D2.14 DESCRIPTION METRICS FOR TRAFFIC INTERACTIONS

Primary Author(s)	MSc. Michiel Braat | TNO
	Dr. Pedro Huertas Leyva | UNI
	MSc. Jeroen Manders | TNO
	MSc. Fabian Russ | IKA
	Dr. Arturo Tejada | TNO
	Dr. Xinwei Wang | TUD

Related Work Package	WP2
Version/Status	2 | Final Document
Issue date	10/06/2022
Deliverable type	R
Dissemination Level	PU

Project Acronym	SAFE-UP
Project Title	proactive SAFEty systems and tools for a constantly UPgrading road environment
Project Website	www.safe-up.eu
Project Coordinator	Núria Parera | Applus IDIADA
Grant Agreement No.	861570

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Co-Authors

Name	Organisation
MSc. Michiel Braat	Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO)
Dr. Pedro Huertas Leyva	Università degli Studi di Firenze (UNI)
MSc. Jeroen Manders	TNO
MSc. Fabian Russ	Institut für Kraftfahrzeuge RWTH Aachen University (IKA)
Dr. Arturo Tejada	TNO
Dr. Xinwei Wang	Delft University of Technology (TUD)

Document Distribution

Version	Date	Distributed to
1	27/06/2022	Coordination Team
2	30/06/2022	Submission in the EC System
2	06/03/2023	Approved by the EC

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Copyright statement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Executive summary

The SAFE-UP project aims to proactively address the novel safety challenges of the future mobility systems through the development of tools and innovative safety methods that lead to improvements in road transport safety.

Future mobility systems will rely on partially and fully automated vehicles to reduce traffic collisions and casualties by removing causal factors like driver distraction, fatigue or infractions and by reacting autonomously to emergency situations. On the other hand, they may introduce new collision risk factors or risky behaviours when interacting with other traffic participants .

SAFE-UP's Work Package 2 will further the understanding of the impact of vehicle automation technologies on safety by leveraging newly developed behavioural traffic simulation tools. These tools will allow one to simulate specific road networks with a variable proportion of automated vehicles to non-automated traffic participants (including human-drivers, pedestrians, cyclists, and powered two-wheelers). The simulation models will be detailed enough to realistically recreate the effects of unexpected events (like surprise cut-ins). In this way, one will be able to determine whether these technologies induce changes (positive or negative) in surrogate indicators of traffic safety.

An important surrogate indicator of safety is the occurrence of safety-critical interactions in different driving scenarios. Building on the workplan laid down on the first version of this report (Deliverable 2.5), each partner from Task 2.2 provides a detailed description of the severity metrics they developed and outline their plans to use these metrics in Task 2.5.

A complete account of these metrics can be found in Section 2, and the partners' main conclusions and recommendations can be found in Section 3. A short overview of the work is provided here:

- TNO developed two types of metrics for motor vehicle interactions based on two separate techniques: unsupervised anomaly detection and driver profiling. The former is best suited for analysing highway interactions while the latter works best for urban interactions.
- TUD developed a metric for motor vehicle interactions based on artificial field theory. The metric, called probabilistic driving risk field, was calibrated using NDD data and validated using simulation studies and NDD data. TUD showed that their metrics can more accurately predict crashes than time-to-collision-base predictors.
- IKA analysed existing surrogate metrics of safety developed to assess motor vehicle interactions to understand their suitability in characterizing VRU interactions. Along its own metric for pedestrian interactions (deceleration to safety time), IKA recommended a number of metrics to characterize pedestrian and cyclist interactions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

- UNI developed a new metric to characterize longitudinal interactions for PTWs called the motorcycle crash potential index. This metric was then used to identify critical events in naturalistic driving data.

Through this work, a number of common challenges and conclusions were identified:

- There is limited access to well-processed naturalistic driving data that can be used to calibrate and validate severity safety metrics.
- There is no ground-truth for severity metrics, which makes it hard to compare the outcomes of specific metrics in absolute terms. For instance, two metrics based on completely different principles could declare the same interaction as critical while awarding that interaction very different severity values.
- A comparative analysis of several severity metrics should be conducted across the EU and an agreement or convention should be reached with multiple stakeholders (OEMs, government regulators, academia, etc.) on how criticality/severity should be defined and measured.
- The aforementioned convention should lead to the creation of a “severity ground truth”. Since such a ground truth would be location based, a concerted effort should be made to acquire NDD from enough representative locations.

Finally, the work in T2.2 yielded high-quality results that have been documented in at least two journal publications and one conference paper.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement 861570.

Table of contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

List of figures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

List of tables

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

List of abbreviations

Abbreviation	Meaning
CPI	Crash Potential Index
D	Deliverable
DRAC	Deceleration Rate to Avoid Crash
DST	Deceleration to Safety Time
LD	Lateral Distance
LTTC	Lateral Time To Collision
MADR	Maximum Available Deceleration Rate
MCPI	Motorcycle Crash Potential Index
ML	Machine Learning
NDD	Naturalistic driving data
PCA	Principal Component Analysis
PDF	Probability Density Function
PDRF	Probabilistic Driving Risk Field
PET	Post-Encroachment Time
PICUD	Potential Index for Collision with Urgent Deceleration
PSD	Proportion of Stopping Distance
PTW	Powered Two-Wheelers
SMoS	Surrogate Measures of Safety
SotA	State-of-the-Art
SSAM	Surrogate Safety Assessment Model
T	Task
THW	Time Headway
TTC	Time-To-Collision
VRU	Vulnerable road user
WP	Work Package

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

1. Introduction

1.1 The EU Project SAFE-UP

The SAFE-UP project aims to proactively address the novel safety challenges of the future road mobility environment by developing tools and innovative safety methods, leading to improvements in road transport safety.

Future mobility systems are expected to make use of vehicles with full or partial automation of the driving task, the so-called SAE L3/4/5 vehicles (SAE, 2018). By supporting (or even replacing human) drivers during the driving task, such vehicles may help improve road safety by removing some of the known sources of collisions (e.g., driver distraction) or by taking control during critical situations. On the other hand, automated vehicles may introduce new collision risk factors (e.g., increased distraction during transition of control) or induce new risky behaviours in other traffic participants (Hamilton, 2019).

The true impact of vehicle automation technologies on road safety will become apparent in the decades to come, as it depends on social and market trends that are difficult to forecast (like technological developments in sensors for automated vehicles, market penetration and acceptance of automation technologies, etc.).

Through the work in Work Package (WP) 2, SAFE-UP will further the understanding of the future impact of vehicle automation technologies by leveraging newly developed behavioural traffic simulation tools. These tools, currently under development by SAFE-UP's partners in Tasks (T) 2.3 and 2.4 (see Deliverables (D) 2.9-2.12 for details), will allow one to simulate specific road networks with a variable proportion of vehicles equipped with automation technologies. By analysing the simulation results, one will be able to determine whether these technologies induce changes (positive or negative) in indicators of safety.

As explained in the first version of this report (D2.5 "Description metrics for traffic interactions"), an important indicator of safety is the occurrence of safety-critical driving interactions or their precursors (i.e., interactions that have the potential to become safety-critical). The impact on safety can also be seen as a change on the number of traffic participants interacting safely in traffic.[1]

T2.2 aimed at identifying and developing metrics that allow one to classify driving interactions as either safe or as safety-critical (these are the two extremes of the "severity" spectrum). Such metrics can be used to develop collision avoidance algorithms, to test the capabilities of specific vehicle automation technologies or, as in the case of T2.5, to identify safety critical interactions in micro-simulations.

[1] As shown in Figure 1, interactions not declared "safe" are not directly "safety-critical" (and vice versa).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

This document presents a detailed description of several new metrics developed in this project by different partners according to the work plan outlined in D2.5. These metrics play an important role in the contribution of WP2 to the goals of the project.

1.2 Objective of this Report

This report is the final version of D2.5 “Description metrics for traffic interactions”, which provided the following information:

- An extensive overview of severity metrics from the perspectives of motor vehicles, PTWs (powered two-wheelers) and VRUs (vulnerable road users)
- Recommendations of which metrics to use to identify safety-critical interactions and gaps in the literature
- Initial work plans for the T2.2 partners.

Deliverable 2.14 builds on this foundation to provide the following information:

- A detailed description of the severity metrics developed by different T2.2. partners.
- Preliminary plans on how these metrics would be used in T2.5 to analyse micro-simulation results produced by the simulation environment developed in T2.4.[2]

1.3 Report Organization

The rest of the report is organized as follows: Section 2 is the main section of this document. It provides a basic terminology, a well-established approach to relate traffic interactions and safety, which helps the reader to understand the coverage of the different metrics developed by the partners, and a list of common surrogate metrics of safety available in the literature. It also contains the details on the metrics developed by the T2.2 partners. This is followed by our conclusions and recommendation in Section 3.

[2] A detailed description of the initial simulation analysis plans from T2.5 are presented in D2.8 “Analysis of simulation results and identification of future safety-critical traffic interactions”.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement 861570.

2. Partner Contributions

This section presents the metrics developed by each partner in T2.2 (TNO, TUD, IKA, UNIFI). As explained below, each partner has focused their work on different kinds of interactions: Motor vehicle interactions (TNO, TUD), VRU interactions (IKA), and PTW interactions (UNI).

2.1 Basic Terminology and Positioning of Partner's Work

For convenience, this section provides a shortened version of the common terminology presented in D2.5 that is used by all partners to describe their work. It also provides a view on how to position the work by the T2.2 partners from the metric application point of view; and a brief list of common surrogate metrics of safety used throughout the text.

2.1.1 Microscopic traffic simulation

A microscopic traffic simulation is a form of agent-based simulation wherein "…traffic flow is based on the description of the motion of each individual vehicle composing the traffic stream" (Barceló, 2010).

In such simulations, a section of (a possibly real) road network is first constructed, including infrastructural elements like lanes, roundabouts, lane marking, traffic lights and signals, etc. The network is then populated by individual traffic participants (motor vehicles, VRUs, etc.), whose behaviour is prescribed by models (e.g., car following (Alexiadis, et al., 2004)). The individual participant then traverses the network and interacts with each other following their own individual behaviours. The models are generally calibrated in such a way that metrics such as traffic flow and density reach measured levels in the real world.

While traversing the network, different participants will interact with each other under different scenarios. A scenario describes the traffic, infrastructure and environment for the simulation and consists of a sequence of scenes. A scene describes a snapshot that encompasses the mobile and immobile elements of the traffic, infrastructure and environment, and the relations between these elements.

The scenarios give rise to interactions among traffic participants. When one or more interactions are safety-critical (e.g., if an intervention is needed to avoid a collision) the scenario is considered to be a safety-critical scenario.

An interaction is a driving situation where the behaviour of at least two traffic participants can be interpreted as being influenced by their intention of occupying the same region of space at the same time in the near future (Defining interactions: a conceptual framework for understanding interactive behaviour in human and automated road traffic, 2020). Interactions in which only the behaviour of one traffic participant is affected are called reactions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

2.1.2 Driving Interactions, Severity and Surrogate Metrics of Safety

Although driving interactions are very frequent in traffic, collisions are very rare. Intuitively, one can think of interactions as possessing a level of criticality (i.e., severity): Those with low severity have a lower chance to lead to collisions, and those with a higher severity have a higher chance to lead to collisions. A useful way to visualize the relationship between driving interactions and severity is shown in Figure 1. This figure shows the frequency distribution of driving interactions according to their severity.

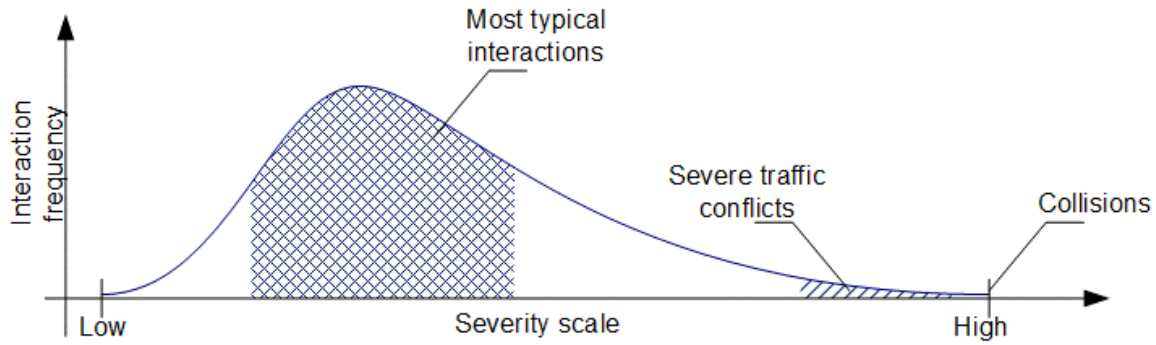

Figure 1 Driving interaction frequency distribution as a function of their severity. Adapted from (Estimating the severity of safety related behaviour, 2006)

Although the general trend shown in Figure 1 has been confirmed in practice, the literature still lacks a generalizable theory of the circumstances that turn particular driving interactions into collisions. A frequent hypothesis is that collisions result from a temporal sequence of events in which a conflict event occurs prior to a crash event (Laureshyn, et al., 2016). Since conflicts (i.e., high-severity interactions) and collisions are aligned on the same continuum of events, the frequency of the low-severity events can be used to predict (statistically) that of high-severity events (i.e., collisions) (Blumenthal, et al., 2020).

Although Figure 1 does not specify what severity is, several methods have been proposed to assign severity to interactions (e.g., time to collision). As explained in detail in D2.5, these methods yield the so-called surrogate metrics of safety (SMoS). The most common SMoS used in this report are defined in Table 1.

Table 1: Definitions of surrogate metrics of safety used in the report. See D2.5 for complete details.

Metric	Unit	Definition
DRAC: Deceleration Rate to Avoid Collision	m/s^2	Differential speed between a following/ response vehicle and its corresponding subject/ lead vehicle divided by their closing time.
CPI: Crash Potential Index	-	Probability that a given vehicle DRAC exceeds its maximum available deceleration rate (MADR) during a given time interval. Driver reaction time can be incorporated in DRAC modified CPI
PET: Post Encroachment Time	s	The time between the moment a road user leaves the area of a potential collision and another road user arrives at it.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

PICUD: Pot. Index for Collision with Urgent Deceleration	m	Distance between the two vehicles considered when they completely stop.
TTC: Time to Collison	s	The time until a collision between the vehicles would occur if they continued on their present course at their present speeds.

These metrics have limitations (e.g., they are only “car” centric, they are not data driven, they cannot be used for path planning purposes, etc. See D2.5 for details). The partners in T2.2 developed new metrics to alleviate some of these limitations. Their work is positioned as shown in Figure 2.

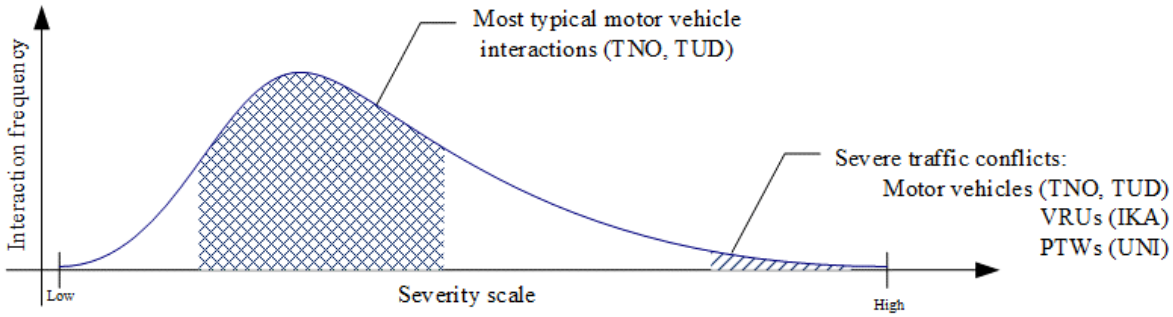

Figure 2 Overview of the positioning of the research work by T2.2 partners.

The partner contributions are presented next.

2.2 Metrics for Motor Vehicle Interactions (TNO)

2.2.1 Motivation

The participation of TNO in WP2 was motivated in part by the need of developing methods to assess the safety of automated vehicles. The higher the automation level, the greater autonomy a vehicle has and the more it resembles an “artificial driver”. Although the automotive industry has large experience (and a long history) testing and assessing the safety of vehicle technology, it has much less experience assessing how safely a vehicle drives in traffic.

Driving is a form of social activity mediated by technology. For a human driver, proving that he/she drive safely requires more than performing correctly particular manoeuvres in specific scenarios. He/she must also show awareness of the driving situation and the impact of his/her actions on other drivers and road users.

It is not clear yet how automated vehicles should prove that they are safe drivers. It has been hypothesized that they should show that they execute the correct manoeuvres for the right reasons (Koopman, et al., 2018). Alternatively, an automated vehicle could be said to

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement 861570.

drive safely if its observable driving behaviour (after it joins traffic) cannot be distinguished from that of other drivers[3].

As shown in Figure 1, the vast majority of current driving interactions (i.e., the most typical interaction) carry low severity. Thus, our approach in this project was to develop techniques that allow one to identify whether a driver (human or automated) drives “typically”.

“Typical” driving varies with the type of location, time of day, weather conditions, etc. Thus, we developed methods to characterize typical driving interactions based on recorded data from specific contexts based on machine learning and statistical profiling techniques trained on publicly-available naturalistic driving data, such as Next Generation Simulation (NGSIM) (USDOT, 2006) and pNEUMA (On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment, 2020).

The first method, presented in Section 2.2.2, is instantaneous and therefore comparable with existing surrogate measures of safety (SMoS). Our metric models whether longitudinal vehicle pairs show (a)typical behaviour by using unsupervised anomaly detection techniques. The second method clusters unique rides (a single trip of a specific driver) based on how they drive. This is often referred to as driver behaviour profiling. Clustering of a ride is done based on how they compare with the norm for each driven road segment. We describe this metric in Section 2.2.3. In Section 2.2.4 we reflect on both metrics and we present our plans for T2.5 in Section 2.2.5.

2.2.2 Unsupervised Anomaly Detection

Modelling what is typical or (a)typical driving behaviour is related to the machine learning field of anomaly detection. An anomaly is a *"condition that deviates from expectations, based, for example, on requirements, specifications, design documents, user documents, standards, or on experience"* (ISO 26262, 2018). We will use anomaly detection as a method to distinguish typical driving behaviour from unsafe interactions. This rests on the assumption that unsafe driving behaviour is a rare occurrence and therefore abnormal or anomalous.

More specifically, we focus on unsupervised anomaly detection. This means:

- Models are focused on modelling the normal, or typical, behaviour. Frequent occurring relationships between the input signals are learned, which are leveraged during testing to define whether something is anomalous or not.
- There is not an explicit assumption about the data other than that it follows a natural distribution between normal and abnormal data. Labelled data is not required to train a model. In our case, a ground truth of atypical driving behaviour

[3] In the foreseeable future, the driving skills of automated vehicles should match those of competent and careful human drivers.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement 861570.

would be unfeasible to make. However, the lack of labelled data makes model validation a difficult task.

- On the one hand, unsupervised anomaly detection methods can detect anomalies not considered a priori. On the other hand, certain expected, infrequent, behaviour can be considered anomalous by the model.

We will cover these three aspects and show our results in the rest of this section. First, we will introduce our naturalistic driving dataset, and how we limit the effects of classifying expected, but low frequent, behaviour as anomalous. Further, we introduce the unsupervised deep neural network model we are going to use for our experiments: the autoencoder. After introducing our validation approach, we end this section by presenting our results.

Naturalistic Driving Data

We will train our model on the US Highway 101 (US-101) and Interstate 80 Freeway (I-80) data from the Next Generation Simulation (NGSIM) dataset (USDOT, 2006) (NGSIM US highway 101 dataset, 2007) (NGSIM interstate 80 freeway dataset, 2006). NGSIM is recorded at 10Hz, using stationary digital video cameras on top of tall buildings. The distance between the camera and the highway, camera angle, and image quality results in the NGSIM dataset containing significant amounts of noise and other limitations (n the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation, 2011) (stimating acceleration and lane-changing 42 dynamics based on ngsim trajectory data, 2008). To minimize these effects, we use a Savitsky-Golay (Smoothing and differentiation of data by simplified least squares procedures., 1964) filter to smooth vehicle positions over time. Further, we follow the approach by Altché et al. (An LSTM network for highway trajectory prediction, 2017), and use a first-degree Savitsky-Golay filter with a window length of 11 (which corresponds to a 1 second time window). The smoothed longitudinal and lateral positions over time are used to recalculate the velocity and acceleration of each vehicle. This is similar to how NGSIM is constructed; positions over time are tracked and other signals are derived from these.

Given the importance of natural data distribution for unsupervised anomaly detection, we do additional data filtering by removing unrealistic longitudinal driving behaviour. First, all same-lane longitudinal vehicle pairs (a leader l and a follower f) are extracted from the dataset. Then, every vehicle pair that violates one of these two rules at least once is removed from the dataset:

1. $\Delta D < 0$, where ΔD is the inter-vehicle-distance between both vehicles. A negative ΔD would indicate an accident has happened, but in NGSIM this occurs frequently as a result of data noise.

