ReliablE in-Vehicle pErception and decisioN-making in complex environmenTal conditionS

Grant Agreement Number: 101069614

D5.3: System SOTIF compliance test report

Document Identification			
Status	Final	**Due Date**	30/06/2025
Version	1.0	**Submission Date**	07/07/2025
Related WP	WP5	**Document Reference**	D5.3
Related Deliverable(s)	D2.3	**Dissemination Level**	PU
Lead Participant	APTIV	**Document Type:**	OTHER
Contributors	ICCS	**Lead Author**	Yogesh Ganesh (APTIV)
		Reviewers	Leonardo González (TECN)
			Anastasia Bolovinou (ICCS)

Funded by the European Union

This project has received funding under grant agreement No 101069614. It is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Document Information

Author(s) – in alphabetical order		
First Name	**Last Name**	**Partner**
Kirsty	Aquilina	APTIV-DE
Kien Cuong	Nguyen	APTIV-FR
Krystian	Sycz	APTIV-PO
Bill	Roungas	ICCS

Document History			
Version	**Date**	**Modified by**	**Modification reason**
0.1	17/04/2025	APTIV	First overall Deliverable Draft showing APTIV's preliminary contributions – some sections still have some parts missing.
0.2	15/05/2025	APTIV	Second overall Deliverable Draft showing APTIV's preliminary contributions – most of the sections are complete but some still have some parts missing.
0.3	10/06/2025	APTIV	Deliverable Draft for review by ICCS and TECN. Included ICCS contribution in Section 6.1.5
0.4	20/06/2025	ICCS, TECN	Reviews submitted from ICCS and TECN
0.5	27/06/2025	APTIV	Integrated review comments.
1.0	30/06/2025	ICCS	Final version to be submitted

Quality Control		
Role	**Who (Partner short name)**	**Approval Date**
Deliverable leader	Yogesh Ganesh (APTIV)	30/06/2025
Quality manager	Panagiotis Lytrivis (ICCS)	30/06/2025
Project Coordinator	Angelos Amditis (ICCS)	30/06/2025

Executive summary

Developing toward safety of the intended functionality, or SOTIF, entails considering what conditions might cause hazards at the vehicle level, even in the absence of system, hardware, or software malfunctions. These conditions could be limitations and weaknesses on any aspect of an ADS stack – Sensing, perception, planning, control, or HMI. Additionally, they could be conditions of the ODD – Scenery elements, environmental conditions, or dynamic elements. Finally, they could entail how the driver of the ADS interacts with the system.

The motivation for this report is twofold. First, it is to discuss approaches supporting a workflow that can provide evidence of the SOTIF of an ADS. Second, it discusses how to find triggering conditions and uses those conditions to create scenarios, which are then fed into the tests that are run to verify and validate an ADS's handling of its known hazardous space.

The deliverable identifies safety goals presented in D2.3 HARA [1] that are SOTIF-relevant and explains how to create vehicle-level metrics that link to those safety goals. Recent statistics from a number of sources were identified for each SOTIF-relevant safety goal, and vehicle-level metrics were calculated to support a "better than" factor of 10 compared to current vehicle accident rates.

Considering the architecture of this project (D2.2 [2]), triggering conditions at the feature, sensor, algorithm, actuator, and driver levels were considered, and SOTIF relevant scenarios related to those triggering conditions were created.

Subsequently, using the vehicle-level metrics and SOTIF relevant scenarios, the deliverable discusses the derivation of system- and subsystem-level metrics. This is done through two means –1) Exploration of bounds for steering, braking, and accelerating using calculation- and simulation-based analyses, and 2) direct derivation from the vehicle-level metrics using a quantitative fault tree analysis (FTA) method.

For metrics derived from the calculation- and simulation-based analyses, bounds were put in place for steering, braking, and accelerating considering that more challenging conditions, such as wet roads and tight curves in roadway, are in ODD for the ADS. In fact, some of the simulation-based analyses demonstrated a possible need to either hand back over to the driver or restrict speed in low-friction conditions given that the simulation vehicle could not traverse the roadway successfully. Safety requirements are defined using the outcome of this system and subsystem-level analysis.

The quantitative FTA method presented in the deliverable shows how vehicle-level metrics are broken down to module level metrics by integrating SOTIF triggering conditions and scenarios of interest. The findings of the FTA can then be used : 1) as

input to the safety requirements in terms of module metrics, 2) to add more safety requirements if the required vehicle level metric are not met, 3) guide test protocol creation if prioritization is required and 4) to improve KPI definitions or derive KPI goal values of modules to ensure safety is respected.

Finally, a test protocol is presented and is recommended as input to WP6 for verification and validation purposes – For example, the test protocol challenges the system to operate in a variety of the triggering conditions to qualify the perception robustness at identifying an output degradation.

Table of contents

List of tables

List of figures

Abbreviations & Acronyms

Abbreviation / Acronym	Description
AI	Artificial Intelligence
ADS	Automated Driving System
ALARP	As Low as Reasonably Practicable
AV	Automated Vehicle
CAV	Connected and Automated Vehicle
CP	Collective Perception
CAM	Cooperative Awareness Messages
CPM	Collective Perception Messages
DNN	Deep Neural Network
Dx.y	Deliverable x.y in EVENTS project
FN	False Negative
FP	False Positive
FTA	Fault Tree Analysis
GAMAB	“Globalement Au Moins Aussi Bon"
GNSS	Global Navigation Satellite System
EC	European Commission
EXP	Experiment
FDIS	Final Draft International Standard
FOV	Field of View
HARA	Hazard Analysis and Risk Assessment
ISO	International Organization for Standardization
ISE	Input signal error
KPI	Key Performance Indicator
MEM	Minimal Endogenous Mortality
ML	Machine Learning
ODD	Operational Design Domain
OSE	Output signal error

Abbreviation / Acronym	Description
RHWF	Random Hardware Fault
SA	Self-Assessment
SAE	Society of Automotive Engineers
SGx	Safety goal x
SOTA	State-of-the-Art
SOTIF	Safety of the Intended Functionality
SPEC	Specification
STPA	System Theoretic Process Analysis
Tx.y	Task x.y in EVENTS project
TSR	Traffic sign recognition
TTL	Time to Leave (time from when an error is induced to when tire crosses lane marker)
UC	Use Case
V2X	Vehicle-to-Everything
VRU	Vulnerable Road User
WP	Work Package

1. SOTIF for ADS testing introduction

1.1 Standards interplay

Throughout the course of this project, the following standards have guided the partners on the approach for testing toward SOTIF for ADS.

- ISO 21448:2022, “Road vehicles — Safety of the intended functionality”: This standard provides the main guidance and workflow for SOTIF from project initiation through field operation. The focus for the intent of this task is on the portion of the standard that supports development phase activities.
- ISO 26262:2018, “Road vehicles — Functional safety”: This standard provides the workflow for functional safety from project initiation through decommissioning.
- ISO/PAS 8800:2024, “Road Vehicles — Safety and artificial intelligence”: This standard builds on the principles from ISO 21448 and ISO 26262, providing a lifecycle for safety of artificial intelligence (AI) systems and providing guidance at the AI system, AI component, and data levels.
- ISO 34502:2022, “Road vehicles — Test scenarios for automated driving systems — Scenario based safety evaluation framework”: This standard provides guidance on deriving critical scenarios and turning those scenarios into test protocols that can be executed. It also suggests methods for executing those test protocols.
- ISO 34503:2023, “Road Vehicles — Test scenarios for automated driving systems — Specification for operational design domain”: This standard provides a taxonomy for ODD, which has supported in the systematic identification of triggering conditions and scenarios.
- ISO/FDIS 34505, “Road Vehicles — Test scenarios for automated driving systems — Scenario evaluation and test case generation”: This standard is based on the principles from ISO 34502 and ISO 34503, identifying which characteristics should be identified and which results should be tracked for a given test protocol.

1.2 Strategies to define reasonable acceptance criteria and validation targets for vehicle metrics

This report discusses approaches supporting a workflow that can provide evidence of the SOTIF of an ADS. It starts at the HARA (output of D2.3 [1]), explaining how to identify safety goals that are SOTIF-relevant and how to create vehicle-level metrics

that link to those safety goals. The report discusses the derivation of system- and subsystem-level metrics through two means –1) Exploration of bounds for steering, braking, and accelerating using calculation- and simulation-based analyses and 2) direct derivation from the vehicle-level metrics using a quantitative FTA method. Furthermore, it discusses how to find triggering conditions and use those conditions to create scenarios, which are then fed into the tests that are run to verify and validate an ADS's handling of its known hazardous space.

1.3 Importance of this document for WP6

The work presented in this deliverable can support WP6 as follows:

1. Provide a test protocol which verifies the safety requirements derived in this deliverable which can be used where applicable and/or possible in WP6 data collection task. Therefore, The test protocols presented in this section can augment or help define the ones presented in D6.1 where safety analysis is performed on both vehicle and sub-system level.

2. Demonstrate the methodology shown in Section 6 to improve KPI definitions or derive KPI goal values which are aligned with the derived safety goal vehicle metrics (Section 3).

1.4 Document structure

The rest of this report is structured in a manner that closely follows the workflow given in Figure 1:

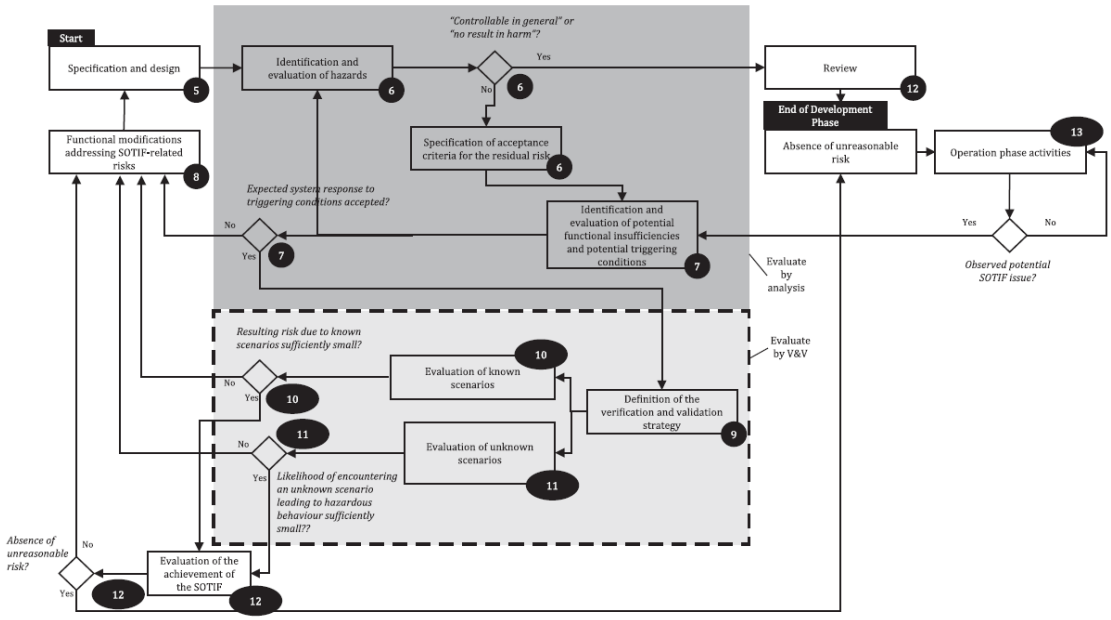

Figure 1: SOTIF workflow (Figure 10 of ISO 21448 [3])

Each section represents an activity or set of activities in this workflow. Each section details the approach taken for the activity, summarizes the output of the approach, and where appropriate, discusses how those outputs link to the activities amongst the partners.

The sections and how they link to the workflow activities are as follows:

Table 1: Mapping of sections to SOTIF workflow

Section	**Workflow activity from ISO 21448**
SOTIF considerations from HARA	Identification and evaluation of hazards
Vehicle-level metrics	Specification of acceptance criteria for the residual risk
Scenario identification	Identification and evaluation of potential functional insufficiencies and potential triggering conditions
Derivations of system- and subsystem-level metrics using calculations and simulations	Identification and evaluation of potential functional insufficiencies and potential triggering conditions
Derivation of subsystem target metric allocations	Functional modifications addressing SOTIF-related risks; Specification and design
Test protocols	Identification and evaluation of potential functional insufficiencies and potential triggering conditions; Definition of the verification and validation strategy; Evaluation of known scenarios

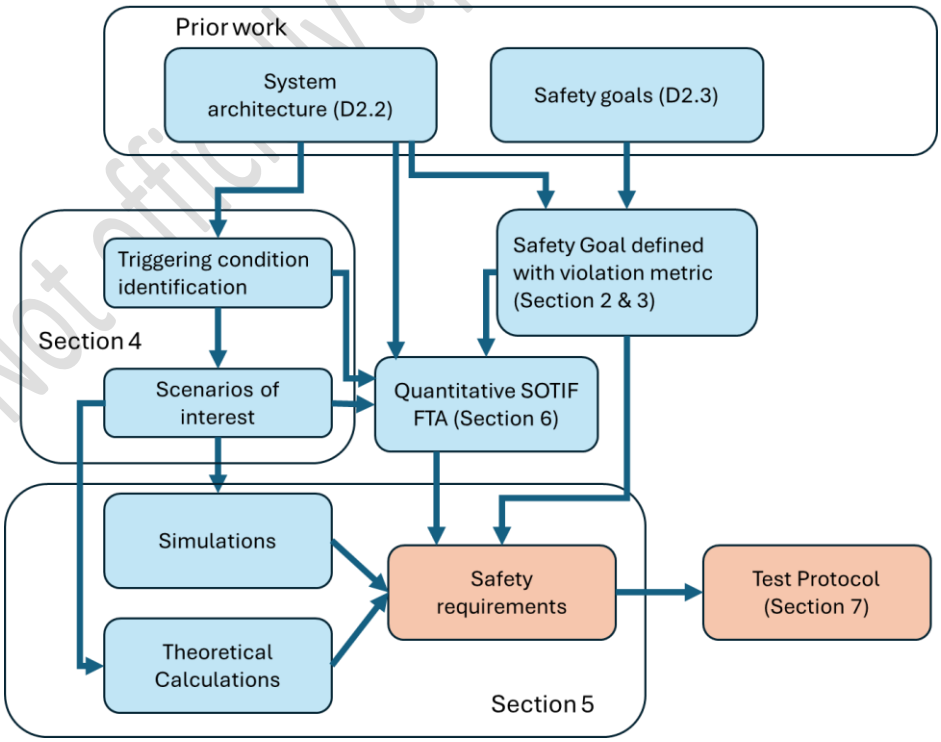

Figure 2: Interaction between sections of this document and previous deliverables.

Figure 2 shows how each section of this document feeds into the next. The orange blocks show the main outcomes of this study. The EVENTS architecture and safety goals derived in previous deliverables (D2.2 [2] and D2.3 [4] respectively) are used as input to this deliverable. Section 2 and Section 3 refine the HARA outputs to focus on SOTIF elements and derive vehicle level acceptance criteria metric assigned to each SOTIF relevant safety goal. Subsequently, Section 4 uses the system architecture to derive triggering conditions and scenarios of interest. These are then used as input to the simulation and theoretical calculation used to derive SOTIF safety requirements. The vehicle level metrics, triggering conditions and scenarios of interest are used as input to the FTA presented in Section 6. The outputs of the FTA can be used to either show that new safety requirements are needed as the vehicle level metric cannot be reached or to quantify safety requirements. Additionally, the FTA can be used also to derive new performance indicators with a direct link to safety or to derive performance indicator goal values; this is not shown in the image above as it is an additional output of FTA not directly linked to the safety requirements. Finally, the derived safety requirements are used to derive a test protocol that can verify them.

It is important to note that this deliverable is limited in scope, such that certain activities were not considered, either in part or in full. These activities include the following, with the second through fifth activities being contingent on the first:

- Evaluation of unknown scenarios: This particular activity is dependent on the availability of a fully integrated system on which the project partners can simulate, resimulate, or drive many miles within the extent of the ODD. The EVENTS project focuses on some important edge cases of interest (called experiments) and not on the entire ODD, making the in-depth evaluation of unknown scenarios out of scope.
- Evaluation of the achievement of the SOTIF
- Review
- Absence of unreasonable risk
- Operation phase activities

2. SOTIF considerations from HARA

2.1 Purpose

One of the first safety activities performed on a system development is the HARA. The objective of the analysis (HARA) is to systematically identify and evaluate the risks from feature usage in different operational situations. Often, however, the HARA is developed from an ISO 26262 [5] lens, in that corner or edge conditions and human factors may not be considered as deeply. As such, a gap analysis is carried out on the initial HARA (refer to [1]) to uncover SOTIF factors that might not have been originally integrated and to properly account for them in subsequent SOTIF concept activities.

2.2 Approach

Two major aspects to be considered in the gap analysis are:

- Has the entire ODD been represented in the operational conditions, the severities, and the exposures?
- Has driver use and potential misuse been represented in the operational conditions, the severities, and the controllabilities?

To complete the gap analysis, one takes the existing HARA and for every use case and the unwanted behaviors within, evaluates if the answer to both of the above questions is “Yes”. If it is no, additional operational conditions are added and/or severity and controllability ratings are adjusted. This may drive a change in safety goals and their ASIL ratings, which lead to updates to the HARA report and drive subsequent functional safety and SOTIF measures.

In addition to these considerations, some of the safety goals may not necessarily be SOTIF-relevant in that there are no SOTIF factors that could lead to their violation. To make this delineation, one can consider for each safety goal whether any functional insufficiency or a driver misuse can lead to its violation; this is shown in the following subsection.

2.3 Summary of considerations

The initial HARA has been augmented with the SOTIF factors below, which represent environmental, infrastructure, and driver conditions. These can be taken as initial high-level triggering conditions to be analysed in greater detail later in the SOTIF workflow:

- Rain
- Dense fog

- Sandstorm
- Illumination
- Low friction road
- Split-mu road (tires on different surfaces)
- Potholes on the road
- Incorrect driver interaction
- Delayed driver interaction

As a result of this HARA augmentation, one need was identified. That need is a potential driver misuse scenario, **driver requesting ADS activation outside the ODD**, which considers risks induced by feature operation outside the designated bounds due to possible driver misunderstanding or lack of knowledge. This need is considered as part of safety goal 3 (SG3) shown in Table 2, but no further in depth assessment of driver engagement or intention will be considered as this was not part of the scope of the project where the focus was on external challenging conditions. In the EVENTS project, the driver's role during the ADS activation and the control transition to the driver during MRMs was not part of the investigations.

SOTIF relevancy of safety goals was also evaluated, and Table 2 below lists the safety goals that were deemed SOTIF-relevant. The other safety goals in D2.3 could be satisfied by considering internal system safety mechanisms not tied to functional insufficiency, and hence are not considered here.

Table 2: SOTIF-relevant safety goals

Safety goal name	Safety goal definition
SG3	Prevent ADS use outside of ODD
SG4	Prevent insufficient/unintended steering
SG5	Prevent unintended braking on system limit
SG6	Prevent loss or insufficient braking
SG7	Prevent unintended acceleration
SG9	Ensure safe stop in case of no driver takeover

3. Vehicle-level metrics

3.1 Approach using accident statistics

A core element to arguing the SOTIF for a system is establishing how safe is safe enough. Rationale that are often used, per ISO 21448 [3] include:

- GAMAB: A principle that means “globally at least as good as”.
- Positive risk balance: A principle that comprehensively considers the hazards of the new system while allowing tradeoffs to be made (e.g. if a residual risk has increased, it may be acceptable if it is counterbalanced by reductions in other risks).
- ALARP: A principle that means “as low as reasonably practicable”, ALARP recognizes that zero risk is not possible.
- MEM: A principle that is based on not increasing the death rate in society.

For the calculation of the vehicle-level metrics for this project, the partners selected the GAMAB approach (with exception of SG3), which entailed finding relevant accident types for each of the SOTIF-relevant safety goals. Those accident types per safety goal (refer to D2.3 [1] for their derivation) are as follows:

Table 3: Relevant passenger vehicle accident types per safety goals

Safety goal	Relevant passenger vehicle accident type
*SG3: Prevent ADS use outside of ODD	N/A
SG4: Prevent insufficient/unintended steering	Lane departure collisions
SG5: Prevent unintended braking on system limit	Rear-end collisions due to factors other than inattention by following driver or weather
SG6: Prevent loss or insufficient braking	Rear-end collisions
SG7: Prevent unintended acceleration	Rear-end collisions
SG9: Ensure safe stop in case of no driver takeover	Collisions due to drowsiness

**** SG3 did not have a relevant passenger vehicle accident type and thus was handled in slightly different manner than the rest of the safety goals.***

After identifying accident types, the corresponding statistics were identified and where appropriate, these statistics were normalized, and a “better than” factor of 10 was applied.

3.2 Goal-by-goal metrics breakdown

3.2.1 SG3

For SG3, the Random Hardware Fault (RHWF) target from ISO 26262 [2] was directly taken for the metric with the assumption that ODD detection within an ADS bears similarity to a hardware component that has some level of randomness to it. As SG3 carries an ASIL D rating, the metric is 1.00E-8 1/h.

3.2.2 SG4

The vehicle-level metric for SG4 was calculated as follows using readily available United States data on lane departure collisions from 3 separate sources, which are assumed to be comparable to European Union data:

Table 4: Calculations for SG4 vehicle-level metric

Item	Value	Unit
Number of crashes per year	5,930,496 [6]	crashes/year
Percentage of fatalities due to roadway departures	51% [7]	
Number of roadway departures per year	3024553	crashes/year
Hours driven in United States per year	9.30E+10 [8]	hours/year
Hours between crashes due to roadway departures	3.07E+04	hours/crash
"Better than" factor	10	
Hours between crashes due to roadway departures, improved	3.07E+05	hours/crash
Crash rate	3.25E-06	/h

3.2.3 SG5

The vehicle-level metric for SG5 was calculated as follows using readily available United States data on rear-end collisions from 3 separate sources, which are assumed to be comparable to European Union data:

Table 5: Calculations for SG5 vehicle-level metric

Item	Value	Unit
Number of crashes per year	5,930,496 [6]	crashes/year
Number of rear-end crashes per year	1700000	crashes/year
Percentage of rear-end crashes due to factors other than inattention by following driver or weather	28% [9]	
Number of rear-end crashes per year due to factors other than inattention by following driver or weather	476000	crashes/year
Hours driven in United States per year	9.30E+10 [8]	hours/year
Hours between rear-end crashes due to factors other than inattention by following driver or weather	1.95E+05	hours/crash
"Better than" factor	10	
Hours between rear-end crashes due to factors other than inattention by following driver or weather, improved	1.95E+06	hours/crash
Crash rate	5.12E-07	/h

This calculation considered data on rear-end crashes specifically due to factors other than inattention by following driver or weather due to the following reasons:

- If a crash occurred due to the following vehicle's driver being inattentive, the responsibility for the crash would be on that driver.
- Weather conditions that are challenging enough such that a following vehicle would crash into the ego-vehicle regardless of if the braking is intended or unintended (e.g. icy roads) are not considered here.

3.2.4 SG6 and SG7

The vehicle-level metric for SG6 and SG7 was calculated as follows using readily available United States data on rear-end collisions from 2 separate sources, which are assumed to be comparable to European Union data:

Table 6: Calculations for SG6 vehicle-level metric

Item	Value	Unit
Number of crashes per year	5,930,496 [6]	crashes/year
Number of rear-end crashes per year	1700000	crashes/year
Hours driven in United States per year	9.30E+10 [8]	hours/year
Hours between rear-end crashes	5.47E+04	hours/crash
"Better than" factor	10	
Hours between rear-end crashes, improved	5.47E+05	hours/crash
Crash rate	1.83E-06	/h

3.2.5 SG9

The vehicle-level metric for SG9 was calculated as follows using readily available United States data on collisions due to drowsiness from 3 separate sources, which are assumed to be comparable to European Union data:

Table 7: Calculations for SG9 vehicle-level metric

Item	Value	Unit
Number of crashes per year	5,930,496 [6]	crashes/year
Percentage of crashes due to drowsy driver	18% [10]	
Number of crashes per year due to drowsy driver	1043767.296	crashes/year
Hours driven in United States per year	9.30E+10 [8]	hours/year
Hours between crashes due to drowsy drivers	8.91E+04	hours/crash
"Better than" factor	10	
Hours between crashes due to drowsy drivers, improved	8.91E+05	hours/crash
Crash rate	1.12E-06	/h

3.3 Summary

The vehicle-level metrics calculated in Sections 3.2.1 to 3.2.5 are considering accidents which resulted in fatalities, injuries and property damage. Converting the number of accidents which only considered fatalities and injuries we get 1,707,112 accidents [6] for 2.8E12 miles driven in 2022 [8] resulting in 3.79E-7 per km which would be comparable the overall value outlined in D2.3 [1] for European statistics which was 3.85E-7 per km. This shows that there is a correlation between the statistics used for these values in Table 8 and D2.3 [1] increasing our confidence that they are also applicable for a vehicle used in the European region.

A recap of the vehicle-level metrics can be found below:

Table 8: Summary of vehicle-level metrics

Safety goal	**Vehicle-level metric (1/h)**
SG3: Prevent ADS use outside of ODD	1.00E-8
SG4: Prevent insufficient/unintended steering	3.25E-6
SG5: Prevent unintended braking on system limit	5.12E-7
SG6: Prevent loss or insufficient braking	1.83E-6
SG7: Prevent unintended acceleration	1.83E-6
SG9: Ensure safe stop in case of no driver takeover	1.12E-6

4. Scenario identification

4.1 Approach

Per ISO 21448 [3], a scenario is defined as a “description of the temporal relationship between several scenes in a sequence of scenes, with goals and values within a specified situation, influenced by actions and events”.

Similarly, a triggering condition is defined as a “specific condition of a scenario that serves as an initiator for a subsequent system reaction contributing to either a hazardous behaviour or an inability to prevent or detect and mitigate a reasonably foreseeable indirect misuse”. Focus on identifying triggering conditions is critical for 2 main reasons:

- The triggering conditions can be used up front during the concept phase to help develop measures to achieve the SOTIF.
- The triggering conditions can be applied as part of the scenario identification to generate test protocols for the verification and validation phases, exercising the system beyond nominal conditions.

