This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 814761

D3.6

Enhanced toolbox of recommended data collection tools, monitoring methods and interventions including thresholds for the safety tolerance zone

Safe tolerance zone calculation and interventions for driver-vehicle-environment interactions under challenging conditions

iDREAMS

Project identification

Grant Agreement No	814761
Acronym	i-DREAMS
Project Title	Safety tolerance zone calculation and interventions for driver-vehicle-environment interactions under challenging conditions
Start Date	01/05/2019
End-Date	30/04/2022
Project URL	www.idreamsproject.eu

Document summary

Deliverable No	D3.6
Deliverable Title	Enhanced toolbox of recommended data collection tools, monitoring methods and interventions including thresholds for the safety tolerance zone
Work Package	3
Contractual due date	31/08/2021
Actual submission date	14/08/2021
Nature	Report
Dissemination level	Public
Lead Beneficiary	Loughborough University
Responsible Author	Rachel Talbot
Contributions from	Rachel Talbot, Graham Hancox - Loughborough University (Lough). Tom Brijs, Kris Brijs, Veerle Ross - Universiteit Hasselt (UHasselt). Christos Katrakazas, Eva Michelaraki - National Technical University of Athens (NTUA). Bart De Vos- DriveSimSolutions (DSS) Chiara Gruden, Matjaž Šraml - Univerza V Mariboru (UM) Catia Gaspar, André Lourenço, Carlos Carreiras - Cardioid Technologies Lda (Cardio ID). Amir Pooyan Afghari , Eleonora Papadimitriou - Technische Universiteit Delft (TU Delft) Kui Yang, Alam Md Rakibul, Christelle Al Haddad, Constantinos Antoniou (TUM)

Please refer to the document as:

Talbot, R. et al. (2021). Enhanced toolbox of recommended data collection tools, monitoring methods and interventions including thresholds for the safety tolerance zone. Deliverable 3.6 of the EC H2020 project i-DREAMS

Revision history (including peer review & quality control)

Version	Issue date	% Complete[1]	Changes	Contributor(s)
V0.1	03/06/2021	80%	Initial deliverable structure and content formatting	Graham Hancox – content provided by authors listed above
V0.2	15/06/2021	95%	Addition of intro, conclusion and exec summary; editorial suggestions	Rachel Talbot
V0.3	1/7/2021	100%	Minor edits / improvements – Review version	Rachel Talbot, Christos Katrakazas, Chiara Gruden
V0.4	11/8/2021	100%	Responses to review comments and IP review	Tom Brijs, Christos Katrakazas, Rachel Talbot
V1.0	12/8/2021	100%	Final version	

Disclaimer

Copyright

[1] *i-DREAMS quality assurance process:*

Due date – 3 months: ready for internal review
Due date – 2 months: start review
Due date – 1 month: end review
Due date – 5 days: ready for submission / coordinator submits

Table of contents

List of Figures

List of Tables

Glossary and abbreviations

Word / Abbreviation	Description
ADAS	Advanced Driving Assistance System
DBN	Dynamic Bayesian Networks
ECG	Electrocardiogram
FCW	Forward Collision Warning
GPS	Global Positioning System
IMU	Internal Measurement Unit
KSS	Karolinska Sleepiness Scale
LCD	Liquid-Crystal Display
LDW	Lane Departure Warning
LSTMs	Long Short-Term Memory Networks
MMWG	Mathematical Model Working Group
OSM	Open Street Maps
PCW	Pedestrian Collision Warning
PPG	Photoplethysmogram
STZ	Safety Tolerance Zone
UFCW	Urban Forward Collision Warning
WP	Work Package

Executive Summary

The i-DREAMS project aims to establish a framework for the definition, development, testing and validation of a context-aware safety envelope for driving called the "Safety Tolerance Zone' (STZ). Taking into account driver background factors and real-time risk indicators associated with the driving performance as well as the driver state and driving task complexity indicators, a continuous real-time assessment will be made to monitor and determine if a driver is within acceptable boundaries of safe operation. Moreover, safety-oriented interventions will be developed to inform or warn the driver in real-time as well as on an aggregated level after driving, through an app-and web-based gamification coaching platform (post-trip intervention).

The conceptual framework of the i-DREAMS platform integrates aspects of monitoring (such as context, operator, vehicle, task complexity and coping capacity), to develop a Safety Tolerance Zone for driving. In-vehicle interventions and post-trip interventions will aim to keep the drivers within the Safety Tolerance Zone as well as provide feedback to the driver. This conceptual framework will be tested in simulator studies and three stages of field trials in Belgium, Greece, Germany, Portugal, and the United Kingdom with over 600 participants representing car, bus, truck, and rail drivers.

This deliverable (D3.6) is an update of Deliverable 3.2, the second deliverable of the Operational Design Work Package of i-DREAMS. The original aim of D3.2 was to provide a more concrete description of the STZ by defining variables, values and thresholds associated with each phase of the STZ. D3.2 also made the initial steps in identifying mathematical models which had the potential to explain and analysis the resulting data – both from a real-time and post trip perspective. D3.2 was published early in the project (beginning of 2020) and since then there have been a number of further developments. These include, algorithm creation, technology updates and fitting test vehicles with the full technology to test performance of both the technology and STZ calculations. In addition the analysis work packages have commenced – WP6, analysis of risk factors and WP7, Evaluation of safety interventions. Together with the Mathematical Model Working Group (MMWG), a group formulated as part of WP3 to focus on evaluating and identifying the most appropriate analysis methodologies, WP6 and WP7 leaders have refined their plans based on the available data.

This deliverable therefore constitutes an update to selected sections of D3.2. The original authors have worked with the WP3 partnership to update original text and added new text to this deliverable to reflect the above described developments in WP4, 6 and 7 as well as the MMWG work.

The variables proposed in D3.2 have been confirmed with those available using the iDREAMS platform as developed by WP4. This has resulted in a list of variables that can be measured for which mode and that can be used to calculate STZ phases.

The real-time warning strategies for the four performance objectives (Headway, Illegal overtaking, Speeding, Fatigue) that can be assigned variable thresholds are defined and threshold ranges are assigned to each STZ phase. For each of these four strategies additional variables can be used as indicators and/or modifiers and the types of real-time warnings are outlined.

Driving style, in terms of 'normal' (STZ normal phase) and 'abnormal' (STZ danger and avoidable accident phase) is discussed and it was concluded that it is necessary to account for the possibility of the driver being in a 'normal' driving style for one performance indicator and an 'abnormal' driving style for another.

A key aspect of defining the STZ, is measuring task complexity and (driver) coping capacity with safe driving defined as when these two dimensions are in balance. Therefore also which variables are associated with each of these are defined, alongside the method and frequency of recording, which mode are applicable and whether real-time or post-trip modelling methodologies are required for analysis.

Alternative definitions of risk are discussed and described that relate to the STZ phases or the detection of an 'event' (discrete variables). In addition, ways in which the overall risk during a period of time are defined e.g. a composite STZ value or proportion of time spent in a STZ phase (continuous variables)

Finally detailed descriptions of the relevant mathematical models (Dynamic Bayesian Network, Long Short-Term Memory, Discrete Choice Models, and Structural Equation Models) are provided with an explanation as to when they could be used for analysis. This depends on the variable type (discrete, continuous) and when the associated values are calculated (real-time or post-trip) were provided. For each model, the relevant independent variables or risk definitions that can feed into the model were defined and the relevant equations/functions were defined.

Over the next six months, on road field trials will be conducted for the passenger car, bus and truck mode and simulator trials for the rail mode. Any learning from the simulator and field trials or changes to the platform that relate to this deliverable will be documented in the WP7 and WP6 deliverables that will be published at the end of the project.

1 Introduction

1.1 About the project

The overall objective of the i-DREAMS project is to set up a framework for the definition, development, testing and validation of a context-aware safety envelope for driving ('Safety Tolerance Zone'-STZ), within a smart Driver, Vehicle & Environment Assessment and Monitoring System (i-DREAMS). Taking into account driver background factors and real-time risk indicators associated with the driving performance as well as the driver state and driving task complexity indicators, a continuous real-time assessment will be made to monitor and determine if a driver is within acceptable boundaries of safe operation. Moreover, safety-oriented interventions will be developed to inform or warn the driver in real-time in an effective way as well as on an aggregate to give real timed level after driving through an app and web-based gamified coaching platform. Figure 1 summarises the conceptual framework, which will be tested in a simulator study and three stages of on-road trials in Belgium, Germany, Greece, Portugal, and the United Kingdom with a total of 600 participants representing car driver, bus driver, truck drivers and rail drivers.

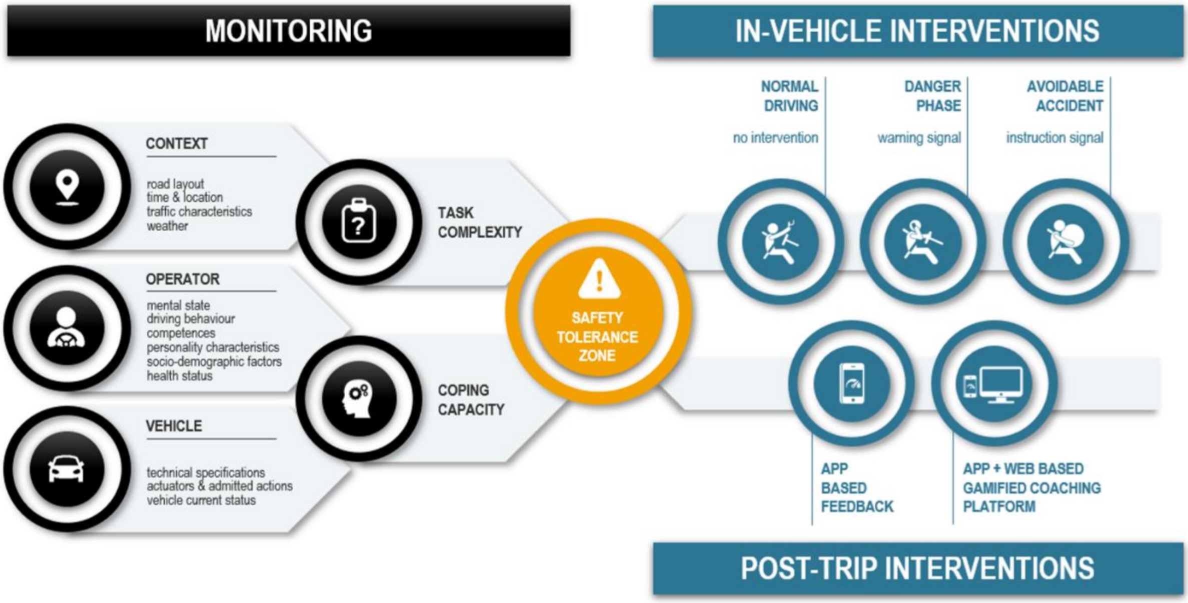

Figure 1: Conceptual framework of the i-DREAMS platform.

The key output of the project will be an integrated set of monitoring and communication tools for intervention and support, including e.g., in-vehicle assistance and feedback and notification tools as well as a gamified platform for self-determined goal-setting working with incentive schemes, training, and community building tools. The technology that will be implemented includes a customised LCD capacitive touch display that communicates with the CardioID Gateway to receive the status of the STZ, giving real-time audio and visual alerts. It will also allow for driver identification upon vehicle start-up. Information coming to the CardioID Gateway is from a context-aware road monitoring system (Mobileye), and electrocardiogram (ECG), or photoplethysmography (PPG) technology (CardioWheel/ Wristband), as well as an application installed on the user's phone to monitor hand-held phone usage.