2. $|A_l| > 4 \vee |A_f| > 4$, where A_l and A_f are the leading and following vehicle accelerations. This filters out sudden changes in position measurements that lead to very high acceleration or deceleration.

To prevent us from classifying normal, non-severe, but low-frequent behaviour as atypical, we limit ourselves to longitudinal driving behaviour. More specifically, every timestep from

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

the remaining longitudinal vehicle pairs is used as a sample if $\frac{\Delta D}{V_f} <= 4$, where $\frac{\Delta D}{V_f}$ correspond to the Time Headway (THW) between both vehicles (Evans, 1991) for that timestep. This prevents samples with high inter-vehicle distances, which are rare in the rush hour traffic of NGSIM, to be considered anomalous by the model. Further, in our selection of input features for the autoencoder model, we do not include any lateral features such as position within the lane and lateral velocity/acceleration. This prevents lane change behaviour, which is infrequent in our remaining data, from becoming classified as anomalous.

The data that is used as input for the anomaly detection model solely consists of longitudinal velocity, acceleration and distance features. For velocity V and acceleration A we use both the leading l and following f vehicle values as features, as well as their relative values ($\Delta V = V_f - V_l$, and $\Delta A = A_f - A_l$). Further, we include the inter-vehicle-distance ΔD. Since we want to balance acceleration, velocity, and inter-vehicle-distance to be as important as the others for our model, we triple the importance of the ΔD feature during modelling. In summarizing, the following seven features are used for our model:

Velocity features:

- V_l
- V_f
- ΔV

Acceleration features:

- A_l
- A_f
- ΔA

Distance features:

- ΔD

Autoencoder

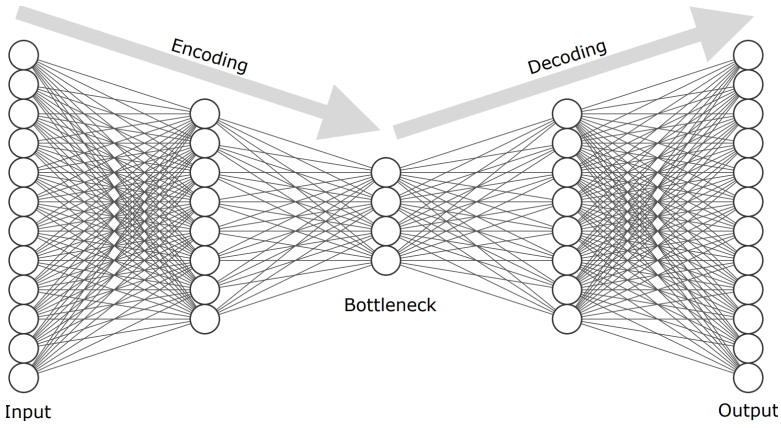

Figure 3. A typical encoding-decoding autoencoder architecture.

Figure 3 shows a typical autoencoder structure: It is a deep neural network that uses an encoding-decoding architecture to learn a mapping from the input to itself (Nonlinear principal component analysis using autoassociative neural networks, 1991):

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

$$\hat{X} = D(E(X))$$

where X is the input data, E is the encoder that encodes X into a smaller latent space (bottleneck). D is the decoder that decodes the latent representation back to a reconstruction of the input $\hat{X}$.

The goal of an autoencoder is to train the encoder and decoder to minimize the difference between the actual input X and reconstructed input $\hat{X}$ by using the l_2-norm:

$$\min_{E,D} ||X - D(E(X))||^2.$$

During training, an autoencoder minimizes the loss over all its training samples. Since data outliers (i.e., anomalies and input noise) appear with low frequency in the training set, the autoencoder does not prioritize their reconstruction while training. After training, the autoencoder reconstructs with the highest precision the most common characteristics of the indicators applied to it. This allows one to readily identify anomalies by simply comparing the inputted and reconstructed indicators and determining when the reconstruction error exceeds acceptable levels. This makes autoencoders well suited for anomaly detection applications (Anomaly detection with robust deep autoencoders, 2017) (Variational autoencoder based anomaly detection using reconstruction probability, 2015) (Outlier detection with autoencoder ensembles, 2017) and an excellent choice to model driving behaviour, as severe traffic conflicts are rare compared to typical driving interactions.

More specifically, our autoencoder encodes the 7 inputs features to 2 dimensions via three dense layers: 7 -> 128 -> 32 -> 2. The decoder reverses the encoder architecture to output 7 dimensions: 2 -> 32 -> 128 -> 7. The autoencoder hyperparameters are optimized toward minimizing the l2-norm reconstruction error for this architecture. We use ReLU (Imagenet classification with deep convolutional neural networks, 2012) activations and batch normalization (Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015) on all hidden layers. The model is trained for 100 epochs with a batch size of 1024 and uses Adam (Adam: A method for stochastic optimization, 2014) with default parameters as an optimizer.

Validation Approach

Whilst the goal of our approach is to monitor safe driving behaviour, in all its meaning, we cannot validate our modelling approach with the undefined concept of safe driving behaviour. Instead, we validate our approach to objective risk based on four different risk metrics. For every metric, we test how good the autoencoder is at defining severe traffic situations based on their frequency of occurrence. We compare our autoencoder results with the following four risk metrics:

- Time to collision (TTC) (Near miss determination through use of a scale of danger, 1972) (Traffic conflicts technique: state-of-the-art, 1996) models the time in seconds left before two vehicles collide, assuming nobody takes evasive action and speed differences are maintained.
- Time headway (THW) (Evans, 1991) defines the time between the two vehicles passing the same location. More specifically, we model the time headway as the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

time it takes between the rear bumper of the leading vehicle and the front bumper of the following vehicle to pass the same point.

- Deceleration rate to avoid collision (DRAC) (Use of speed limiters in cars for increased safety and a better environment, 1991) is defined as the minimal required deceleration rate the following vehicle has to apply to avoid a crash with the leading vehicle.
- Potential index for collision with urgent deceleration (PICUD) (Traffic conflict analysis and modeling of lane-changing behavior at weaving section, 2001) is defined as the distance between two vehicles when they completely stop assuming that the leading vehicle applies its emergency brake. We follow Uno et al. (A microscopic analysis of traffic conflict caused by lane-changing vehicle at weaving section, 2002) in picking the PICUD parameters: a reaction time of $1.0s$ and a deceleration rate of $3.3m/s^2$.

Results

Figure 4 below shows the results of our experiments, where we compare the percentile reconstruction errors (higher equals more anomalous) of our autoencoder with TTC, THW, DRAC, and PICUD. Percentile values of the reconstruction error are taken to better visualize how they compare to the SMoS.

From Figure 4 we can observe an inverse relationship between SMoS frequency and autoencoder reconstruction error.

On the one hand, this leads to the correct tendencies, where the most severe cases have the lowest frequency and highest reconstruction errors. And this shows for all of our varied selection of SMoS (TTC, THW, DRAC & PICUD). On the other hand, this also results in infrequent and non-severe behaviour leading to higher reconstruction errors when compared to the most behaviours. The plots for THW and PICUD show this well.

When there is no ground truth on safe driving behaviour, the lack of interpretability of the autoencoder metric troubles validation efforts. There is no theoretical guarantee that higher reconstruction errors indicate more severe situations. Additionally, with this approach, one has to consider the many other dimensionalities of the data, especially those unrelated to severity, where you'd expect similar outcomes: a lower frequency leads to higher reconstruction errors. For the presented case, where we only focus on longitudinal vehicle pairs on highway driving taken from similar circumstances (i.e.: similar traffic flow & congestion), we can limit our data dimensionality, and therefore model, to be on a scale of severity. However, when applying this methodology to the much more multi-modal, vibrant, and varying context of urban driving, this became an infeasible task.

Whilst unable to solve with this approach, the requirements of a more theoretically grounded approach and the ability to cope with the complexities of urban driving, led us to explore and develop a metric based on driver profiling as described in the next section.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

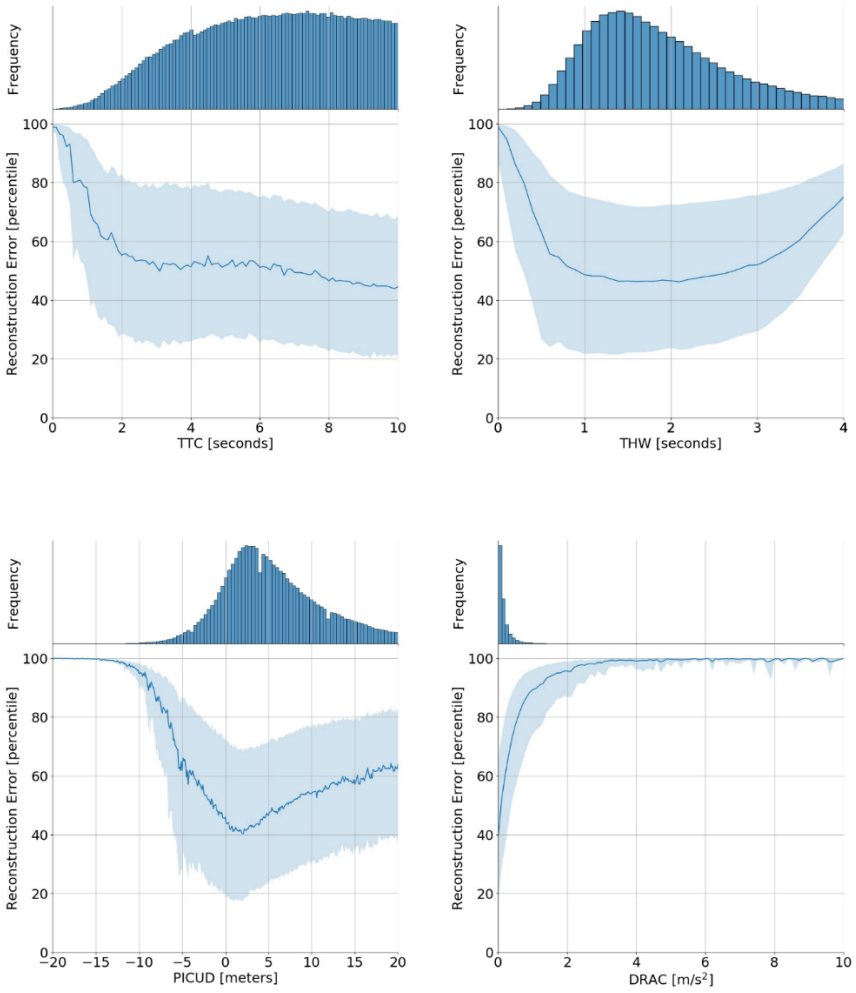

Figure 4. Autoencoder percentile reconstruction errors distribution for TTC, THW, DRAC, and PICUD. We plot the median reconstruction error within the interquartile range bandwidth. On top of each graph the marginal distribution of the corresponding risk metric is visualized.

2.2.3 Driver Profiling

Driver Behaviour Profiling by Clustering

Driver behaviour profiling is defined as *"a normalized score of driver behaviour which serves as a proxy for assessing crash risk by combining a number of measures of risk on a common scale within and across drivers."* (Driver behaviour profiles for road safety analysis, 2015). We focus on behaviour as a collection of interactions. This is different from the previous metric which focuses on singular interactions between a leading and following vehicle. This has as advantage that this measure is aggregated over different driving modalities, which makes the influence of these modalities less severe on the outcome than in the previous metric.

Behavior profiling relies on extracting features from data that are known to be proxies for critical scenarios. These features can then be used to create a driver profile per driver. Then

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

we can cluster these driver profiles in different clusters to make sense of the different kinds of profiles that are found in the data.

Our approach is based on the work of Warren et al. (Clusters of driving behavior from observational smartphone data, 2019) who cluster drivers based on how they compare to the norm across unique road segments. We believe selecting road segments is important to cope with the variety of urban road networks. In the rest of this section we describe our feature selection approach, naturalistic driving dataset, how we select road segments, calculate and cluster driver profiles. Finally, our clustering results are presented and interpreted.

Feature selection

Selecting profiling features is an important first step. To be able to describe and analyse clustering results, we need an in-depth understanding of the chosen features and how they are related to crash risk. Also, each feature should correlate with the actual behaviour.

For this metric, we pick velocity as our first feature. Higher velocities are associated with an increase in crash risk (Travelling speed and the risk of crash involvement volume 2-case and reconstruction details, 1997) (MoRT&H, 2018) (The Power Model of the relationship between speed and road safety: update and new analyses, 2009) (The effects of drivers' speed on the frequency of road accidents, 2000). Besides this, higher velocities are correlated with the impact velocity of a crash, leading to more severe crashes (Speed limits, enforcement, and health consequences, 2012) (Traffic safety dimensions and the power model to describe the effect of speed on safety, 2004).

Further, we select acceleration as a feature. Aggressive acceleration and deceleration events, when a driver applies more force to the pedal than during normal driving, are associated with a larger risk of collision (Driver behaviour profiles for road safety analysis, 2015) (Using naturalistic driving data to assess variations in fuel efficiency among individual drivers, 2010).

Our driving profiling techniques are based on statistics of speed and acceleration over different road segments. These variables are leading measures of safety, as they correlated significantly (in the statistical sense) with hard braking and (in-turn with) rear-end collisions (see Figure 5) (Safe Enough: Approaches to Assessing Acceptable Safety for Automated Vehicles, 2020).

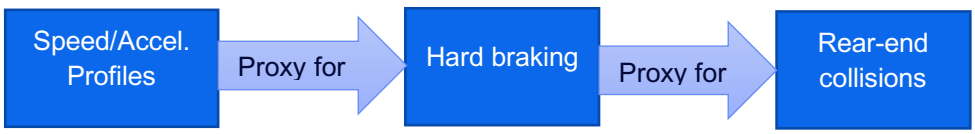

Figure 5: Leading measures and safety

Naturalistic Driving Data & Road Segmentation

We use the pNEUMA dataset for developing this metric (On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment, 2020). The pNEUMA dataset consist of half a million vehicles that have been collected from Athens

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

during morning peak-hour by the usage of drones. The dataset is recorded at 25 frames per second.

To prevent our model from finding disturbances within the data (either resulting from data quality or physical impossibilities), we use a Savitsky-Golay (Smoothing and differentiation of data by simplified least squares procedures., 1964) filter to smooth vehicle velocities over time. We use a first-degree Savitsky-Golay filter with a window length of two seconds (51 samples). The smoothed velocity is also used to calculate the acceleration of each vehicle.

Besides this, we filter out all the non-car-like vehicles from the dataset. This means that we only consider the vehicles with the 'car' or 'taxi' class.

We only use data from the main one-way roads in the dataset, because they are similar to the roads that will be analyzed in the simulations of the Task 2.5 .

The selected roads are divided up into segments. The aim for this is, just as in Warren et al. (Clusters of driving behavior from observational smartphone data, 2019), to compare the behaviour of the drivers per road segment. This is done to cope with the different kind of behaviours that span the different kind of road segments in the urban environment. This means that we cut up a road in different segments, dividing them based on where traffic lights, pedestrian crossings and intersections are on the roads.

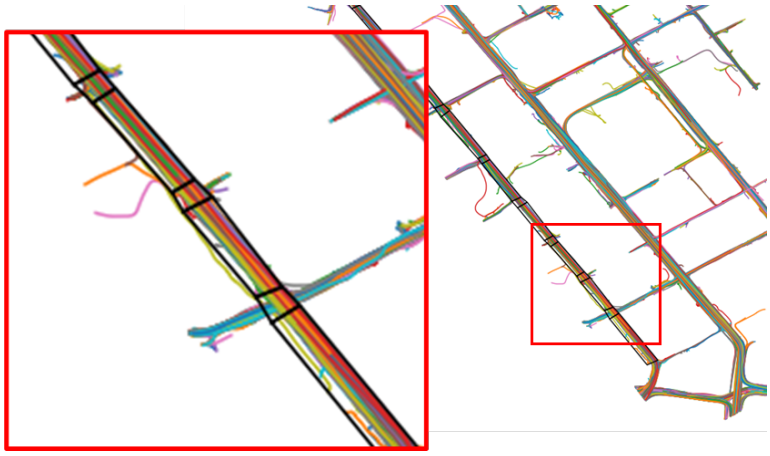

Figure 6: Part of the routes selected from the pNEUMA dataset. The black squares show how the segments are defined. They are defined based on the road structure.

Driver Profiles and Clustering

The driver profiles are constructed for each driver. As we are unable to connect multiple trips of one driver, we consider each trip as a separate driver. We define a trip $t \in T$ as a sequence of segments $t = (s_1, s_2, \dots, s_i, \dots, s_n)$, where each segment $s_i \in S$ the set of defined segments.

For every segment $s \in t$ and every trip $t \in T$, we calculate the following features:

- The mean of the trip's velocity within that segment $v_\mu^{s,t}$
- The standard deviation of the trip's velocity within that segment $v_\sigma^{s,t}$
- The mean of the trip's acceleration within that segment $a_\mu^{s,t}$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

- The standard deviation of the trip's acceleration within that segment $a_{\sigma}^{s,t}$

By using for both the speed and acceleration not only the mean but also the standard deviation as features, we also include the variety of the speed and acceleration values in the segment.

When calculating these features we remove all measurements of the vehicles when the speed at that timepoint is lower than 0.1 m/s. The time a drive stands still in certain segments can vary greatly in a segment based on whether, for example, a traffic light is red or green. If we, for example, have the speed profiles of two different trips in the same segments as in Figure 7, the behaviour when driving is the same, only the time the driver stood still is different. To make sure these two similar profiles also have similar features, we remove the measurement in a segment when the car is standing still.

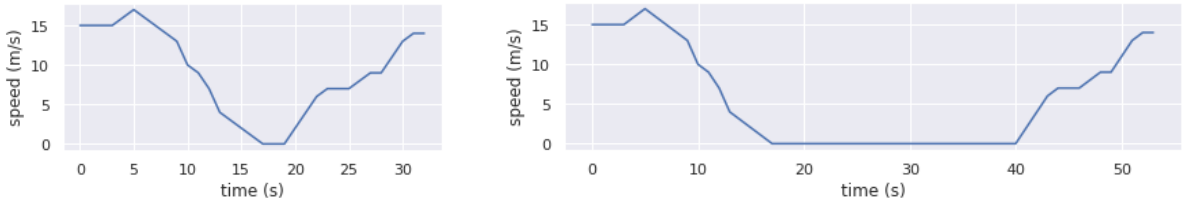

Figure 7: Two hypothetical speed measurements in a segments where the only difference is the time the cars stood still in front of a traffic light.

Each calculated feature $x \in \{v_{\mu}, v_{\sigma}, a_{\mu}, a_{\sigma}\}$ is then normalised per segment according to

$$x_{normal}^{s,t} = \frac{x^{s,t} - \overline{x^{s}}}{\sigma_{x}^{s}}$$

Where:

$\overline{x^{s}}$ is the mean of feature x of all trips in segment s.

σ_{x}^{s} is the standard deviation of feature x of all trips in segment s.

This is done to compare the behaviour of a driver in that segment to the other drivers within that segment. As shown in Figure 8, this makes these segment-based features relative to the population of trips that drove through that segment. For example, the red car in Figure 8 has a relatively low mean velocity, but a high acceleration.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

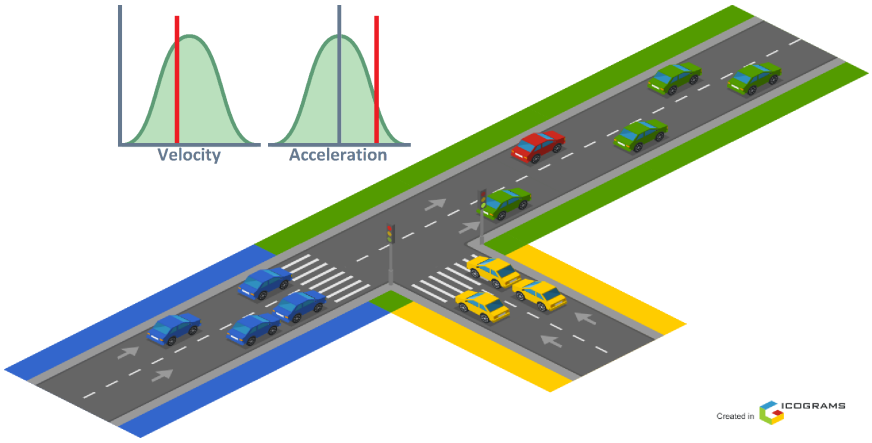

Figure 8: Visualization of comparison of ego vehicle (red) average feature values versus the norm for its current segment (green). Blue and yellow indicate different road segments. Image partially created with icograms.com.

Then, for each trip t and each feature $x \in \{v_\mu, v_\sigma, a_\mu, a_\sigma\}$ all normalised values per segment are aggregated to the whole trip as follows:

$$x^t = \frac{1}{N} \sum_{s}^{s \in t} x_{normal}^{s,t}$$

Such that for every trip, we have the feature vector $(v_\mu^t, v_\sigma^t, a_\mu^t, a_\sigma^t)$ as a representation of that trip driver profile. Using the driver profiles for each trip we use unsupervised clustering to discover different types of driver profiles. We use the K-means (Least squares quantization in PCM, 1982) algorithm for clustering. The elbow method is used to determine the number of clusters.

