PRISSMA Project
Plateforme de Recherche et d'Investissement pour la Sûreté
et la Sécurité de la Mobilité Autonome
04/2021 - 04/2024

[L3.3] SECOND CONTROLLED ENVIRONMENT TEST CAMPAIGN AS PART OF THE GLOBAL HOMOLOGATION PROCESS FOR AUTONOMOUS MOBILITY

PROTOCOLE DE LA SECONDE CAMPAGNE D'ESSAIS EN ENVIRONNEMENT CONTRÔLÉ S'INSCRIVANT U SEIN DU PROCESSUS GLOBAL D'HOMOLOGATION DE LA MOBILITÉ AUTONOME

Main authors : R. REGNIER (LNE), A. PIPERNO (UTAC), M. DESSAINT (UTAC) , C. SERBOUH (UTAC), J. BOURET (UTAC), A. RENZAGLIA (INRIA), J-B HOREL (INRIA), K. MANN (INRIA), C. LAUGIER (INRIA), E. CHATEAUROUX (TRANSPOLIS), Y. NOUET (TRANSPOLIS), P. DUTHON (CEREMA), M. BOUDALI (LNE), A. BEN-DAOUED (CEREMA), C. SEGONNE (CEREMA), Y. TAHIROU (VALEO), A. SEBA (IGN), P. BOUQUET (IGN), R. DENIS (VALEO)

Keywords: AI based vehicles homologation, repeatability tests, robustness tests, AI overfitting tests, anticipation tests, closed road tests, test bench, Simulation, evaluation, AI systems, validation methods, verification process, evaluation protocol, scenario generation, metrics, KPI, performance indicators, simulation tools, sensor models, traffic simulation, graphical and physic engines, simulation environments, proof-of-concept

Abstract. This document describes the final state of the implementation of proofs-of-concept (POC) that aim at demonstrating the use of simulation tests during the homologation and certification processes of autonomous vehicles. Several POC are currently being developed within the PRISSMA project and their particular ongoing work is presented separately.

Résumé. Ce document décrit l'état final de la mise en œuvre des preuves de concept (POC) qui visent à démontrer l'utilisation des tests de simulation lors des processus d'homologation et de certification des véhicules autonomes. Plusieurs POC sont actuellement en cours de développement dans le cadre du projet PRISSMA et leurs travaux particuliers en cours sont présentés

Table des matières

Chapter 1: Introduction of the PRISSMA METHOD

In the dynamic landscape of autonomous vehicles, securing certification for embedded artificial intelligence (AI) is pivotal. This process not only ensures the safety of these systems but also validates their optimal performance and adherence to industry standards. This introduction outlines the comprehensive steps of the methodology developed for the homologation of embedded artificial intelligence in autonomous vehicles, with the specific focus on on-track and bench testing. It is an extension of Deliverable 1.5 on the PRISSMA project's overall protocol and requirements and will complement Deliverable 2.7 on its simulation counterpart. The methodology aims to ensure a thorough and rigorous evaluation that will be available in five distinct proof-of-concept (POC) scenarios, assuring safety, performance, and compliance of autonomous systems. In contrast to Deliverable 2.7, the approach taken for this part was to keep a common framework without setting a precise methodology in stone. Indeed, as the 5 POCs are drastically different, we had to ensure that our approach could include these different procedures. The methodology framework is as follows:

1. Evaluation Protocol:

The assessment of each specific function or task follows a structured protocol comprising four essential parts. Each of these parts contributes to a comprehensive understanding of the capabilities of the embedded AI, covering aspects such as the description of the function to be evaluated, scenarios or associated evaluation databases, requirements, metrics, and evaluation criteria, as well as a detailed description of the conducted trials.

2. Presentation of Testing Environments and Means:

This section highlights the testing environments and means implemented for on-track and bench trials. It includes detailed descriptions of tracks, testing benches, driving robots, sensors, and other elements crucial to the evaluation process. Subsections of this part will be adapted based on the specific type of POC under evaluation.

3. III. Participation Instructions:

Although optional, this section provides essential information to potential clients for certification and operators conducting the trials. It may include safety instructions, client-provided hardware requirements, data exchange formats, result communication procedures, and other relevant information.

