDRO Police Accident Report Form

FIGURE A.1: FEDRO Police Accident Report Form - Page 1

Allgemeine Angaben

Quelle | Unfall-Nr. | Unfalltyp | Hauptursache

Unfalldatum (Tag, Monat, Jahr) | Unfallzeit (Std. : Min.) | Sachschaden (in CHF)

Beteiligte

Objekte | Total Personen | Getötete | Lebensbedrohlich Verletzte | Erheblich Verletzte | Leicht Verletzte

Unfallort und -lokalisierung

Kanton | BFS-Gemeinde-Nr. | 410 innerorts | 411 ausserorts | +

Koordinaten . / . | Autobahn / Autostrasse (Bezeichnung, km, m, Richtung)

Gemeinde

PLZ / Ortschaft

Strasse / Haus-Nr.

Strassenabschnitt | Fahrbahn / Richtung

Strassenart
- 430 Autobahn
- 431 Autostrasse
- 432 Hauptstrasse
- 433 Nebenstrasse
- 434 AB-Nebenanlage
- 439 andere
- 440 Einfahrt Autobahn / -strasse
- 441 Ausfahrt Autobahn / -strasse
- 442 Rampe in Verzweigung
- 443 Einbahnstrasse
- 449 andere

Verkehrsbedingungen +
- 450 schwach
- 451 rege
- 452 stark
- 453 stockende Kolonne
- 454 stehende Kolonne
- 459 andere

Zonensignalisation
- 460 keine
- 461 Tempo-30-Zone
- 462 Begegnungs-Zone
- 463 Fussgänger-Zone
- 469 andere

Höchstgeschwindigkeit

km/h
- 470 temporär
- 471 Wechselsignalisation
- 479 andere

Unfallstelle
- 480 gerade Strecke
- 481 Kurve
- 482 Platz
- 483 Parkplatz
- 484 Rastplatz
- 485 Kreuzung
- 486 Kreisverkehrsplatz
- 487 Einmündung
- 489 andere
- 500 Ein- / Ausfahrt Parkplatz / Liegenschaft
- 501 Einmündung Feldweg
- 502 Einmündung Fussweg
- 503 Einmündung Radweg
- 504 Pannenstreifen
- 505 Streifen in Fahrbahnmitte
- 506 Radweg
- 507 Radstreifen
- 508 Trottoir
- 509 Parkfeld
- 510 Verkehrsberuhigung
- 511 Schutzinsel
- 512 Fussgängerstreifen
- 513 Haltestelle
- 519 andere

Strassenzustand
- 520 trocken
- 521 feucht
- 522 nass
- 523 Schneematsch
- 524 verschneit
- 525 vereist
- 529 andere
- 530 ölig, schmierig
- 531 verschmutzt
- 532 Rollsplitt / Sand
- 533 reduzierter Winterdienst
- 534 Schlaglöcher
- 535 Spurrinnen
- 539 andere

Strassenanlage
- 550 eben
- 551 Gefälle
- 552 Steigung
- 553 Kuppe
- 559 andere
- 560 Brücke / Überführung
- 561 Tunnel
- 562 Unterführung
- 563 Baustelle
- 564 Umleitung
- 569 andere

Witterung
- 580 schön
- 581 bedeckt
- 582 Regen
- 583 Schneefall
- 584 vereisender Regen
- 585 Hagel
- 589 andere
- 590 starker Wind
- 591 Nebel
- 592 Sonnenblendung

Verkehrsregelung
- 600 keine
- 601 LSA in Betrieb
- 602 LSA gelbblinkend
- 603 LSA nicht in Betrieb
- 604 in Betrieb
- 605 nicht in Betrieb
- 606 Handzeichengabe
- 609 andere
- 610 LSA auf Anmeldung

Vortrittsregelung
- 490 keine
- 491 Fussgängerstreifen
- 492 kein Vortritt, signalisiert
- 493 Rechtsvortritt
- 494 Stoppstrasse
- 495 Tram-Vortritt
- 499 andere

Bahnübergang +
- 570 kein
- 571 unbewacht
- 572 nur Blinklicht
- 573 Blinklicht und Schranke
- 579 andere
- 540 Blinklicht nicht in Betrieb
- 541 Schranke offen
- 549 andere

Lichtverhältnis
- 620 Tag
- 621 Dämmerung
- 622 Nacht
- 623 unbekannt

Sicht
- 630 keine Beeinträchtigung
- 631 Sichtbehinderung
- 632 unbekannt

Strassenbeleuchtung +
- 640 keine
- 641 ausser Betrieb
- 642 punktuell
- 643 durchgehend
- 649 andere

FIGURE A.2: FEDRO Police Accident Report Form - Page 2

Titelblatt-Rückseite:

Unfallhergang (Bericht)

Unfallskizze

Fahrstreifenbreiten in der Anfahrtsrichtung, Signalisation, Markierung (Richtungspfeile usw.), Strassennamen.
Objekte 1, 2, 3 usw. gemäss Unfallhergang und Objektblätter:

Legende:

Bewegungsrichtung vorwärts
Bewegungsrichtung rückwärts
unvorsichtiges Wegfahren
stehend

rutschend
schleudernd
indirekt beteiligtes Objekt
Standort (Zeugen usw.)