Results

Figure 9 shows the four different features of the diver profiles plotted together. Each dot is a different trip. In Table 2 we can see the correlation between the different features of the diver profiles. We see that there is a strong positive correlation between the acceleration standard deviation and velocity standard deviation. Besides that, there is a negative correlation between the mean velocity and both the standard deviation of the velocity and the standard deviation of the acceleration.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Table 2: Correlation between the features of the diver profiles

	Mean velocity	Standard deviation velocity	Mean acceleration	Standard deviation acceleration
Mean velocity	1	-0.34	0.09	-0.23
Standard deviation velocity	-0.34	1	0.06	0.84
Mean acceleration	0.09	0.06	1	0.01
Standard deviation acceleration	-.023	0.84	0.01	1

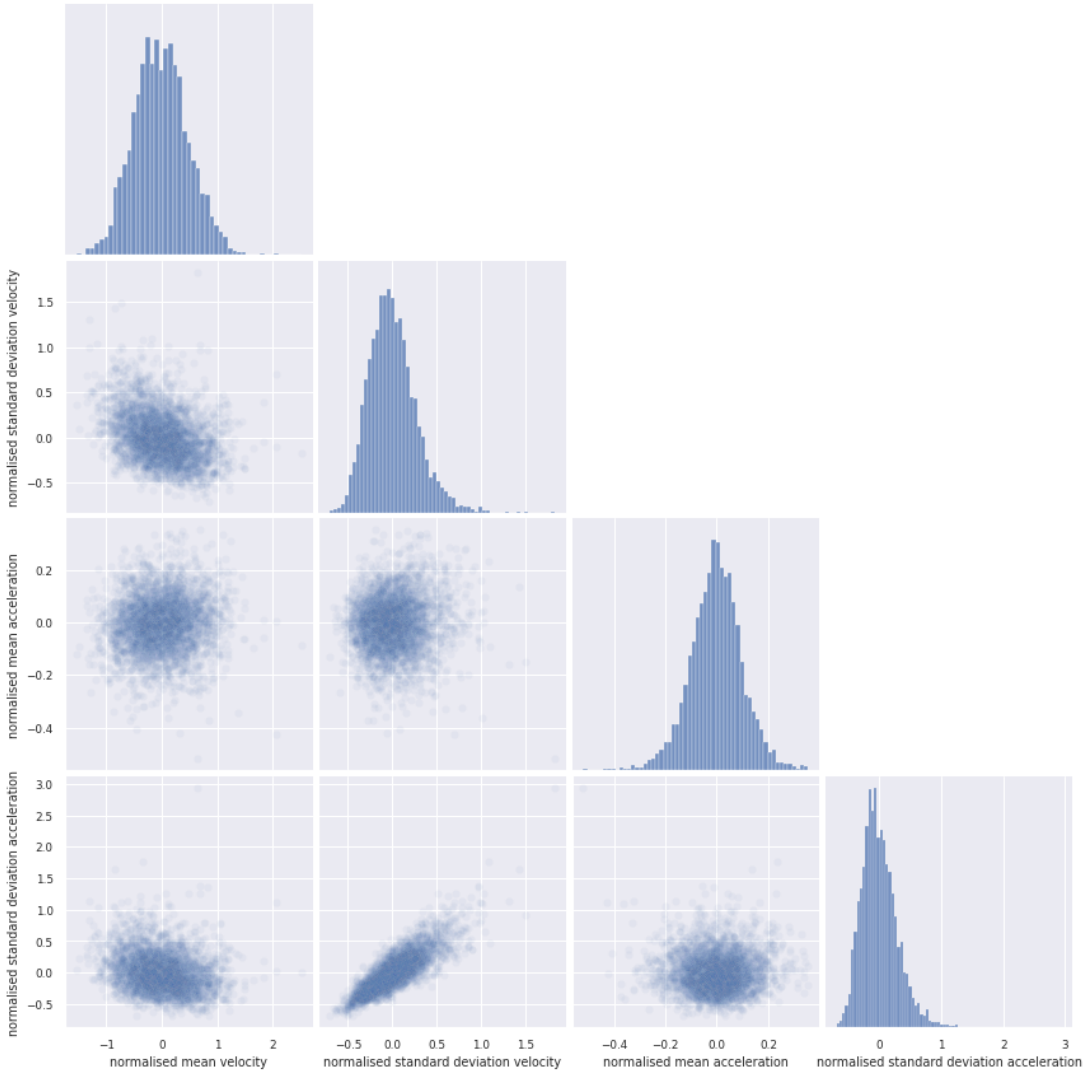

Figure 9: Plot of all driver profiles based on their four features and the combinations. Each dot is one trip.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Using the elbow method, four clusters were found in the data. Figure 10 shows these clusters for the four features. The purple cluster consists of 1210 trips, the red cluster includes 1216 trips, the blue cluster 998 trips, and the green cluster 521 trips. Table 3 shows the mean for each feature in each cluster.

Table 3: Mean of each driver profile feature for each cluster.

Cluster	Mean velocity	Standard deviation velocity	Mean acceleration	Standard deviation acceleration
Purple	-0.67	0.03	-0.64	-0.01
Red	0.69	-0.93	-0.10	-0.87
Blue	0.29	0.29	0.89	0.27
Green	-0.77	1.69	-0.10	1.66

For the four clusters, we can make the following observations:

- The purple cluster consists of profiles with an average standard deviation of the velocity and acceleration per segment, but it has a low average speed and acceleration.
- The red cluster is characterised by a below-average standard deviation of velocity and acceleration, but has a high average velocity. We see in Figure 11 that most of these trips happen earlier in the morning with the peak of the distribution at a starting time between 8:00 and 8:30.
- The blue cluster has all features at above-average values.
- The green cluster is mainly characterised by an above-average velocity standard deviation and above-average acceleration standard deviation. But with that, it also has a low mean velocity. This could include trips in which the driver has to start and stop a lot. What is also very noticeable is that in Figure 11 this cluster is skewed to later starting times. With the peak of the distribution around 9:30 to 10:00.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

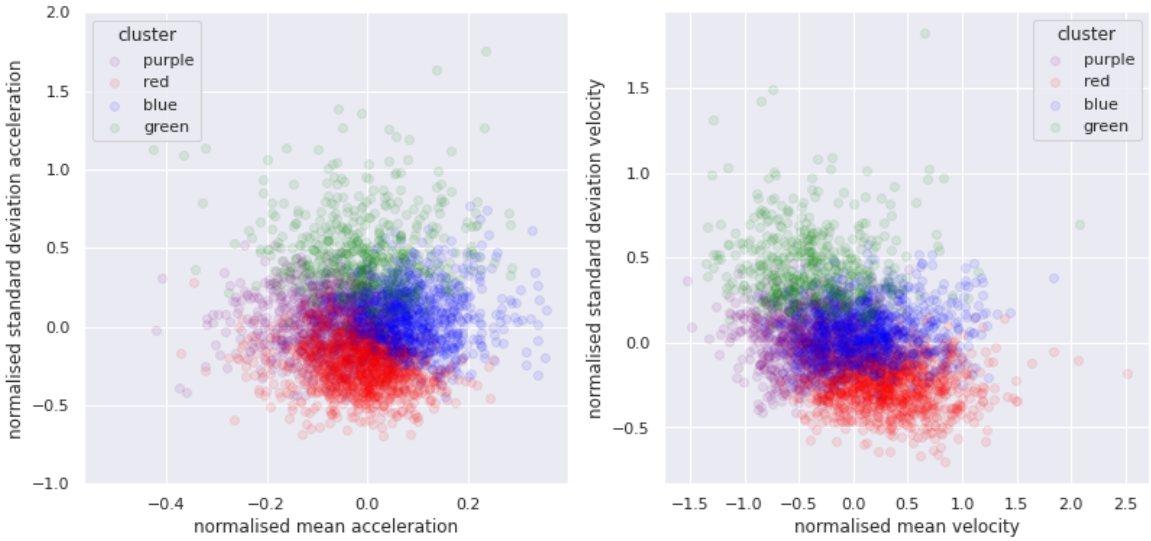

Figure 10: Plot of all driver profiles clustered in four clusters. Each dot is one trip.

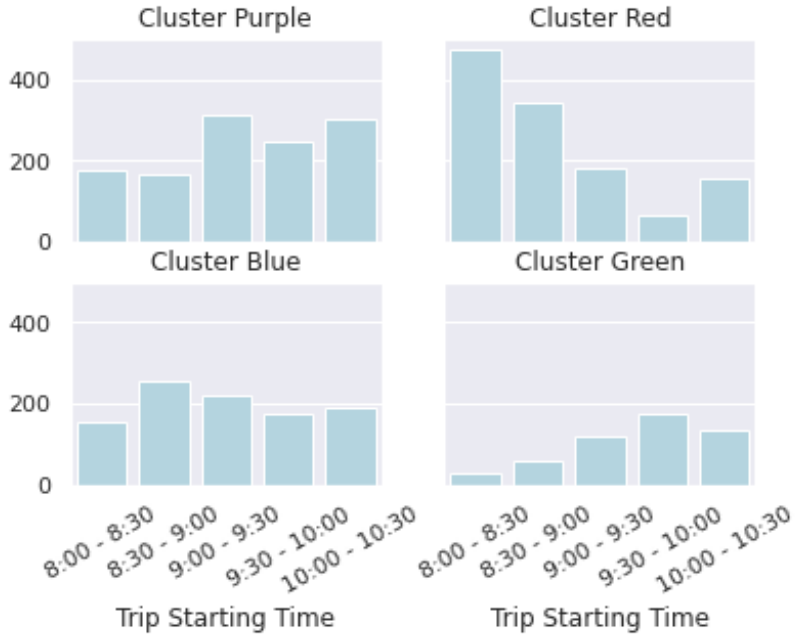

Figure 11: starting time of different trips per cluster.

Discussion and conclusion

Based on these made observations we could interpret the clusters in the following way:

The purple cluster has trips in it with driver profiles with average or lower than average features. Because high values for these features are proxies for unsafe driving, it is possible to interpret this cluster as a cluster that consists of divers with a typical or safe driving style.

In contrast to that, the blue cluster consists of profiles which on average score higher on the different features. As a high score on these features is a proxy for safety-critical driving, we could say that this cluster consists of the more unsafe trips that are recorded in the dataset.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

However, as we only look at the relative measurement of these features it is not possible to say anything about the absolute difference between the safety of driver profiles. It only shows behavioural tendencies.

For the red and green clusters, we see something else happening. Both clusters are correlated with the time of driving. Drivers in the red clusters seem to have made their trip between 8:00 and 9:30, while the green cluster has the majority of the drives later in the morning, between 9:00 and 10:30. This shows that these profiles are dependent on a variable that is also dependent on the time of the trip. This can be something like the heaviness of the traffic, or the number of other kinds of road users like scooters.

This means that these clusters should not be interpreted as reflecting only the drivers' behaviour, as it also reflects a road modality (e.g. amount of traffic). This problem could be alleviated if we have more data than only one trip per driver. If we have more trips per driver, a driver will encounter also a lot of different traffic amounts and thus this variable will be not as present in the driver profiles. Another way to remove this variable from influencing the driver profile is to only consider trips in a specific traffic amount. This is an important lesson to keep in mind when using the driver profiling in the simulations in Task 2.5.

We can conclude that using driver profiling to validate the driving behaviour of drivers can be useful. We were able to find two clusters which could be interpreted as containing safe driving behaviour and unsafe driving behaviour. But we also see that there is ample opportunity to further develop the method, as it can still be challenging to capture the drivers' behaviour independent of the driving modalities like the kind of road or the traffic situation.

2.2.4 Reflection

Automated vehicles with increasingly higher levels of automation will need to prove in the future that they follow safe driving behaviour principles. However, it is yet undefined what safe driving means. We reason that when an automated vehicle drives like a typical human being, it has the correct driving behaviour and its majority of driving will be safe (Figure 1).

Two methods were introduced to model the typicality of human driving. Our first method, as described in Section 2.2.2, is based on unsupervised anomaly detection. By detecting the abnormal situations in longitudinal vehicle following, atypical and potentially dangerous situations can be detected. Whilst this works well in a controlled environment such as rush-hour highway driving, handling the multimodality of urban driving and the lack of a theoretically justified model interpretation are the largest drawbacks of this method.

Our second method, as presented in Section 2.2.3, tries to overcome these drawbacks by using driver profiling. This method looks for differences in the behaviour of drivers during their entire trip by constructing driver profiles. Subsequently, we cluster all profiles and find common driver profiles. Due to our selection of features being aligned with the safety of driving, we can indicate which driver profiles are considered safe or unsafe. Our study shows that our approach managed to find driver profiles indicating certain driving behaviour. However, the multimodality of driving, especially related to traffic congestion, lead to time-specific driving profiles. We believe more data (i.e.: multiple trips per driver) could limit these effects. Due to being better interpretable, promising initial results, and the option to simulate

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

sufficient data per driver, we picked this metric for our work in T2.5 as described below (Section 2.2.5).

2.2.5 Preliminary plans for T2.5

This method for driver behaviour profiling described in Section 2.2.3 will be used on the simulations to describe and cluster the behaviour of the agents within the simulation. For each variation of the simulation, with and without AVs, we will cluster the agents based on the previously described features that are correlated with unsafe driving behaviour. With these profiles two separate analyses will be done.

Firstly, we will look at profile clusters within one simulation. Here the main objective is to validate if there is a correlation between parameters of initialized agents which influence the safety of the agents' behaviour and the found driver clusters.

Secondly, the profiles from clusters of different variations of simulations will be compared. The aim is to see if replacing part of the agents for AVs will change the driving behaviour of both the AVs as well as the other agents in the simulation. In the driving profiling method, this will be indicated by a change of clustering.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

2.3 Metrics for Motor Vehicle Interactions (TUD)

2.3.1 Motivation

Traditionally, traffic safety studies for interactions between motor vehicles (car-to-car interactions in short) relied on the records of reported vehicle crashes, field operational test data and naturalistic driving data, etc (Dingus, et al., 2016; Mullakkal-Babu, Wang, He, Arem, & Happee, 2020). Meanwhile, based on the simulated vehicle trajectories, the car-to-car interactions can be quantified by various metrics regarding different aspects, e.g., potential time to collision, collision probability and crash severity. Such measures are known as SMoS (Gettman & Head, 2003; Laureshyn, et al., 2016), and a prominent example is TTC. These metrics indicate a potential conflict between two road users.

However, during on-road driving, the neighbouring traffic environment, especially for the neighbouring vehicles, can vary dynamically. The SMoSs for motor vehicle interactions, which are defined based on the predicted motions of interacting vehicles, are suitable for this purpose, i.e., they can be calculated at each moment during an encounter. The subject vehicle, whose safety is to be assessed, is not certain about the future motion of its neighbouring vehicles and consequent crash outcome. Uncertainty, therefore, is an inherent component of the driving risk estimate. Most SMoS do not typically account for this uncertainty.

We propose an approach named Probabilistic Driving Risk Field (PDRF) to assess the driving risk for car-to-car interactions. The approach uses vehicle states (including vehicle position, velocity and mass) as input and takes into account the probability of motion predictions of neighbouring vehicles.

2.3.2 Technical Description

Using the paradigm of Artificial Field Theory (Dunias, 1996), we propose an approach named PDRF to assess the driving risk for car-to-car interactions. The proposed SMoS PDRF takes into account the probability of motion predictions of neighbouring vehicles, and consists of a crash severity term and a collision probability term (ISO 26262, 2018). The subject and neighbouring vehicle's possible positions and associated probabilities at discrete future time steps are predicted. The collision probability is modelled as the likelihood of overlap in the predicted spatial configurations of the subject and neighbour vehicle, under the motion prediction results. The risk can be estimated for a single time step or over multiple future time steps, depending on the required temporal resolution.

To calculate the collision probability in PDRF, the motions of the subject and neighbouring vehicles are predicted separately. For the subject vehicle, if a motion planner is equipped and provides a planned trajectory, its future motion of the subject can be known accordingly; otherwise, we propagate its future motion based on its current vehicle state.

The dynamic state of a neighbouring vehicle is denoted by the position (x, y) of its center of mass and the velocity (v_x, v_y) along the longitudinal and lateral directions, resulting in

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

$$\begin{cases} \frac{d}{dt}\begin{pmatrix} x \\ v_x \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{pmatrix} x \\ v_x \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} a_x \\ \frac{d}{dt}\begin{pmatrix} y \\ v_y \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{pmatrix} y \\ v_y \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} a_y \end{cases},$$

where a_x and a_y are the accelerations along with the longitudinal and lateral directions respectively. The dynamic state is to be propagated from the current time instant t to a future time $t + t_f$.

The neighbouring vehicle motion is also subject to several physical constraints, including non-holonomic behaviour, backward motion prohibition, and acceleration range limitation.

- $-0.17v_x \leq v_y \leq 0.17v_x$ representing the non-holonomic behaviour of the vehicle. This condition assumes that the vehicle heading angle is approximately represented as $\beta = arctan\frac{v_x}{v_y}$ and is bounded as $|\beta| \leq 10^\circ$, during motorway driving.
- $v_x \geq 0$ representing the strictly forward movement.
- $a_x^{min} \leq a_x \leq a_x^{max}$ representing the feasible acceleration range that is restricted by the engine power and brake torque limitations.

The probability functions of acceleration per vehicle can be estimated by treating acceleration signals as a random variable (Analyzing human driving data an approach motivated by data science methods, 2016). Given the predicted future motions of the neighbouring vehicles, PDRF further assumes their accelerations have a normal distribution under a specific mean (typically set as 0 since we do not have trajectory planning information of the neighbouring vehicle) and standard deviation (Analysis of effects of driver/vehicle characteristics on acceleration noise using GPS-equipped vehicles, 2010). For a single future time instant, the collision probability can now be calculated as the double integral of the probability density functions (PDF) of the neighbouring vehicle within the intersection area.

Here, the value of the acceleration variability distribution for a particular value of acceleration is interpreted as the relative likelihood of occurrence. We consider "collision likelihood" defined as the relative likelihood of a neighbouring vehicle n applying acceleration:

$$a_{x,n} = \frac{\delta x - \delta v_x t_f - 0.5 a_{x,s} t_f^2}{0.5 t_f^2},$$

$$a_{y,n} = \frac{\delta y - \delta v_y t_f - 0.5 a_{y,s} t_f^2}{0.5 t_f^2},$$

where $a_{x,n}/a_{x,s}$ and $a_{y,n}/a_{y,s}$ denote the accelerations of the neighbouring vehicle n / subject vehicle s in longitudinal and lateral directions, δx and δy the relative spacing, δv_x and δv_y the relative velocity in longitudinal and lateral directions, and t_f the future prediction horizon. Considering $a_{x,n}$ and $a_{y,n}$ as random variables, the collision likelihood c can be defined as follows:

$$c = \mathcal{N}\left(\frac{\delta x - \delta v_x t_f - 0.5 a_{x,s} t_f^2}{0.5 t_f^2} | \mu_x, \sigma_x\right) \cdot \mathcal{N}\left(\frac{\delta y - \delta v_y t_f - 0.5 a_{y,s} t_f^2}{0.5 t_f^2} | \mu_y, \sigma_y\right),$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

where $\mathcal{N}$ is the PDF of normal distributions, and its parameters μ denotes the mean and σ denotes the standard deviation. Note these parameters can either be pre-defined through a calibration process, or dynamically updated using advanced deep neural networks for probabilistic motion predictions (Probabilistic Risk Metric for Highway Driving Leveraging Multi-Modal Trajectory Predictions, 2022).

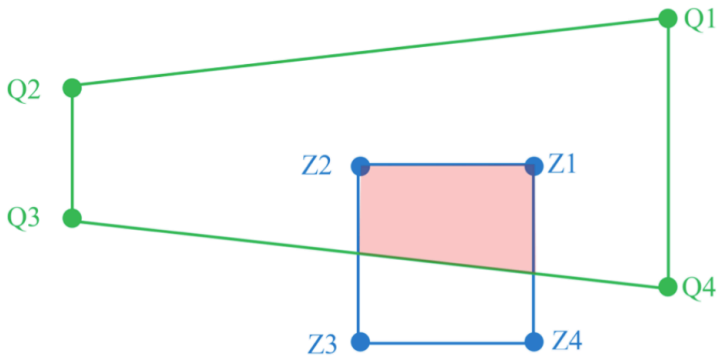

Figure 12. Geometric representation of the polygons Q, Z and their overlap O (area shaded in pink).

However, the vehicle possesses a geometry and a crash is the overlap in geometric boundaries of vehicles. Further considering the collision probability of the vehicle model and imposing the motion constraints, the boundary of the reachable state of n at time $t + t_f$ can thus be represented as a quadrilateral polygon *Q* (see Figure 12). Using the predicted position of s at $t + t_f$ along with the geometry of the two vehicles, we define the potential collision zone: *Z*. Thereby, the collision probability is non zero if there exists an overlap in *Q* and *Z*. The overlapping region is another polygon denoted by *O* defined in the spatial domain. The overlap *O* in the spatial domain has to be converted to acceleration domain denoted a^O for collision probability estimation by using the following relation.

$$a_x^c = \frac{x^c - x_n - v_{x,n}t_f}{0.5t_f^2}, a_y^c = \frac{y^c - y_n - v_{y,n}t_f}{0.5t_f^2} \tag{1}$$

where x^c and y^c denote the corner positions of *O*, a_x^c and a_y^c denote the corresponding corners of a^O. Eq. (1) specifies a linear relationship between acceleration coordinates and the spatial coordinates, and therefore a^O is also a quadrilateral polygon. Then the collision probability c can be obtained by integrating the joint acceleration variability distribution over a^O as follows:

$$c = \mathcal{N}(da_x, da_y).$$

More details of the collision probability calculation can be found in (Probabilistic field approach for motorway driving risk assessment, 2020).