Though there are multiple approaches to identify scenarios, the focus of this project is on creating scenarios from known triggering conditions. For this, we applied the following top-level approach:

Figure 3: Scenario identification approach

4.2 Triggering condition identification

Following the approach in Figure 3, we created a list of triggering conditions relevant to all facets of the functional and physical architecture (assuming one high-level EVENTS architecture as seen in D2.3 [1]). This list was compiled through various sources, including expert knowledge of the ADS, expert knowledge of components within the system, and benchmarking. Analyses, verification results, validation results, and real-world operation are other ways to determine triggering conditions for a system, but are not considered for this report as they are contingent on the availability of a more mature system and/or components within the system.

Regarding expert knowledge of the system and/or components within the system, the key questions posed to the expert sources were:

- What are known insufficiencies of the system, sensor, algorithm, actuator, driver, and/or user and what are all the conditions that cause those insufficiencies to come up?
- Are there items listed in the ODD taxonomy (shown in Figure 4) provided by ISO 34503 [11] that also can be considered as triggering conditions?

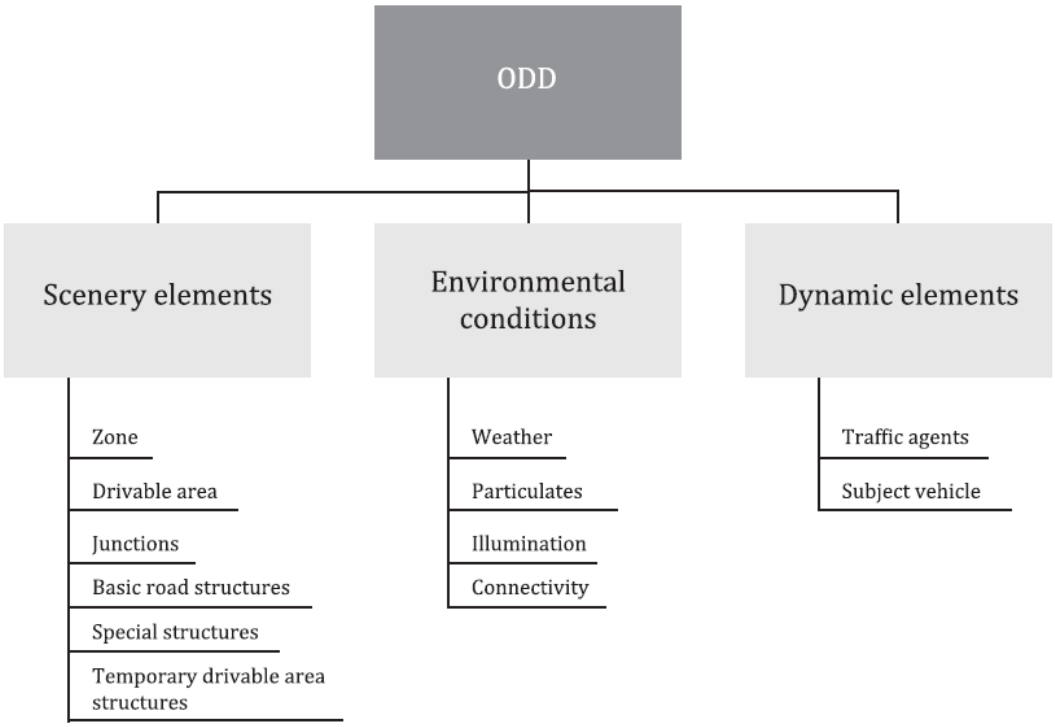

Figure 4: Top-level ODD taxonomy (Figure 2 of ISO 34503 [11])

Regarding benchmarking, this was done from different sources including, but not limited to:

- Accident and recall databases for issues caused by comparable features or components within the feature
- Human factors databases to understand how drivers or users use and misuse vehicles
- Academic papers on technology being used
- Market comparison reports
- Regulations impacting infrastructure or traffic over time
- News articles

4.3 Scenarios of interest

Using the stated approach, we formed the following scenarios:

Table 9: Scenarios of interest per safety goals

Safety goal	Relevant triggering conditions	Scenario
SG3: Prevent ADS use outside of ODD	Sensing/perception: Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast, inaccurate map data, antenna blockage, signal interference, satellite blockages (infrastructure, geological, or vegetation), roadways that change direction at certain times, tiered roadways	Scenario 1: Engaged system driving from in-ODD conditions to out-of-ODD conditions, in the presence of the listed scenery elements and/or environmental conditions Scenario 2: Driver trying to engage system while actively in out-of-ODD conditions, in the presence of the listed scenery elements and/or environmental conditions
SG4: Prevent insufficient/unintended steering	Sensing/perception: Degraded lane markings, covered lane markings Planning/control/actuation: Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, sudden tight turns, strong winds	Engaged system driving in the presence of the listed scenery elements and/or environmental conditions
SG5: Prevent unintended braking on system limit	Sensing/perception: Mud, snow, wet road, water splash/spray, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast	Engaged system driving in the presence of the presence of a rear vehicle and the listed scenery elements and/or environmental conditions
SG6: Prevent loss or insufficient braking	Sensing/perception: Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast Planning/control/actuation: Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, uphill conditions,	Engaged system driving in the presence of a lead object and the listed scenery elements and/or environmental conditions

	downhill conditions, strong winds	
SG7: Prevent unintended acceleration	Sensing/perception: Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast Planning/control/actuation: Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, uphill conditions, downhill conditions, strong winds	Engaged system driving in the presence of a lead object and the listed scenery elements and/or environmental conditions

5. Derivations of system- and subsystem-level metrics using calculations and simulations

5.1 Approach

For the SOTIF of a system, vehicle-level metrics are not the only bounds. System- and subsystem-level metrics that confine the system's dynamic behaviour (such as limits on lateral and longitudinal velocities and accelerations) are also relevant.

Generating the requirements for these metrics often entails characterizing vehicles while pushing them dynamically, which can be risky to do in the real-world, even with a trained test driver. Additionally, scenarios can be physically difficult to set up or be challenging to scale up. Given these challenges, alternatives are doing calculations or using a dynamics simulation tool in an exploratory fashion to do the characterizations.

For the intent of this deliverable, CarMaker was used and five separate simulation-based analyses were set up with different motivations in support of a subset of the SOTIF-relevant safety goals. These are summarized below:

Table 10: Analysis motivations per safety goals

Safety goal	Analysis motivation
SG3: Prevent ADS use outside of ODD	To determine how soon driver takeover and/or safe stop is needed
SG4: Prevent insufficient/unintended steering	To determine how soon the ego-vehicle leaves the lane with different steering errors
SG5: Prevent unintended braking on system limit	To determine collision times assuming varying conditions (errors on braking, following vehicle distance)
SG6: Prevent loss or insufficient braking	To determine collision times assuming varying conditions (errors on braking, lead vehicle distance and system response timing)
SG7: Prevent unintended acceleration	To determine collision times assuming varying conditions (errors on acceleration, lead vehicle distance)
*SG9: Ensure safe stop in case of no driver take over	N/A

***SG9 was not considered for a simulation-powered analysis as the bounds on lateral and longitudinal behaviour from the other safety goals were assumed to cover it.**

5.2 Simulation tool approach for exploring system and sub-system metrics

To perform the exploratory simulations to define the system and sub-system metrics, especially those related to the vehicle's on-road dynamic handling, the following parts of the simulation software (in this deliverable, IPG CarMaker Version 12.0) are leveraged:

- Road surface and geometrical modelling
- Vehicle dynamic modelling
- IPG built-in driver and Simulink to control the vehicle motion
- Ideal sensors for ground-truthing and measurements

The following sections will give more details about the above items, showing how CarMaker was used to obtain the required outputs for each simulation.

5.2.1 Road surface and geometrical modelling

In the considered simulations setups, the roads are limited to flat multi-laned roads that are straight or curved to left/right in some of its portion. For the curved road segments, their radius of curvature is a parameter that can be changed as per simulation setup.

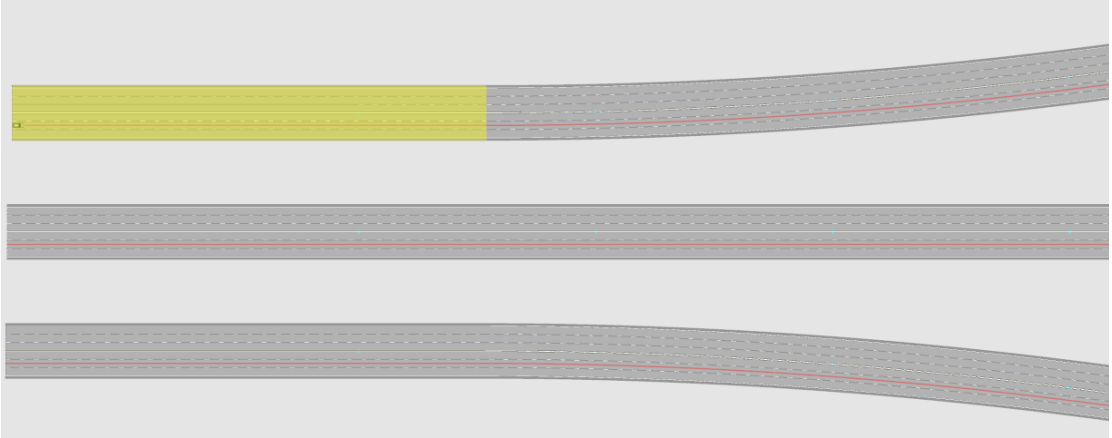

Figure 5: Flat roads with curvature for curved segments set as a test parameter

For some simulation setups, the width of the lane served as a test parameter and was varied as lateral excursions from the lane were being studied.

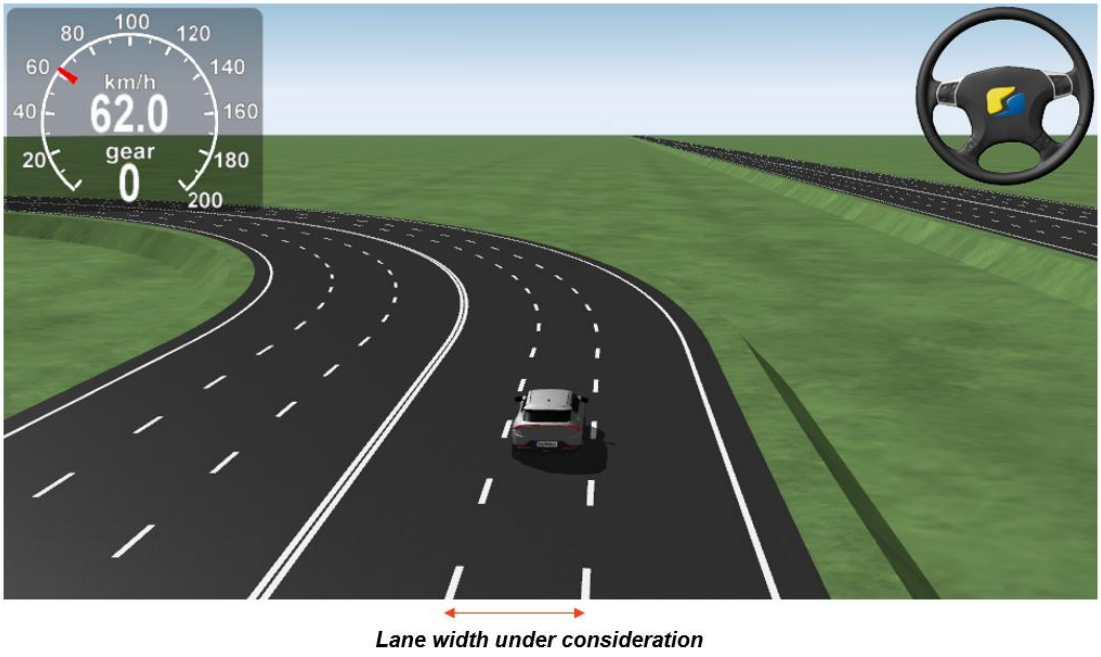

Figure 6: Lane width under consideration as a test parameter

Each road in the simulation setups has an associated friction parameter that will be fed into a tire model (that will be described in the following section) to create overall dynamic effects on the tested vehicle.

5.2.2 Vehicle dynamic modelling

The vehicle model used in this exploration study is a generic electric vehicle with a rigid body structure (derived from Kia EV6 model provided in IPG CarMaker 12.0). The "RealTime Tire" model is used for the four wheels (we use IPG CarMaker default model [12], Section 18.1 of the reference manual). This tire model takes, among other

quantities, the friction coefficient between tire and road as inputs to produce, among other outputs, the sideslip angle, the turn slip, and the longitudinal slip. This helps determining the overall dynamic effects on the tested vehicle.

For the steering, the model “Static Steer Ratio” is used with a rack travel to steering pinion angle of 85.598 rad/m (we use IPG CarMaker default model [12], Section 15.3).

5.2.3 IPG built-in driver and interaction with Simulink

For most exploratory simulation setups, the simulation tool’s built-in driver is used for lateral and longitudinal manoeuvres such as keeping in lane, keeping constant speed, accelerating, or decelerating. Some more complex simulations, however, require the injection of steering errors into the vehicle motion control. For these, Simulink shall be used as shown in Section 5.3.2.

5.2.4 Ideal sensors for ground-truthing and measurements

For ground-truthing and measurements during the simulations, the following sensors are added to the ego-vehicle:

1. Collision sensors are used to label when a collision happens.
2. Ideal object sensors are used for measuring relative distance and speed to other traffic users / vehicles.
3. Road line sensors are mounted on the four corners of the vehicle to detect when the ego-vehicle departs from the lane.

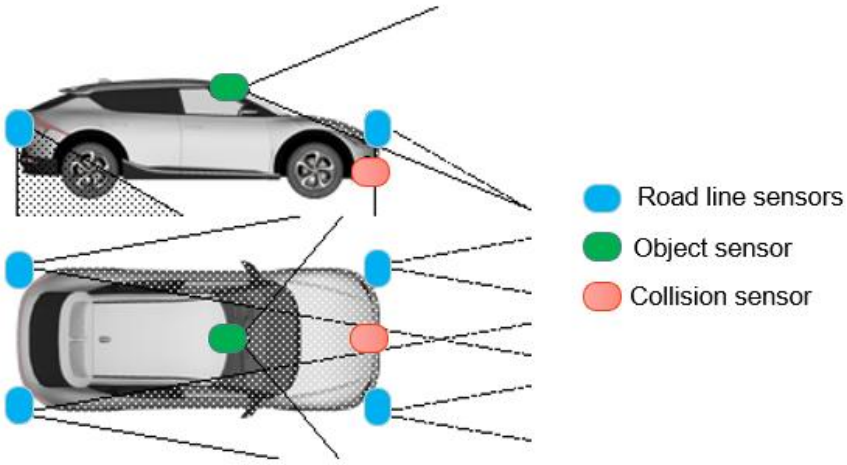

Figure 7: Ideal sensors for ground-truthing and measurements

5.3 Goal-by-goal simulation setup and outputs

5.3.1 SG3

Prior to running the simulation, a basic calculation was run to identify how much time it would take for the ego-vehicle to come to a full stop in ideal friction conditions. We are assuming that the ego-vehicle is travelling at a maximum speed of 130 kph for highways or 50 kph for urban roads, the driver takes 15 s to react to a takeover request (supported by literature [13]), and the maximum longitudinal deceleration is -4 m/s^2 (using the deceleration value for a minimum risk manoeuvre from R157 [14]). Therefore, it takes approx. 9 s at 130 kph or 3 s at 50 kph to come to a full stop, implying the perception module would have to identify an out-of-ODD zone 24 s at 130 kph or 18 s at 50 kph ahead of the zone.

The relevant information for the simulation setup for SG3 is as follows:

Table 11: Simulation setup information for SG3

ODD elements of interest	Zones identified as acceptable to operate in per ODD, road friction, weather
Vehicle and accessories	Passenger vehicle
Roadway	Straight road, curved road (tightest curvature at maximum speed)
Information to record in analysis	Curvature, friction, stopping distance
Simulation setup	Determine stopping distance at maximum speed within ODD and with varying road conditions (wet/dry road, straight/curved road).

For this exploratory simulation, the ego-vehicle is set up to follow a road (straight or curved to the left) with a given friction coefficient at a maximum allowed velocity. At a given time, the ego-vehicle shall attempt to brake at 4 m/s^2 with its deceleration controller. The outcome to be observed per simulation run is the stopping distance. Table 12 below summarises the different variations of the simulation setup.

Table 12: Variation of the simulation setup for SG3

Variation	Range	Variant
Curvature (1/m)	[0, 1/20]	[0, 1/1000, 1/800, 1/600, 1/500, 1/400, 1/300, 1/200, 1/100, 1/50, 1/30, 1/20]
Ego cruising speed (kph)	[25, 130]	[130, 130, 130, 130, 130, 125, 108, 88, 62, 44, 34, 27][(*)]
Road friction	[0.2, 1]	[0.2, 0.22, 0.25, 0.25, 0.3, 0.33, 0.35, 0.37, 0.39, 0.4, 0.41, 0.42, 0.45, 0.5, 0.7, 1]
Ego deceleration (m/s^2)	[4]	[4]

(*)Each speed in this array corresponds to one curvature in the above array. They are determined based on maximum allowed speed and maximum allowed lateral acceleration (3 m/s^2)

Figure 8 shows the outcomes of the simulations on the straight road. It depicts a constant stopping distance (204 m) when the friction coefficient varies between 1 and 0.37 and starts to increase up to 312 m (108 m increase for 130 kph vehicle speed) when friction coefficient goes lower than that (down to 0.2). These results confirm known observations that at lower levels of friction the required stopping distance increases; these numbers need to be taken into consideration when deciding how in advance the systems needs to know it would be going out of an ODD zone.

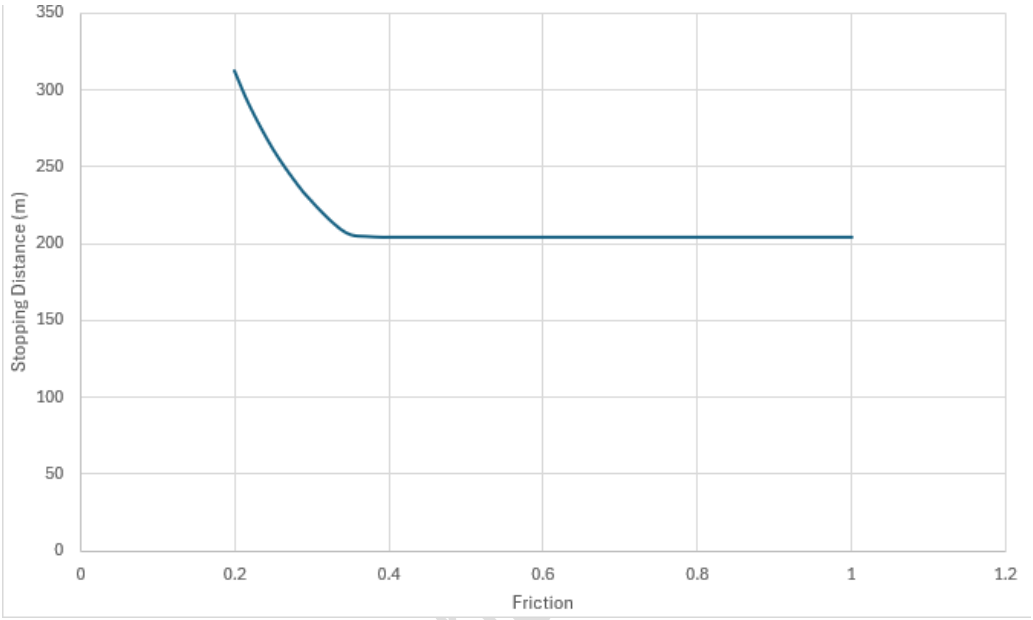

Figure 8: Stopping distance-friction relationship on straight road

For all curved roads in consideration (curvature from 1/1000 to 1/20), the ego-vehicle goes out of the road edge when friction goes below certain threshold μ_{orb}. This threshold increases as the curvature increases up-to 1/400 and then decrease again above such a curvature as shown in Table 13. It is important to remember that tither curves are transversed at much lower speeds thus enabling lower friction coefficient to be transversed without going out of the lane.

Table 13: Friction coefficient under which ego-vehicle goes out of road borders

Curvature (1/m)	Test speed (kph)	Friction coefficient under which ego-vehicle goes out of road borders μ_{orb}
1/1000	130	0.35
1/800	130	0.38
1/600	130	0.4
1/500	130	0.43
1/400	125	0.43

1/300	108	0.4
1/200	88	0.3
1/100	62	0.3
1/50	44	0.23
1/30	34	0.2
1/20	27	0.2

5.3.2 SG4

A rough calculation using an instantaneous heading error was run to identify what heading error would correlate to a given TTL in ideal friction conditions assuming that the ego-vehicle is travelling at its maximum allowed speed per the lateral acceleration limits of 3 m/s^2 (value from R79 [15]):

- 100 m curvature → 17 m/s
- 200 m curvature → 24 m/s
- 300 m curvature → 30 m/s
- 400 m curvature → 35 m/s
- 500 m curvature → 39 m/s

To find the heading error, we first calculated the distance to leave the lane, represented in Figure 9 below by a straight line of length L, from the TTL. As an example, for a 0.5 s TTL and assuming the vehicle was traveling around a 100 m curvature, the distance to leave the lane would be 0.5 s * 17 m/s = 8.5 m. Using this distance L and the lateral distance from the vehicle to the outside edge of the lane, represented below by b, it is then possible to make an approximation of the heading error as the angle between 2 equal lengths in an isosceles triangle. Heading error is equivalent to ɣ/2. The results of this calculation are shown in Figure 10.

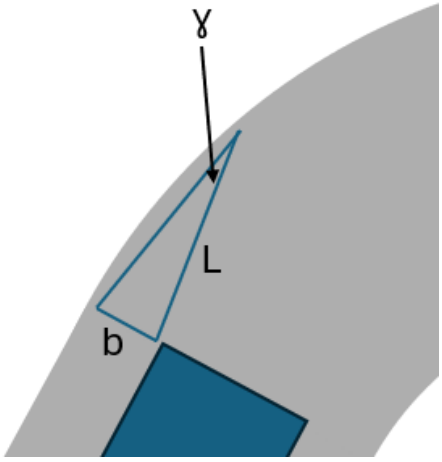

Figure 9: Heading error visualization

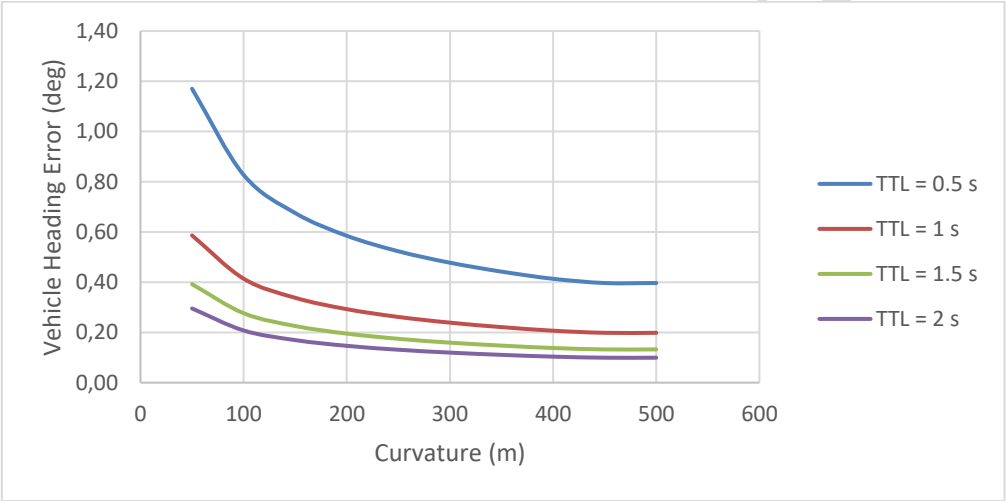

Figure 10: Heading error vs. lane curvature

As the rough calculation only used a heading error applied at an instant in time, a simulation was defined to capture the cumulative effects of a heading error over a period of time due to any number of factors (e.g. lane perception error, actuator error, etc.). The relevant information for the simulation setup for SG4 is as follows:

Table 14: Simulation setup information for SG4

ODD elements of interest	Curves, narrow lanes, weather
Vehicle and accessories	Passenger vehicle
Roadway	Varying curvatures (250, 500, 750, 1000 m) and varying lane widths (2.5, 3.5 m)
Information to record in analysis	Curvature, time to depart (time of curve start to time of any part of first wheel crossing line), lane width, friction

Simulation setup	Determine lane departure time when going maximum speed within ODD and with varying road (curvatures, widths, wet/dry road).

The following simulations need to use sweeping parameters whose concrete values are then presented in a corresponding table; these variables will be shown as italic text inside of square brackets when used in text.

For this exploratory simulation, the ego-vehicle is set up to drive first on a straight road segment with a speed of [*cruising-speed*] kph(*) before engaging into a curved road segment with a curvature of [*curvature*] 1/m with the same cruising speed. This cruising speed is set to be the maximum allowed velocity ensuring the lateral acceleration does not exceed 3 m/s^2 on curved segment.

The simulation is organized into two steps as follows:

1. Follow the lane centre with perfect perception and record the steering commands.
2. Replay the same recorded commands, but inject a steering command inaccuracy when engaging into the curved road segment.