Fussgänger/in
Kollisionsstelle
Nordpfeil

FIGURE A.3: FEDRO Police Accident Report Form - Page 3

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Strassen ASTRA
Office fédéral des routes OFROU
Ufficio federale delle strade USTRA
Uffizi federal da vias UVIAS

Unfallaufnahmeprotokoll

Objektblatt

Allgemeine Angaben

Quelle ☐☐ Unfall-Nr. ☐☐☐☐☐☐☐ Objekt-Nr. ☐☐☐ Objekt-Kategorie ☐ 650 Fahrzeug ☐ 651 Fussgänger/in ☐ 652 Nichtgenügen der Meldepflicht (auch Führerflucht)

+

Anzahl Personen ☐☐☐ Hauptverursacher/in ☐ 660 ja ☐ 661 nein Ursache(n)

Fahrzeug-Immatrikulation

Kennzeichen Fahrzeug — Land — Kt. — Nummer — Nr.-Zusatz +

Marke ____ Typ ____ Farbe ____

Stamm-Nr. ☐☐☐.☐☐☐.☐☐☐

Bauartbedingte Höchstgeschwindigkeit: ≤30 ☐ 670; ≤45 ☐ 671; ≤80 ☐ 672; unbeschränkt ☐ 673

Anhänger / Auflieger
- 680 ☐ kein
- 681 ☐ bis 750 kg
- 682 ☐ über 750 kg

Kennzeichen Art
- ☐ 690 weiss Motorwagen
- ☐ 691 weiss Motorrad
- ☐ 692 gelb
- ☐ 693 grün (Landwirtschaft)
- ☐ 694 blau (Arbeitsfahrzeug)
- ☐ 695 braun (Ausnahmefahrzeug)
- ☐ 696 schwarz (Militärfahrzeug)
- ☐ 697 keine (z.B. Fahrrad)
- ☐ 698 ausländisch
- ☐ 699 andere

Fahrzeugart

- ☐ 710 Personenwagen bis 3.5 t bis 9 Plätze
- ☐ 711 Kleinbus bis 3.5 t über 9 Plätze
- ☐ 712 Lieferwagen bis 3.5 t
- ☐ 713 Sattelschlepper bis 3.5 t
- ☐ 714 Gesellschaftswagen über 3.5 t bis 17 Plätze
- ☐ 715 Gesellschaftswagen über 3.5 t über 17 Plätze
- ☐ 716 Linien- / Trolleybus
- ☐ 717 Lastwagen bis 7.5 t
- ☐ 718 Lastwagen über 7.5 t
- ☐ 719 Sattelschlepper bis 7.5 t
- ☐ 720 Sattelschlepper über 7.5 t
- ☐ 721 Motorfahrrad (ohne 731)
- ☐ 722 Motorrad bis 50 cm³ und max. 4 kW
- ☐ 723 Motorrad bis 125 cm³ und max. 11 kW
- ☐ 724 Motorrad bis 25 kW
- ☐ 725 Motorrad über 25 kW
- ☐ 726 leichtes Motorfahrzeug bis 550 kg (3- oder 4-rädrig)
- ☐ 727 leichtes Motorfahrzeug über 550 kg (3- oder 4-rädrig)
- ☐ 728 Tram
- ☐ 729 Bahn +
- ☐ 730 Fahrrad
- ☐ 731 Fahrrad mit elektrischer Tretunterstützung
- ☐ 732 Landwirtschaftliches Motorfahrzeug
- ☐ 733 unbekannt
- ☐ 738 andere motorisierte Fahrzeuge ____
- ☐ 739 andere nicht motorisierte Fahrzeuge ____
- ☐ 740 Trike
- ☐ 741 Quad
- ☐ 742 Invalidenfahrzeug
- ☐ 743 Fahrzeugähnliches Gerät (FäG)