The expected crash severity s considering the vehicle mass and velocity is constructed as

$$s = 0.5M\beta^2(\Delta V)^2 \tag{2}$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

where M is the mass of subject vehicle, $\beta = \frac{M_n}{M_n+M}$ the mass ratio with M_n denoting the mass of neighbouring vehicle, and ΔV the relative velocity between subject and neighbouring vehicles. In practice, the information of neighbouring vehicles can be obtained by various sensors and corresponding processing algorithms in the subject vehicle, or via communications between vehicles. The establishment of crash severity in Eq. (2) is under an assumption that the collision is inelastic without secondary crashes due to bounce back, indicating both the vehicles would move together after the crash. Besides, the relative velocity between the vehicles is calculated using the current velocities at time instant t.

The PDRF metric is finally calculated as the product of the collision probability and the expected crash severity,

$$\mathrm{PDRF} = c \cdot s,$$

where c is the predicted collision probability at time $t + t_f$.

Calibration and validation

To calibrate and validate the proposed car-to-car interaction SMoS PDRF, we consider near-crash scenarios from both simulations and naturalistic driving data.

A set of safety-critical hard braking events were simulated. The subject vehicle travels on a lane, and a leading vehicle travels in the front on the same lane both starting with a constant longitudinal velocity. Each simulated event lasts for 15 seconds. At t = 6 s, the lead vehicle abruptly applies a constant deceleration of 5 m/s^2 to stop. The subject vehicle maintains constant velocity at the beginning, since we aim to evaluate the driving risk when the subject does not take any actions. At the final stage of the event, the ego brakes with 5 m/s^2 to stop just at the end of the event. In practice, the subject vehicle could react earlier in line with the obtained driving risk, as we discuss later. We vary 3 parameters of this scenario. The initial spacing between the two vehicles is set as 20, 40, 60 and 80 meters. The initial velocity of the preceding vehicle is varied between 5–30 m/s with an interval of 1 m/s, and the maximum velocity of the two vehicles corresponding to each of the initial spacing is 10 m/s, 16 m/s, 23 m/s and 30 m/s respectively. This results in 1217 simulations (36, 144, 361 and 676 simulations for different initial spacing respectively). To avoid crashes before hard braking, the minimum initial time headway between the two vehicles is set to 3 s. Therefore, the initial velocity of the subject is varied from a minimum of 5 m/s to a maximum of 30 m/s. The simulated scenarios thus include both crash and non-crash events. To set a ground truth, the crash events are labelled as Positive risk, and the non-crash events are Negative risk. In line with preliminary experiments and to achieve a higher crash prediction accuracy, TTC < 3 s and PDRF > 100 Joules are then selected as the threshold to identify risk events. We quantify the accuracy of risk estimates in terms of the True Positive, True Negative, False Positive and False Negative risk description.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

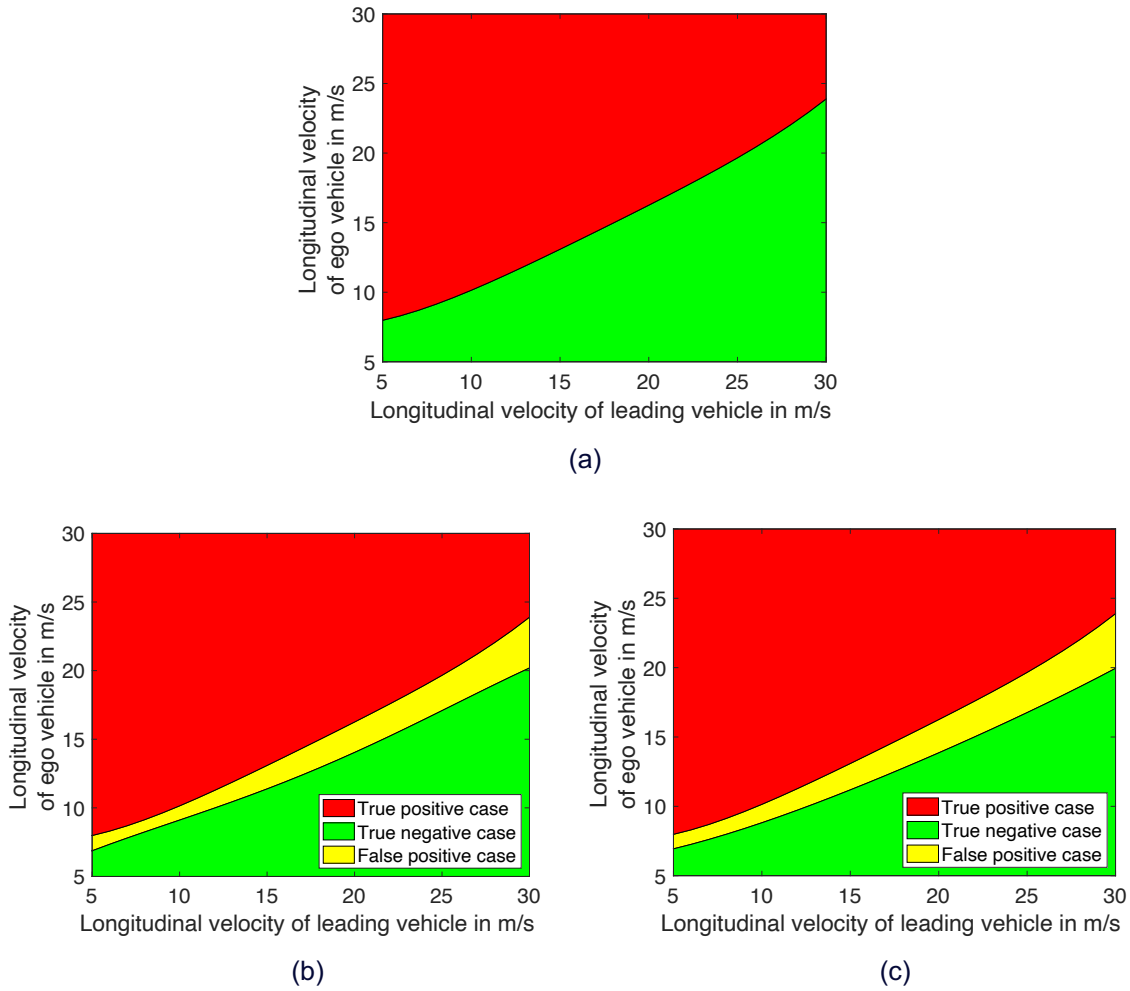

Figure 13. Overview of risk description by PDRF and TTC in hard braking scenario simulations. (a) Simulations in which preceding vehicle (80 m ahead) brakes hard. As a ground truth, crashes are in red (Positive risk) and no crashes are in green (Negative risk). (b) Risk description with PDRF. (c) Risk description with TTC.

As seen in Figure 13, both PDRF and TTC result in few false-positive cases, particularly in those scenarios in which the relative velocity between the subject and preceding vehicle is relatively small. These false positives represent high risk interactions leading to near-miss events. The acceleration distribution of the preceding vehicle is calibrated as $\mathcal{N}(a_x|0,2)$ and $\mathcal{N}(a_y|0,0.2)$.

We further validate PDRF using real-world hard braking event during trip number 8427 in the 100 vehicles naturalistic study (Neale, Dingus, Klauer, Sudweeks, & Goodman, 2005). The subject vehicle is following the lead vehicle fairly closely on a two-lane road when the preceding vehicle slows down to stop. The subject vehicle brakes to avoid hitting the preceding vehicle in the rear. Both TTC and PDRF risk estimates qualitatively reflect the event narration.

During the initial phase (0 to 4 s), the subject was closely following the preceding vehicle indicating an unsafe interaction. During this phase, both PDRF risk with different prediction time step t_f(see Figure 14(d)) and TTC (see Figure 14(e)) indicate the existence of risk. Thereafter the preceding vehicle slows down to stop; this unsafe development is described

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

as a gradual descent in TTC (time 4–6 s) and as a temporally adjacent rise in the severity (Figure 14(b)) and the PDRF risk descriptions (t_f=4, 3 and 2 s). Evaluating the multiple descriptions of risk (see Figure 14(d)) and crash probability (see Figure 14(c)) reveals how this situation evolves in terms of driving risk. The low peak in crash probability with t_f=2 s indicates marginal chances of an imminent crash (i.e. at the next 2 s). Moreover, a high-risk peak with t_f=3 s around 5.6 s and a subsequent lower risk peak with t_f=2 s suggests that the hard braking, which started around 5 s, is sufficient to evade the danger. In combination, these observations indicate that the braking was effective to avoid the danger. The absence of risk with t_f=1 s indicates the collision was successfully avoided at least until the end of the observation. PDRF risk with t_f=3 s attained the highest estimate and its moment of maximum risk (see Figure 14(c)) is closer to the moment when the subject begins the braking around 5 s (see Figure 14(f)). It can be seen that the PDRF risk model could qualitatively reflect the event narration via modelling both the crash possibility and the crash severity, and its risk description was consistent with that by TTC.

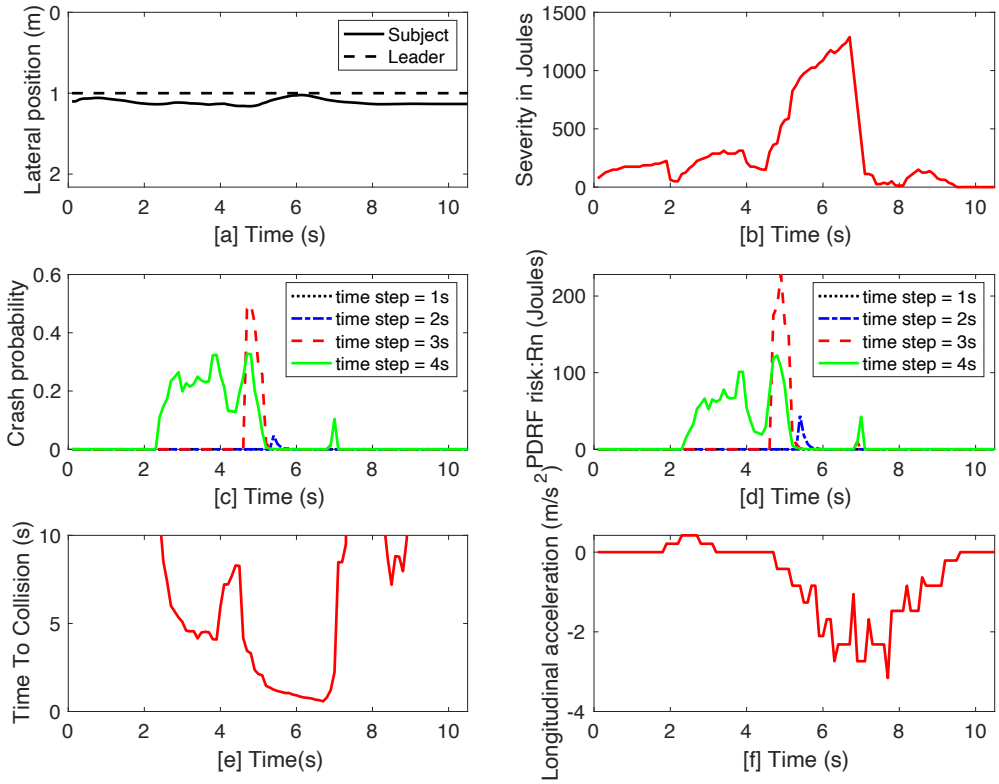

Figure 14. Risk estimates of an encounter in which the subject vehicle avoids a rear-end crash by braking.

The prediction time horizon t_f influenced the PDRF risk descriptions. Notably, the PDRF risk estimate with t_f=3 s yielded a single peak, attained the maximum value, and was temporally closest to the moment when the driver initiated the evasive manoeuvre. The proposed PDRF will be further validated with safety-critical events created in the simulation platform in T2.5.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

2.3.3 Preliminary plans for T2.5

The identification and/or development of criticality metrics for driving interactions is one of main goals of T2.2. In T2.5, such metrics can be used to identify safety critical interactions through simulations. Specifically, for the developed SMoS PDRF for car-to-car interactions, the general procedure to employ PDRF is as follows.

- Generate benchmark simulation scenarios for calibrations. The safety-critical car-to-car interactions have been identified in advance by manual labelling or other classic SMoSs, e.g., Surrogate Safety Assessment Model (SSAM) module.
- Pre-process trajectory files from simulation platform Aimsun. To enable offline simulation analysis, the generated trajectory output with labels should be further processed before using as input for PDRF.
- Calibrate parameters of the PDRF, including prediction time horizon, normal distribution parameters, and the threshold for safety-critical event identifications. Check whether the calibrated PDRF has a comparable/advanced performance compared to metrics in SSAM.
- Employ PDRF in line with the overall simulation analysis procedure designed in T2.5. Ideally, PDRF will be employed to identify car-to-car interactions for a set of scenarios, where initial traffic settings, especially the penetration rate of AVs, vary. Distributions of the PDRF corresponding to different trajectory settings are to be obtained, and their corrections are to be analysed. In doing so, certain particular properties in future traffic scenarios involving AVs are expected to be revealed.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

2.4 Metrics for VRU Interactions (IKA)

2.4.1 Motivation

Unlike in the area of car-to-car interactions, the literature available showed a low level of maturity looking at VRU-to-traffic interactions. Nevertheless, research into high-risk situations for pedestrians and cyclists is of particular relevance, as they are especially vulnerable in potential collisions (Cyclists' anger experiences in traffic: the cycling anger scale. Transportation research part F: traffic psychology and behaviour, 62, 564-574, 2019). Collisions with motorized traffic pose the greatest risk, which is why the IKA focuses on VRU-to-car interactions. For this reason, in T2.2 several measures based on the car-to-car interaction literature that are capable of describing the risk of VRU-to-car interactions were selected. These will be described in detail in the following section. The goal is to apply already known measures in the field of VRUs and thus to develop them further within the scope of T2.5.

2.4.2 Technical Description

In the SAFE-UP project, the IKA did not want to focus just on the observable behaviour but also wanted to consider the emergence of VRU's behaviour. For this reason, an extensive literature review was conducted at the beginning of T2.2. As described in deliverable D2.5, the literature shows different aspects of human behaviour than can have an impact on the emergence of safety critical traffic situations. First of all, the VRU's subjective evaluation of the situation (e.g. the perceived risk/ perceived criticality) is relevant if you look at Car-VRU interactions (Kovácsová, et al., 2019; Oron-Gilad, et al., 2020). Furthermore, the individual emotional state can influence human behaviour and their decisions in traffic situations (Dittrich, 2020). Regarding the observable measures, one can consider physiological values (e.g. heart rate; (Doorley, et al., 2015)) or objectivley measured VRU behaviour like trajectories (Abadi, et al., 2019) reaction times (Oron-Gilad, et al., 2020), speed (Abadi, et al., 2019) and aggressive behaviour (Dittrich, 2020) as indicators for upcoming potential critical traffic situations.

Since these factors are primarily indicators of the occurrence of a critical situation due to VRU behaviour, this approach was combined with objective-based measures. For this reason, the participant study in T2.3 was not only used to collect data for the calibration of the predefined pedestrian and bicyclist model but to investigate the influence of the emotional state frustration on the actual human's behaviour as well. Furthermore, one can find the mentioned human factors in the chosen simulation model behaviour. For example, recent research showed that the evaluation of the situation is crucial for the crossing decision of pedestrians (Zhuang, et al., 2020) and therefore determines the criticality of crossing situations as measured using the deceleration to safety time (DST), which is used for our simulation model. Even if the deeper investigation of these correlations between human factors and the resulting traffic behaviour were not in the focus of this project, this example shows the importance of considering this approach for further research projects.

The chosen measures to assess the criticality of the resulting traffic interactions between VRUs and cars are described in detail in the following.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

To evaluate the critical interactions between vehicles and VRU's, the known metrics from a car-to-car interaction are picked up and adapted to the use case based on the specific situations of the simulation (Lu, et al., 2021). For an unambiguous and direction independent determination of the individual metrics, a moving coordinate system is chosen. This coordinate system corresponds to the reference line coordinate system defined in OpenDRIVE, which follows the course of the road (ASAM, 2021). The "s coordinate" corresponds to the longitudinal coordinate and the "t coordinate" to the lateral. The velocities v_s and v_t are understood similarly. For a better understanding, the individual parameters are shown in Figure 15Figure 17.

For the pedestrian, the metrics from Table 4 have been selected to identify and analyse potential critical interaction with a vehicle. The focus is only on scenarios in which the pedestrian crosses the road. Displayed are the metrics with their units and the mathematical relationship. Below is the description of the metrics and the corresponding modifications:

- Deceleration to Safety Time (DST):
  The DST describes how much an object has to decelerate/accelerate in order to reach the conflict point just after the other object has left it *(Hupfer, 1997; Kotte, et al., 2018)*. This evaluation metric is suitable for scenarios in which the trajectories of the two objects cross and can thus be directly transferred to a situation in which a pedestrian crosses the road.

- Post Encroachment Time (PET):
  PET describes the time difference between a pedestrian ($time_{Ped}$) leaving the crossing zone and a vehicle ($time_{Veh}$) entering it. In this case, the trajectories of the pedestrian and the vehicle cross.

- Time Head Way (THW):
  THW describes the time gap to a preceding object. In our application, the interpretation of THW is modified so that the pedestrian is a standing object on the road. Thus, the THW between the approaching vehicle and the standing pedestrian is evaluated. The calculation of the metric is valid as long as the pedestrian is walking on the road. The modified THW is here theoretically equivalent to the Time to Collision where the pedestrian has no speed.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

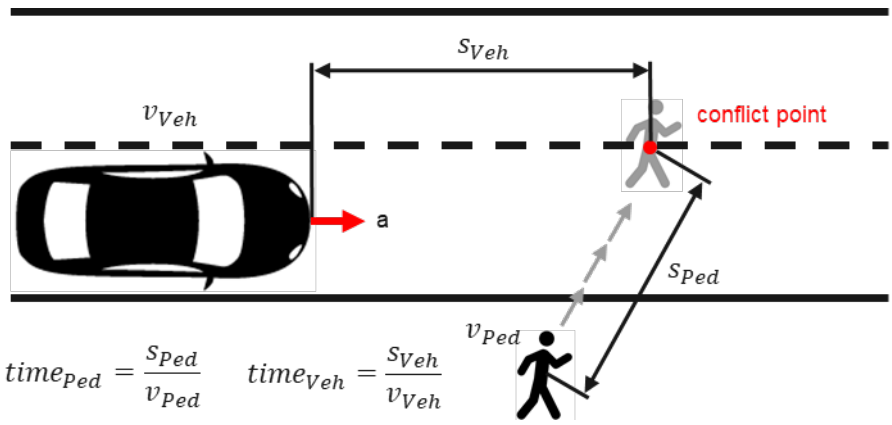

Figure 15: Schematic illustration of the used parameters and variables for a vehicle to pedestrian interaction

Table 4: Evaluation metrics for the potential collision with a pedestrian

Metrics	Unit	Equation
Deceleration to Safety Time (DST)	$\frac{m}{s^2}$	$DST = a = \frac{2 * (s_{Veh} - v_{Veh} * time_{Ped})}{time_{Ped}^{\ 2}}$
Post Encroachment Time (PET)	s	$PET = time_{Ped} - time_{Veh}$
Time Head Way (THW)	s	$THW = time_{Veh} = \frac{s_{Veh}}{v_{Veh}}$

For the bicyclist, the metrics from Table 5 were selected to evaluate the simulation. Just as for the pedestrian, the units and corresponding mathematical formulas for calculation are shown in this figure. The gray rows include potential longitudinal collisions and the green lateral collisions. Below is the explanation and interpretation of the individual metrics:

- Time to Collison (TTC):
  The remaining time to an imminent collision between a vehicle and a cyclist, if the current driving condition (with respect to velocity, direction etc.) does not change.
- Time Head Way (THW):
  Description of time gap from a vehicle to the preceding cyclist.
- Post Encroachment Time (PET):
  The difference between the time a cyclist leaves a conflict point ($\mathrm{time_{Bike}}$) until the time a vehicle arrives to this point ($\mathrm{time_{Veh}}$).
- Deceleration rate to avoid collision (DRAC):
  The rate at which a vehicle must decelerate to avoid a probable collision with a cyclist. Equation applies to vehicles traveling on the same path.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

- Proportion of stopping distance (PSD):
  The ratio of distance available between two objects and the distance that is required to avoid a collision with the maximum available deceleration rate (MADR). MADR depends on vehicle type and environmental conditions such as pavement skid resistance.

