www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

School of Transportation Sciences

Master of Transportation Sciences

Master's thesis

Risk event characterization in i-DREAMS for trucks with a special focus on time, weather and road type

Walta Brhane Abay

Thesis presented in fulfillment of the requirements for the degree of Master of Transportation Sciences, specialization Traffic Safety

SUPERVISOR :
Prof. dr. Tom BRIJS

MENTOR :
De heer Muhammad Wisal KHATTAK

2021
2022

School of Transportation Sciences

Master of Transportation Sciences

Master's thesis

Risk event characterization in i-DREAMS for trucks with a special focus on time, weather and road type

Walta Brhane Abay

Thesis presented in fulfillment of the requirements for the degree of Master of Transportation Sciences, specialization Traffic Safety

SUPERVISOR :

Prof. dr. Tom BRIJS

MENTOR :

De heer Muhammad Wisal KHATTAK

PREFACE

This master thesis represents the findings of the research investigation on risk event characterization in i-DREAMS for trucks with a special focus on time, weather, and road type, completed by Walta Brhane Abay, Master's student of Transportation Sciences (Traffic Safety), under the supervision of Professor Tom Brijs and Mr. Wisal Khattak. The initial inspiration for this research study stemmed from a curiosity to comprehend the characteristics of risky events based on time of day, weather, and road type, as only a limited number of risky events in a limited number of studies were investigated in previous works.

I want to express my sincerest gratitude to Professor Tom Brijs and Mr. Wisal Khattak for their insightful comments and assistance with my Master's thesis during the research and writing processes. Additionally, I want to thank the i-DREAMS project for providing me with all the necessary data to accomplish this master's thesis. The Flemish Inter-University Council (Vlaamse Interuniversitaire Raad/VLIR-UOS), which provided me with full financial support for my Master's degree at Hasselt University, is also deserving of appreciation.

SUMMARY

Along with its many benefits, road transportation also faces serious safety concerns like crash-related fatalities and injuries. The human factor, vehicle, and environment are the dominant factors that affect road safety (Arumugam & Bhargavi, 2019; Komackova & Poliak, 2016), with the human factor majorly contributing to road safety's negative consequences (Arumugam & Bhargavi, 2019; Barnard et al., 2016). Human factor encompasses physiological errors brought on by driver fatigue and drowsiness, as well as behavioral errors such as speeding, drunk driving, aggressive driving, and distracted driving (Arumugam & Bhargavi, 2019). A thorough understanding of road user behavior (drivers) is required to identify the underlying causes of the adverse consequences and suggest possible prevention measures. Thus, with an emphasis on three factors: time of the day, weather, and road type, this study aimed to characterize risky driving events experienced by truck drivers in Belgium.

The study's research questions focused on identifying risky driving events that reflect when and where truck drivers are most likely to show risky driving events and figuring out the relationship between the risky events. Eleven input variables related to risky driving events and three characterizing factors are used to address these research questions. To analyze these variables, characterizing elements (such as time of day and weather) were examined initially using the elbow method to ascertain the number of clusters before being further analyzed using k-means clustering to generate statistically significant distinct clusters. A road network shapefile was acquired from available sources to address the characterization of risky events involving road type. Multivariate analysis of variance was employed to determine the influence of time of day, weather, and road type on the prevalence of risky events (dependent variables). A correlation test was also conducted on the risky events to determine whether they occurred concurrently. Additionally, kernel density estimation is used to estimate the density of risky events based on the total number of risky events and three severity-based risky events (low, medium, and high), grouping based on time of day, weather, and type of road.

Based on the number of clusters obtained from the elbow method (four clusters for the time of the day and three clusters for weather), significantly distinct groups were formed using k-means clustering, where the cluster center for the four clusters of time of the day were morning, midday, afternoon/evening, and night/early morning) and for weather adverse, average, and clear weather conditions. Further analysis of the clusters based on the eleven risky events showed that the time of day substantially impacted fatigue-related risky events, especially those of medium severity. In addition, truck drivers were more likely to experience fatigue in the afternoon or evening than at midday. It was found that morning had the highest density of risky events, followed by midday and afternoon/evening (which are about the same), and overnight/early morning had the lowest density. Also, it was revealed that weather significantly affected total speeding, vulnerable road user collision avoidance, and low fatigue events. Further analysis indicated that more speeding events appeared in adverse than average weather conditions, whereas more fatigue events occurred in clearer than average weather conditions. Besides, density distribution based on weather clusters revealed that clear weather conditions had the highest density of events, followed by average and adverse weather conditions, respectively.

Moreover, road type significantly affected events involving speeding, acceleration, deceleration, steering, and tailgating. Compared to motorways, primary and secondary roads had higher speeding and deceleration events. Additionally, acceleration and steering events were higher on primary, secondary, and tertiary roads than on motorways, while tailgating events were the opposite. Also, it was shown that there were more speeding and steering events on primary roads than on trunk roads and more acceleration events on trunk roads than on motorways. Besides, tailgating events were more common on trunk roads than secondary and tertiary roads and primary than tertiary roads. Furthermore, tertiary roads and motorways had much higher densities of risky events.

The risky event correlation test revealed a higher likelihood that steering events will occur concurrently with either acceleration, deceleration, or speeding events. Besides, a strong correlation between speeding and tailgating events was found. There was also a probability that drivers would exhibit steering events when tailgating events occur. Additionally, fatigue events were highly correlated with speeding and tailgating events. Moreover, lane discipline and deceleration events were strongly correlated.

Generally, this study attempted to provide better insights into truck drivers' risky driving behaviors based on time, weather, and road type. This will serve as a foundation for further research on risky driving events by integrating additional characteristics left out of this study's scope and a more in-depth analysis of the factors included. The findings of this study may also serve as an inspiration for future investigations into safety precautions for risky driving events.

TABLE OF CONTENTS

PREFACE i
SUMMARY ii
TABLE OF CONTENTS iv
LIST OF FIGURES vii
LIST OF TABLES viii
LIST OF ABBREVIATIONS ix
1 INTRODUCTION 1
1.1 General 1
1.2 Overview of the i-DREAMS project 1
1.2.1 Data collection tools 2
1.2.1.1 Technologies for vehicle and driver state monitoring 2
1.2.1.2 Technologies for environment monitoring 3
1.2.2 Technologies utilized for intervention 4
1.2.2.1 Technologies utilized for real-time intervention 4
1.2.2.2 Technologies utilized for post-trip intervention 4
1.2.3 Targeted risky events for safety interventions 5
1.3 Problem statement 6
1.4 Research question 7
1.5 Objectives of the study 7
1.5.1 General Objective 7
1.5.2 Specific objectives 7
2 LITERATURE REVIEW 9
2.1 Review of Related Works 9
2.1.1 Road layout effects 9
2.1.2 Traffic condition Effect 12
2.1.3 Weather effect 14
2.1.4 Time of day effect 16
2.1.5 The combined effect of road layout, time of the day, weather, and traffic characteristics 18
2.1.6 Relationship between the risky events 21
2.2 Review of research methods 22
3 METHODOLOGY 25

3.1 General 25
3.2 Thesis Data 26
3.2.1 Participants and data collection procedure 26
3.2.2 Data employed 27
3.2.3 Overview of the risky driving events employed 29
3.2.4 Overview of the data obtained from the i-DREAMS project 30
3.2.5 Data obtained from open sources 31
3.2.5.1 Weather data 31
3.2.5.2 Road type data 32
3.3 Data processing 32
3.3.1 Organizing and processing risky events by time of the day 32
3.3.2 Organizing and processing risky events by weather 33
3.3.3 Organizing and processing risky events by location 33
3.3.3.1 Buffering 34
3.3.3.2 Join attributes by nearest 35
3.4 Method of data analysis 35
3.4.1 K-means Clustering 35
3.4.2 Elbow method 36
3.4.3 Multivariate analysis of variance (MANOVA) 37
3.4.3.1 Follow-up analysis 38
3.4.4 Correlation test 39
3.4.5 Kernel density estimation (KDE) 39
3.4.5.1 Categorization of hotspots 40
4 DATA ANALYSIS AND RESULTS 41
4.1 Descriptive statistics 41
4.1.1 Risky events description 41
4.1.2 Cluster description 42
4.1.2.1 Number of clusters 42
4.1.2.2 Cluster centre 43
4.1.3 Road type description 44
4.2 Characterization of risky events 46
4.2.1 Multivariate analysis of risky events 46
4.2.1.1 Analysis of risky events based on time of the day 46

4.2.1.2 Analysis of risky events based on weather conditions 47
4.2.1.3 Analysis of risky events based on road type 49
4.2.2 Correlation test 53
4.2.2.1 Correlation test based on total risky events 53
4.2.2.2 Correlation test based on low risky events 54
4.2.2.3 Correlation test based on medium risky events 54
4.2.2.4 Correlation test based on high risky events 54
4.2.3 Density estimation of risky events 55
4.2.3.1 Severity-based distribution of risky events 55
4.2.3.2 Distribution of risky events based on severity and characterizing factors 58
5 DISCUSSION 61
5.1 Time-based characterization 61
5.2 Weather-based characterization 62
5.3 Road type based characterization 62
5.4 Co-occurrence of risky events 63
5.5 Limitations 64
5.6 Future research 64
5.7 Implications 65
6 CONCLUSION 67
REFERENCES 69
APPENDICES 75
Appendix-I 75
Appendix - II 76
Appendix – III 91

LIST OF FIGURES

FIGURE 1 Conceptual framework of the i-DREAMS platform with monitoring (pillar I-left) and interventions (pillar II-right) (Kaiser et al., 2020)....... 2
FIGURE 2 Illustration of the warning stages (Katrakazas, Michelaraki, Yannis, Kaiser, Brijs, et al., 2020)... 4
FIGURE 3 i-DREAMS technologies for post-trip intervention (a) smartphone app (b) web-based platform (i-DREAMS project, 2020). 5
FIGURE 4 Relative change in injury crashes associated with an increase in the event counts: (a) braking events, (b) speed alert events (Gitelman et al., 2018). 10
Figure 5 Overview of the study's methodology....... 25
FIGURE 6 Safety promoting goals and performance objectives of the i-DREAMS project (K. Brijs et al., 2020). 30
FIGURE 7 Risky events and variables obtained from the i-DREAM project. 31
FIGURE 8 Data matching flow chart. 33
FIGURE 9 Buffering (a) risky events before buffering (b) risky events after buffering. 34
FIGURE 10 Frequency of risky events based on safety-promoting goals. 42
FIGURE 11 Elbow graph (a) Time of the day and (b) Weather. 43
FIGURE 12 Percentage of clustered risky events based on (a) Time of the day and (b) Weather....... 44
FIGURE 13 Belgium road network....... 45
FIGURE 14 Percentage of risky events based on road type. 45
FIGURE 15 Density distribution of total risky events. 55
FIGURE 16 Density distribution of low risky events....... 56
FIGURE 17 Density distribution of medium risky events. 57
FIGURE 18 Density distribution of high risky events....... 57
FIGURE 19 Density of risky events based on severity and time of the day. 58
FIGURE 20 Density of risky events based on severity and weather clusters. 59
FIGURE 21 Density of risky events based on severity and road type. 59

LIST OF TABLES

TABLE 1 Linkage of risky events and road layout........ 11
TABLE 2 Parameters used to define traffic condition........ 13
TABLE 3 Linkage of risky events and traffic conditions........ 14
TABLE 4 Research findings of previous studies regarding the impact of adverse weather conditions on traffic safety........ 15
TABLE 5 Linkage of risky events and weather........ 16
TABLE 6 Research findings on the effect of time of the day on truck crash occurrence........ 17
TABLE 7 Linkage of risky events and time of the day........ 18
TABLE 8 Research findings on the characterization of risky events based on different factors........ 19
TABLE 9 Research findings on the correlation between risky events........ 21
TABLE 10 Review of research methods........ 23
TABLE 11 Description of phases and duration overview........ 27
TABLE 12 Timetable of field experiment for both groups........ 28
TABLE 13 Weather parameters and definition........ 32
TABLE 14 Descriptive statistics of risky events (performance objectives)........ 41
TABLE 15 Clustering based on time of the day........ 43
TABLE 16 Clustering based on weather........ 44
TABLE 17 Between-subject effects: impact of time of the day on medium severity dependent variables 46
TABLE 18 Multiple comparisons: differences in medium fatigue events based on time of the day........ 47
TABLE 19 Between-subject effects: impact of weather on the dependent variables based on total events........ 48
TABLE 20 Multiple comparisons: differences in total risky events based on weather clusters........ 48
TABLE 21 Between-subject effect: impact of weather on low severity dependent variables........ 49
TABLE 22 Multiple comparisons: differences in low fatigue events based on weather clusters........ 49
TABLE 23 Between-subject effects: impact of road type on the dependent variables based on total events........ 50
TABLE 24 Between-subject effects: impact of road type on low severity dependent variables........ 51
TABLE 25 Between-subject effects: impact of road type on medium severity dependent variables........ 51
TABLE 26 Between-subject effects: impact of road type on high severity dependent variables........ 52
TABLE 27 Correlation test on total risky events based on time of the day........ 53
TABLE 28 Correlation test on low severity dependent variables based on time of the day........ 54
TABLE 29 Correlation test on medium severity dependent variables based on time of the day........ 54
TABLE 30 Correlation test on high severity dependent variables based on time of the day........ 55

LIST OF ABBREVIATIONS

AADT	Annual Average Daily Traffic
Adj-MAPE	Adjusted Mean Absolute Percentage Error
ADT	Average Daily Traffic
ANOVA	Analysis of variance
CO	Change Objectives
CSV	Comma Separated Values
DBSCAN	Density Based Spatial Clustering of Applications with Noise
df	degree of freedom
ECG	Electrocardiogram
EU	European Union
EUTRACO	European Transport Company
FCA	Forward Collision Avoidance
GIS	Geographic Information System
GLM	General Linear Model
GODB	Group Op De Beeck
GPS	Global Positioning System
HCPC	Hierarchical Clustering on Principal Components
i-DREAMS	smart Driver and Road Environment Assessment and Monitoring System
IMOB	Institute for Mobility
KDE	Kernel Density Estimation
MANOVA	Multivariate Analysis of Variance
NASA	National Aeronautics and Space Administration
NB	Negative Binomial
NKDE	Network Kernel Density Estimation
PAI	Prediction Accuracy Index
PCA	Principal Component Analysis
PO	Performance Objectives
POWERS	Prediction of Worldwide Energy Resource
QGIS	Quantum Geographic Information System
RMSE	Root Mean Square Error
SD	Standard Deviation
Sig.	Statistical significance
SO	Safety Outcomes
SPG	Safety Promoting Goals
SSCP	Sum of Squares and Cross Products
Std. Error	Standard Error
STZ	Safety Tolerance Zone
WHO	World Health Organization

x

1 INTRODUCTION

1.1 General

The rapid progressions in transportation automation in the digital age have created new challenges, changing the structure of interaction between vehicle, driver, and driving context and necessitating a better understanding of the human factors influencing driver behavior. Among these three factors, several aspects of driver behavior have been repeatedly shown in the literature to be essential for the safe operation of transportation systems (T. Brijs et al., 2020).

Arumugam and Bhargavi (2019) and Fitzharris et al. (2017) described that fatigue, drowsiness, and attention/distraction are significant road safety concerns. Paredes et al. (2018) stated that reducing stress and increasing psychological wellness transforms daily driving time into a conscious experience. Besides, Melinder (2007) studied that socio-cultural factors (e.g., religion, wealth) impact the values related to road safety. Generally, driving behavior, traffic, land use, and demographics affect road safety (Komackova & Poliak, 2016). The introduction and development of connected technology and the adoption of big data are transforming every industry (Arumugam & Bhargavi, 2019). Road traffic behavior and protection have significantly improved due to modern information technology. The development patterns of the internet, traffic, and knowledge have been unavoidable for future road traffic systems (Qu et al., 2019).

The intelligent transport system enables automation of the processes of collecting context data on-road incidents and processing them in real-time with the goal of dynamic response to changes in the transportation situation (Malygin et al., 2018). The best possible use of these opportunities would enable the European Union (EU) and the rest of the world to promptly resolve emerging problems and navigate new technologies to meet its optimistic road safety targets (T. Brijs et al., 2020). The following section describes an overview of the i-DREAMS project (smart Driver and Road Environment Assessment and Monitoring System), which runs from 2019 to 2022 and is financed by the European Union’s Horizon 2020 research and innovation program.

1.2 Overview of the i-DREAMS project

The i-DREAMS project is a European project in which multiple partners from universities, companies, industries, businesses, and so on across Europe collaborate to set up a framework for the definition, development, testing, and validation of a context-aware safety envelope for driving known as the ‘Safety Tolerance Zone (STZ).’ The conceptual framework in figure 1 incorporates aspects of monitoring, in-vehicle interventions, and post-trip interventions to establish an STZ and provide guidance to the driver. The main concerns of the i-DREAMS modes are car, truck, bus, tram, and train (Kaiser et al., 2020). Truck will be the subject of this study.

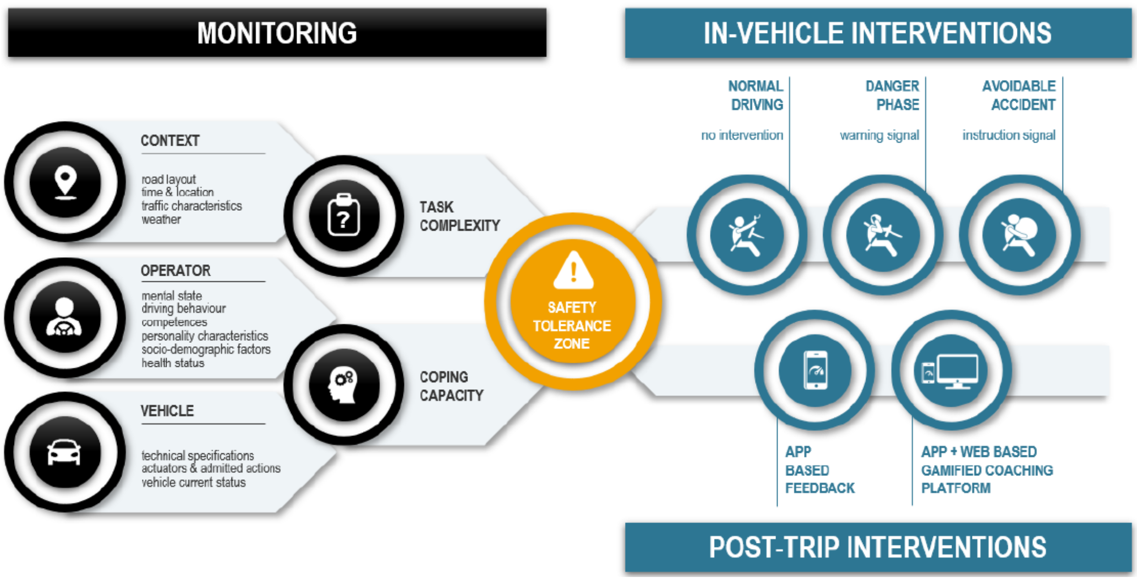


FIGURE 1 Conceptual framework of the i-DREAMS platform with monitoring (pillar I-left) and interventions (pillar II-right) (Kaiser et al., 2020).

The i-DREAMS platform, on the one hand, monitors information from the driving context, driver, and vehicle, to make an estimation of task complexity, and coping capacity. On the other hand, it analyzes and interprets that data in real-time and assists the driver during the trip by offering warnings or alerts about potentially dangerous events that might occur on the road, as well as coaching the driver after the trip using a smartphone app. This means an app is installed on the driver's smartphone and, in this study, the truck drivers. Each driver can access their driving behavior records, ratings, past trips, and other information and receive tips and feedback on improving their driving behavior (Katrakazas, Michelaraki, Yannis, Kaiser, Brijs, et al., 2020).

In addition to the app, the i-DREAMS project also has a web-based platform, usually used by a coach in a transportation company where truck drivers are employees or members. The company coach (driver coach) follows up on the behavior of a group of drivers using the web-based platform (Katrakazas, Michelaraki, Yannis, Kaiser, Brijs, et al., 2020).

1.2.1 Data collection tools

The core point of the i-DREAMS project is to set up a framework for STZ. One of the basic aspects which play a great role in assuring the practical implementation of STZ is monitoring the three components, the driving context (environment), driver, and vehicle using in-vehicle technology. The following topic discusses the selected technologies for measurements in the i-DREAMS project and their description.

1.2.1.1 Technologies for vehicle and driver state monitoring

The equipment for the driver, vehicle, and environment monitoring is selected by considering different factors such as accuracy, validity, suitability for trucks, usability, and user acceptability (Katrakazas, Michelaraki, Yannis, Kaiser, Senitschnig, et al., 2020). The selected technology is mounted in the truck to gather data on the driver's state, such as fatigue and distraction (hand-held mobile phone use while driving), and details about driving context like reading speed signs, weather conditions, and headway distance. It also collects information about the vehicle related to fuel consumption, when it started to operate, and its duration (Katrakazas, Michelaraki, Yannis, Kaiser, Senitschnig, et al., 2020). The following are the selected technologies used to measure vehicle and driver state (driver capacity) in the context of trucks.

a. CardioWheel

The technology in the truck contains a steering wheel cover mounted over the truck's existing steering wheel. This uses an electrocardiogram (ECG) to detect the driver's sleepiness or fatigue when both hands are in contact with the truck's steering wheel. It also enables to extract the driver's ID and know if the driver is using both hands on the steering wheel (Katrakazas, Michelaraki, Yannis, Kaiser, Senitschnig, et al., 2020).

b. Gateway

The Gateway, a kind of device in the dashboard, is another hardware mounted in the vehicle. The Gateway has an accelerometer and gyroscope inside and also gathers data from other i-DREAMS devices (e.g., CardioWheel, Mobileye, and Dashcam). The Gateway uses the collected data to calculate vehicle trajectory and speed, namely Global Positioning System (GPS), trip start and duration, time of the day, brake usage, acceleration, and deceleration. It also monitors the status of the windscreen wiper (either on or off), the time when the vehicle's ignition is turned on or off, as well as the frequency and severity of overspeeding. The Gateway also has a central computer that measures the driver's position in the STZ, triggers relevant in-vehicle interventions in real-time, and uploads data for post-processing and post-trip interventions to the i-DREAMS cloud platform via Wi-Fi or 3G/4G (Katrakazas, Michelaraki, Yannis, Kaiser, Senitschnig, et al., 2020).

1.2.1.2 Technologies for environment monitoring

The technology monitoring the environment (driver context) focuses on task demand. Like the driver and vehicle state monitoring, the technologies used to monitor the driver context are selected based on accuracy, validity, suitability, usability, and acceptability (Katrakazas, Michelaraki, Yannis, Kaiser, Senitschnig, et al., 2020).

a. Mobileye

Two cameras are mounted on the vehicle. Mobileye is one of the cameras which looks outside or in front of the truck for speed limits, time headways and time-to-collision, pedestrians and bikes in front of the driver, lane departure, non-overtaking zones, rain conditions depending on windscreen wipers, and poor visibility. However, Mobileye does not store videos; it only detects and interprets information (Katrakazas, Michelaraki, Yannis, Kaiser, Senitschnig, et al., 2020).

b. Dashcam

The second camera installed in the truck's dashboard is called Dashcam. This camera only stores road scene videos when Mobileye warnings are produced or when severe incidents are observed (Katrakazas, Michelaraki, Yannis, Kaiser, Senitschnig, et al., 2020).

1.2.2 Technologies utilized for intervention

This topic gives an overview of the i-DREAMS project technologies for real-time feedback and post-trip interventions. After comparing and contrasting the available instruments and assessing their adoption and effectiveness in the i-DREAMS mode truck, the following technologies are selected.

1.2.2.1 Technologies utilized for real-time intervention

The effectiveness of how in-vehicle alerts are configured and offered to vehicle drivers is critical to the successful performance of real-time interventions. Display, timing, and information are the three fundamental criteria (specific design features) that decide the usefulness of in-vehicle messaging. The display should draw the driver's attention, the timing should be correct, and the details given should be meaningful. After weighing the feasibility and acceptability of various available solutions, the nomadic device is chosen as the most convenient solution for real-time intervention. Both visual and auditory interventions are available for this unit, and these in-vehicle real-time alerts are presented on a separate display mounted on the vehicle's dashboard or windshield. The display shows the visual alerts while a speaker in the Gateway triggers the audio signals associated with alerts (K. Brijs et al., 2020).

Based on the data collected using the technology installed in the truck, the algorithm's outcome may be that no danger is detected, in which case the system does not have a reason to alert the driver. When danger is detected, the system should warn the driver before the threat becomes larger and larger, resulting in a crash. Depending on the severity of the risk, the type of alert can be either an early or late warning. For more intense or extreme risks, the driver is alerted visually and using an auditory signal with a warning sound that alerts the driver to take action, unless the driver can end up in a collision (K. Brijs et al., 2020). Figure 2 shows the recommended stage of the warning system for real-time intervention depending on the driver's situation.

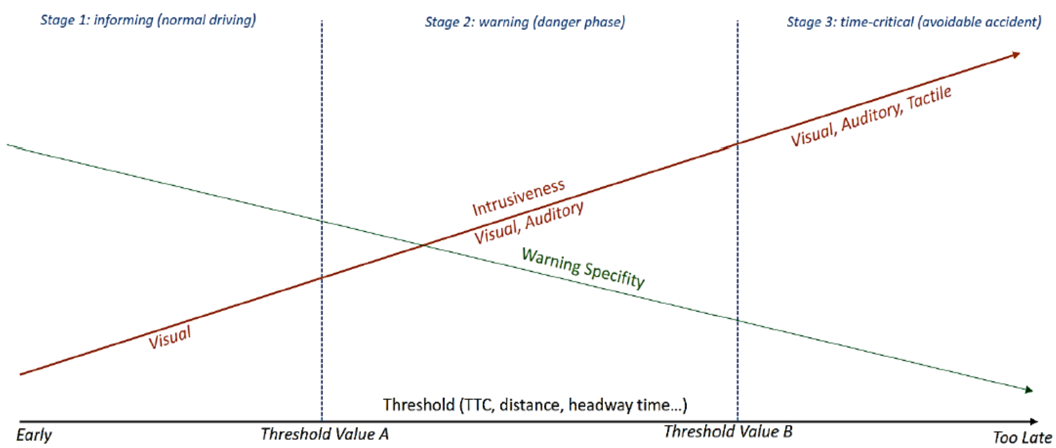


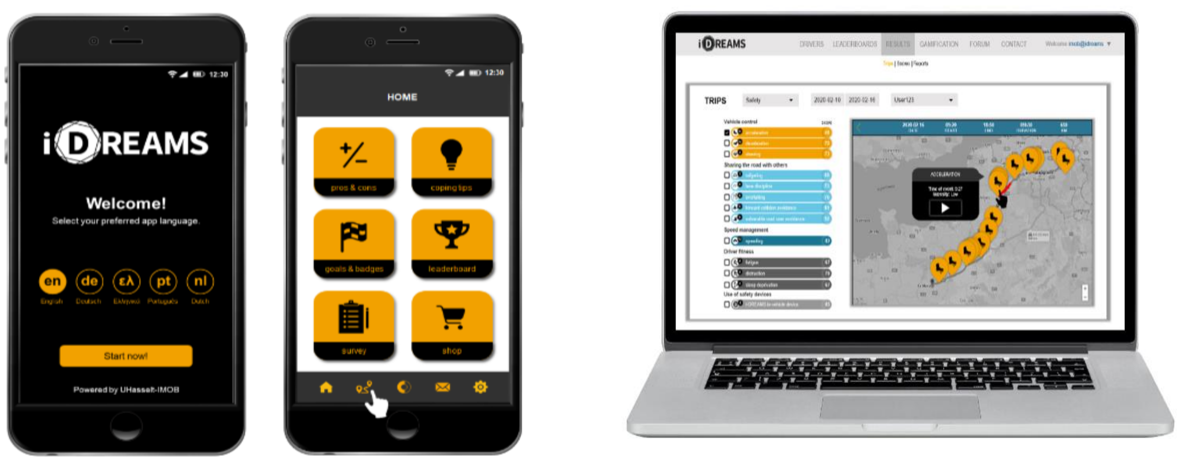
FIGURE 2 Illustration of the warning stages (Katrakazas, Michelaraki, Yannis, Kaiser, Brijs, et al., 2020).

1.2.2.2 Technologies utilized for post-trip intervention

The technologies, applications, and schemes used in the post-trip interventions of the i-DREAMS project are selected based on the capability of changing driver behavior and enhancing knowledge, attitudes, perception, and, eventually, safety performance. The intervention tools used for the post-trip interventions are a smartphone app (for the driver) and a web-based platform (for the driver mentor/coach). The smartphone app provides information, guidance, and notifications to drivers, while the web-based platform offers information about the behavior of a group of drivers to the driver coach (company coach) (K. Brijs et al., 2020).

Smartphone app and web-based platform

These two instruments are used to show digital road map data. The truck drivers and coaching company get all the information at the end of the day, an overview of annotated trip data with geolocated risk-related events, and recorded road video data on the i-DREAMS web platform and smartphone app. The smartphone app offers trip details such as trip duration and length, speeding, distraction, harsh brakes, accelerations, steering, and driving at dangerous hours for truck drivers. The truck drivers get feedback based on the collected data and their scores. On the other hand, the web-based website offers reports on the truck driver's driving actions for driver coaching (the transport company in which the truck drivers are employees) (Katrakazas, Michelaraki, Yannis, Kaiser, Brijs, et al., 2020). Figure 3 shows the two instruments of i-DREAMS technologies for post-trip intervention.