1.2 Deliverable overview and report structure

This deliverable (D3.6) is an update of Deliverable 3.2, the second deliverable of the Operational Design Work Package of i-DREAMS. The original aim of D3.2 was to provide a more concrete description of the STZ by defining variables, values and thresholds associated with each phase of the STZ. D3.2 also made the initial steps in identifying mathematical models which had the potential to explain and analysis the resulting data – both from a real-time and post trip perspective. One of the key roles of D3.2 was to provide the project with early indications of the best inputs to algorithms etc by comparing the output of WP2 which examined the state of the art for monitoring and associated technology to the available technology as defined in WP4, technical implementation.

D3.2 was published early in the project (beginning of 2020) and since then there have been a number of further developments. These include, algorithm creation, technology updates and fitting test vehicles with the full technology to test performance of both the technology and STZ calculations. In addition the analysis work packages have commenced – WP6, analysis of risk factors and WP7, Evaluation of safety interventions. Together with the Mathematical Model Working Group (MMWG), a group formulated as part of WP3 to focus on evaluating and identifying the most appropriate analysis methodologies, WP6 and WP7 leaders have refined their plans based on the available data.

This deliverable therefore constitutes an update to selected sections of D3.2. The original authors have worked with the WP3 partnership to update original text and added new text to this deliverable to reflect the above described developments in WP4, 6 and 7 as well as the MMWG work.

1.2.1 Content of deliverable

D3.6 is divided into three main technical sections, each designed to update key aspects of D3.2. Section 2, i-DREAMS performance objectives and the Safety Tolerance Zone phases, presents the final selection of driver performance objectives, the associated variables and how these relate to the three phases of the STZ. Where appropriate, thresholds are stated and distinctions are made between the four i-DREAMS modes: Private Cars, Buses, Trucks and Rail (trains and trams). This is an update of D3.2 sections 3.1.4 and 3.1.5.

Section 3, Intervention strategies, focuses on when interventions are triggered, how driving style can be incorporated into STZ calculations and details the variables relating to task complexity and coping capacity and how they are measured for each mode. This is also related to whether these variables require a real-time or post trip modelling technique. This is an update of D3.2, sections 3.2.1 and 3.2.2.

The final technical section, section 4, describes the four mathematical models that were identified in D3.2 and how these can be applied in the i-DREAMS risk analyses. This section presents definitions of risk as an output variable and distinguishes between real-time and post-trip data analyses. The section also sets out some requirements for evaluating intervention effectiveness.

1.2.2 The influence of the COVID-19 pandemic

This deliverable, as originally planned, would have included a brief description of the relevant learning gained from the simulator studies and the first phase of the field trials. Unfortunately the project suffered substantial delays both in terms of the technology being ready (component shortages required last minute substitutes, redesigns and additional testing) and access to participants – all test site countries (Belgium, Germany, Greece, Portugal, UK) were subject to restrictions in terms of travel and social contact causing. As of July 2020, simulator studies have only been possible in Belgium and Germany and the field trials have only just started at some sites. This means no meaningful evaluation can be included here. As all other updates could be made, the decision was made not to delay this deliverable and any significant learning or updates from the trials will be reported within WP6 and WP7 deliverables.

The reader of this Deliverable must be informed that some information in this document, such as exact threshold values and detailed intervention algorithm logic, were purposely left out to protect the unique IP generated in the project and to maximise the potential for commercial valorisation afterwards. Additional details can be requested from the authors.

2 i-DREAMS performance objectives and the Safety Tolerance Zone phases

2.1 Aim of the section

This section aims to review the efforts made, and improvements achieved, during the project. It will highlight the progress made in the selection of indicators, the definition of real-time interventions and the thresholds of the indicators used. All these elements have been updated and finalised in the past months in relation to the real needs for the implementation of the Safety Tolerance Zone (STZ).

2.2 Definition of the Safety Tolerance Zone and its phases

The i-DREAMS intervention aims to effectively increase driving safety by assisting the driver in his/her driving task. To achieve this purpose, the Safety Tolerance Zone concept has been developed (Table 1), in which three different driving phases can be identified: normal, danger and avoidable accident phase. As set out in Deliverable 3.2 (Katrakazas et al, 2020), the normal driving phase represents the conditions in which a crash is unlikely to occur, i.e., the crash risk is low. During this phase, the driver can successfully adapt his/her behaviour to meet the task demand. The danger phase is characterised by changes in normal driving that indicate that a crash may occur, therefore, the crash risk is increased. Finally, the avoidable accident phase occurs when a collision scenario develops but there is still time for the driver to intervene and avoid the crash. The need for action is more urgent than in the danger phase and if the driver does not adapt his/her behaviour to the current circumstances, a crash is very likely to occur.

Table 1: Phases of the STZ

Phases of STZ	Description
Normal driving phase	Crash risk is minimal
Danger phase	Risk of crash increases as internal / external events occur
Avoidable crash phase	Crash is very likely to occur if no preventative action taken by driver

The fundamental goal of the i-DREAMS platform is to keep the driver in the normal driving phase for as long as possible and, where this is not possible, to prevent the transition from the danger to the avoidable accident phase. To this end, the platform combines both real-time and post-trip interventions which, respectively, aim to nudge and coach the driver The platform is a warning based driver assistance system, it does not actively intervene physically in any way with the driving task. The abstract concept of the STZ is operationalised at the level of performance objectives. To estimate in which STZ phase the driver is in and which interventions should be provided, the i-DREAMS platform uses two modules. First, it uses the monitoring module, which takes measurements related to the context, the operator, and the vehicle, to derive the demands of the driving task and the driver's ability to cope with these demands. This module estimates at which stage of the STZ the driver is operating at any given time. More specifically, the monitoring module registers driving behaviour related to a list of performance objectives as shown in Figure 2[2] (from Deliverable 3.3, Brijs et al., 2020). For these different performance objectives, events are detected. Second, the in-vehicle intervention module is responsible for keeping the driver within the normal phase of the STZ, either by providing a warning or alert during the trip (real-time intervention) or by providing feedback about the journey after the completion of the driving task (post-trip intervention). In case of real-time interventions, a different type of in-vehicle warning is being delivered to the driver depending on the severity of the detected event.

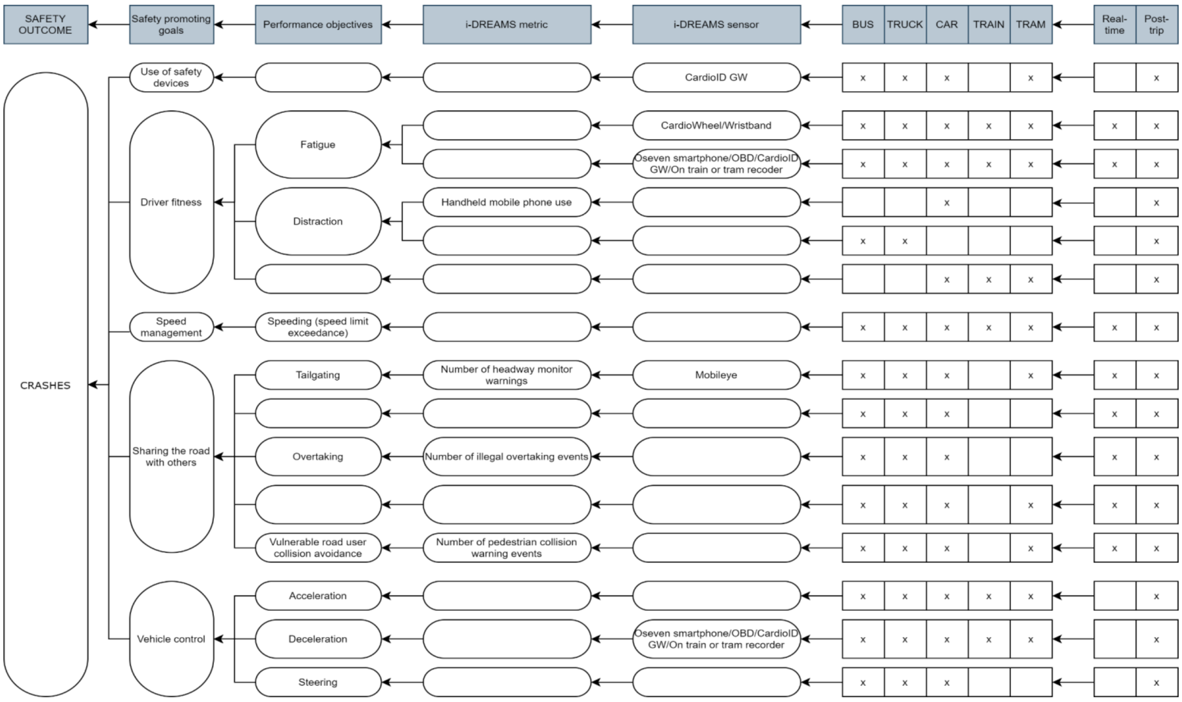

Figure 2: Safety promoting goals and related parameters

For the real-time interventions, a nudging approach is used since the driver has little time to think about his actions. This approach uses heuristics (i.e., mental shortcuts) and manipulation of cues within a social or physical environment to activate unconscious thought processes involved in human decision making. The delivered type of real-time intervention depends on the retrieved STZ phases: in the normal driving phase, no intervention is required. When it is detected that the driver has entered the danger phase, a warning or an indication should be given. Meanwhile, in the avoidable accident phase, a more specific intervention is required such as an intrusive warning signal (accompanied or not by an instruction) that prompts the driver to take decisive action. With respect to the post-trip interventions, nudging is being reinforced by a coaching platform that operates outside the context of a trip.

[2] Some information in Figure 2 is purposely left out for reasons of confidentiality

2.3 Comparison between the initially considered indicators and the final selection

In Deliverable 3.2, a comprehensive overview of the state of the art regarding driving performance indicators was developed. Covering all driving situations and indicators would go beyond the scope of the current project. It has already been established that extensive data from both driving simulators and field tests are needed to cover most driving scenarios and to capture episodes of abnormal driving. Therefore, in the course of the project, the need for a selection of specific indicators became evident during the practical implementation of the algorithm and the use of the platform at different stages. Some of these indicators confirm those discussed in D3.2, while others were added to provide a more complete description of the driving task and driver skills. Table 1 shows a comparison between the indicators reviewed in D3.2 and those effectively used for the STZ calculation.

Table 2: Comparison between the variables reviewed in D 3.2 and the available variables effectively operationalised in i-DREAMS platform

Variables considered in D 3.2 after literature review and the operationalised variables defined by WP4.					
Deliverable D 3.2	**Implemented in i-DREAMS platform for STZ calculation**				
			Availability per mode		
Description	**Source**	**Description**	**Cars**	**Trucks/Buses**	**Rail**
Headway time	Mobileye (AWS)	Headway time	*	*	*
		Vehicle ahead detected	*	*	*
Pedestrian collision warning (PCW)		Pedestrian collision warning (PCW)	*	*	*
Urban forward collision warning (UFCW)		Urban forward collision warning (UFCW)	*	*	*
Forward collision warning (FCW)		Forward collision warning (FCW)	*	*	*
Left lane departure warning		Left lane departure warning	*	*	
Right lane departure warning		Right lane departure warning	*	*	
		Low visibility indicator	*	*	*
		Time of day indicator	*	*	*
Speed exceedance[1] and speed at turns indication[2]		Speed limit sign recognition	*	*	
	Mobileye (Cars)	Wipers indicator	*	*	*
Longitudinal acceleration / deceleration		Braking indicator	*	*	*
Speed exceedance[1] and speed at turns indication[2]		Speed	*	*	*
		Left turn signal indicator	*	*	
		Right turn signal indicator	*	*	
	GPS	Location (latitude and longitude)	*	*	*
Speed exceedance[1] and speed at turns indication[2]		Speed	*	*	*
Heading		Heading	*	*	*
Driver attention level (sleepiness level)	CardioWheel	Sleepiness (from ECG signal)		*	
		Driver change detection (from ECG signal)		*	
Hands on wheel		Hands on wheel detection		*	
Steering wheel accelerometer		Steering wheel dynamics		*	
	Wristband	Sleepiness (from PPG signal)	*		*
	OSeven app	handheld mobile phone use	*	*	
Harsh acceleration	Gateway	Harsh acceleration / braking / cornering (via IMU)	*	*	*
Long driving hours		Trip duration timer	*	*	*
[1]Speed exceedance is processed based on speed limit indication and measured vehicle speed; [2]Speed turns indication is processed on turn indication activation and speed exceedance.					