Angaben zum Fahrzeug

- ☐ 700 Verbrennungsmotor
- ☐ 701 elektrischer Antrieb
- ☐ 702 handgeschaltet
- ☐ 703 Automat
- ☐ 704 Crash Recorder
- ☐ 705 ABS / ESP
- ☐ 706 4x4
- ☐ 707 Schneeketten
- ☐ 708 Spikes
- ☐ 709 andere ____

Anprall (max. 3)

- ☐ 750 Leitschranke
- ☐ 751 Insel / Inselpfosten
- ☐ 752 Schild / Mast / Pfosten
- ☐ 753 Zaun / Mauer / Geländer
- ☐ 754 Baum
- ☐ 755 Bahnschranke
- ☐ 756 korrekt parkiertes Fahrzeug
- ☐ 757 steigende Böschung
- ☐ 758 fallende Böschung
- ☐ 759 andere ____
- ☐ 760 Tier ____

Abblendlicht

- ☐ 770 ja
- ☐ 771 nein
- ☐ 772 unbekannt

Angaben zum/zur Lenker/in oder Fussgänger/in

Eigentumsverhältnis

- ☐ 780 Halter/in
- ☐ 781 Familienmitglied
- ☐ 782 Drittperson
- ☐ 783 Geschäftswagen
- ☐ 784 unbekannt

Lenker/in

- ☐ 790 Privatfahrer/in
- ☐ 791 Mietfahrer/in
- ☐ 792 Lernfahrer/in
- ☐ 793 Berufsfahrer/in
- ☐ 794 Militärfahrer/in
- ☐ 795 Strolchenfahrer/in
- ☐ 796 unbekannt

Ablenkung

- ☐ 800 keine
- ☐ 801 Telefon ohne Freisprechanlage
- ☐ 802 Telefon mit Freisprechanlage
- ☐ 803 Gerät (Navigation / Radio)
- ☐ 804 Mitfahrer/in
- ☐ 805 Tier
- ☐ 809 andere ____

Fahr- / Gehzweck

- ☐ 810 Taxi
- ☐ 811 Arbeitnehmertransport
- ☐ 812 öffentlicher Verkehr
- ☐ 813 Land- / Forstwirtschaft
- ☐ 814 SDR- / ADR-Transport
- ☐ 815 Geschäfts- / Gütertransport
- ☐ 816 Schulweg
- ☐ 817 Schülertransport
- ☐ 818 Arbeitsweg
- ☐ 819 Freizeit / Einkauf
- ☐ 820 Ferien- / Tagesreise
- ☐ 821 Kurierdienst
- ☐ 822 unbekannt

Vertrautheit mit der Strecke

- ☐ 830 keine / gering
- ☐ 831 gut
- ☐ 832 unbekannt

Angaben zum Führerausweis

- ☐ 840 Führerausweis vorhanden
- ☐ 841 Führerausweis entzogen
- ☐ 842 kein Führerausweis
- ☐ 843 Lernfahrausweis
- ☐ 844 nicht notwendig (z.B. Radfahrer/in, FäG)
- ☐ 845 unbekannt
- ☐ 850 mit gültiger ADR-Bescheinigung
- ☐ 851 auf Probe
- ☐ 852 Lernfahrt falsch begleitet

seit — Tag — Monat — Jahr — Land — Kategorie

Auflagen ____ PIN

Einfluss von Alkohol / Arznei- / Betäubungsmittel bei dem/der Lenker/in oder Fussgänger/in