(a) (b)

FIGURE 3 i-DREAMS technologies for post-trip intervention (a) smartphone app (b) web-based platform (i-DREAMS project, 2020).

1.2.3 Targeted risky events for safety interventions

The indicators or warnings of risky driving behaviors are aimed at multiple risk factors that could be effectively tracked while mitigating the chance of an accident or injury. However, there is a difference between targeted variables since the i-DREAMS project has both real-time and post-trip interventions. The real-time intervention focuses on variables that can be effectively tracked in real-time, and the feedback helps to reduce the likelihood of a crash occurring. In-vehicle interventions are usually targeted towards a mental state such as fatigue, drowsiness, attention or distraction, stress, emotions, or the driver's overall workload.

Also, physiological measurements such as heart rate variability, skin conductance, skin temperature, breathing rate, or electroencephalogram signals are used in in-vehicle interventions. This implies the driver's state (e.g., fatigue, distraction) is taken into account to determine the necessity of a warning and when this warning should be given. For example, when the driver is driving in bad weather, and he/she is sleepy or distracted, a tailgating warning will be triggered more rapidly than when the driver is not distracted, and the weather is good. In other words, the warnings for different risks are implemented such that they are adaptive and take into account all the available information about the driver, the vehicle, and the environment. Besides, immediate action is needed in events like lane discipline, the location of reckless incidents, and maintaining a reasonable distance from the vehicle ahead.

On the other hand, the post-trip intervention helps enhance overall driving behavior. As a result, post-trip feedback is typically focused on the frequency of severe events (acceleration, braking, or cornering), distracted driving, or other reckless events during peak hours or dangerous night hours, while eco-driving techniques are also given special attention. The i-DREAMS targeted risky events for the transport mode truck are sleepiness, hand-held mobile phone use while driving, acceleration, deceleration, steering, tailgating, lane departure, overtaking, potential collision events, speeding, and checking the score after the trip using the smartphone app (Katrakazas, Michelaraki, Yannis, Kaiser, Brijs, et al., 2020).

1.3 Problem statement

Road transport is essential for exchanging goods and people from one place to another or between different countries. Regrettably, road transportation also has many significant adverse effects. Crash-related fatalities and injuries are the particular consequences of road transport (Barnard et al., 2016). Also, traffic-related deaths reached 1.35 million in 2016 (World Health Organization [WHO], 2018). Besides, according to the European Commission (2018) findings, traffic accidents across the EU killed about 25.600 people and injured over 1.4 million people in 2016.

Road traffic accidents are a serious threat to human life (Qu et al., 2019). The most critical determinants which affect road safety are the human factor, vehicle, and environment (Arumugam & Bhargavi, 2019; Komackova & Poliak, 2016). The human factor is responsible for most of the negative impact on road safety (Arumugam & Bhargavi, 2019; Barnard et al., 2016). The human factor can be physiological or behavioral. Physiological mistakes are happening due to driver fatigue and drowsiness. Behavioral errors include distracted driving, drunk driving, aggressive driving, road rage, hard acceleration, hard braking and cornering, and speeding (Arumugam & Bhargavi, 2019).

As a result, a thorough understanding of road user behavior is needed to identify the root causes of the negative effects and recommend viable strategies for mitigating them. For instance, how do road users behave in various situations, how and where do usual traffic patterns or ordinary behavior turn into critical incidents or crashes, and what factors influence the driving style and related issues (Barnard et al., 2016). Anderson (2009) stated a more thorough understanding of indications of causal effects could be attained by identifying road accident hotspots and appending value-added data. This study also recommended that more research into road accident analysis in creating a relationship between driving risk and geospatial features is needed. Besides, Mohaymany et al. (2013) suggested that it is essential to integrate high-risk events and spatial properties to understand traffic collisions better and improve road safety.

Thus, this study developed a characterization of risky events based on time of day and weather conditions to identify dangerous driving scenes that describe when and where drivers are most likely to make mistakes and determine the correlation between the risky events. This study also integrated risky events with geospatial features, particularly road type.

1.4 Research question

Studies have been conducted to understand better the risk of driving from several perspectives, such as road layout, time of day, and weather conditions. Gitelman et al. (2018), Hassan et al. (2016), and Yuan et al. (2021) studied the occurrence of risky events on various road layouts. Besides, different studies looked at the impact of time of the day on risky events such as speeding (Tseng et al., 2016), fatigue (Anund et al., 2017), and drowsy driving (McCartt et al., 2000). Moreover, Chen and Zhang (2016) mentioned that weather conditions significantly affect the frequency of risk events. The following basic research questions were used to address the problem statement mentioned:

a. What impact does the time of day have on risky events?
   e.g., Are the risky events more likely to happen at night or during the day? (e.g., speeding)
b. What happens to the risky events when the road type changes?
   e.g., Are the risky events more likely to happen on motorways or primary roads? (e.g., acceleration)
c. What is the impact of weather conditions on risky events?
   e.g., Are the risky events more likely to happen in adverse or clear weather conditions? (e.g., distraction)
d. What is the correlation between the risky events?
   e.g., Is there a link between speeding and following close behavior (tailgating), or is there a link between headway alerts and forward collision alerts?

1.5 Objectives of the study

1.5.1 General Objective

The study's general objective was to identify and characterize risky events in i-DREAMS for trucks.

1.5.2 Specific objectives

The study aimed to accomplish the following specific objectives:

- To Identify and characterize risky events based on the time of day.
- To Identify and characterize risky events based on road type.
- To Identify and characterize risky events based on weather conditions.
- Determining the correlation between the risky events.

2 LITERATURE REVIEW

2.1 Review of Related Works

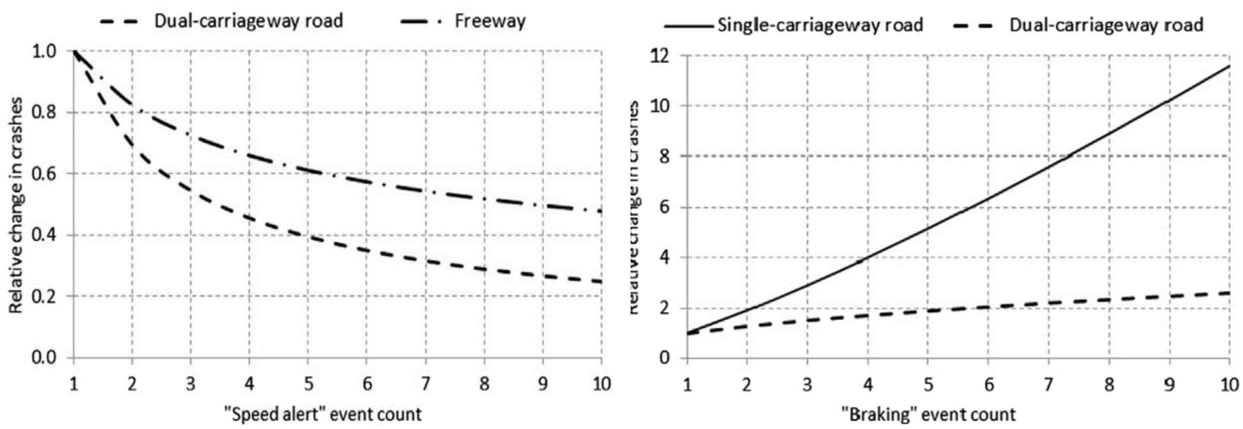
Studies on truck-involved crashes have gotten much attention worldwide (Brodie et al., 2009; Gates et al., 2013; Häkkänen & Summala, 2001; Lombardi et al., 2017). Many factors, including roadway geometric features, traffic circumstances, environmental conditions, driver traits, and vehicle characteristics, affect truck-involved crashes, particularly crash frequency and severity (Chang & Chien, 2013; Dong et al., 2015; Lemp et al., 2011; Yang et al., 2019; Zhu & Srinivasan, 2011). Risk factors, including driving behaviors, road locations, weather conditions, accident types, and the number of vehicles involved, were highly connected to crash severity (Khorashadi et al., 2005; Xie et al., 2012; Zhu & Srinivasan, 2011).

In addition, other studies have attempted to quantify the risk of driving under various conditions based on various criteria, including road layout, time of day, environmental conditions, and traffic flow. In this regard, previous studies have revealed that crashes that happened at different time periods, region types, and route segments have varied characteristics regarding roadway and environmental-related risk variables. Khorashadi et al. (2005), for example, looked at the severity of truck driver injuries in urban and rural areas. Uddin and Huynh (2017) investigated the severity of crash injuries in rural and urban regions under various lighting conditions. The results demonstrated significant risk variables changes under varied lighting conditions and area types combinations. In different time conditions, Dong et al. (2014), Osman et al. (2016), (Pahukula et al., 2015), and Zou et al. (2017) studied the risk factors leading to crash frequency and severity at intersections and work zones. They discovered that intersection facilities, traffic flows at different times, and lighting conditions in work zones were all linked to the severity of crash injuries.

2.1.1 Road layout effects

Previous research has explored various safety issues related to trucks based on the impact of the road layout. For instance, Dong et al. (2014) analyzed the effect of lane number, lane width, and intersection on truck accidents. Chang and Chien (2013) studied the impact of the tunnel, horizontal curve, grade section, straight section, and others (e.g., interchange) on driver injury severity. Yuan et al. (2021) looked into the impact of road-related variables (roadway alignment, roadway grade, surface condition, surface type, and speed limit) and the number of lanes on different groups of truck drivers. Besides, Simon et al. (2009) found that intersections account for 21% of fatalities and 43% of injuries in accidents involving at least one passenger car in Europe. Despite being explicitly targeted, intersection accident mechanisms warrant further examination.

Based on a study by Gitelman et al. (2018) on the relationship between crashes, infrastructure characteristics, and events, different event types have varied relationships with infrastructure characteristics and varying effects on crashes. At least for some event types, infrastructure characteristics such as indicators of junction proximity, length of road sections, change in road width, road type, right and left shoulder width, lane width, horizontal radius, and vertical grade were found to have a significant impact. According to the study's multivariable models, road type, junction proximity, section length, and road shoulder widths were the most influential road factors for the driving events studied. However, compared to speed events, a large portion of the infrastructure attributes had the opposite influence on braking events. Junction proximity or geometry constraints typically increase braking events and decrease speed events. Contrarily, improved road conditions and driving on longer portions without at-grade junctions are linked to fewer braking events and more speed alert events. Due to these events' opposite nature, such impact disparities appear plausible. Figure 1 also depicts the relationships between event counts and relative changes in injury crashes for road types where the two variables were determined to have a significant link. Hence, increased braking occurrences are linked to a slight rise in crashes on dual-carriageway roads and a significant increase on single-carriageway roads, as shown in figure 4 (a). On the other hand, an increase in speed events is linked to a slight reduction in freeway crashes and a greater drop on dual-carriageway roads, as shown in figure 4 (b).



(a) (b)

FIGURE 4 Relative change in injury crashes associated with an increase in the event counts: (a) braking events, (b) speed alert events (Gitelman et al., 2018).

In addition, numerous studies have looked at the effect of road layout on the occurrence of risky events. According to Abdel-Aty et al. (2011), head-on and rear-end crashes are more frequent on high-speed roads, roads with no sidewalks, undivided roads, and two-lane rural roads. Fatigue on expressways and long steep grade roads (Chen & Zhang, 2016), overtaking and speeding on two-lane motorway networks (Nagy & Sandor, 2012), and rear-end collisions on freeways (Zhao & Lee, 2018) are also among the risky events studied. Moreover, table 1 summarizes research findings for road layout features linked to different risky events in various studies.

TABLE 1 Linkage of risky events and road layout

References	Road layout	Risky events
Dong et al. (2014)	Increasing lane width both on major and minor roads	Rear end or rear underride collision
	Decreasing lane width	Improved lane keeping, more accurate steering behavior, and reduction in driving speed
	A higher number of a left turn lane at intersection	Sideswipe collision
	Skewed-angle crossroads	lower acceleration rate and higher lane departure
Dong et al. (2015)	Highway	Higher speed
Gitelman et al. (2018)	Junction proximity or geometric constraints	higher braking event and lower speed alert event
	Improved road condition and driving on longer sections without at-grade junctions	Lower braking event and higher speed alert event
Häkkänen and Summala (2001)	Two-lane highway	Head-on collision
Harb et al. (2007)	Unsignalized intersection	Rear-end collision
Jansen and Wesseling (2018)	Urban areas	Harsh braking event
Ketabi et al. (2011)	Two-way highway	Overtaking
Khorashadi et al. (2005)	Urban settings	Side-impact collision
Mahmud et al. (2021)	Two-lane two-way highway	Overtaking
Pahukula et al. (2015)	Large median	Speeding
	Wide shoulder width	Lane discipline
Pokorny et al. (2017)	Urban areas, particularly intersections	Truck bicycle accident (vulnerable user collision)
Thiffault and Bergeron (2003)	Monotonous and low demanding road environment (for a long time on the straight and smooth road)	Fatigue, large lane deviation, and large steering
Ting et al. (2008)	Straight, uneventful, and long roads	Fatigue
Uddin and Huynh (2017)	A higher number of lane	Increase lane changing
	Freeway	Rear-end crash risk
Xie et al. (2012)	Interstate highways	Higher average speed
Yuan et al. (2021)	Lane planning (lane width)	Failure to drive in the proper lane
Zellmer (2013)	Interstate and highways	Tailgating

2.1.2 Traffic condition Effect

To date, several research studies have looked into several road safety issues related to the impact of traffic conditions/environment on the frequency and severity of truck-involved crashes (Dong et al., 2014; Dong et al., 2015; Khorashadi et al., 2005; Uddin & Huynh, 2017; Zou et al., 2017). Dong et al. (2014) looked into the impact of traffic conditions in terms of traffic volume, namely the number of vehicles involved in an accident. The findings from the study suggest that traffic volume has a favorable impact on the frequency of all types of crashes. According to this study, the frequency of truck-involved and car-truck-involved crashes rises as the truck percentage rises. Also, the study found that truck-car collisions involve more trucks than truck–truck and single-truck collisions.

Similarly, Dong et al. (2015) also found traffic conditions contributed to the severity of truck-involved crashes. According to this study, truck-involved crashes on roads with smaller traffic volumes had a significantly higher fatality risk. This result was expected because a reduced traffic flow means a higher possibility of higher speeds, which contributes positively to the severity of the incident. Besides, the number of vehicles involved in a crash, vehicle occupancy, inside city limit, and traffic control was found to significantly affect injury severity in accidents involving large trucks (Khorashadi et al., 2005). Also, it was revealed that when a traffic sign/signal or stop sign is present at the crash location, the crash is more severe (more than 50 percent of the crashes occurred when the traffic signals were on). Aggressive driving at traffic signals (e.g., tendency to speed up when the traffic light turns red) and complexity of traffic signs/signals location (e.g., intersections) have been noted as contributing factors (Zou et al., 2017). According to Uddin and Huynh (2017), traffic volume was adversely associated with injury severity (major and possible/no injury) in rural and urban locations, with daylight and nighttime being significant for urban areas but just daylight conditions in rural ones. One probable explanation is that drivers become more cautious during periods of heavy traffic flow and at night.

Generally, the previous studies have attempted to characterize traffic conditions in terms of traffic volume and speed limit (Chan, 2017; Chang & Chien, 2013; Dong et al., 2014; Dong et al., 2015; Golob et al., 2008; Guo et al., 2010; Khorashadi et al., 2005; Shi et al., 2016; Uddin & Huynh, 2017; Wang et al., 2013; Zheng, 2012) and traffic controls (Dong et al., 2015; Khorashadi et al., 2005; Zou et al., 2017). Whether to investigate crash frequency or severity, most studies defined traffic conditions as nearly similar, with some variables in common and others distinct (see table 2).

TABLE 2 Parameters used to define traffic condition

Reference	Parameters
Dong et al. (2014)	AADT and speed limit
Dong et al. (2015)	AADT, speed limit, and traffic controls
Golob et al. (2008)	Volume level (speed and density), number of vehicles and other parties involved, and movement of vehicles before a collision
Guo et al. (2010)	ADT, signal coordination along a corridor, and speed limit
Khorashadi et al. (2005)	Number of vehicles involved in the crash, vehicle occupancy, inside city limit, and traffic controls
Shi et al. (2016)	Volume per lane, average speed, and average occupancy
Uddin and Huynh (2017)	AADT and speed limit
Wang et al. (2013)	Total delay, AADT, average vehicle speed, and speed limit
Zou et al. (2017)	Traffic volume (high occupancy vehicles) and traffic controls (traffic signal, yield sign, none, and other)

Aside from the impact of traffic conditions on the frequency and severity of truck-involved crashes, research has been conducted to link traffic conditions to several risky events. Dong et al. (2014) found that the car crash rate reduces as the percentage of trucks increases. One possible explanation is that, as the number of trucks rises, the frequency of lane changes and car overtaking reduces for a given vehicle density. On the other hand, the frequency of truck-involved crashes and car–truck-involved crashes rises as the truck percentage rises. The reason could be that the likelihood of colliding with a truck increases when the fraction of trucks in a traffic stream of a given volume increases. Moreover, table 3 lists different risky events linked with different traffic condition characteristics.

TABLE 3 Linkage of risky events and traffic conditions

References	Traffic condition characteristics	Risky events
Ahmed and Ghasemzadeh (2018)	Free-flow speed conditions	Speeding and longer average headway time
Dong et al. (2014)	Percentage increase in trucks	Reduces lane changing and overtaking by cars
Dong et al. (2015)	Lower traffic volume	Higher speed
Golob and Recker (2003)	Heavily congested, stop-and-go traffic	Rear-end collision
	Stable traffic (low volume and high steady speeds)	Weaving collision
Osman et al. (2016)	Lower traffic volume	Higher speed
Tarko et al. (2011)	High traffic volume	Increase lane changing and cutting in events Lowers speeding and tailgating events
Thiffault and Bergeron (2003)	Low-traffic loads	Fatigue
Wu and Thor (2015)	Volatile traffic flow	Sudden speed change
Yang et al. (2018)	High traffic density	Increases lane change frequency, high instantaneous acceleration, high overtaking frequency, and lower overtaking headway
Zhao and Lee (2018)	Heavy vehicle following a car	Lower average speed but higher risk of braking
	Abrupt speed change	Higher rear-end collision
Zou et al. (2017)	Traffic light turning red	speeding

2.1.3 Weather effect

Weather has been found to have a substantial impact on road safety. In most cases, adverse weather worsens dangerous driving conditions and considerably increases the number of collisions. The detrimental effect of adverse weather (fog, rain, snow, ground blizzards, slush, and strong wind) on driving behavior, visibility, pavement conditions, and driving performance has increased the collision rate (Ahmed & Ghasemzadeh, 2017; Peterson et al., 2008). For instance, Hermans et al. (2006) investigated the impact of 17 climatic factors on the hourly number of injuries. Out of the 17 climatic factors, the presence of precipitation has the highest effect. Besides, a study by Qiu and Nixon (2008) reviewed and analyzed research findings (34 papers and 78 recordings) to examine the interaction between weather and road safety. According to the results, crash rates often rise during precipitation, with snow having a higher impact on crash incidence than rain: snow can raise crash rates by 84% and injury rates by 75%. In addition, as indicated in table 4, several previous studies have found that crash rates increased significantly in various adverse weather conditions.

TABLE 4 Research findings of previous studies regarding the impact of adverse weather conditions on traffic safety

Reference	Research findings
Ahmed et al. (2018)	During inclement weather (severe wind or snow), truck-involved crashes occur 19% more frequently than when there are no truck-involved crashes. Truck-involved crashes happened more frequently when the road conditions were not dry, such as when there was ice or frost.
Andrey et al. (2003)	Rainy or snowy weather raised crash frequency by 75%. Average injured persons involved in crashes increased by 45% when it rains or snows. Snowfall had a greater impact on the incidence of crashes than rainfall.
Dong et al. (2015)	Inclement weather, such as fog or windy weather, increases the risk of a severe injury.
Hermans et al. (2006)	Out of 17 climatic factors, precipitation had the highest effect on the number of injuries
Keay and Simmonds (2005)	The probability of a crash was 0.7 times higher in rainy weather than in dry weather. Under wet conditions, the duration since the last rainfall increased the crash risk.
Khattak and Knapp (2001)	Injury and non-injury crash rates increased by 11 and 21 times during snow events, respectively. Snow weather conditions decreased the risk of injury crash compared to non-injury crash conditional on crash occurrence.
Offei and Young (2014)	Snowy and icy road conditions caused 32.4% and 28.6% of the crashes, respectively. Dry roads caused about 30.3% of the accidents. Clear weather caused 50% of the accidents. Snowy weather conditions contributed to around 20% of the crashes, while strong winds, fog, rain, dust, and hail contributed the remaining percentage.
Qiu and Nixon (2008)	Rainy weather increases crash rates by 71% and injury crashes by 49%. Snowy weather raises crash rates by 84% and injury crashes by 75%.

Concerning truck-involved risky events, Chen and Zhang (2016) investigated that weather conditions play a significant role in the frequency of risky events, particularly fatigue-related crashes. This study categorized weather conditions based on visibility (good and bad) and adversity (fine and adverse). The result suggests that adverse weather and poor visibility are highly associated with fatigue-induced crashes involving trucks, mainly because trucks must brake over longer distances on slippery roads. Likewise, Chipman and Jin (2009) found that poor driving circumstances, such as hot, rainy, or foggy weather, as well as noise, can all influence a driver's focus, making fatigue more likely. Table 5 shows an overview of reviewed linkages between different adverse weather conditions and risky events.

TABLE 5 Linkage of risky events and weather

Reference	Weather condition	Risky event
Abdel-Aty et al. (2011)	Visibility obstruction due to fog and smoke	Head-on and rear-end collision
Ahmed and Ghasemzadeh (2018)	Rainy weather condition	Longer average headway time and reduced speed
Chen and Chen (2011)	Snow or slush road surface	Difficulty in vehicle control
Chen and Zhang (2016)	Adverse weather conditions such as wet pavement and decreased visibility	Fatigue
Das et al. (2019)	Foggy weather condition	Lower lane-keeping ability
Golob and Recker (2003)	Wet roads	Lane-change maneuvers
	Dry roads	Rear-end collision
Kilpeläinen and Summala (2007)	Snow	Decreases traffic flow speed
Pahukula et al. (2015)	Clear weather condition	Higher speeding
Peng et al. (2017)	Reduced visibility due to fog	Reduced headway and higher speeding
Tarko et al. (2011)	Adverse weather conditions (rain and snow)	Decreases speeding and tailgating
Zheng et al. (2018)	Fog and severe crosswinds	Trucks hard to control
	Icy road surface	Lower speed

2.1.4 Time of day effect

Several studies looked at the impact of the time of day on crash severity and frequency. For instance, Pahukula et al. (2015) researched that clear weather nights and nights with no illumination result in no injury or crash with severe consequences. Also, Zheng et al. (2018) found that the time of day is one of the contributing factors impacting the severity of truck crashes. According to this study, the most dangerous time is early morning (3 a.m. - 6 a.m.), when single and multiple fatality crashes are more likely to occur. Also, fatal crashes are more likely on weekends, but non-fatal crashes are more likely on Fridays.

In contrast, according to Wang and Prato (2019), model findings on time of day perspective revealed that only the night time, namely between midnight and 6 a.m., was related to moderate increases in injury (1.5%) and fatality (3.3%) probabilities. There was no discernible change in injury or fatality probability between weekdays, weekends, or holidays. Table 6 also summarized related research findings concerning the effect of time of day on the occurrence of truck accidents.

TABLE 6 Research findings on the effect of time of the day on truck crash occurrence

Reference	Time of the day
Brodie et al. (2009)	Crushes were common on weekdays, particularly Friday, Monday, and Wednesday. Crashes occurred most frequently between 10:00 a.m. and 12 noon and midnight and 2:00 a.m., respectively.
Islam and Hernandez (2013)	Approximately 76.2% of crashes were in the dark (poor street lighting).
Khorashadi et al. (2005)	Morning (5:31–8:00 a.m.) accidents in urban areas are 37.1% less likely to result in a severe/fatal injury (relative to other periods), compared to a mere 4.3% reduced risk of a severe/fatal injury in rural areas during the same period.
Lombardi et al. (2017)	Peak crush hours for older drivers were 12:00 - 04:00 p.m., with 79.8% of crashes happening during daylight hours, and peak crash days were generally consistent from Tuesday to Friday. Peak crash hours for younger drivers were 03:00 - 07:00 p.m., with 58.4 percent happening during daylight hours, while peak crash days were Friday and Saturday.
Offei and Young (2014)	The majority of the accidents (64%) occurred during daylight hours. Around 18% of accidents occurred at dawn or dusk, with darkness and unlighted conditions accounting for 15% of all accidents.
Osman et al. (2016)	Daytime crashes (6:00 a.m. - 6:00 p.m.) are associated with higher severe results in the event of a crash. Compared to no injury, traveling during peak hours was found to be related to a lower likelihood of injury.
Pokorny et al. (2017)	In urban regions, 35%, 28 %, 35%, and 2% of truck bicycle accidents occurred, but in rural areas, 0%, 14%, 66%, and 10% occurred. Morning (6 a.m. - 10 a.m.), midday (10 a.m. - 3 p.m.), late afternoon (3 p.m. - 9 p.m.), and night (9 p.m. - 12 a.m.) are the time of day used to describe the percentage distribution of truck bicycle accidents for both urban and rural areas, consequently.
Wang and Prato (2019)	Midnight and 6 a.m. were related to moderate increases in injury (1.5%) and fatality (3.3%) probabilities. No visible change in injury or fatality probability between weekdays and weekends or holidays was found.
Zheng et al. (2018)	The most dangerous time was early morning (3 a.m. - 6 a.m.), when single and multiple fatality crashes were more likely to occur. Fatal crashes were more likely on weekends, but non-fatal crashes were more likely on Fridays
Zhu and Srinivasan (2011)	Crashes occurred under dark but lighted conditions (7:30 p.m. to 5:30 a.m.) relative to day-light and dark periods. Weekday crashes are likely to be less severe than weekend crashes.

Most studies looked at the impact of time of day on the frequency of crashes or the severity of crash injuries. Only a few researchers have attempted to determine the effect of the time of day on the occurrence of risky events. Pahukula et al. (2015) found that between 10:00 a.m. and 3:00 p.m., free-flowing characteristics like speeding and changing lanes contributed to large truck-involved crashes. Also, according to Chen and Zhang (2016), fatigue-related crashes are more likely to happen overnight and early in the morning (12 a.m. to 6 a.m.), possibly due to the heightened sleepiness and fatigue associated with the human circadian cycle during these times. Table 7 lists different periods associated with different risky events in various studies.

TABLE 7 Linkage of risky events and time of the day

References	Time of the day	Risky events
Abdel-Aty et al. (2011)	Night without street lighting	Head-on and rear-end crashes
Chen and Zhang (2016)	Overnight and early morning hours (12 midnight to 6 a.m.)	Fatigue
Chipman and Jin (2009)	Night	Speeding
Dingus et al. (2006)	Afternoon and evening	Fatigue
Golob and Recker (2003)	Daylight	Rear-end collision
Nagy and Sandor (2012)	Rush hour	Overtaking and speeding
Osman et al. (2016)	Evening time (6 p.m. – 12 a.m.)	Higher speed
Pahukula et al. (2015)	Between 10:00 a.m. and 3:00 p.m.	Speeding and lane changing
Wu et al. (2016)	Night	Speeding, cutting in, acceleration and deceleration

2.1.5 The combined effect of road layout, time of the day, weather, and traffic characteristics

The contribution of road layout, time of the day, weather, and traffic characteristics to risky road events have been studied based on the different combinations. Golob and Recker (2003) evaluated the correlations between traffic accidents on urban freeways and the traffic flow layout, considering weather and lighting circumstances. According to this study, multiple vehicle crashes caused by weaving maneuvers are more likely to occur on wet roads during daylight than on dry or wet roads during darkness. On the other hand, rear-end crashes are more likely to occur during daylight hours on dry roads. Also, (Osman et al., 2016) mentioned the combined effect of traffic volume and time of the day, which found that evening crashes are likely linked with decreased visibility and higher speeds due to lower traffic volume. Table 8 summarizes research findings on any combination of the four parameters (road layout, time of day, weather, and traffic characteristics) associated with different risky events.