- Lateral Time to Collison (LTTC):
  Same as TTC only adapted to a potential lateral collision. Accordingly, not longitudinal distances and velocities are considered here, but the respective lateral components.

- Lateral Time Head Way (LTHW):
  Same as THW only adapted to a potential lateral collision.

- Lateral Distance (LD):
  Absolute lateral distance between cyclist and vehicle that becomes relevant during overtaking manoeuvers.

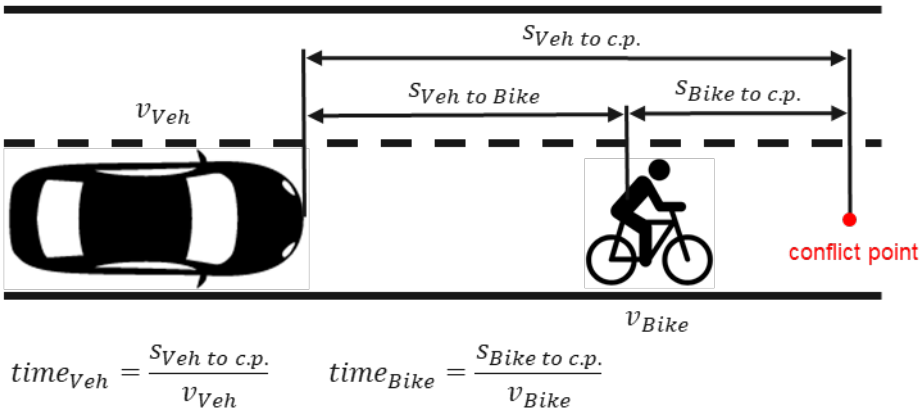

Figure 16: Schematic illustration of the used parameters and variables for a vehicle to cyclist interaction (longitudinal)

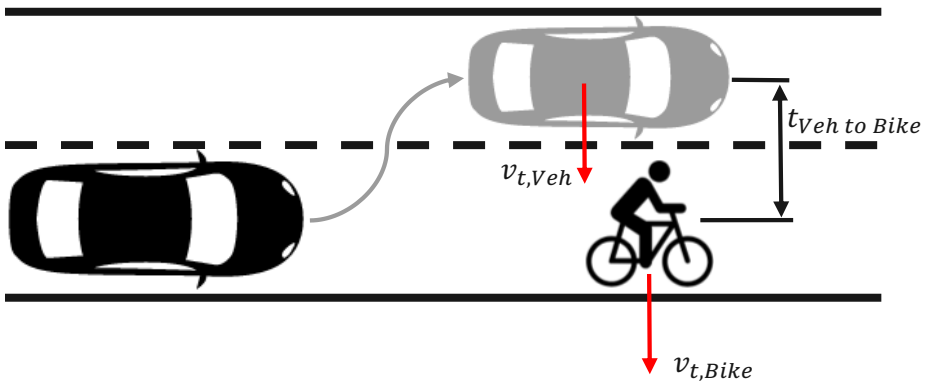

Figure 17: Schematic illustration of the used parameters and variables for a vehicle to cyclist interaction (lateral)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Table 5: Evaluation metrics a the potential collision with a cyclist

Metrics	Unit	Equation
Time to Collison (TTC)	s	$TTC = \frac{ds}{dv} = \frac{s_{Veh\ to\ Bike}}{v_{Veh} - v_{Bike}}$
Time Head Way (THW)	s	$THW = \frac{s_{Veh\ to\ Bike}}{v_{Veh}}$
Post Encroachment Time (PET)	s	$PET = time_{Veh} - time_{Bike}$
Deceleration rate to avoid collision (DRAC)	$\frac{m}{s^2}$	$DRAC = \frac{(v_{Veh} - v_{Bike})^2}{2 * s_{Veh\ to\ Bike}}$
Proportion of stopping distance (PSD)		$PSD = 2 * MADR * \frac{s_{Veh\ to\ Bike}}{v_{Veh}^2}$
Lateral Time to Collison (LTTC)	s	$LTTC = \frac{dt}{dv_t} = \frac{t_{Veh\ to\ Bike}}{v_{t,Veh} - v_{t,Bike}}$
Lateral Time Head Way (LTHW)	s	$LTHW = \frac{t_{Veh\ to\ Bike}}{v_{t,Veh}}$
Lateral Distance (LD)	m	$LD = t_{Veh\ to\ Bike}$

2.4.3 Preliminary plans for T2.5

Based on the previously described metrics, the evaluation and analysis of the scenarios is performed in T2.5 in post-processing. For this purpose, the following steps will follow:

- Importing and converting the log files of the simulation
- Identification of the conflict partners and calculation of the critical values of the evaluation metrics
- Interpretation and visualization of the critical conflict scenariosp08

Due to the large amount of simulation data, which is generated by a large road network and a large number of road users, the evaluation of the data should be fully automated and runtime optimized. For this purpose, a tool will be developed that automatically analyses the corresponding file format from the simulation. All interactions between vehicles and VRUs will be searched in the data and analysed according to their most critical interaction. The critical interactions are evaluated based on the selected metrics from T2.2 and matched with corresponding thresholds identified in T2.5. Based on this analysis, the interpretation between the critical metrics and the model behaviour will follow.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

2.5 Metrics for PTW Interactions (UNI)

2.5.1 Motivation

The literature shows only a few models of PTW driving behaviour for specific manoeuvres, and that beyond the thresholds already used for car-to-car interactions metrics are mostly unavailable to determine the severity risk of different interactions between PTWs and other road users (e.g. PTW-to-car).

Our severity metric is necessary because, to detect safety-critical scenarios of PTW-to-car interactions, the metrics must consider the actual behaviour of PTW riders and the specific characteristics of their interactions with cars. The main features that make PTW behaviour different is their narrow width and their higher capacity to swerve what makes that the conventional metrics such as TTC and PET have to be analysed in a different perspective. Figure 18 shows an example of the different behaviour of PTW rider compared to car driver comparing the trajectories of Cars and PTWs on a sector of an urban road from Athens extracted from pNeuma dataset. Figure 18 shows that whereas car drivers have a clear lane-discipline behaviour, PTW riders circulate in a more unpredictable way without following the lane-discipline.

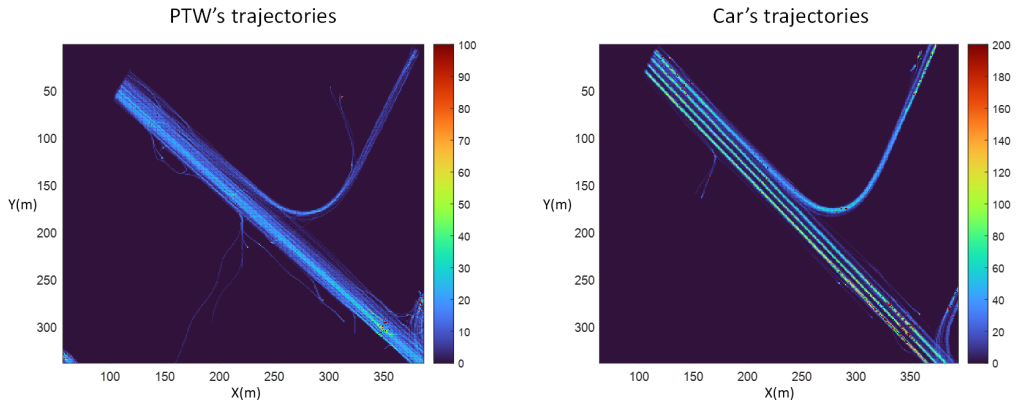

Figure 18. Left: Trajectory of 190 PTWs; Right: Trajectory of 303 Cars (Color represents the frequency of the trajectories)

To develop better simulation models capable of measuring the risk of traffic conflicts in a comprehensive perspective, a better understanding of the unique patterns of interaction of PTWs with other road users is needed. The SMoS for PTWs has been developed using naturalistic driving data and will estimate the severity of traffic interactions for scenarios involving vehicles moving in the same direction as PTWs. The conflicts selected in our analysis will be those in which the PTW has to brake as an evasive response to a rear-end collision.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

2.5.2 Technical Description

Two different approaches have been used to derive the SMoS that required different approaches. The first one classifies the severity of traffic conflicts by making the best use of the 'maximal' data observed with kinematics-based measures such as brake deceleration or yaw rate and the driver's use of controls with the PTW via brake, throttle and steer. This first classification uses in-vehicle naturalistic data collected by a PTW (local view) from Florence roads collected during 2BESAFE project (Rider Behavioral Patterns in Braking Manoeuvres, 2016). The clusterisation of the events has been developed using an unsupervised classification algorithm based on K-means and two-step methods. The second approach uses bird's-eye view naturalistic data collected from Athens urban roads in pNeuma project (On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment, 2020) using the spacing with traffic mix, speeds and type of road user as factors of the real-time safety metrics.

Unsupervised classification method. Florence data.

Data were collected as part of the EU funded research project 2BESAFE (www.2besafe.eu). The study collected naturalistic riding data from five riders for a period ranging from 1 up to 2 months to investigate near-missed events and other dangerous riding conditions (Rider Behavioral Patterns in Braking Manoeuvres, 2016). The set of signals registered was comprised of: linear acceleration (three components); roll, yaw and pitch angles and velocities; longitudinal speed; brake activation; brake pressures (front and rear circuits); throttle position; steering angle; GPS position; turn signals; video: 2 cameras positioned to capture the frontal environment (required a minimum 90° field of view) and the rider's head. Data collection was performed with a scooter (Piaggio Beverly Tourer 300ie). This choice reflected the fact that in the Florence vehicle area most of PTWs are scooters used for daily commuting. The main limitation of these naturalistic data is that information of the environment is limited to what video-cameras are able to register. This means that in most cases some actors involved are not captured as well as the distances to other users, as the recorded videos mainly function to validate traffic conflicts or to code the events with expert analysts in a non-automatic and time-consuming task. Filtered data were processed to identify all braking events recorded during normal and safety relevant riding conditions. In total 3573 braking events, characterized by 36 parameters, were identified (Table 6).

After an exploratory analysis of the data, the selected braking events to identify safety critical scenarios were those where the initial speed was higher than 20 km/h and with a maximum deceleration higher than 2 m/s^2. Looking at the distribution of the *velocity* (Figure 19), we assumed that the peak on 20 km/h was related to low speed and congested traffic and that there is low risk of conflict under those circumstances. Similarly, for the *maximum brake deceleration*, a clear difference was found among braking events with deceleration above and under $2m/s^2$. Thus, we assumed based on the study with cars of Hydén (Traffic conflicts technique: state-of-the-art, 1996), that braking events with less than $2m/s^2$ were related to events without conflict or with a small reaction necessary to avoid the conflict.

All these events, corresponding to 66% of all the 3573 braking events collected, were directly considered as 'normal' events. The second subset of braking events, which is the one above the determined thresholds and corresponds to 34% of the cases, was analysed for cluster

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

analysis using the full set of dynamic variables in order to identify the "less normal" behaviour, being associated with possible safety critical events.

Table 6. Description of the 36 parameters used to characterize the braking event.

Parameter	No. of var.	Description
Type of braking event	1	Identifies the type of braking action: only front or rear braking; combined braking.
Relative position of braking action	1	Applies to combined braking events. Sequence of activation of the front and rear circuits discriminating between the following different actions: a) action starts with rear or front braking only and ends respectively with front or rear braking only (intersected braking events); b) action starts and ends with rear braking only (front in rear braking); c) action starts and ends with front braking only (rear in front braking).
Duration	3	Overall braking event duration, and duration for each of the front and rear circuits.
Braking pattern	2	Shape of the braking action (respectively for the front and rear circuit): single, double or multiple peaks.
Start shift	1	Time shift of the pressure onset between the front and rear circuit (positive in case the braking action starts with the front circuit).
Maximum pressure	3	Maximum pressure value in each circuit and their sum.
Position of maximum pressure	2	Position in time of the maximum pressure in each circuit. The position is expressed as a percentage of the overall duration of the event (100% corresponds to the end of the event).
Shift of the maxima	1	Time shift between the maximum pressure of the front and rear circuit.
Velocity	3	Longitudinal velocity at the start and at the end of the braking event, and their difference.
Deceleration	2	Maximum deceleration and mean deceleration within the event.
Position of max. deceleration	1	Position in time of the maximum deceleration, expressed as a percentage of the overall duration of the event (100% corresponds to the end of the event).
Pressures at max. deceleration	2	Pressures in the front and rear circuits corresponding to the maximum deceleration.
Yaw rate	4	Yaw rate at the beginning of the braking event and at maximum deceleration, and their absolute values.
Roll rate	4	Roll rate at the beginning of the braking event and at maximum deceleration, and their absolute values.
Roll angle	3	Roll angle at the beginning of the braking event, at maximum deceleration and their difference.
Steering at max. deceleration	1	Steering position corresponding to the maximum deceleration.
Throttle	2	Throttle position at the start and at the end of the braking event.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

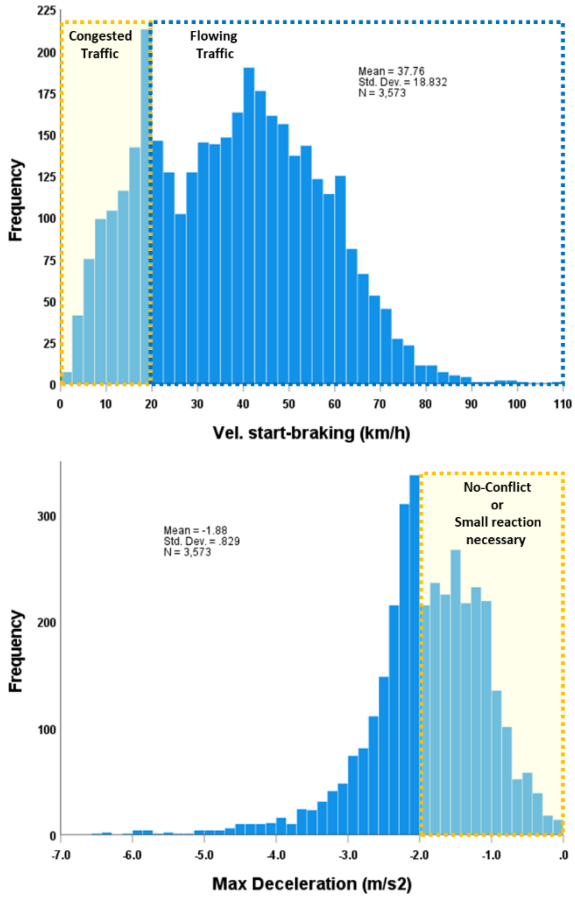

Figure 19. Distribution of velocity at start-braking and maximum deceleration of the 3573 braking events

Following a principal component analysis (PCA) of the different parameters characterizing the braking events, four features were selected for the unsupervised classification algorithm: *maximum deceleration*, *maximum yaw rate*, *throttle position* when start braking and *brake duration*. Two clustering algorithms were used to test the stability of the process: two-step and K-means. The final cluster analysis was performed with the two-step algorithm developed by Chiu et al. (A robust and scalable clustering algorithm for mixed type attributes in large database, 2001). The range in the number of clusters was selected based on the Schwarz Bayesian Criterion (BIC) and the change in BIC between adjacent number of clusters. After each processing, clusters were screened in order to verify if they were meaningful for behaviour identification.

After a two-step cluster analysis of the data using the four features from the PCA, three different clusters related to three different behaviours during braking events have been identified (Table 7). Cluster 3, referred to as the Potential Risk cluster, which accounted for 6.2% of the data set, represented the events with the highest *maximum deceleration* and *yaw rate* and was associated with events that were potentially safety critical. Cluster 1 and 2 are associated to '*normal*' events (safe) with short and long duration respectively and are labelled as *Normal-Short and Normal-Long*. The scatter plots of Figure 20 show clearly how *maximum deceleration* and *yaw rate* are determinant in identifying the Potential Risk group (cluster 3) and how braking duration differentiates the clusters 1 and 2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Table 7. Results of the cluster analysis

Cluster	Variable	Mean	(s.d.)
1: *Normal-Short* **(n=822)** **(23.0%)**	brake duration (s)	3.83	1.63
	max deceleration (m/s^2)	-2.42	0.33
	max yaw rate (rad/s)	0.04	0.04
	throttle (ini. position) (deg.)	2.06	1.21
2: *Normal-Long* **(n=172)** **(4.8%)**	brake duration (s)	8.82	4.82
	max deceleration (m/s^2)	-2.51	0.50
	max yaw rate (rad/s)	0.05	0.04
	throttle (ini. position) (deg.)	5.43	4.79
3: *Potential Risk* **(n=223)** **(6.2%)**	brake duration (s)	4.28	2.37
	max deceleration (m/s^2)	-3.52	1.02
	max yaw rate (rad/s)	0.16	0.13
	throttle (ini. position) (deg.)	2.50	1.79
Low deceleration or Low speed Brake event **(n=2356)** **(65.9%)**	max deceleration: <2m/s^2 or Velocity start-braking: <20km/h		

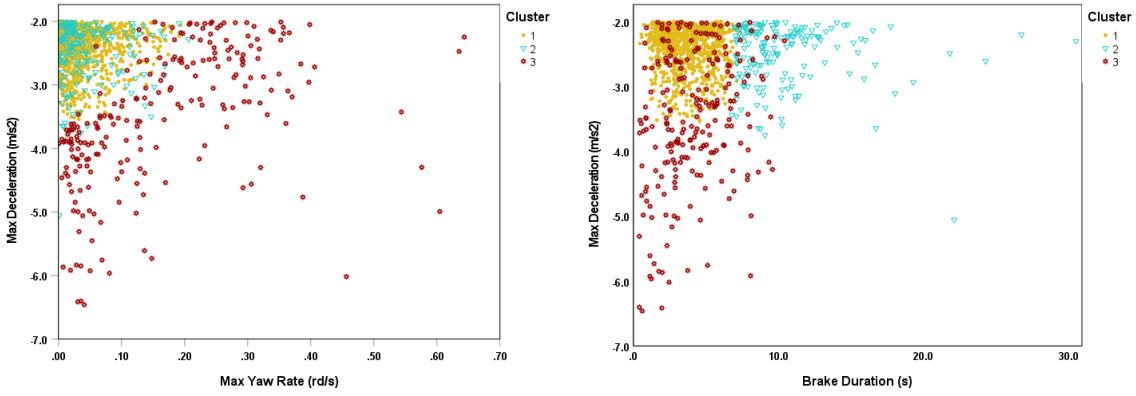

Figure 20. Scatter plot of the features maximum deceleration vs. maximum yaw rate (left) and brake duration with the three clusters identified. Cluster '3' represents the braking events with higher risk.

Finally, a discriminant analysis was conducted to obtain the predictive equations able to categorize the different clusters (specially cluster 3 – *Potential Risk*) using as independent variables the four features previously defined. These equations are used to categorise the dependent variable *Cluster* and to identify cluster 3 associated to a potential safety risk scenario. Figure 21 shows the coefficients of the discriminant functions. Function 1 (representing 57.6% of the variance) is related to *max deceleration* and *max yaw rate* and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

represents the *Risk Score,* thus confirming the assumption made earlier. Function 2 is related to *brake duration* and *throttle position* and is represented as *Duration Score*. Figure 21, with a graph of individual cases on the discriminant dimension, shows how function-1 defines mostly cluster 3 and function-2 mostly differentiates cluster 1 from cluster 3.

Feature	Function	
	1 (57.6%)	2 (42.4%)
brake duration	.137	**.917**
max deceleration	**.895**	-.180
max yaw rate	**-.858**	.101
throttle (ini. position)	.201	**.781**

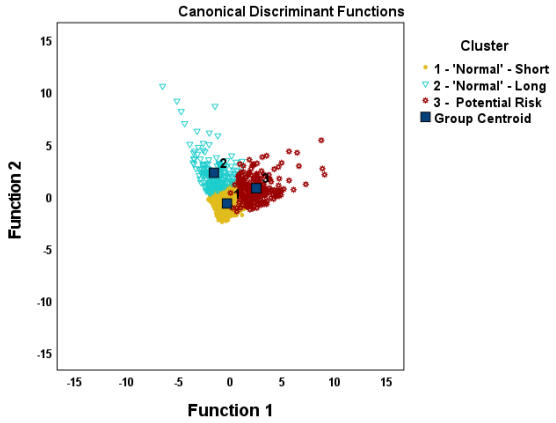

Figure 21. Left: Standardized Canonical Discriminant Function Coefficients; Right: Cluster distributions and Centroid of Canonical Discriminant Functions.

The classification results (Table 8) reveal that 95% of braking events were classified correctly into *Normal*-Short, *Normal*-Long or Potential Risk clusters. Braking events with *Potential risk* were classified with an accuracy of 92.4%. The cross validated classification showed that overall 94.9% were correctly classified

Table 8. Classification results table

		Risk or not	Predicted Group Membership			Total
			Normal-Short	*Normal*-Long	Potential Risk	
Original Group Membership [a]	Count	1 – *Normal*-Short	788	2	32	822
		2 - *Normal*-Long	6	162	4	172
		3 – Potential Risk	16	1	206	223
	%	1 – *Normal*-Short	95.9	.2	3.9	100.0
		2 - *Normal*-Long	3.5	94.2	2.3	100.0
		3 – Potential Risk	7.2	.4	92.4	100.0
Cross-validated[bc]	Count	1 – *Normal*-Short	788	2	32	822
		2 - *Normal*-Long	7	161	4	172
		3 – Potential Risk	16	1	206	223
	%	1 – *Normal*-Short	95.9	.2	3.9	100.0
		2 - *Normal*-Long	4.1	93.6	2.3	100.0
		3 – Potential Risk	7.2	.4	92.4	100.0

a. 95.0% of original grouped cases correctly classified.
b. In cross validation, each case is classified by the functions derived from all cases other than that case.
c. 94.9% of cross-validated grouped cases correctly classified.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Motorcycle Crash Potential Index (MCPI). Athens data.