The following diagram (Figure 11) illustrates the 2 above steps:

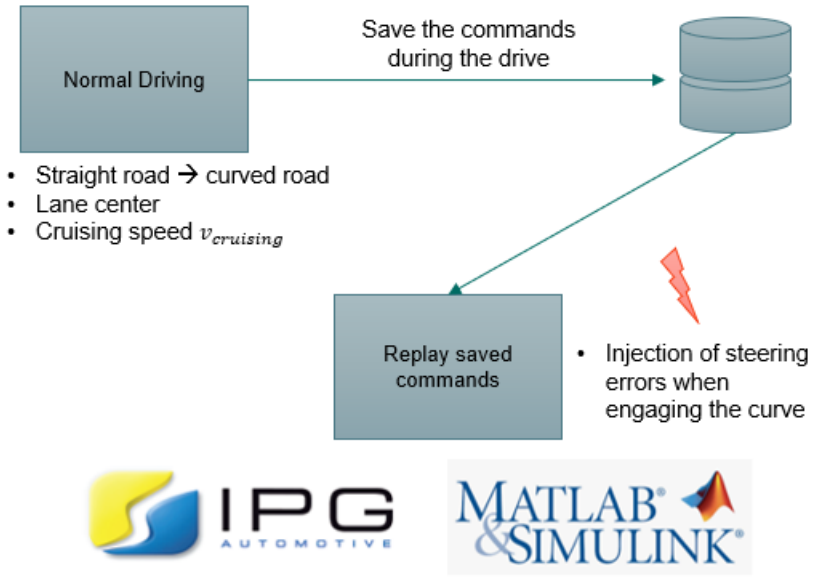

Figure 11: Injection of steering errors to emulate road perception inaccuracy

Table 15 below summarizes the different variations of the simulation setup. The outcome to be observed is the lane departure time or TTL.

Table 15: Variation of the simulation setup for SG4

Variation	Range	Variant
Road Curve Type	NA	[Straight, Left Curved]
Curvature [*curvature*] (1/m)	[0, 1/20]	[0, 1/1000, 1/800, 1/600, 1/500, 1/400, 1/300, 1/200, 1/100, 1/50, 1/30, 1/20]
Road friction	[0.2, 1]	[0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]
Cruising speed [*cruising-speed*] (kph)	[25, 130]	[130, 130, 130, 130, 130, 125, 108, 88, 62, 44, 34, 27][(*)]
Soft deceleration (m/s^2)	[-4]	[-4]
Lane width (m)	[2.5, 3.5]	[2.5, 3.5]
Max steering error (deg)	[-0.5, 0.5]	[-0.5, -0.3, -0.1, 0, 0.1, 0.3, 0.5]

(*)Each speed in this array corresponds to one curvature in the above array. They are determined based on maximum allowed speed and maximum allowed lateral acceleration

Due to limitations in time, these simulations were run with a braking event occurring due to safety measures already built into the software in the loop, and for this reason, the results are not directly transferrable to the original intent of the simulation, which is focused on TTL with heading error in absence any change to the longitudinal control. However, there are some indirect yet meaningful insights that were still gathered, which we discuss below.

Insight 1: In absence of steering error, lane departures will still occur at certain combinations of friction and curvature.

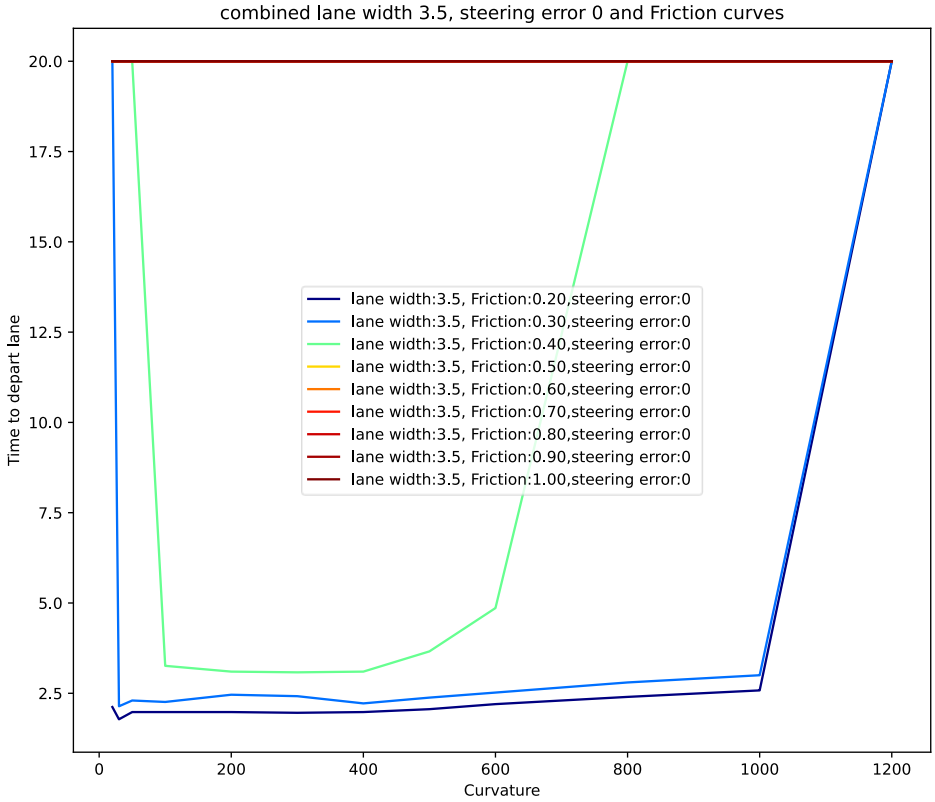

Figure 12: TTL-curvature relationship with no steering error and varying friction.

From Figure 12, one can gather that when lane width is of 3.5 m and no heading error (representing here by no steering error), the host vehicle stays in lane for all the curvature under consideration for friction coefficient greater or equal to 0.5 (a curvature of 1200 represents a straight road). For the friction coefficient below that (0.4), the host vehicle starts to leave the lane in the case of 100 m-600 m range of curvature. As friction goes lower, the host vehicle will leave the lane on a wider range of road curvatures; and the TTL goes also lower with this variation (TTLs are clamped to a maximum value of 20 s to in make it easier to read on Figure 12).

Insight 2: At or below a friction coefficient of 0.3, the TTLs are comparable regardless of heading error induced.

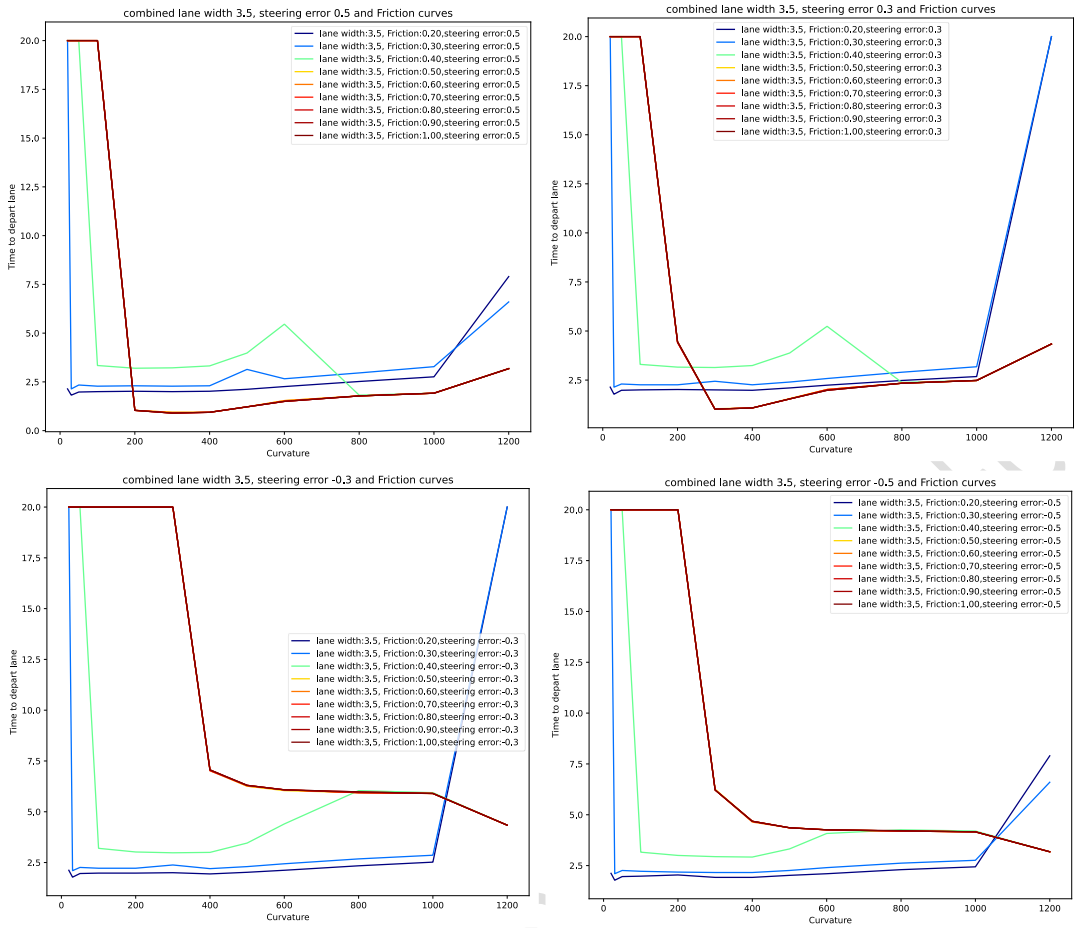

Figure 13: TTL-curvature relationship with varying steering errors and varying friction.

Figure 13 shows that for the friction coefficient smaller or equal to 0.3, the TTL are very much similar between the different steering errors. This can be explained by the fact that, under the effect of low friction, the host vehicle starts to slide, following more its free motion than its steered motion.

Based on the above insights, residual risk can be evaluated. What this means is that those working on the project can evaluate if the risks due to low friction are acceptable or if further measures, such as limitation of operation of the ADS in certain temperatures, are required. The exploratory simulations identified exactly what coefficients of friction lead to lane departure events and how quickly those lane departure events happened, which both are relevant pieces of information for the design of such measures.

5.3.3 SG5

Prior to running the simulation, a basic calculation was run to identify how much time it would take for a following vehicle to collide with the ego-vehicle in ideal friction conditions assuming that the vehicles had an initial time gap of 0.8 s and that the ego-vehicle was then braking for some duration of time at -12 m/s^2; representing an

unintended braking (SG5 violation). Figure 14 is assuming the following vehicle decelerates at -12 m/s^2 after 1.5 s, accounting for the following vehicle driver's reaction to the ego-vehicle braking. Values of distance that are negative indicate the occurrence of a collision. This calculation serves as a sanity check to help design the simulation.

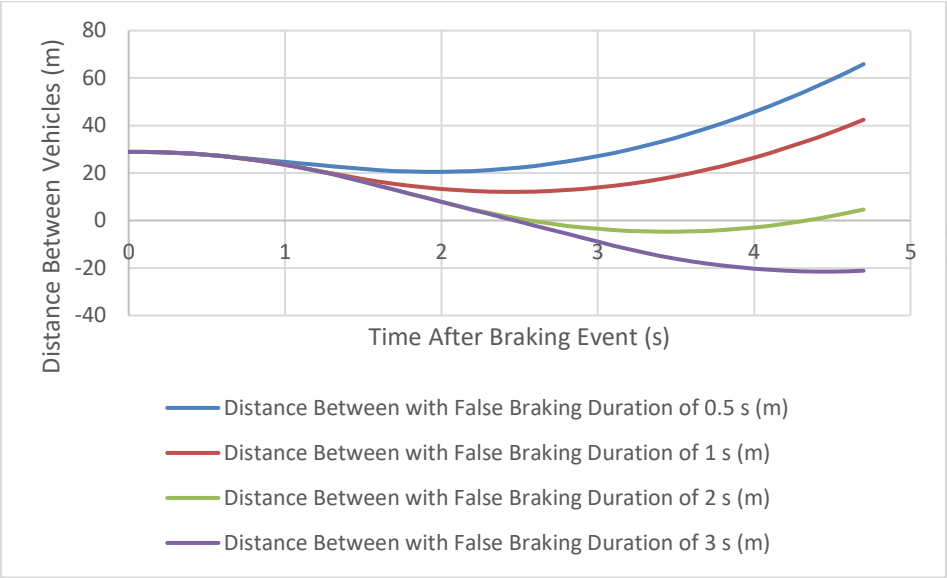

Figure 14: Distance between vehicles vs. time after braking event

The relevant information for the simulation setup for SG5 is as follows:

Table 16: Simulation setup information for SG5

ODD elements of interest	All
Vehicle and accessories	Passenger vehicle
Roadway	Straight road
Information to record in analysis	Following distance, time to collision (global minimum), braking duration, friction
Simulation setup	Determine collision time when going maximum speed within ODD if braking discontinues after X s, assuming wet/dry road conditions. Following vehicle is Y s (Y = 1, 2, 3) behind ego-vehicle travelling same maximum speed. Following vehicle keeps this speed for entirety of test run. Ego-vehicle brakes at maximum braking for X s (X = 0.5, 1, 1.5, 2) and then accelerates back up to original speed at maximum acceleration.

For this exploratory simulation, a vehicle is set up to follow the ego-vehicle at the same speed of 130 kph on a straight road. The initial time gap between the two vehicles is of [*initial-time-gap*] s. Given a triggering event, ego-vehicle starts to brake at [*soft-deceleration*] m/s^2 for a duration of [*braking-time*] s. If during this braking time, ego-vehicle speed reaches a threshold of 62 kph, it shall brake at [*hard-deceleration*] m/s^2. After some time (defined as a test parameter [*braking-time*]), the driver recognizes the fault braking situation (false positive braking); the driver takes over then the control and accelerates at 2.5 m/s^2. During all these events, the following vehicle keeps its speed constant (130 kph). Our choice of [*soft-deceleration*] value is 5 m/s^2 and this comes from our safety concept: performing full/hard brake is not done at high speed as it could lead to a destabilisation event (>62 kph).

Table 17 below summarizes the different variations of the simulation setup for this SG5. The outcome to be observed is the global minimum of the time to collision.

Table 17: Variation of the simulation setup for SG5

Variation	Range	Variant
Road friction [*road-friction*]	[0.2, 1]	[0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1]
Ego cruising speed (kph)	[130]	130
Soft deceleration [*soft-deceleration*] (m/s^2)	[-5]	[-5]
Hard deceleration [*hard-deceleration*] (m/s^2)	[-12]	[-12]
Soft-Hard Threshold (kph)	[62]	[62]
Initial time gap [*initial-time-gap*] (s)	[1, 3]	[1, 2, 3]
Braking time [*braking-time*] (s)	[0.5, 2]	[0.5, 1, 1.5, 2]
Ego acceleration (m/s^2)	[2.5]	[2.5]

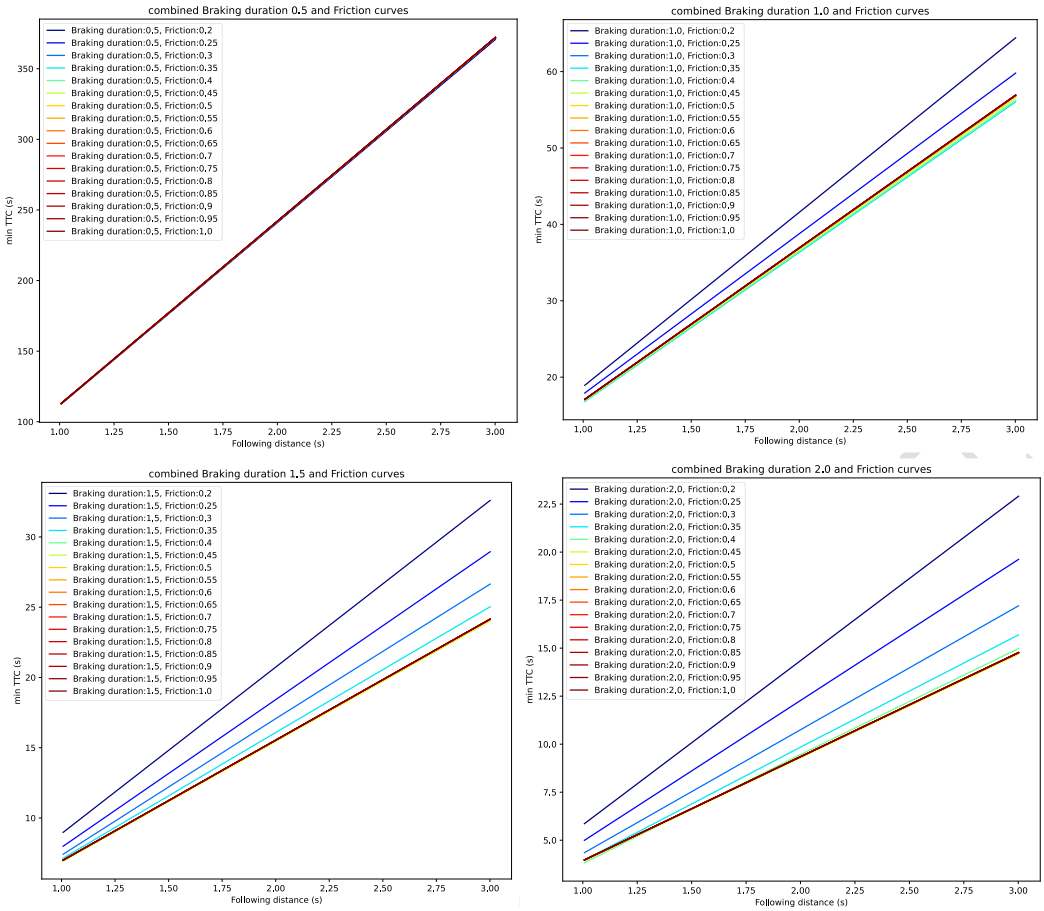

Figure 15: Minimum TTC for varying braking durations and frictions

Figure 15 shows that is a linear relationship between the initial time gap (or the initial following distance in seconds computed by dividing the distance by the velocity of the following vehicle) and the worst case TTC with larger initial time gaps resulting in larger TTC. Additionally, as one would expect the longer the ego-vehicle braking the smaller the TTC. Moreover, looking at the friction coefficient, it can be seen that the smaller the friction coefficient the larger the TTC. This is due to the ego-vehicle not reaching the desired level of deceleration during the deceleration event thus leaving a larger gap between it and the following vehicle.

Another important outcome of this simulation is that, irrespective of the friction coefficient, any false positive (FP) brake duration of less than or equal to 1 s does not result in critical TTC values. The smallest TTC values seen in these cases are above 15 s which is usually not considered a critical scenario. Additionally, no collisions were recorded in this simulation setup with the smallest TTC values below 5 s (approx. 3.8 s) seen for the FP braking of 2 s duration (refer to Figure 15 bottom right panel with following distance of 1 s).

For this simulation setup, the ego-vehicle was braking using the soft deceleration value (5 m/s^2) as it never reached the velocity threshold required for hard deceleration

in the considered simulation runs. The high TTC values seen in this simulation are also due to the dynamics of the ego-vehicle which require some time before the value of 5 m/s^2 can be reached so the average deceleration value reached in these simulation for the requested time duration had a magnitude smaller than 5.

TTCs smaller than 5 s are only seen for FP braking durations greater or equal to 2 s which agrees with Figure 14 where only durations above 2 s result in collision.

A FP braking can be caused by FP error from the perception modules, such an error will be considered Section 6.

5.3.4 SG6

Prior to running the simulations for this safety goal, a basic calculation was run to obtain an initial indicator of how much time it would take for the ego-vehicle to collide with a lead vehicle in ideal friction conditions assuming that:

- Prior to the lead vehicle braking, both vehicles were going 130 kph
- Prior to the lead vehicle braking, the vehicles were a distance of 0.8 s, or 29 m, apart
- The lead vehicle has stopped by braking at -12 m/s^2
- The ego-vehicle has attempted to match the above deceleration but has some error (ego braking lower than lead vehicle braking by some amount)

The x-intercepts for each curve shown in the figure below represent the times to collision for different braking errors.

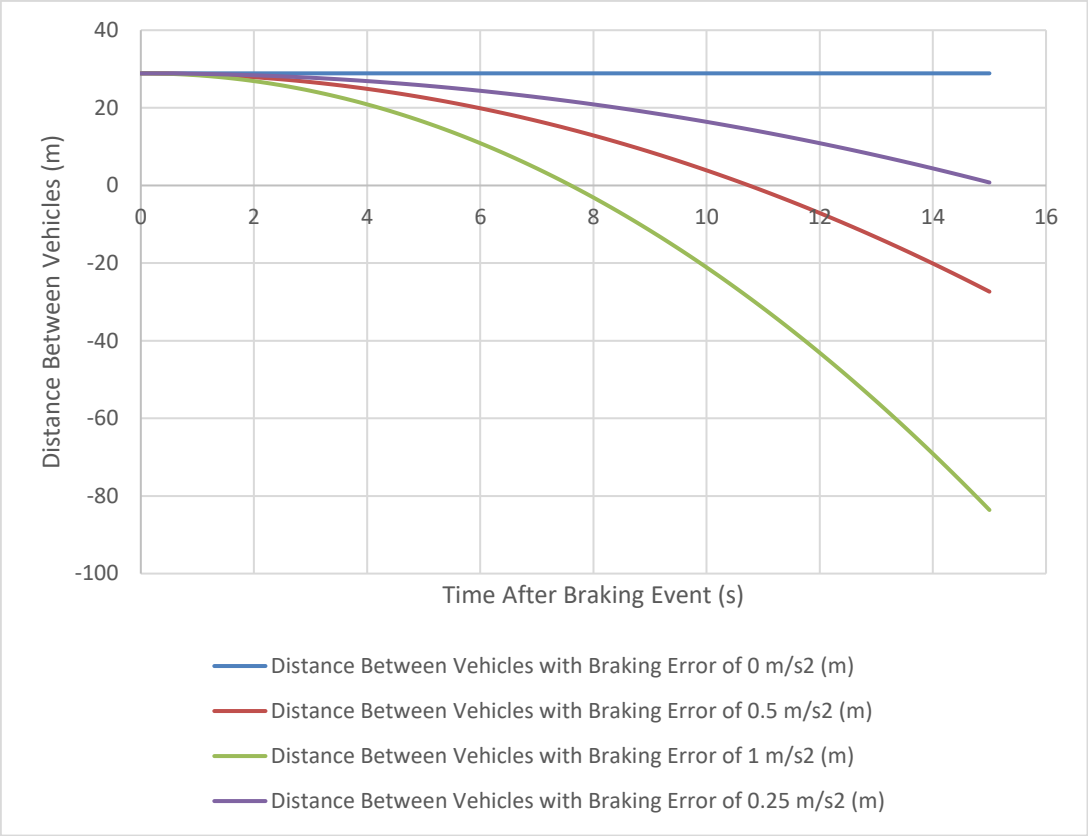

Figure 16: Distance between vehicles vs. time after braking event with ego braking error

The relevant information for the simulation setup for SG6 are as follows:

Table 18: Simulation setup information for SG6

ODD elements of interest	All
Vehicle and accessories	Passenger vehicle
Roadway	Straight road
Information to record in analysis	Following distance (initial time gap in s), time to collide (global minimum), braking deceleration error or delay, friction
Simulation setup	Determine collision time when going maximum speed within ODD and braking is insufficient by X m/s^2, assuming wet/dry road conditions. ego-vehicle is Y s (Y = 1, 2, 3) behind leading vehicle travelling maximum speed. Leading vehicle brakes at maximum braking. ego-vehicle brakes at maximum braking with deceleration error (minus X m/s^2 (X = 0.25, 0.5, 0.75) OR with Z s delay (Z = 0.25, 0.5, 0.75, 1).

For this exploratory simulation, the ego-vehicle is set up to follow a lead vehicle on a straight road with the same speed of 130 kph. Given a triggering event, the lead vehicle brakes with a deceleration of [*lead-deceleration*] m/s^2. To mitigate the risk of collision, the ego-vehicle starts to react only after a delay of [*react-delay*] s. Its reaction consists of a soft brake of [*soft-deceleration*] m/s^2 until it reaches 62 kph, and hence a hard brake with a deceleration of [*hard-deceleration*] m/s^2. A typical value for [*soft-deceleration*] is 4 m/s^2 and for [*hard-deceleration*] is 12 m/s^2. Table 19 below summarises the different variations of the simulation setup for this SG6. The outcome to be observed is the global minimum of the time to collision.

Table 19: Variation of the simulation setup for SG6.

Variation	Range	Variant
Road friction [*road-friction*]	[0.2, 1]	[0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1]
Cruising speed (kph)	[130]	130
Soft deceleration [*soft-deceleration*] (m/s^2)	[-4, -5]	[-4.25, -4.5, -4.75, -5]
Hard deceleration [*hard-deceleration*] (m/s^2)	[-11, -12]	[-11.25, -11.50, -11.75, -12]
Soft-Hard Threshold (kph)	[62]	[62]
Distance to lead (m)	[36, 108] + ego length	[36, 72, 108] + ego length
Lead deceleration [*lead-deceleration*] (m/s^2)	[-7, -8]	[-7.8]
Reaction delay [*react-delay*] (s)	[0, 1]	[0, 0.25, 0.5, 0.75, 1]

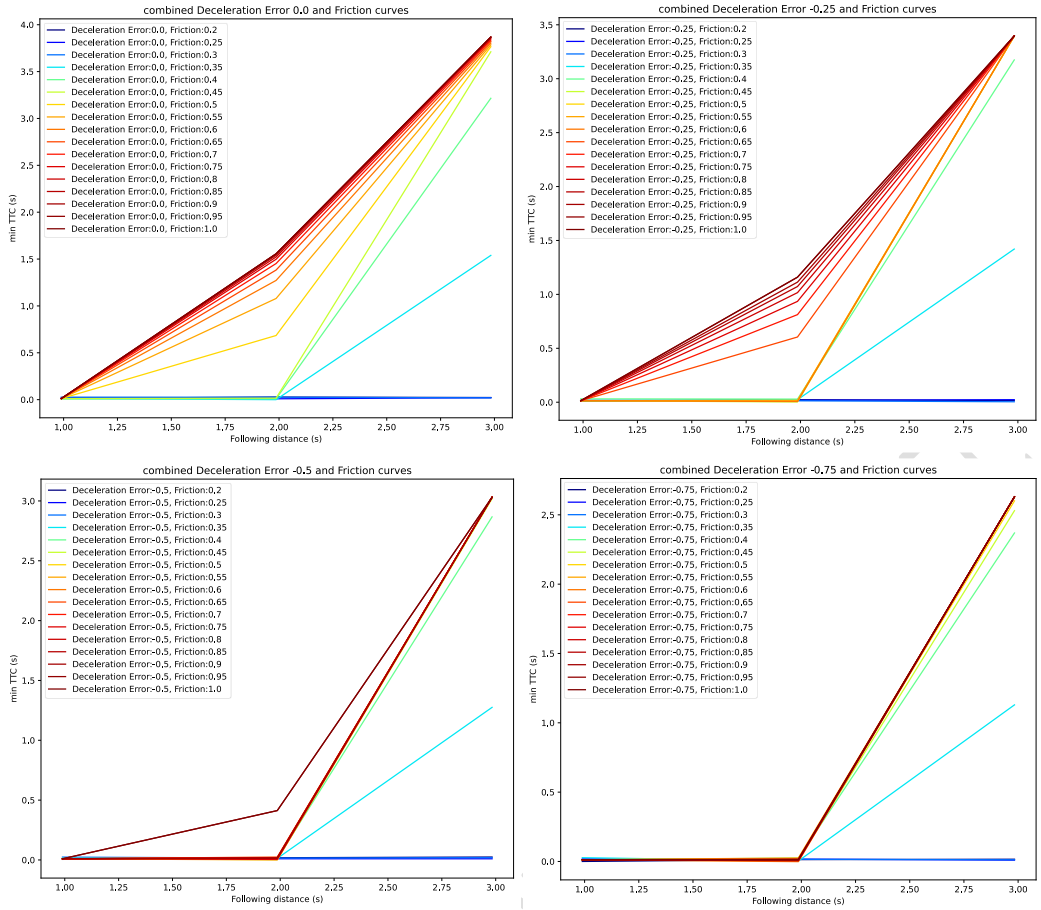

Figure 17: Minimum TTC for varying braking errors (with no delay), time gaps, and friction coefficients.