It can be seen from Table 1 that Deliverable D 3.2 did not specifically identify which variables were appropriate for cars, which for buses/trucks and which for rail. This distinction was possible as the project progressed and is indicated in Table 8, Table 9 and Table 10The resumed indicators are closely related to the types of interventions provided to drivers, which are described in the next sections.

2.4 Defining the interventions

As mentioned above, the purpose of real-time interventions is to keep vehicle operators within the normal phase of the STZ or avoid the transition from the danger to the avoidable accident phase. Real-time interventions are triggered based on crucial inputs from the implementation of the STZ.

Depending on is the respective STZ phase, these in-vehicle interventions work differently. In the normal driving phase, the i-DREAMS platform's monitoring module is active and keeps track of task complexity and available coping capacity in real-time. However, in the normal driving phase no warnings are issued to the vehicle operator, as there is no indication that a collision scenario is evolving. When the vehicle operator enters the danger phase, the intervention module of the i-DREAMS platform generates a warning message to alert the driver, in order to prevent him/her from entering a collision scenario. When the vehicle operators are in the avoidable accident phase, the intervention module of i- DREAMS sends a warning message to the operators in order to prompt an instinctive reaction. Each risk scenario (forward collision, over speeding) has its own specific symbol and sound-symbol that changes in intrusiveness (size, sound level, intensity) according to the STZ stage. These in-vehicle interventions are tailored to the specific characteristics and regulations of the transport mode and country.

During the project, several improvements were made regarding the practical definition of to-be-used real-time interventions, indicators, and thresholds.

Four different groups of real-time intervention strategies were proposed in WP4:

- Headway warning (Tailgating)
- Illegal overtaking warning
- Speeding warning
- Fatigue warning.

These strategies take into consideration indicators to estimate task demand and driver capacity to modify the thresholds accordingly and to provide the right intervention. The main indicators and associated thresholds used by each strategy are listed below.

On top of these four context-based interventions, there are also interventions with hard-coded thresholds. These are:

- Lane departure warning
- (Urban) Forward Collision warning
- Pedestrian Collision warning
- Distraction warning (phone in hand)

2.4.1 Real-time headway warning strategy

The headway intervention strategy is essentially based on comparing the current time headway with a time headway threshold that is constantly updated.

The threshold is calculated as follows:

Threshold Value (THW) = Base Threshold Value (THW) + Penalty value (THW)

The variable threshold is calculated by adding a penalty value (to make the intervention trigger sooner) to a base threshold value. The base threshold value is set based on vehicle speed, i.e. a lookup table that links time headway thresholds to vehicle speed is used. The penalty value is based on indicators that estimate task complexity (weather, time of day) and driver capacity (distraction, KSS score, trip duration). The result is a variable time-headway threshold with boundaries of:

- 2.2s – 1.0s time headway for the threshold between normal driving and dangerous driving
- 1.2s – 0.6s time headway for the threshold between dangerous driving and avoidable accident.

This means that a time headway below 0.6s will always be considered as being in the avoidable accident stage and a time headway above 2.2s will always be considered as normal driving. Table 2 summarises the typology and range of the considered thresholds.

Table 3: Thresholds applied in the real-time headway warning strategy

Threshold Variable	Time headway (s)
Threshold modifiers	Vehicle speed, weather, time of day, distraction, KSS score (sleepiness), trip duration.
Threshold boundaries dangerous driving	1.0s – 2.2s
Threshold boundaries avoidable accident	1.2s – 0.6s

2.4.2 Real-time illegal overtaking warning strategy

This real-time strategy warns the driver who attempts an illegal overtaking manoeuvre, intended as the manoeuvre of crossing the median line, in a no-overtaking road section, as indicated by a no-overtaking road sign. The idea behind this strategy is that it aims to discourage the driver from performing the illegal overtaking manoeuvre. Moreover, it should discourage the driver from performing illegal overtaking manoeuvres in the near future.

The first step in the intervention strategy is to detect if an illegal overtaking manoeuvre is taking place. The intervention will only work on road sections where an overtake manoeuvre is not allowed and when vehicle speed is above 35 km/h, to avoid false-positive manoeuvre detection on intersections and other manoeuvring at low speeds. The algorithm detects an overtaking manoeuvre based on usage of the turn signal, lane crossing, steering behaviour (yaw rate) and acceleration. A minimal **acceleration** of 0.2 m/s² is also required for any manoeuvre to be considered as an overtaking manoeuvre.

In the next step, when an illegal overtaking manoeuvre has been detected, threshold curves are used. There are different curves, the correct curve is selected based on transport mode (Car, Bus, Truck, Tram) and weather conditions (wet/dry). The curves identify speed and acceleration values at which driving can be considered normal or abnormal. Different curves are used because in general accelerations are lower for heavy vehicles like buses and trucks compared to cars. Together with the threshold curves, contextual indicators (driving duration, distraction, KSS score) are used to determine the intervention type. The variables needed to detect the manoeuvre, the minimal conditions, the thresholds variable, and modifiers are listed in Table 3.

Table 4: Thresholds and minimal conditions utilised in the illegal overtaking warning strategy

Variables for manoeuvre detection	Overtaking restriction, vehicle speed, acceleration, turn signal usage, lane crossing, steering behaviour,
Minimal conditions	Vehicle speed > 35 km/h Acceleration > 0.2 m/s^2
Threshold variable	acceleration
Threshold modifiers	Weather, transport mode, KSS score, trip duration, distraction

2.4.3 Real-time speeding warning strategy

This warning strategy proposes driving speed thresholds for interventions, that are related to contributing factors that may negatively impact road safety, such as weather conditions, time of day, high-risk hours, fatigue, distraction, and sleepiness. Each one of these factors has a specific value and influence the proposed thresholds following the formula:

Proposed driving speed threshold for interventions = normal driving speed limits * contributing factors

The threshold boundaries that result from this formula and the combination of contributing factors that are being used are:

- (1.0325 * speedlimit) -> (1.1 * speedlimit) for the threshold between normal driving and dangerous driving
- (1.0475 * speedlimit) -> (1.15 * speedlimit) for the threshold between dangerous driving and avoidable accident stages.

The real time speed strategy is summarised in Table 5.

Table 5: Threshold ranges for real time speed strategy

Threshold Variable	Driving speed
Threshold modifiers	Weather conditions, time of day, high-risk hours, fatigue, sleepiness, distraction
Threshold boundaries dangerous driving	(3.25% - 10%) to (4.75%-15%) above speed limit
Threshold boundaries avoidable accident	Driving speed exceeds 4.75%-15% above speed limit

2.4.4 Real-time fatigue warning strategy

Fatigue is considered in this project as the inability to continue a task that has been going on too long and is distinguished from sleepiness, which is defined as the need to fall asleep. Nevertheless, these two states are closely related, which is why thresholds for both fatigue and sleepiness are implemented in this warning strategy. Fatigue is measured as the driving duration, while sleepiness is measured using the KSS. Regarding fatigue, it is worth mentioning that the driving time for private drivers can last 5 hours or more, while the driving time for professional drivers is limited to 4.5 hours, as required by the regulations of the European Union. A system of thresholds for fatigue and KSS were therefore defined as reflected in Table 4.

Table 6: Fatigue and KSS thresholds for the real-time fatigue warning strategy

Fatigue	Drive duration < 2 hours: normal driving
	Drive duration = medium: danger phase
	Drive duration = long: avoidable accident phase
	Drive duration = very long: avoidable accident phase
KSS	KSS = low: normal driving
	KSS = medium: danger phase
	KSS = high: avoidable accident phase
	(-1) if KSS is unknown
Driving duration for professional drivers	4.5 hours

However, instead of the STZ thresholds being fixed across all drivers, the i-DREAMS fatigue intervention algorithm takes into account additional contributing factors, i.e., age and gender which modify the threshold levels for Drive duration (i.e. medium, long, very long = f(age, gender)). This is choice is motivated by previous research (Filtness et al. 2012; Campagne et al 2004) showing indications of greater vulnerability of male and younger drivers under prolonged driving. See Pilkington-Cheney et al. (2021) for a more detailed explanation of the fatigue warning strategy.

2.4.5 Real-time interventions for rail

As far as real-time interventions for the rail mode are concerned, they are generally similar to those for road modes. However, due to the specificities of this transportation mode, the warning strategy for illegal overtaking cannot be applied and the thresholds for headway warning interventions should be defined in cooperation with the tramway operator.

Recently some additional information has been obtained from the tramway operator, which permit to relate travel speed (in relation to speed limit) to the different stages of the safety tolerance zone. If travel speed results within the posted speed limit, the operator is driving within the normal driving phase; the driver enters the danger phase, and a warning should be given, if the travel speed is over the speed limit by +3 km/h; finally, an urgent warning is needed, when the travel speed is +5 km/h over the speed limit. This threshold defines the avoidable accident phase.

2.5 The connection between the thresholds and the three STZ phases

Unlike previous deliverables, in which the relationship between the three phases of the STZ and the thresholds was only proposed theoretically, it has now been made concrete. As can be seen in Table 5, each of the developed warning strategies has its own thresholds for its specific indicators, which change according to the STZ phase.

Table 7: Connection between the three STZ phases and the thresholds for the 4 warning strategies

	Real-time headway warning strategy	**Real-time illegal overtaking warning strategy**	**Real-time speeding warning strategy**	**Real-time fatigue warning strategy**
NORMAL DRIVING PHASE	THW > variable threshold (1.0s – 2.2s)	acceleration ≤ 0.2m/s^2 OR speed < 35km/h OR turn signal and LDW indicator = 0	driving speed < variable threshold 1 (3.25% - 10% above speed limit)	DD*<2 hrs AND KSS = low
DANGER PHASE	THW < variable threshold 1 (1.0s – 2.2s) AND THW > variable threshold 2 (1.2s – 0.6s)	acceleration ≥ 0.2 m/s^2 and other indicators (KSS, etc.) are in normal ranges	driving speed between variable threshold 1 (3.25% - 10% above speed limit) and variable threshold 2 (4.75% - 15% above speed limit)	DD<2 hrs AND KSS=medium; DD=medium AND KSS=low;
AVOIDABLE ACCIDENT PHASE	THW < variable threshold 2 (1.2s – 0.6s)	Acceleration > 0.2 m/s^2 and other indicators (KSS, etc.) are in abnormal ranges	driving speed > variable threshold 2 (4.75% - 15% above speed limit)	DD<2 hrs AND KSS=medium or low; DD=medium AND KSS=medium; DD=medium AND KSS=high; DD=long AND KSS=low; DD=long AND KSS=med/high; DD=very long
**DD = Driving duration, THW = Time Headway*				

3 Intervention strategies

3.1 Recommendations on triggering interventions

As described in previous deliverables of the project (e.g., Deliverable 3.3; Brijs et al., 2020), safety-oriented interventions will be developed to prevent drivers from getting too close to the boundaries of unsafe operation and to bring back the driver into the Safety Tolerance Zone (STZ). There are four different intervention stages of the on-road trials which will be offered. To begin with, in the first stage (i.e. baseline), no interventions will be provided. In the second stage, only real-time interventions (real-time alerts and notifications) will be implemented with the help of an in-vehicle warning system. Then, in the third stage, real-time interventions and post-trip interventions will be provided analysed in order to assess their effectiveness on driving behaviour (e.g., safety-critical events, near misses etc.) and driver state, on the basis of the methodology set out in Deliverable 7.1 (Katrakazas et al., 2020). Lastly, in order to increase the impact of interventions on driver safety, in the fourth stage, both real-time interventions and gamification elements of post-trip interventions will be given, since both are complementary.