The metric developed for PTW-to-car interactions, labelled as Motorcycle Collision Potential Index (MCPI), is based on the Crash Potential Index (CPI) defined in (Calibration and validation of simulated vehicle safety performance at signalized intersections., 2008) for vehicle-to-vehicle conflicts where vehicles were mostly cars. The metric considers multiple factors such as TTC, time headway and DRAC and stablishes the risk index based on a distribution of MADR for different PTW rider performances. One of the reasons for the difficulty of implementing the CPI as a risk metric is the need for on-the-field measurements with comprehensive and detailed vehicle tracking data.

pNeuma Data

The naturalistic data used corresponds to the pNeuma dataset (On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment, 2020), which represents an urban environment (city of Athens) that includes a significant amount of PTW users to develop our metrics. The dataset contains detailed trajectories of tracked vehicles from the weekday morning rush hour (between 8:00 and 10:30), with a time frequency of 0.04 s (25 frames per second). The main limitation of these naturalistic data is that vehicles' lengths and widths are not provided, so considering that the trajectories are related to the centroid of the vehicle the distances front-bottom vehicles have to be estimated based on the average dimensions (Table 9).

Table 9. Dimensions assigned to the vehicles for the MCPI analysis

Vehicle	Length (m)	Width (m)
Car	5	2
PTW	2.0	1
Bicycle	1.8	0.8
BUS	12.5	4
Medium Vehicle	5.83	2.67
Heavy Vehicle	12.5	3.33

To develop the MCPI safety metric, it was decided to use two sectors with different characteristics that may represent different PTW rider behaviour with different types of conflicts. By doing so, the variability of PTW rider behaviour is also better captured and a more robust metric can be defined. Two one-way urban road sections of 500-550 meters length with different characteristics have been selected: i) a one-lane section; ii) a 4-lanes section (Figure 22). The two sections correspond to different travel speed profiles with PTW as Figure 23 shows. These two sectors were selected because they were similar to those of the network *Town 7* used in AIMSUN for T2.5. On the one hand, the 4-lane section corresponds to an urban road with an average speed of 50 km/h as that of *Town 7* network for microsimulation without traffic congestion. On the other hand, on the 1-lane section, PTW speed is more similar to that of a congested traffic scenario and, in addition, PTWs are less likely to overtake because the road is narrower and, consequently, they maintain the target condition of our study, i. e. the car-following condition, for a longer time.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

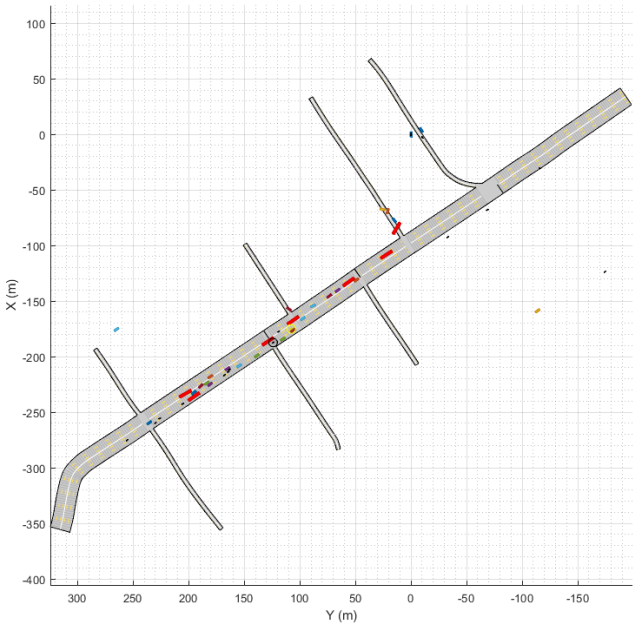

Figure 22. Representation of the simulation and trajectories of the 4-Lanes urban road section

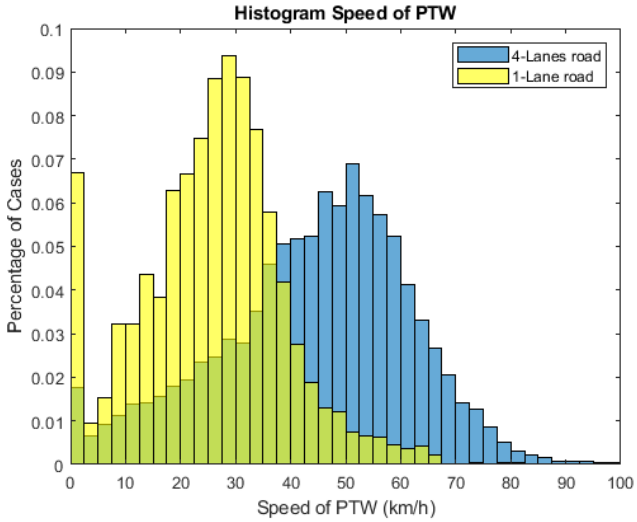

Figure 23. Distribution of travel speed of PTWs: 4-Lanes urban road vs. 1-Lane urban road

A subset of the naturalistic data with the characteristic detailed in Table 10 has been used to develop and assess the performance of the MCPI risk metrics.

Table 10. Characteristics of the data used for the MCPI development

Sector	Traffic-volume	Average Speed for PTWs	Number of days	Time	car-following (PTW) cases	Length
1-Lane	Low	50 km/h	5	360min	1200	540m
4-Lanes	High	30 km/h	4	270min	98	550m

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

MCPI

MCPI is defined as the probability that a PTW vehicle DRAC exceeds its maximum available deceleration rate MADR$^{(PTW)}$ during a given time interval. The MCPI index is obtained using an equation of the form

$$MCPI_i = \frac{\sum_{t=ti_i}^{tf_i} P\left(MADR^{(PTW)} \leq DRAC_{i,t}\right) \cdot \Delta t}{T_i}$$

Where:

$$MCPI_i = crash\ potential\ index\ for\ PTW\ i$$
$$DRAC_{i,t} = deceleration\ rate\ required\ to\ avoid\ the\ crash$$
$$MADR^{(PTW)} = truncated\ distribution\ of\ PWT\ rider\ performance\ during\ emergency\ braking$$
$$ti_i = initial\ simulated\ time\ interval\ for\ vehicle\ i$$
$$tf_i = final\ simulated\ time\ interval\ for\ vehicle\ i$$
$$\Delta t = simulation\ time\ interval$$
$$T_i = total\ event\ time\ for\ vehicle\ i$$

The simulation algorithm updates the DRAC at each time interval (0.1 seconds) according to the response of the drivers in the preceding time interval. As an example, if at a certain time interval (ti_k) the driver of the PTW applies a deceleration rate greater than the DRAC, then at the next time interval (ti_{k+1}) the DRAC becomes lower, assuming that the speed of the leading vehicle is constant. Similarly, if the leading vehicle starts braking or increases deceleration, the DRAC will be reduced for a constant braking deceleration of the PTW that follows the vehicle.

To consider the braking performance of PTW riders with different abilities, the MADR is included as a stochastic component where PTW riders are expected to act differently during a braking event with a specific DRAC. The analysis in this task considers PTWs operating under good conditions, i.e. dry road and daylight. For these set of conditions, we used the distribution of emergency braking in field experiments with initial speeds from 40 to 60 km/h coming to a full stop that were performed by 13 riders of different rider skills (Emergency braking performance of motorcycle riders: skill identification in a real-life perception-action task designed for training purposes, 2019) obtained from previous project MOTORIST of the FP7 of the European Union. Figure 24 represents the kernel distribution of the deceleration achieved in the 149 emergency events, the adjusted normal distribution, and the example of the probability of collision for a PTW in a conflict with a DRAC of 7m/s^2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

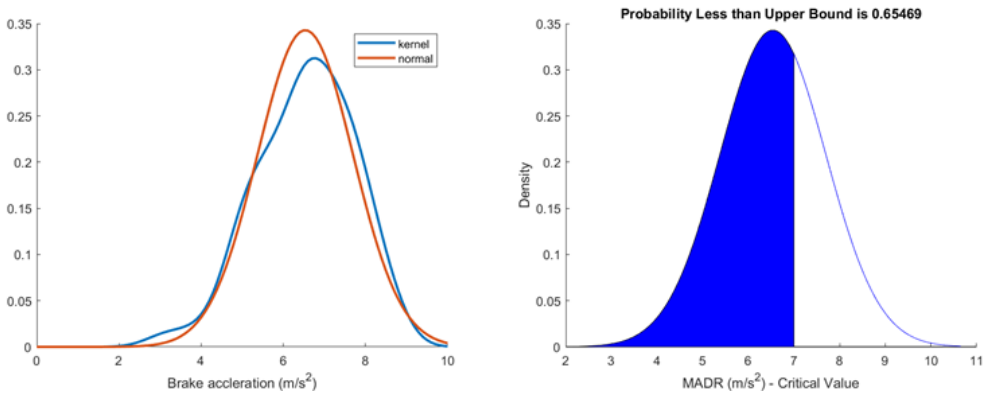

Figure 24. (left) Distribution of decelerations for emergency braking tests (n=149 test); (right) probability of crash of 65.4% for a DRAC of 7m/s2

Identifying car-following events

The algorithm developed examined all braking events occurring on the two sections of urban road selected for this work. Potential lead vehicles were then identified during each brake application following an algorithm for car-following events (Methodology for identifying car following events from naturalistic data, 2014) summarized below.

First, to be considered car-following behaviour by the PTW rider, the object ahead had to have a time headway of less than 3 s (Research Board, 2010). Then, those objects that were within 2m of the vehicle trajectory during 65% of the brake application were identified as potential lead vehicles. For multiple potential lead vehicles, the object with a distance from the vehicle trajectory of less than 1.5 m and/or the object closest to the PTW vehicle was chosen as the lead vehicle. Once the lead vehicle was identified, the safety distances and the distributions of TTC, PET and DRAC during the 3 seconds before and 3s after brake application were identified.

Due to the uncertainty associated with the assumed path width of the vehicle in front, for the MCPI calculation we have classified the overlap as *Some Overlap* (percentage of overlap less than 100%) and *Full* Overlap (see Figure 25) based on the approach of previous studies with cars (Chen, 2016). In the case of *Full Overlap* the MCPI consider the crash probability as a function of the DRAC and $MADR^{(PTW)}$ as defined above. While for cases with *Some Overlap,* given the small width of PTWs and their high manoeuvrability to swerve, the MCPI applies a correction factor proportional to the overlap degree. Following the same process, considering the frequent non-lane discipline behaviour observed on the PTW riders, we had to stablish a range to the PTW path to stablish when a leading object was a potential object to collide (and thus with its corresponding TTC and MCPI value). When the PTW is out of the path of the lead vehicle (set to 1.5m distance between PTW and car centroids), then, the car-following condition is considered ended, the TTC is not evaluated and the instantaneous MCPI is reduced to no risk values.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

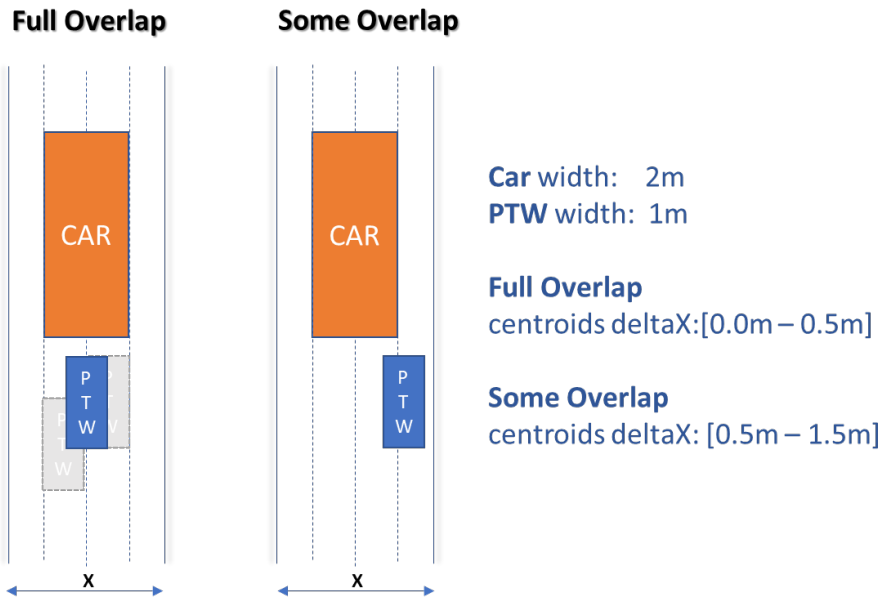

Figure 25. Example of Full and Some Overlap of path for PTW following a Car.

Results

Figure 26 shows a histogram of the results of the MCPI analysis for 1-lane and 4-lanes sections. The results of the MCPI of both sections are comparable despite the different characteristics of the traffic. Contrary to what might be expected, the 1-line section involved a higher rate of dangerous interactions (almost 20% of the car-following events with braking) than the 4-line section (approx. 8%). The results suggests that the PTWs in a narrower road take more risk keeping safety distances and overtaking, despite riding with lower speeds.

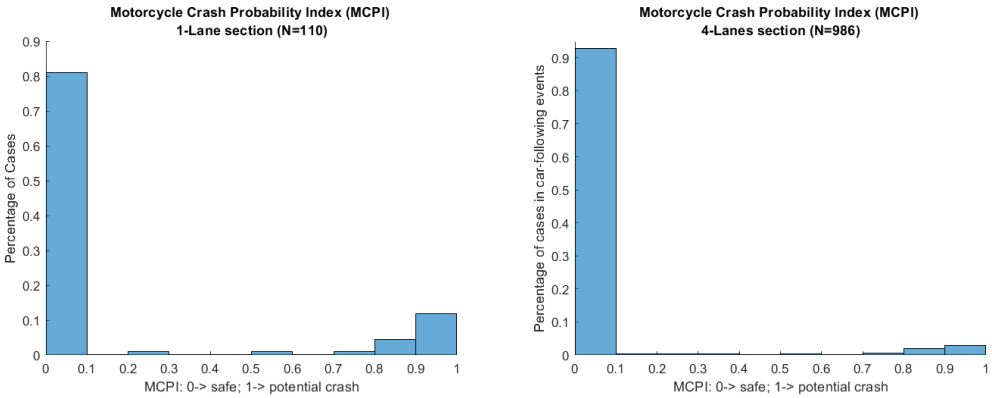

Figure 26. MCPI distribution for braking events in sector 4-Lanes

As commented previously, vehicle dimensions were not fully measured in the pNeuma dataset but were estimated using the average size of each vehicle. Considering that PTWs do not usually ride in the centre of the lane, their fast-swerving capability and the short safety distances taken by riskier riders, the MCPI is sensitive to cases where the actual size of the lead car differs from the estimated size. In particular, in cases where the actual length of the lead car is less than the *a priori* estimated 5 metres in length or 2 metres in width, if the PTW rider tailgates the lead vehicle at a distance less than difference between the actual and the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

estimated length, the MCPI metric will consider a crash. This is the reason why the MCPI values get values of 1 in some cases, even though no collision has been identified. Figure 27, where time 0 is the instant of maximum braking, shows an example of such a case, in which the PTW reaches a very short headway distance before overtaking the lead car at instant -0.5s. In the figure, Y and X distance represents the longitudinal and lateral distance respectively between the PTW and the identified Most Important Object (MIO), distance time=0s corresponds to the highest brake deceleration instant. Therefore, the MCPI in these cases will not identify an actual crash, but a high-risk manoeuvre that could lead to a crash, regardless of the fact that specific events such as the one in this example end without a collision.

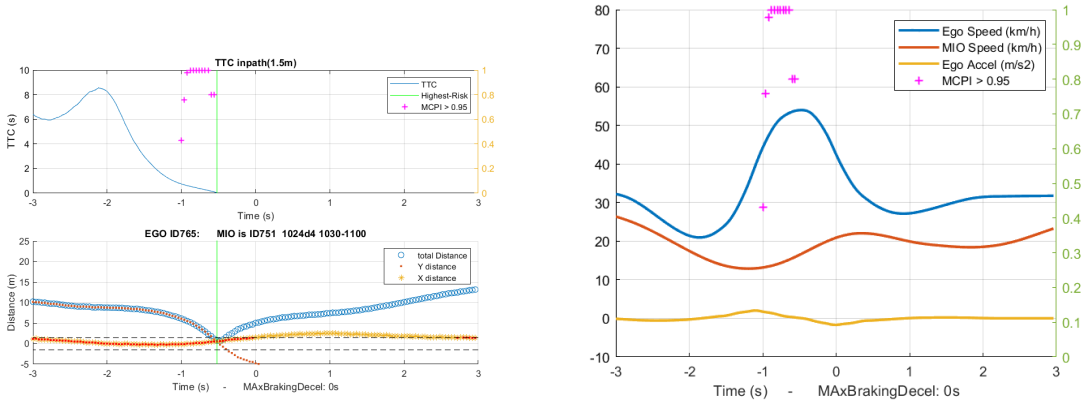

Figure 27. Example of overtaking with overlapping path

Figure 28 and Figure 29 show the MCPI distribution of each dataset grouped into 30-minute sets. The results show that there are not significant changes in the MCPI to be noted between the different datasets.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