Figure 17 shows that there are collisions (TTC smaller than 0.1 s) in the simulation for low values of friction and there are always collisions for an initial time gap of 1 s at any friction value and at any braking error value. This agrees with the theoretical calculation shown in Figure 16 that for small time gaps braking value errors in these ranges lead to collision.

Additionally, the larger the braking error the more collisions are observed in the simulations with all runs for the braking error of -0.75 m/s^2 ending in collision for an initial time gap of 2 s. The initial time gap of 3 s shows to be the safest with no collisions for friction coefficients greater than or equal to 0.35. In contrast to the SG5 simulation, it is important to note that in this simulation setup low values of friction are making the situation more dangerous as the ego-vehicle is unable to reach the high value of decelerations to avoid a collision.

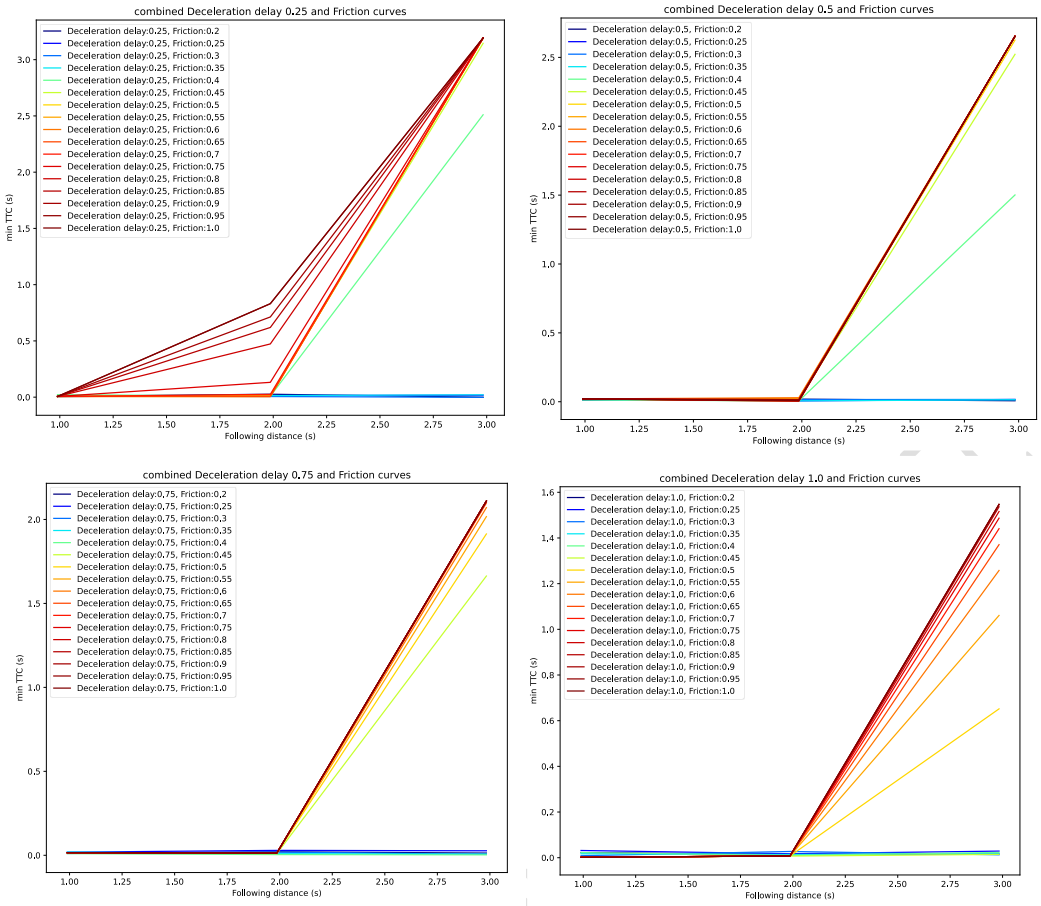

Figure 18: Minimum TTC for varying braking delays, time gaps, and friction coefficients

Figure 18 shows that delays greater or equal to 0.5 s lead to collisions for time gaps smaller or equal to 2 s irrespective of the friction coefficient. For a time gap of 2 s there are no collisions when the friction coefficients are larger than or equal to 0.5 for no delay and 0.75 for a delay of 0.25 s.

Once again as seen in Figure 17 the 3 s time gap proved to be the safest with no collision for friction coefficient larger than or equal to:

- 0.35 for no delay,
- 0.4 for delays of 0.25 s and 0.5 s,
- 0.45 for 0.75 s delay and 0.5 for 1 s delay.

These values of time gaps agree with the UK highway code recommendation [16] to keep at least 2 s to the leading vehicle in good weather conditions.

Considering both the 2 s and 3 s time gap we can see that the larger the delay in braking the higher the friction coefficient needs to be to avert collision, thereby restricting the road conditions which would be safe for the ADS.

Additionally, for the 1 s delay case with the 3 s time gap, the minimum TTC for simulations without collisions are all smaller than those seen for the same friction coefficient for the largest braking value error (-0.75 m/s^2) shown in Figure 17 suggesting that a delay in response can have more dire consequences than an error in the braking value. Furthermore, with an error of -0.75 m/s^2 the ego-vehicle was able to brake safely at lower friction coefficients compared to the 1 s delay scenario showing that the former can work on a larger range of road conditions compared to the latter. A braking delay can be caused by false negative (FN) error from the perception modules, such an error will be considered Section 6.

5.3.5 SG7

Prior to running the simulation, a basic calculation was run to identify how much time it would take for the ego-vehicle to collide with a lead vehicle in ideal friction conditions assuming that:

- Prior to the lead vehicle braking, both vehicles were going 130 kph.
- Prior to the lead vehicle braking, the vehicles were a distance of 0.8 s, or 29 m, apart.
- The lead vehicle has sped away at 3 m/s^2.
- The ego-vehicle has attempted to match the above acceleration but has some error.

Values of distance that are negative in Figure 19 indicate the occurrence of a collision. The x-intercepts for each curve below represent the time to collision.

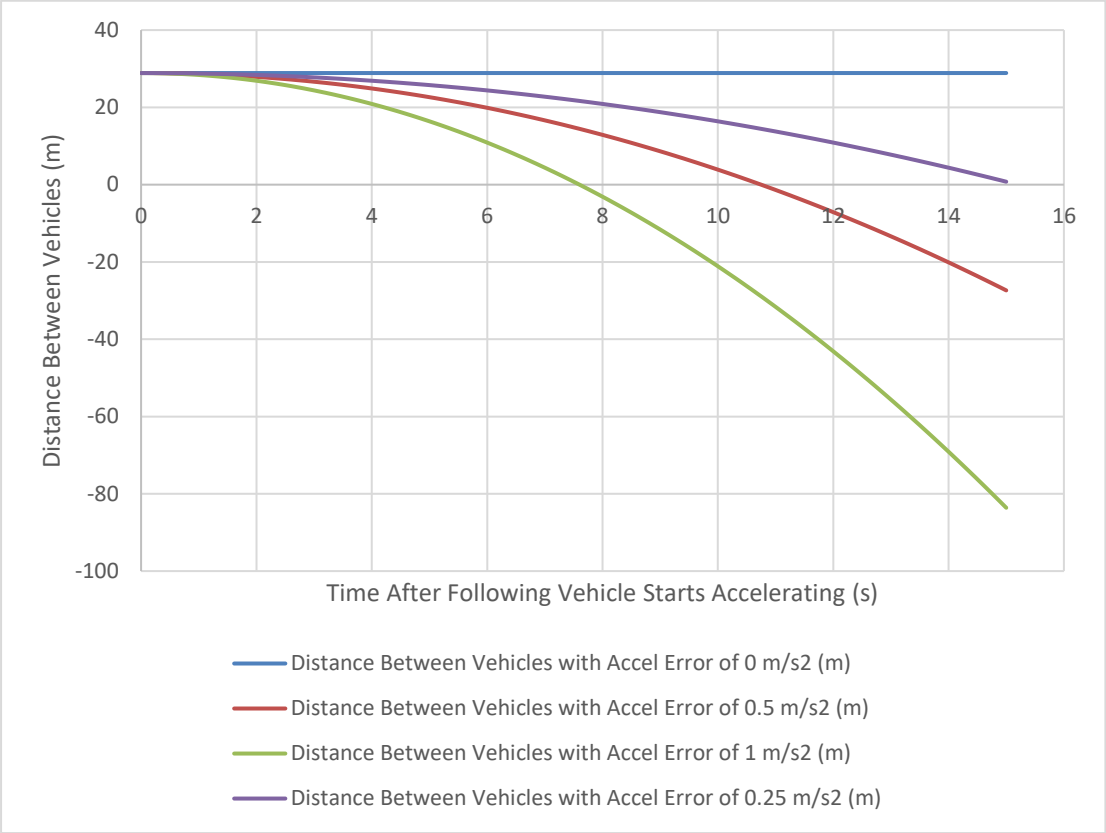

Figure 19: Relative distance vs. time after acceleration event with ego acceleration error

The relevant information for the simulation setup for SG7 are as follows:

Table 20: Simulation setup information for SG7

ODD elements of interest	All
Vehicle and accessories	Passenger vehicle
Roadway	Straight road
Information to record in analysis	Following distance, time to collide (global minimum), throttle error, friction
Simulation setup	Determine collision time when going X m/s within ODD and throttle is incorrect, assuming wet/dry road conditions. Ego-vehicle is Y s (Y = 1, 2, 3) behind leading vehicle travelling X m/s. Leading vehicle accelerates away at Z m/s^2 (Z = 1, 2, 3). Distance between ego-vehicle and lead vehicle is still over the desired setpoint, so the ego-vehicle also accelerates. Ego-vehicle accelerates with error (plus X m/s^2 (X = 0.25, 0.5, 0.75 above lead vehicle).

For this exploration, the ego-vehicle is set up to follow a lead vehicle on a straight road with the same speed of 80 kph and with a time gap of [*initial-time-gap*] s. Given a triggering event (e.g. road ahead free up), the lead vehicle starts to accelerate with [*lead-vehicle-acceleration*] m/s^2 until it reaches the cruising speed of 130 kph which is also the desired set speed for the ego-vehicle. At the same time, ego-vehicle also accelerates but with [*ego-accel-error*] m/s^2 higher than lead vehicle acceleration. Here below is a table summarizing the different variations of the simulation setup for this SG7. The outcome to be observed is the global minimum of the time to collision.

Table 21: Variation of the simulation setup for SG7

Variation	Range	Variant
Road friction	[0.2, 1]	[0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1]
Cruising speed phase 1 (kph)	[80]	[80]
Lead vehicle acceleration [*lead-vehicle-acceleration*] (m/s^2)	[1, 2.25]	[1, 1.5, 2.25]
ego acceleration error [*ego-accel-error*] (m/s^2)	[0.25, 0.75]	[0.25, 0.50, 0.75]
Cruising speed phase 2 (kph)	[130]	[130]
Initial time gap [*initial-time-gap*] (s)	[1, 3]	[1, 2, 3]

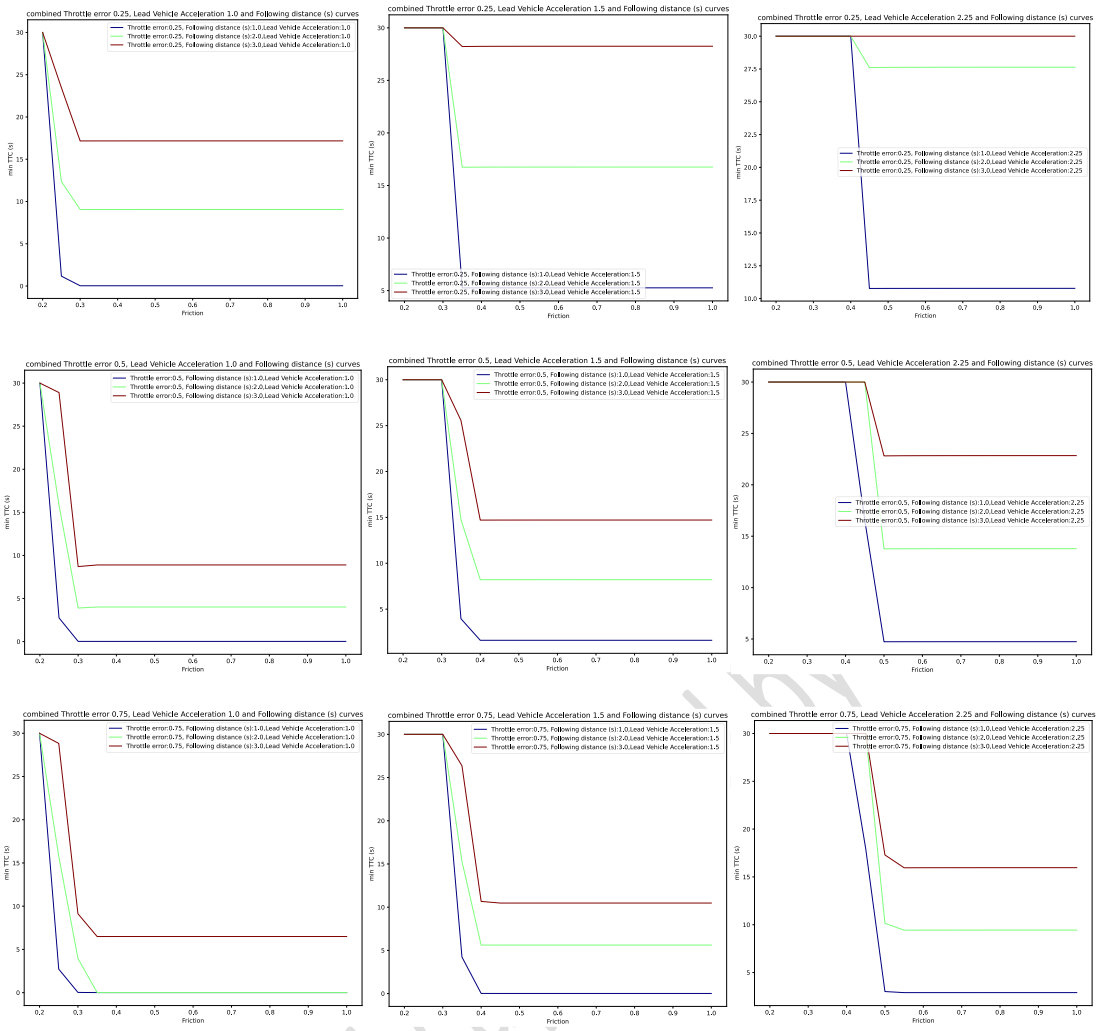

Figure 20: Minimum TTC (s) versus friction coefficient in varying conditions

Figure 20 shows the general tendency that low friction helps reduce frontal collision due to throttle errors. This is mainly because low friction does not permit the ego-vehicle to reach the requested erroneous acceleration, thus reducing the frontal collision risks. This figure shows also that, with a given level of acceleration (e.g. lead vehicle acceleration of 2.25 m/s^2, throttle errors of 0.25 m/s^2), roughly the same friction threshold (0.4 in this case) would trigger this increase in TTC.

The roughly-equal distances between the curves on high-friction area show the close-to-linear relationship between the following distance and the minimum TTC, except when collision happens or when the minimum TTC exceeds the maximum value of 30 s used to make Figure 20 more readable . On the low-friction area, the above-mentioned relationship might be different but as the two vehicles are far from collision, no further investigation is needed within this study.

With the same following distance, the same throttle error would induce lower minimum TTC when the acceleration level reduces (figures from right column to left

column). In other words, TTC becomes more critical as the relative throttle error increases (relative throttle error = $\frac{throttle\ error}{acceleration\ level}$).

The Figure 20 shows also that following distance of 1 s is the most susceptible to collision (58 over 153 cases in consideration and this can happen to all three throttle errors 0.25, 0.5, 0.75 m/s^2); the following distance of 2 s is less subjected to collision (14 over 153 cases, and only happens with throttle errors of 0.75 m/s^2); whereas the following distance of 3 s is considered safer as it never leads to collision in this simulation setup.

5.4 Summary of findings

The requirements along with their rationales can be found below:

Table 22: Safety requirements resulting from simulations and calculations

Safety Goal	Requirement	Rationale
SG3	System shall detect upcoming unmapped zone before 24 s of reaching that zone if in speed zone between 50 and 130 kph, or 18 s if in speed zone below 50 kph.	This requirement assumes that the driver has a 15 s opportunity to react to a takeover request, and the maximum longitudinal deceleration is -4 m/s^2. 25 s or 18 s accounts for the amount of time needed to stop the vehicle before the unmapped zone if the driver does not respond to the takeover request.
	System shall detect construction zones before 24 s of reaching that zone if in speed zone between 50 and 130 kph, or 18 s if in speed zone below 50 kph.	
SG4	Vehicle heading shall be correct within TBD degrees of commanded vehicle heading. Note: Due to time constraints on the simulation tied to this, the refined value is a future work.	This requirement assumes that the ego-vehicle is at the maximum speed for a curve of 500 m, 39 m/s (defined by maximum lateral acceleration of 3 m/s^2). Higher errors lead to or come close to a lane excursion event.
SG5	After actuating emergency braking, system shall abort braking if target has not been detected for a duration of 0.5 s.	Based on the simulation, it was found that aborting braking after 0.5 s gives the ego-vehicle the opportunity to continue without collision from a following vehicle, assuming it takes that vehicle's driver 1.5 s to react based on human factors averages with some padding for unexpected situations [17].
	System shall account for both forward and rear threats when calculating braking command.	The harm induced by a following vehicle running into the ego-vehicle due to a false positive braking event could be higher than the harm induced if a false positive lead vehicle was real.
	System shall meet TBD false positive rates toward its perception modalities.	Multiple perception modalities working with each other can help to address

		each modality's insufficiencies. Placing a false positive metric on each modality helps to ensure that the top-level metric for this safety goal is met.
SG6	Braking shall be correct within 0.25 m/s^2 of braking command.	This requirement assumes that ego and lead vehicles are going maximum speed of 130 kph, with the ego-vehicle trailing at a following distance of 0.8 s. The lead vehicle stops at -12 m/s^2, and the ego-vehicle reacts immediately, but with degraded deceleration. Time to collision is 15 s, which overlaps the 15 s opportunity a non-attentive driver has to take over.
	System shall meet TBD false negative rates toward its perception modalities.	Multiple perception modalities working with each other can help to address each modality's insufficiencies. Placing a false negative metric on each modality helps to ensure that the top-level metric for this safety goal is met.
SG7	Throttle shall be correct within 0.2 m/s^2 of acceleration command.	This requirement assumes that ego and lead vehicles are going speed of 120 kph, with the ego-vehicle trailing at a following distance of 0.8 s. The lead vehicle speeds away at 12 m/s^2, and the ego-vehicle reacts immediately, but with incorrect throttle. Time to collision is 15 s, which overlaps the 15 s opportunity a non-attentive driver has to take over.

6. Derivation of subsystem and module-level target metric allocations using FTA

6.1 Approach

Decomposing a vehicle-level safety goal violation metric to lower-level targets such as subsystem and module targets is an important step of a safety analysis. This in-depth study can help to understand what each module needs to achieve in order for the system metrics to be respected. For our decomposition approach, we will be using fault tree analysis (FTA) to break down a vehicle-level safety goal violation event (our high-level undesired event) into basic events representing subsystem and module-level errors.

Figure 21: OR gate (left) and AND gate (right).

An FTA represents a physical system as a diagram using logical gates (most commonly used being AND and OR gates shown in Figure 21) to link between different events that can cause a high-level undesired event. The OR gate means that the probabilities of the events are added whilst the AND gate means they will be multiplied. Events are represented using either circles (basic events that require no further decomposition) or rectangles (events which are derived from other events). An FTA is done by first understanding the system of interest, then by constructing the FTA structure and finally by doing qualitative and quantitative evaluation [18].

First, we start by looking into the EVENTS architecture to get an understanding of our system and how an error would propagate through it (see 6.1.1). This propagation path is then captured in the FTA structure (see 6.1.2). We also include SOTIF-related triggering conditions presented in Section 4.3 as part of the FTA. Subsequently, we explain how to perform a quantitative evaluation of the FTA; this is done by assigning error rates or probabilities to each basic event in the FTA to ensure the vehicle-level safety goal violation rate presented in Table 8 (Section 3.2) is respected. Afterwards, we correlate the error rate assigned to each basic event linked to a perception subsystem to a KPI goal value (see 6.1.3). This approach is summarised in Figure 22. For the decision/motion planning subsystem, a simpler approach is presented in Section 6.1.4. We also propose some further performance metrics which could be used in WP6 (see Section 6.1.5). Finally, we present an example of how to use FTA to derive a module-level KPI metric (Section 6.1.6) and end with a summary of our findings in Section 6.2.

Figure 22: Method for vehicle-level metric decomposition into subsystem/module level metric

The decomposition from vehicle-level to module-level of a metric can be used to refine safety requirements with specific values. Additionally, the FTA can be used to check if further safety requirements need to be added in case the high-level metric cannot be met.

6.1.1 Overall EVENTS project detailed architecture

In D2.2 [2], the consortium presented a high-level architecture for the EVENTS projects as shown in Figure 23. This section aims to present a more detailed architecture showing how modules inside each of the main blocks seen in Figure 23 are connected to each other. This is done to gain a better understanding of the system to assist with FTA.

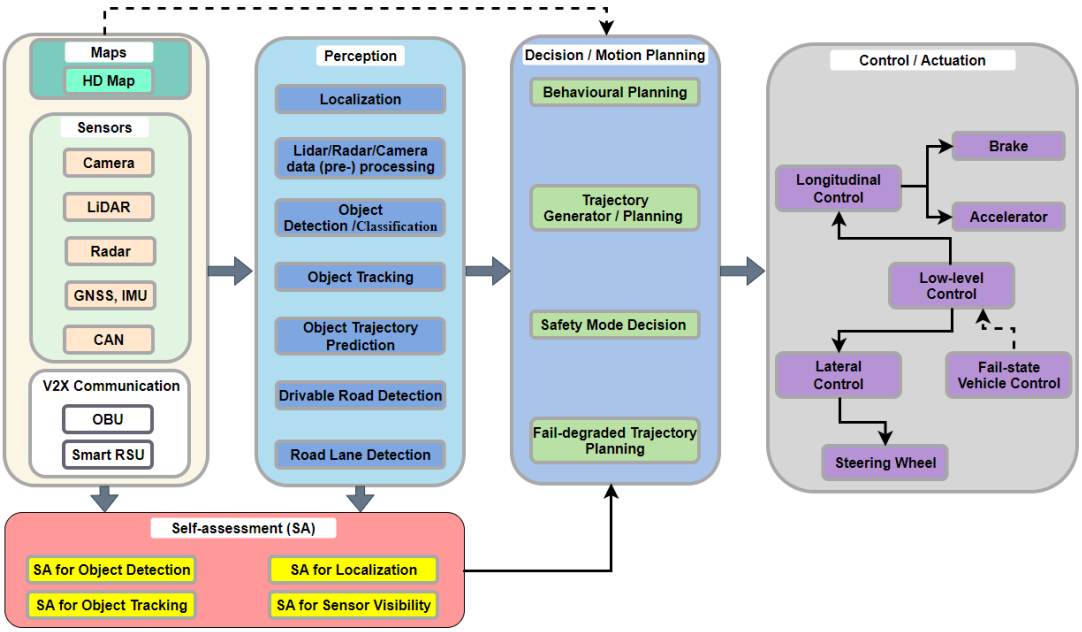

Figure 23: EVENTS architecture taken from D2.2 [2]

Figure 24 shows the interconnections between the various modules used inside of the perception subsystem. Some modules are used by many other modules, such as the localization module and the object detection and classification module. Here, we assume the classification part to have fewer safety-related consequences compared to the detection of the objects, and false positive (FP) and false negative (FN) perception errors will be the focus of the FTA [19] .