With regards to real-time interventions, there are four different intervention strategies implemented for each transport mode (cars, buses, and trucks). Firstly, a real-time headway warning strategy has been developed which takes into account some other important indicators, such as distraction, Karolinska Sleepiness Score (KSS), fatigue, driving duration, braking as well as the interaction between them. In addition, a detailed methodology of a real-time illegal overtaking warning strategy has been proposed. In this strategy, several task complexity and coping capacity indicators have been taken into account, such as left turn, left direction, lane departure warning for left lane, speed, acceleration, weather conditions, wiper activation, time since trip started (in hours) per driving duration and KSS. Furthermore, a real-time speeding warning strategy has been provided, containing some other important risk factors of task complexity (i.e., weather conditions, road layout, time of the day) and coping capacity (i.e., fatigue, distraction, sleepiness). It should be noted that in this intervention strategy, the combination of task complexity and copying capacity is also investigated. Finally, a real-time fatigue warning strategy has been developed. It is worth mentioning that driving performance indicators, such as, driving duration, KSS and loss of sleep, or other demographic characteristics delivered from questionnaires (e.g., gender, age) are available. Thus, based on the outputs derived from each real-time intervention strategy, the corresponding thresholds will be provided.

Taking all the aforementioned strategies into account, different warning triggering thresholds have been proposed for the three STZ levels (i.e., normal phase, dangerous phase, avoidable accident phase). In the normal phase, no warnings will be provided to the driver. In the dangerous phase, visual warnings, possibly a symbol that would indicate improving driving behaviour will be given. Lastly, in the avoidable accident phase, visual warnings with different colour accompanied by an auditory alert that would indicate improving driving behaviour immediately will be triggered.

3.1.1 Driving style recognition and STZ concept

It has been discussed in the previous deliverables (Section 3.2 of Deliverable 3.2; Katrakazas et al., 2020) that the driving style recognition can be considered as a critical input in the framework that controls the intervention triggering mechanism for different risk situations. However, and in addition to the intervention-triggering framework, the driving style recognition and the factors affecting it can be a critical input for the data analysis framework too. The aim of this latter framework is to identify the level of the STZ which the driver is within at any given moment and to understand the factors contributing to the variation of the level of the STZ for each risk situation. The effects of various factors, however, may be reflected in different driving styles. For example, while driving at the speed limit may be associated with the "normal driving phase" of the STZ in rear-end collision risk situations when the driver is not fatigued, the same factor may be associated with the "danger phase" of the STZ when the driver is fatigued. This moderation effect may appear in other "abnormal" driving styles too and may influence the effects of risk factors on the STZ. As a result, it is of great importance to take driving style recognition into account in the data analysis framework. To do so, the driving style factors are recognised (i.e., controlled for) within the mathematical models for data analysis in i-DREAMS (i.e., machine learning models and statistical models) by including them as features and/or independent variables (elaborate discussions about these models will be presented in the next chapter).

3.1.2 i-DREAMS technology and risk indicators for driving styles with recommendations of threshold values

Determining driving style within a driving simulator experiment would mostly be impracticable since it would require numerous hours of driving for each individual to clearly identify their respective driving style. Therefore, it is foreseeable that the required data may only be available as input for on-road trials within i-DREAMS. Given the i-DREAMS technology, limited risk situations (accident types) can be covered such as rear-end collision, or collision with pedestrians ahead. Lane departure warning is also available to cover head-on collision in situations, where driving is carried out on two-lane roads. It is, therefore, recommended that a driving style or behaviour recognition and their indicators are generalised and applicable for all risk situation that one is able to capture with the available technology. An overview of the available measurements (risk indicators) that are going to be used in the experiments in order to identify the STZ or the normal/abnormal driving behaviour in all transport modes (i.e., cars, buses, trucks and rails) is presented below.

Table 8: Recorded variables to indicate/measure task complexity

Variable relevant to task complexity	Method of recording	Frequency of recording	Relevant modes	Real-time/Post-trip modelling
Time of day	OSeven app, CardioID Gateway (day, night), Dashcam	Recorded per second	Car, truck, bus, rail	Real-time/ Post-trip
Wipers on/off	Mobileye	Event	Car, truck, bus, tram	Real-time/ Post-trip
Low visibility indicator	Mobileye	Event	Car, truck, bus, rail	Real-time/ Post-trip
Road environment	CardioID Gateway GPS, road type info from OSM if available (i.e., urban, suburban, rural, highway), Mobileye, OSeven (i.e., speed limits, GPS signal)	50 - 100 meters	Car, truck, bus, rail	Post-trip

Variable relevant to task complexity	Method of recording	Frequency of recording	Relevant modes	Real-time/Post-trip modelling
Road geometric configuration	CardioID Gateway GPS (i.e., spirals, geometric design, straight road, curves, sharp bend, narrow/wider lanes)	Needs map matching	Car, truck, bus, rail	Real-time/ Post-trip
Trip duration	OSeven app (i.e., start/end trip time) for cars, CardioID Gateway	Recorded per trip. A trip is engine on to engine off (engine off <5 mins = continuation of same 'trip')	Car, truck, bus, rai	Post-trip
Traffic density	Gateway + CAM Roadway scene video base event, road scene video (high traffic volume, low traffic volume)	Event	Car, truck, bus	Post-trip

Table 9: Recorded variables to indicate/measure the direct indicators of coping capacity

Variable relevant to coping capacity	Method of recording	Frequency of recording	Relevant modes	Real-time/ Post-trip modelling
Electrocardiogram (ECG)	CardioWheel	Periodically sampled signal, with a sampling frequency of 1000 Hz	Truck, bus	Real-time/ Post-trip
Inter-Beat Intervals (IBI)	CardioWheel (bus/truck) Wristband (car/rail)	Sequence of events indicating the time interval between successive heartbeats (about one per second)	Car, truck, bus, rail	Real-time/ Post-trip
Sleepiness Detection / Sleepiness	CardioWheel (truck/bus), Wristband (car/rail)	Time window of 2 minutes and 30 seconds	Car, truck, bus, rail	Post-trip/ Available in real-time but in vehicle computation needed for cars/trucks/buses
Time from the start of the trip	CardioID Gateway	Event (indicating driver fatigue level based on the length of time driving)	Car, truck, bus, rail	Real-time/ Post-trip
Hands-On Detection	CardioWheel	Event (indicating the detection of the driver's hands on the steering wheel)	Truck, bus	Real-time/ Post-trip
Driver Change Detection	CardioWheel (truck/bus), CardioID Gateway (Car)	Event (indicating a driver change has been detected from the ECG signal). For cars if engine is switched off the gateway/display will ask for confirmation of driver	Car, truck, bus	Post-trip for cars (available in real-time but in vehicle computation needed for trucks/buses)

Variable relevant to coping capacity	Method of recording	Frequency of recording	Relevant modes	Real-time/ Post-trip modelling
Driver identification	CardioWheel, Cardio Gateway	Event (identity of each driver, as stated at the beginning of the trip on the intervention device)	Car, truck, bus	Post-trip for cars (available in real-time but in vehicle computation needed for trucks/buses)

Table 10: Recorded variables to indicate/measure the mental indicators of coping capacity

Variable relevant to coping capacity	Method of recording	Frequency of recording	Relevant modes	Real-time/ Post-trip modelling
Fatigue Detection	CardioID Gateway	Indicating driver fatigue detection indicators; KSS and time from the start of the trip	Car, truck, bus, rail	Post-trip/ Available in real-time but in vehicle computation needed for cars/trucks/buses
Distraction Detection	CardioWheel, OSeven app	Indicating the detection of the driver's hands on the steering wheel or detecting the duration of mobile phone use	Car, truck, bus	Real-time/ Post-trip

Table 11: Recorded variables to indicate/measure the driver behavioural indicators of coping capacity

Variable relevant to coping capacity	Method of recording	Frequency of recording	Relevant modes	Real-time/ Post-trip modelling
Harsh acceleration	OSeven app for cars, CardioID Gateway	Event (indicating occurrence of harsh acceleration)	Car, truck, bus, rail?	Post-trip/ Real-time only by CardioID gateway
Harsh braking	OSeven app for cars, CardioID Gateway	Event (indicating occurrence of harsh braking)	Car, truck, bus, rail?	Post-trip/ Real-time only by CardioID gateway
Harsh Cornering	CardioID gateway	Event (indicating occurrence of harsh cornering)	Car, truck, bus, rail?	Real-time can be implemented if needed
Vehicle speed and speeding (i.e., start time of speeding (hh:mm), speeding duration (sec), average speed over speed limit (km/h), percentage of driving time above speed limit (%), location of speeding section on the map)	OSeven app, Mobileye, CardioID gateway	Satellite-based geolocation data, about one sample per second, open street maps	Car, truck, bus, tram	Post-trip/ Real-time only by Mobileye
Mobile phone use (i.e., start time of mobile use (hh:mm), mobile use duration (sec), location of mobile use on the map)	OSeven app (an integrated handheld mobile phone app will be created for	Event (with the provision of detecting the duration of mobile phone use)	Car (truck, bus, rail under development)	Real-time/ Post-trip

Variable relevant to coping capacity	Method of recording	Frequency of recording	Relevant modes	Real-time/ Post-trip modelling
	different transport modes)			
Gyroscope IMU	CardioID gateway	Periodically sampled gyroscope inertial signal, with sampling frequency of 119 Hz	Car, truck, bus	Real-time/ Post-trip
Time headway	Mobileye	Satellite-based geolocation data, about one sample per second	Car, truck, bus, tram	Real-time/ Post-trip
Headway level	Mobileye	Recorded only when changing from one level to another	Car, truck, bus, tram?	Real-time/ Post-trip
Speed Limit Indication (SLI)	Mobileye	Event - based	Car, truck, bus	Real-time/ Post-trip
Forbidden Overtaking Sign	Mobileye	Event - based	Car, truck, bus	Real-time/ Post-trip
Turn indicator activation/deactivation	Mobileye	Event - based	Car, truck, bus	Real-time/ Post-trip
Pedestrian Ahead	Mobileye	Event - based	Car, truck, bus, tram	Real-time/ Post-trip
Pedestrian Collision Warning (PCW)	Mobileye	Event - based	Car, truck, bus, tram	Real-time/ Post-trip
Vehicle Ahead Detected	Mobileye	Event - based	Car, truck, bus, tram	Real-time/ Post-trip
Forward Collision Warning (FCW)	Mobileye	Event based	Car, truck, bus, tram?	Real-time/ Post-trip
Urban Forward Collision Warning (UFCW)	Mobileye	Event based	Car, truck, bus, tram?	Real-time/ Post-trip
Lane Departing Warning (left/right)	Mobileye	Event based	Car, truck, bus	Real-time/ Post-trip

It should be noted that data related to operator competence, personality characteristics and the socio-demographic background will be collected via survey questionnaires. In particular, the various relevant indicators which can be defined are:

- Competencies, measured on the basis of metrics on risk assessment, attention regulation, self-appraisal
- Personality, measured on the basis of metrics on adventure-seeking, disinhibition, experience-seeking, boredom susceptibility
- Sociodemographic profile, measured on the basis of age, gender, experience, socio-economic status, nationality, ethnicity, cultural identity

Table 12: Participants' data that could indicate/measure coping capacity

Relevant information	Questionnaires to be used/questions	Method of recording	Relevant modes	Real-time/ Post-trip modelling
Demographic data (age, gender)	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Driving experience (year of attaining driving licence)	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Annual Mileage	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Professional driver	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
ADAS user	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Fatigue	The Fatigue Questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Sleepiness & sleep quality	Epworth Sleepiness Scale Pittsburgh Sleep Quality Index	Questionnaires	Car, truck, bus, rail	Post-trip
Speeding	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Tailgating	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Tendency to Distractions	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Accident involvement at driver's fault (3 previous years)	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip
Traffic offence -violations record (3 previous years)	i-DREAMS participant entry questionnaire	Questionnaires	Car, truck, bus, rail	Post-trip

4 Modelling considerations for analysis

The literature review presented in Section 4.3 of Deliverable 3.2 (Katrakazas et al., 2020) revealed that four modelling approaches could be more suitable for modelling the STZ within the i-DREAMS project. In particular, Dynamic Bayesian Networks (DBNs), Long Short-Term Memory models (LSTMs), Discrete Choice Models (DCMs) and Structural Equation Models (SEMs) were deemed the most appropriate. While this literature review provided a good understanding of the potential modelling candidates in i-DREAMS and the selected models seem plausible, there are still some open issues that need to be considered for model selection. For example, the suggested models may be confronted with additional limitations considering the different types of data being collected in i-DREAMS. In addition, several new limitations have been identified with additional deeper investigations into these models. For example, it is noted that LSTM is not able to incorporate the inter-relationship between variables into real-time predictions (endogeneity) and SEM is not suitable for analysing discrete dependent variables. As a result, and prior to discussing the selected mathematical models, it seems necessary to map these models to the research questions in i-DREAMS.

The mapping of the models to research questions depends on three dimensions for data analysis in i-DREAMS: (1) the purpose of data analysis –being prediction or explanatory analysis, (2) the time element of data analysis –being real-time or post-trip, and (3) the variable type of risk indicators –being discrete or continuous (as it may be necessary to test alternative definitions of risk in addition to the three-level STZ definition; please see the next section for more discussion about this). The mathematical model to be used in i-DREAMS depends on a combination of these three dimensions. This section aims to provide the updated formulation of these models.

When the purpose of data analysis is the prediction of risk (e.g., prediction of the STZ phases), the data should be analysed in real-time. Prediction of risk after the trip has completed may not have an application in i-DREAMS. As thoroughly discussed in the previous sections, machine learning algorithms are proper analytical methods for real-time data analysis. However, the type of machine learning algorithms to be used certainly depends on the type of risk indicators being discrete or continuous. The Long Short-Term Memory (LSTM) deep neural networks and Dynamic Bayesian Network (DBN) models are suitable for prediction of discrete and continuous indicators of risk.

When the purpose of data analysis is explanatory analysis, the data should be analysed after the trip has been completed because the explanatory analysis in i-DREAMS is primarily done for identifying relationships between driving behaviour (at an aggregate level) and risk. As thoroughly discussed in Section 4.3 of deliverable D3.2, statistical models are suitable for explaining the underlying mechanisms of risk and so are proper analytical methods for post-trip data analysis. However, the type of statistical models to be used depends on the type of risk indicators too. Structural Equation Models (SEM) are only suitable for continuous dependent variables i.e., risk indicators. When the dependent variable is discrete, discrete choice models (DCM) are needed.

The type of risk indicator to be used as the dependent variable plays a pivotal role in selecting the type of mathematical model for data analysis. As such, five alternative indicators of risk are first described and a brief description of each algorithm is then presented, followed by an explicit description of the proposed models.

4.1 Alternative indicators of risk in i-DREAMS

The STZ with three discrete phases (*level 1*: safe driving phase, *level 2*: danger phase, and *level 3*: avoidable crash phase) has been defined as an indicator of risk in i-DREAMS. While this definition of risk is useful for triggering and applying real-time interventions, it may not present a complete picture of risk. A driver could theoretically fall within an overall phase of the STZ although his/her driving performance indicators may fall within a different phase of the STZ. For example, a driver who has been driving for less than two hours may engage in tailgating behaviour with a time headway of 1.6 seconds. According to the defined thresholds (please refer to the synthesis of risk factors for interventions in i-DREAMS in the previous section), the fatigue state of this driver falls within the “normal” phase of the STZ and yet his/her time headway falls within the “danger” phase of the STZ. While the current definition of risk in i-DREAMS is suitable for triggering the real-time warning (headway warning in this example), it is not able to determine how likely it is that the driver will be involved in a risky event after all. This limitation is a direct consequence of not defining the overall state of the STZ for the driver. Other alternative definitions of risk are thus necessary to address this limitation in i-DREAMS.

The STZ could be defined as an overall composite risk variable (possibly a weighted sum of all STZ levels for different risk factors) (Hermans et al., 2008; Gitelman et al., 2010) in addition to a single risk variable for each risk factor. This composite variable can then be used to shed more light on the overall state of driving.

Risk can also be defined based on a more generic definition of the probability of occurrence of rare events (Songchitruksa and Tarko, 2006), and in this context, it is often considered as a product of intensity and duration of risk exposure. This is the case in epidemiology, for example, the analogy of COVID-19 risk modelling, where the risk of infection is estimated as a product of the intensity of exposure (e.g., keeping the 1.5 meter distance) and the duration of exposure (e.g., less than a certain duration), aiming to minimise the probability of such a rare event (Liu et al., 2020). In i-DREAMS, these rare events could be near-misses which are much more frequent than crashes. Hence, the probability of near-misses (as an indicative dependent variable) could be another indicator of risk which can be obtained as a function of indicators of task complexity/coping capacity (intensity) and the time that is spent in each phase of the STZ or the proportion of each phase of STZ during the drive duration (exposure).

It is worth mentioning that this definition of risk has an elegant validating application in i-DREAMS. The STZ is hypothesised a prior and thus the notion of defining thresholds to separate three phases of the STZ in each risk factor requires validation (one may argue that the STZ could have additional phases). Yet, the STZ is latent by nature and so it is necessary to validate the presumed STZ levels with an observable indicator of risk – preferably a surrogate safety measure such as a near-miss. For example, a high likelihood of a near-miss despite driving in a safe STZ level in a particular risk factor may be indicative of the need to revisit the STZ and its corresponding thresholds for that risk factor.

In addition, and borrowing from the above approach, the proportion of each phase of the STZ during the drive could itself be an indicator of risk too. Such a continuous indicator of risk may be highly useful for post-trip explanatory data analysis and can link risk with driving behaviours.

The time that is spent in each phase of the STZ can also be a useful indicator of risk for real-time prediction purposes. For example, the time that is spent in each level of the KSS score may be considered in triggering the real-time fatigue/sleepiness intervention because a driver who has remained in a high KSS score for a prolonged period of time should be warned more severely than a driver who has had a high KSS score for a short period of time. The same rationale may be used for speeding warnings. While the current real-time intervention strategies are based on the previous knowledge from the literature and expert judgment and do not use such information (mainly because we need to first trigger interventions to be able to obtain data), the time that is spent in each level of the STZ can be used for adjusting the duration/frequency/pitch of warnings (for certain risk factors such as sleepiness and speeding) later in the project. For example, the collected data (from simulator experiments or field trials) can be used to predict if a driver will be in a KSS score of 7 ("danger" phase) for a prolonged period of time and if yes, then the initial presumed frequency/shape/pitch of sleepiness/fatigue warning may be adjusted.

Overall and depending on the definition of risk, its indicators may be discrete or continuous. These combinations of risk definitions and indicators are shown in Table 11.

Table 13: Definition of risk and the corresponding indicator variables

Definitions of risk	
Discrete variable	**Continuous variable**
Alternative 1: Safety tolerance zone (1: normal phase, 2: danger phase, 3: avoidable crash phase)	**Alternative 2:** Composite STZ (weighted sum of discrete levels of STZ of all risk factors)
Alternative 3: Rare event (0: no event, 1: near-miss)	**Alternative 4:** Proportion of each safety tolerance zone during the drive for each risk factor
	Alternative 5: Time spent in each safety tolerance zone for each risk factor

4.2 Modelling updates

Considering risk as a dependent variable in i-DREAMS, the type of mathematical model to be used for data analysis highly depends on the definition of risk adopted in each case. Therefore, further updates about the four selected mathematical models (DBN, LSTM, DCM, and SEM) are presented in the following with respect to the definition of risk adopted in each case (Table 13). A schematic overview of modelling approaches to be considered for the analysis is given in Figure 3.

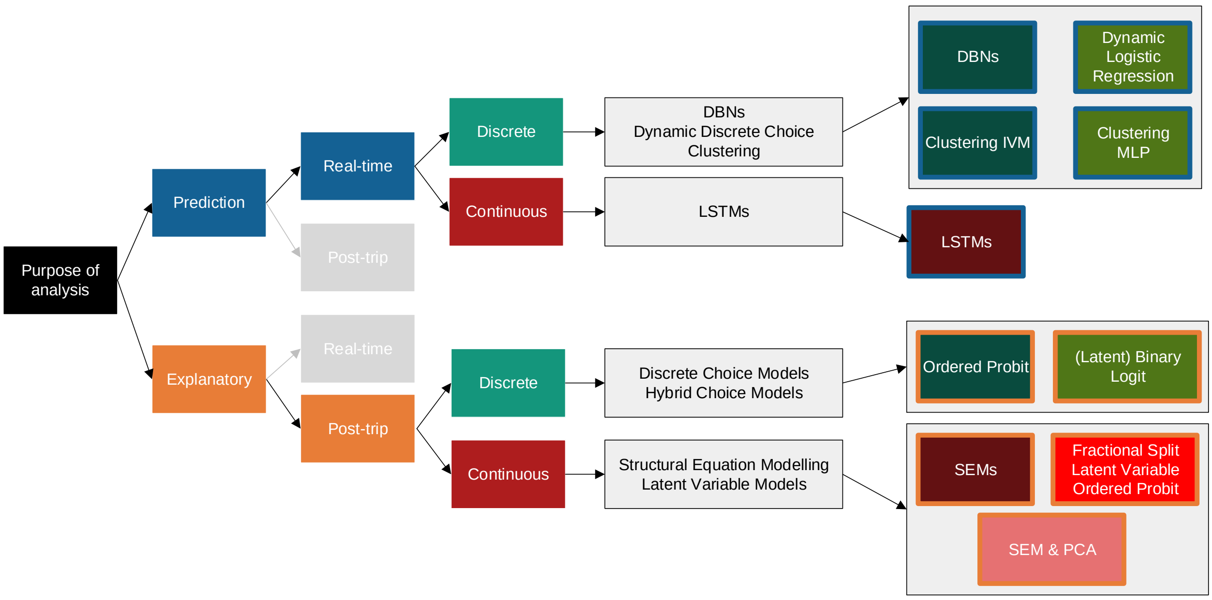

Figure 3: Schematic overview of modelling approaches considered for the analysis of risk factors

4.2.1 Mathematical models for real-time data analysis

4.2.1.1 Dynamic Bayesian Network

Hypothesis

The first hypothesis that should be checked refers to the STZ levels as the difference between task complexity and coping capacity. To begin with, it is hypothesised that a situation is risky if the level of task complexity is different from the level of coping capacity. For example, if the driving is task is difficult and the operator state is decreased, then risk is probable. In order to identify risk, the level of task complexity as well as the level of coping capacity need to be predicted and compared. As a result, the hypothesis forms a real-time multi-level classification problem, where the dependent variable takes the form of a category representing the difference of task complexity and coping capacity. Task Complexity variables (X_1) and coping capacity variables (X_2) can be used to identify individual levels of coping capacity and task complexity, and can also be supplemented by other indicators to predict Y. The relationship between the variables and their causal relationship can be depicted in the following flowchart in Figure 4:

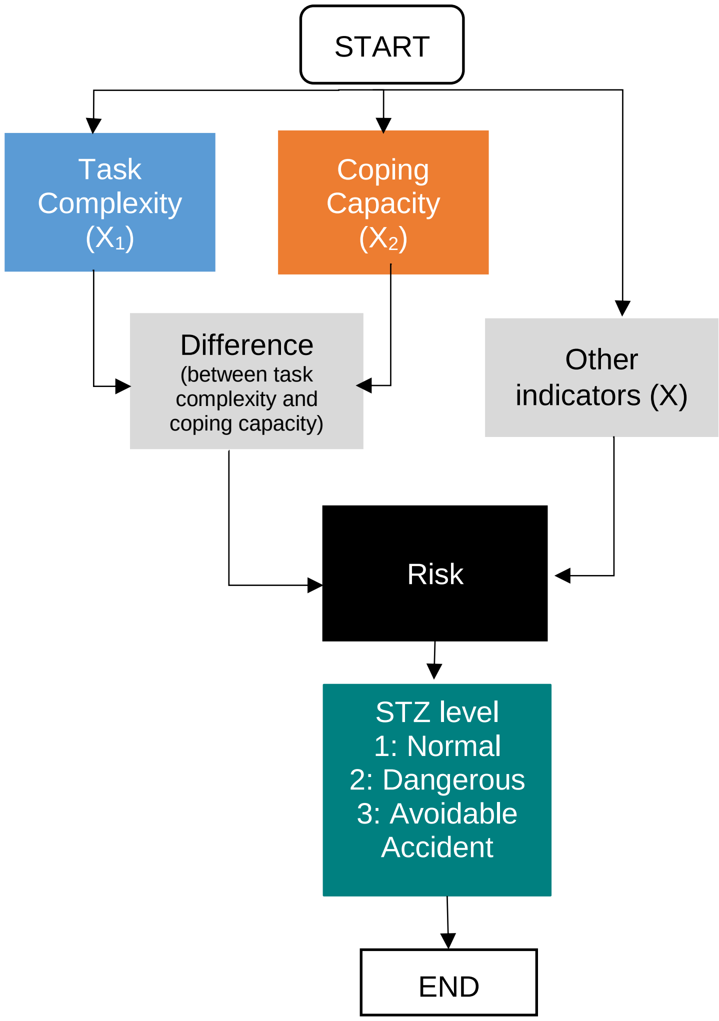

Figure 4: Flowchart associated with the first hypothesis

Independent Variable Screening

Independent variables can be included in the analysis in four categories:

(i) Observed indicators of task complexity: Discrete variables (time of day, wipers on/off, low visibility indicator, road environment, road geometric configuration, Traffic density), Continuous variables (trip duration, start/end trip time),

(ii) Observed indicators of coping capacity: Discrete variables (Hands-on detection, driver change detection, driver identification), Continuous variables (Electrocardiogram (ECG), Inter-Beat Intervals (IBI), sleepiness detection, time from the start of the trip, harsh acceleration, harsh braking, harsh cornering, vehicle speed and speeding, mobile phone use, gyroscope IMU, magnetometer IMU, time headway, headway level, speed limit indication (SLI), forbidden overtaking sign, turn indicator activation/deactivation, pedestrian ahead, pedestrian collision warning (PCW), vehicle ahead detected, forward collision warning (FCW), urban forward collision warning (UFCW), lane departing warning (left/right))

(iii) Latent variables of task complexity: Discrete variables (Environment complexity, road complexity, traffic / time complexity), Continuous variables (Duration / traffic complexity)

(iv) Latent variables of coping capacity: Continuous variables (Mental capacity, driver behaviour risk)

Unit of Analysis

Individual specific analysis (for each driver). Data aggregation for real-time applications 30-second data/ 1-minute data

Model Specification

The raw sensor measurements will be observed. By filtering these raw measurements, the Context-Operator-Vehicle (COV) indicators will become available, so they will be used to determine the coping capacity and task complexity at each time moment. Hence, the two layers of coping capacity and task complexity depend on the COV indicators. Finally, as the operator's capacity indicates the ability of the driver to operate safely with regards to the task imposed, the operator's capacity depends on the complexity of the task. The proposed DBN structure along with the proposed characteristics to estimate task complexity and coping capacity is depicted in Figure 5.

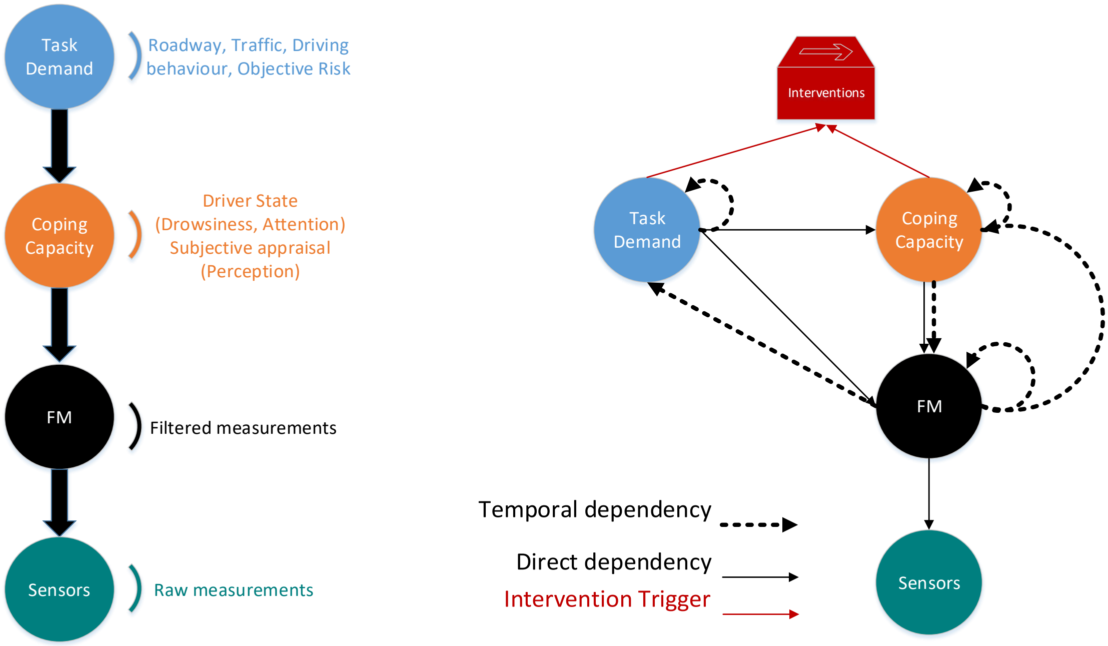

Figure 5: The proposed DBN for STZ modelling

The proposed DBN can be described by the joint distribution:

$$P(TC^{0:T}, CC^{0:T}, FM^{0:T}, Z^{0:T})$$
$$= P(TC_0, CC_0, FM_0, Z_0) \prod_{t=1}^{T} P(TC_t|TC_{t-1}FM_{t-1})\, P(CC_t|TC_tCC_{t-1}FM_{t-1})\, P(FM_t|FM_{t-1}TC_tCC_tCC_{t-1}) P(Z_t|FM_t)$$
$$t \in \mathbb{N} \; and \; t \leq T \qquad (1)$$

where:

- TC: Task Complexity
- CC: Coping Capacity
- FM: Filtered COV Measurements
- Z: Raw measurements
- t: current time step

- T: Total time of measurements

Parametric forms

Task Complexity: The expected task complexity $P(TC_t|TC_{t-1}FM_{t-1})$ is derived from the previous task complexity and the available indicators on environment variables (i.e., time of day, wipers on/off, low visibility indicator, road environment, road geometric configuration and traffic density).

$P(TC_t|TC_{t-1}FM_{t-1}) = f(Environment, Vehicle\ variables, TC_{t-1})$ (2)

Coping Capacity: Coping capacity $P(CC_t|TC_tCC_{t-1}FM_{t-1})$ can be estimated through functions that correlate the effect of task complexity on coping capacity (Faure et al., 2016) modified by a factor to take the previous coping capacity into account.

$P(CC_t|TC_tCC_{t-1}FM_{t-1})= f(Driver,\ TC_t, CC_{t-1})$ (3)

Filtered Measurements: $P(FM_t|FM_{t-1}TC_tCC_tCC_{t-1})$ is the probability of the indicator values based on the current task complexity and coping capacity, as well as their previous values and the previous coping capacity, can be mapped based on the specific scenarios that will be tested in the simulators. In that way, specific ranges of values or task complexity - and coping capacity-specific measurements along with their corresponding probabilities of appearance can be identified.

Raw measurements: For the probability of the raw measurements $P(Z_t|FM_t)$ a sensor model based on Agamennoni et al. (2011), and the Student t-distribution can be followed.

In order to identify the different STZ levels, a comparison between the layers of task complexity and coping capacity will be made. The following probability is proposed to be inferred in order to identify avoidable accident or dangerous STZ levels. It should be mentioned that this probability refers to situations that task complexity and coping capacity are beyond normal operations (i.e., increased or high task complexity with decreased or low coping capacity) given the available sensor observations.

P(TC ≠normal ∪ CC ≠normal |Sensors) (4)

Examples of the different STZ levels according to task complexity and coping capacity are highlighted in Table 12. It can be observed that low coping capacity leads to Avoidable Accident or Dangerous phase, decreased coping capacity leads to Dangerous or Normal phase, while high coping capacity leads to Normal phase, regardless the other layers of task complexity.

Table 14: Different STZ levels according to task complexity and coping capacity

Task Complexity	Coping Capacity	STZ Level
High	Low	Avoidable Accident
High	Decreased	Dangerous
High	High	Normal
Increased	Low	Avoidable Accident
Increased	Decreased	Dangerous
Increased	High	Normal
Low	Low	Dangerous
Low	Decreased	Normal

Task Complexity	Coping Capacity	STZ Level
Low	High	Normal

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

$$likelihoodi = \pi(xi)^{yi}(1-\pi(xi))^{(1-\,yi)} \qquad (5)$$

where:

- xi is the covariate vector
- π(xi) is the probability of the event for the i^{th} subject which has covariate vector xi
- yi is the multiple dependent variable representing the risk probability which has the outcomes y=0 (STZ: Normal Phase), y=1 (STZ: Dangerous Phase) and y=2 (STZ: Avoidable Accident Phase)

The logistic regression equation is:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \qquad (6)$$

where:

- β_0 is the intercept
- βi is a coefficient for the explanatory variable xi

In addition, similarly to the frequentist approach, taking the exp(β) provides the odds ratio for one unit change of that parameter.

4.2.1.2 Long Short-Term Memory (LSTM)

Hypothesis

The second research hypothesis which should be checked refers to the time spent in STZ as a real-time regression/forecasting problem). For the second hypothesis, the following processing will be applied. We have a specific risk factor (i.e., task complexity or coping capacity) along with the corresponding measurements and metrics for each variable. At each time, we target a specific risk factor (i.e., STZ levels of each risk factor are known) but we can also use other important variables (e.g., weather conditions, distraction, etc.) in the same model in order to make the prediction. The entire dataset will be split into train and test set. Based on these indicators, we need to predict the risk, and therefore, the time spent in each STZ level (i.e., Normal, Dangerous, Avoidable Accident). The problem is a real-time regression problem and can be solved by the LSTM formulation. In order to make sure that the risk calculated is reliable, we should perform a good level of forecast accuracy for all the STZ levels. For instance, if we can produce a good prediction for the "Avoidable Accident" phase, it should be made clear that we can produce a good prediction for the "Normal" phase, as well. This implies that the level of the STZ should be known beforehand, otherwise this hypothesis needs to be supplemented by a classification problem or a clustering one. The flowchart associated with this hypothesis is shown in Figure 6.