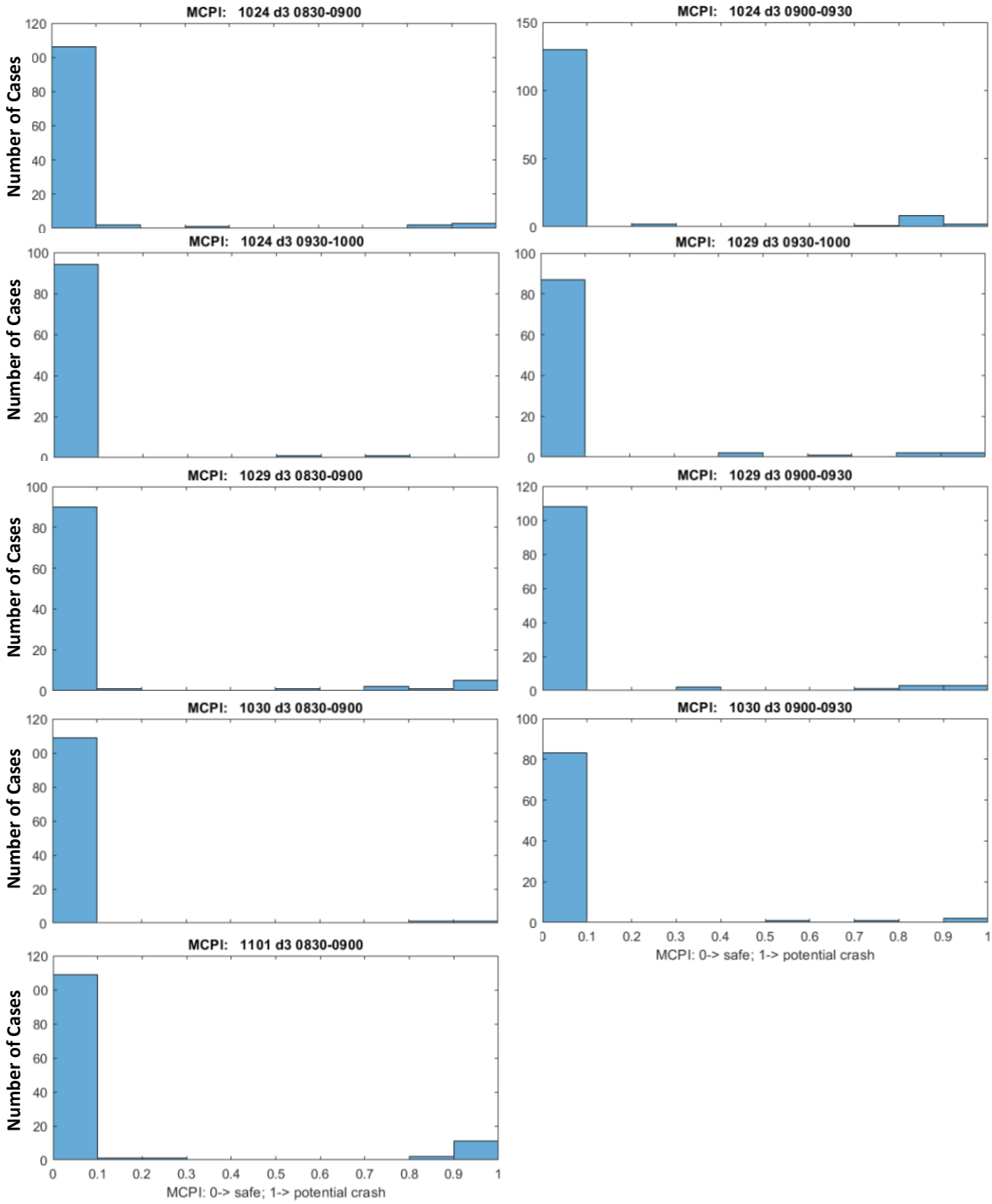

Figure 28. 4-Lanes section: MCPI by set of 30min data analysed

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

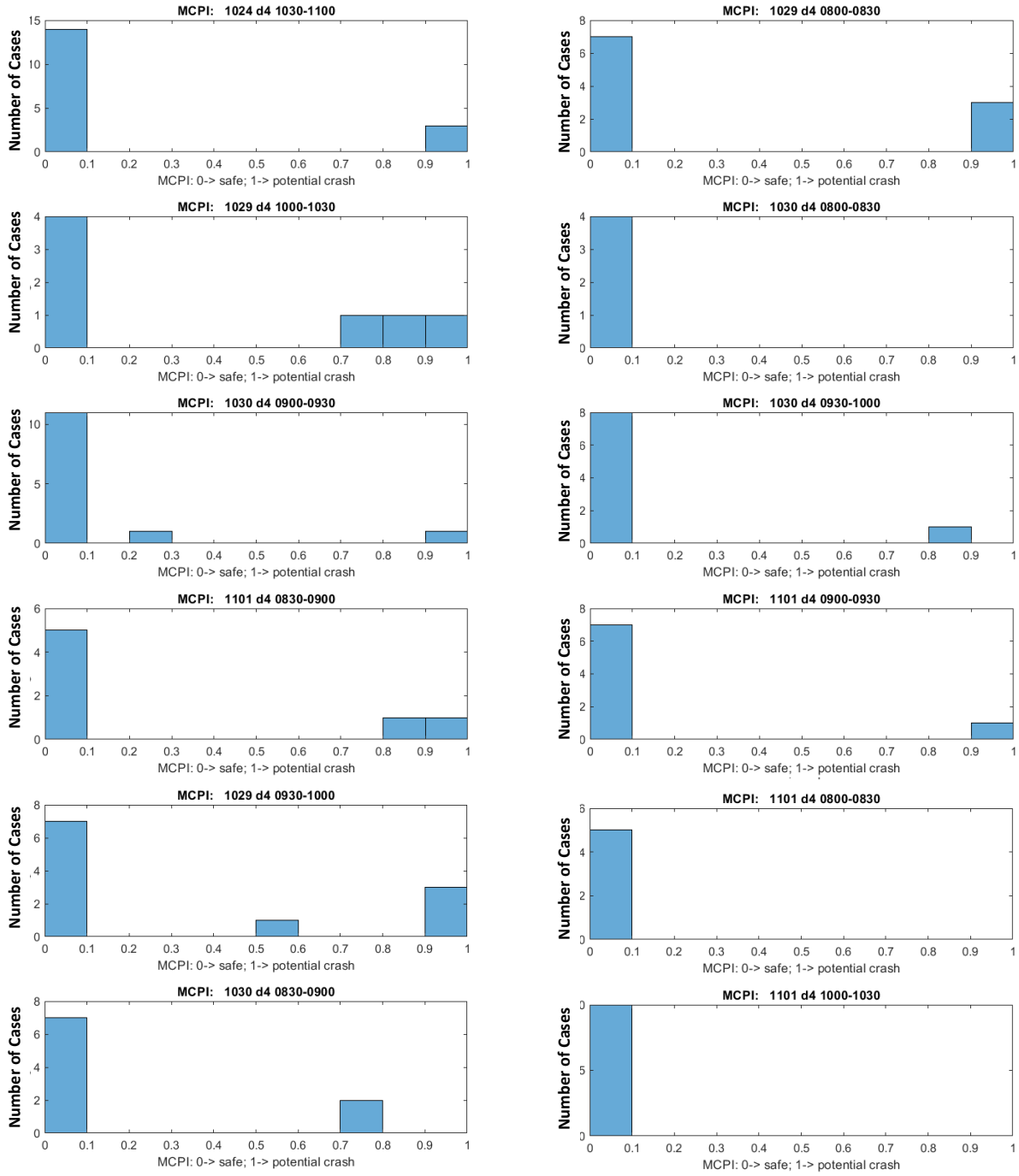

Figure 29. 1-Lane section: MCPI by set of 30min data analysed

Validation

The safety critical scenarios identified can be reviewed through simulation can be reviewed by performing a simulation with driving scenarios created with Matlab® R2022a, where the trajectories extracted from the naturalistic driving data over time and the road are programmed. In the absence of data on actual vehicle dimensions, the dimensions of vehicles such as PTWs, cars and buses are assumed to be cubic in shape and with a default value for length, width and height corresponding to that of a typical vehicle in its class. Thus,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

validation is carried out by using the information of the trajectories of the road users involved and running a simulation of the critical events encountered. The simulation of the events allows a qualitative assessment of the risk of collision based on the behaviour of the actors involved in the conflict. Figure 30 and Figure 32 show two examples of safety critical scenario identified.

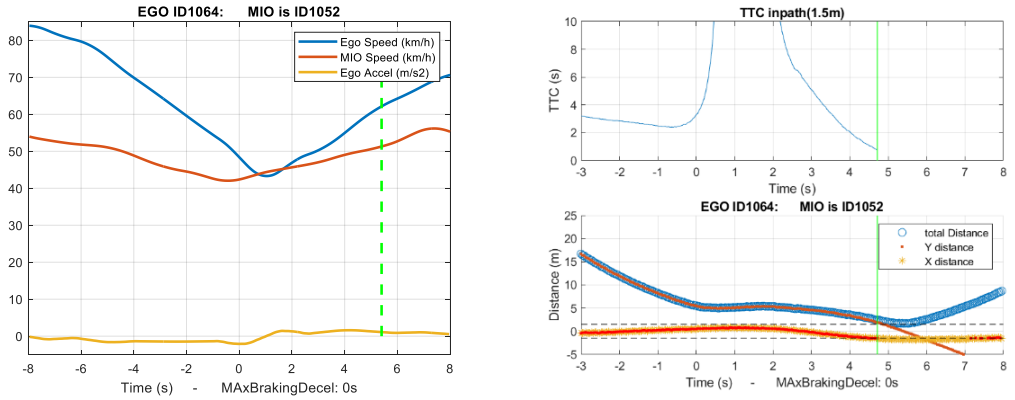

Figure 30. Car-following + overtaking with short safety distance. Ego vehicle (PTW) and MIO (car)

Error! Reference source not found. shows a case of overtaking with short safety distance. The graph from the left shows how the PTW (ego) reduces the speed from 83km/h to less than 50km/h during the car-following and then accelerates to overtake the car (MIO). As can be seen in Figure 30 (right) representing the distances, the overtaking is done between seconds 5 and 6. In this case, the overtaking has been done (total distance moving to negative values) with a lateral distance between centroids of 1.7m, that for a width car of 2m and a width PTW of 1m would represent a minimum lateral distance (0.2m). Figure 31 shows a sequence of snapshots of the car-following plus overtaking commented above.

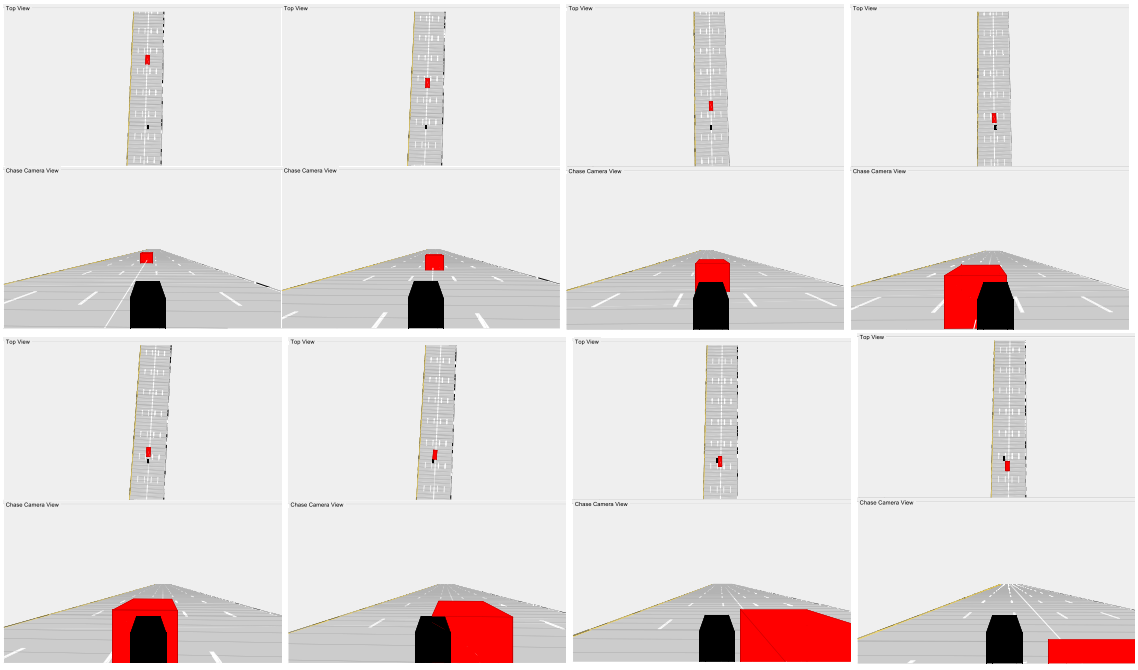

Figure 31. Simulation of real case of car-following + overtaking with short safety distance. Ego vehicle (PTW) in black and MIO (car) in red.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

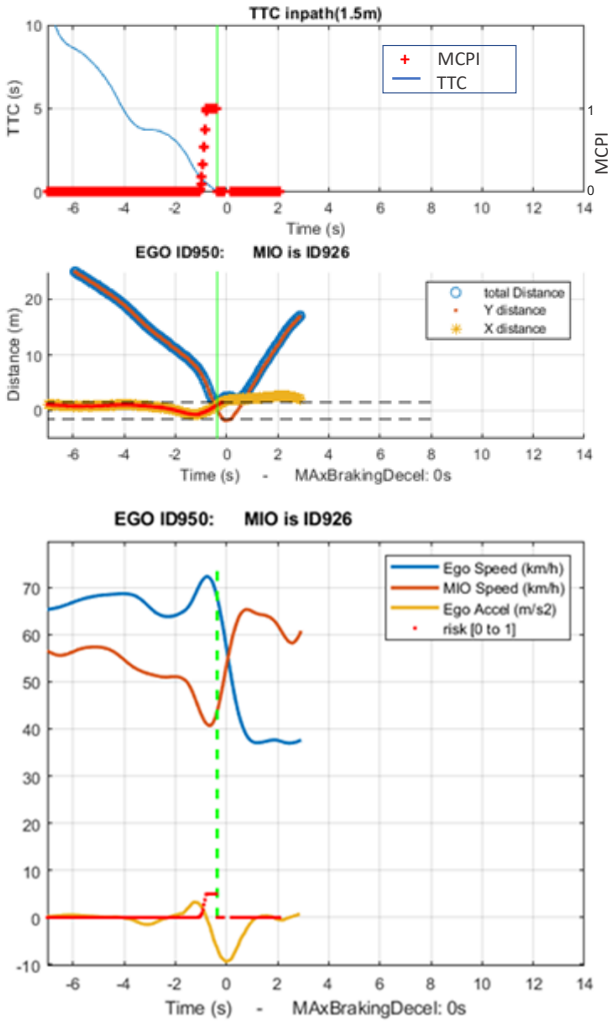

Figure 32. Example of High brake deceleration to avoid rear-end in sector B (4-lanes).

Finally, we use as example the case of hard brake to avoid rear-end collision presented in Figure 32 where at certain point the PTW (ego) must achieve a very high deceleration to avoid a rear-end collision. The simulation of the event confirmed that it corresponds to a near-miss traffic conflict where the ego PTW is running fast and overtaking other road users without lane-discipline until it has to brake achieving the minimum headway distance (Figure 33) after a peak of deceleration of almost 9m/s^2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

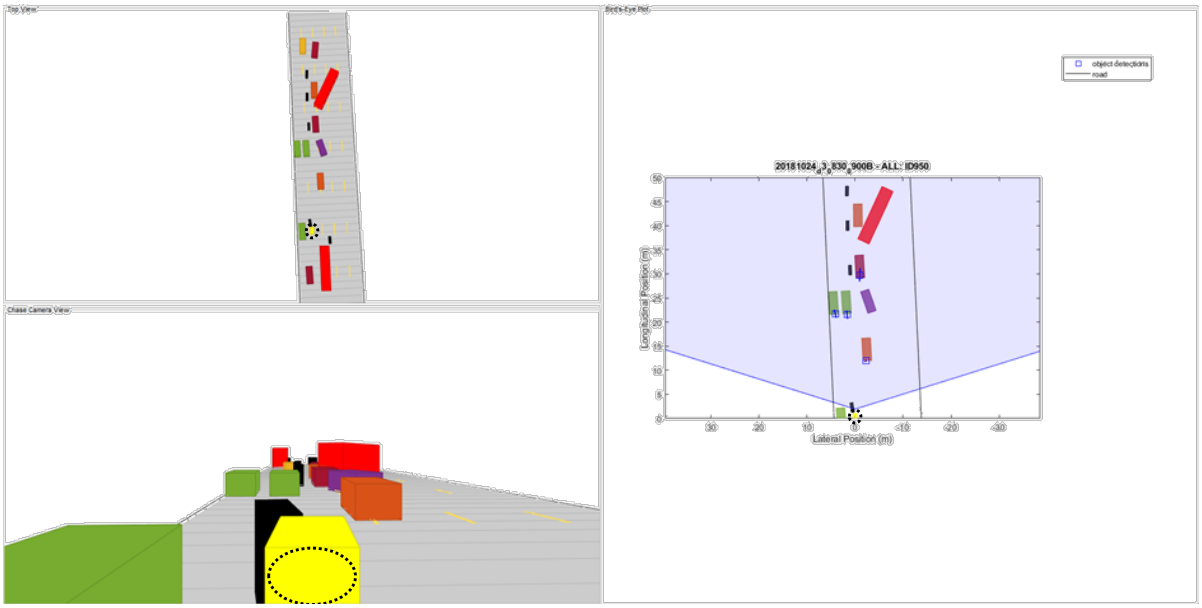

Figure 33. Simulation with three views of event with high CPI in 4-Lanes sector, PTW ego vehicle is represented in yellow.

2.5.3 Preliminary plans for T2.5

The MCPI metric will be used in the different micro-simulations in AIMSUN Next performed in T2.5, in order to assess the effect of the inclusion of autonomous vehicles compared to the reference scenarios without autonomous vehicles on: a) the number of PTW-to-car conflicts encountered; b) the typology of conflicts.. The data extracted from the road users during the microsimulations corresponds to the road users trajectories, speeds and road network. No application of the metrics based on cluster analysis can be done because the microscopic simulation lacks of full information of the dynamics (e.g. yaw rate) or actuators of the PTW (e.g. throttle). It is important to note that the vehicle models embedded in the AIMSUN simulation platform operate in a collision-free environment, so that AIMSUN forces unachievable decelerations (i.e. greater than 1g) to avoid collisions that would occur in the real world.

The comparation with and without autonomous vehicle can be done analysing the conflicts of different sectors of the road network (e.g. those from Figure 34. In order to find more traffic conflicts where the motorcycle had to brake hard, a delay of 2.5s in the response time of the PTW riders was forced in 25% of the events. As Figure 35 shows, the MCPI identified 2 out of 67 cases with car-following events as potential crash for the simulations with delay in the response time forced and 0 out of 42 cases for the simulation with no delays in the response time of the PTWs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

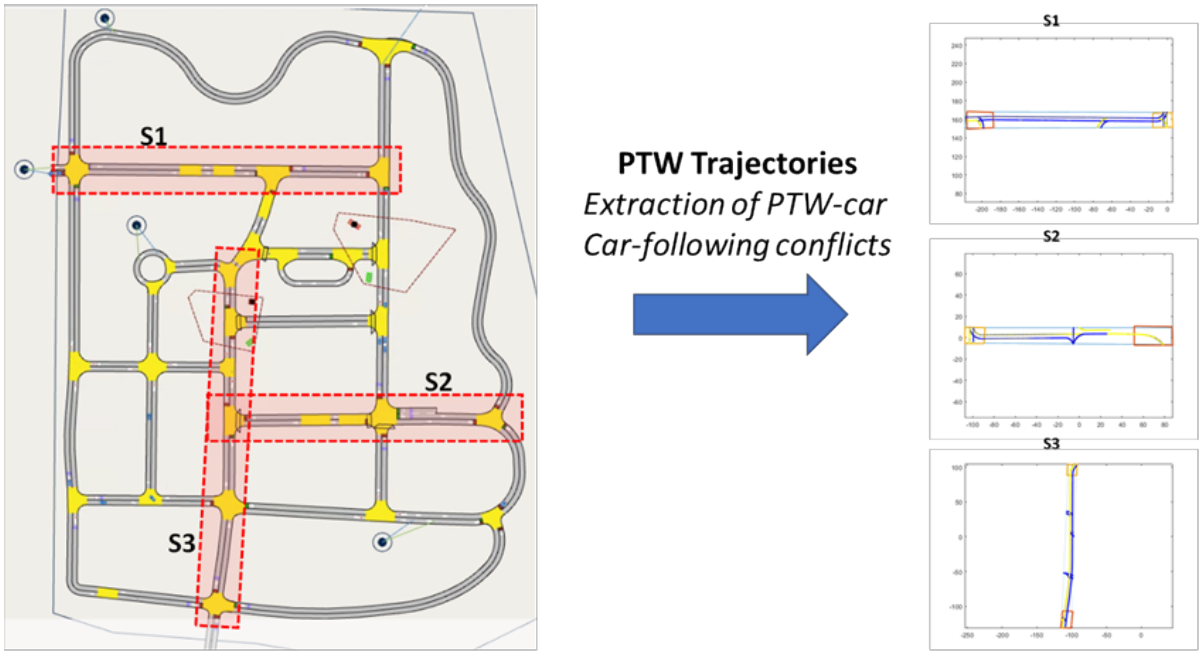

Figure 34. Identification of sectors S1, S2 and S3 of Town 7 that will be analysed during T2.5

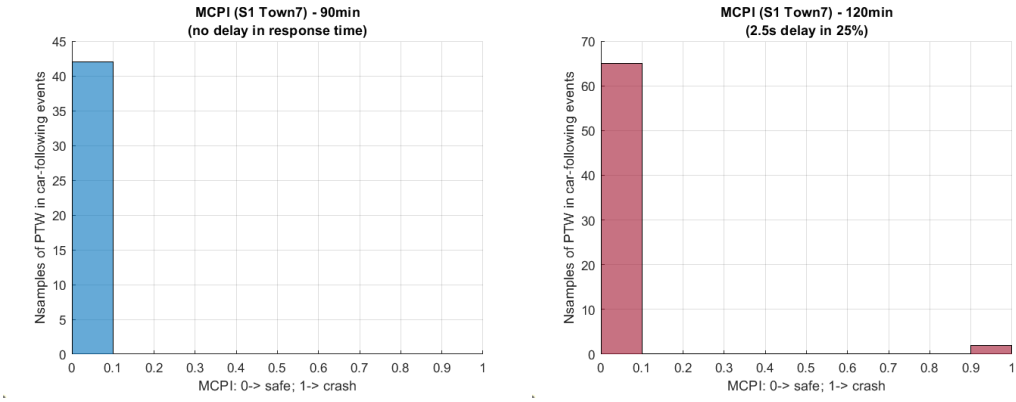

Figure 35. MCPI in section S1 for two different conditions.

For a second validation to check how MCPI works to identify conflicts where the car crosses the path of the PTWs, we generated a network scenario consisting of an X-intersection where the PTWs had to give the right of way (see Figure 36) and the PTW violated the right of way of the cars.

As expected, the frequency of cases with the MCPI identifying a crash is high.. In 68 out of 243 cases (28%) where the PTW crossed the intersection with right of way violation, the MCPI identify a crash conflict (Figure 37). Figure 38 shows one case of those identified where on the one hand the car (MIO) is driving at constant speed and crossing the intersection in front of the PTW and on the other hand the maximum deceleration of the PTW (ego) to avoid the collision in the last instant is higher than 1g (i.e. potential crash).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

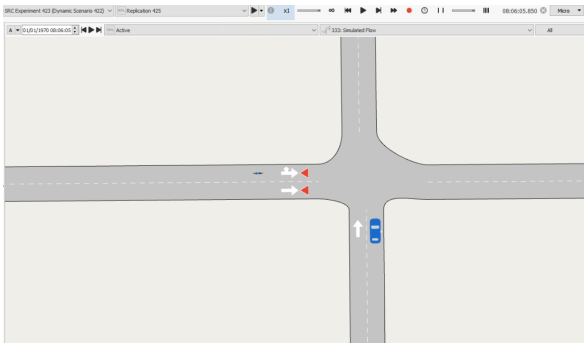

Figure 36. Intersection scenario for simulation with yield sign.