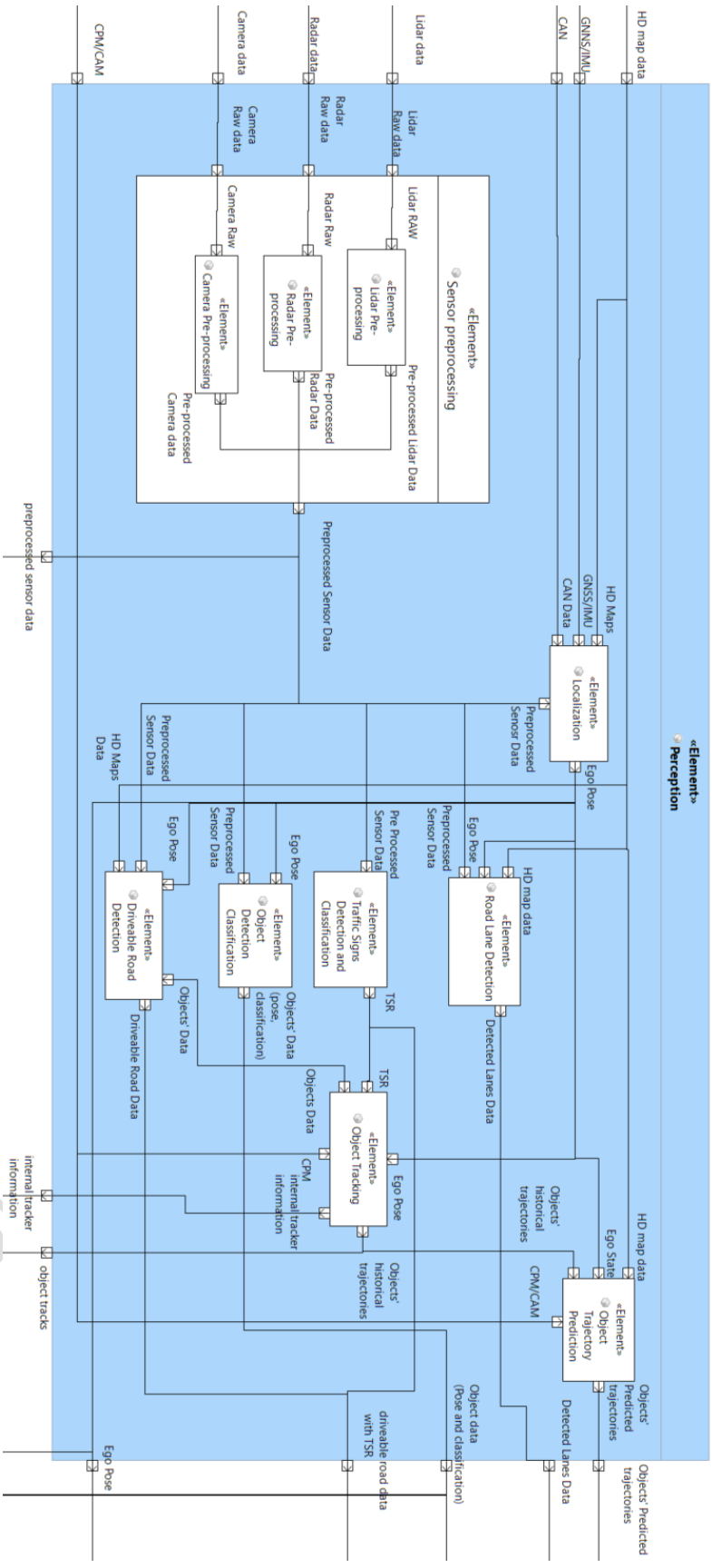

Figure 24: Detailed perception architecture for EVENTS with V2X via CPM/CAM input.

The following assumptions are made regarding the modules inside the perception subsystem:

1. Localisation: Assumed to be required by the EVENTS project ADS assuming GNSS and IMU inputs. Additionally, EXP8 provides localisation information using radar as main input.

2. Road lane Detection: Assumed to be required by the EVENTS project ADS. EXP4 provides lane detection using vision as main input.

3. Driveable Road Detection: Assumed to be required by the EVENTS project ADS. EXP8 provides scene segmentation using radar as main input; this is considered as part of driveable road detection in this analysis.

4. Traffic Signs Detection and Classification: Assumed to be required by the EVENTS project ADS. EXP4 detect and classify signs using vision as input. TSR (traffic sign recognition) will be used to refer to the output of this module, here TSR does not include alerting the driver about current traffic sign information.

5. Object Detection and Classification: Assumed to be required by the EVENTS project ADS. Main contributors are EXP1 using lidar as input, EXP5 using lidar as input, EXP6 using radar as input and EXP7 using lidar as input.

6. Object Tracking: Assumed to be required by the EVENTS project ADS. Main contributors are EXP1, EXP3 using V2X with CPM and EXP5. The output of this module are object tracks (also referred to objects' historical trajectories in Figure 24).

7. Object Trajectory Prediction: Assumed to be required by the EVENTS project ADS. Main contributors EXP1, EXP2 using V2X with CPM/CAM and EXP5.

The above assumptions are important for the FTA in order to know where to put the SOTIF triggering conditions dependent on the sensing input modality of each module.

Figure 25 shows all the modules that compose the self-assessment subsystem. We are assuming different kinds of self-assessment which are then captured at different levels in the FTA. Some self-assessment modules are assumed to provide frame-by-frame error information, while others are assumed to have a longer time frame. Additionally, Augmented Perception is not shown as part as the on board self-assessment but it shown as an input to the decision and motion planning subsystem and considered as self-assessment.

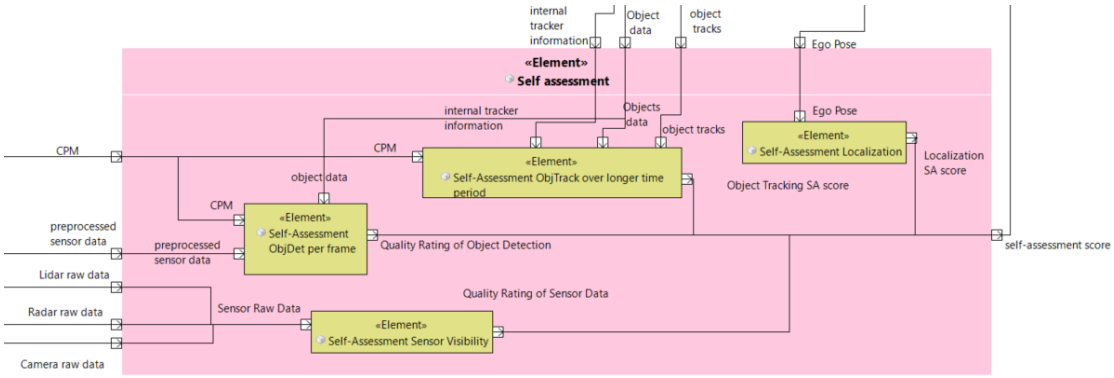

Figure 25: Detailed self-assessment architecture for EVENTS

The following assumptions are made regarding the modules inside the self-assessment subsystem:

1. A sensor visibility self-assessment is available.
2. A self-assessment per frame is available for the object detection and classification module. Main contributor is EXP2 (with V2X), EXP5 and EXP7.
3. A self-assessment over longer time period is available for the object tracking module. Main contributor EXP3.
4. A self-assessment per frame for GNSS localisation is available for the localisation module. Main contributor is EXP7.

Figure 26 shows the detailed architecture for the modules that comprise the decision and motion planning subsystem. This was included for completeness as the internal modules of this subsystem are not required in our FTA (as we will be focusing on perception), only the inputs to the decision and motion planning subsystem are required for further analysis. One of the inputs is augmented perception. Augmented perception (EXP2) using V2X object-level information from agents able to exchange CPMs serves as the self-assessment for object trajectory prediction in the extended field of view.

For this safety proof of concept, we focused more on perception module error derivation and not on decision making; this is a good starting point as some literature argues it is one of the most significant factors for safety [20]. However, decision making module errors are still important and could also be derived in the same approach but we leave this for future work.

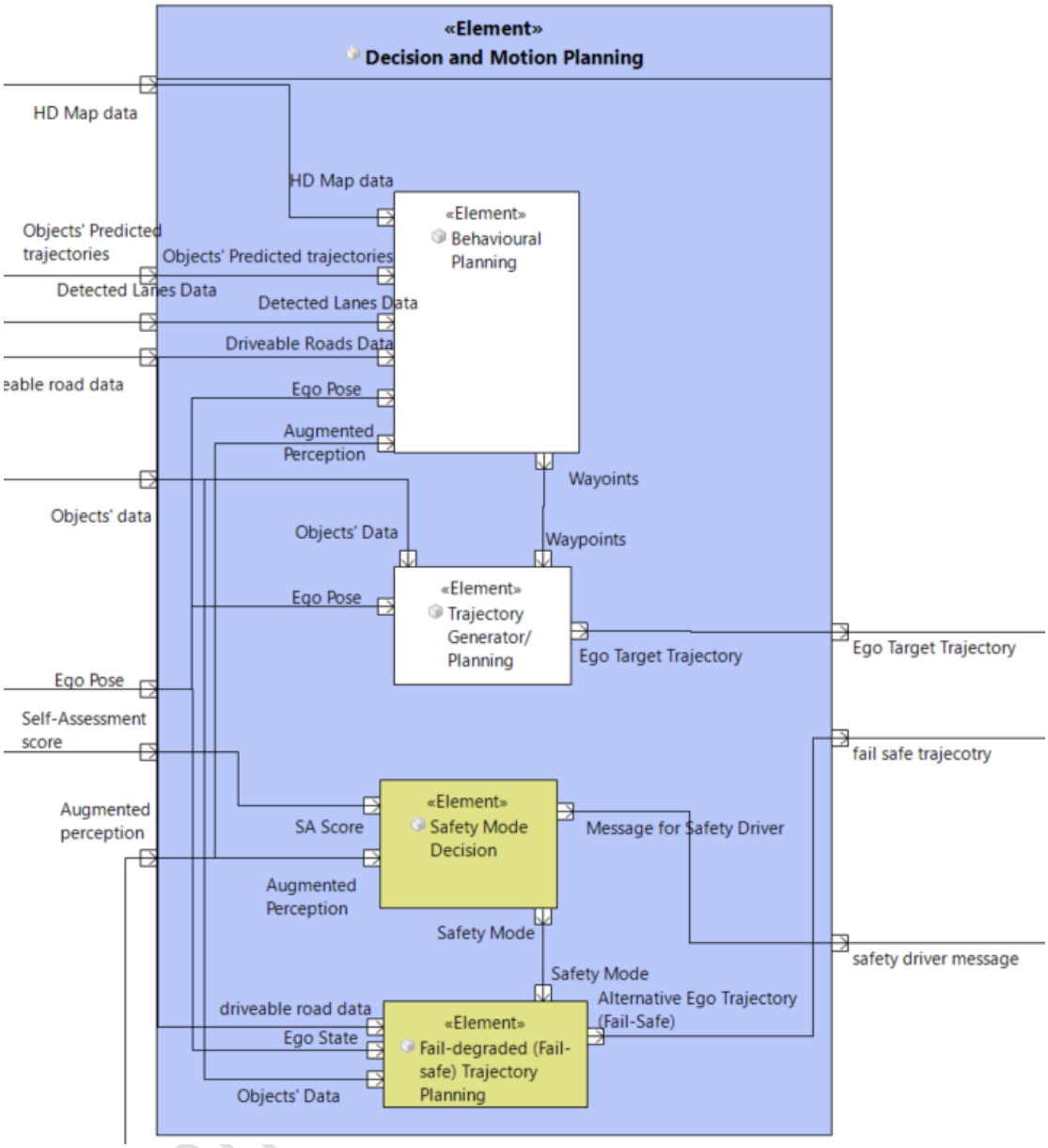

Figure 26: Detailed decision/motion planning architecture for EVENTS

6.1.2 FTA for error rate decomposition

The FTA shown in this deliverable assumes that there is a harmonised architecture for EVENTS with all components or modules developed in the different experiments working together to implement an automated driving system (ADS). The self-assessment modules shown in Figure 25 reduce the occurrence of a fault and are treated like a safety mechanism for the system; we will not be decomposing these further to include SOTIF triggering events.

Looking at the output ports of the perception and self-assessment modules in the previous section, it can be seen which inputs directly influence the decision making output. These are the objects' predicted motions, detected lane boundaries, object data (object pose and classification coming from the object detection/classification module), driveable road data with TSR, ego pose (the output of localisation module), self-assessment score and augmented perception (these last two inputs are captured via AND gates to the corresponding modules). Additionally, almost all of the self-assessment modules (apart from augmented perception) are using the same input sensors as the modules that they are protecting, so using an AND gate makes the assumption that the self-assessment module and the rest of the modules do not have an error caused by the same factor simultaneously. Errors in these modules will be used as basic events in our FTAs.

The system theoretic process analysis (STPA) guidewords explained in D2.3 were also used to help in the FTA design, challenging us to think about if an error occurs because something was provided:

1. Not at all or not when needed,
2. With incorrect intensity,
3. With incorrect delivery,
4. With incorrect duration,
5. Too early or late.

The guidewords in point 1, point 4, and point 5 all lead to FP and FN errors in perception modules; these errors will be the focus of our FTA. Those in points 2 and point 3 were not considered to focus on the most critical errors for this proof of concept.

The FTA shown in Figure 27 is considering the violation of SG5 ,"Prevent unintended braking on system limit". This is a collapsed view of the FTA, with each branch subsequently shown in detail from Figure 28 until Figure 38. For all views, pink events represent errors due to inherent algorithm insufficiencies, while blue events indicate errors due to triggering conditions or occurrence probabilities of relevant scenarios.

The FTA has been done using Medini Analyze 2024 R1 (ANSYS 2024).

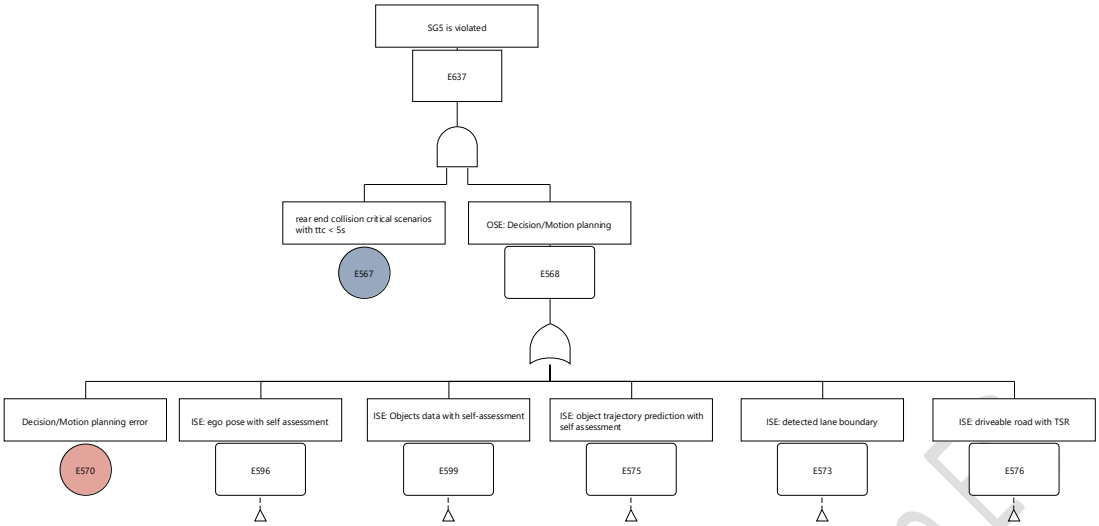

Figure 27: Top-level SG5 FTA

The first basic event seen at the top of the FTA (Figure 27) is the occurrence of relevant scenarios related to the safety goal violation. The use of relevant scenario occurrence rate can be seen in literature [20, 21] and in the SOTIF ISO Annex C [3] for error rate decomposition from higher-level metrics. Basic events due to relevant scenarios or SOTIF triggering conditions are shown in blue whilst basic events due to module inherent insufficiencies are shown in pink. All intermediate events (shown by rectangles) are perception related and a basic event has been used for the decision/motion planning error as this will not be decomposed further in this analysis with our focus being on perception. Output signal error (OSE) implies that the output of the module has an error. This can be due to two things: 1) The module itself produced an error or 2) there was an error in its input marked with input signal error (ISE). All of the events used in this FTA could cause the system to perform unintended braking (SG5 violation).

The rest of this subsection will cover in detail each intermediate event shown in Figure 27 showing the tree structure until basic events are reached and will explain how to assign error rates and provide recommendations on how to apply this to other safety goals. This is shown in the following table:

Table 23: Structure of the rest of subsection

Subsection segment name	Summary
ISE: ego pose with self-assessment	Figure 28 and Figure 29 show the tree structure for this intermediate event and expose the ego pose error basic event, radar triggering conditions and GNSS triggering conditions. More information about GNSS-based triggering conditions is provided here.
ISE: objects data with self-assessment	Figure 30, Figure 31 and Figure 32 show the tree structure for this intermediate event and expose the object data error basic event and radar and lidar triggering conditions. More information about radar and lidar triggering conditions is provided here.
ISE: objects trajectory prediction with self-assessment	Figure 33, Figure 34 and Figure 35 show the tree structure for this intermediate event and expose the object trajectory prediction error basic event, object tracking error basic event and its link to the ISE: objects data with self-assessment intermediate event.
ISE: detected lane boundary	Figure 36 shows the tree structure for this intermediate event and exposes the lane boundary error basic event and its link to vision based triggering conditions and ISE: ego pose with self-assessment.
ISE: driveable road with TSR	Figure 37 and Figure 38 show the tree structure for this intermediate event and expose the driveable road error basic event, TSR error basic event and radar and vision based triggering conditions. More information about vision based triggering conditions is provided here.
Error rate assignments and adaptations for other safety goals	Here we explain how to assign errors rates to the various basic events to achieve the acceptable vehicle level metric. Additionally, we provide some recommendation on how to use this method when considering other safety goals. Namely this involves finding different relevant scenarios and removing or adding relevant triggering conditions.

ISE: ego pose with self-assessment

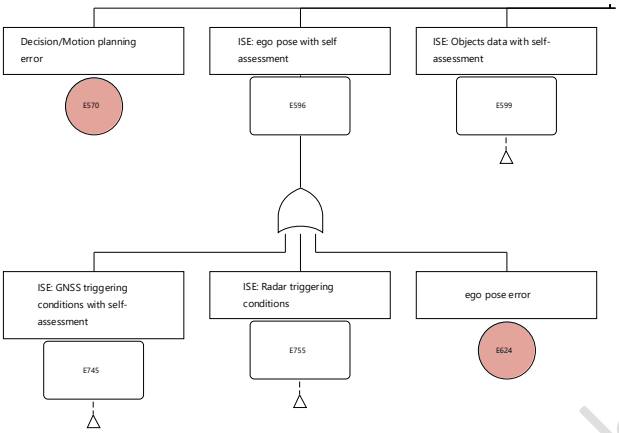

Figure 28: SG5 FTA with ego pose estimation error

Figure 28 shows the FTA for ego pose estimation errors. Jumps of the estimation of the ego-vehicle location on the map might lead to unnecessary braking to be applied by the system. The localisation self-assessment module is only considering errors due to GNSS, and that is why the self-assessment is only ANDed to that part of the FTA shown in Figure 29. The radar triggering conditions are also applicable to other modules and will be explained later on in the section.

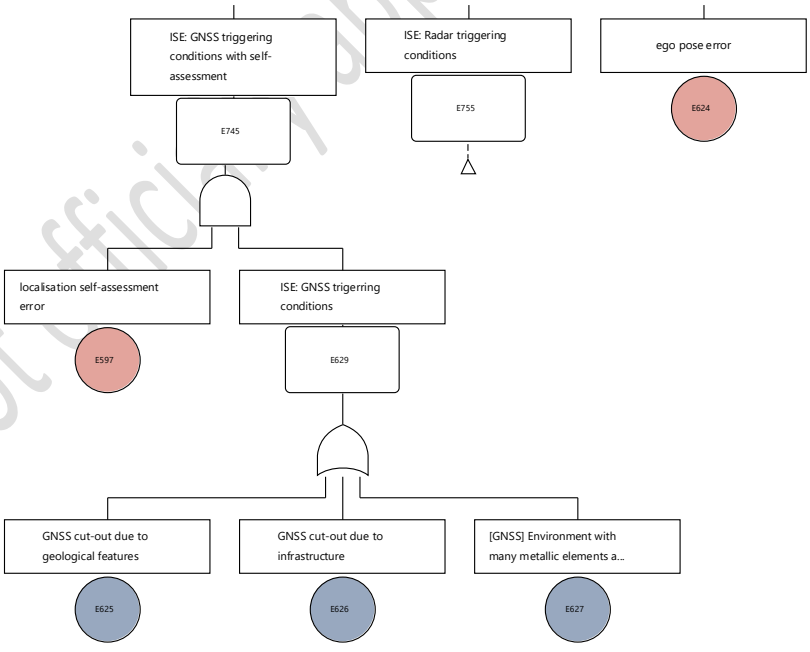

Figure 29: SG5 FTA with GNSS triggering conditions

Figure 29 shows triggering conditions that lead to an increase in the expected error rate of the ego pose (localisation module) related to GNSS. We are assuming that the

IMU should not have any significant SOTIF-related errors and is not considered in this study. The following triggering conditions are expected to lead to errors:

- GNSS cut-out due to infrastructure: When we are in urban canyons (narrow streets surrounded by high buildings [22]) areas or inside tunnels, the GNSS signal is not properly received by the ego-vehicle leading to inaccuracies.
- GNSS cut-out due to geological features: When we are in naturally occurring canyons, valleys or other naturally occurring geological features, the GNSS signal is not properly received by the ego-vehicle leading to inaccuracies.
- Environments with many metallic elements and objects: When inside these environments, we can expect degradation of the GNSS signal.

ISE: objects data with self-assessment

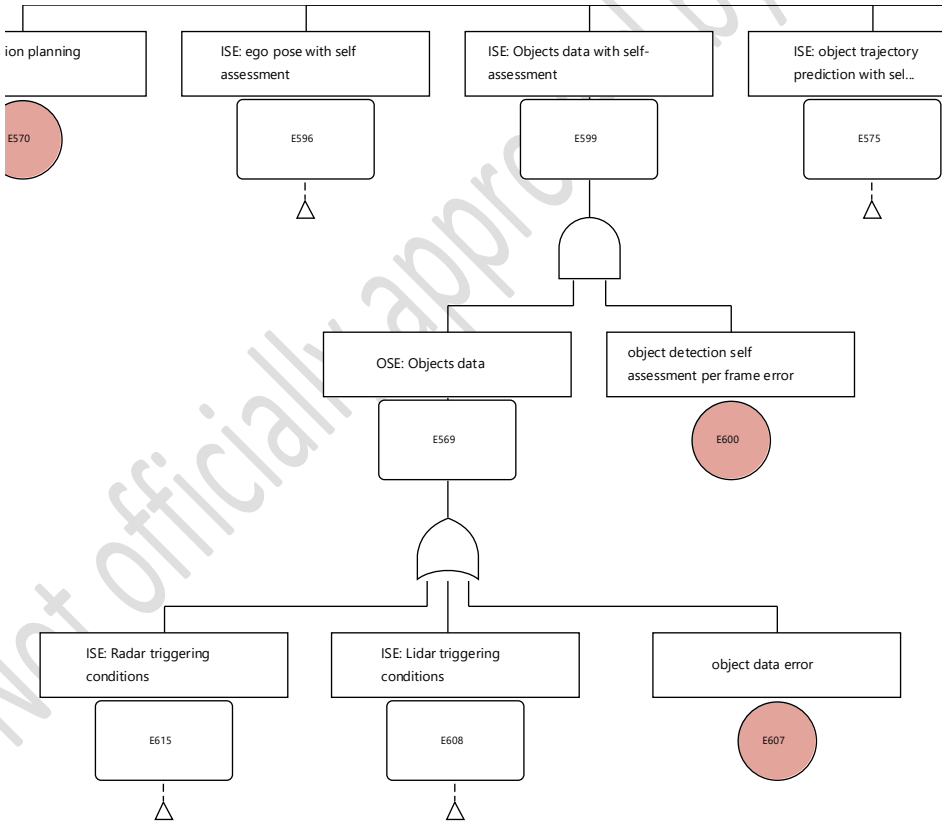

Figure 30: SG5 FTA with object data error

Figure 30 shows the expanded FTA structure for errors related to the object data with self-assessment. The correctness of the object data depends on the ego pose and also on the input sensing modalities as shown in the architecture diagram (Figure 24). It was decided not to consider ego pose errors in this part of the FTA as we can assume

that global localisation errors, which are more sensitive to triggering conditions, do not have a big influence on object data perception as the vehicle states in local coordinates obtained through the IMU should be enough. Therefore, we only considered the input sensing modalities' triggering conditions as part of the FTA. The FTAs for the triggering conditions are expanded in Figure 31 and Figure 32. Only radar and lidar triggering conditions are considered as vision is not currently used for object detection and classification in the EVENTS project.