TABLE 8 Research findings on the characterization of risky events based on different factors

References	Road layout	Traffic characteristics	Weather	Time of the day	Risky events
Abdel-Aty et al. (2011)	High-speed roads, undivided roads, roads with no sidewalks, and two-lane rural roads	-	-	Night without street lighting	Head-on and rear-end crashes
Ahmed and Ghasemzadeh (2018)	Highway	Free-flow speed conditions	Clear weather condition	-	Speeding and lower headway time Higher average acceleration and deceleration
Chen and Chen (2011)	Rural highway	-	Snow or slush road surface	-	Difficulty in vehicle control
Chen and Zhang (2016)	Expressway and long steep grade	-	Adverse weather condition	12 a.m. - 6 a.m.	Fatigue
Dong et al. (2014)	Urban signalized intersection	Percentage increase in trucks	-	-	Reduced lane changing and overtaking by cars
Golob and Recker (2003)	Urban freeways	-	Wet roads	Daylight	Weaving maneuvers
			Dry roads	Daylight	Rear-end collisions
Kilpeläinen and Summala (2007)	Two-lane main highways outside urban areas	-	Snow	-	Decreases traffic flow speed
Nagy and Sandor (2012)	Two-lane motorway network	-	-	Rush hour	Overtaking and speeding
Osman et al. (2016)	-	Lower traffic volume		Evening	Higher speed
Pahukula et al. (2015)	Urban areas	Lower traffic volume		10 a.m. – 3 p.m. (Mid-day)	Speeding and lane changing

		-	Clear weather conditions	-	Speeding
Tarko et al. (2011)	Urban freeways	High traffic volume	-	-	Increase lane changing and cutting in events Lowers speeding and tailgating events
		-	Rain and snow	-	Decreases speeding and tailgating
Thiffault and Bergeron (2003)	Monotonous and low demanding road environment	Low-traffic loads	-	-	Fatigue
Wu and Thor (2015)	Freeway	Volatile traffic flow	-	-	Distraction
Yang et al. (2018)	Freeway	Higher traffic density	-	-	Increases lane changing, acceleration, overtaking, and decreases over taking headway
Zhao and Lee (2018)	Freeway	Heavy vehicle following car	-	-	Lower average speed, but higher risk of braking
		Abrupt speed change	-	-	Rear-end collision risk
Zou et al. (2017)	Urban areas	Traffic light turning red	-	-	Speeding

2.1.6 Relationship between the risky events

Researchers have linked many factors such as road layout, weather, time of day, and traffic characteristics to risky events, but few have looked at the links between the risky events. According to Chen and Zhang (2016), compared to non-speeding, truck drivers who exhibited speeding behaviors were responsible for 71.07 and 79.98% of all fatigue-related truck crashes in Jiangxi and Shaanxi, respectively. This study also found that, in contrast to safe headway, driving at unsafe headway was responsible for 31.50, and 36.52% of all fatigue-related truck crashes in Jiangxi and Shaanxi, respectively. Wu and Thor (2015) identified an approach for establishing associations between factors contributing to crashes when evaluating crash sequences for rear-end collisions and found that distraction is associated with deceleration in the sequence.

Also, Harb et al. (2007) conducted a correlation test between deceleration rates, response reaction time, speed, gap distance, and gap time while studying the contribution of light-truck vehicles in rear-end crashes. Reaction response time was shown to be negatively correlated with speed (Pearson correlation = -.498, P = .001) and positively correlated with gap distance (Pearson correlation = .497, P < .001) and time headway (Pearson correlation = .583, P= .001). The deceleration rate is positively correlated with speed (Pearson correlation = .366, P= .02) but negatively with gap distance (Pearson correlation = -.604, P < .001) and time headway (Pearson correlation = -.602, P < .001). According to the study, the relationships between reaction response time and deceleration rates with speed, time headway, and gap distance are inherent. Besides, table 9 lists research findings on the correlation between risky events.

TABLE 9 Research findings on the correlation between risky events

References	Correlated risky events	Effect
Chen and Zhang (2016)	Over-speeding and risky following with fatigue	Positive
Golob and Recker (2003)	Rear-end collision with high variations in relatively low speeds	positive
	Weaving collisions with high steady speeds	Positive
Harb et al. (2007)	Reaction response time with speed	Negative
	Reaction response time with gap distance and time headway	Positive
	Deceleration rate with speed	Positive
	Deceleration rate with gap distance and time headway	Negative
Thiffault and Bergeron (2003)	Lane deviation and steering with fatigue	Positive
Ting et al. (2008)	Average headway, extreme steering, edge-line crossing, standard deviation of lateral position with reaction time	Positive
Wu and Thor (2015)	Distraction with deceleration	Positive
Zhou and Zhang (2019)	Over-speeding with fatigue	Positive
	Stable control acceleration and braking with speeding	Positive

2.2 Review of research methods

From a methodological standpoint, Chiou et al. (2017) proposed and compared two methods, namely the clustering approach (that is, k-means) and the multivariate approach, to identify essential attributes affecting crash frequencies at different times of the day so as to offer effective time-specific countermeasures. According to the study, the former uses a crash count model to estimate the total number of crashes and a clustering model to separate segments into clusters based on their crash frequency distribution patterns by time of day. The latter considers crash frequencies at various times of the day as target variables, with the potential correlation between them. The clustering approach performs well in terms of Adjusted Mean Absolute Percentage Error (Adj-MAPE) and Root Mean Square Error (RMSE), where both values are used to evaluate the goodness of fit. The study's clustering approach is divided into two sections. The first stage involves applying widely used Poisson and Negative binomial (NB) models to estimate total crashes on each segment. The second stage involves grouping freeway segments into defined clusters based on their crash frequency distribution patterns throughout the day. Each cluster's average time-of-day crash frequency distribution is then employed to describe the segments that make up that cluster.

To determine the key factors contributing to a traffic accident, a study by Kumar and Toshniwal (2015) proposed a framework based on k-modes clustering and an association rule mining algorithm. The study's explanatory variables, such as accident type, road type, lightning on the road, and road feature, are clustered using the k-modes method. This study used association rule mining to produce rules for each cluster as well as the complete data set, taking strong rules with high lift values for the analysis. In a study conducted by Wu et al. (2016), to make the most of GPS data collected by transportation businesses and explore the potential rules of commercial vehicle driver behavioral characteristics, events related to driving behavioral traits are extracted according to GPS data attributes based on factor analysis, and eight parameters of driving behavioral traits are transformed into a few aggregated variables containing clear information about driving behavior. Using these variables as indicators, hierarchical clustering is used to conduct a cluster analysis of commercial vehicle driver behavioral characteristics in the selected case.

Zhou and Zhang (2019) presented an intriguing technique for analyzing potentially dangerous driving behaviors of commercial truck drivers. The study applied a density based spatial clustering of applications with noise (DBSCAN) to different types of truck drivers who are first identified using principal component analysis (PCA). The DBSCAN approach counts the number of points in a fixed-radius neighborhood to estimate density. If two points are in the same neighborhood, they are linked. To find a cluster, DBSCAN starts with an arbitrary point p and finds all points in the dataset that are density-reachable from p. If p is a core point, a cluster is formed. If p is a cluster boundary point indicating no points density-reachable from p, then DBSCAN uses the same procedure for the next unclassified point. When all points in the dataset have been allocated to a cluster or recognized as noise, the DBSCAN algorithm completes (Kazemi-Beydokhti et al., 2017).

Li et al. (2019) introduced an innovative approach for investigating spatiotemporal distributions of individualized driving accidents and identifying dangerous driving scenes where drivers are more likely to make mistakes using Geographic Information System (GIS) and mobile sensing techniques. The study first detected driving errors using smartphone sensors. Then, geostatistical detected errors with road networks and driving trajectories were analyzed to identify driving error hotspots. Next, a “scenic tuple” was formed to represent operating errors, then individualized dangerous driving scenes were extracted.

The significant similarities among an individual's driving errors are investigated using two clustering methods, namely k-modes clustering and Hierarchical clustering on principal components (HCPC).

Another study, Harirforoush and Bellalite (2019), suggested a two-step integrated approach for finding traffic accident hotspots on a road network. A spatial analysis method named network kernel density estimation (KDE) is used in the first step. The critical crash rate is then used in a network screening process as the second step. The Prediction Accuracy Index (PAI) is applied to test the crash trends of the collected data in the first step. Since crash clustering does not imply that a site is a hotspot, the critical crash rate is then used to establish an accurate comparison.

Similarly, Anderson (2009) applied KDE and k-means clustering to profile road accident hotspots. This study presents a methodology for studying the spatial patterns of injury-related road accidents using GIS and Kernel Density Estimation and a clustering methodology for creating a classification of road accident hotspots using environmental data and results from the first section. Table 10 also highlights the research methodologies employed in previous studies to analyze spatial-temporal data and the clustering approach.

TABLE 10 Review of research methods

Reference	Year of data collected	Methodology for spatial pattern	Method of clustering
Li et al. (2019)	One month	GIS and KDE	k-modes clustering and HCPC
Harirforoush and Bellalite (2019)	Three years	GIS and KDE	KDE-method, PAI, and critical crash rate
Anderson (2009)	Five years	GIS and KDE	k-means clustering
Mohaymany et al. (2013)	Three years	GIS and Network KDE (NKDE)	NKDE
Pljakić et al. (2019)	Three years	GIS and NKDE	NKDE

3 METHODOLOGY

3.1 General

Literature related to the study's objective is assessed and carefully condensed as part of the methodology to bring the notion into the study's purpose. The literature review was conducted with the study's research questions in mind, emphasizing the findings of the studies reviewed and the methods used to analyze related research issues. Following the literature review, a detailed description of the data types used and how they are structured to address the research questions is carried out. Finally, a thorough explanation and description of the data analysis process, including the steps used to organize, categorize and analyze the data acquired, and a brief description of the statistical test utilized, is also part of chapter 3. Besides, an overview of the study's methodology is shown in figure 5 below.

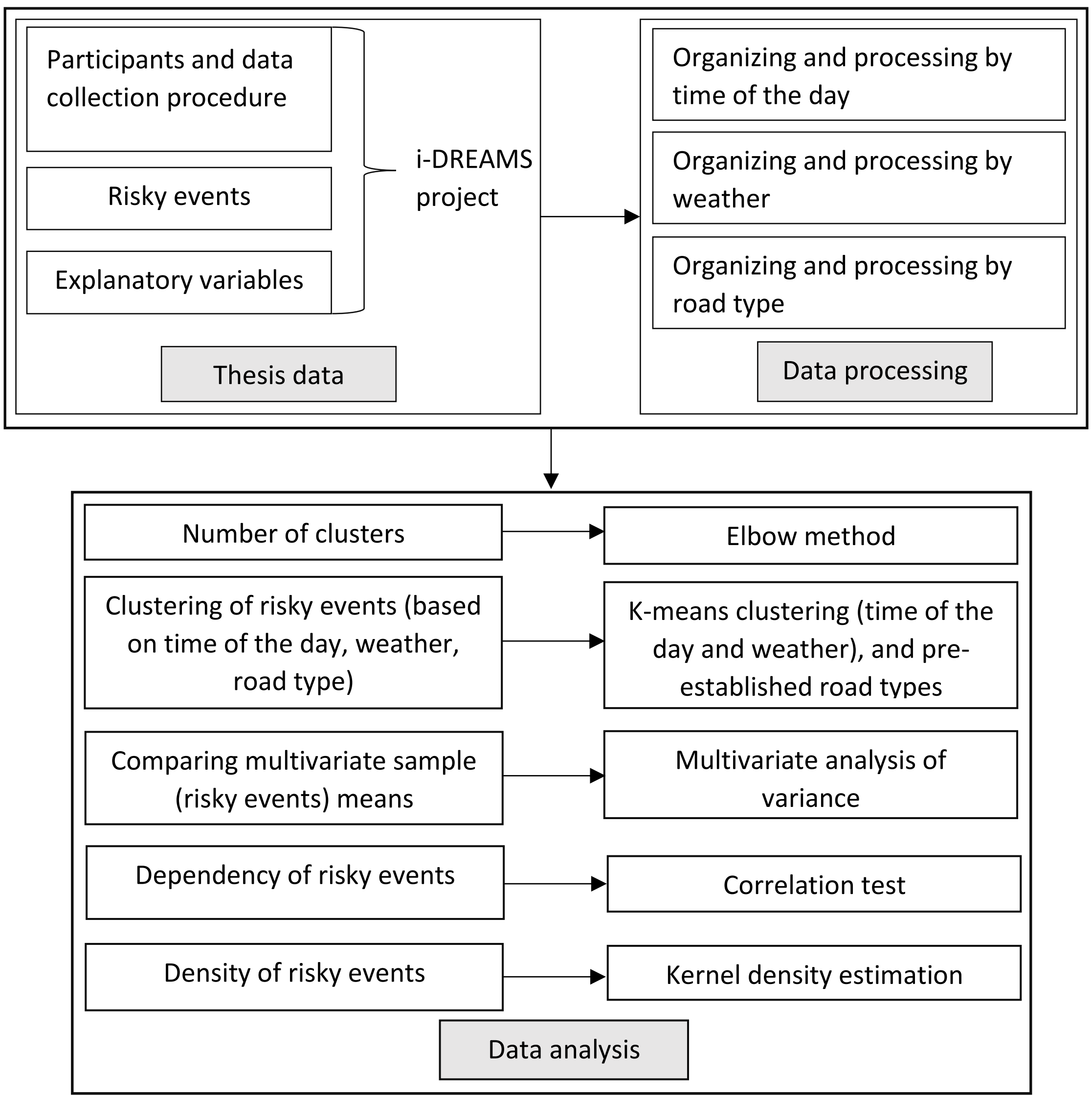


Figure 5 Overview of the study's methodology

3.2 Thesis Data

3.2.1 Participants and data collection procedure

The current thesis used the data collected through a naturalistic driving study of the i-DREAMS project conducted by a consortium of a research team led by Hasselt University Transportation Research Institute (IMOB). The i-DREAMS project developed a new in-vehicle monitoring system to collect continuous naturalistic data (about driver, vehicle, and environment). It also provides real-time intervention to held drivers within the safety tolerance zone (the detail for how the data is collected and what technologies are utilized is discussed in chapter 1).

The field experiment of the i-DREAMS project was set to take place in five European nations (Belgium, Germany, Greece, Portugal, and the United Kingdom), with an experimental protocol developed to collect data at each stage. The field experiment consisted of four stages and was scheduled to last a total of 12 months for all participants in all participating countries. Stage 1 was simulation (total duration: two months), stage 2 was pilot testing (total duration: one month), stage 3 was baseline measurement (total duration: two months), and stage 4 was intervention testing (total duration: seven months). Stages 2, 3, and 4 take place on the road, while the first stage was planned to be conducted in a driving simulator. Drivers who participate in the pilot study (stage 2) are not retained for stage 3 and stage 4 in order to avoid influencing the results of the real experiment (stages 3 and 4). They have already used the i-DREAMS technology platform in the pilot testing and thus were not eligible to participate in the main study.

All the drivers who participated in the study's intervention period (stage 4) also participated in the baseline measurement phase (stage 3). Stage 3 (baseline measurement) had only one phase, and stage 4 (in-vehicle and post-trip interventions) had three phases. In total, data in stage 3 and stage 4 were collected in four phases;

- Phase 1: Baseline measurement (no interventions), in which driving performance is recorded, but no i-DREAMS intervention technology is in use, i.e., their performance is monitored, but no alerts are received.
- Phase 2: Real-time intervention via an in-vehicle warning system.
- Phase 3: Real-time intervention via an in-vehicle warning system and post-trip feedback via a smartphone app.
- Phase 4: Real-time intervention via an in-vehicle warning system and post-trip feedback via a smartphone app combined with a gamified web platform.

Besides, table 11 lists the timing (the length of time each phase took) applied for each phase to carry out the field experiment of the i-DREAMS project in Belgium.

TABLE 11 Description of phases and duration overview

Phase	Description	Duration per participant
1	Baseline measurement (no interventions)	Four weeks
2	Real-time intervention via an in-vehicle warning system.	Four weeks
3	Real-time intervention via an in-vehicle warning system and post-trip feedback via a smartphone app	Four weeks
4	Real-time intervention via an in-vehicle warning system and post-trip feedback via a smartphone app combined with a gamified web-platform	Six weeks

3.2.2 Data employed

The primary subjects of the study were the i-dreams project truck drivers who took part in the on-road field experiment in Belgium. There were 75 truck drivers scheduled to participate in the on-road field experiment. They were split into two groups, with 38 drivers in the first group (G1) and 37 in the second group (G2). The need to divide the drivers was due to logistical reasons, requiring less equipment to be acquired than if everyone was tested simultaneously. Even though more than 20 truck drivers took part in the on-road field study, only 16 truck drivers' data were included in this study because that was the available data at the time for analysis. The first group was named European Transport Company (EUTRACO), while the second was called Group Op De Beeck (GODB). Each participant (truck driver) was given a unique code for communication and planning purposes, and information such as vehicle code, driver name, and address was logged. Table 12 shows the two groups of truck drivers (EUTRACO as G1, GODB as G2), as well as the week numbers (W1, W2, etc.) during which each group participated in the i-DREAMS project.

The first group (G1) of eight drivers began baseline measurements in week 1 (W1) for four weeks (from W1 to W4), after which they completed all intervention programs phase 2 (P2), phase 3 (P3), and phase 4 (P4) from W5 to W18. The same equipment was installed in G2 during the field experiment for 18 weeks (W1-W18). For the 18 weeks of field experiment W1 to W18, G2 followed the same procedure as G1 (P1, P2, P3, P4).

In general, the naturalistic data collected by the i-DREAMS project from the truck drivers who took part in the baseline measurement (phase 1) in Belgium are part of this thesis.

TABLE 12 Timetable of field experiment for both groups

Phase	Group	Week and date					
		W1	W2	W3	W4		
1	G1	9/20 - 9/26/2021	9/27 - 10/3/2021	10/4 - 10/10/2021	10/11 - 10/17/2021		
	G2	11/29 – 12/5/2021	12/6 – 12/12/2021	12/13 – 12/19/2021	12/20 – 12/26/2021		
		W5	W6	W7	W8		
2	G1	10/18 - 10/24/2021	10/25 - 10/31/2021	11/1 - 11/7/2021	11/8 - 11/14/2021		
	G2	12/27 – 1/2/2022	1/3 – 1/9/2022	1/10 – 1/16/2022	1/17 – 1/23/2022		
		W9	W10	W11	W12		
3	G1	12/20 - 12/26/2021	12/27 - 1/2/2022	1/3 - 1/9/2022	1/10 - 1/16/2022		
	G2	1/24 – 1/30/2022	1/31 – 2/6/2022	2/7 – 2/13/2022	2/14 – 2/20/2022		
		W13	W14	W15	W16	W17	W18
4	G1	1/17 - 1/23/2022	1/24 - 1/30/2022	1/31 - 2/6/2022	2/7 - 2/13/2022	2/14 - 2/20/2022	2/21 - 2/27/2022
	G2	2/21 – 2/27/2022	2/28 – 3/6/2022	3/7 – 3/13/2022	3/14 – 3/20/2022	3/21 – 3/27/2022	3/28 – 4/3/2022

3.2.3 Overview of the risky driving events employed

The i-DREAMS intervention, as mentioned in Chapter 1, aimed to effectively increase driver safety by assisting the driver in his/her driving task. As discussed in K. Brijs et al. (2020), four levels of driver safety were targeted to improve driver safety. Safety outcomes (SO) are at the highest level targeted by the i-DREAMS interventions (e.g., the likelihood of crash occurrences such as frontal crashes, side crashes, or rear crashes). Safety promoting goals (SPG) is at the second-highest level. These are the behaviors that should change to achieve the desired safety outcomes. The performance objectives (PO) are at the second-lowest level. These are the more specific behaviors or behavioral parameters that should be changed to achieve the safety-promoting objectives. The change objectives (CO) are at the bottom of the hierarchy. These are the underlying behavioral factors that should change for the performance goals to be met.

Of the four levels of driver safety, behaviors that describe safety-promoting goals and performance objectives were covered in this study. Vehicle control, sharing the road with others, speed management, driving fitness, and the use of safety devices are all behaviors frequently monitored in the context of safety-promoting interventions. The i-DREAMS platform set out to achieve five safety-promoting goals, four of which are included in this research. The following is a list of the four safety-promoting goals (K. Brijs et al., 2020):

- SPG1: Vehicle control performance (expressed as a numerical score) for trucks equipped with and subjected to the i-DREAMS interventions.
- SPG2: Performance in terms of sharing the road with others (expressed as a numerical score) for trucks equipped with and exposed to the i-DREAMS interventions.
- SPG3: Speed management performance (expressed as a numerical score) for trucks equipped with and exposed to the i-DREAMS interventions.
- SPG4: Performance in terms of driving under conditions where one is fit enough (expressed as a numerical score) for trucks equipped with and exposed to the i-DREAMS interventions.

Performance objectives are the more specific activities or behavioral parameters that must change to achieve the safety-promoting aims. More specific and relevant (surrogate) measurements will need to be proposed for properly operationalizing objectives specified at this highest degree of impact (K. Brijs et al., 2020). This research has four safety-promoting goals and eleven performance targets, as shown in figure 6, where several performance objectives are linked to one safety-promoting goal. Vehicle control consists of three performance objectives: acceleration, deceleration, and steering. Tailgating, overtaking, lane discipline, forward collision avoidance (FCA), and vulnerable road user collision avoidance (VRUCA) are the five performance objectives of road sharing. Besides, fatigue and distraction are incorporated as performance objectives of driver fitness, also known as health. Finally, speeding is the only performance objective that falls within the category of speed management.

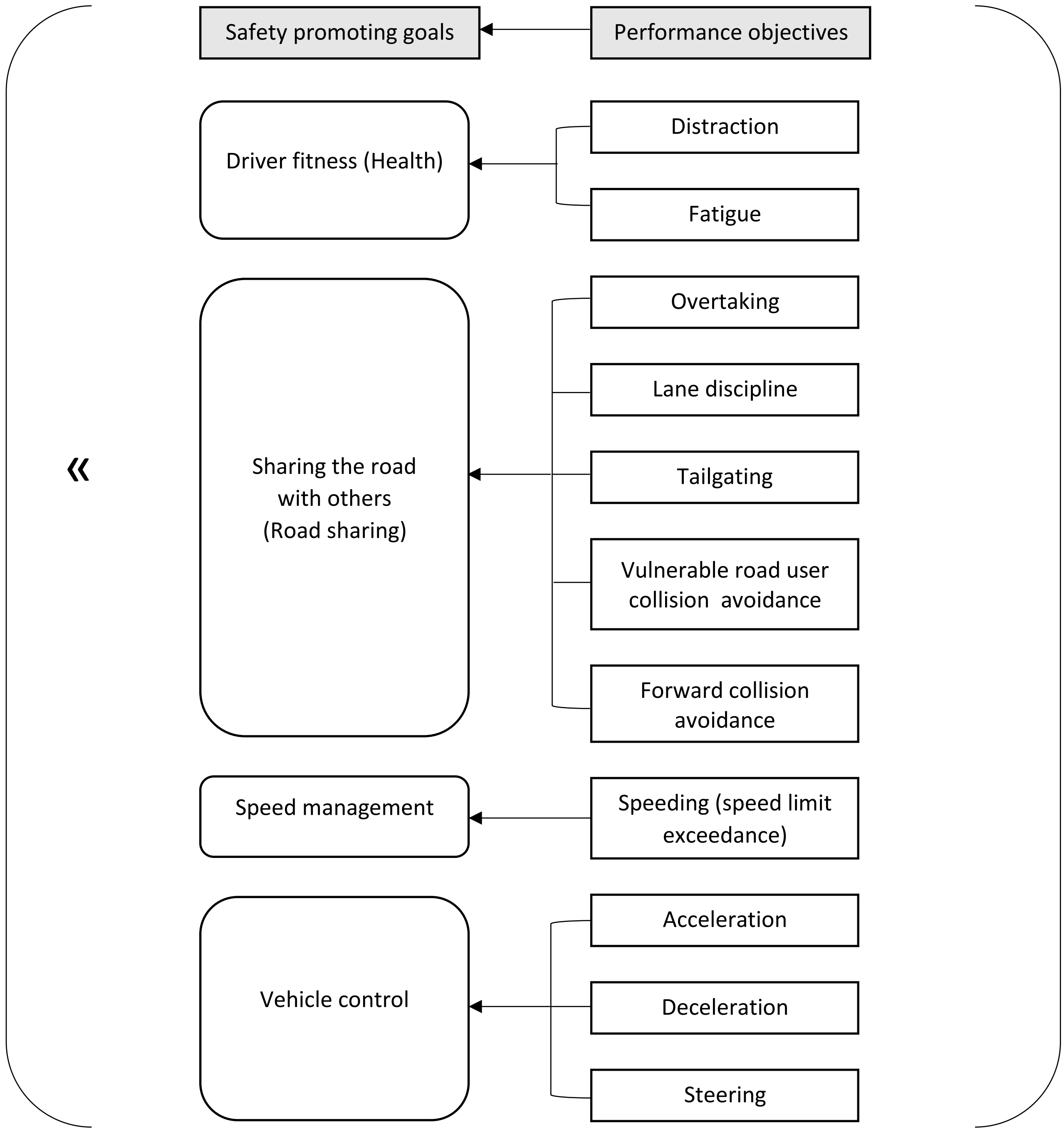


FIGURE 6 Safety promoting goals and performance objectives of the i-DREAMS project (K. Brijs et al., 2020).

3.2.4 Overview of the data obtained from the i-DREAMS project

The factors considered to characterize risky driving events were weather, road type, and time of day, as stated in the study's objective section. Except for weather data and road type shapefile, the data used in this study was obtained from the i-DREAMS project.

Figure 7 depicts the types of risky events considered to be characterized and the elements used during data processing. A data file was also provided containing the date ranges for each phase.

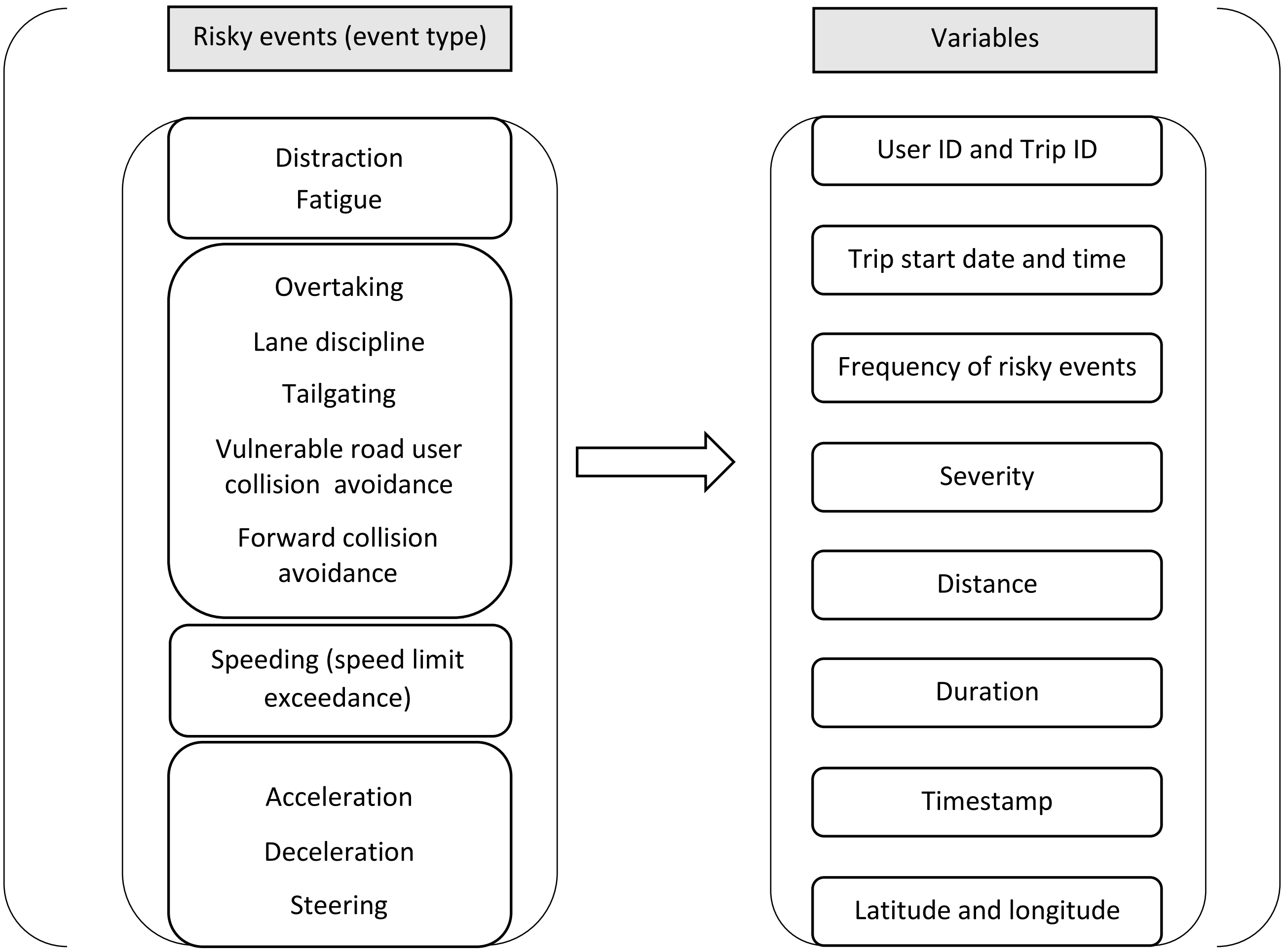


FIGURE 7 Risky events and variables obtained from the i-DREAM project.