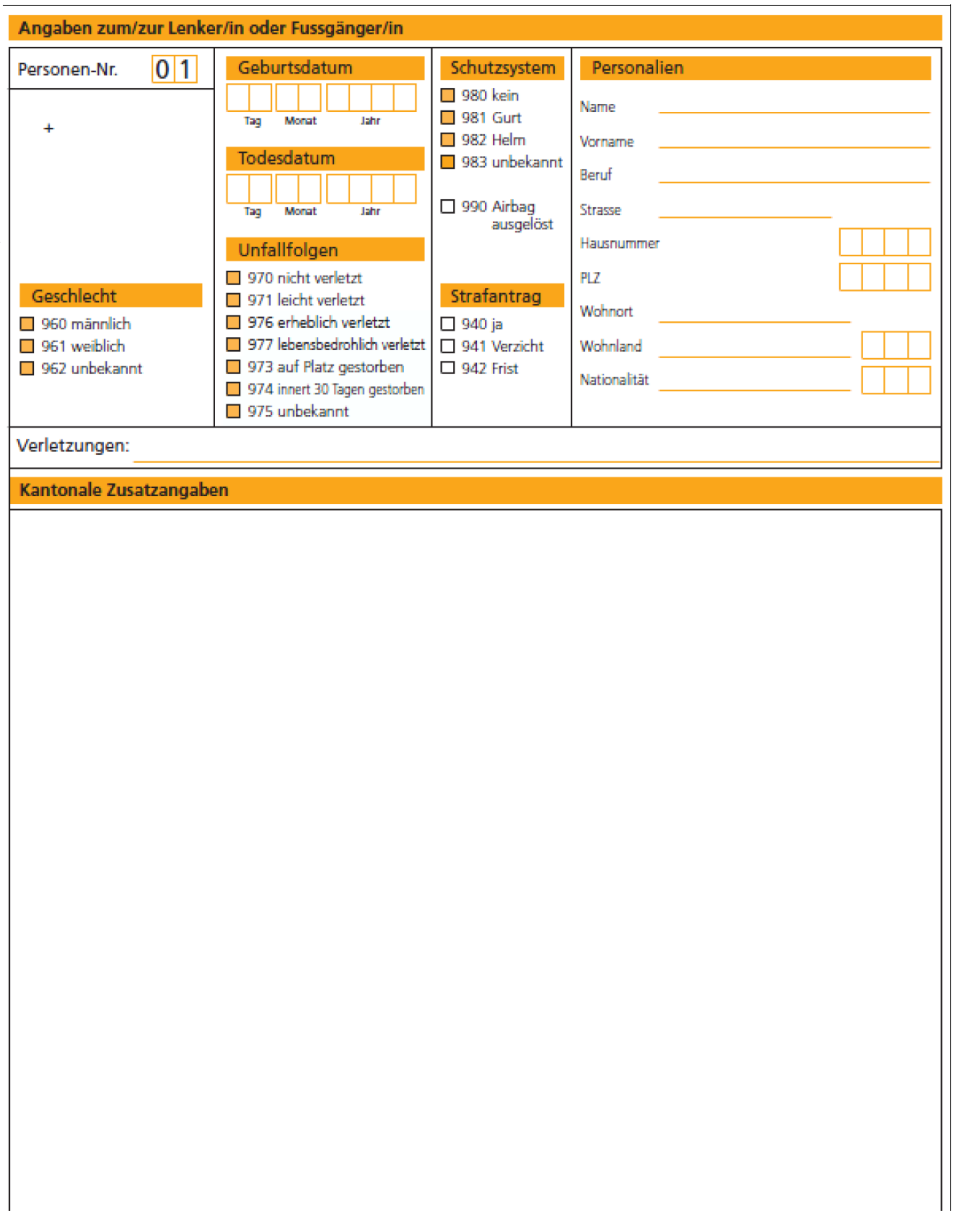
Verdacht auf Alkohol: ☐ ja (860) ☐ nein (861) +

Resultat Atemtest ☐.☐☐ ‰

Verdacht auf Arzneimittel: ☐ ja (880) ☐ nein (881) ☐ Erg.pos. (882)

Verdacht auf Betäubungsmittel: ☐ ja (890) ☐ nein (891) ☐ Erg.pos. (892)

FIGURE A.4: FEDRO Police Accident Report Form - Page 4



Angaben zum/zur Lenker/in oder Fussgänger/in

Personen-Nr. 0 1

+

Geschlecht
- 960 männlich
- 961 weiblich
- 962 unbekannt

Geburtsdatum

Tag Monat Jahr

Todesdatum

Tag Monat Jahr

Unfallfolgen
- 970 nicht verletzt
- 971 leicht verletzt
- 976 erheblich verletzt
- 977 lebensbedrohlich verletzt
- 973 auf Platz gestorben
- 974 innert 30 Tagen gestorben
- 975 unbekannt

Schutzsystem
- 980 kein
- 981 Gurt
- 982 Helm
- 983 unbekannt
- 990 Airbag ausgelöst

Strafantrag
- 940 ja
- 941 Verzicht
- 942 Frist

Personalien

Name

Vorname

Beruf

Strasse

Hausnummer

PLZ

Wohnort

Wohnland

Nationalität

Verletzungen:

Kantonale Zusatzangaben

Appendix B

Accident Hotspot Warning Signs and Text Combinations

Sign Image	English Text	German Text
	Dangerous Crossroad	Gefährliche Kreuzung
	Dangerous Junction	Gefährliche Einmündung
	Disregarding Right of Way	Vortritt wird missachtet
	Rear-end Collisions	Auffahrunfälle
	Control Speed	Auf Geschwindigkeit achten
	Swerving Accidents	Schleuder- und Selbstunfälle
	Control Speed	Auf Geschwindigkeit achten
	Dangerous Roundabout	Gefährlicher Kreisel
	Control Speed	Auf Geschwindigkeit achten
	Disregarding Right of Way	Vortritt wird missachtet