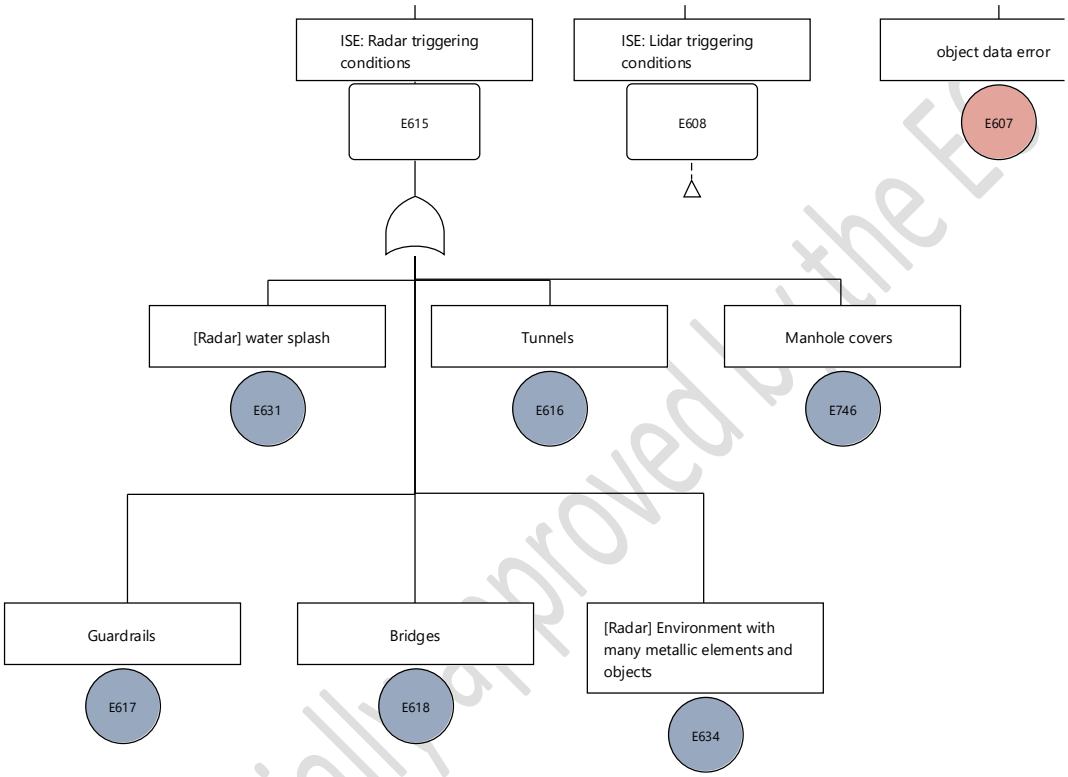

Figure 31: SG5 FTA with radar triggering conditions

The way that the FTA is built is assuming that each module has an inherent error rate due to the algorithm insufficiencies (shown by the pink event in Figure 31 above); here we would assign inherent error rate related to FP errors. Subsequently, we are assuming that those errors due to insufficiencies of the algorithm are exacerbated by triggering conditions (shown by the blue event in Figure 31 above) which increase the error rate via the OR gate; only the triggering conditions most likely to cause insufficiencies for this particular modality are considered. This approach permits the computation of two different error rates depending on whether the test dataset used for evaluation contains triggering conditions or not. Assuming the test set does not contain any triggering condition scenarios, then the inherent error rate should be used for metric estimation, while if the test data set contains triggering conditions, the summation of the two types of error rates can be used. This division is especially important since some triggering conditions can be rare and might not be readily

available in all datasets [23]. It may be noted that here, the input events to gates are considered to be independent of each other; this means all the triggering conditions depicted in the FTA are assumed to be independent of each other.

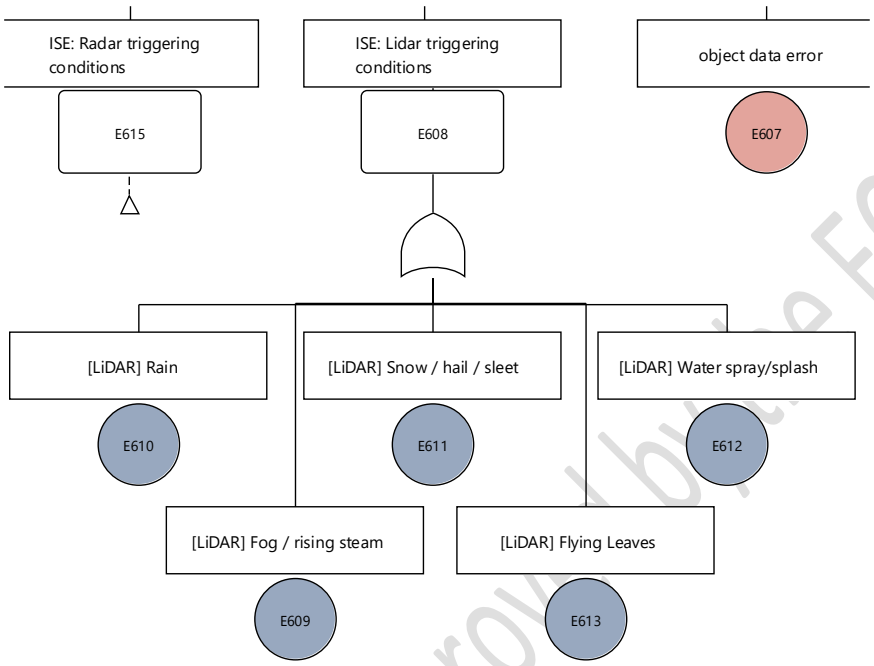

Figure 32: SG5 FTA with lidar triggering conditions

The triggering conditions shown in Figure 31 and Figure 32 are considered relevant to SG5 violation as they can all lead to detecting objects when there are none:

- Water spray (small droplets that remain suspended in the air for a long time before falling to the road resembling a fog cloud [24]) can create ghost objects for lidar as it creates additional detections, the higher the water level on the road the larger the number of detections that are observed [25]. Water spray and splash depend on many factors including water depth on the road which in turn depends on rain intensity and rain duration; these two factors could be used as a way to approximate the probability of occurrence of water spray and splash [26]. Water splash is different from water spray as it creates large drops that fall towards the road in a ballistic manner [24]. Having a wet road is a pre-requisite for water splash/spray so we can assume an occurrence probability between in the same order of magnitude as 1% and 10% which is the probability used for wet road according to ISO26262 [5].

- Flying leaves might cause false detections. To experience this, the ego-vehicle would need to be driving on a road lined with trees that drop leaves during the change of seasons.

- Tunnels can cause ghost objects due the reflections off the walls and the ceiling [27]. In the Netherlands, there are 10 km [28] of motorway tunnels (which constitute about 0.5% of all motorway length in the Netherlands) but this varies between different European countries. For instance, in Italy having there is a total 2600 km [29] of tunnels and the total length of motorway tunnels in which there was an accident in 2018 was about 199 km (so about 2.87% of the motorway length as there are 6943 km of motorway in Italy [30]). This shows the large variance between different countries so simply using one value for very different regions might lead to an overestimation or underestimation of the occurrence probability.

- Guardrails are a typical reflective surface for radar which can cause ghost objects [31]. They can also cause false negative detections.

- Bridges can cause FP detection if the object driving on the bridge are assigned as targets moving directly in front of the ego-vehicle or if the bridge itself is marked as an obstacle. Bridges can also cause FN detections at a distance.

- Metal dense environments can cause false detections or false negative for radar.

- Snow/hail/sleet: Snow can cause both false detections and missed detections for lidar [32].

- Rain can cause false detections in the space between the lidar and the target [33]. The number of days per year with at least 1 mm of precipitation for Europe (50 countries) is around 100 to 150 days whilst the number of heavy precipitation (>20 mm) is much lower for 2021 with around 6 days for most of the region and some specific regions getting around 18 days and some smaller regions getting 30 days [34]. Therefore, the probability of occurrence varies depending on the region with heavy rain between 1.6% and 8.2%.

- Fog/rising steam can cause attenuation and backscattering of the lidar signal [35] leading to a reduction in detections (relevant to FN) and also noise due to reflections caused by fog (relevant to FP) [33]. The mean number of days with visibility less than 200 m in some European countries in winter can go up to 15 days per half year (8.2% of 365 days) with summer months having most countries with 3 days (1.6%) per half year (computed over 1976 and 2006 period); the number of days depends on the European country and time of the year [36].

A study [37] presented a 10000 km dataset collected by driving an instrumented vehicle in Germany, Sweden, Denmark, and Finland in winter of which 0.7% was collected in dense fog (<100 m visibility), 1.6% in rain conditions and 26.2% in snow conditions. These numbers can also be used as an estimate for the triggering conditions occurrence probability for these conditions specifically in winter. Looking at the ISO21448 mission profile [3], they assume 5% for fog, 5% for snow, 5% for heavy rain, 7% for rain. This shows that it is difficult to get one specific value for an occurrence probability and that a range is more suitable per region. If the system is expected to work in multiple regions, the worst case scenario should be considered or scaled by the expected time the system would be spending in each region.

Most of these triggering conditions are mentioned in the SOTIF standard table with scenario factors (Table B.3 [3]). A large number of triggering conditions are related to weather conditions; this is not surprising and is also corroborated by literature as previous studies [38, 39] have found that bad weather conditions are linked to self-driving system failures.

However, the occurrence of a triggering condition does not necessarily mean that an error will occur. Apart from the occurrence probability of such condition, it is also required to include a probability that the triggering condition leads to an error in the specific module. For instance, from the above list of triggering conditions we notice that guardrails are very common in a typical ADS ODD on highway implying this will have a very high occurrence probability. However, the final error probability needs to be the multiplication of this occurrence probability and the likelihood that it would cause a ghost object. Additionally, this means that the influence of the triggering conditions on each module might be different and that is why even though the name of the source of the error might be the same (e.g. guardrail) different events are used in the tree. For instance, the same triggering condition might have a different error factor on the ego pose or the object data or the driveable road estimation, so they are in fact different events in our tree.

ISE: objects trajectory prediction with self-assessment

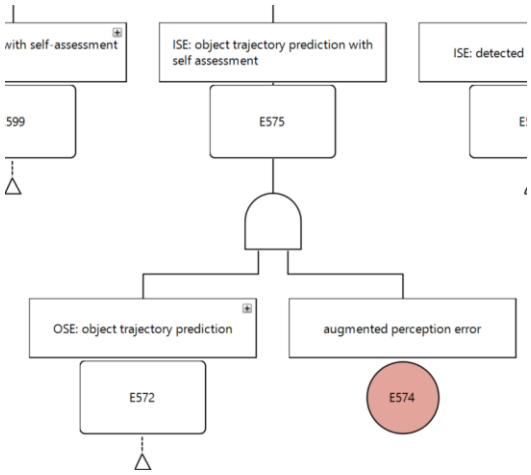

Figure 33: SG5 FTA with object trajectory prediction error with self-assessment

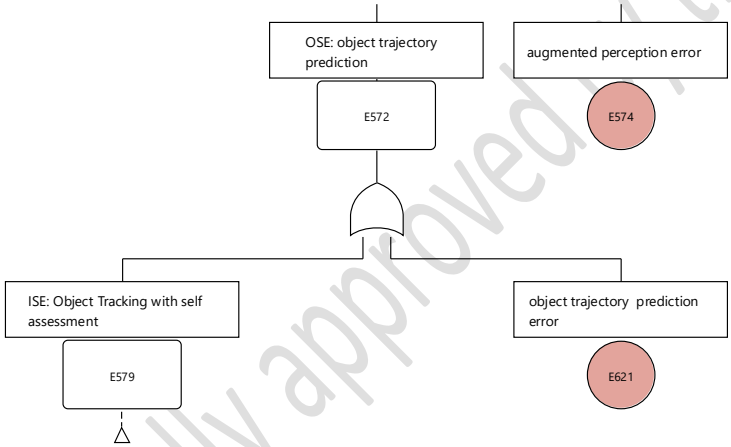

Figure 34: SG5 FTA with object trajectory prediction error

Figure 33 and Figure 34 show the FTA structure for the object trajectory prediction error. An incorrect object trajectory prediction might lead to unnecessary braking applied from the system, thus leading to SG5 violation. For this to happen, both the prediction and the self-assessment provided by the augmented perception would need to fail. Object trajectory prediction errors might be caused inside this module or through any of its inputs, thus the need to include object tracking errors in the FTA. Object tracking is not directly used by the decision/motion planning module and thus its effects are confined to the object trajectory prediction errors. The FTA for object tracking can be seen in Figure 35. Object tracking depends on object data so they are part of its FTA structure and will inherit its error.

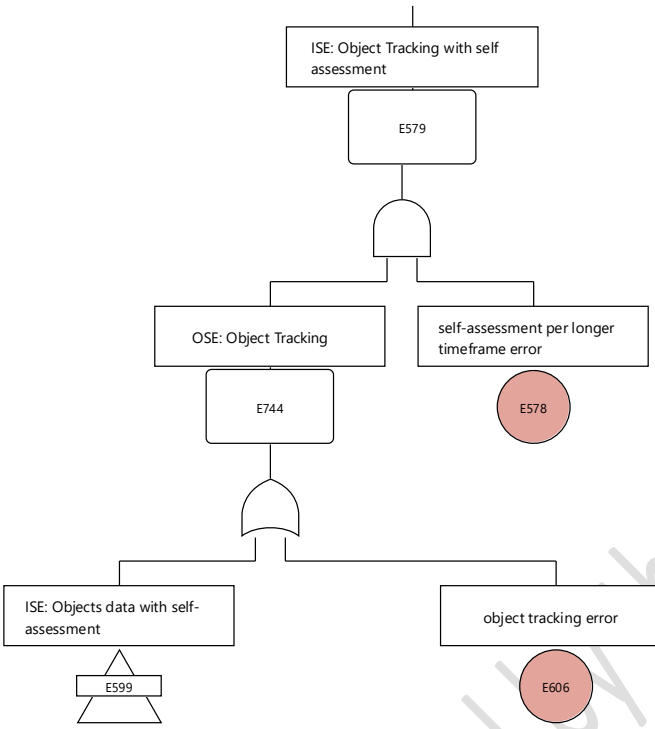

Figure 35: SG5 FTA with object tracking error with self-assessment

ISE: detected lane boundary

Figure 36 shows the FTA for the detected lane boundary error; this error is included as a mistaken lane boundary might lead to erroneously assigning a perceived object to the ego lane, leading to unnecessary braking. Incorrect global positioning might also impact lane assignment so it is shown in the FTA. We added an AND gate to show we can add a scaling factor to the influence of an ego pose error on the lane boundary error as this gives the option to reduce its impact if the module already has some safeguards in place. This part of the tree structure does not include any self-assessment.

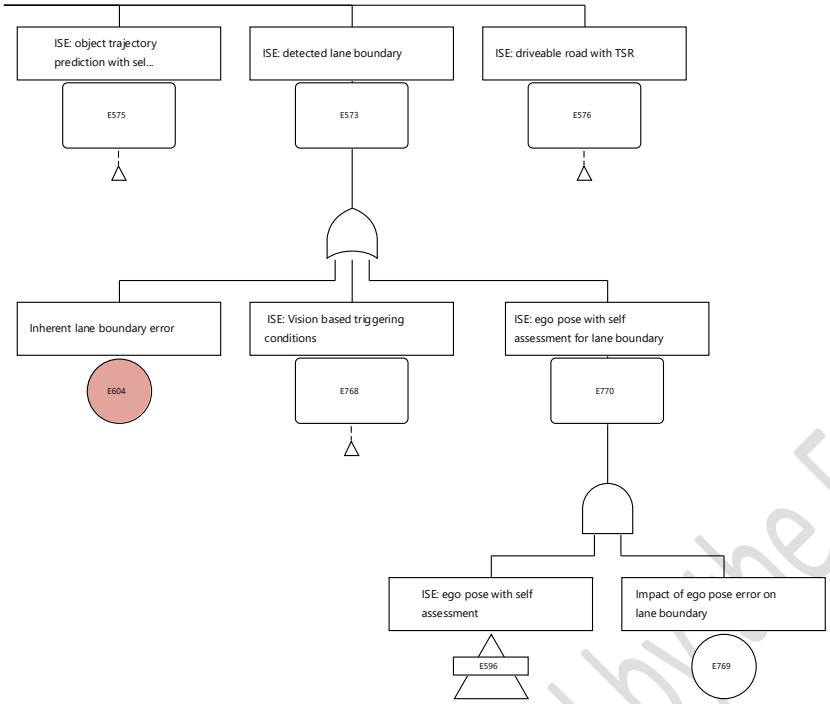

Figure 36: SG5 FTA with detected lane boundary error

ISE: driveable road with TSR

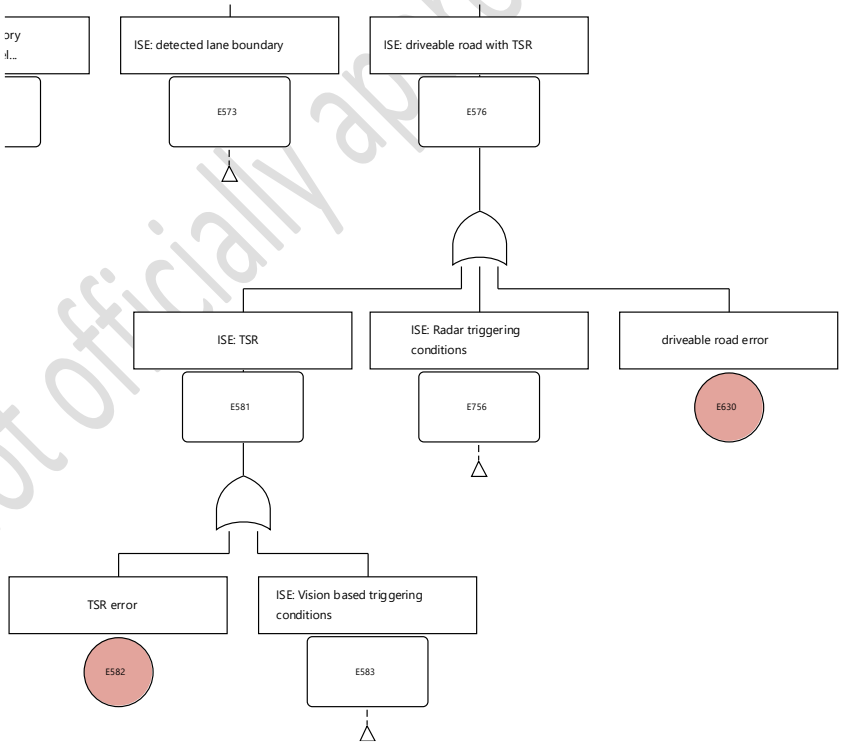

Figure 37: SG5 FTA with TSR driveable road error

Figure 37 shows the FTA structure for the driveable road area module with traffic sign recognition. An error in this module would lead to erroneously braking for some

detected road infrastructure which is not there or could lead to harshly braking to follow an incorrect traffic speed limit sign. This part of the tree structure does not include any self-assessment. This module depends on both radar and vision in the EVENTS project thus including both kinds of triggering conditions in the FTA structure. Figure 38 shows the FTA structure for vision triggering conditions. The considered conditions are:

- Fog/rising steam, rain, snow/hail/sleet and darkness, all of which can disrupt visibility range [40] which is directly applicable to FN errors. These triggering conditions can also lead to FP and FN errors as the quality image is heavily degraded leading to incorrect lane marking detections (e.g. snow deposits on roads [32]) and incorrect speed sign recognition (reporting a lower speed than the real one).

- Rain at night with artificial lighting is a challenge for vision resulting in an increase in lateral position error outliers and wet roads are in general challenging due to the water reflecting the light [41].

- Excessive amount of artificial light sources and sudden bright light might lead to contrast issues, both of which cause image degradations leading to both FN and FP errors.

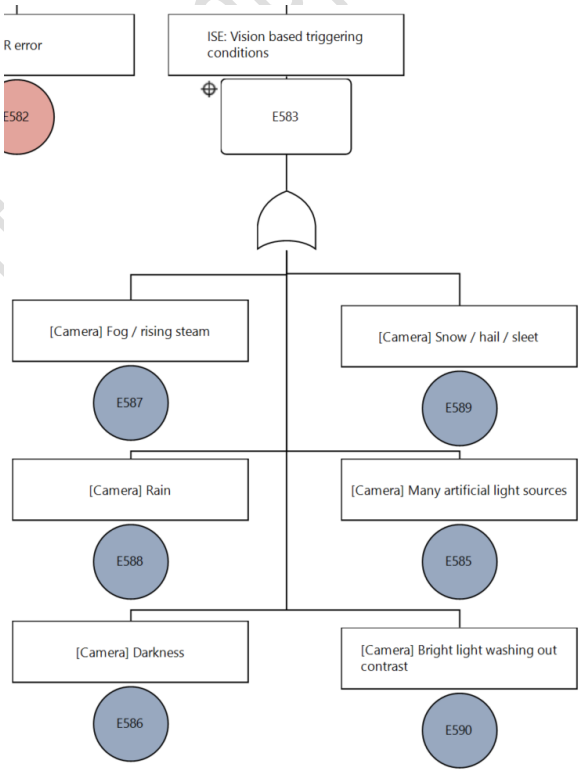

Figure 38: SG5 FTA with vision triggering conditions

Error rate assignments and adaptations for other safety goals

Once the FTA is constructed, the error rates for each of the basic events needs to be populated. This can be done as follows:

1. A value for the blue events in the FTA is estimated via literature review (or data analysis if data is available). This can be a very challenging task as it can be difficult to find the correct statistics in available literature and any publicly available reports. It is important to note that simply putting the occurrence probability of a certain triggering condition is not enough. The actual probability that it would cause an error in the system should instead be put in the FTA (multiply the occurrence probability by the probability it will cause an error in the module).

2. The vehicle level error rate (safety goal violation metric) is distributed amongst the various pink basic events shown in the FTA. The error rate can be divided equally or using an importance weighting depending on the influence of the module to the final safety goal violation. It is important to note that this error represents the expected error without triggering conditions influencing the module. Finally, any probabilities used to show the impact of modules on each other need to be set (e.g. white basic event in Figure 36).

3. The FTA probabilities are evaluated.

4. Step 2 and 3 are repeated with adjustments to the distributions until the safety goal violation metric is equal or less than the given acceptable vehicle level error rate in the beginning of this deliverable. In the case that this is not achieved, additional safety measures are put in place to reduce the impact of any of the basic events.

The approach was explained in detail for SG5. A similar approach can be done for other safety goals by changing the relevant scenario event at the start of the FTA and only considering the modules relevant to that specific safety goal violation.

SG6 is about not braking enough or not braking at all, which means it has a strong similarity to SG5 as the relevant scenario in Figure 27 would still apply. The main difference compared to the FTA for SG5 would be the nature of the error on the modules. For SG5, we are interested in errors that would cause objects to be assigned as our lead vehicle when either they are not or there is no object there while the opposite is true for SG6. For SG6, we are not braking for a lead vehicle that is there, either because it was not detected or not correctly assigned to our lane. Going step by step through the FTA for SG6, we notice that the main differences would boil down to the relevant triggering conditions. For instance, for the radar triggering conditions

we would need to be remove from the FTA the manhole cover (radar triggering condition) as this usually do not cause missed detections (Figure 31).

The following triggering conditions would need to be added for all sensing modalities:

- Sudden tight turns, which might lead to object suddenly appearing in FOV not being tracked immediately but with a delay (FN error).
- Mud, which causes partial blockages which if not immediately flagged by the self-assessment module might also lead to a FN error.

Additionally, we would need to add the sensor visibility self-assessment as this is relevant for SG6.

SG7 is about applying too much acceleration and the reasoning outlined above for SG6 would also apply here as missing a lead vehicle is one of the possible causes of too much acceleration.

SG4 is about applying too much steering leading to lateral collisions. This is different from the previous goals and would require a few changes to the FTA. The first change to be done would be in Figure 27 as a new relevant scenario would be needed. Weather related triggering conditions should still be applicable. For example, a benchmarking report found that simulated rain resulted in 69% of the runs exiting the current lane [42]. Additionally, any lane marking degradation due to bad visibility or snow deposits would influence this safety goal.

6.1.3 Correlation of module error rate to perception validation targets

Error definition

The first step to correlate an error rate to a validation target for a certain module is to define what is an error for that specific module. A error from a module should be something that can lead to an error at the vehicle level. For the scope of this deliverable, which is a safety analysis proof of concept, we will focus on errors which come from FN and FP perception outputs; these kind of mistakes can lead to critical outcomes [19] but other kind of mistakes should also be analysed in a complete SOTIF analysis of a system. However, a single mistaken frame does not directly imply a error; another additional condition to consider is the duration of these mistakes. An error can be defined as a sequence of consecutive mistakes as is implied in ISO/PAS 8800 Section 9.5.1 [43].

Definitions of perception error are as follows:

- A FN error is when the system missed a critical target for more than x seconds.

- A FP error is when the system created a critical target when one does not exist for more than x seconds.
- A FN self-assessment error is when the system missed to report error for more than x seconds.
- A FP self-assessment error is when the system reported an error when one does not exist for more than x seconds. However, this does not lead to a safety concern as it would simply revert the system to the minimum risk manoeuvre which is the safe state of our system. Therefore, this error will not be considered for the safety analysis.

Looking at the L3 regulation R157 [14] for automated driving systems, we can obtain an estimate for a reasonable estimate of the duration of the error. In the regulation, it is assumed that for the competent and careful driver model it takes the driver 0.4 s to perceive a risk and 0.75 s to start acting on the risk. Therefore, we decided to set the critical fault duration to be 0.5 s as our automated system should then be able to do decision making much faster than 0.75 s. Additionally, looking at the simulation results seen in Figure 18, we can see that all simulations with a time gap of 2 s end in collision for delays greater or equal to 0.5 s further showing that 0.5 s is a good candidate for a FN error threshold value. From the simulation and theoretical analysis in 5.3.3, we found that a threshold value of about 2 s would be more applicable for FP errors.

Scaling of the error

The above error definition mentions the word “critical”. This means that to transform it to a perception KPI which uses the full field of view of the sensors we would need to scale it up. The critical region of interest depends on the specific safety goal which for instance would be the zone in front of the ego-vehicle for SG5, SG6 and SG7.

Assuming the metrics cannot be computed solely for the critical zone but have to be computed on a 360 degree FOV around the ego-vehicle (refer to Figure 39), the error can be scaled accordingly permitting a larger number of errors per hour. The scaling can be done by multiplying the error rate for the critical region derived using the FTA and then using it as the input to Step 4 below. The error can be higher in the non-safety critical region, for instance an additional factor of 2 could be used, thus permitting twice the amount of errors in the non-critical regions thereby increasing the overall permissible error rate. However, this would not guarantee the correct error rate in the critical region and thus metrics looking specifically at the critical region are preferrable.