3.2.5 Data obtained from open sources

3.2.5.1 Weather data

In addition to the data obtained from the i-DREAMS project, weather parameters were obtained through an open-source called National Aeronautics and Space Administration (NASA) Prediction of Worldwide Energy Resource (POWERS) data access viewer. Six parameters were collected from the open-source:

- The temperature at 2 meters
- Dew/frost point at 2 meters
- Relative humidity at 2 meters
- Precipitation corrected
- Surface pressure
- Wind speed at 10 meters

The six weather parameters were derived hourly, using Brussels's latitude and longitude as a location reference. Table 13 also includes a definition for each of the six meteorological parameters based on the NASA POWER data access viewer (2022):

TABLE 13 Weather parameters and definition

Parameter	Definition	Unit
Temperature at 2 meters	The average air (dry bulb) temperature at 2 meters above the surface of the earth.	Degree Celsius (C)
Dew/frost point at 2 meters	The dew/frost point temperature at 2 meters above the surface of the earth.	Degree Celsius (C)
Relative humidity at 2 meters	The ratio of actual partial pressure of water vapor to the partial pressure at saturation, expressed in percent.	Percent (%)
Precipitation corrected	The bias corrected average of total precipitation at the surface of the earth in water mass (includes water content in snow).	Millimeters per hour (mm/hour)
Surface pressure	The average of surface pressure at the surface of the earth.	Kilopascal (kPa)
Wind speed at 10 meters	The average of wind speed at 10 meters above the surface of the earth.	Meter per second (m/s)

3.2.5.2 Road type data

In addition to weather data, a shapefile of Belgium's road network that included a description of the road type was obtained from an open-source named DIVA-GIS (DIVA-GIS, 2022). Although the file downloaded contained a fair bit of unclear information that was not required for this thesis, it had been cleaned up and processed to the point where it could be used as an input for this study. The road types that were used in this study to characterize the occurrence of risky events are as follows:

- Motorway
- Primary roads
- Secondary roads
- Tertiary roads
- Trunk roads

3.3 Data processing

3.3.1 Organizing and processing risky events by time of the day

The i-DREAMS project provided a two-part data file packaged in excel as a Comma Separated Values (CSV) file for each group. The first data file contained trip events, which primarily described the frequency of occurrence of each risky event per trip, as well as communication and planning data, which included the dates when each phase began and ended. Also, user ID, trip ID, trip start date and time, duration, and distance were all included in the first data file. The second data file included location coordinates (latitude and longitude) where each risky event occurred. Besides, trip ID, event type, and timestamp were part of the second data file. The risky events were classified into three levels in both data files (low, medium, and high).

Following the acquisition of the data file, it was necessary to commence data processing through filtering, where the four phases could be separated. However, because each trip's date and time were presented as a single integrated row of data in a single column, it was necessary to separate them before starting the filtering process. As a result, the trip start date and start time were separated and assigned to distinct columns. Then, using the date ranges listed in table 12 for both groups, the filtering process was employed to extract data for each phase from the big data file, resulting in four separate sheets, one for each phase.

3.3.2 Organizing and processing risky events by weather

The data organization process was continued after obtaining all six weather data sets on an hourly basis via the open-source NASA POWER data access viewer. The data was extracted using the period between the start of phase one of group one (September 20, 2021) and the end of phase one of group two (April 3, 2022). The extracted data includes dates when a field experiment was conducted and dates when one was not since the data was not explicitly retrieved for the days when only a field experiment was conducted. As a result, as shown in figure 8, a matching process was carried out based on the date and timestamp where each risky event occurred. This helped return the weather data to each trip ID by making an exact match of the date and an approximate match of the timestamp.

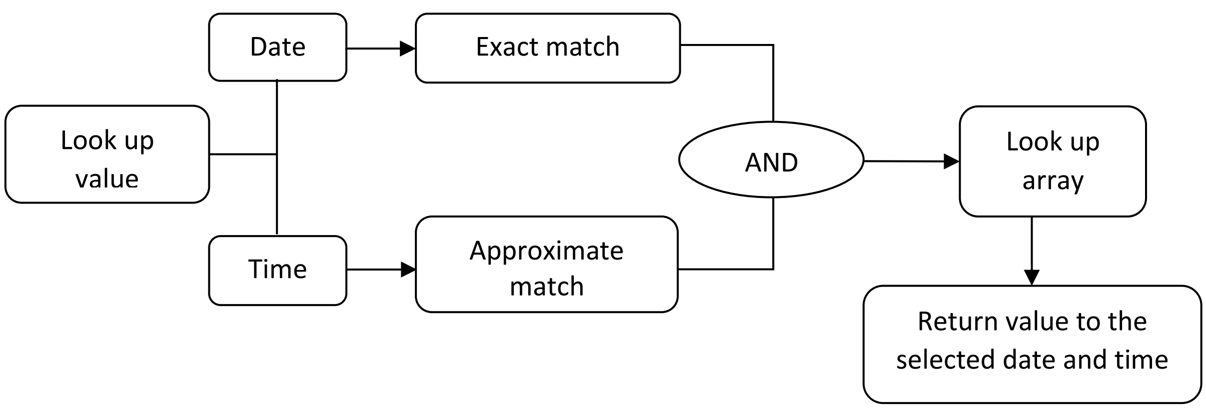


FIGURE 8 Data matching flow chart.

3.3.3 Organizing and processing risky events by location

A data file of risky events based on location was provided, containing each risky event's coordinates (latitude and longitude). This was a more detailed version of the data for the number of risky events per trip. For example, if a trip has an "N" number of risky events, the trip was divided into "N" rows, each row having the coordinates (latitude and longitude) for the corresponding risky event. The intent of having the spot where the risky events happened was to characterize these events based on road type. However, the data only had coordinates, with no information about the road layout (road type) where the coordinates were located. It was difficult to characterize the risky events without a road-type description. Hence, a Belgium road network shapefile with road type description was downloaded from an open source called DIVA-GIS (DIVA-GIS, 2022).

3.3.3.1 Buffering

The risky event data of phase one and the Belgium road network shapefile were first imported into Quantum Geographic Information System (QGIS) 3.22 to assign the road type to each coordinate. However, in addition to Belgium, the phase's layer included coordinates of risky events in the Netherlands, Germany, Luxembourg, and France, necessitating a buffering process. Buffering allows drawing circular boundaries around points or rectangular boundaries on either side of lines or around the outside of polygons. Hence, a buffer was created around the perimeter of the polygon generated from a shapefile representing Belgium's administrative boundary. The newly-created Belgium map layer with buffering was used as a reference feature. Then all of the risky events inside the Belgium map layer were selected and exported as a new shapefile. Figure 9 (a) shows the distribution of risky events before buffering, whereas figure 9 (b) shows the area in which the risky events included for this study, exclusively risky events that occurred in Belgium, which are acquired after the buffering process.

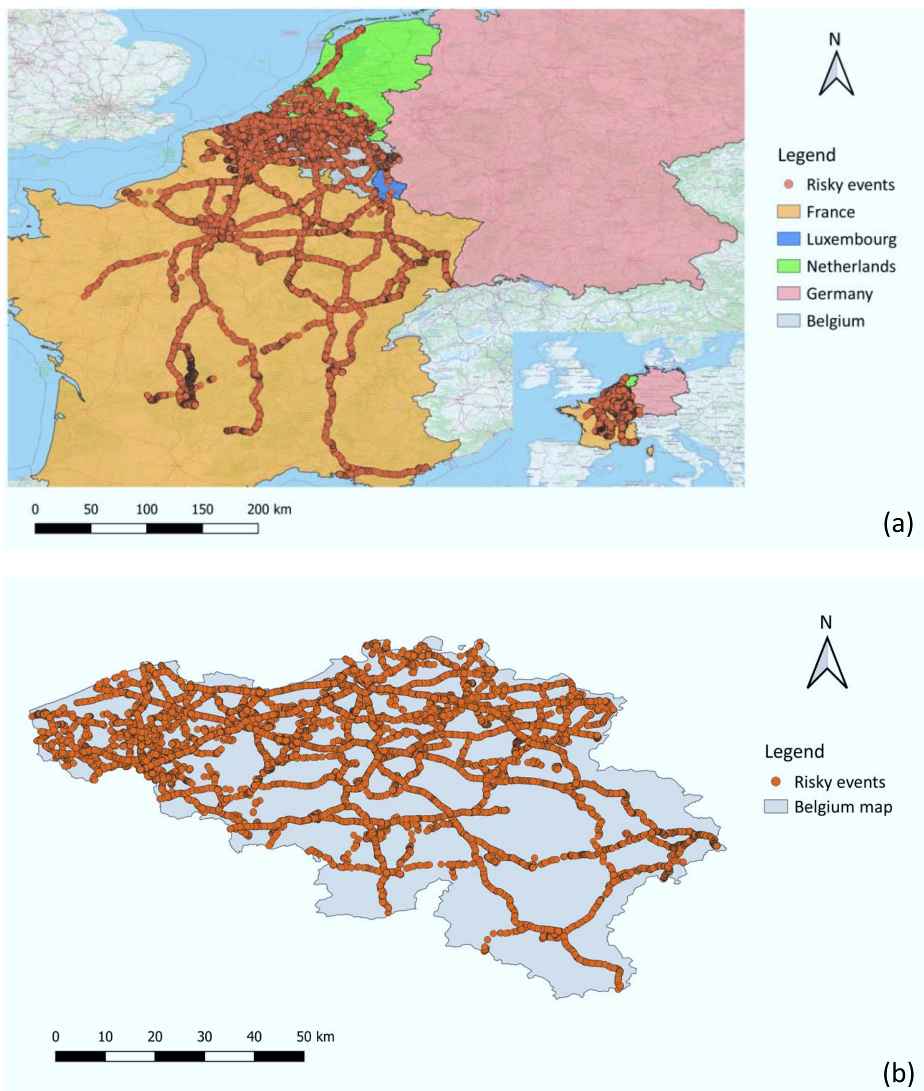


FIGURE 9 Buffering (a) risky events before buffering (b) risky events after buffering.

3.3.3.2 Join attributes by nearest

After all phase one risky event occurrences were delimited to the Belgium map, the road type attribution was done to define the location of the risky events. Before doing so, the geometric validity of the Belgium road network derived from the open-source had to be validated. As a result, a validity check on the geometries of a vector layer (Belgium road network) was performed using an algorithm (from QGIS version 3.22). The algorithm divides geometries into three groups (valid, invalid, and error), then generates a vector layer with the features from each category. The valid output layer only contains valid features (no topological errors), the invalid output layer contains all of the algorithm's invalid features, and the error output layer is a point layer that points to where the invalid features were found.

During the geometric validity check of the road network, six error outputs were discovered and automatically fixed using the built-in technique for fixing geometry errors. This technique tries to make a valid representation of an invalid geometry without losing any input vertices. After the geometry errors have been fixed, attributes are allocated to the closest position using the technique called join attributes by nearest. The nearest attribute assignment algorithm takes an input vector layer and creates a new vector layer that is an extended version of the original one, with more attributes in its attributable table. The additional attributes and values emanated from a second vector layer, where features are linked by locating the closest features in each layer.

The selected attributes from the nearest feature are included in the output features, as well as new attributes for the distance to the nearest feature, the feature's index, and the coordinates of the closest point on the input feature (feature X, feature Y) to the matched nearest feature, as well as the coordinates of the closet point on the matched feature (nearest X, nearest Y). Only the features indicating the road type were taken from the newly generated attributes, and the rest were cleaned. As a result, the final attribute table for analysis included the trip ID, risky event type, timestamp, the six weather parameters, road type, severity, and coordinates (latitude and longitude).

3.4 Method of data analysis

The data obtained from the i-DREAMS project was analyzed using the statistical methods listed below, which were chosen following literature reviews related to the current study, the type of data obtained from the i-DREAMS project, and the corresponding objectives formulated.

3.4.1 K-means Clustering

K-means clustering is one of the most straightforward unsupervised learning techniques to handle the well-known clustering problem. It divides things into clusters based on their similarities and differences from objects in other groups (Nirmal, 2019). According to James et al. (2013), the required number of clusters K must first be specified to perform K-means clustering. The K-means algorithm will then place each observation in precisely one of the K clusters. The K-means clustering method results from an easy-to-understand mathematical problem. Firstly, some notation is defined. Assume that sets $C_1,\ldots,C_k$ represents sets of indices for the observations in each cluster. These sets satisfy two requirements:

a) $C_1 \cup C_2 \cup \ldots \cup C_k = \{1, \ldots, n\}$. So each observation is a part of at least one of the K clusters.

b) $C_k \cap C_{k'} = \emptyset$ for all k ≠ k'. In other words, no observation belongs to more than one cluster, indicating that the clusters are non-overlapping.

If the i^{th} observation is found in the k^{th} cluster, then $i \in C_k$. According to the K-means clustering theory, a good clustering is one for which the within-cluster variation is as minimal as possible. The within-cluster variance for cluster C_k is a measure $W(C_k)$ of the extent by which observations within a cluster vary from one another. Though minimizing within-cluster variance is a logical idea, we must first define it to make it actionable. The most common method, the squared Euclidean distance, is used in the current study, where it is mathematically defined as follows:

$$W(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} \left(x_{ij} - x_{i'j}\right)^2 \tag{1}$$

Where $|C_k|$ represents the number of observations in the k^{th} cluster. This means that the within-cluster variance for the k^{th} cluster is equal to the sum of all the pairwise squared Euclidean distances between the observations in the kth cluster divided by the total number of observations in the k^{th} cluster.

Combining the idea of minimizing the variation within cluster and equation (1) gives the optimization problem that defines K-means clustering,

$$\underset{C_1, \dots, C_k}{minimize} \left\{ \sum_{k=1}^{k} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} \left(x_{ij} - x_{i'j}\right)^2 \right\} \tag{2}$$

Besides, the following approach addresses the K-means optimization problem mentioned in equation (2), which involves lowering the value of the objective in equation (2):

a) Give each observation a random number between 1 and K. These serve as the initial cluster assignments for the observations.
b) Continue iterating until the cluster assignments stop changing:
   - Calculate the cluster centroid for each of the K clusters. The vector of the p feature means for the observations in the kth cluster is the centroid of the kth cluster.
   - Assign each observation to the cluster with the closest centroid (where closest is defined using Euclidean distance).

3.4.2 Elbow method

The elbow method focuses on how much (percentage) variance can be explained as a function of the number of clusters. This approach is based on the premise that one should select the number of clusters to ensure that adding another cluster doesn't significantly improve the modeling of the data. The elbow method yields a graph of the percentage of variance explained by the clusters against the number of clusters. The first clusters will provide a lot of information, but eventually, the marginal gain will decline sharply, giving the graph a tilt. At this point, the appropriate 'K,' or the number of clusters, is selected, leading to the term "elbow criterion" (Bholowalia & Kumar, 2014). In this study, the number of K clusters to use for K-Means algorithm data grouping was chosen using the elbow criterion approach, where the elbow method is expressed using the Sum of Squared Error as follows (Irwanto et al., 2012):

$$SSE = \sum_{K=1}^{K} \sum_{x_i \in S_K} \| x_i - C_K \|_2^2 \qquad (3)$$

Where K is the number of clusters formed, xi is the data present in each cluster, CK is the cluster centroid, and SSE is the sum of the average Euclidean Distance of each point against the centroid. The idea is to start with K=2 and raise it by 1 in each step as you calculate your clusters and the training cost. At some value of K, the cost starts to decline sharply, and as you increase K more, the cost plateaus, indicating that this is the ideal K value. After doing this, there will be more clusters, but some old clusters will be quite close to the new ones.

3.4.3 Multivariate analysis of variance (MANOVA)

MANOVA is a general linear model (GLM) technique mainly designed to assess multi-variable statistical models. With MANOVA, it is possible to discriminate between two or more different groups using a combination of quantitative variables. When there are several dependent variables, i.e., when Analysis of Variance (ANOVA) is insufficient to identify group differences, MANOVA can be used as a replacement (Andy, 2009). The objective of this study is to ascertain the influence of time of day, weather, and road type on the prevalence of risky events (dependent variables), with the time of day having four clusters (groups), the weather having three clusters, and road type having five groups. To take advantage of MANOVA in assessing group differences using multiple variables, the study used clusters (groups) as predictors or fixed factors, risky events categorized into low, medium, and high severity, as well as additional total risky events as dependent variables.

This study included eleven risky events (dependent variables), as was previously described (see section 3.2.2). All 11 dependent variables were considered for computing the group (cluster) difference based on the total risky events, but only seven of the risky events were considered based on the low, medium, and high severity. This is because risky events, including distraction, forward collision avoidance, vulnerable road user collision avoidance, and lane discipline, were only recorded as total risky events. In light of this, the MANOVA design was an 11 X 11 matrix for comparison based on total risky events, whereas it was 7 X 7 based on low, medium, and high risky events.

The test statistic (F-ratio) for MANOVA is obtained by comparing the ratio of systematic to unsystematic variance for various dependent variables. The hypothesis sum of squares and cross products matrix (or hypothesis SSCP) is the matrix that depicts the systematic variance (or the model sum of squares for all variables). The error sum of squares and cross-products matrix (or error SSCP) is the matrix that represents the unsystematic variation (or residual sums of squares for all variables). This F-ratio represents the model's performance in terms of how good compared to how bad it is (how much error there is). In addition to the F-ratio, the measure of association strength, which defines the proportion of the overlapping variance between the independent variable and the first combination of dependent variables, was reported using Wilks's lambda (Λ) and the Pillai-Bartlett trace (V) (Andy, 2009). Wilks's lambda (Λ) is applied when the homogeneity of variance assumption is upheld, and Pillai's trace (V) is used when the assumption is violated.

Wilks's lambda is calculated using the product of the unexplained variance on each of the variates (see equation (4), where the symbol ∏ is identical to the summation symbol (Σ) except that it signifies multiply rather than add up). Therefore, Wilks's lambda measures the proportion of error variance to total variance for each variate (Andy, 2009):

$$\Lambda = \prod_{i=1}^{s} \frac{1}{1+\lambda_i} \tag{4}$$

The Pillai's trace V value is calculated using equation (5), where s is the number of variates and λ is the eigenvalue for each discriminant variate (Andy, 2009).

$$V = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i} \tag{5}$$

3.4.3.1 Follow-up analysis

The F-ratio primarily reveals whether the data-fitted model explains more variation than unrelated variables, but it does not tell which of the dependent variables accounts for the difference between the group. This prompted the adoption of a follow-up analysis (further analysis to identify which dependent variable(s) resulted in a significant difference), where the significance value of the initial test of the null hypothesis was used to determine whether or not follow-up analysis was necessary to be conducted. Hence, a follow-up analysis was not performed when a non-significant MANOVA (i.e., true null hypothesis) was achieved. On the other hand, a separate ANOVA was used for the follow-up analysis if the MANOVA output revealed statistically significant differences between the groups, where a p-value of less than 0.05 was regarded as a significant value.

Additionally, the results of independent ANOVA (between-subject effects) only reveal which dependent variables differ significantly between the groups, not where the differences between the groups reside or which group differs significantly from the others. Therefore, it was necessary to conduct further analysis to determine which groups differed after conducting separate ANOVA. This required comparing each group (as if running multiple t-tests), but using post hoc, a feature of MANOVA, made this task easier. In addition to pairwise comparison (comparing all possible pairings of the groups), the post hoc test determines if there is a positive or negative difference between the groups, i.e., whether the value of the variable under consideration is higher or lower than in one group than in another. There are two ways to do a post hoc test in a MANOVA: assuming equal or unequal variance. Based on Andy (2009), this study used Tukey's procedure for equal sample size and similar group variance and the Games-Howell procedure for any lingering doubts about the equality of group variance to conduct the post hoc test.

3.4.4 Correlation test

The risky events were subjected to a correlation test using a bivariate test to determine their simultaneous occurrence. Based on severity levels, a correlation test was conducted between the dependent variables (risky events), in which Spearman correlation analysis was utilized to calculate the correlation coefficient. This correlation coefficient measures how closely two variables are related linearly. The correlation coefficient value spans from absolute value +1 to -1, in which the ranges are interpreted as follows (Ratner, 2009):

- 0 implies the absence of a linear relationship.
- +1 denotes a perfect positive linear relationship, meaning that as one variable's values increase, the other variable also does so according to an exact linear rule.
- -1 denotes a perfect negative linear relationship, in which when one variable's values rise, the other variable's values fall according to an exact linear rule.
- Values between 0 and 0.3 (0 and -0.3) imply a weak positive (negative) linear relationship through a fragile linear rule.
- Values between 0.3 and 0.7 (-0.3 and -0.7) denote a moderately positive (negative) linear relationship through a fuzzy-firm linear rule.
- Values between 0.7 and 1.0 (-0.7 and -1.0) denote strong positive (negative) linear relationship through a firm linear rule.

Generally, when the correlation coefficient is closer to the absolute value of one, then the variables are highly correlated; when the value is closer to zero, the variables are less correlated. In the current study, a strong correlation was considered as one with a correlation coefficient above 0.5.

3.4.5 Kernel density estimation (KDE)

Kernel density estimation is one of the effective methods for calculating the spread of risk of an accident (Anderson, 2009). KDE entails covering each point with a symmetrical surface, calculating the distance between the point and a reference location using a mathematical function, and then adding the values for all the surfaces for that reference location. For successive points, the same process is repeated. This enables applying a kernel to each observation, and adding the individual kernels provides the density estimate (smoothly continuous intensity surface) for the distribution of accident spots. The intensity is maximum at the point event center and gradually drops until it reaches zero at the radius of the research circle (Fotheringham et al., 2000). The density at a definite location is calculated by equation (6):

$$f(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{d_i}{h}\right) \qquad (6)$$

where f(x, y) represents the estimated density at the given location (x, y), n stands for the number of observations, h is the bandwidth or kernel size, K is the kernel function, and d_i is the distance between the location (x, y) and the location of the ith observation. The KDE method yields a raster result that is displayed as a grid of cells. The key variables that affect the KDE technique are cell size and bandwidth. The selection of bandwidth is highly subjective (Anderson, 2009).

3.4.5.1 Categorization of hotspots

The KDE is applied to estimate the density of risky events based on the total and the three severity-based risky events. Hence, the dataset of risky events was first split into four separated data files based on total, low, medium, and high risky events. The KDE is then applied to each data file in which each risky event was grouped according to the time of the day, weather, and road type. It makes sense that an area with many trips would encounter higher risky events. For example, if the number of trips on motorways was higher than on primary roads, there would also be a higher number of risky incidents on motorways. Therefore, it will be difficult to conclude that the frequency of risky events is higher on motorways, as this could be due to a higher number of trips on motorways. In this study, the risky events in the high-density areas were extracted and standardized by dividing the distance traveled in that particular high-density area to derive a valid conclusion.

Besides, in the analysis using KDE, because there is no index associated with statistical significance, hotspots were categorized using equal intervals. As a result, the risky event density was divided into five categories: very low, low, medium, high, and very high. The risky events that fell within the high and very high-density range were then employed for this study.

4 DATA ANALYSIS AND RESULTS

The study's findings are divided into two sections. The first section presents the findings of the descriptive analysis carried out on the 11 risky events (variables) based on the severity level. This section also briefly explains k-means clustering results, including the number of clusters obtained using the elbow method on weather and time of day, and their cluster centers. Besides, a description of the type of road considered to characterize the risky events based on road layout is covered in this section. The findings from the multivariate analysis of variance, correlation test on the dependent variables, and kernel density estimation are presented in the second part. This section is further divided into three parts depending on the three factors.

4.1 Descriptive statistics

4.1.1 Risky events description

In general, this study utilized the data from 16 truck drivers of the i-DREAMS project, taking 799 trips and recording a total of 68,775 risky events that occurred during those trips. Table 14 gives descriptive statistics of the frequency distribution of the 11 risky events (performance objectives) in terms of overall risky events and the three severity levels (low, medium, and high). Out of the total risky events recorded, steering and tailgating had the highest frequency of occurrence, with 23,071 (33.6%) and 20,993 (30.5%), respectively. Variables like vulnerable road user collision avoidance, overtaking, and forward collisions have the lowest rates of occurrence, with frequencies of 1 (0.0%), 12 (0.0%), and 90 (0.1%), respectively. Also, throughout these 68,776 risky events, there were no low-risk events related to speeding, tailgating, or overtaking (see table 14).

TABLE 14 Descriptive statistics of risky events (performance objectives)

Risky events	Low	Medium	High	Total
Speeding	0 (0.0)	530 (30.5)	1208 (69.5)	1738 (2.5)
Acceleration	6590 (63.0)	3364 (32.2)	498 (4.8)	10452 (15.2)
Deceleration	8017 (93.3)	534 (6.2)	44 (0.5)	8595 (12.5)
Steering	15333 (66.5)	7446 (32.3)	292 (1.3)	23071 (33.6)
Tailgating	0 (0.0)	17057 (81.3)	3936 (18.8)	20993 (30.5)
Overtaking	0 (0.0)	4 (33.3)	8 (66.7)	12 (0.0)
Fatigue	149 (48.2)	105 (34.0)	55 (17.8)	309 (0.4)
Lane discipline				3283 (4.8)
FCA				90 (0.1)
VRUCA				1 (0.0)
Distraction				231 (0.3)

Similarly, descriptive statistics were conducted on the frequency distribution of the risky events depending on the four safety-promoting goals. As depicted in figure 10, of all risky events detected, vehicle control accounted for the substantial portion, 42118 (61.2%); the lowest percentages were for health (driver fitness) and speed management, at 540 (0.8%) and 1738 (2.5%), respectively. Also, neither speed management nor health contributed to the occurrences of low risky events. Generally, speed management and road sharing contributed the most for high and medium risky events, respectively, contributing 69.5% and 70.0% each. While vehicle control made the most considerable contributions to the occurrence of both low and total risky events, contributing 71.1% and 61.2% each. Health was the least in all the severity levels. The lack of the three severity levels for one of its performance objectives, distraction, may account for the health's little impact.

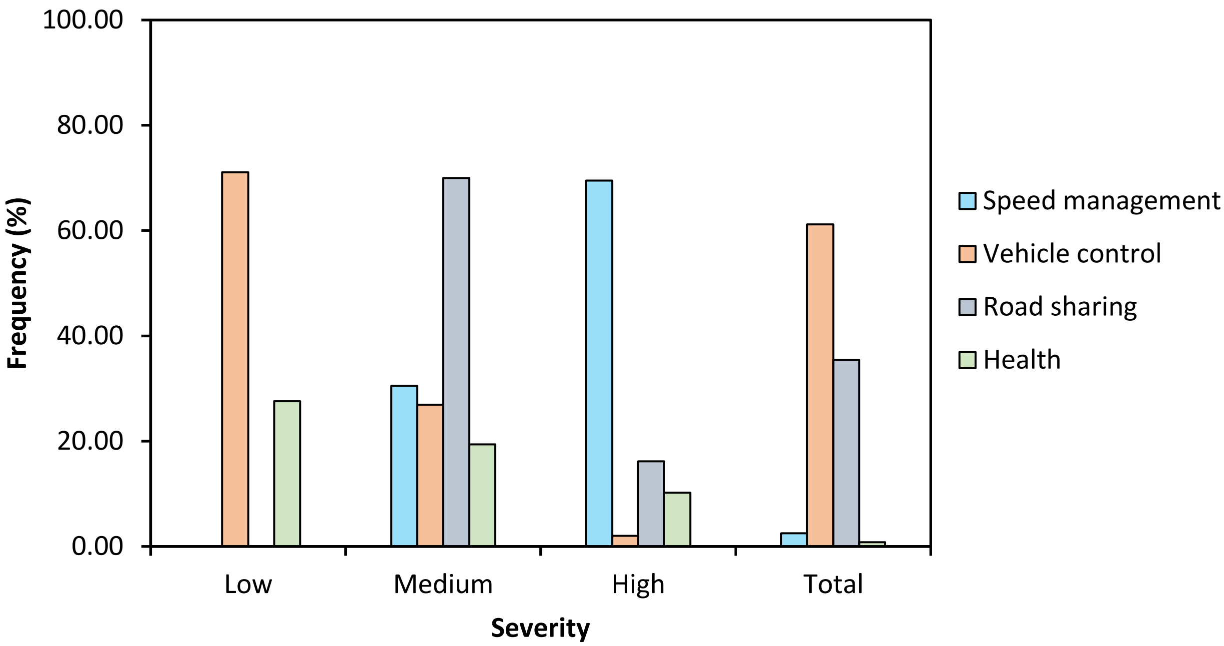
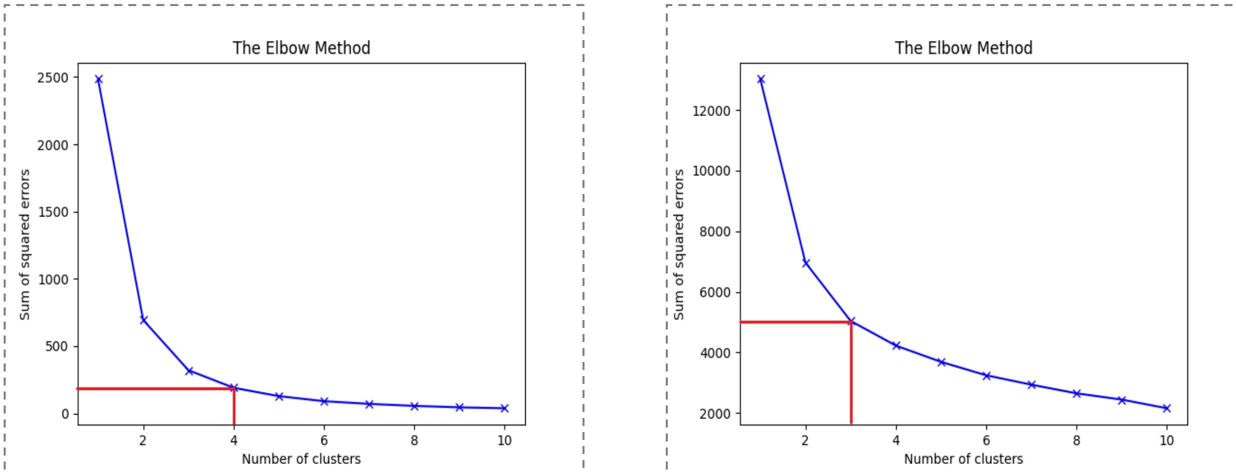


FIGURE 10 Frequency of risky events based on safety-promoting goals.