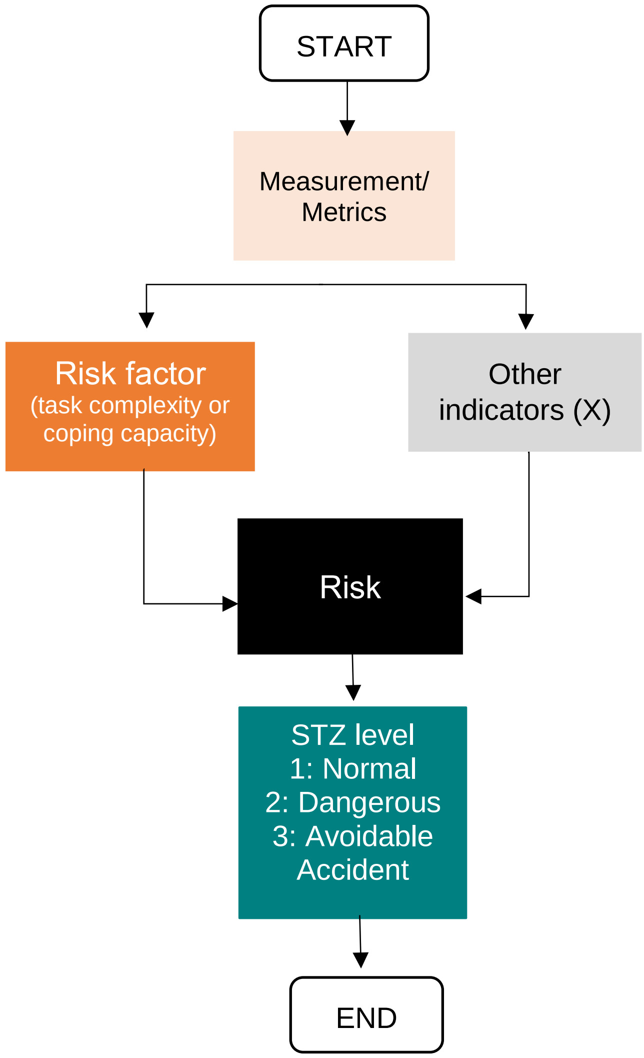

Figure 6: Flowchart associated with the second hypothesis

Independent Variable Screening

Independent variables can be included in the analysis in four categories:

(i) Observed indicators of task complexity: Discrete variables (time of day, wipers on/off, low visibility indicator, road environment, road geometric configuration, Traffic density), Continuous variables (trip duration, start/end trip time),

(ii) Observed indicators of coping capacity: Discrete variables (Hands-on detection, driver change detection, driver identification), Continuous variables (Electrocardiogram (ECG), Inter-Beat Intervals (IBI), sleepiness detection, time from the start of the trip, harsh acceleration, harsh braking, harsh cornering, vehicle speed and speeding, mobile phone use, gyroscope IMU, magnetometer IMU, time headway, headway level, speed limit indication (SLI), forbidden overtaking sign, turn indicator activation/deactivation, pedestrian ahead, pedestrian collision warning (PCW), vehicle ahead detected, forward collision warning (FCW), urban forward collision warning (UFCW), lane departing warning (left/right))

(iii) Latent variables of task complexity: Discrete variables (Environment complexity, road complexity, traffic / time complexity), Continuous variables (Duration / traffic complexity)

(iv) Latent variables of coping capacity: Continuous variables (Mental capacity, driver behaviour risk)

Unit of Analysis

Individual specific analysis (for each driver). Data aggregation for real-time applications 30-second data/ 1-minute data

Model Specification

With regards to the second proposed LSTM model, the problem of defining the STZ levels becomes more straightforward, since LSTMs as a sub-category of Deep Neural Networks act like "black-boxes" (Xu et al., 2013) and thus the only input that needs to be provided to the model are labelled time series data. The proposed approach using LSTMs is given in Figure 7.

It should be mentioned that at the current time, there is no information about abnormal driving situations and identification. If abnormal driving is detected, then the influence of abnormal driving could be added into coping capacity so that it is included in STZ calculation. Collected historical measurements from the i-DREAMS technologies can be used as input for an unsupervised learning approach grouping together measurements correlated with normal operation of a vehicle and those departing from normal driving behaviour. The detection of abnormal driving may thereby become a valuable input to the STZ LSTM model.

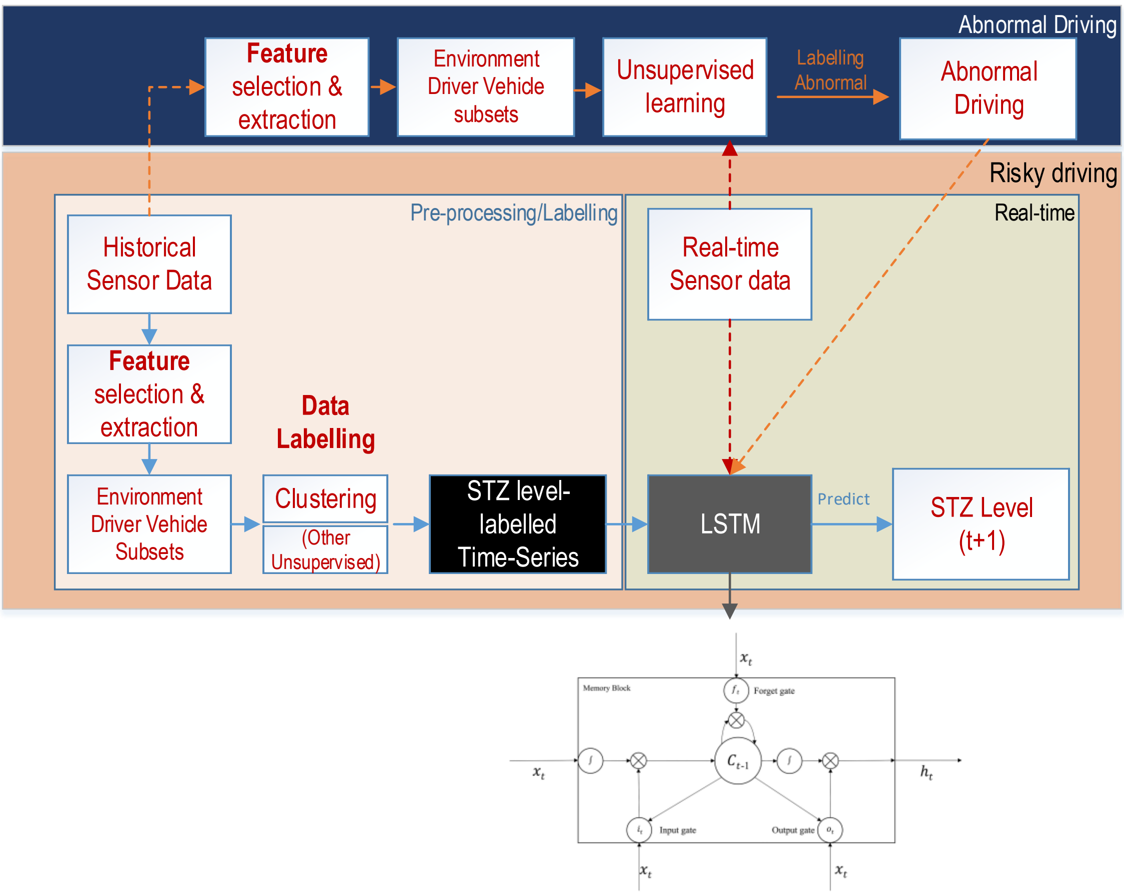

Figure 7: STZ modelling using LSTMs

4.2.2 Mathematical models for post-trip data analysis

4.2.2.1 Discrete indicators of risk

Discrete choice models (DCMs) are the most common statistical approach to model discrete indicators of risk (i.e., alternatives 1 and 3 in Table X). These models rely on the maximum utilisation theory in economics (Hensher et al., 2005) stating that among many alternatives, individuals select the alternative (i.e., discrete category) that maximises their utility. Thus, the first step in formulating DCMs is defining a utility for each discrete alternative. This utility will not have a physical meaning but is rather an auxiliary term to determine the probability of selecting an alternative over the other alternatives. Depending on the nature of the discrete variable being nominal (i.e., rare event/no rare event) or ordered (i.e., STZ levels), DCMs can take the form of either unordered or ordered.

Unordered Discrete Choice Models

Let Y be a discrete dependent variable with s nominal categories (e.g., s=0: no rare event, s=1: rare event). The utility of the s^{th} category (U_s) is stated as:

$$U_s = \beta_s X_s + \varepsilon_s$$

where β_s are estimable parameters (including the intercept), X_s are explanatory variables (e.g., sociodemographic factors, vehicle type, etc.) and ε_s is the random error term assumed to be identically and independently distributed across observations and describing the random part of the utility. Assuming that ε_s is generalised extreme value distributed (McFadden, 1981), the probability of the s^{th} category can be presented as:

$$P(\mathrm{Y} = s\ ) = \frac{e^{(\beta_S X_S)}}{\sum_{j=1}^{S} e^{(\beta_j X_j)}}$$

The likelihood of occurring the s^{th} category across all individuals can then be determined by the product of the above equation over the entire observations. This model is referred to as the *multinomial logit discrete choice model* in the statistical and econometrics literature (Hensher et al., 2005).

When the dependent variable has only two categories (s=2), the above model reduces to the binary logit model. This model can be used to determine the probability of a rare event (e.g., a near-miss) in Table X. Additional variants of this model such as *latent variable binary logit model* may also be useful depending on the hypothesis between risk, task complexity and coping capacity. .

Ordered Discrete Choice Models

Let Y be a discrete dependent variable with s ordered categories (e.g., $S = 1$ if normal driving, $S = 2$ if danger phase, and $S = 3$ if avoidable accident phase). In ordered discrete choice models, the actual category of the dependent variable (Y_s) is associated with an underlying latent variable ($Y_s{}^*$). This latent variable is then mapped to the actual categories by thresholds (τ) and using the following linear function:

$$Y_s^* = \kappa X_s + \delta_i \quad \text{and} \quad Y_s = S \quad \text{if} \quad \tau_{s-1} < Y_s^* < \tau_s$$

where κ is the vector of parameters, X_s is the vector of covariates for the s^{th} category and δ_i is the random error term. To estimate the latent propensity of the dependent variable, it is assumed that:

$$E(Y_s|X_s) = H_s(.),\ 0 \leq H_s(.) \leq 1,\ \ \sum_{s=1}^{S} H_s = 1$$

where $H_s(.)$ is the probability density function for the discrete category *s*. Depending on the distributional assumption for the probability of error terms, $H_s(.)$ can take standard normal or standard logistic probability density functions for the *ordered probit* or *ordered logit discrete choice models*, respectively. Maximum likelihood estimation approach is used to estimate this log-likelihood function.

This model can be used to determine the probabilities of each level of the STZ in Table X. Additional variants of this model such as *latent variable ordered probit (logit) model* may also be useful depending on the hypothesis between risk, task complexity and coping capacity .