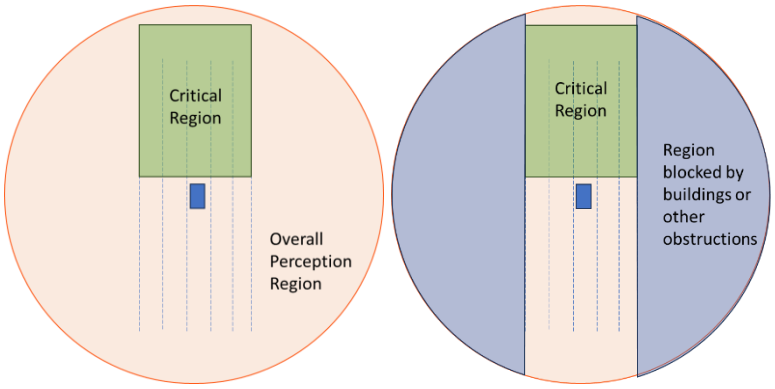

Figure 39: Different scaling configurations for the error rates

Error rate and probability conversion

For our analysis, it is convenient to also have a value for the probability of error alongside the error rate. A constant error rate can be converted into a probability by using an exponential distribution [44]. For instance, taking the error rate of SG5 (5.12E-7/h) from 3.3, we get an error probability p by setting λ = 5.12E-7 in the following equation:

$$p = 1 - e^{-\lambda t}$$

Equation 1: error probability

The variable t represents time so the probability of error increases with time. For instance, if we would like to know the probability of error for 1 hour, we would set t = 1 which would result in a probability of 5.15E-7. Additionally, assuming a total lifetime of a car, for instance, of around 15 years with an average yearly driving time of 366 hours [8], we can estimate a total lifetime of 5490 hours per car. Setting t = 5490, we would get a failure probability of 0.0028.

Error rate to recall/precision mapping

Once we have a definition for error we can link the error rate derived per module using the FTA to the specific mistake metric of interest. For instance, the error rate could then be linked to the number of FN frames per hour of driving or FNR. The following steps need to be done to obtain the FNR metric:

1. Estimate how many true critical targets would be observed in 1 hour of driving.
2. Set a false negative rate (FNR), i.e. the probability that there is a missed detection defined by no_missed_detections/no_of_ground_truth_detections (FN/(FN+TP)).
3. Estimate the correlation between probability of error or error rate to the FNR. The higher the sampling rate of the object detection, the higher the permitted

FNR for a specific error rate. For this study, we assumed a sampling rate of 10 Hz as lidar is used as sensing modality for most of the object detectors. This means we set N=5.

4. Repeat Steps 3 and 4 to find the highest FNR that satisfies the acceptable error rate.

A similar approach can be used to estimate the precision value with the main difference being that the equation for Step 2 and 3 above would be FP/(FP+TP) (or false discovery rate). Additionally, instead of N=5, we can set N= 20 as we are assuming it would take 2 s of consecutive FP mistake to cause an error.

For this work, we made a rough assumption that mistakes (FN or FP) happen independently between two frames. With this in place, we computed the correlation between the error rate/probability of error and the metric of interest (FNR or false discovery rate) using 2 methods:

1. **Theoretically**: Computing the probability of at least one error using the recursive equations found in [45]. The recursive method computes for each new draw the probability of not getting a consecutive success using the probability of not getting consecutive successes for the previous 2 draws. The number of draws is set to the total number of estimated critical objects in all captured frames in one hour n_p. This means we are computing the probability that in 1 hour of driving we have one or more errors.

2. **Simulations:** Running simulations by drawing a success or error using a Bernoulli distribution with the probability of error set to the chosen metric. The simulation simulates an hour of driving by drawing a random number multiple times equal to n_p. This is then repeated for 3 million times (simulating 3 million hours) to get a more accurate value for the error rate (no of errors/no of hours). This 3 million is estimated using the equation presented in 7.1 to compute the number of samples required to capture a specific error percentage with a given degree of confidence; here we used a confidence of 95% and used an error probability of 1E-6 with the reliability set to 1-1E-6. This also agrees with estimating a population proportion of 1E-6 mean error rate using binomial confidence intervals (Clopper-Pearson Method [46, 4]) as we get a lower bound of 0 and upper bound of 1.23E-6 when assuming the case when no errors are observed.

We are assuming there are 100000 instances of a critical vehicle in an hour when using a cycle time of 10Hz; therefore n_p is set to 100000. The 1000000 instances were estimated by considering that in [47] there were 2 objects per frame (0.1 missed

object per frame with a recall of 95% leads to 0.1*0.95/0.05 + 0.1 objects using TP/TP+FN) in the critical zone (36 m in front of ego-vehicle by 17 m width), scaling this (72000 object in 36000 frames assuming 10 Hz) for a 50 m critical zone, we get 100000 objects.

For different traffic densities and ODD, we might have a different amount of positive class occurrence. Changing the positive class per hour, we can analyse the variation in the KPI value for precision and recall.

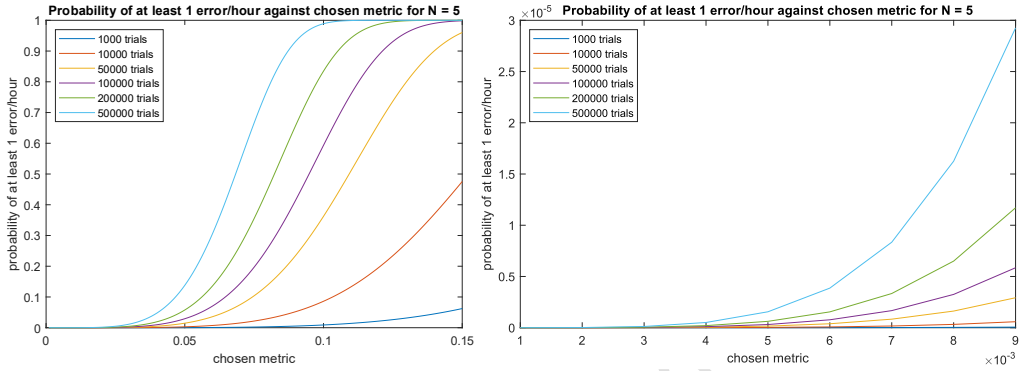

Figure 40: Theoretical probability computation with N=5. The figure on the right is a zoomed-in version of the figure on the left.

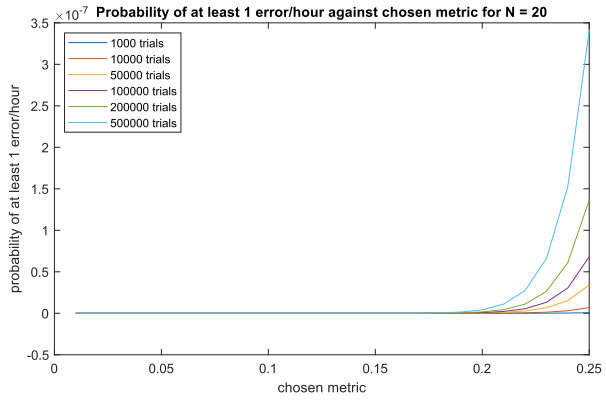

Figure 41: Theoretical probability computation with N=20.

Figure 40 and Figure 41 show that the higher the amount of trials, which in this case represent the number of critical targets (positive class), the higher the probability an error might occur. This method is based on our assumption that the FNR or false discovery rate ratios are related to the actual probability of making an error.

It can also be seen that the higher the duration of the consecutive mistake, the less likely an error can happen. For instance, if we are assuming that we need to have a 2 s FP error to cause any safety critical event (Figure 41), any precision value larger or equal to 83% would have an error probability on the order of 1E-13.

For smaller values of error duration (Figure 40 assuming chosen metric is FNR), we can see that any TPR value smaller or equal to 87.2% would result in the occurrence of 5 consecutive mistakes (1 error) in an hour when presented with 500,000 critical targets. The TPR is more forgiving for the 100,000 case with a value of 82.2%. Therefore, any TPR close to these values would not be acceptable. On the other hand, for the same number of critical targets, to get a probability of error close to 1E-10, we would need to have a TPR of 99.9%. For a probability of error close to 3E-7, we would need a TPR of 99.5% (Figure 40 right) and a TPR 99.3% for an error close to 1E-6 (on the same range of the probability requested by the safety acceptance criteria); these are quite challenging TPR values to achieve.

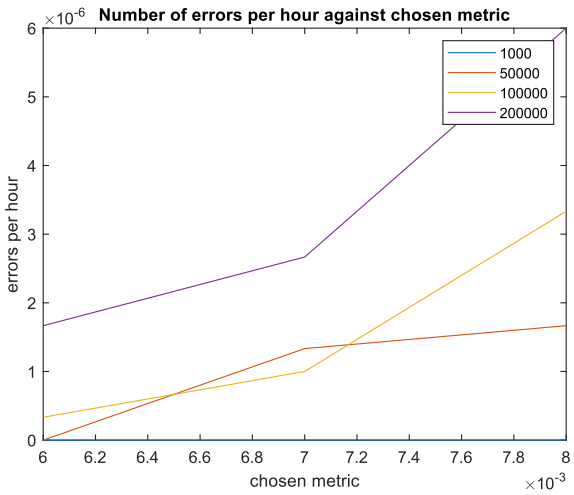

Figure 42: Results from 3 million repetitions with N=5 for different number of targets/hour.

Figure 42 shows the estimated mean error rate per hour using the simulation method. Comparing it to the probability of error estimated using the theoretical method, we can see that similar results are obtained as shown in the Table 24 below:

Table 24: Comparison between the theoretical and simulation estimates.

FNR	Positive Class per hour	Simulation Mean Error per hour	Error per hour (inverse of Equation 1 $\lambda = -\ln(1-p)$)
0.006	100000	3.33e-7	7.73e-7
0.007	100000	1e-6	1.67e-6
0.008	100000	3.33e-6	3.25e-6
0.006	200000	1.67e-6	1.55e-6
0.007	200000	2.67e-6	3.34e-6
0.008	200000	6e-6	6.5e-6

The theoretical result tends to overestimate the error rate compared to the simulation-estimated mean, but both are similar so the theoretical approach can be used as it requires fewer computations.

It is important to note that both the mathematical and simulation approach have limitations. Additionally, they are both assuming each frame is independent from the previous frame, which might not hold in real life scenarios. More sophisticated methods will be explored in future work.

Once a concrete FN or FP metric is obtained, we match it with the KPIs considered in EVENTS Deliverable 6.1 [48]. Examples of this are shown in Table 25 and Table 26.

Table 25: Perception metrics that could be quantified using this methodology

Perception module	**Performance metrics**	**Relationship to derived recall or precision metric**
Object data (pose and classification)	Average Precision (Exp1) Recall/Precision (EXP 6)	Direct relationship to the estimated Recall and Precision. For AP the best Precision Recall combination could be used.
Object Trajectory prediction	Miss rate (EXP 5), manoeuvre classification FP/FN metrics (EXP7)	Incorrect prediction could lead to wrong lane assignment. There is a direction relationship to the estimated Recall and Precision.
Detected lane boundary	Distance error of lane boundary (EXP 4)	If this KPI is given a large distance error threshold it can be linked to the estimated Recall and Precision in this section. Large errors in lane boundary can lead to incorrect lane assignment.
Driveable road with TSR	TSR FP FN metrics (EXP 4), Accuracy of segmentation via IoU% (EXP8)	Direct relationship to the estimated Recall and Precision for TSR. The relationship of IOU to FP and FN metrics would need to be extracted (this would require further work).
Localisation error	Large mean squared error (EXP8)	If this KPI is given a large distance error threshold it can be linked to the estimated Recall and Precision in this section.

Table 26: Self-assessment metrics that could be quantified using this methodology

Self-assessment module	Error	Relationship to derived recall or precision metric
Object detection per frame	FN self-assessment fault (Error detection rate EXP 7)	Direct relationship to the estimated Recall.
GNSS-based self-assessment	Horizontal protection level larger error (EXP 7)	If this KPI is given a large distance error threshold it can be linked to the estimated Recall.
Object tracking per longer timeframe	FN metrics (Error detection rate EXP3)	Direct relationship to the estimated Recall. Self-assessment FP not considered.
Augmented perception	FN metrics (EXP 2)	Direct relationship to the estimated Recall. Self-assessment FP not considered.

New safety KPI definitions

The easiest and more accurate way to make sure that the KPI satisfies the safety goal would be to revise the KPI definition to match the error definition as explained above. This is illustrated in the following KPI definitions:

1. Number of N consecutive critical FPs per hour should be less or equal to the number found in the FTA.
2. Number of N consecutive critical FNs per hour should be less or equal to the number found in the FTA.

The term “critical” means that the error needs to happen in the critical region for that specific safety goal as mentioned above.

Assuming that during the EVENTS projects the amount of driving hours in the evaluation dataset is limited, the KPI can only be satisfied with a specific value of confidence (refer to 7.1).

Additionally for datasets that are made of specific scenarios and not continuous driving, the failure rate from the FTA can be transformed into a probability value (as seen above) which then would transform the recommended KPI definition as follows:

1. The percentage of scenarios with an error (N consecutive critical FP) should be less or equal to the number found in the FTA.

2. The percentage of scenarios with an error (N consecutive critical FN) should be less or equal to the number found in the FTA.

This assumes that there is at most 1 error in each scenario, which should be reasonable as scenarios time lengths are on the order of a few minutes.

6.1.4 Decision and motion planning validation target metrics

From literature, it was found that planning modules can use a framework like "Responsibility-Sensitive Safety" (RSS) [19] that can mitigate errors when assuming that they are receiving a correct perception input [20].

The following equation [19] can be used as the target value for the minimum longitudinal distance to obstacles mentioned as evaluation metric in D6.1 (EXP1, EXP8) or as a starting point to any distance related metric (EXP2, EXP4):

$$d_{\min} = \left[v_r \rho + \frac{1}{2} a_{\max,\text{accel}} \rho^2 + \frac{\left(v_r + \rho a_{\max,\text{accel}}\right)^2}{2a_{\min,\text{brake}}} - \frac{v_f^2}{2a_{\max,\text{brake}}} \right]_+,$$

where v_r and v_f are the velocities of the rear and front vehicle, ρ is the response time for maximum braking, $a_{\max,\text{brake}}$ of the front car and for a maximum acceleration $a_{\max,\text{accel}}$ of the rear car. The rear car can be assumed to be the ego-vehicle, which will then brake using $a_{\min,\text{brake}}$. Therefore, assuming the values for each of the velocities and accelerations can be computed/measured or assumed, one can compute the minimum safe distance for any scenario with 2 cars.

An equation for the lateral distance is also provided as follows:

$$d_{\min} = \mu + \left[\frac{v_1 + v_{1,\rho}}{2} \rho + \frac{v_{1,\rho}^2}{2a_{\min,\text{brake}}^{\text{lat}}} - \left(\frac{v_2 + v_{2,\rho}}{2} \rho - \frac{v_{2,\rho}^2}{2a_{\min,\text{brake}}^{\text{lat}}} \right) \right]_+,$$

where v_1 and v_2 are the lateral velocities of two neighboring cars, μ is the final lateral distance, $v_{1,\rho} = v_1 + \rho a_{\max,\text{accel}}^{\text{lat}}$, $v_{2,\rho} = v_2 - \rho a_{\max,\text{accel}}^{\text{lat}}$, $a_{\max,\text{accel}}^{\text{lat}}$ is the lateral acceleration, and $a_{\min,\text{ brake}}^{\text{lat}}$ is the lateral braking value.

6.1.5 Other evaluation metrics

While we considered a subset of evaluation metrics for the subsystems, it is important to note that this activity can be expanded to consider other evaluation metrics. The following table includes other relevant evaluation metrics for ML-based ADFs (some of these metrics where already presented in D6.1 [48] but are still shown for completeness sake):

Table 27: Evaluation metrics for ML-based ADFs

Category	Metric name	Description	Applies to	Unit / Format
Perception	mAP (mean Average Precision)	Average precision over multiple classes and IoU thresholds	Object detection	Percentage (%)
	IoU (Intersection over Union)	Overlap between predicted and actual regions	Semantic segmentation	Value between 0 and 1
	False Positive Rate (FPR)	% of times the model incorrectly detects an object	All perception modules	Percentage (%)
	False Negative Rate (FNR)	% of missed detections (critical for VRUs)	All perception modules	Percentage (%)
	Latency	Processing time per frame	Sensor fusion / perception	Milliseconds (ms)
	Uncertainty Calibration (Brier Score)	Measures how well predicted probabilities reflect true likelihoods	Probabilistic models	Value between 0 and 1
Planning	Success Rate	Percentage of successful, goal-reaching episodes	Trajectory planning	Percentage (%)
	Collision Rate	% of scenarios that ended in a simulated or real collision	Motion planning	Percentage (%)
	Time-to-Collision (TTC)	Time remaining before predicted collision under current trajectory	Planning / control	Seconds (s)
	Path Deviation	Deviation between planned and actual path	Trajectory tracking	Meters (m)
Control	Tracking Error	Mean absolute error between reference and	Lateral/longitudinal ctrl	Meters (m) / Degrees (°)

		executed control values		
	Control Overshoot	Extent to which output exceeds target	Steering/braking control	% or value in units
	Settling Time	Time taken for system to stabilize after a command	Control	Seconds (s)
System Safety	Functional Insufficiency Rate	Frequency of ML-induced insufficiencies leading to risk	System-level SOTIF eval.	Events per hour
	ODD Sensitivity Score	Performance variation across different ODD slices (e.g., fog, rain, night)	Entire stack	Qualitative / % degradation
	Adversarial Robustness Score	Sensitivity to adversarial input perturbations	Perception / control	Accuracy drop / %
	Scenario Coverage Score	% of known scenario categories tested	All	Percentage (%)
Explainability	SHAP/ALE Influence Score	Feature attribution strength for critical decisions	ML-based planning	Relative ranking / plot

6.1.6 Example of module level metric derivation

Here we will show an example of how to use the quantitative FTA method to provide a goal value for EXP6. For EXP6, the worst safety related mistake would be to detect an overdriveable object as an obstacle (a FP) as this would lead to unnecessary braking. To demonstrate our decomposition we will be giving arbitrary error values to the events shown in our fault tree in Section 6.1.2.

A value of 5.12E-7 errors per hour is the target for the top level event as per Section 3.3 and a probability of 20% was given to the top scenario of interest (obtained from Table 1 of [20]). All the blue events which were in the path of a self-assessment block were given a value of 1E-7 error rate per hour and the pink modules 1E-6. All the blue events which were not in the path of a self-assessment block where given a value of 5E-8 error rate per hour and the pink events 5E-7 (lane boundary error and ego pose

error basic events). Driveable road error and TSR error (E630 and E582 respectively) were set to 2.5E-7 and 2E-7 respectively as they were assumed to be less important for the safety goal. The ego pose error impact event (Figure 36 E769) was given a probability of 0.1 to reduce the impact on the lane boundary error FTA. All self-assessment blocks were set to 1E-6. The top decision/motion planning error pink event was set to 1E-9 (small arbitrary value given as it was not the focus of our current FTA as explained in previous sections). Through this example it is already clear the importance of having self-assessment modules as without them the available error rate for modules and acceptable susceptibility to triggering conditions is diminished making the required performance even more stringent. Giving these error rates to the events in the tree the high level error metric was achieved (value close to but smaller than 5.12E-7 was obtained for safety goal event). The mission time used to compute all the probabilities was set to 10000 hours on Medini Analyze.

The FP error rate per hour available for the object data module is 1E-6. This value needs to be in turn divided by the possible FP errors that could happen, one possibility is have the object detection block output an obstacle where there is none and another is for the overdriveability classifier to classify an overdriveable object as an obstacle. As the occurrence of the latter event is less likely than a generic FP a smaller budget is given to EXP6, with 1E-7 error. In literature we see that perception level errors derived from vehicle level errors keep the same order of magnitude (e.g 1E-7 for vehicle level and 5E-7 for perception level [20]). This shows our method agrees with literature and also provides for more detailed decomposition as here we started with a vehicle level metric of 5.12E-7 and ended with a perception module error of 1E-6 which was further reduced for the overdriveability use case. The order of magnitude decrease seen in our method is due to the self-assessment blocks assumed in the FTA.

The safety KPI for EXP6 would become the following:

The percentage of scenarios with an error (40 consecutive FP in critical zone in front of ego-vehicle e.g. 50 m) should be less or equal to 0.1%.

N=40 was computed by assuming a consecutive FP error of 2 s with a sampling rate of 50 ms. The value of 0.1% was achieved by assuming the probability of failure over the lifetime of 1 car (set to 10000 hours in this FTA study example) and assuming the scenarios considered in the test set cover such a lifetime. This might be a fair assumption in the debris use case as we are not expecting to encounter many debris in the lifetime of a car.

This KPI is very stringent when only assuming a radar-based perception as a few cm mistake in the height component of the object might result in a sustained FP classification; even having two such problematic scenarios out of a 1000 scenarios would fail this KPI. In order to achieve this KPI other sensing modalities such as camera

or lidar would be required in order to confirm whether an obstacle is in front of the ego-vehicle. Additionally, it is important to note that the available errors were distributed arbitrarily amongst all the modules and triggering conditions events without any specific importance scaling so to the true error budget for this KPI could be in fact higher than the one shown here. This value is derived to show our proposed method in practice and is not considered binding for WP6.

The KPI already provided in D6.1 for EXP6 is the precision; assuming we only consider it for the critical zone of 50 m we could use the method shown in Section 6.1.3 to go from the failure probability of 1-exp(-1E-7)= 1E-7 to a false alarm rate of 0.5 which would imply a goal value of at least 0.5 (see Figure 43). We also run the analysis assuming the whole lifetime of a car (1 billion trials assuming 100000 critical objects) and this still shows that a false alarm rate of 0.5 (see Figure 43) would be enough to satisfy the error probability (0.001 mentioned above). This is quite a low and easily achievable value as here we are assuming each frame is independent and that the probability of making a FP estimation is governed by the false alarm rate. Unfortunately, both assumptions could be incorrect especially when the N of interest is large like in this case. We recommend thus to redefine the KPI to match the error which is safety critical as was done in the previous paragraph.

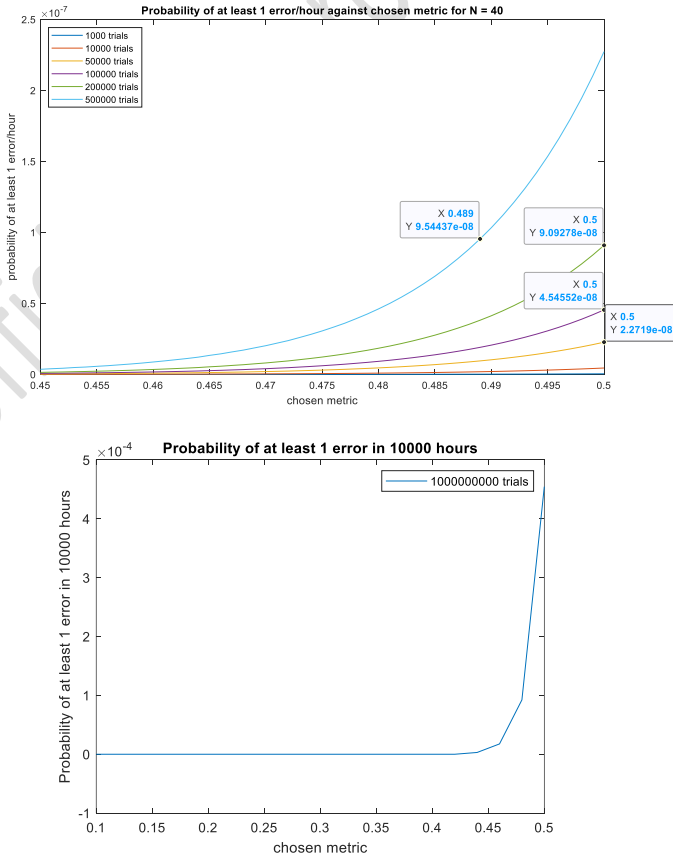

Figure 43: Theoretical mappings of error probabilities (top: per hour, bottom: per vehicle lifetime) vs. false alarm rate for N = 40

6.2 Summary of findings

Section 6 presented a methodology to systematically decompose a high level vehicle error metric to a module level error metric. The chosen method to obtain this decomposition is a quantitative FTA. This is an important tool to be used in safety analysis to understand how the vehicle level error can be mapped to the components making up the system. To properly perform FTA, an in depth knowledge of the system architecture, ODD and module limitations is required.

The findings of the FTA can then be used as: 1) input to the safety requirements in terms of module metrics (e.g. FP rates), 2) to add more safety requirements if the required vehicle level metric cannot be reached 3) guide test protocol creation as the FTA can help highlight which conditions should be given a higher priority and 4) to improve KPI definitions or derive KPI goal values of modules to ensure safety is respected. The last two points could be used to support WP6.

To perform this analysis, we assumed we had access to a harmonised EVENTS project architecture (Section 6.1.1) which is then used to go from an identified SOTIF-related safety goal violation (SG5 was used to demonstrate the method) to the events causing its violation being either the EVENTS modules error (e.g. object data errors) or SOTIF triggering conditions (e.g. bad weather conditions having a negative impact on some perception modalities). The focus of our FTA in this study was on perception modules. Subsequently, we also explained how to populate an FTA with errors rates to be able to perform a quantitative FTA (Section 6.1.2). Moreover, we also went on to provide a definition for what it means to have an error (this usually is linked to a minimum time duration) and provided recommendations on how to define KPIs that are more directly linked to safety goals (Section 6.1.3).

We also recommended some KPIs for decision-making modules which were not derived using FTA (Section 6.1.4) and also provided some additional metrics that could be used to support WP6 (Section 6.1.5).

The previous subsection (Section 6.1.6) shows an example of how to use the presented FTA method. In future work each error rate used in the FTA needs to be further researched and matched to the limitations of each module to provide more accurate decompositions. In the example, arbitrary values are used to show the main outcomes of the FTA. Even with arbitrary error values, the FTA already provides some useful observations of the system under study, such as the huge impact that self-assessment modules have on error decomposition.