4.1.2 Cluster description

4.1.2.1 Number of clusters

The number of clusters based on time of the day and weather was calculated by analyzing the timestamps of the risky events and the six meteorological parameters, respectively. Figures 11 (a) and (b) show the elbow graph for the time of day and weather parameters, respectively. Four clusters were selected based on the time of day, and three were selected based on weather factors.



(a) (b)

FIGURE 11 Elbow graph (a) Time of the day and (b) Weather.

4.1.2.2 Cluster center

K-means clustering was employed to determine cluster centers for the time of day using the timestamp at which the risky events were recorded and four elbow-derived numbers of clusters. Table 15 shows the cluster center computed using k-means clustering for each cluster, with the time range description based on a study by Pokorny et al. (2017).

TABLE 15 Clustering based on time of the day

Cluster	Cluster center	Time range	Time of the day
1	4:34:38	00:00 - 7:00	Night/Early morning
2	12:24:48	11:00 - 15:00	Midday
3	16:18:45	15:00 - 00:00	Afternoon/evening
4	8:37:06	7:00 - 11:00	Morning

Similarly, the weather's cluster center was determined using k-means clustering, and variables including temperature, dew point, humidity, pressure, precipitation, and wind were used. Also, to compute the cluster center, three numbers of clusters yielded by the elbow technique were used. The cluster centers for each cluster are shown in Table 16, and the cluster center for each cluster is defined as follows:

- Cluster 1 (adverse weather condition): Low temperature, dew point, pressure, wind, moderate precipitation, and high humidity.
- Cluster 2 (average weather condition): High precipitation, moderate temperature, dew point, humidity, pressure, and wind.
- Cluster 3 (clear weather condition): High temperature, dew point, pressure, and wind, along with little precipitation and humidity.

TABLE 16 Clustering based on weather

Cluster	Temperature	Dew point	Humidity	Precipitation	Pressure	Wind
1	5.31	4.63	95.22	0.1	101.2	4.61
2	10.03	7.04	81.67	0.11	101.44	5.2
3	16.22	9.25	63.57	0.04	101.69	5.51

Additionally, each case (risky events) was grouped into a cluster based on how differently the observations (timestamps for cases involving time of day and the six metrological elements for cases involving weather) varied within each cluster (see section 3.4.1). Figure 12 (a) and (b) show the percentage of risky events grouped according to the time of day and weather, respectively. In the case of time of the day, nearly all of the clusters have the same number of risky events, with cluster 4 (morning) having the most 19477 (27.8%) and cluster 3 (afternoon/evening) having the least 21.7% risky events (see figure 12 (a)). The clusters were designed to maximize the differences between the cases (timestamp) in various clusters, $F(3, 68771) = 277805$, $p < .01$. While in the case of weather-based clustering, the proportion of risky events categorized in each cluster differed widely. Figure 12 (b) demonstrates cluster 1 has the highest percentage of risky events, 44586 (64.8%), followed by cluster two, 14666 (21.3%), and cluster three, 9532 (13.8%). Like time-based clustering, weather-based clustering was formed to maximize the differences between the cases in various clusters, $F(2, 68772) = 846.88$, $p < .01$.

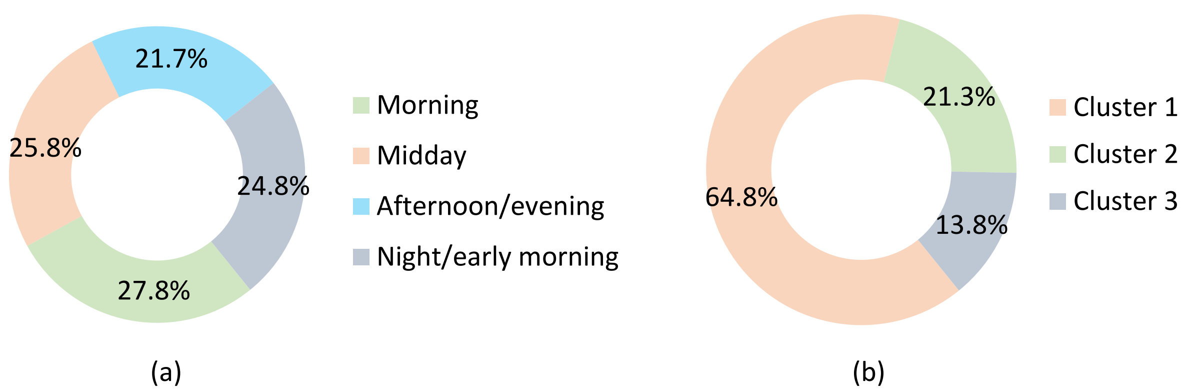


FIGURE 12 Percentage of clustered risky events based on (a) Time of the day and (b) Weather.

4.1.3 Road type description

Five categories of roads are used to analyze risky events based on road layout: motorways, primary roads, secondary roads, tertiary roads, and trunk roads. Figure 13 exhibits the distribution of the five types of roads throughout the Belgium road network. After the risky events were loaded into QGIS and superimposed on the road network, the algorithm called join attributes by nearest (see section 3.3.3.2) was used to join them with the closest road type.

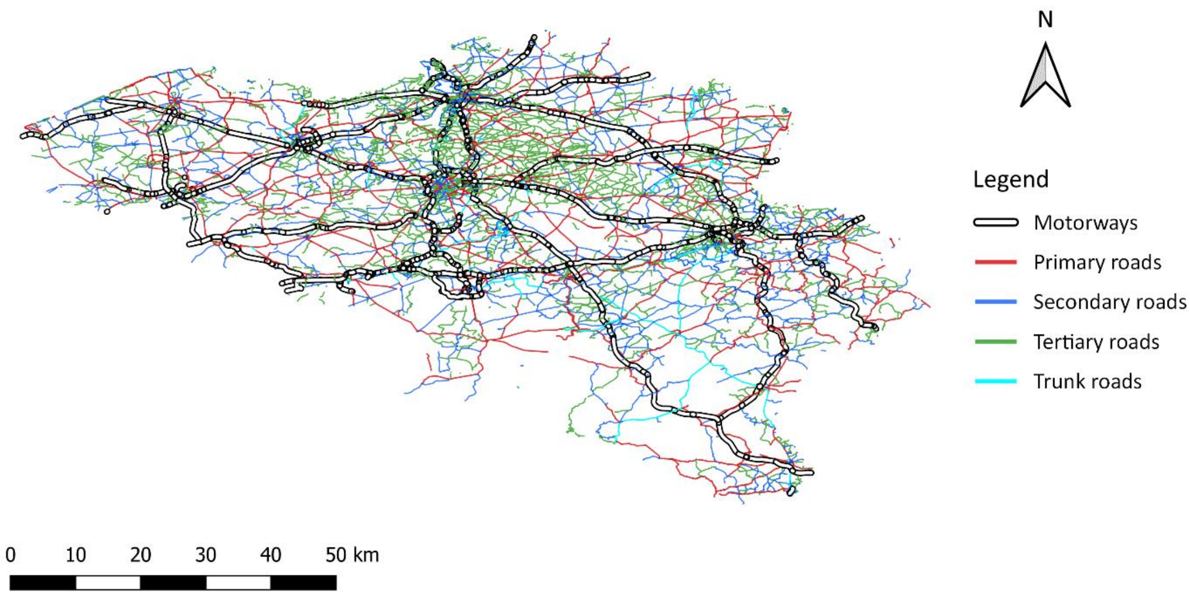


FIGURE 13 Belgium road network.

The algorithm (join attributes by nearest) results in a joined layer containing risky event occurrence per road type, where the description of the percentage distribution of the risky event on the five road types is shown in figure 14. Most risky events were reported to have taken place on motorways (46.4%), while the fewest risky events were found on tertiary and trunk roads (8.6% and 10.3%, respectively). This proportion of risky events may have occurred since truck drivers were the study's focus. Given the size of the vehicle, it is clear that truck drivers favor driving on motorways. Additionally, due to traffic congestion and their size, heavy trucks are prohibited from traveling on some minor roads, which may also account for the lower frequency of risky events on tertiary roads.

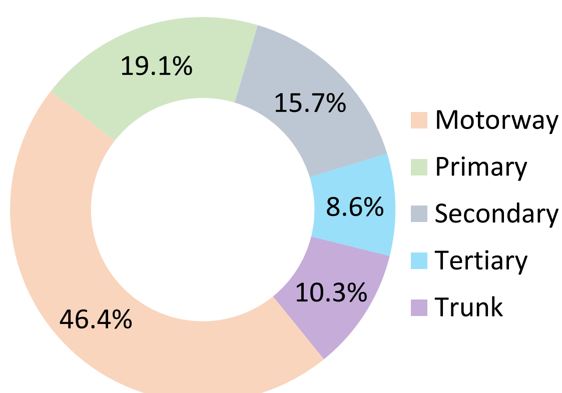


FIGURE 14 Percentage of risky events based on road type.

4.2 Characterization of risky events

The classification of risky events based on the time of day, the weather, and the type of road are explained in the following sections. The frequency of risky events per trip was standardized to 100 kilometers for each of the three aforementioned factors. For instance, using the four clusters that were created based on the time of day, the risky events that occurred on each cluster were divided by the distance traveled on each cluster, and then the result was standardized to 100 kilometers. A similar strategy was used for both the type of road and the weather. As a result, MANOVA, correlation test, and KDE results, which are covered in more detail in the following sections, were applied after standardizing the number of risky events by the distance traveled.

4.2.1 Multivariate analysis of risky events

The results of the multivariate analysis of risky events based on the time of day, weather, and type of road are presented in the following sections. Each of the previously listed factors is followed by a detailed presentation of the results from the multivariate test, between-subject effects (univariate ANOVA), and multiple comparisons (post hoc test).

4.2.1.1 Analysis of risky events based on time of the day

Analyzing the risky events based on time of the day revealed that the time of day had no statistically significant effect on the likelihood of total risky events, $V = .576$, $F(30, 150) = 1.189$, $p > .05$, low risky events, $V = .202$, $F(12, 168) = 1.011$, $p > .05$, and high risky events, $V = .307$, $F(21, 159) = .864$, $p > .05$. However, the variation in time of day resulted in a significant difference on the occurrence of the medium risky events with $V = .558$, $F(21, 159) = 1.73$, $p = .031$. Besides, table 17 shows further analysis of the dependent variables for the medium severity.

TABLE 17 Between-subject effects: impact of time of the day on medium severity dependent variables

Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Speeding	1.068	3	.356	2.107	.109
Acceleration	28.649	3	9.55	.27	.847
Deceleration	1.188	3	.396	.774	.513
Steering	91.852	3	30.617	.198	.897
Tailgating	2060.451	3	686.817	2.498	.069
Overtaking	.001	3	0	1.761	.165
Fatigue	.167	3	.056	5.75	.002

As shown in Table 17 above, time of the day had a statistically significant effect on medium severity fatigue events, $F(3, 57) = 5.75$; $p < .01$. Moreover, a follow-up analysis was carried out on fatigue with Games-Howell post hoc test, as shown in Table 18 below in the multiple comparisons table.

TABLE 18 Multiple comparisons: differences in medium fatigue events based on time of the day

(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Morning						
	Midday	.036	.028	.581	-.040	.113
	Afternoon/evening	-.107	.046	.128	-.237	.023
	Night/early morning	.014	.030	.964	-.067	.096
Midday						
	Afternoon/evening	-.143	.044	.024	-.270	-.016
	Night/early morning	-.022	.028	.861	-.097	.054
Afternoon/ evening	Night/early morning	.121	.046	.070	-.008	.250

Based on observed means.
The error term is Mean Square(Error) = .010.
* The mean difference is significant at the .05 level.

The table above exemplifies that, on average, medium fatigue event occurrences were significantly less frequent during midday than in the afternoon or evening ($p = .024$), but it did not differ significantly between morning and midday ($p = .581$), morning and afternoon/evening ($p = .128$), and morning and night/early morning ($p = .964$). Also, the difference in medium fatigue events between midday and night/early morning ($p = .861$) and between afternoon/evening and night/early morning ($p = .07$) was not statistically significant.

4.2.1.2 Analysis of risky events based on weather conditions

The risky events were put through a MANOVA to determine whether there was a relationship between weather clusters and the frequency of risky events. Weather was seen to have a statistically significant effect on the frequency of total risky events, $\Lambda = .359$, $F(20, 54) = 1.805$, $p = .044$, low risky events, $\Lambda = .576$, $F(8, 66) = 2.619$, $p = .015$, and medium risky events, $\Lambda = .463$, $F(12, 62) = 2.426$, $p = .012$. However, there was no statistically significant impact of weather on the frequency of high severity risky events, $V = .519$, $F(14, 62) = 1.55$, $p > .05$. In addition, a follow-up analysis was carried out on total risky events, low and, medium severity risky events. As shown in Table 19 below, it was appeared that weather had a statistically significant effect on total speeding events, $F(2, 36) = 3.542$; $p = .039$, and vulnerable road user collision avoidance events, $F(2, 36) = 3.443$; $p = .043$.

TABLE 19 Between-subject effects: impact of weather on the dependent variables based on total events

Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Speeding	7.922	2	3.961	3.542	.039
Acceleration	1623.171	2	811.586	2.533	.093
Deceleration	140.841	2	70.421	1.825	.176
Steering	3161.359	2	1580.679	1.718	.194
Tailgating	750.713	2	375.357	.833	.443
Overtaking	.001	2	.001	.142	.868
VRUCA	.359	2	.18	3.443	.043
FCA	.49	2	.245	.629	.539
Fatigue	.019	2	.009	.606	.551
Distraction	6.457	2	3.229	.067	.935

Again, a further analysis was conducted on total speeding and vulnerable road user collision avoidance with the Tukey HSD post hoc test, as shown in Table 20 below. On average, there were substantially more speeding incidents in cluster 1 than in cluster 2 (p = .032), but there was no statistically significant difference between cluster 1 and cluster 3 (p = .377) or between cluster 2 and cluster 3 (p = .761). Despite the between-subject effect showing that weather had a statistically significant impact on vulnerable road user collision avoidance, multiple comparisons revealed that there was no statistically significant difference in the frequency of VRUCA between clusters 1 and 2 (p = .065), clusters 1 and 3 (p = .976), and clusters 2 and 3 (p = .121)

TABLE 20 Multiple comparisons: differences in total risky events based on weather clusters

Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Speeding	Cluster 1	Cluster 2	.986*	.374	.032	.073	1.9
		Cluster 3	.648	.479	.377	-.524	1.819
	Cluster 2	Cluster 3	-.339	.479	.761	-1.51	.833
VRUCA	Cluster 1	Cluster 2	.188	.081	.065	-.009	.385
		Cluster 3	-.022	.104	.976	-.275	.232
	Cluster 2	Cluster 3	-.21	.104	.121	-.463	.044

Based on observed means.

The error term is Mean Square(Error) = 48.210.

* The mean difference is significant at the .05 level.

Similarly, a follow-up test on low severity risky events has been conducted, as shown in Table 21 below. It appeared that weather had a statistically significant effect on low fatigue events, F (2, 36) = 3.303; p = .048.

TABLE 21 Between-subject effect: impact of weather on low severity dependent variables

Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Acceleration	497.193	2	248.596	1.901	.164
Deceleration	132.831	2	66.415	1.894	.165
Steering	1184.25	2	592.125	2.277	.117
Fatigue	.133	2	.066	3.303	.048

Additional analysis on low fatigue events was conducted using the Tukey HSD post hoc test, as shown in Table 22 below. A statistically significant difference existed between clusters 2 and 3 (p = .045), and the mean difference revealed that cluster 2 had a reduced likelihood of low fatigue events than cluster 3. However, no statistically significant difference was found between clusters 1 and 3 (p = .414), nor between clusters 1 and 2 (p = .280).

TABLE 22 Multiple comparisons: differences in low fatigue events based on weather clusters

(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Cluster 1	Cluster 2	.078	.050	.280	-.045	.200
	Cluster 3	-.082	.064	.414	-.239	.075
Cluster 2	Cluster 3	-.160*	.064	.045	-.317	-.003

Based on observed means.
The error term is Mean Square(Error) = .020.
* The mean difference is significant at the .05 level.

Moreover, the medium severity dependent variables have been subjected to a follow-up test. The result suggested that weather significantly impacted medium fatigue events, F (2, 36) = 5.158; p = .011. Additional analysis of medium fatigue events using the Tukey HSD post hoc test revealed a statistically significant difference between clusters 1 and 2 (p = .017), with the mean difference indicating that medium fatigue events were more common in cluster 1 than in cluster 2. As opposed to this, there was no statistically significant difference between clusters 1 and 3 (p = .991) or clusters 2 and 3 (p = .056).

4.2.1.3 Analysis of risky events based on road type

Analyzing the risky events based on road type revealed that road type had a statically significant impact on the total risky events, V = .982, F(40, 276) = 2.245, p < .001, low risky events, V = .478, F(16, 300) = 2.543, p < .001, medium risky events, V = .827, F(24, 292) = 3.172, p < .001, and high risky events, V = .619, F(28, 288) = 1.884, p = .006. A follow-up study was also conducted on the three severity-based risky events as well as the total number of risky events. Table 23 shows between-subject effects based on total risky events, where type of road had a substantial impact on the dependent variables, including speeding, F (4, 75) = 3.851; p = .007, acceleration, F (4, 75) = 5.689; p < .001, deceleration, F (4, 75) = 3.571; p < .001, steering, F (4, 75) = 6.857; p < .001, and tailgating, F (4, 75) = 10.708; p < .001.

The statistically significant effect discovered on speeding, acceleration, deceleration, steering, and tailgating was further examined using multiple comparisons of the Games-Howell post hoc test to see where that significant impact was experienced based on the types of roads. Unlike primary and secondary roads, motorways had considerably fewer total speeding events ($p = .013$ and .043, respectively). Additionally, there were substantially fewer total acceleration events on motorways than on primary, secondary, and trunk roads ($p < .001$, .003, and .043, respectively). In comparison to primary and secondary roads, motorways had significantly fewer total deceleration events ($p = .007$ and .009, respectively). Furthermore, there were considerably fewer total steering events on motorways compared to primary roads ($p = .003$), secondary roads ($p = .015$), and tertiary roads ($p = .003$). It was observed that total steering events on primary roads were significantly higher than on trunk roads ($p = .043$). Compared to secondary and tertiary roads, motorways had significantly more incidents of total tailgating events ($p < .001$). Moreover, the total number of tailgating events was substantially greater on trunk roads than on secondary and tertiary roads ($p = .022$ and .005, respectively), and on primary roads than on tertiary roads ($p = .015$).

TABLE 23 Between-subject effects: impact of road type on the dependent variables based on total events

Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Speeding	843.697	4	210.924	3.851	.007
Acceleration	46260.04	4	11565.01	5.689	<.001
Deceleration	40575.81	4	10143.95	3.571	.01
Steering	267240.2	4	66810.06	6.857	<.001
Tailgating	28425.16	4	7106.29	10.708	<.001
Overtaking	.203	4	.051	.678	.609
Fatigue	1.554	4	.389	.766	.551
Distraction	9.669	4	2.417	.824	.514
FCA	3.706	4	.927	1.927	.115
Lane discipline	980.096	4	245.024	1.347	.261

Table 24 displays the between-subject effects observed after a follow-up analysis was conducted for the significant impact of road type revealed on low risky events. The road type had a substantial impact on the dependent variables, including acceleration, $F(4, 75) = 7.219$; $p < .001$, deceleration, $F(4, 75) = 3.864$; $p = .007$, and steering, $F(4, 75) = 6.808$; $p < .001$. Additionally, further analysis revealed that motorways had considerably fewer low acceleration events than primary, secondary, tertiary, and trunk roads ($p < .001$, .002, .012, and .032, respectively). Low steering events were also significantly lower on motorways than on primary, secondary, and tertiary roads ($p < .001$, .01, and .003, respectively). There were significantly fewer low deceleration events on motorways compared to primary ($p = .004$) and secondary ($p = .006$) roads.

TABLE 24 Between-subject effects: impact of road type on low severity dependent variables

Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Acceleration	19888.83	4	4972.208	7.219	<.001
Deceleration	33653.99	4	8413.497	3.864	.007
Steering	102803.4	4	25700.85	6.808	<.001
Fatigue	.319	4	.08	.459	.766

A follow-up analysis was done to see whether the type of road significantly affected medium risky events, as shown in Table 25. According to the observed between-subject effects, the road type significantly influenced the dependent variables, including speeding, $F(4, 75) = 4.366$; $p = .003$, acceleration, $F(4, 75) = 3.685$; $p = .009$, steering, $F(4, 75) = 5.226$; $p < .001$, and tailgating, $F(4, 75) = 11.088$; $p < .001$. Further analysis revealed that primary roads had significantly more instances of medium speeding than motorways ($p = .012$) and trunk roads ($p = .038$). Additionally, it was found that both primary and secondary roads had considerably higher occurrences of medium acceleration events than motorways ($p = .004$ and $.01$, respectively). Unlike the primary, secondary, and tertiary roads, motorways had considerably fewer medium steering events ($p = .045$, $.035$, and $.009$, respectively). On the other hand, it was observed that there were significantly more medium tailgating events on motorways than on secondary ($p < .001$) and tertiary roads ($p < .001$). Medium tailgating was also shown to be substantially less on secondary roads than on trunk roads ($p = .015$) and tertiary roads than on primary and trunk roads ($p = .012$ and $.003$, respectively).

TABLE 25 Between-subject effects: impact of road type on medium severity dependent variables

Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Speeding	138.343	4	34.586	4.366	.003
Acceleration	4599.784	4	1149.946	3.685	.009
Deceleration	336.491	4	84.123	1.673	.165
Steering	34054.99	4	8513.747	5.226	<.001
Tailgating	18825.66	4	4706.414	11.088	<.001
Fatigue	1.305	4	.326	2.36	.061

The follow-up analysis of high-risk events showed that the road type had a substantial impact on the dependent variables, including speeding, $F(4, 75) = 2.544$; $p = .046$, and tailgating, $F(4, 75) = 7.577$; $p < .001$ (see table 26). Despite the road type significantly affecting high speeding incidents, the post hoc test revealed that there were only marginally more significant speeding events on primary and secondary roads than on a motorway. However, a post hoc analysis of high tailgating events revealed that motorways experienced more incidents than primary, secondary, and tertiary roads ($p = .048$, $.013$, and $.006$, respectively).

TABLE 26 Between-subject effects: impact of road type on high severity dependent variables

Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Speeding	321.428	4	80.357	2.544	.046
Acceleration	117.98	4	29.495	.783	.54
Deceleration	1.281	4	.32	.737	.57
Steering	231.462	4	57.866	1.877	.123
Tailgating	1066.199	4	266.55	7.577	<.001
Overtaking	.241	4	.06	.835	.507
Fatigue	.47	4	.117	1.239	.302

4.2.2 Correlation test

The bivariate correlations of the target variables were examined. The target variables' Pearson correlation coefficient and descriptive statistics (mean and standard deviation) are demonstrated below based on the severity level.

4.2.2.1 Correlation test based on total risky events

Based on the total risky events, the Pearson correlation coefficient revealed a strong positive linear relation between fatigue and tailgating, $r(16) = .994$, $p < .001$, steering and acceleration, $r(16) = .901$, $p < .001$, speeding and tailgating, $r(16) = .881$, $p < .001$, fatigue and speeding, $r(16) = .865$, $p < .001$, steering and deceleration, $r(16) = .848$, $p < .001$, lane discipline and deceleration, $r(16) = .758$, $p < .001$, and steering with speeding, $r(16) = .723$, $p = .002$. Besides, fatigue with FCA, $r(16) = .693$, $p = .003$, deceleration with speeding, $r(16) = .680$, $p = .004$, FCA with tailgating, $r(16) = .634$, $p = .008$, and deceleration with acceleration, $r(16) = .593$, $p = .015$ were strongly correlated (see table 27).

TABLE 27 Correlation test on total risky events based on time of the day

Risky events	M	SD	1	2	3	4	5	6	7	8	9	10	11
1. Speeding	13.393	19.564	--										
2. Acceleration	224.518	390.456	.447	--									
3. Deceleration	145.273	284.257	.680**	.593*	--								
4. Steering	328.097	478.467	.723**	.901**	.848**	--							
5. Tailgating	301.036	927.258	.881**	.204	.307	.441	--						
6. Overtaking	.035	.075	-.111	-.223	-.177	-.230	-.112	--					
7. VRUCA	.002	.007	-.164	-.121	-.121	-.158	-.075	-.123	--				
8. FCA	.506	.930	.460	.069	.247	.306	.634**	-.095	-.129	--			
9. Lane discipline	14.300	26.096	.292	.381	.758**	.533*	-.144	.248	-.146	-.156	--		
10. Fatigue	4.971	14.487	.865**	.181	.314	.433	.994**	-.121	-.083	.693**	-.165	--	
11. Distraction	.563	1.588	-.113	-.185	-.121	-.207	-.120	-.010	.014	-.204	-.150	-.096	--

** P < .01.
* P < .05.

4.2.2.2 Correlation test based on low risky events

Similarly, a Pearson correlation coefficient based on the low risky events was derived. The four variables (VRUCA, FCA, distraction, and lane discipline) were collected depending on the total events, so only seven of the total dependent variables were considered this time. Additionally, the correlation test did not include dependent variables with zero frequency. Table 28 shows the Pearson correlation coefficients between the dependent variables for low severity. Steering had a strong positive correlation with acceleration, $r(16) = .934$, $p < .001$ and deceleration, $r(16) = .754$, $p < .001$.

TABLE 28 Correlation test on low severity dependent variables based on time of the day

Risky events	M	SD	1	2	3	4
1. Acceleration	141.445	239.506	--			
2. Deceleration	139.199	282.006	.496	--		
3. Steering	185.843	244.823	.934**	.754**	--	
4. Fatigue	2.223	7.249	.339	.290	.466	--

** p < .01.

4.2.2.3 Correlation test based on medium risky events

A strong positive linear relation between steering and speeding, $r(16) = .880$, $p < .001$ and steering with acceleration, $r(16) = .825$, $p < .001$, as shown in Table 29 below. Additionally, fatigue with deceleration, $r(16) = .694$, $p = .003$, tailgating with deceleration, $r(16) = .652$, $p = .006$, tailgating with speeding, $r(16) = .585$, $p = .017$, and acceleration with speeding, $r(16) = .520$, $p = .039$ were strongly correlated.

TABLE 29 Correlation test on medium severity dependent variables based on time of the day

Risky events	M	SD	1	2	3	4	5	6	7
1. Speeding	6.375	11.897	--						
2. Acceleration	75.104	151.410	.520*	--					
3. Deceleration	5.766	9.737	.272	-.132	--				
4. Steering	139.902	239.242	.880**	.825**	.254	--			
5. Tailgating	240.348	724.921	.585*	-.020	.652**	.397	--		
6. Overtaking	.015	.038	-.093	-.152	-.101	-.177	-.084	--	
7. Fatigue	.572	1.201	-.155	-.130	.694**	.008	-.029	.154	--

** p < .01.
*p < .05.

4.2.2.4 Correlation test based on high risky events

Table 30 presents the results of a Pearson correlation test based on high-severity events. A strongly positive association has been found between fatigue and tailgating, $r(16) = .996$, $p < .001$, tailgating and speeding, $r(16) = .980$, $p < .001$, fatigue and speeding, $r(16) = .977$, $p < .001$ and steering with deceleration, $r(16) = .964$, $p < .001$.

TABLE 30 Correlation test on high severity dependent variables based on time of the day

Risky events	M	SD	1	2	3	4	5	6	7
1. Speeding	7.017	10.514	--						
2. Acceleration	7.969	19.873	-0.163	--					
3. Deceleration	0.308	0.795	-0.115	-0.154	--				
4. Steering	2.352	5.229	-0.113	-0.126	.964**	--			
5. Tailgating	60.688	202.63	.980**	-0.117	-0.09	-0.089	--		
6. Overtaking	0.019	0.069	-0.022	-0.085	0.021	-0.049	-0.066	--	
7. Fatigue	2.176	7.231	.977**	-0.126	-0.031	-0.042	.996**	-0.079	--

**p < .01.
*p < .05.

4.2.3 Density estimation of risky events

The outputs of kernel density estimation of the risky events based on total, low, medium, and high risky events are presented in the following sections. In addition, the distribution of risky events is explored in relation to the time of the day, weather, and road type.