4.2.2.2 Continuous indicators of risk

Structural Equation Models (SEM) are suitable for continuous indicators of risk (i.e., alternatives 2, 4 and 5 in Table X). These models are estimated using ordinary least squares (OLS) approach. Let Y_i be a continuous indicator of risk. A structural equation modelling approach is used to correlate this dependent variable to the independent variables. The SEM consists of two components: a structural equation and measurement equations. The structural equation is a regression model capturing the relationship between variables:

$$Y_i = \beta X_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

where β_i are estimable parameters (including the intercept), X_i are explanatory variables (e.g., demographics, coping capacity and task complexity) and ε_i is the random error term assumed to be normally distributed across observations and describing the random part of the structural equation.

The measurement equations, on the other hand, are concerned with how well various measured exogenous indicators measure latent variables. In other words, and in estimating the above structural equation, the latent variables (e.g., latent risk, latent task complexity, latent coping capacity) can be measured (i.e., measurement equation) using a linear additive combination of certain observed indicators. However, many of these indicators often have high autocorrelation with one another. To address this problem, the Principal Component Analysis (PCA) can be used to summarise the observed indicators into orthogonal variables (i.e., principal components) that are not correlated. The PCA creates a set of new variables, referred to as *principal components (PC)*, each of which is a linear and orthogonal combination of the original variables in such a way that each orthogonal combination captures the maximum variability in the original set of variables and has the minimum autocorrelation with other linear combinations. The principal components are then arranged based on their decreasing contribution to the total variance of the original set of explanatory variables: the first principal component explains the highest variability in the explanatory variables; the second principal component explains the second-highest variability in the explanatory variables, and so forth (the cumulative contribution of all principal components is equal to 1). These principal components can then be used in the analysis as indicators of the original latent variables. The number of principal components to be used in the model depends on the specific research objective, though the common practice is to use all principal components with Eigenvalues greater than one (Tipping and Bishop, 1999).

Assuming that ε_i is normally distributed, the structural equation can be estimated using generalised least squares or maximum likelihood estimation approaches.

4.3 STZ labelling and preparation of training datasets

In the previous sections, the modelling approaches presented, treat risk and STZ as the dependent variable in a supervised classification or regression setup. In order, however, to perform such a modelling analysis, STZ labels should be known a priori for the models, so that each analysis technique, with its strengths and limitatations, can efficiently predict the level of risk at each time moment and especially for the real-time intervention strategy.

In order to label data for representing STZ levels, the approach that the project could follow is the following:

1. Obtain pre-defined thresholds for defining normal, dangerous and avoidable accident phases, as these are described in section 2.4 and the warning strategies. For example, the speeding strategy takes into account different contributing factors according to the coping capacity and task complexity measurements available and the fatigue warning adjusts the thresholds according to driving duration and the age and gender of the driver.
2. Use these a-priori thresholds on gathered data with regards to headway, illegal overtaking, speeding and fatigue and label the gathered data. For example, if data on overtaking restriction, speed, acceleration, turn signal, lane crossing and steering are collected, and the corresponding speed and acceleration are above 35km/h with acceleration greater than $0.2 m/s^2$, the data for the data collection period will be marked as “dangerous” for the dependent STZ of illegal overtaking.
3. If the thresholds do not provide three levels of the STZ, then data could be normalized and values within the 90% C.I. of the dependent variable could represent normal driving, with 95% C.I. being the dangerous phase and all other values being the avoidable accident phase.
4. As also mentioned in the LSTM approach description (i.e. section 4.2.1.2), unsupervised learning and clustering could be used to automatically label the three phases for each risk factor. In that way only independent variables (i.e. task complexity and coping capacity measurements) will be fed into Principal Component Analysis (Mahahan et al, 2020), k-means (Yang et al, 2021) or t-SNE (Yang et al, 2021), in order to obtain three clusters, representing the three levels of STZ for each risk factor. The clustering validation can then be performed by using the silhouette coefficient (Yap and Cats, 2021) or the Dunn index (Nawrin and Rahman, 2017).

Following the four aforementioned steps, datasets gathered from both the simulation and pilot on-road trials can be labelled with regards to the STZ level and can be used for training the algorithms for real-time interventions. Further validations of the clusters obtained can also be achieved in the post-trip analysis phase.

4.4 Evaluating intervention effectiveness

Following the design of the assessment methodology, the most important step to assess the effectiveness of safety interventions is the organisation of the back-office database, which will provide all necessary data for the realisation of the individual evaluations. The back-office database will also assist in performing comparisons among countries and different transport modes (i.e., cars, trucks, buses, and rails), which subsequently will enhance the intervention performance evaluation and the quality of the assessment results. Since the naturalistic driving experiments have not started yet for all countries, the crucial aspect is the collaboration as well as the interaction among partners from each country who are going to access and analyse the data with the backend database. During the four different stages of the evaluation of safety interventions (especially, in the second stage: real-time interventions, in the third stage: real-time and post-trip and in the fourth stage: both real-time interventions and gamification elements of post trip interventions), good coordination is required to collect the data needed and build a representative sample.

Qualitative data will also be collected using questionnaires at the end of the trials. This will inform where issues arose with equipment and the study design in general. The participants will have been exposed to the equipment for a long time, so this feedback is expected to be detailed, varied and useful.

5 Conclusions

This deliverable aimed to update selected sections of D3.2 to reflect subsequent project developments.

The variables proposed in D3.2 were confirmed with those available using the iDREAMS platform as developed by WP4. This resulted in a list of variables that can be measured and for which mode that can be used to calculate STZ phases.

The real-time warning strategies for the four performance objectives (Headway, Illegal overtaking, Speeding, Fatigue) that can be assigned variable thresholds were defined and threshold ranges were assigned to each STZ phase. The deliverable went on to explain that for each of these four strategies additional variables are used as indicators and/or modifiers and the types of real-time warnings were outlined.

The deliverable also discussed driving style, in terms of ‘normal’ (STZ normal phase) and ‘abnormal’ (STZ danger and avoidable accident phase) and that it is necessary to account for the possibility of the driver being in a ‘normal’ driving style for one performance indicator and an ‘abnormal’ driving style for another.

A key aspect of defining the STZ, is measuring task complexity and (driver) coping capacity with safe driving defined as when these two dimensions are in balance. The deliverable therefore also defines which variables are associated with each of these, the method and frequency of recording, which mode are applicable and whether real-time or post-trip modelling methodologies are required for analysis.

Alternative definitions of risk were discussed and described that relate to the STZ phases or the detection of an ‘event’ (discrete variables). In addition, ways in which the overall risk during a period of time were defined e.g. a composite STZ value or proportion of time spent in a STZ phase (continuous variables)

Finally detailed descriptions of the relevant mathematical models (Dynamic Bayesian Network, Long Short-Term Memory, Discrete Choice Models, and Structural Equation Models) were provided with an explanation as to when they could be used for analysis. This depends on the variable type (discrete, continuous) and when the associated values are calculated (real-time or post-trip) were provided. For each model, the relevant independent variables or risk definitions that can feed into the model were defined and the relevant equations/functions were defined.

5.1 Next Steps

The information included in this deliverable is being utilised in two main ways:

The information in this deliverable will be utilised is in the analysis work packages, WP6 (Analysis of risk factors) and WP7 (Evaluation of safety interventions). The models described will be utilised and further developed during the analysis, for example to identify which factors most influence task complexity, coping capacity and how the STZ phases relate to this as well as evaluating how effective the real-time and post trip interventions were on behaviour change.

The definitions of values and variables have already been included in the various algorithms that have been implemented by WP4 in the i-DREAMS platform. Over the next six months, on road field trials will be conducted for the passenger car, bus and truck mode and simulator trials for the rail mode. The on road trial is split into two phases which will allow small alterations to be made to the i-DREAMS platform at the end of phase one if necessary. Any learning from the simulator and field trials or changes to the platform that relate to this deliverable will be documented in the WP7 and WP6 deliverables that will be published at the end of the project.

6 References

Brijs, K., Brijs, T., Ross, V., Donders, E., Yves Vanrompay, Y., Geert Wets, G., . . . Gaspar, C. (2020). Toolbox of recommended interventions to assist drivers in maintaining a safety tolerance zone. Deliverable 3.3 of the EC H2020 project i-DREAMS.

Campagne, A, Pebayle, T, Muzet, A, 2004 A Correlation between driving errors and vigilance level: influence of the driver's age. Physiology & Behavior, 80, pp. 515-524

Filtness, A. J., Reyner, L. A. and Horne, J. A. 2012 "Driver sleepiness—Comparisons between young and older men during a monotonous afternoon simulated drive," Biol. Psychol, vol. 89, no. 3, p. 580-583,.

Gitelman, V., Doveh, E. and Hakkert, S., 2010. Designing a composite indicator for road safety. Safety Science, 48(9), pp.1212-1224.

Hensher, D. A., Rose, J. M., Rose, J. M. and Greene, W. H. 2005. Applied choice analysis: a primer, Cambridge university press.

Hermans, E., Van den Bossche, F. and Wets, G., 2008. Combining road safety information in a performance index. Accident Analysis & Prevention, 40(4), pp.1337-1344.

Katrakazas, C., Michelaraki, E., Yannis, G., Cuenen, A., Brijs, K., Brijs, T., Fitness, A., . . . Gruden, C. (2020). Methodology for the evaluation of interventions. Deliverable 7.1 of the EC H2020 project i-DREAMS.

Katrakazas, C., Michelaraki, E., Yannis, G., Kaiser, S., Senitschnig, N., Ross, V., . . . Taveira, R. (2020). Toolbox of recommended data collection tools and monitoring methods and a conceptual definition of the safety tolerance zone. Deliverable 3.2 of the EC H2020 project i-DREAMS.

Liu, P., Beeler, P. and Chakrabarty, R.K., 2020. Dynamic interplay between social distancing duration and intensity in reducing COVID-19 US hospitalizations: A "law of diminishing returns. Chaos: An Interdisciplinary Journal of Nonlinear Science, 30(7), p.071102.

Mahajan, V., Katrakazas, C., & Antoniou, C., 2020. Prediction of Lane-Changing Maneuvers with Automatic Labeling and Deep Learning. Transportation Research Record, 2674(7), 336–347. https://doi.org/10.1177/0361198120922210

McFadden, D. 1981. Econometric models of probabilistic choice. In: Manski, D., McFadden (Eds.), A Structural Analysis of Discrete Data with Econometric Applications. The MIT Press, Cambridge, MA.

Nawrin, S, Rahman, M. R, Akhter, S. 2017. Exploring K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System. International Journal of Advanced Computer Science and Applications, 8(3), 264-272.

Pilkington-Cheney, F., Pooyan Afghari, A., Filtness, A., Papadimitriou, E., Lourenco, A., Brijs, T. (2021). The i-DREAMS intervention strategies to reduce driver fatigue and sleepiness for different transport modes. Proceedings of the 7th International IEEE Conference on Models and Technologies for Intelligent Transportation Systems Institute of Electrical and Electronic Engineers TBC

Songchitruksa, P. and Tarko, A.P., 2006. The extreme value theory approach to safety estimation. Accident Analysis & Prevention, 38(4), pp.811-822.

Tipping, M.E., and Bishop, C.M., 1999. Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3), pp.611-622.

Yang K., Al Haddad C., Yannis G., Antoniou C., 2021, Driving Behavior Safety Levels: Classification and Evaluation, Proceedings of the 7th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (16-17 June 2021)

Yang, K, Rakibul, A; Al Haddad, C; Ezzati Amini, R; Antoniou, C. (2020). An algorithm for safety tolerance zone calculation. Deliverable 4.2 of the EC H2020 project i-DREAMS.

Yap, M., Cats, O. Predicting disruptions and their passenger delay impacts for public transport stops. Transportation 48, 1703–1731, 2021. https://doi.org/10.1007/s11116-020-10109