7. Test protocols

7.1 Approach

As discussed previously, the activities of ISO 21448 [3] are focused on reducing the hazardous scenarios encountered in a system's usage. Some of this area is known, and we account for them by putting safety measures in place. The rest of this area is unknown, and we seek to uncover these unknown hazardous areas through verification and validation activities. The below image from ISO 21448 [3] conveys this reduction:

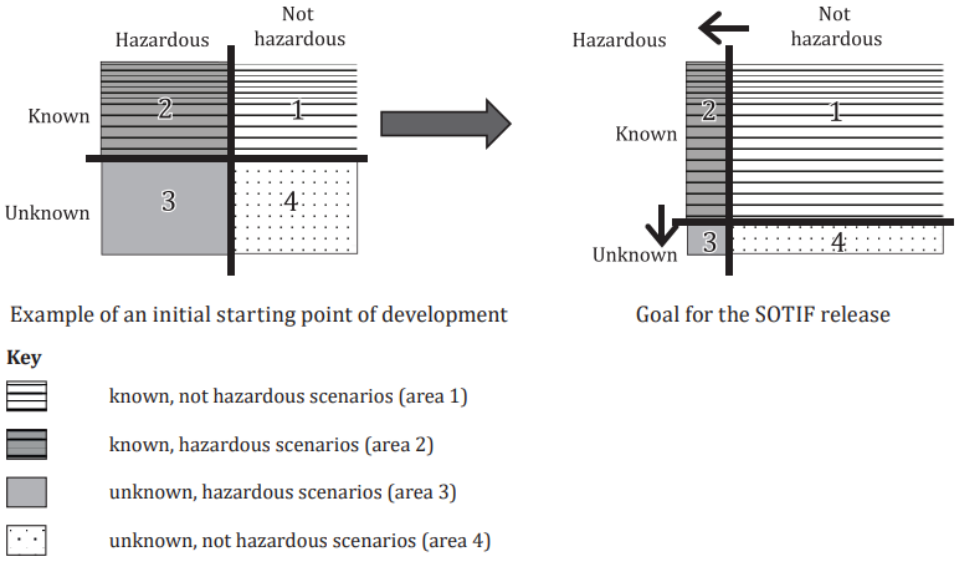

Figure 44: SOTIF scenario visualisation (Figure 6 from ISO 21448 [3])

The test protocols discussed in this section generally cover verifying and validating the known hazardous scenarios. However, to some extent, they also partially address unknown hazardous scenarios as they consider permutations of conditions that may have not been considered previously.

We used the results of Sections 5.4 (requirements from the safety goals) and 4.3 (scenarios from the safety goals) to generate the test protocols. A summary of this can be seen below in Table 28:

Table 28: Summary of requirements and scenarios

Safety Goal	Requirement	Scenario
SG3	System shall detect upcoming unmapped zone before 25 s of reaching that zone if in speed zone between 50 and 130 kph, or 18 s if in speed zone below 50 kph.	Scenario 1: Engaged system driving from in-ODD conditions to out-of-ODD conditions, in the presence of the listed scenery elements and/or environmental conditions Scenario 2: Driver trying to engage system while actively in out-of-ODD conditions, in the presence of the listed scenery elements and/or environmental conditions
	System shall detect construction zones before 24 s of reaching that zone if in speed zone between 50 and 130 kph, or 18 s if in speed zone below 50 kph.	
SG4	Vehicle heading shall be correct within TBD degrees of commanded vehicle heading. Note: Due to time constraints on the simulation tied to this, the refined value is a future work.	Engaged system driving in the presence of the listed scenery elements and/or environmental conditions
SG5	After actuating emergency braking, system shall abort braking if target has not been detected for a duration of 0.5 s.	Engaged system driving in the presence of the presence of a rear vehicle and the listed scenery elements and/or environmental conditions
	System shall account for both forward and rear threats when calculating braking command.	
	System shall meet TBD false positive rates toward its perception modalities.	
SG6	Braking shall be correct within 0.25 m/s^2 of braking command.	Engaged system driving in the presence of a lead object and the listed scenery elements and/or environmental conditions
	System shall meet TBD false negative rates toward its perception modalities.	
SG7	Throttle shall be correct within 0.2 m/s^2 of acceleration command.	Engaged system driving in the presence of a lead object and the listed scenery elements and/or environmental conditions

The top-level approach was as follows for each requirement:

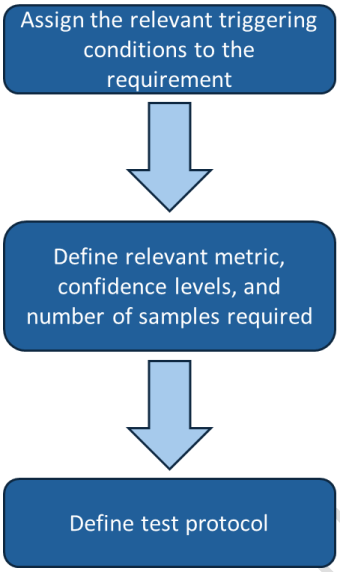

Figure 45: Top-level approach for test protocol generation

With regards to confidence levels, a value of 50% was assumed for all test protocols. From this value, we could calculate a number of samples required by using the following equation [49]:

$$N = \frac{\ln(1 - C)}{\ln(R)}$$

Equation 2: Calculation of number of samples required

Where N = number of samples, C = confidence level, and R = reliability which is the probability that your sample matches the required specification or alternatively 1-probability of samples not respecting the specification [50] . The probability of samples not respecting the specification can be computed using the exponential distribution (explained in Equation 1 in Section 6.1.3) if only a constant failure or error rate is available.

The test protocols shown in the next section can be used to aid the work in WP6. The test protocols are tied to safety requirements some of which are directed to a specific subsystem e.g. perception; this implies that EVENTS experiments that are only considering perception can still refer to such test protocols.

7.2 Summary of test protocols

7.2.1 SG3

Table 29: Test protocol for SG3, first requirement

ID	Req_A3.1
Requirement	System shall detect upcoming unmapped zone before 25 s of reaching that zone if in speed zone between 50 and 130 kph, or 18 s if in speed zone below 50 kph.
Relevant triggering conditions	Areas where GPS could cut out due to geological features and large infrastructures
Related metric	The time at which the unmapped zone is detected has to be greater than X s for Y% of the N1 samples. The Y% represents the error probability required to compute the denominator of the sample computation equation (Equation 2).
Desired confidence level	50%
Samples	N1 roadways, with N1 determined from related metric and confidence level
Test protocol	VEHICLE AND ACCESSORIES: Representative model of vehicle for simulation, actual vehicle for on-road test, data recorder SPLIT BETWEEN SIMULATION AND ON-ROAD TEST: Find N1 roadways where the roadway goes from mapped to unmapped (with > 1 min worth of mapped roadway). Approximately 10% of the roadways should cover some distribution of areas where GPS signal is not available due to geological features and large infrastructures (e.g. tunnels, tall buildings, long stretches of overpass). Drive an engaged system through these roadways and measure how far from the unmapped zone the system detects it.

Table 30: Test protocol for SG3, second requirement

ID	Req_A3.2
Requirement	System shall detect construction zones before 24 s of reaching that zone if in speed zone between 50 and 130 kph, or 18 s if in speed zone below 50 kph.
Relevant triggering conditions	Cones, posts, concrete barriers, construction workers, construction lighting, lanes with temporary marking, temporary signage, drop in road surface due to resurfacing
Related metric	The time at which the unmapped zone is detected has to be greater than X s for Y% of the N2 samples.
Desired confidence level	50%
Samples	N2 roadways, with N2 determined from related metric and confidence level

Test protocol	VEHICLE AND ACCESSORIES: Representative model of vehicle for simulation, actual vehicle for on-road test, data recorder SPLIT BETWEEN SIMULATION AND ON-ROAD TEST: Find N2 roadways that cover a distribution of the following: Cones, posts, concrete barriers, construction workers, construction lighting, lanes with temporary marking, temporary signage, drop in road surface due to resurfacing. Drive an engaged system through these roadways and measure how far from the construction zone the system detects it.

Table 31: Test protocol for SG3, third requirement

ID	**Req_A3.3**
Requirement	System shall have capability to detect perception degradations beyond 20%, (where degradations entail a shift outside of specified accuracy).
Relevant triggering conditions	Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast
Related metric	The time at which the 20% degradation is detected has to be smaller than X s for Y% of the N3 hours.
Desired confidence level	50%
Samples	N3 hours, with N3 determined from related metric and confidence level
Test protocol	VEHICLE AND ACCESSORIES: Representative model of vehicle for simulation, actual vehicle for on-road test, data recorder, ground truthing equipment SPLIT BETWEEN SIMULATION AND ON-ROAD TEST: Drive for N3 hours with cumulative hours representing these conditions: Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast. Drive an engaged system and measure how much time the system takes to detect its perception degradations beyond 20%.

7.2.2 SG4

Table 32: Test protocol for SG4

ID	**Req_A4.1**
Requirement	Steering shall be correct within TBD degrees of commanded heading of the vehicle.
Relevant triggering conditions	Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, sudden tight turns, strong winds
Related metric	The heading error has to be smaller than X deg for Y% of the N hours.

Desired confidence level	50%
Samples	N hours, with N determined from related metric and confidence level
Test protocol	VEHICLE AND ACCESSORIES: Representative model of vehicle for simulation, actual vehicle for on-road test, data recorder, ground truthing equipment SPLIT BETWEEN SIMULATION AND ON-ROAD TEST: Drive for N hours, with cumulative hours representing all of the following conditions: Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, sudden tight turns, strong winds. Drive an engaged system and measure actual heading vs. commanded heading of the vehicle.

7.2.3 SG5

Table 33: Test protocol for SG5, first requirement

ID	**Req_A5.1**
Requirement	After actuating high magnitude braking (>= 5 m/s^2), system shall abort braking if target has not been detected for a duration of 0.5 s.
Relevant triggering conditions	Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, uphill conditions, downhill conditions, strong winds
Related metric	The time it takes to end emergency braking has to be lower than X s for Y% of the N1 events
Desired confidence level	50%
Samples	N1 events, with N1 determined from related metric and confidence level
Test protocol	VEHICLE AND ACCESSORIES: Representative model of vehicle for simulation, data recorder SIMULATION: Create N1 events that induce hard braking (>= 5 m/s^2) for both pedestrian and vehicle ghost targets in a variety of conditions representing the following: Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, uphill conditions, downhill conditions, strong winds. Remove relevant target upon start of emergency braking and determine how long from start of emergency braking it takes to end emergency braking.

Table 34: Test protocol for SG5, second requirement

ID	**Req_A5.2**
Requirement	System shall account for both forward and rear threats when calculating braking command.

Relevant triggering conditions	N/A
Related metric	The impact speed has to be lower than X kph.
Desired confidence level	N/A
Samples	N2 events, with N2 determined number of parameters varied in simulation
Test protocol	VEHICLE AND ACCESSORIES: Representative model of vehicle for simulation, data recorder SIMULATION: Create N2 events in simulation that induce hard braking (>= 5 m/s^2) with ego-vehicle between 2 vehicles. Items to vary include the following: - Distance between lead and ego-vehicle - Distance between ego and following vehicles - Response delay of T1 (T1 = [0.5:0.5:2]) seconds for ego-vehicle - Response delay of T2 (T2 = [0.5:0.5:2]) seconds for following vehicle Desired outputs are as follows: - Delta speed between lead and ego-vehicle at impact - Delta speed between ego and following vehicle at impact

Table 35: Test protocol for SG5, third requirement

ID	**Req_A5.3**
Requirement	TBD FP metric toward a given perception modality
Relevant triggering conditions	Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast.
Related metric	The number of FP objects shall be less than X per hour.
Desired confidence level	50%
Samples	N3 hours, with N3 determined from related metric and confidence level
Test protocol	VEHICLE AND ACCESSORIES: Actual vehicle for on-road test, data recorder, ground truthing equipment SIMULATION AND ON-ROAD TEST: Drive for N hours (with cumulative hours representing these conditions: Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast Drive an engaged system and count number of objects that are sent to the planning algorithm component because they were ghost objects.

7.2.4 SG6

Table 36: Test protocol for SG6, first requirement

ID	Req_A6.1
Requirement	Braking shall be correct within 0.25 m/s^2 of braking command.
Relevant triggering conditions	Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, uphill conditions, downhill conditions, strong winds
Related metric	The braking error has to be smaller than X m/s^2 for Y% of the N hours.
Desired confidence level	50%
Samples	N hours, with N determined from related metric and confidence level
Test protocol	VEHICLE AND ACCESSORIES: Representative model of vehicle for simulation, actual vehicle for on-road test, data recorder, ground truthing equipment SPLIT BETWEEN SIMULATION AND ON-ROAD TEST: Drive for N hours, with cumulative hours representing all of the following conditions: Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, uphill conditions, downhill conditions, strong winds. Drive an engaged system and measure actual vs. requested braking.

Table 37: Test protocol for SG6, second requirement

ID	Req_A6.2
Requirement	TBD FN metric toward a given perception modality
Relevant triggering conditions	Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast
Related metric	The number of FN objects shall be less than X per hour.
Desired confidence level	50%
Samples	N3 hours, with N3 determined from related metric and confidence level
Test protocol	VEHICLE AND ACCESSORIES: Actual vehicle for on-road test, data recorder, ground truthing equipment SIMULATION AND ON-ROAD TEST: Drive for N hours (with cumulative hours representing these conditions: Mud, snow, wet road, wet windshield, hail, high EMI, lots of surrounding metal, rain, fog, rising steam, darkness, lots of artificial light, bright light washing out contrast Drive an engaged system and count number of objects that are not sent to the planning algorithm component because they were missed.

7.2.5 SG7

Table 38: Test protocol for SG7

ID	Req_A7.1
Requirement	Throttle shall be correct within 0.2 m/s^2 of acceleration command.
Relevant triggering conditions	Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, uphill conditions, downhill conditions, strong winds
Related metric	The acceleration error has to be smaller than X m/s^2 for Y% of the N hours.
Desired confidence level	50%
Samples	N hours, with N determined from related metric and confidence level
Test protocol	VEHICLE AND ACCESSORIES: Representative model of vehicle for simulation, actual vehicle for on-road test, data recorder, ground truthing equipment SPLIT BETWEEN SIMULATION AND ON-ROAD TEST: Drive for N hours, with cumulative hours representing all of the following conditions: Bald tires, low-mu roads (gravel, snowy, wet, and sandy roads), split-mu roads, potholes, uphill conditions, downhill conditions, strong winds. Drive an engaged system and measure actual vs. requested acceleration.

8. Conclusions

Developing toward SOTIF entails considering what conditions might cause hazards at the vehicle level, even in the absence of system, hardware, or software malfunctions. These conditions could be limitations and weaknesses on any aspect of an ADS stack – Sensing, perception, planning, control, or HMI. Additionally, they could be conditions of the ODD – Scenery elements, environmental conditions, or dynamic elements. Finally, they could entail how the driver of the ADS interacts with the system.

In this deliverable, we presented a step-by-step approach to SOTIF and how we might identify and address the factors above, starting from the outputs of the HARA (D2.3) and providing new outputs in the form of:

1. A revised HARA with respect to SOTIF considerations and identification of SOTIF-relevant safety goals.
2. New vehicle-level metric recommendations aligned to each SOTIF-relevant safety goal and based in random hardware failure rates and accident rates.
3. Identification of scenarios to tie to exploratory simulations and test protocols.

4. Exploratory calculations and simulations to help establish bounds on vehicle dynamic controls.
5. A quantitative FTA methodology that enables derivation of module-level metrics from the vehicle-level ones and highlights where one might need adjustment of metrics and/or addition of safety measures.
6. Safety requirements based on results from the exploratory simulations and FTAs.
7. A test protocol which is presented and is recommended as input to WP6 for verification and validation purposes.

Throughout the course of this deliverable, there were a number of assumptions that we made due to the limitations that we encountered at the vehicle level and also at the subsystem level. These include, but are not constrained to, the following:

- Limited data on insufficiencies for all types of modules (sensing, planning, perception, control).
- Low resolution on signal interfaces between modules.
- Lack of availability of logs from individual modules and from integrated modules for characterization.
- Inability to simulate with full stack for project.
- Out-of-scope nature of human factors aspect.

Based on the above limitations, we made assumptions where needed to show the SOTIF methodology using knowledge from prior work and literature. If this project is to be taken to full development, integration, and deployment, the expectation is that we would revisit these assumptions and limitations to ensure full coverage of the ODD and of the intended functionality.

References

[1] Y. Ganesh et al., “Vehicle System Hazard Analysis & RiskAssessment,” 2023. [Online]. Available: https://www.events-project.eu/wp-content/uploads/2024/10/D2.3.pdf. [Accessed 15 April 2025].

[2] A. Ohazulike et al., “EVENTS D.2.2 Full Stack Architecture & Interfaces,” 2023. [Online]. Available: https://www.events-project.eu/wp-content/uploads/2024/10/D2.2.pdf. [Accessed 9 April 2025].

[3] ISO-21448, “Road vehicles — Safety of the intended functionality,” 2022.

[4] “Binomial proportion confidence interval,” [Online]. Available: https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval. [Accessed 10 June 2025].

[5] ISO 26262, “Road vehicles — Functional safety,” 2018.

[6] J. Christy Bieber, “Car Accident Statistics For 2025,” 2024. [Online]. Available: https://www.forbes.com/advisor/legal/car-accident-statistics/. [Accessed 16 April 2025].

[7] “Roadway Departure Safety,” 2023. [Online]. Available: https://highways.dot.gov/safety/RwD. [Accessed 16 April 2025].

[8] FTS, “American Driving Survey: 2022,” 2023. [Online]. Available: https://aaafoundation.org/wp-content/uploads/2023/09/202309_2022-AAAFTS-American-Driving-Survey-Brief_v3.pdf. [Accessed 16 April 2025].

[9] J. M. Sullivan and M. J. Flannagan, “Risk of Fatal Rear-End Collisions: Is There More to It Than Attention?,” in *Driving Assessment Conference 2*, 2003.

[10] B. C. Tefft, “Drowsy Driving in Fatal Crashes, United States, 2017–2021,” 2024. [Online]. Available: https://aaafoundation.org/drowsy-driving-in-fatal-crashes-united-states-2017-2021/#:~:text=The%20National%20Highway%20Traffic%20Safety,drowsy%20driving%20in%20motor%20vehicle. [Accessed 16 April 2025].

[11] ISO 34503, “Road Vehicles — Test scenarios for automated driving systems — Specification for operational design domain,” 2023.

[12] I. CarMaker, *Reference Manual Version 12.0.1.*

[13] N. Merat, A. H. Jamson, F. C. Lai, M. Daly and O. M. Carsten, “Transition to manual: Driver behaviour when resuming control from a highly automated vehicle,” ScienceDirect, 2014.

[14] “UN Regulation No 157 – Uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping Systems,” 2024.

[15] "UN Regulation No. 79 - Uniform provisions concerning the approval of vehicles with regard to steering equipment," 2025.

[16] "The Highway Code," [Online]. Available: https://www.highwaycodeuk.co.uk/control-of-the-vehicle.html. [Accessed 24 April 2025].

[17] [Online]. Available: https://www.jdpower.com/cars/shopping-guides/when-driving-what-is-the-average-reaction-time.

[18] W. S. Lee, D. L. Grosh, F. A. Tillman and C. H. Lie, "Fault Tree Analysis, Methods, and Applications - A Review," *IEEE Transactions on Reliability,* Vols. R-34, no. 3, pp. 194-203, 1985.

[19] S. Shalev-Shwartz, S. Shammah and A. Shashua, "On a formal model of safe and scalable self-driving cars," *arXiv preprint arXiv:1708.06374,* 2017.

[20] F. Oboril, C. Buerkle, A. Sussmann, S. Bitton and S. Fabris, "Mtbf model for avs-from perception errors to vehicle-level failures," in *2022 IEEE Intelligent Vehicles Symposium (IV)*, 2022.

[21] R. Yu, C. Wang, Y. Sui and Y. Zhang, "Decomposition and Quantification of Sotif Requirements for Perception Systems of Autonomous Vehicles," *Preprint Available at SSRN: https://ssrn.com/abstract=4598460 or http://dx.doi.org/10.2139/ssrn.4598460.*

[22] W. Chen, C. Zhang, Y. Peng, Y. Yao, M. Cai and D. Dong, "Enhancing GNSS positioning in urban canyon areas via a modified design matrix approach," *IEEE Internet of Things Journal,* vol. 11, no. 6, pp. 10252-10265, 2024.

[23] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[24] G. Dumas and J. Lemay, "Splash and Spray Measurement and Control: Recent Progress in Quebec," *The Aerodynamics of Heavy Vehicles: Trucks, Buses, and Trains. Lecture Notes in Applied and Computational Mechanics,* vol. 19, p. 533–547, 2004.

[25] J. R. V. Rivero, T. Gerbich, B. Buschardt and J. Chen, "The effect of spray water on an automotive LIDAR sensor: A real-time simulation study," *IEEE Transactions on Intelligent Vehicles,* vol. 7, no. 1, pp. 57-72, 2021.

[26] G. W. Flintsch, L. Tang, S. Katicha, E. de Leon, H. Viner, A. Dunford, K. Nesnas, F. Coyle, P. Sanders, R. Gibbons, B. Williams, D. Hargreaves, T. Parry, K. K. McGhee, R. M. Larson and K. L. Smith, "Splash and Spray Assessment Tool Development Program," 01 10 2014. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/40866. [Accessed 30 April 2025].

[27] S. Gupta and R. R. Debata, "Tunnel Detection for Automotive Radar using Object Elevation Trends and Gaussian Filtering," in *2022 1st International Conference on Informatics (ICI)*, Noida, India, 2022.

[28] "The road safety of motorway tunnels," 2011. [Online]. Available: https://swov.nl/system/files/publication-downloads/fs_tunnels_uk_archived.pdf. [Accessed 16 April 2025].

[29] A. Pireddu and S. Bruzzone, "An analysis of the influence of tunnel length and road type on road accident variables," *Rivista di statistica ufficiale,* vol. 2, pp. 71-102, 2021.

[30] ERSO, "Motorways 2018," [Online]. Available: https://road-safety.transport.ec.europa.eu/document/download/931bdf28-1328-4694-a839-4c692a293cf5_en?filename=ersosynthesis2018-motorways.pdf. [Accessed 28 April 2025].

[31] F. Kraus, N. Scheiner, W. Ritter and K. Dietmayer, "The Radar Ghost Dataset – An Evaluation of Ghost Objects in Automotive Radar Data," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, 2021.

[32] Y. Zhang, A. Carballo, H. Yang and K. Takedav, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 196, pp. 146-177, 2023.

[33] K. Montalban, C. Reymann, D. Atchuthan, P.-E. Dupouy, N. Riviere and S. Lacroix, "A Quantitative Analysis of Point Clouds from Automotive Lidars Exposed to Artificial Rain and Fog," *Atmosphere,* vol. 12, no. 6, 2021.

[34] RCC-CM, "Annual Bulletin on the Climate in WMO Region VI - Europe and Middle East - 2021," 2024. [Online]. Available: https://www.dwd.de/EN/ourservices/ravibulletinjahr/bulletin_2021.pdf?__blob=publicationFile&v=16. [Accessed 29 April 2025].

[35] M. Hahner, C. Sakaridis, D. Dai and L. Van Gool, "Fog Simulation on Real LiDAR Point Clouds for 3D Object Detection in Adverse Weather," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[36] G. J. van Oldenborgh, P. Yiou and R. Vautard, "On the roles of circulation and aerosols in the decline of mist and dense fog in Europe over the last 30 years," *Atmospheric Chemistry and Physics,* vol. 10, no. 10, pp. 4597-4609, 2010.

[37] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer and F. Heide, "Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[38] C. Lv, D. Cao, Y. Zhao, D. J. Auger, M. Sullman, H. Wang, L. M. Dutka, L. Skrypchuk and A. Mouzakitis, "Analysis of autopilot disengagements occurring during autonomous

vehicle testing," *IEEE/CAA Journal of Automatica Sinica,* vol. 5, no. 1, pp. 58-68, 2018.

[39] C.-G. Roh and I.-J. Im, "A Review on Handicap Sections and Situations to Improve Driving Safety of Automated Vehicles," *Sustainability,* vol. 12, no. 14, 2020.

[40] C.-G. Roh, J. Kim and I.-J. Im, "Analysis of impact of rain conditions on ADAS," *Sensors,* vol. 20, no. 23, p. 6720, 2020.

[41] H. Li, N. Bamminger, Z. F. Magosi, C. Feichtinger, Y. Zhao, T. Mihalj, F. Orucevic and A. Eichberger, "The effect of rainfall and illumination on automotive sensors detection performance," *Sustainability,* vol. 15, no. 9, 2023.

[42] AAA, "Environmental Effects (Rain) on ADAS Sensors Full Research Report – Oct 2021," 2021. [Online]. Available: https://newsroom.aaa.com/asset/environmental-effects-rain-on-adas-sensors-full-research-report-oct-2021/. [Accessed 29 April 2025].

[43] ISO/PAS8880, "Road vehicles- Safety and artificial intelligence," 2024.

[44] Minitab, "Hazard functions in reliability analysis," [Online]. Available: https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/reliability/supporting-topics/distribution-models/hazard-functions/. [Accessed 13 May 2025].

[45] A. Krieger, "On the probability of n consecutive successes out of N tries," *IEEE transactions on aerospace and electronic systems,* Vols. AES-20, no. 6, pp. 835--835, 1984.

[46] M. Thulin, "The cost of using exact confidence intervals for a binomial proportion," *Electron. J. Statist,* vol. 8, no. 1, pp. 817 - 840, 2014.

[47] C. Buerkle, F. Oboril, J. Jarquin, F. Pasch and K.-U. Scholl, "Safe perception: On relevance of objects for vehicle safety," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021.

[48] B. Roungas et al., "D6.1 Experimental Procedures and," [Online]. Available: https://www.events-project.eu/wp-content/uploads/2024/07/EVENTS_D6.1_Evaluation_methodology_and_scope_v1.0.pdf. [Accessed 12 May 2025].

[49] Minitab, "How Many Samples Do You Need to Be Confident Your Product Is Good?," [Online]. Available: https://blog.minitab.com/en/the-statistical-mentor/how-many-samples-do-you-need-to-be-confident-your-product-is-good.

[50] S. S. Wang, T. A. Canida, J. D. Ihrie and S. J. Chirtel, "Sample Size Determination for Food Sampling," *Journal of Food Protection,* vol. 86, no. 9, 2023.