4.2.3.1 Severity-based distribution of risky events

This section clarifies how risky events are distributed depending on their total, low, medium, and high severity. Thus, KDE was used to analyze all 68775 risky events (the total number of risky events), and the resulting heatmap of the risky events' distribution is presented in figure 15. The heatmap shows that there was a very high density of risky events in Belgium's east (labeled as 'A') and high density in its north regions (labeled as 'B'). The remaining areas (C, D, E, and others) had low densities of risky events. Risky events ranging from 7082 to 9513 were found in the very high-density area, roughly colored red on the heatmap. In contrast, the high-density area included risky events ranging from 4721 to 7081, and the remaining hotspots contained zero to 4720 risky events (low to medium density). Besides, based on the hotspots marked in figure 15, the density distribution of total risky events, from low to high, ranked as E-D-C-B-A.

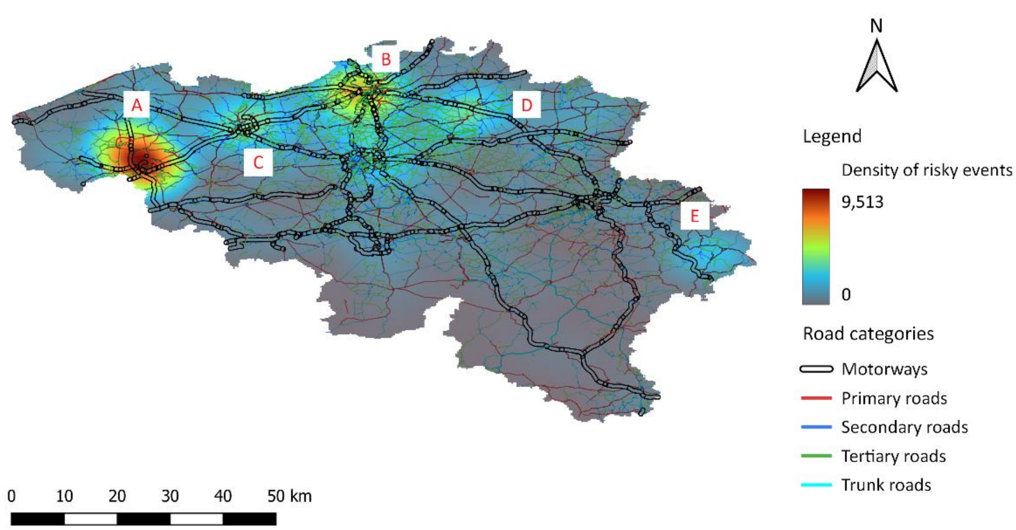


FIGURE 15 Density distribution of total risky events.

Figure 16 displays a heatmap of risky events with low severity, which made up 30089 (44%) of all risky events. Similar to the total risky events, very high densities of risky events were discovered in Belgium's east (designated as 'A') and north (designated as 'B') regions. Hotspots with very high density covered a range of risky events from 3650 to 4868. Also, locations with high density had risky events ranging from 2433 to 3649, while very low to medium density hotspots had risky events ranging from zero to 2432. In addition, the ranking for low to high density is E-C-D-B-A based on the hotspots shown in figure 16.

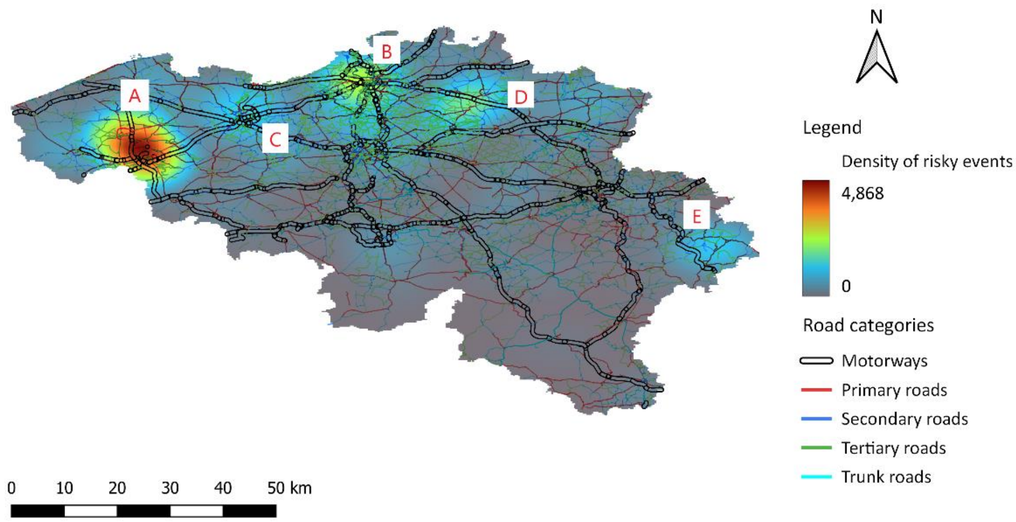


FIGURE 16 Density distribution of low risky events.

Furthermore, the kernel density estimation for events of medium severity includes 30089 (44%) risky events, which resulted in the heatmap depicted in Figure 17. Compared to the results of total and low risky events, the heatmap of medium severity risky events showed one extra hotspot. Similar to high-density hotspots of total risky events, a very high density of risky events was found in the east part of Belgium (location 'A'), and a high density was found in the north part of Belgium (location 'B'). Based on the identified hotspots, the density distribution of risky events of medium severity ranges from E-D-F-C-B-A (low to high). Furthermore, hotspots with very high density included risky events between 2837 and 3784, while high density included between 1891 and 2836, and the remaining hotspots with low to medium density included between zero and 1890 risky events.

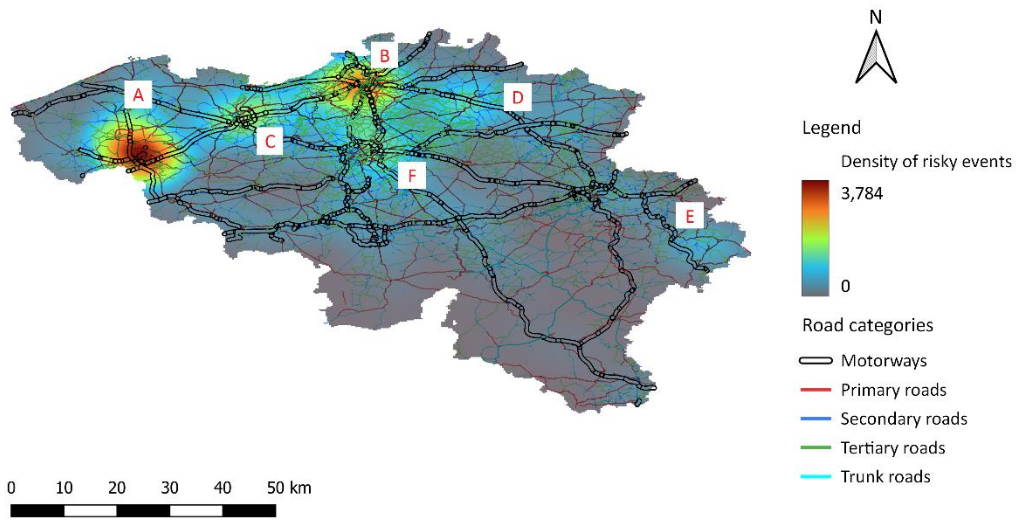


FIGURE 17 Density distribution of medium risky events.

Compared to the density distribution of the risky events covered in the preceding discussion, the kernel density estimation of high risky events exhibited a distinct heatmap. Figure 18 below depicts the hotspot for extremely high density of risky events in Belgium's north (location 'B') and the hotspot for high density of risky events in Belgium's east (location 'A'). In addition, hotspots 'C' and 'F' had medium to high densities of risky events, while the remaining locations had very low to low densities. The range of risky events was 0 to 345 for areas with very low to medium densities, 520 to 693 for very high densities, and 346 to 519 for high densities. Besides, the density hotspots were arranged in hierarchical order from low to high as E-D-F-C-A-B.

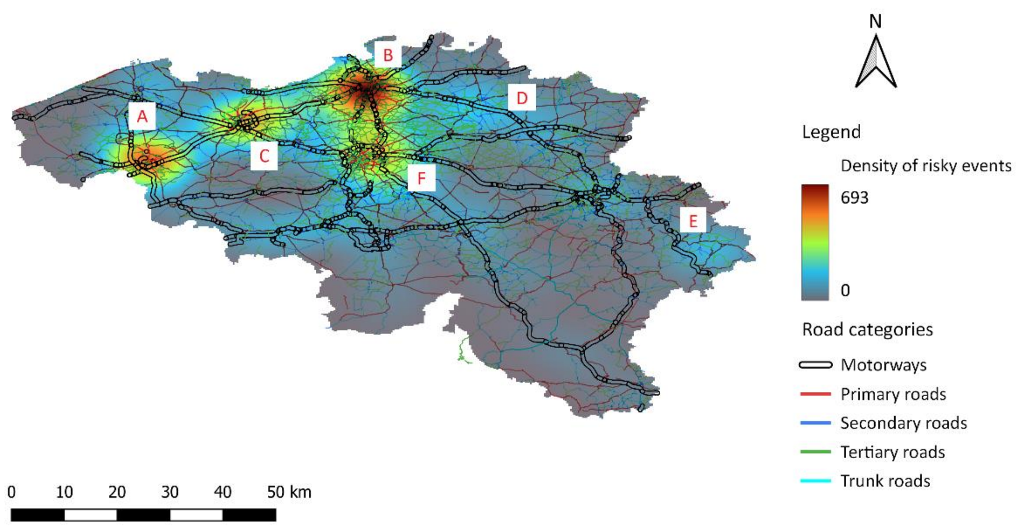


FIGURE 18 Density distribution of high risky events.

4.2.3.2 Distribution of risky events based on severity and characterizing factors

The severity-based risky events kernel density estimation was extracted and further analyzed based on the characterizing factors such as the time of day, the weather, and the type of road. The average density distribution of risky events per kilometer and time of the day is shown in figure 19. The density of risky events in the morning was greater than the remaining clusters of time of the day for the low, medium, and total risky events, except for the high severity. In contrast, the density of risky events for low, medium, and total risky events was nearly the same during the midday and afternoon/evening. The lowest density of risky events was seen in those that occurred at night or early morning across all severity categories. Besides, it was observed that midday has the highest density of risky events with high severity, followed by afternoon/evening. In general, Based on the total risky events, morning (7.68) had the highest average density of risky events per kilometer, followed by afternoon/evening (6.86), midday (6.73), and night/early morning (7.68).

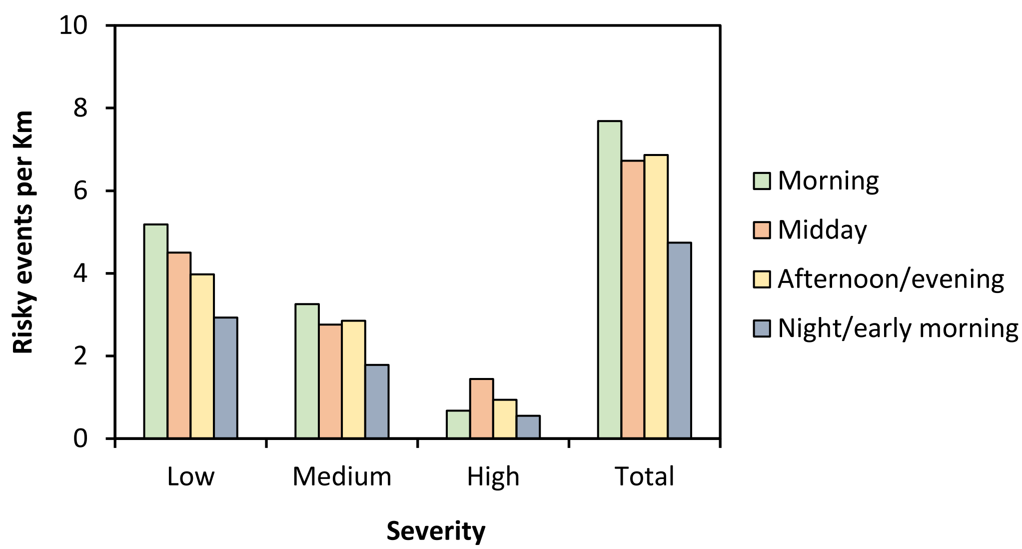


FIGURE 19 Density of risky events based on severity and time of the day.

Figure 20 illustrates the average density of risky events per kilometer based on severity level and weather cluster. Except for high severity, cluster three had the highest average density of risky events per kilometer, followed by clusters two and one, respectively. Cluster two encountered the highest density of risky events in cases of high severity, whereas cluster one experienced the least density of risky events across all categories. Besides, based on the total number of risky events, cluster three (9.73) had the highest average density per kilometer, followed by cluster two (7.46) and cluster one (5.67).

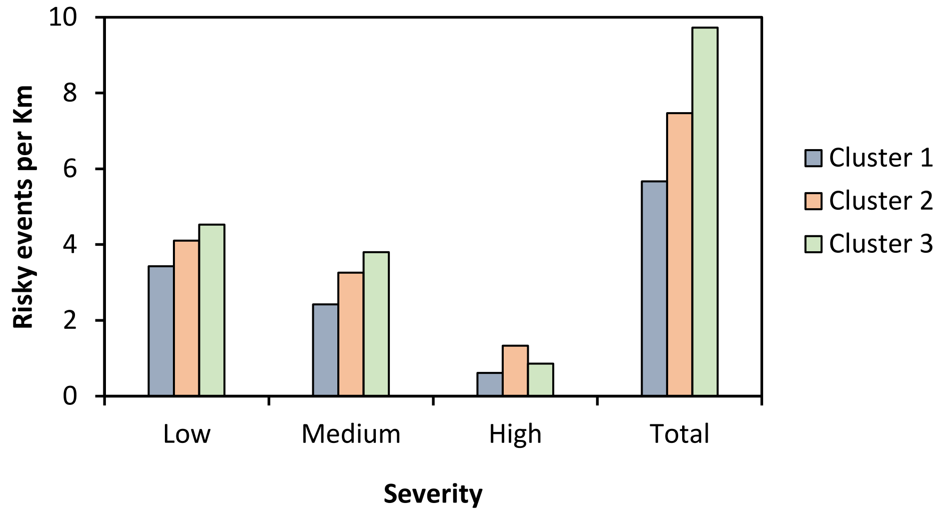


FIGURE 20 Density of risky events based on severity and weather clusters.

Likewise, the density of risky events was characterized according to the type of road, with figure 21 showing the average density of risky events per kilometer for each road type. It was shown that the average density of risky events in tertiary roads was significantly higher than in the other four road types for low, medium, and total risky events. However, it was only barely higher in the case of high severity events. Conversely, in all severity categories, trunk roads had the lowest average density of risky events. Compared to primary, secondary, and trunk roads, the average density of risky events on motorways was significantly greater for low, medium, and total risky events. Besides, a roughly similar average density of risky events occurred across all severity categories on primary and secondary roads. In terms of the total number of risky events, tertiary roads had an average density of 26 events per kilometer, compared to 11.22 on motorways, 5.21 on secondary roads, 5.1 on primary roads, and 3.54 on trunk roads.

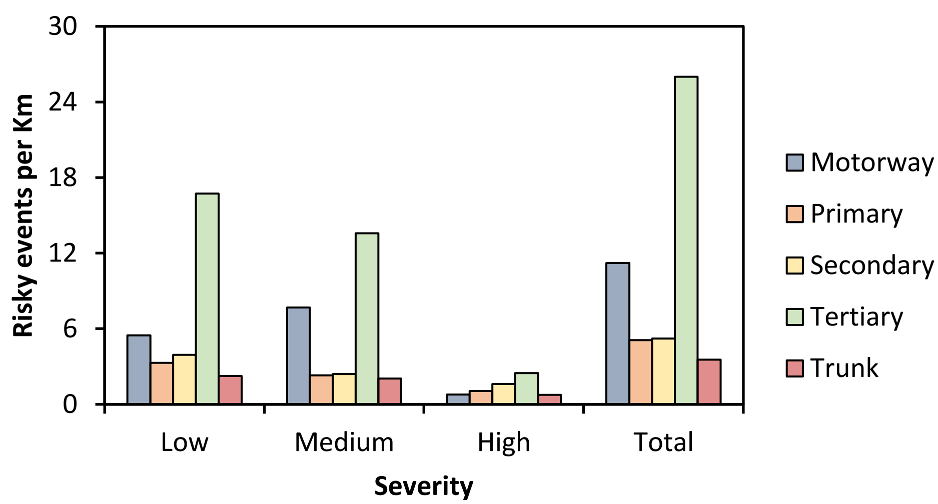


FIGURE 21 Density of risky events based on severity and road type.

5 DISCUSSION

This section presents a brief discussion of the relation between the study's objectives and findings. The discussion emphasized the importance of characteristics including time of day, weather, and type of road for the incidence of risky events and how it accords with previous findings. It also highlighted the interdependence between the risky events.

Previous studies had attempted to characterize risky events based on various variables, including road layout, time of the day, weather, and traffic flow. The severity level, however, was not considered, and the characterization was limited to a small number of risky events. As a result, the main goal of this study was to characterize risky events according to time of day, weather, and type of road. It also contained three risky event severity levels, eleven dependent factors (risky events), and three independent variables (time of the day, weather, and road type). To do this, three stages were taken: firstly, a cluster was formed within each independent variable, and a multivariate analysis of variance was employed to examine group differences; secondly, a bivariate correlation was used to determine the interdependence of the variables; and lastly, a kernel density estimation was done to look at the distribution of the risky events.

In general, the findings of this study indicated that the type of road had a significant influence on several risky events, followed by the weather and the time of day. In the case of total risky events and a medium severity-based correlation test, multiple dependent variables were also more correlated, suggesting a significant likelihood that several risky events may occur simultaneously. The correlated dependent variables were too few for risky events of low severity, possibly because only four dependent variables were recorded for this severity level. Furthermore, in terms of density distribution, cluster three from weather conditions, tertiary roads from the type of road, and morning from time of day had the highest densities of risky events.

5.1 Time-based characterization

The multivariate analysis of risky events yields that the time of day substantially impacted fatigue-related risky events, especially those of medium severity. Truck drivers are also more prone to afternoon or evening fatigue than midday. This result is in line with that of Dingus et al. (2006), which found contrary to what would be anticipated due to circadian rhythm effects, fatigue difficulties are more acute in the afternoon and evening hours. A promising study by Hartley (2000) revealed that fatigue accounts for 36% of crashes before dawn, declining to 4% in the morning, and rising again in the afternoon and evening. In a slightly contradictory finding, Chen and Zhang (2016) found that fatigue events are most common overnight and early morning.

In terms of the distribution of risky events, this study found that morning had the highest density, followed by midday and afternoon/evening, roughly the same, and the overnight/early morning had the lowest density. Although crashes are not the same as risky events, it is apparent that risky situations lead to a crash (existence of a causal relationship between crashes and risky events). Brodie et al. (2009) found that crashes happened most frequently between 10:00 a.m. and 12 noon and between midnight and 2:00 a.m., respectively. The former is consistent with the results of the current study, but the latter does not.

Additionally, this study is congruent with the findings of Offei and Young (2014), which found that most accidents (64%) happened during the daytime.

5.2 Weather-based characterization

Although the multivariate test indicated that weather significantly impacted speeding, fatigue, and vulnerable road user collision avoidance, the between-subject effect (further analysis) on VRUCA didn't reveal any statistically significant differences between the three weather clusters. This study found a statistically significant difference in the total number of speeding incidents between adverse (cluster one) and average (cluster two) weather conditions, with the mean difference indicating that adverse weather conditions had more speeding incidents than average weather conditions. Despite the unfavorable weather conditions in cluster one, this study's findings revealed that drivers frequently exhibited speeding incidents. This may be because drivers tend to drive faster in severe weather. After all, they want to avoid traveling in inclement weather for an extended period. This result is in line with research by Peng et al. (2017), who found a significant speed rate under low visibility conditions. Contrarily, Pahukula et al. (2015) found higher speeding events under clear weather conditions, Zheng et al. (2018) lower speed under icy road surfaces, and Tarko et al. (2011) lower speed under bad weather conditions.

The number of fatigue events was also substantially different between clusters two and three, with the mean difference indicating that cluster three has more fatigue events than cluster two. Cluster three had clear weather compared to the other weather clusters. As a result, long-distance driving in favorable conditions (clear weather) increases the likelihood that drivers may grow weary and lose concentration, necessitating intervention strategies that prevent drivers from becoming fatigued and ensure safe driving. In contrast to the current study's findings, Chen and Zhang (2016) research showed that inclement weather, such as wet pavement and reduced visibility, raises the probability of fatigue incidents.

Besides, density distribution based on weather cluster showed that cluster three (clear weather condition), cluster two (average weather condition), and cluster one (adverse weather condition) had high to low density, respectively. The increased traffic flow during clear weather conditions (Cools et al., 2010) may be related to cluster three's high density of risky events. Both contradictory and congruent results have been found concerning this. According to Offei and Young (2014), clear weather conditions were to blame for 50% of accidents, snowy weather conditions for 20% of collisions, and the remaining 30% were caused by strong winds, fog, rain, dust, and hail. On the other hand, Keay and Simmonds (2005) found that the likelihood of a crash was 0.7 times higher in rainy conditions compared to dry weather.

5.3 Road type based characterization

Speeding, acceleration, deceleration, steering, and tailgating were generally the risky events (dependent variables) that were most significantly influenced by road type. Speeding events were higher on primary and secondary roads than on motorways. Additionally, primary roads experienced more speeding events than trunk roads. This might be due to the speed limit getting higher as you move from local roads to arterials, causing drivers on the latter to go beyond the posted limit. According to Ryan et al. (2019), speeding events on primary roads were considerably more than on motorways. Conversely, it was found that interstate highways had higher speeds (Dong et al., 2015; Xie et al., 2012).

Unlike the primary, secondary, tertiary, and trunk roads, motorways' acceleration events were less frequent. Deceleration events were higher on primary and secondary roads than on motorways. Additionally, it was demonstrated that there were fewer steering events on motorways than on primary, secondary, and tertiary roads and less on trunk roads than on primary roads. Motorways have more restricted access and higher mobility than other types of roads, leading to a lesser likelihood of changing speed and a lower possibility of experiencing acceleration, deceleration, and steering events. Drivers are less likely to change speed in motorways than on other types of roads because motorways have restricted accessibility and higher mobility, enabling drivers to drive at uniform speed for longer distances. This raises the possibility of experiencing fewer reduced acceleration, deceleration, and steering events. In contrast, increased accessibility features like junction proximity or geometry restraints on the other road types will have dizzying effects such as frequent steering effects and speed changes, causing acceleration and deceleration events. According to Gitelman et al. (2018), driving on longer portions without at-grade junctions and improved road conditions reduces braking events, whereas junction proximity or geometric limits raise braking alerts.

Motorways had a higher incidence of tailgating events than primary, secondary, and tertiary roads. The current study also discovered that tailgating events were substantially more common on the trunk than on secondary and tertiary roads. Similar findings were made by Zellmer (2013), showing interstates and highways had a more significant number of tailgating events. Furthermore, tertiary roads and motorways had much higher densities of risky events. Although it may seem clear that risky event occurrence is less likely to occur frequently on tertiary roads due to evidence of reduced traffic flow, the density of risky events per unit of distance traveled was considerably higher.

5.4 Co-occurrence of risky events

The risky event correlation test revealed a higher likelihood that steering events will occur concurrently with either acceleration, deceleration, or speeding events. Depending on a situation, there is a chance that acceleration, deceleration, or speeding events could happen when maneuvering and navigating are necessary. However, using the steering wheel is also necessary to maneuver or navigate the different sections of a road. Due to these events' interdependence, steering may occur in conjunction with acceleration, deceleration, or speeding.

Besides, a strong correlation between speeding and tailgating events was found. When there is a speeding event, which is associated with aggressive driving, there is a chance that drivers will engage in risky behaviors, especially tailgating, because speeding drivers are more likely to lose control of their vehicles. There is also a chance that drivers will show steering events because they will try to avoid the reduced headway with the car in front of them. Additionally, fatigue events were highly correlated with speeding and tailgating events. This result was in line with the findings of Chen and Zhang (2016), indicating that fatigue-related truck crashes were associated with over-speeding (71.07 and 79.98%) and risky following (31.50 and 36.52%). Similarly, Zhou and Zhang (2019) found a positive correlation between overspeeding and fatigue events. Moreover, the current study found that lane discipline and deceleration events were strongly correlated.

5.5 Limitations

The following limitations have affected the process and quality of the current study:

Firstly, although a sample size of approximately 75 truck drivers was intended for this study, only 16 drivers were used, which is too little compared to the expectations. Hence, the sample size does not accurately reflect the population of truck drivers in Belgium, but it can be used as a baseline to determine what kind of patterns emerged using that sample size.

Secondly, baseline data collection (phase one) was conducted between September 2021 and December 2021 (summer to fall). As this period does not include the conditions typified by winter (January to March) and spring (April to June), it may lead to a biased conclusion regarding the pattern of behavior of truck drivers. Similarly, driving behavior may also alter depending on the time of day during various seasons. It is therefore essential to note that the result regarding characterizing risky events based on weather and time of day was confined to summer and fall, which are not all-encompassing seasons.

Thirdly, the data on road type gleaned from available sources lacked sufficient detail and was obscure. Although the data underwent extensive cleaning and processing to the point where it could be used for this study, there were some ambiguous details. In addition, data on weather conditions were retrieved from open sources using Brussels as a national representative. Thus, the lack of information about the weather at the exact location where the risky events occurred could induce a deviation from reality.

The number of risky events has to be standardized depending on trip exposure (trip distance) before characterizing the risky events. While the trip distance for each participant was provided, it was challenging to determine the distance traveled on each cluster. For instance, it was vital to know how far a trip had been on each type of road when it involved multiple roads. An approximate distance was calculated by connecting the timestamps of each risky event because the distance traveled on each type of road was unavailable. This case was also similarly applied for the time of day and weather cases by determining the distance traveled per time of day and weather cluster, respectively. Moreover, it was intended for the current study to incorporate variables for traffic conditions to characterize risky events, but this was not possible owing to a lack of data. Thus a literature review was conducted instead.

5.6 Future research

Future studies can take advantage of the opportunity presented by the findings and limitations of the current study to investigate further how risky events can be characterized. This will deliver detailed information regarding driving behavior patterns under various weather and time of day conditions. The current study evaluated the impact of several road types, including motorways, primary, secondary, tertiary, and trunk roads. However, road layout is a wide notion that can be described in various ways. Therefore, future research can be expanded to encompass aspects of road geometry, including junction proximity or geometry constraints and settlement (urban and rural).

Also, future studies should examine the characteristics of traffic conditions, such as free-flowing and congested traffic, as well as the impact of speed limits on driving behavior. It is also crucial to take the sample size into account. This is because taking sample size into account allows researchers to manage the possibility of reporting a false-negative finding (Type II error) or to predict the precision of the results their experiment will produce (Biau et al., 2008). Therefore, future studies need to include a sufficient number of drivers because a high number of participants lowers the likelihood of a biased result. In addition, the current study used supplementary data from open sources, mainly information about road types and weather conditions. However, to reduce the risk of drawing conclusions that are deviated from reality, future research should use data explicitly collected during the occurrence of risky events instead of acquiring from open sources.

5.7 Implications

The results of this study could be very significant since they could be used to improve the i-DREAMS framework, which is defining, building, testing, and validating a context-aware safety envelope for driving known as the "Safety Tolerance Zone" (STZ). Understanding the time of day when a risky event occurs, the weather when a risky event occurs, and the type of road where a risky event occurs plays a crucial part in defining the environment and driving behavior. The observed characteristics of the risky events will subsequently be used as the foundation for modifying and developing the intervention strategies used in the i-DREAMS project.

This study found several risky events connected to characteristics of the time of day, weather, and types of roads. It also identified the most correlated risky events that could occur concurrently. This will drive future studies to look further into the risky events to reveal other characteristics. Additionally, based on the results of the current study, Belgium's transportation and traffic department may allow universities, research centers, nongovernmental organizations, or private companies to provide creative solutions to lessen the effects of these characteristics found in the study. As was already noted, the i-DREAMS project will use this to modify the intervention technologies and strategies used to bring the drivers into the safety tolerance zone.

6 CONCLUSION

This study aimed to characterize risky events based on three variables: time of day, weather, and road type, with the hope of contributing crucial information to the i-DREAMS project and future research. Prior studies have attempted to assess the risk of driving in different situations based on variables, such as road type, time of day, and environmental conditions, but they only considered a small number of risky events. In addition, previous research did not consider characterizing risky events according to varying degrees of severity. Thus, this paper presented a more comprehensive characterization of risky events by considering three different levels of severity as well as the aggregate of them.

In this study, four clusters of time of the day (morning, midday, afternoon/evening, and night/early morning), three clusters of weather (clear, average, and adverse weather conditions), and five road types (motorway, primary, secondary, tertiary, and trunk roads) were employed. While the five road types were obtained from open sources, the elbow method and K-means clustering were used to determine the number of clusters and cluster centers for the time of day and weather.

The findings showed that the time of day significantly affected fatigue-related risky events, particularly those of medium severity. Also, fatigue events were more frequent in the afternoon or evening than in midday. Besides, the heatmap of risky events demonstrated that morning hours were found to have the highest density of risky events, followed by midday and afternoon/evening, which were found to have approximately the same density, while overnight/early morning had the lowest density.