Sign Image	English Text	German Text
	Dangerous Curve	Gefährliche Kurve
	Control Speed	Auf Geschwindigkeit achten
	Disregarding Right of Way	Vortritt wird missachtet
	Parking Accidents	Parkierunfälle
	Disregarding Right of Way	Vortritt wird missachtet
	Caution Bikes	Auf Zweiräder achten
	Caution Cyclists	Auf Fahrräder achten
	Caution Motorbikes	Aur Motorräer achten
	Danger Changing Lanes	Gefahr bei Fahrstreifenwechsel
	Control Speed	Auf Geschwindigkeit achten

Sign Image	**English Text**	**German Text**
	Caution Pedestrians	Auf Fussgänger achten
	Dangerous Pedestrian Crossing	Gefährlicher Fussgängerstreifen
	Disregarding Right of Way	Vortritt wird missachtet
	Caution Trucks	Auf Lastwagen achten
	Caution Animals	Auf Tiere achten
	Caution Trains/Trams	Auf Zug/Tram achten
	Dangerous Railroad Crossing	Gefährlicher Bahnübergang
	Disregarding Right of Way	Vortritt wird missachtet
	Dangerous Railroad Crossing	Gefährlicher Bahnübergang
	Disregarding Right of Way	Vortritt wird missachtet

Sign Image	English Text	German Text
	Caution Dangerous Area	Achtung gefährliche Stelle
	Disregarding Right of Way	Vortritt wird missachtet
	Caution Buses	Auf Busse achten
	Disregarding Right of Way	Vortritt wird missachtet
	Disregarding Traffic Light	Ampel wird missachtet
	Dangerous Tunnel	Gefährlicher Tunnel

Curriculum Vitae

Personal Information

Benjamin William Ryder
Stampfenbachstrasse 67
8006 Zürich
Switzerland

Tel.: (+41) 078 928 59 11
Email: bryder@ethz.ch
Email: benryder1988.br@gmail.com

Born:. 05. 10. 1988 in Bedford, United Kingdom
Nationality: British

Education

2015 – Present **ETH Zürich, Switzerland**
Ph.D. candidate and doctoral researcher
Chair of Information Management of Prof. Dr. Fleisch

Benjamin is part of the Connected Car team in the Bosch IoT Lab, a collaboration between ETH Zürich, the University of St. Gallen, and the Bosch Group. His work focuses on making data-powered road safety a reality by combining traffic accident analysis, real-time data from connected vehicles, and in-vehicle warnings of dangerous locations for drivers.

2008 – 2012 **Imperial College London, United Kingdom**
Master of Engineering in Computing

2002 – 2007 **Sharnbrook Academy, United Kingdom**
AS- and A-Levels:
Maths (A), Law (A), Physics (B), Further Maths (B)

Professional Experience

2012 – 2015	**BAE Systems Applied Intelligence NetReveal** Technical Consultant
2012	**Hamlyn Centre for Robotic Surgery** UROP Summer Research Placement
2010 & 2011	**Deutsche Bank Group** Technical Analyst Internship and Work Placement

Programming Languages

Object Oriented	Java, C & C++
Functional	Python, Perl, Haskell, Prolog
Web & Mobile	HTML, CSS, JavaScript, PHP, Android
Databases	Cassandra, MySQL, Postgres, SAS

Honors & Awards

2017	**Supervised ETH Medal Master Thesis** Student: Mr. Philipp Egolf
2016	**Co-Authored Best Poster** 6th International Conference on the Internet of Things
2015	**Co-Authored Best Poster** 5th International Conference on the Internet of Things
2011	**Elected Public Relations Officer** Imperial College Gaelic Athletics Club
2009	**Elected Vice President** Imperial College Welsh Society
2007	**Elected Head Boy** Sharnbrook Academy

Other Interests

Travelling	Working Holiday Gap year between college and university, travelling to America, Raratonga, Australia and New Zealand.

	Worked on a winery and as a waiter in Australia and as part of the operations team for the Mount Hutt Ski Area in New Zealand
Fitness	Enjoys a variety of sports including snowboarding, surfing, and climbing
Education	Part time tutoring of 1st and 2nd year bio-engineering students in programming with C and C++
Startups	Completed the Innosuisse Startup Campus Business Concept and Venturelab ICT Business Creation courses. Self launched a selection of websites, e.g. timealign.org and mypreprint.com, and as part of the Ph.D. project released the safe driving Android App Avertu
Other	Also interested in photography, philosophy and good coffee

Zürich, November 7, 2018