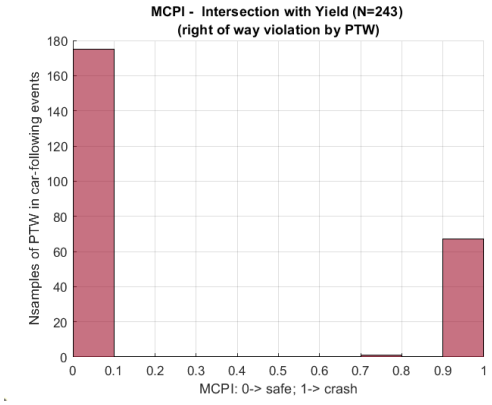

Figure 37. MCPI for scenario with right of way violation at intersection

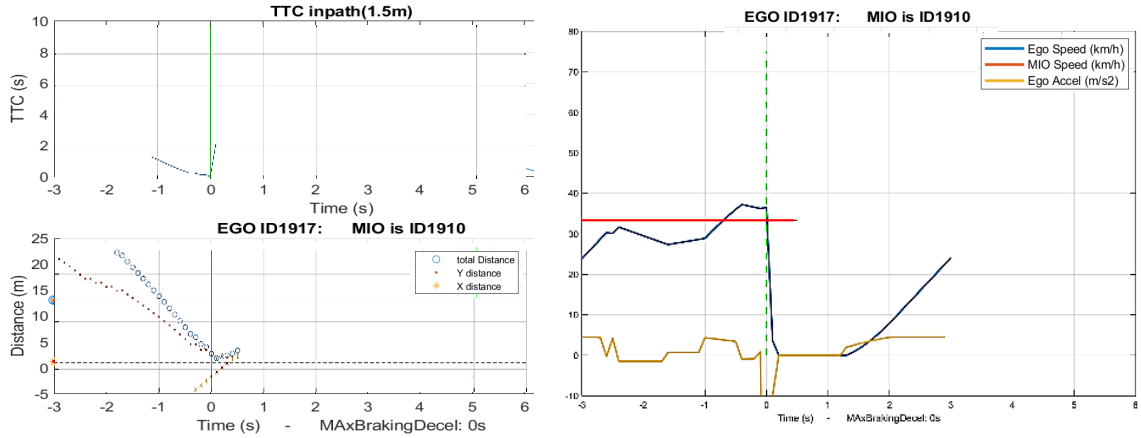

Figure 38. Example of MCPI = 1 where AIMSUN makes PTW (Ego) brake to more than 1g to avoid the crash with the crossing car(MIO) at intersection.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

3. Conclusions & Recommendations

The T2.2 partners have developed a number of metrics to understand the severity of traffic interactions. Their work was focussed separately in three types of interactions: Motor vehicles (TNO, TUD), VRUs (IKA), and PTWs (UNI).

Each partner provided in Section 2 a detailed account of the derivation of their metrics, on how they were calibrated and validated, and their main observations about the benefits of their metrics and their next step. Here we provide a summary of their conclusions.

TNO developed two types of metrics for motor vehicle interactions based on two separate techniques: unsupervised anomaly detection and driver profiling. The former is based on machine learning principles and was shown to work when applied to driving interactions in controlled environments such as rush-hour highway driving. The technique was not well suited for urban setting containing multimodal traffic. The second method, driver profiling, is able to identify clusters of drivers with different profiles (e.g., aggressive, typical) in urban traffic. The initial results show that the profiles are more easy to interpret than unsupervised anomaly detection from a safety perspective. For this reason, profiling will be used in T2.5 to perform a differential analysis between simulations without and with automated vehicles (AVs) to test the hypothesis of whether the introduction of AVs change the profile clusters (see Section 2.2.5).

TUD developed another metric for motor vehicle interactions based on artificial field theory. The metric, called probabilistic driving risk field (PDRF), was calibrated using NDD data and validated using simulations studies and NDD data. In the former, a critical driving scenario was set up (hard braking by a lead vehicle, requiring the vehicle under test to perform an emergency stop) and simulated multiple events by varying initial longitudinal velocities for the ego and lead vehicles. Only some of the events led to crashes. Both TTC and PDRF were used to predict (before the potential crash moment) whether the scenario would become a crash or not. The study showed that PDRF more accurately predicted the crash events than TTC. This was also confirmed using NDD data instead of simulated data (Probabilistic field approach for motorway driving risk assessment, 2020; Probabilistic Risk Metric for Highway Driving Leveraging Multi-Modal Trajectory Predictions, 2022). In T2.5, the PDRF distributions from simulations without and with AVs will be computed, compared and analysed to determine the effect of AVs on safety. See Section 2.3.3 for details.

IKA analysed existing surrogate metrics of safety developed to assess motor vehicle interactions to understand their suitability in characterizing VRU interactions. For pedestrians, IKA determined that deceleration to safety time and pedestrian-centric post encroachment time and time headway are the most relevant metrics. For cyclists, IKA proposed cyclist-centric variants of well-known longitudinal interaction metrics such as time to collision and time headway, as well a new versions of such metrics suitable for lateral interactions. These metrics and their associated criticality thresholds will be used to identify critical interactions with VRUs in the T2.5 simulations, and an analysis will be conducted to determine whether they are related to the presence of AVs (see Section 2.4.3).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

UNI developed two new metrics for PTW-to-car interactions based on two separate techniques: Unsupervised classification method and motorcycle crash potential index (MCPI). The first metric, using in-vehicle naturalistic data, identified three different behaviours during braking events, where one was referred as *Potential Risk* cluster. The MCPI metric characterizes longitudinal interactions for PTWs and stablishes the risk index based on a distribution of MADR for different PTW rider performances. The former metric was calibrated using NDD from Athens The ability of this metric to identify critical events by relating the rate of deceleration required to avoid a collision to the braking capabilities of the vehicle and the actual braking abilities of the rider was confirmed by simulation studies including two types of critical interaction involving hard braking of the motorcycle. In T2.5 These metrics will be used to determine whether the presence of AVs influence the number of critical interactions with PTWs (see Section 2.5.3).

Although the partners worked independently on their metrics, they found a number of common challenges:

- There was limited access to well-processed NDD data that allows for accurate measurement of vehicle sizes, and inter-vehicle distances. This made it difficult to estimate accurate certain metrics including collision risk.
- A standard toolchain to acquire and process infrastructure-based data (e.g., video of traffic) in order to extract vehicle characteristics (e.g., size) and trajectories is not available. As a consequence, several partners had to develop their own custom methods to process NDD.
- The results based on algorithms trained on NDD are difficult to generalize. Although driving behaviour may be uniform within countries, it may vary across countries. Hence, the calibration of some of the metrics may need to be updated in order to be applicable to situation other than the ones represented by the training data.
- There is no ground-truth for severity metrics. That is, severity (or safety, or risk) are not fundamental properties but conventions. That makes it hard to compare the outcomes of specific metrics in absolute terms. For instance, two metrics based on completely different principles could declare the same interaction as critical while awarding that interaction very different severity values. This leads to two important, and maybe overly ambitious, recommendations:
- A comparative analysis of several severity metrics should be conducted across the EU to determine appropriate acceptability/criticality thresholds. Further, an serious attempt should be pursued with multiple stakeholders (OEMs, government regulators, academia, etc.) to reach an agreement or convention on how criticality should be defined and measured.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

- The aforementioned effort should lead to creation of a “severity ground truth”. Since such a ground truth would be location based, a concerted effort should be made to acquire NDD from enough representative locations across Europe.
- The successful use of the metrics developed in T2.2 for simulation analysis depends on how well these simulation can be calibrated. If they cannot be well calibrated, the T2.2 metrics will not be to produce numerically realistic results. In such case, these metrics can only be used for comparative analysis.

Finally, the work in T2.2 yielded high-quality results that have been documented in at least two journal publications and one conference paper:

- de Gelder, E.; Adjenughwure, K.; Manders, J.; Snijders, R.; Paardekooper, J.-P.; Op den Camp, O.; Tejada, A. & De Schutter, B., “PRISMA: A Novel Approach for Deriving Probabilistic Surrogate Safety Measures for Risk Evaluation”, Accident Analysis and Prevention, under review, 2021
- X. Wang, J. Alonso-Mora, and M. Wang, “Probabilistic risk metric for highway driving leveraging multi-modal trajectory predictions,” IEEE Transactions on Intelligent Transportation Systems, 2022. (DOI: 10.1109/TITS.2022.3164469)
- X. Wang, Z. Li, J. Alonso-Mora, and M. Wang. "Prediction-Based Reachability Analysis for Collision Risk Assessment on Highways." IEEE Intelligent Vehicles Symposium, Aachen, Germany (2022). (https://arxiv.org/pdf/2205.01357.pdf)

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement 861570.

References

A microscopic analysis of traffic conflict caused by lane-changing vehicle at weaving section. **Uno, Nobuhiro, et al. 2002.** 2002. Proceedings of the 13th Mini-EURO Conference-Handling Uncertainty in the Analysis of Traffic and Transportation Systems, Bari, Italy. pp. 10–13.

A robust and scalable clustering algorithm for mixed type attributes in large database. **Chiu, T., et al. 2001.** 2001. Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 263–268.

Abadi, M. G., et al. 2019. Factors impacting bicyclist lateral position and velocity in proximity to commercial vehicle loading zones: Application of a bicycling simulator. *Accident Analysis and Prevention.* April 2019, pp. pp. 29-39.

Adam: A method for stochastic optimization. **Kingma, Diederik P. en Ba, Jimmy. 2014.** 2014, arXiv preprint arXiv:1412.6980.

Alexiadis, V, Jeannotte, K en Chandra, A. 2004. *Traffic Analysis Toolbox, Volume I: Traffic Analysis Tools Primer.* sl : US Department of Transportation, 2004. FHWA-HRT-04-038.

An LSTM network for highway trajectory prediction. **Altché, Florent en de La Fortelle, Arnaud. 2017.** 2017. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). pp. 353–359.

Analysis of effects of driver/vehicle characteristics on acceleration noise using GPS-equipped vehicles. **Ko, Joonho, Guensler, Randall en Hunter, Michael. 2010.** 2010, Transportation Research Part F: Traffic Psychology and Behaviour, pp. 21-31.

Analyzing human driving data an approach motivated by data science methods. **Wagner, Peter, et al. 2016.** 2016, Chaos, Solitons & Fractals, pp. 37-45.

Anomaly detection with robust deep autoencoders. **Zhou, Chong en Paffenroth, Randy C. 2017.** 2017. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 665–674.

ASAM. 2021. ASAM OpenDRIVE®. [Online] 03 Aug 2021. https://www.asam.net/standards/detail/opendrive/.

Barceló, Jaume. 2010. Models, Traffic Models, Simulation, and Traffic Simulation. *Fundamentals of Traffic Simulation.* New York : Springer, 2010.

Batch normalization: Accelerating deep network training by reducing internal covariate shift. **Ioffe, Sergey en Szegedy, Christian. 2015.** 2015, arXiv preprint arXiv:1502.03167.

Blumenthal, Marjory S, et al. 2020. *Safe Enough: Approaches to Assessing Acceptable Safety for Automated Vehicles.* Santa Monica, CA : RAND Corporation, 2020.

Calibration and validation of simulated vehicle safety performance at signalized intersections. **Cunto, F en Saccomanno, F.F. 2008.** 3, 2008, Accident analysis & prevention, Vol. 40, pp. 1171-1179.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Chen, R. 2016. *Driver behavior in car following-the implications for forward collision avoidance.* s.l. : (Doctoral dissertation, Virginia Tech), 2016.

Clusters of driving behavior from observational smartphone data. **Warren, Josh, Lipkowitz, Jeff en Sokolov, Vadim. 2019.** 2019, IEEE Intelligent Transportation Systems Magazine, pp. 171--180.

Cyclists' anger experiences in traffic: the cycling anger scale. Transportation research part F: traffic psychology and behaviour, 62, 564-574. **Oehl, M, Brandenburg, S en Huemer, A. K. 2019.** 2019.

Defining interactions: a conceptual framework for understanding interactive behaviour in human and automated road traffic. **Markkula, G, et al. 2020.** 6, 2020, Theoretical Issues in Ergonomics Science, Vol. 21, pp. 728-752.

Dittrich, M. 2020. *Persuasive Technology to Mitigate Aggressive Driving - A Human-centered Design Approach.* Würzburg, Germany : Inaugural-Dissertation, 2020.

Doorley, R., et al. 2015. Analysis of heart rate variability amongst cyclists under perceived variations of risk exposure. *Transportation Research Part F.* 2015, pp. pp. 40-54.

Driver behaviour profiles for road safety analysis. **Ellison, Adrian B, Greaves, Stephen P en Bliemer, Michiel CJ. 2015.** 2015, Accident Analysis & Prevention, pp. 118--132.

Driver crash risk factors and prevalence evaluation using naturalistic driving data. **Dingus, T. A., et al. 2016.** 2016, Proceedings of the National Academy of Sciences, pp. 2636-2641.

Dunias, P. 1996. *Autonomous robots using artificial potential fields.* 1996.

Emergency braking performance of motorcycle riders: skill identification in a real-life perception-action task designed for training purposes. **Huertas-Leyva, P., et al. 2019.** 2019, Transportation Research Part F: Traffic Psychology and Behaviour, Vol. 63, pp. 93-107.

Estimating the severity of safety related behaviour. **Svensson, Asse en Hydén, Christer. 2006.** 2, 2006, Accident Analysis and Prevention, Vol. 38, pp. 379-385.

Evans, Leonard. 1991. *Traffic safety and the driver.* sl : Science Serving Society, 1991.

Hamilton, Isobel Asher. 2019. Uber says people are bullying its self-driving cars with rude gestures and road rage. *Business Insider.* [Online] 13 06 2019. [Citaat van: 15 04 2021.] https://tinyurl.com/pw7s7cs8.

Hupfer, C. 1997. *Computergestutzte Videobildverarbeitung zur Verkehrssicherheitsarbeit am Beispiel von Fussgangerquerungen an Stadtischen Hauptverkehrsstrassen.* 1997.

Imagenet classification with deep convolutional neural networks. **Krizhevsky, Alex, Sutskever, Ilya en Hinton, Geoffrey E. 2012.** 2012, Advances in neural information processing systems, Vol. 25, pp. 1097–1105.

ISO 26262, ISO. 2018. ISO 26262-1:2018, Road vehicles -- Functional Safety -- Part 1: Vocabulary. 2018.

Koopman, Philip en Wagner, Michael. 2018. *Toward a framework for highly automated vehicle safety validation.* sl : SAE Technical Paper, 2018.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Kotte, J. en Pütz, A. 2018. *Methodology and Results for the Investigation of Interactions Between Pedestrians and Vehicles in Real and Controlled Traffic Conditions.* Wiesbaden : Springer Fachmedien Wiesbaden, 2018.

Kovácsová, N, de Winter, J en Hagenzieker, M. 2019. What will the car driver do? A video-based questionnaire study on cyclists' anticipation during safety-critical situations. *Journal of Safety Research.* 2019, pp. pp. 11-21.

Laureshyn, A., et al. 2016. *Review of current study methods for VRU safety - Appendix 6.* s.l. : Deliverable 2.1 Part 4, 2016.

Least squares quantization in PCM. **Lloyd, Stuart. 1982.** 1982, IEEE transactions on information theory, pp. 129--137.

Lu, C., et al. 2021. Performance evaluation of surrogate measures of safety with naturalistic driving data. *Accident Analysis & Prevention, Volume 162,.* 2021.

Methodology for identifying car following events from naturalistic data. **Kusano, Kristofer D., Montgomery, Jade and Gabler, Hampton C. 2014.** 2014, IEEE Intelligent Vehicles Symposium, Proceedings, pp. 281-285.

MoRT&H, Ministry of Road Transport & Highway. 2018. *Road Accidents in India.* sl : https://morth.nic.in/sites/default/files/Road_Accidednt.pdf, 2018.

n the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation. **Punzo, Vincenzo, Borzacchiello, Maria Teresa en Ciuffo, Biagio. 2011.** 2011, Transportation Research Part C: Emerging Technologies, vol. 19, no. 6, pp. 1243–1262.

Neale, V. L., et al. 2005. *An overview of the 100-car naturalistic study and findings.* sl : National Highway Traffic Safety Administration, 2005.

Near miss determination through use of a scale of danger. **Hayward, John C. 1972.** sl : Pennsylvania State University University Park, 1972.

NGSIM interstate 80 freeway dataset. **Halkias, John en Colyar, James. 2006.** 2006, Federal Highway Administration (FHWA), Tech. Rep. FHWA-HRT-06-137.

NGSIM US highway 101 dataset. **Colyar, James en Halkias, John. 2007.** 2007, Federal Highway Administration (FHWA), Tech. Rep. FHWA-HRT-07-030.

Nonlinear principal component analysis using autoassociative neural networks. **Kramer, Mark A. 1991.** 1991, AIChE journal, pp. 233--243.

On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment. **Barmpounakis, Emmanouil and Geroliminis, Nikolas. 2020.** 2020, Transportation research part C: emerging technologies, pp. 50--71.

—. **Barmpounakis, Emmanouil en Geroliminis, Nikolas. 2020.** 2020, Transportation research part C: emerging technologies, pp. 50--71.

Oron-Gilad, T en Meir, A. 2020. Understanding complex traffic road scenes: The case of child-pedestrians' hazard perception. *Journal of Safety Research.* February 2020, pp. pp. 111-126.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

Outlier detection with autoencoder ensembles. **Chen, Jinghui, et al. 2017.** 2017. Proceedings of the 2017 SIAM international conference on data mining. pp. 90–98.

Probabilistic field approach for motorway driving risk assessment. **Mullakkal-Babu, F. A., et al. 2020.** 2020, Transportation Research Part C: Emerging Technologies, p. 102716.

Probabilistic Risk Metric for Highway Driving Leveraging Multi-Modal Trajectory Predictions. **Wang, Xinwei, Alonso-Mora, Javier and Wang, Meng. 2022.** 2022, IEEE Transactions on Intelligent Transportation Systems, pp. 1-14.

Research Board, Transportation. 2010. *Highway Capacity Manual (HCM2010).* Washington D.C. : Transportation Research Board of the National Academies, 2010.

Rider Behavioral Patterns in Braking Manoeuvres. **Baldanzini, Niccolò, et al. 2016.** 2016, Transportation Research Procedia, pp. 4374–4383.

SAE International. 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. 2018. J3016_201806.

Safe Enough: Approaches to Assessing Acceptable Safety for Automated Vehicles. **Blumenthal, Marjory S., et al. 2020.** sl : RAND, 2020, RAND Corporation.

Smoothing and differentiation of data by simplified least squares procedures. **Savitzky, Abraham en Golay, Marcel J. E. 1964.** sl : ACS Publications, 1964, Analytical chemistry, Vol. 36, pp. 1627–1639.

Speed limits, enforcement, and health consequences. **Elvik, Rune. 2012.** 2012, Annual review of public health, pp. 225--238.

stimating acceleration and lane-changing 42 dynamics based on ngsim trajectory data. **Thiemann, C, Treiber, M en Kesting, A. 2008.** 2008, 87th Transportation Research Board Annual, vol. 43.

Surrogate Safety Measures from Traffic Simulation Models. **Gettman, D. and Head, L. 2003.** 2003, Transportation Research Record, pp. 104–115.

The effects of drivers' speed on the frequency of road accidents. **Taylor, Marie C, Lynam, DA en Baruya, A. 2000.** 2000, Transport Research Laboratory Crowthorne.

The Power Model of the relationship between speed and road safety: update and new analyses. **Elvik, Rune. 2009.** 2009.

Traffic conflict analysis and modeling of lane-changing behavior at weaving section. **Iida, Yasunori, et al. 2001.** 2001. Proceedings of Infrastructure Planning. Vol. 24, pp. 305–308.

Traffic conflicts technique: state-of-the-art. **Hydén, C. 1996.** sl : Transportation Dept., Univ. Kaiserslautern Germany, 1996, Traffic safety work with video processing, Vol. 37, pp. 3–14.

Traffic safety dimensions and the power model to describe the effect of speed on safety. **Nilsson, Goran and others. 2004.** 2004.

Travelling speed and the risk of crash involvement volume 2-case and reconstruction details. **Kloeden, CN, et al. 1997.** 1997, Adelaide: NHMRC Road Accident Research Unit, The University of Adelaide.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

USDOT. 2006. *NGSIM -- Next Generation Simulation.* sl : US Department of Transportation, 2006.

Use of speed limiters in cars for increased safety and a better environment. **Almqvist, Sverker, Hydén, Christer en Risser, Ralf. 1991.** 1991, Transportation Research Record.

Using naturalistic driving data to assess variations in fuel efficiency among individual drivers. **LeBlanc, David J, Sivak, Michael en Bogard, Scott E. 2010.** 2010, University of Michigan, Ann Arbor, Transportation Research Institute.

Variational autoencoder based anomaly detection using reconstruction probability. **An, Jinwon en Cho, Sungzoon. 2015.** 2015, Special Lecture on IE, Vol. 2, pp. 1–18.

Zhuang, X, et al. 2020. *Pedestrian estimation of their crossing time on multi-lane roads. Accident Analysis & Prevention, 143, 105581.* 2020.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.