Weather-based characterization showed that weather significantly impacted the frequency of total risky events and risky events of medium and low severity, particularly on speeding and fatigue events. It was found that there were more speeding events in adverse weather conditions (cluster one) than in average weather conditions (cluster two) and high fatigue events in cluster three (clear weather) than in cluster two. Additionally, the heatmap of the risky events based on weather showed that cluster three (clear weather condition) had the highest density of events, followed by cluster two (average weather condition) and cluster one (adverse weather condition).

Road type-based characterization showed that road type significantly affected total risky events and risky events of a low, medium, and high severity, particularly those involving speeding, acceleration, deceleration, steering, and tailgating events. Speeding events were more frequent on primary than trunk roads and higher on primary and secondary roads than on motorways. Also, primary and secondary roads had more deceleration events than motorways. Besides, motorways had fewer acceleration events than primary, secondary, tertiary, and trunk roads. Steering events were higher on primary, secondary, and tertiary roads than on motorways. Similarly, primary roads had more steering events than trunk roads. There were more tailgating events on primary, trunk, and motorways than tertiary roads. Likewise, tailgating events on motorways and trunk roads were higher than on secondary roads. Furthermore, the heatmap of risky events based on road type revealed that tertiary roads and motorways had a substantially greater density of risky events.

The risky event correlation test showed that steering events were more likely to happen simultaneously with acceleration, deceleration, or speeding events. Also, a strong correlation between speeding and incidents of tailgating was observed. Additionally, there was a possibility that drivers may exhibit steering events while tailgating. Fatigue events were strongly correlated with speeding and tailgating events. The current study also found a strong association between lane discipline and deceleration events.

REFERENCES

Abdel-Aty, M., Ekram, A.-A., Huang, H., & Choi, K. (2011). A study on crashes related to visibility obstruction due to fog and smoke. *Accident Analysis & Prevention*, *43*(5), 1730-1737.

Ahmed, M. M., Franke, R., Ksaibati, K., & Shinstine, D. S. (2018). Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models. *Accident Analysis & Prevention*, *117*, 106-113.

Ahmed, M. M., & Ghasemzadeh, A. (2017). *Exploring the impacts of adverse weather conditions on speed and headway behaviors using the SHRP2 naturalistic driving study data*.

Ahmed, M. M., & Ghasemzadeh, A. (2018). The impacts of heavy rain on speed and headway Behaviors: An investigation using the SHRP2 naturalistic driving study data. *Transportation research part C: emerging technologies*, *91*, 371-384. https://doi.org/10.1016/j.trc.2018.04.012

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, *41*(3), 359-364.

Andrey, J., Mills, B., Leahy, M., & Suggett, J. (2003). Weather as a chronic hazard for road transportation in Canadian cities. *Natural hazards*, *28*(2), 319-343.

Andy, F. (2009). Discovering statistics using SPSS. In.

Anund, A., Fors, C., & Ahlstrom, C. (2017). The severity of driver fatigue in terms of line crossing: a pilot study comparing day-and night time driving in simulator. *European Transport Research Review*, *9*(2), 1-7.

Arumugam, S., & Bhargavi, R. (2019). A survey on driving behavior analysis in usage based insurance using big data. *Journal of Big Data*, *6*(1), 1-21.

Barnard, Y., Utesch, F., van Nes, N., Eenink, R., & Baumann, M. (2016). The study design of UDRIVE: the naturalistic driving study across Europe for cars, trucks and scooters. *European Transport Research Review*, *8*(2), 1-10.

Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, *105*(9).

Biau, D. J., Kernéis, S., & Porcher, R. (2008). Statistics in brief: the importance of sample size in the planning and interpretation of medical research. *Clinical orthopaedics and related research*, *466*(9), 2282-2288.

Brijs, K., Brijs, T., Ross, V., Donders, E., Vanrompay, Y., Wets, G., Dirix, H., Katrakazas, C., Yannis, G., & Kaiser, S. (2020). D3. 3 Toolbox of recommended interventions to assist drivers in maintaining a safety tolerance zone.

Brijs, T., Brijs, K., Kaiser, S., Talbot, R., Lourenco, A., Antoniou, C., Yannis, G., Avenoso, A., & Wets, G. (2020). i-DREAMS: an intelligent driver and road environment assessment and monitoring system.

Brodie, L., Lyndal, B., & Elias, I. J. (2009). Heavy vehicle driver fatalities: Learning's from fatal road crash investigations in Victoria. *Accident Analysis & Prevention*, *41*(3), 557-564.

Chan, C.-Y. (2017). Advancements, prospects, and impacts of automated driving systems. *International journal of transportation science and technology*, *6*(3), 208-216.

Chang, L.-Y., & Chien, J.-T. (2013). Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. *Safety science*, *51*(1), 17-22.

Chen, C., & Zhang, J. (2016). Exploring background risk factors for fatigue crashes involving truck drivers on regional roadway networks: a case control study in Jiangxi and Shaanxi, China. *SpringerPlus*, *5*(1). https://doi.org/10.1186/s40064-016-2261-y

Chen, F., & Chen, S. (2011). Injury severities of truck drivers in single-and multi-vehicle accidents on rural highways. *Accident Analysis & Prevention*, *43*(5), 1677-1688.

Chiou, Y.-C., Sheng, Y.-C., & Fu, C. (2017). Freeway crash frequency modeling under time-of-day distribution. *Transportation research procedia*, *25*, 664-676. https://doi.org/10.1016/j.trpro.2017.05.450

Chipman, M., & Jin, Y. L. (2009). Drowsy drivers: The effect of light and circadian rhythm on crash occurrence. *Safety science*, *47*(10), 1364-1370.

Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *Weather, Climate, and Society*, *2*(1), 60-68.

Das, A., Ghasemzadeh, A., & Ahmed, M. M. (2019). Analyzing the effect of fog weather conditions on driver lane-keeping performance using the SHRP2 naturalistic driving study data. *Journal of safety research*, *68*, 71-80. https://doi.org/10.1016/j.jsr.2018.12.015

Dingus, T. A., Neale, V. L., Klauer, S. G., Petersen, A. D., & Carroll, R. J. (2006). The development of a naturalistic data collection system to perform critical incident analysis: An investigation of safety and fatigue issues in long-haul trucking. *Accident Analysis & Prevention*, *38*(6), 1127-1136.

DIVA-GIS. (2022). *Spatial Data Download*. Retrieved May 5 from https://www.diva-gis.org/datadown

Dong, C., Clarke, D. B., Richards, S. H., & Huang, B. (2014). Differences in passenger car and large truck involved crash frequencies at urban signalized intersections: An exploratory analysis. *Accident Analysis & Prevention*, *62*, 87-94.

Dong, C., Richards, S. H., Huang, B., & Jiang, X. (2015). Identifying the factors contributing to the severity of truck-involved crashes. *International journal of injury control and safety promotion*, *22*(2), 116-126.

European Commission. (2018). Annual Accident Report 2018. *European Commission, Directorate General for Transport*. https://ec.europa.eu/transport/road_safety/system/files/2021-07/asr2018.pdf

Fitzharris, M., Liu, S., Stephens, A. N., & Lenné, M. G. (2017). The relative importance of real-time in-cab and external feedback in managing fatigue in real-world commercial transport operations. *Traffic injury prevention*, *18*(sup1), S71-S78.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative geography: perspectives on spatial data analysis*. Sage.

Gates, J., Dubois, S., Mullen, N., Weaver, B., & Bédard, M. (2013). The influence of stimulants on truck driver crash responsibility in fatal crashes. *Forensic science international*, *228*(1-3), 15-20.

Gitelman, V., Bekhor, S., Doveh, E., Pesahov, F., Carmel, R., & Morik, S. (2018). Exploring relationships between driving events identified by in-vehicle data recorders, infrastructure characteristics and road crashes. *Transportation research part C: emerging technologies*, *91*, 156-175.

Golob, T. F., Recker, W., & Pavlis, Y. (2008). Probabilistic models of freeway safety performance using traffic flow data as predictors. *Safety science*, *46*(9), 1306-1333.

Golob, T. F., & Recker, W. W. (2003). Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. *Journal of transportation engineering*, *129*(4), 342-353.

Guo, F., Wang, X., & Abdel-Aty, M. A. (2010). Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis & Prevention*, *42*(1), 84-92. https://doi.org/10.1016/j.aap.2009.07.005

Häkkänen, H., & Summala, H. (2001). Fatal traffic accidents among trailer truck drivers and accident causes as viewed by other truck drivers. *Accident Analysis & Prevention*, *33*(2), 187-196.

Harb, R., Radwan, E., Yan, X., & Abdel-Aty, M. (2007). Light truck vehicles (LTVs) contribution to rear-end collisions. *Accident Analysis & Prevention*, *39*(5), 1026-1036. https://doi.org/10.1016/j.aap.2007.01.007

Harirforoush, H., & Bellalite, L. (2019). A new integrated GIS-based analysis to detect hotspots: a case study of the city of Sherbrooke. *Accident Analysis & Prevention*, *130*, 62-74.

Hartley, L. (2000). Fatigue and driving. *International Encyclopedia of Ergonomics and Human Factors-3 Volume Set*, 446.

Hassan, H. M., Sarhan, M., Eng, P., Garib, A., & Hussain Al Harthei, M. (2016). Drivers' Time Headway Characteristics and Factors Affecting Tailgating Crashes.

Hermans, E., Brijs, T., Stiers, T., & Offermans, C. (2006). The impact of weather conditions on road safety investigated on an hourly basis. Washington, DC: Proceedings of the 85thAnnual meeting of the Transportation Research Board,

i-DREAMS project. (2020). *Presentations*. Retrieved April 11 from https://idreamsproject.eu/wp/presentations/

Irwanto, I., Purwananto, Y., & Soelaiman, R. (2012). Optimasi Kinerja Algoritma Klasterisasi K-Means untuk Kuantisasi Warna Citra. *Jurnal Teknik ITS*, *1*(1), A197-A202.

Islam, M., & Hernandez, S. (2013). Large truck–involved crashes: Exploratory injury severity analysis. *Journal of transportation engineering*, *139*(6), 596-604.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jansen, R. J., & Wesseling, S. (2018). Harsh Braking by Truck Drivers: A Comparison of Thresholds and Driving Contexts Using Naturalistic Driving Data. Proceedings of the 6th Humanist Conference, The Hague, The Netherlands,

Kaiser, S., Eichhorn, A., Aigner-Breuss, E., Pracherstorfer, C., Katrakazas, C., Michelaraki, E., Yannis, G., Pilkington-Cheney, F., Talbot, R., & Hancox, G. (2020). D2. 1 State of the art on monitoring driver state and task demand.

Katrakazas, C., Michelaraki, E., Yannis, G., Kaiser, S., Brijs, K., Ross, V., Dirix, H., Neven, A., Paul, R., & Donders, E. (2020). D2. 2 Technologies for safety interventions and assessment of their effectiveness.

Katrakazas, C., Michelaraki, E., Yannis, G., Kaiser, S., Senitschnig, N., Ross, V., Adnan, M., Brijs, K., Brijs, T., & Talbot, R. (2020). D3. 2 Toolbox of recommended data collection tools and monitoring methods and a conceptual definition of the Safety Tolerance Zone.

Kazemi-Beydokhti, M., Ali Abbaspour, R., & Mojarab, M. (2017). Spatio-temporal modeling of seismic provinces of Iran using DBSCAN algorithm. *Pure and Applied Geophysics*, *174*(5), 1937-1952.

Keay, K., & Simmonds, I. (2005). The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. *Accident Analysis & Prevention*, *37*(1), 109-124.

Ketabi, D., Barkhordari, A., Mirmohammadi, S. J., & Mehrparvar, A. H. (2011). Aberrant behaviors and road accidents among Iranian truck drivers, 2010. *Health promotion perspectives*, *1*(2), 130.

Khattak, A. J., & Knapp, K. K. (2001). Interstate highway crash injuries during winter snow and nonsnow events. *Transportation Research Record*, *1746*(1), 30-36.

Khorashadi, A., Niemeier, D., Shankar, V., & Mannering, F. (2005). Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. *Accident Analysis & Prevention*, *37*(5), 910-921.

Kilpeläinen, M., & Summala, H. (2007). Effects of weather and weather forecasts on driver behaviour. *Transportation research part F: traffic psychology and behaviour*, *10*(4), 288-299.

Komackova, L., & Poliak, M. (2016). Factors affecting the road safety. *Journal of Communication and Computer*, *13*, 146-152.

Kumar, S., & Toshniwal, D. (2015). A data mining framework to analyze road accident data. *Journal of Big Data*, *2*(1). https://doi.org/10.1186/s40537-015-0035-y

Lemp, J. D., Kockelman, K. M., & Unnikrishnan, A. (2011). Analysis of large truck crash severity using heteroskedastic ordered probit models. *Accident Analysis & Prevention*, *43*(1), 370-380.

Li, X., Goldberg, D. W., Chu, T., & Ma, A. (2019). Enhancing driving safety: Discovering individualized hazardous driving scenes using GIS and mobile sensing. *Transactions in GIS*, *23*(3), 538-557.

Lombardi, D. A., Horrey, W. J., & Courtney, T. K. (2017). Age-related differences in fatal intersection crashes in the United States. *Accident Analysis & Prevention*, *99*, 20-29.

Mahmud, S. S., Ferreira, L., Hoque, M. S., & Tavassoli, A. (2021). Overtaking risk modeling in two-lane two-way highway with heterogeneous traffic environment of a low-income country using naturalistic driving dataset. *Journal of safety research*.

Malygin, I., Komashinskiy, V., & Korolev, O. (2018). Cognitive technologies for providing road traffic safety in intelligent transport systems. *Transportation research procedia*, *36*, 487-492.

McCartt, A. T., Rohrbaugh, J. W., Hammer, M. C., & Fuller, S. Z. (2000). Factors associated with falling asleep at the wheel among long-distance truck drivers. *Accident Analysis & Prevention*, *32*(4), 493-504.

Melinder, K. (2007). Socio-cultural characteristics of high versus low risk societies regarding road traffic safety. *Safety science*, *45*(3), 397-414.

Mohaymany, A. S., Shahri, M., & Mirbagheri, B. (2013). GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. *Geo-spatial Information Science*, *16*(2), 113-119.

Nagy, E., & Sandor, Z. (2012). Overtaking ban for heavy goods vehicle in Hungary on the national motorway network. *Pollack Periodica*, *7*(1), 83-95.

NASA POWER data access viewer. (2022). *Higher resolution daily time series, renewable energy community* Retrieved April 29 from https://power.larc.nasa.gov/data-access-viewer/

Nirmal, S. (2019). Comparative study between k-means and k-medoids clustering algorithms. *J. Classif*, *6*, 839-844.

Offei, E., & Young, R. (2014). *Quantifying the impact of large percent trucks proportion on rural freeways*.

Osman, M., Paleti, R., Mishra, S., & Golias, M. M. (2016). Analysis of injury severity of large truck crashes in work zones. *Accident Analysis & Prevention*, *97*, 261-273.

Pahukula, J., Hernandez, S., & Unnikrishnan, A. (2015). A time of day analysis of crashes involving large trucks in urban areas. *Accident Analysis & Prevention*, *75*, 155-163.

Paredes, P. E., Zhou, Y., Hamdan, N. A.-H., Balters, S., Murnane, E., Ju, W., & Landay, J. A. (2018). Just breathe: In-car interventions for guided slow breathing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(1), 1-23.

Peng, Y., Abdel-Aty, M., Shi, Q., & Yu, R. (2017). Assessing the impact of reduced visibility on traffic crash risk using microscopic data and surrogate safety measures. *Transportation research part C: emerging technologies*, *74*, 295-305.

Peterson, T. C., McGuirk, M., Houston, T. G., Horvitz, A. H., & Wehner, M. F. (2008). Climate variability and change with implications for transportation. *Transportation Research Board*, *90*(2.3).

Pljakić, M., Jovanović, D., Matović, B., & Mićić, S. (2019). Identification of Accident Hotspot Locations Using Network Kernel Density. In: September.

Pokorny, P., Drescher, J., Pitera, K., & Jonsson, T. (2017). Accidents between freight vehicles and bicycles, with a focus on urban areas. *Transportation research procedia*, *25*, 999-1007.

Qiu, L., & Nixon, W. A. (2008). Effects of adverse weather on traffic crashes: systematic review and meta-analysis. *Transportation Research Record*, *2055*(1), 139-146.

Qu, Y., Lin, Z., Li, H., & Zhang, X. (2019). Feature recognition of urban road traffic accidents based on GA-XGBoost in the context of big data. *IEEE Access*, *7*, 170106-170115.

Ratner, B. (2009). The correlation coefficient: Its values range between+ 1/− 1, or do they? *Journal of targeting, measurement and analysis for marketing*, *17*(2), 139-142.

Ryan, C., Murphy, F., & Mullins, M. (2019). Semiautonomous vehicle risk analysis: A telematics-based anomaly detection approach. *Risk analysis*, *39*(5), 1125-1140.

Shi, Q., Abdel-Aty, M., & Lee, J. (2016). A Bayesian ridge regression analysis of congestion's impact on urban expressway safety. *Accident Analysis & Prevention*, *88*, 124-137. https://doi.org/10.1016/j.aap.2015.12.001

Simon, M., Hermitte, T., & Page, Y. (2009). Intersection road accident causation: A European view. 21st International Technical Conference on the Enhanced Safety of Vehicles,

Tarko, A. P., Anastasopoulos, P. C., & Zuriaga, A. M. P. (2011). Can education and enforcement affect behavior of car and truck drivers on urban freeways. International Conference on Road Safety and Simulation,

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

Ting, P.-H., Hwang, J.-R., Doong, J.-L., & Jeng, M.-C. (2008). Driver fatigue and highway driving: A simulator study. *Physiology & behavior*, *94*(3), 448-453. https://doi.org/10.1016/j.physbeh.2008.02.015

Tseng, C.-M., Yeh, M.-S., Tseng, L.-Y., Liu, H.-H., & Lee, M.-C. (2016). A comprehensive analysis of factors leading to speeding offenses among large-truck drivers. *Transportation research part F: traffic psychology and behaviour*, *38*, 171-181.

Uddin, M., & Huynh, N. (2017). Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. *Accident Analysis & Prevention*, *108*, 44-55.

Wang, C., Quddus, M., & Ison, S. (2013). A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK. *Transportmetrica A: Transport Science*, *9*(2), 124-148.

Wang, Y., & Prato, C. G. (2019). Determinants of injury severity for truck crashes on mountain expressways in China: A case-study with a partial proportional odds model. *Safety science*, *117*, 100-107.

World Health Organization [WHO]. (2018). *Global status report on road safety 2018: summary*.

Wu, C., Sun, C., Chu, D., Huang, Z., Ma, J., & Li, H. (2016). Clustering of Several Typical Behavioral Characteristics of Commercial Vehicle Drivers Based on GPS Data Mining: Case Study of Highways in China. *Transportation Research Record: Journal of the Transportation Research Board*, *2581*(1), 154-163. https://doi.org/10.3141/2581-18

Wu, K. F., & Thor, C. P. (2015). Method for the use of naturalistic driving study data to analyze rear-end crash sequences. *Transportation Research Record*, *2518*(1), 27-36.

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

Yang, B., Guo, Y., & Xu, C. (2019). Analysis of freeway secondary crashes with a two-step method by loop detector data. *IEEE Access*, *7*, 22884-22890.

Yang, L., Li, X., Guan, W., Zhang, H. M., & Fan, L. (2018). Effect of traffic density on drivers' lane change and overtaking maneuvers in freeway situation—A driving simulator–based study. *Traffic injury prevention*, *19*(6), 594-600.

Yuan, Y., Yang, M., Guo, Y., Rasouli, S., Gan, Z., & Ren, Y. (2021). Risk factors associated with truck-involved fatal crash severity: Analyzing their impact for different groups of truck drivers. *Journal of safety research*, *76*, 154-165.

Zellmer, T. (2013). *Development of stationary and mobile tailgating detection solutions for ground vehicles* Clemson University].

Zhao, P., & Lee, C. (2018). Assessing rear-end collision risk of cars and heavy vehicles on freeways using a surrogate safety measure. *Accident Analysis & Prevention*, *113*, 149-158.

Zheng, Z. (2012). Empirical analysis on relationship between traffic conditions and crash occurrences. *Procedia-Social and Behavioral Sciences*, *43*, 302-312.

Zheng, Z., Lu, P., & Lantz, B. (2018). Commercial truck crash injury severity analysis using gradient boosting data mining model. *Journal of safety research*, *65*, 115-124. https://doi.org/10.1016/j.jsr.2018.03.002

Zhou, T., & Zhang, J. (2019). Analysis of commercial truck drivers' potentially dangerous driving behaviors based on 11-month digital tachograph data and multilevel modeling approach. *Accident Analysis & Prevention*, *132*, 105256. https://doi.org/10.1016/j.aap.2019.105256

Zhu, X., & Srinivasan, S. (2011). A comprehensive analysis of factors influencing the injury severity of large-truck crashes. *Accident Analysis & Prevention*, *43*(1), 49-57.

Zou, W., Wang, X., & Zhang, D. (2017). Truck crash severity in New York city: an investigation of the spatial and the time of day effects. *Accident Analysis & Prevention*, *99*, 249-261.

APPENDICES

Appendix-I

K-means clustering: Time of the day

a. Initial cluster centers

	1	2	3	4
Timestamp	0:06:50	15:27:39	23:08:07	7:47:14

b. **Final cluster centers**

	1	2	3	4
Timestamp	4:34:38	12:24:48	16:18:45	8:37:06

c. **Number of cases in each cluster**

Cluster	1	17028
	2	17710
	3	14890
	4	19147
Valid		68775
Missing		0

d. ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Timestamp	5.310E+12	3	19113456.564	68771	277805.551	.000

K-means clustering: Weather

a. Initial cluster centers

	Temperature	Dew point	Humidity	Precipitation	Pressure	Wind
1	-3.77	-3.78	100	0	102.06	2.52
2	18.72	14.73	77.5	.54	100.76	8.93
3	18.6	7.14	47.12	0	102.5	1.91

b. Final cluster centers

	Temperature	Dew point	Humidity	Precipitation	Pressure	Wind
1	5.31	4.63	95.22	.1	101.2	4.61
2	10.03	7.04	81.67	.11	101.44	5.2
3	16.22	9.25	63.57	.04	101.69	5.51

c. Number of cases in each Cluster

Cluster	1	44586
	2	14666
	3	9523
Valid		68775
Missing		0

d. ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Temperature	512549.858	2	13.802	68772	37134.609	.000
Dew point	98158.388	2	13.742	68772	7143.047	.000
Humidity	4296642.408	2	17.172	68772	250217.023	.000
Precipitation	12.999	2	.039	68772	332.568	<.001
Pressure	1049.804	2	1.454	68772	722.199	.000
Wind	4213.955	2	4.976	68772	846.883	.000

Appendix - II

MANOVA: Time of the day

a. Total risky events

Box's Test of Equality of Covariance Matrices	
Box's M	269.049
F	1.566
df1	110
df2	4399.822
Sig.	<.001

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
Time of the day	Pillai's Trace	.576	1.189	30	150	.247
	Wilks' Lambda	.514	1.199	30	141.6	.238
	Hotelling's Trace	.775	1.206	30	140	.233
	Roy's Largest Root	.460	2.301c	10	50	.026

b. Low risky events

Box's Test of Equality of Covariance Matrices	
Box's M	57.827
F	1.676
df1	30
df2	8324.745
Sig.	.012

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
Time of the day	Pillai's Trace	.202	1.011	12	168	.441
	Wilks' Lambda	.805	1.017	12	143.2	.436
	Hotelling's Trace	.232	1.020	12	158	.433
	Roy's Largest Root	.182	2.552c	4	56	.049

c. Medium risky events

Box's Test of Equality of Covariance Matrices	
Box's M	76.113
F	1.93
df1	28
df2	2301.198
Sig.	.002

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
Time of the day	Pillai's Trace	.558	1.730	21	159	.031
	Wilks' Lambda	.517	1.809	21	147	.022
	Hotelling's Trace	.795	1.881	21	149	.016

Roy's Largest Root	.573	4.340c	7	53	<.001

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Time of the day (Medium risky events)	Speeding	1.068	3	.356	2.107	.109
	Acceleration	28.649	3	9.550	.270	.847
	Deceleration	1.188	3	.396	.774	.513
	Steering	91.852	3	30.617	.198	.897
	Tailgating	2060.451	3	686.817	2.498	.069
	Overtaking	.001	3	.000	1.761	.165
	Fatigue	.167	3	.056	5.750	.002

Multiple comparisons

Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Fatigue	Morning	Night/early morning	.014	.030	.964	-.067	.096
		Midday	.036	.028	.581	-.040	.113
		Afternoon/evening	-.107	.046	.128	-.237	.023
	Midday	Night/early morning	-.022	.028	.861	-.097	.054
		Afternoon/evening	-.143	.044	.024	-.270	-.016
		Morning	-.036	.028	.581	-.113	.040
	Afternoon/ evening	Night/early morning	.121	.046	.070	-.008	.250
		Midday	.143	.044	.024	.016	.270
		Morning	.107	.046	.128	-.023	.237
	Night/early morning	Midday	.022	.028	.861	-.054	.097
		Afternoon/evening	-.121	.046	.070	-.250	.008
		Morning	-.014	.030	.964	-.096	.067

d. High risky events

Box's Test of Equality of Covariance Matrices	
Box's M	101.765
F	2.581
df1	28
df2	2301.198
Sig.	<.001

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
Time of the day	Pillai's Trace	.307	.864	21	159	.637
	Wilks' Lambda	.713	.876	21	146.995	.622
	Hotelling's Trace	.375	.887	21	149	.608
	Roy's Largest Root	.280	2.122c	7	53	.057

MANOVA: Weather

a. Total risky events

Box's Test of Equality of Covariance Matrices	
Box's M	75.076
F	.866
df1	55
df2	2906.379
Sig.	.749

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
Weather	Pillai's Trace	.781	1.793	20	56	.045
	Wilks' Lambda	.359	1.805	20	54	.044
	Hotelling's Trace	1.394	1.813	20	52	.044
	Roy's Largest Root	1.008	2.822	10	28	.015

Tests of Between-Subjects Effects						
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Weather	Speeding	7.922	2	3.961	3.542	.039
	Acceleration	1623.171	2	811.586	2.533	.093
	Deceleration	140.841	2	70.421	1.825	.176
	Steering	3161.359	2	1580.679	1.718	.194
	Tailgating	750.713	2	375.357	.833	.443
	Overtaking	.001	2	.001	.142	.868
	VRUCA	.359	2	.180	3.443	.043
	FCA	.490	2	.245	.629	.539
	Fatigue	.019	2	.009	.606	.551
	Distraction	6.457	2	3.229	.067	.935

Multiple Comparisons							
Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Speeding	Cluster 1	Cluster 2	.987	.374	.032	.073	1.900

		Cluster 3	.648	.479	.377	-.524	1.819
	Cluster 2	Cluster 1	-.987	.374	.032	-1.900	-.073
		Cluster 3	-.339	.479	.761	-1.510	.833
	Cluster 3	Cluster 1	-.648	.479	.377	-1.819	.524
		Cluster 2	.339	.479	.761	-.833	1.510
VRUCA	Cluster 1	Cluster 2	.188	.081	.065	-.009	.385
		Cluster 3	-.022	.104	.976	-.275	.232
	Cluster 2	Cluster 1	-.188	.081	.065	-.385	.009
		Cluster 3	-.210	.104	.121	-.463	.044
	Cluster 3	Cluster 1	.022	.104	.976	-.232	.275
		Cluster 2	.210	.104	.121	-.044	.463

b. Low risky events

Box's Test of Equality of Covariance Matrices	
Box's M	36.258
F	1.433
df1	20
df2	1394.04
Sig.	0.097

Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Weather	Pillai's Trace	.432	2.343	8	68	.027
	Wilks' Lambda	.576	2.619	8	66	.015
	Hotelling's Trace	.721	2.885	8	64	.008
	Roy's Largest Root	.701	5.955	4	34	<.001

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Weather	Acceleration	497.193	2	248.596	1.901	0.164
	Deceleration	132.831	2	66.415	1.894	0.165
	Steering	1184.25	2	592.125	2.277	0.117
	Fatigue	0.133	2	0.066	3.303	0.048

Multiple comparisons

Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Fatigue	Cluster 1	Cluster 2	.078	.050	.280	-.045	.200
		Cluster 3	-.082	.064	.414	-.239	.075
	Cluster 2	Cluster 1	-.078	.050	.280	-.200	.045
		Cluster 3	-.160	.064	.045	-.317	-.003
	Cluster 3	Cluster 1	.082	.064	.414	-.075	.239
		Cluster 2	.160	.064	.045	.003	.317

c. Medium risky events

Box's Test of Equality of Covariance Matrices	
Box's M	60.542
F	.976
df1	42
df2	1206.779
Sig.	.517

Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Weather	Pillai's Trace	.595	2.257	12	64	.019
	Wilks' Lambda	.463	2.426	12	62	.012
	Hotelling's Trace	1.035	2.586	12	60	.008
	Roy's Largest Root	.895	4.774	6	32	.001

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Weather	Speeding	.745	2	.373	2.776	.076
	Acceleration	224.996	2	112.498	2.889	.069
	Deceleration	.141	2	.071	.120	.888
	Steering	489.132	2	244.566	1.165	.323
	Tailgating	742.453	2	371.227	1.389	.262
	Fatigue	.082	2	.041	5.158	.011

Multiple Comparisons

Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Fatigue	Cluster 1	Cluster 2	.092	.032	.017	.015	.169
		Cluster 3	-.005	.040	.991	-.104	.094
	Cluster 2	Cluster 1	-.092	.032	.017	-.169	-.015
		Cluster 3	-.097	.040	.056	-.196	.002
	Cluster 3	Cluster 1	.005	.040	.991	-.094	.104
		Cluster 2	.097	.040	.056	-.002	.196

d. High risky events

Box's Test of Equality of Covariance Matrices	
Box's M	57.15
F	1.521
df1	28
df2	3136.116
Sig.	.039

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
Weather	Pillai's Trace	.519	1.55	14	62	.120
	Wilks' Lambda	.545	1.522	14	60	.131
	Hotelling's Trace	.720	1.492	14	58	.143
	Roy's Largest Root	.478	2.116	7	31	.071

MANOVA: Road type

a. Total risky events

Box's Test of Equality of Covariance Matrices	
Box's M	665.265
F	2.781
df1	165
df2	7804.744
Sig.	<.001

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
Road type	Pillai's Trace	0.982	2.245	40	276	<.001
	Wilks' Lambda	0.28	2.516	40	252.12	<.001
	Hotelling's Trace	1.748	2.819	40	258	<.001
	Roy's Largest Root	1.229	8.481	10	69	<.001

Tests of Between-Subjects Effects						
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Road type	Speeding	843.697	4	210.924	3.851	.007
	Acceleration	46260.04	4	11565.01	5.689	<.001
	Deceleration	40575.81	4	10143.95	3.571	.010
	Steering	267240.2	4	66810.06	6.857	<.001
	Tailgating	28425.16	4	7106.29	10.708	<.001
	Overtaking	.203	4	.051	0.678	.609
	Fatigue	1.554	4	.389	0.766	.551
	Distraction	9.669	4	2.417	0.824	.514
	FCA	3.706	4	.927	1.927	.115
	Lane discipline	980.096	4	245.024	1.347	.261

Multiple Comparisons							
Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound

Speeding	Motorway	Primary	-8.360	2.232	.013	-15.191	-1.528
		Secondary	-8.016	2.553	.043	-15.843	-.189
		Tertiary	-2.728	1.592	.452	-7.561	2.105
		Trunk	-3.050	1.908	.518	-8.869	2.770
	Primary	Motorway	8.360	2.232	.013	1.528	15.191
		Secondary	.344	3.336	1.000	-9.344	10.031
		Tertiary	5.632	2.673	.247	-2.179	13.443
		Trunk	5.310	2.873	.366	-3.036	13.656
	Secondary	Motorway	8.016	2.553	.043	.189	15.843
		Primary	-.344	3.336	1.000	-10.031	9.344
		Tertiary	5.288	2.946	.399	-3.369	13.946
		Trunk	4.967	3.128	.517	-4.155	14.088
	Tertiary	Motorway	2.728	1.592	.452	-2.105	7.561
		Primary	-5.632	2.673	.247	-13.443	2.179
		Secondary	-5.288	2.946	.399	-13.946	3.369
		Trunk	-.322	2.409	1.000	-7.325	6.681
	Trunk	Motorway	3.050	1.908	.518	-2.770	8.869
		Primary	-5.310	2.873	.366	-13.656	3.036
		Secondary	-4.967	3.128	.517	-14.088	4.155
		Tertiary	.322	2.409	1.000	-6.681	7.325
Acceleration	Motorway	Primary	-72.120	14.058	<.001	-115.269	-28.971
		Secondary	-53.258	11.721	.003	-89.140	-17.376
		Tertiary	-43.514	14.607	.059	-88.366	1.338
		Trunk	-31.546	10.077	.043	-62.307	-.785
	Primary	Motorway	72.120	14.058	<.001	28.971	115.269
		Secondary	18.862	18.036	.832	-33.563	71.288
		Tertiary	28.606	20.032	.615	-29.504	86.716
		Trunk	40.574	17.013	.150	-9.114	90.262
	Secondary	Motorway	53.258	11.721	.003	17.376	89.140
		Primary	-18.862	18.036	.832	-71.288	33.563
		Tertiary	9.744	18.466	.984	-43.985	63.473
		Trunk	21.712	15.139	.611	-22.267	65.691
	Tertiary	Motorway	43.514	14.607	.059	-1.338	88.366
		Primary	-28.606	20.032	.615	-86.716	29.504
		Secondary	-9.744	18.466	.984	-63.473	43.985
		Trunk	11.968	17.469	.958	-39.128	63.064
	Trunk	Motorway	31.546	10.077	.043	.785	62.307
		Primary	-40.574	17.013	.150	-90.262	9.114
		Secondary	-21.712	15.139	.611	-65.691	22.267
		Tertiary	-11.968	17.469	.958	-63.064	39.128
Deceleration	Motorway	Primary	-70.355	16.977	.007	-122.734	-17.975
		Secondary	-42.930	10.771	.009	-76.121	-9.738
		Tertiary	-32.476	11.149	.068	-66.837	1.886
		Trunk	-36.770	19.028	.343	-95.488	21.947

	Primary	Motorway	70.355	16.977	.007	17.975	122.734
		Secondary	27.425	20.058	.653	-31.421	86.271
		Tertiary	37.879	20.263	.358	-21.482	97.239
		Trunk	33.584	25.463	.682	-40.333	107.502
	Secondary	Motorway	42.930	10.771	.009	9.738	76.121
		Primary	-27.425	20.058	.653	-86.271	31.421
		Tertiary	10.454	15.440	.960	-34.336	55.244
		Trunk	6.159	21.821	.999	-58.194	70.513
	Tertiary	Motorway	32.476	11.149	.068	-1.886	66.837
		Primary	-37.879	20.263	.358	-97.239	21.482
		Secondary	-10.454	15.440	.960	-55.244	34.336
		Trunk	-4.295	22.010	1.000	-69.100	60.511
	Trunk	Motorway	36.770	19.028	.343	-21.947	95.488
		Primary	-33.584	25.463	.682	-107.502	40.333
		Secondary	-6.159	21.821	.999	-70.513	58.194
		Tertiary	4.295	22.010	1.000	-60.511	69.100
Steering	Motorway	Primary	-146.908	32.281	.003	-246.361	-47.456
		Secondary	-144.435	39.063	.015	-264.867	-24.002
		Tertiary	-75.946	16.724	.003	-127.158	-24.734
		Trunk	-39.185	15.030	.115	-85.122	6.753
	Primary	Motorway	146.908	32.281	.003	47.456	246.361
		Secondary	2.474	50.485	1.000	-144.292	149.239
		Tertiary	70.962	36.091	.314	-35.978	177.903
		Trunk	107.724	35.338	.043	2.460	212.988
	Secondary	Motorway	144.435	39.063	.015	24.002	264.867
		Primary	-2.474	50.485	1.000	-149.239	144.292
		Tertiary	68.489	42.266	.502	-57.880	194.858
		Trunk	105.250	41.624	.125	-19.798	230.298
	Tertiary	Motorway	75.946	16.724	.003	24.734	127.158
		Primary	-70.962	36.091	.314	-177.903	35.978
		Secondary	-68.489	42.266	.502	-194.858	57.880
		Trunk	36.762	22.054	.469	-27.257	100.780
	Trunk	Motorway	39.185	15.030	.115	-6.753	85.122
		Primary	-107.724	35.338	.043	-212.988	-2.460
		Secondary	-105.250	41.624	.125	-230.298	19.798
		Tertiary	-36.762	22.054	.469	-100.780	27.257
Tailgating	Motorway	Primary	26.086	10.946	.150	-5.837	58.008
		Secondary	43.839	9.303	.001	15.775	71.903
		Tertiary	50.212	8.835	<.001	22.971	77.453
		Trunk	11.964	12.479	.871	-24.232	48.160
	Primary	Motorway	-26.086	10.946	.150	-58.008	5.837
		Secondary	17.753	7.155	.133	-3.553	39.060
		Tertiary	24.127	6.534	.015	4.004	44.249
		Trunk	-14.121	10.971	.701	-46.120	17.877
	Secondary	Motorway	-43.839	9.303	.001	-71.903	-15.775

	Primary	-17.753	7.155	.133	-39.060	3.553
	Tertiary	6.373	3.071	.276	-2.997	15.743
	Trunk	-31.875	9.332	.022	-60.030	-3.719
Tertiary	Motorway	-50.212	8.835	<.001	-77.453	-22.971
	Primary	-24.127	6.534	.015	-44.249	-4.004
	Secondary	-6.373	3.071	.276	-15.743	2.997
	Trunk	-38.248	8.866	.005	-65.583	-10.913
Trunk	Motorway	-11.964	12.479	.871	-48.160	24.232
	Primary	14.121	10.971	.701	-17.877	46.120
	Secondary	31.875	9.332	.022	3.719	60.030
	Tertiary	38.248	8.866	.005	10.913	65.583

b. Low risky events

Box's Test of Equality of Covariance Matrices	
Box's M	240.232
F	5.298
df1	40
df2	12408.09
Sig.	<.001

Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Road type	Pillai's Trace	.478	2.543	16	300	.001
	Wilks' Lambda	.576	2.73	16	220.601	<.001
	Hotelling's Trace	.645	2.841	16	282	<.001
	Roy's Largest Root	.454	8.504	4	75	<.001

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Road type	Acceleration	19888.83	4	4972.208	7.219	<.001
	Deceleration	33653.99	4	8413.497	3.864	.007
	Steering	102803.4	4	25700.85	6.808	<.001
	Fatigue	.319	4	.080	.459	.766

Multiple Comparisons

Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Acceleration	Motorway	Primary	-47.682	9.182	<.001	-75.919	-19.445
		Secondary	-34.231	7.311	.002	-56.663	-11.800
		Tertiary	-24.631	6.523	.012	-44.613	-4.650
		Trunk	-20.624	6.255	.032	-39.771	-1.476

	Primary	Motorway	47.682	9.182	<.001	19.445	75.919
		Secondary	13.451	11.618	.775	-20.357	47.259
		Tertiary	23.051	11.139	.262	-9.486	55.587
		Trunk	27.058	10.984	.130	-5.080	59.197
	Secondary	Motorway	34.231	7.311	.002	11.800	56.663
		Primary	-13.451	11.618	.775	-47.259	20.357
		Tertiary	9.600	9.654	.856	-18.427	37.627
		Trunk	13.608	9.475	.610	-13.920	41.135
	Tertiary	Motorway	24.631	6.523	.012	4.650	44.613
		Primary	-23.051	11.139	.262	-55.587	9.486
		Secondary	-9.600	9.654	.856	-37.627	18.427
		Trunk	4.008	8.881	.991	-21.757	29.772
	Trunk	Motorway	20.624	6.255	.032	1.476	39.771
		Primary	-27.058	10.984	.130	-59.197	5.080
		Secondary	-13.608	9.475	.610	-41.135	13.920
		Tertiary	-4.008	8.881	.991	-29.772	21.757
Deceleration	Motorway	Primary	-63.654	14.363	.004	-107.964	-19.344
		Secondary	-40.584	9.769	.006	-70.686	-10.483
		Tertiary	-28.830	9.425	.052	-57.870	.210
		Trunk	-33.556	17.095	.329	-86.309	19.197
	Primary	Motorway	63.654	14.363	.004	19.344	107.964
		Secondary	23.070	17.325	.675	-27.610	73.749
		Tertiary	34.824	17.134	.279	-15.370	85.018
		Trunk	30.098	22.293	.663	-34.685	94.881
	Secondary	Motorway	40.584	9.769	.006	10.483	70.686
		Primary	-23.070	17.325	.675	-73.749	27.610
		Tertiary	11.754	13.516	.906	-27.453	50.961
		Trunk	7.028	19.649	.996	-50.894	64.951
	Tertiary	Motorway	28.830	9.425	.052	-.210	57.870
		Primary	-34.824	17.134	.279	-85.018	15.370
		Secondary	-11.754	13.516	.906	-50.961	27.453
		Trunk	-4.726	19.481	.999	-62.250	52.798
	Trunk	Motorway	33.556	17.095	.329	-19.197	86.309
		Primary	-30.098	22.293	.663	-94.881	34.685
		Secondary	-7.028	19.649	.996	-64.951	50.894
		Tertiary	4.726	19.481	.999	-52.798	62.250
Steering	Motorway	Primary	-94.091	16.816	<.001	-145.912	-42.270
		Secondary	-91.933	23.507	.010	-164.444	-19.422
		Tertiary	-51.500	11.439	.003	-86.669	-16.331
		Trunk	-32.628	14.845	.232	-78.349	13.093
	Primary	Motorway	94.091	16.816	<.001	42.270	145.912
		Secondary	2.158	28.822	1.000	-81.993	86.309
		Tertiary	42.591	20.224	.247	-16.575	101.757
		Trunk	61.463	22.328	.070	-3.364	126.290

	Secondary	Motorway	91.933	23.507	.010	19.422	164.444
		Primary	-2.158	28.822	1.000	-86.309	81.993
		Tertiary	40.433	26.054	.542	-36.976	117.841
		Trunk	59.305	27.718	.235	-22.035	140.645
	Tertiary	Motorway	51.500	11.439	.003	16.331	86.669
		Primary	-42.591	20.224	.247	-101.757	16.575
		Secondary	-40.433	26.054	.542	-117.841	36.976
		Trunk	18.872	18.617	.847	-35.353	73.097
	Trunk	Motorway	32.628	14.845	.232	-13.093	78.349
		Primary	-61.463	22.328	.070	-126.290	3.364
		Secondary	-59.305	27.718	.235	-140.645	22.035
		Tertiary	-18.872	18.617	.847	-73.097	35.353

c. Medium risky events

Box's Test of Equality of Covariance Matrices

Box's M	474.834
F	4.65
df1	84
df2	10746.75
Sig.	<.001

Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Road type	Pillai's Trace	.827	3.172	24	292	<.001
	Wilks' Lambda	.337	3.747	24	245.411	<.001
	Hotelling's Trace	1.518	4.332	24	274	<.001
	Roy's Largest Root	1.186	14.433	6	73	<.001

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Road type	Speeding	138.343	4	34.586	4.366	.003
	Acceleration	4599.784	4	1149.946	3.685	.009
	Deceleration	336.491	4	84.123	1.673	.165
	Steering	34054.99	4	8513.747	5.226	<.001
	Tailgating	18825.66	4	4706.414	11.088	<.001
	Fatigue	1.305	4	.326	2.36	.061

Multiple Comparisons

Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Speeding	Motorway	Primary	-3.363	.890	.012	-6.092	-.634
		Secondary	-2.887	1.043	.088	-6.091	.318

		Tertiary	-1.322	.773	.455	-3.687	1.043
		Trunk	-.484	.257	.349	-1.233	.265
	Primary	Motorway	3.363	.890	.012	.634	6.092
		Secondary	.476	1.354	.997	-3.459	4.412
		Tertiary	2.041	1.159	.415	-1.326	5.408
		Trunk	2.879	.901	.038	.131	5.627
	Secondary	Motorway	2.887	1.043	.088	-.318	6.091
		Primary	-.476	1.354	.997	-4.412	3.459
		Tertiary	1.565	1.281	.739	-2.171	5.300
		Trunk	2.403	1.053	.200	-.818	5.623
	Tertiary	Motorway	1.322	.773	.455	-1.043	3.687
		Primary	-2.041	1.159	.415	-5.408	1.326
		Secondary	-1.565	1.281	.739	-5.300	2.171
		Trunk	.838	.786	.821	-1.549	3.226
	Trunk	Motorway	.484	.257	.349	-.265	1.233
		Primary	-2.879	.901	.038	-5.627	-.131
		Secondary	-2.403	1.053	.200	-5.623	.818
		Tertiary	-.838	.786	.821	-3.226	1.549
Acceleration	Motorway	Primary	-22.336	5.086	.004	-37.912	-6.759
		Secondary	-17.240	4.475	.010	-30.913	-3.566
		Tertiary	-15.340	6.100	.137	-34.066	3.386
		Trunk	-10.192	4.128	.146	-22.783	2.400
	Primary	Motorway	22.336	5.086	.004	6.759	37.912
		Secondary	5.096	6.644	.938	-14.197	24.389
		Tertiary	6.996	7.831	.897	-15.766	29.758
		Trunk	12.144	6.416	.344	-6.518	30.806
	Secondary	Motorway	17.240	4.475	.010	3.566	30.913
		Primary	-5.096	6.644	.938	-24.389	14.197
		Tertiary	1.900	7.449	.999	-19.834	23.634
		Trunk	7.048	5.943	.759	-10.197	24.293
	Tertiary	Motorway	15.340	6.100	.137	-3.386	34.066
		Primary	-6.996	7.831	.897	-29.758	15.766
		Secondary	-1.900	7.449	.999	-23.634	19.834
		Trunk	5.148	7.245	.952	-16.061	26.357
	Trunk	Motorway	10.192	4.128	.146	-2.400	22.783
		Primary	-12.144	6.416	.344	-30.806	6.518
		Secondary	-7.048	5.943	.759	-24.293	10.197
		Tertiary	-5.148	7.245	.952	-26.357	16.061
Steering	Motorway	Primary	-48.557	15.485	.045	-96.217	-.897
		Secondary	-49.402	15.144	.035	-96.004	-2.800
		Tertiary	-23.878	6.202	.009	-42.668	-5.089
		Trunk	-6.195	3.324	.361	-15.946	3.555
	Primary	Motorway	48.557	15.485	.045	.897	96.217
		Secondary	-.845	21.512	1.000	-63.246	61.557

		Tertiary	24.679	16.490	.577	-24.814	74.172
		Trunk	42.362	15.636	.097	-5.544	90.268
	Secondary	Motorway	49.402	15.144	.035	2.800	96.004
		Primary	.845	21.512	1.000	-61.557	63.246
		Tertiary	25.524	16.170	.527	-22.962	74.009
		Trunk	43.207	15.298	.078	-3.648	90.061
	Tertiary	Motorway	23.878	6.202	.009	5.089	42.668
		Primary	-24.679	16.490	.577	-74.172	24.814
		Secondary	-25.524	16.170	.527	-74.009	22.962
		Trunk	17.683	6.570	.089	-1.860	37.225
	Trunk	Motorway	6.195	3.324	.361	-3.555	15.946
		Primary	-42.362	15.636	.097	-90.268	5.544
		Secondary	-43.207	15.298	.078	-90.061	3.648
		Tertiary	-17.683	6.570	.089	-37.225	1.860
Tailgating	Motorway	Primary	18.661	8.885	.247	-7.163	44.486
		Secondary	35.115	7.253	<.001	13.280	56.950
		Tertiary	40.609	6.853	<.001	19.486	61.733
		Trunk	8.877	9.694	.889	-19.243	36.997
	Primary	Motorway	-18.661	8.885	.247	-44.486	7.163
		Secondary	16.454	6.189	.096	-2.036	34.943
		Tertiary	21.948	5.716	.012	4.345	39.551
		Trunk	-9.785	8.927	.807	-35.735	16.166
	Secondary	Motorway	-35.115	7.253	<.001	-56.950	-13.280
		Primary	-16.454	6.189	.096	-34.943	2.036
		Tertiary	5.494	2.517	.233	-2.177	13.165
		Trunk	-26.238	7.304	.015	-48.234	-4.242
	Tertiary	Motorway	-40.609	6.853	<.001	-61.733	-19.486
		Primary	-21.948	5.716	.012	-39.551	-4.345
		Secondary	-5.494	2.517	.233	-13.165	2.177
		Trunk	-31.733	6.907	.003	-53.023	-10.442
	Trunk	Motorway	-8.877	9.694	.889	-36.997	19.243
		Primary	9.785	8.927	.807	-16.166	35.735
		Secondary	26.238	7.304	.015	4.242	48.234
		Tertiary	31.733	6.907	.003	10.442	53.023

d. High risky events

Box's Test of Equality of Covariance Matrices	
Box's M	433.439
F	5.974
df1	56
df2	5784.131
Sig.	<.001

Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Road type	Pillai's Trace	0.619	1.884	28	288	.006
	Wilks' Lambda	0.485	1.986	28	250.205	.003
	Hotelling's Trace	0.861	2.076	28	270	.002
	Roy's Largest Root	0.575	5.918	7	72	<.001

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Road type	Speeding	321.428	4	80.357	2.544	.046
	Acceleration	117.98	4	29.495	.783	.540
	Deceleration	1.281	4	.320	.737	.570
	Steering	231.462	4	57.866	1.877	.123
	Tailgating	1066.199	4	266.55	7.577	<.001
	Overtaking	0.241	4	.060	.835	.507
	Fatigue	0.47	4	.117	1.239	.302

Multiple Comparisons

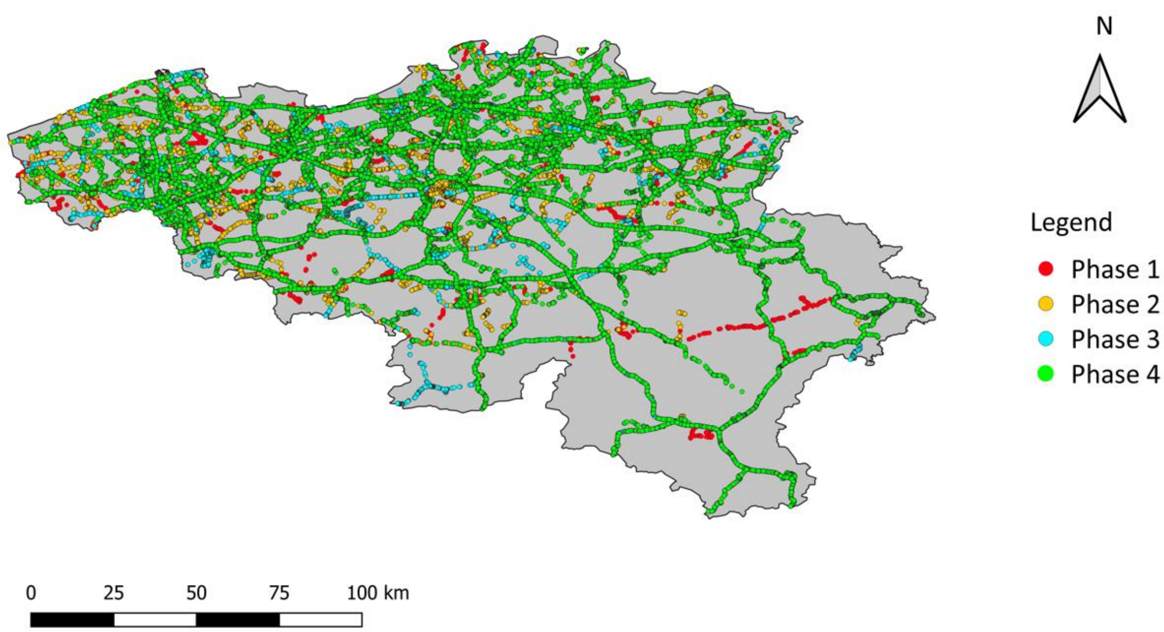
Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Speeding	Motorway	Primary	-4.997	1.673	.058	-10.124	.131
		Secondary	-5.130	1.682	.051	-10.285	.026
		Tertiary	-1.406	.905	.543	-4.133	1.322
		Trunk	-2.566	1.920	.674	-8.460	3.329
	Primary	Motorway	4.997	1.673	.058	-.131	10.124
		Secondary	-.133	2.336	1.000	-6.908	6.641
		Tertiary	3.591	1.856	.329	-1.904	9.085
		Trunk	2.431	2.512	.867	-4.865	9.727
	Secondary	Motorway	5.130	1.682	.051	-.026	10.285
		Primary	.133	2.336	1.000	-6.641	6.908
		Tertiary	3.724	1.864	.299	-1.796	9.244
		Trunk	2.564	2.518	.845	-4.749	9.877
	Tertiary	Motorway	1.406	.905	.543	-1.322	4.133
		Primary	-3.591	1.856	.329	-9.085	1.904
		Secondary	-3.724	1.864	.299	-9.244	1.796
		Trunk	-1.160	2.081	.980	-7.363	5.043
	Trunk	Motorway	2.566	1.920	.674	-3.329	8.460
		Primary	-2.431	2.512	.867	-9.727	4.865
		Secondary	-2.564	2.518	.845	-9.877	4.749
		Tertiary	1.160	2.081	.980	-5.043	7.363
Tailgating	Motorway	Primary	7.424	2.467	.048	.044	14.804
		Secondary	8.724	2.370	.013	1.535	15.913
		Tertiary	9.603	2.285	.006	2.560	16.645

	Trunk	3.088	3.102	.855	-5.913	12.089
Primary	Motorway	-7.424	2.467	.048	-14.804	-.044
	Secondary	1.300	1.156	.792	-2.079	4.678
	Tertiary	2.179	.968	.211	-.783	5.140
	Trunk	-4.337	2.311	.360	-11.225	2.552
Secondary	Motorway	-8.724	2.370	.013	-15.913	-1.535
	Primary	-1.300	1.156	.792	-4.678	2.079
	Tertiary	.879	.685	.704	-1.199	2.957
	Trunk	-5.636	2.207	.123	-12.314	1.042
Tertiary	Motorway	-9.603	2.285	.006	-16.645	-2.560
	Primary	-2.179	.968	.211	-5.140	.783
	Secondary	-.879	.685	.704	-2.957	1.199
	Trunk	-6.515	2.115	.050	-13.033	.003
Trunk	Motorway	-3.088	3.102	.855	-12.089	5.913
	Primary	4.337	2.311	.360	-2.552	11.225
	Secondary	5.636	2.207	.123	-1.042	12.314
	Tertiary	6.515	2.115	.050	-.003	13.033

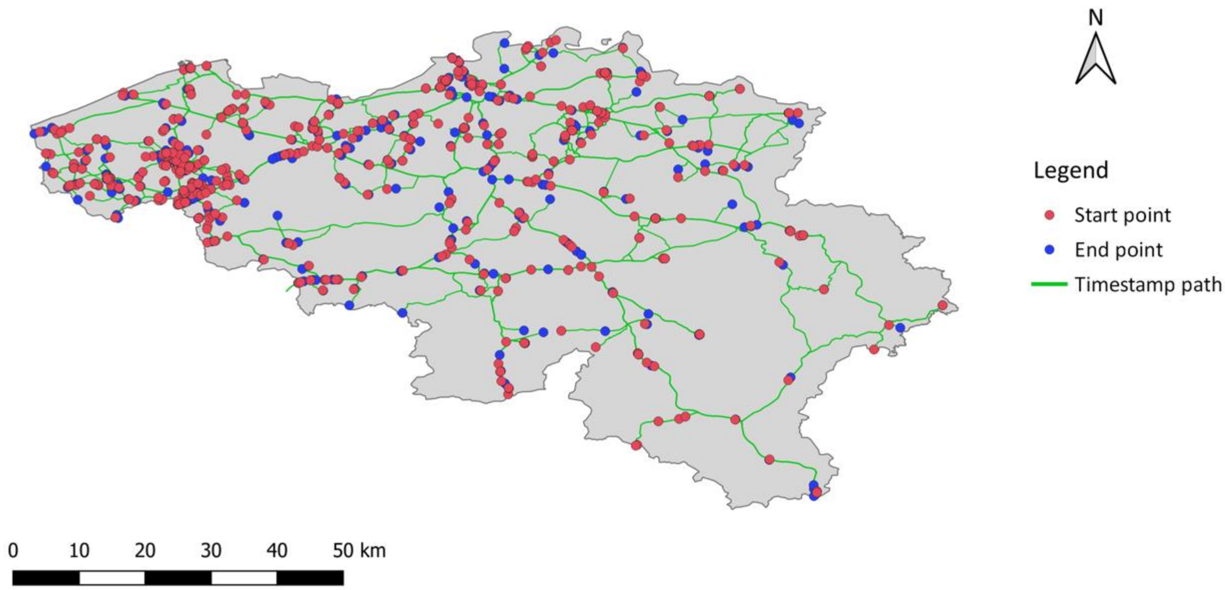
Appendix – III

QGIS outputs

Risky events: All phases after buffering



Phase one: Timestamp path of risky events



Density estimation of risky events

a. Severity and time of the day

Average density of risky events per road type per km				
	Night/early morning	Midday	Afternoon/ evening	Morning
Low	2.931	4.501	3.977	5.184
Medium	1.781	2.759	2.854	3.259
High	.555	1.443	.937	.679
Total	4.746	6.729	6.862	7.685

b. Severity and weather

Average density of risky events per weather cluster per km			
	Cluster 1	Cluster 2	Cluster 3
Low	3.430	4.108	4.528
Medium	2.423	3.256	3.803
High	.613	1.330	.854
Total	5.670	7.465	9.728

c. Severity and road type

Average density of risky events per road type per km					
	Motorway	Primary	Secondary	Tertiary	Trunk
Low	5.466	3.283	3.931	16.717	2.248
Medium	7.674	2.301	2.404	13.567	2.048
High	.779	1.066	1.618	2.474	.743
Total	11.219	5.106	5.214	26.002	3.539