4. IV. Roadmap:

Lastly, a provisional roadmap is established to guide the completion of POCs by early 2024. This roadmap highlights key milestones, anticipated deadlines, and provides an overview of the certification process. It will be continuously adjusted based on progress and results obtained during the POC evaluations.

This integrated methodology offers a holistic approach to the evaluation of autonomous vehicles, ensuring a thorough, reproducible assessment that aligns with safety and performance standards in the ever-evolving field of embedded AI.

Of the 5 POCs selected, two (those of UTAC and TRANSPOLIS) are generalizations of the homologation regulations already in force in the automotive field, and therefore enable us to see in the short term what will be possible to do on track tests to take AI into account. Two are based on the addition of cutting-edge technology to traditional tests (from INRIA/TRANSPOLIS, using augmented reality, and from IGN/VALEO, using precision sensors on tracks), which means we can envisage the benefits of adding new technologies to conventional track tests. Lastly, CEREMA and LNE have demonstrated the potential benefits of bench testing and hybridization with simulation for conditions that are difficult to control on runways (weather degradation in this case). The aim of choosing these 5 POCS is to cover a wide range of track and bench tests that will be implemented in the short and medium term for the POCS, and to show how to develop a protocol for each of these types.

It's worth noting that some of these POCS also have simulation parts. These parts will of course have to respect the protocol set up in deliverable 2.7 for their simulation part.

Chapter 2: UTAC POC: “Increase existing regulations for AI-based ADAS certification (POC IER)”

1. INTRODUCTION

AI based vehicles could have some safety weak points regarding repeatability, robustness, anticipation and overfitting for official known tests. So UTAC PRISSMA WP3 team has built first answers and proposals to adapt or to create new homologation tests scenarios / protocols / testing tools / evaluation metrics for the first WP3 POC tests in UTAC (February to July 2023). These tests have been analyzed and discussed, requirements have been built, and tests and protocols have been confirmed. That is presented in this deliverable. The second WP3 POC tests planed 22 and early April of 2024 will bring confirmation and if necessary fine tuning of them.

Our Inputs are deliverables of PRISSMA WP1 (particularly L1.4), WP2 & WP4 & WP6 (particularly scenarios for virtual/physical/open-road tests) , WP8 (regulation/standards first works); We also preliminary made a review of available vehicles with intelligent & predictive ADAS functionalities, and made a bibliography/state of the art of research works & papers related to « tests for AI & AI for tests » and to AI evaluation tools & metrics in the critical industries (planes, trains,..).

UTAC WP3 chose to test three vehicles (VW Golf 8 with predictive ACC, ZOE NEXYAD « MotorONE » research prototype with AI based anticipation driving, VALEO Drive4U taxi robot) , with three categories of new tests (repeatability & robustness, anticipation, overfitting) , with some existing or new scenarios (standing pedestrian, hidden crossing pedestrian, strong curve/intersection,…) & existing or adapted testing tools (new various pedestrian dummies , …), with new evaluation metrics (measure of performance of Automated emergency breaking but also of anticipation and no-use of emergency maneuvers like AEB,..).

Vehicles Homologation rules, protocols, tests and requirements are very simple, due to economical, technical and harmonization reasons: it cannot require hundreds of tests, neither thousand of physicals and virtual tests, because digital type approval is only in discussion and such amendment for all vehicle functions and regulations will not arrive before 4 years. So, it was very difficult to apply WP1/WP2 recommendations for metrics, needing thousands of tests results; Finally, our results and recommendations are mainly four series of new homologation tests and protocols (for repeatability, robustness, anticipation and overfitting) with 18 scenarios and basic proposals (binary KPI) for new metrics and requirements. Note that this is already quite far ahead compared to on-going regulation and Euro NCAP discussions about AI-based vehicles homologation or evaluation.

1.1 OBJECTIVES

The main objective of WP3 UTAC is to prepare the adaptation of approval tests for AI-based vehicles, and to build relevant and feasible tests and protocols to type approve vehicles with AI.

To do that we had also these two objectives:

- what kind of AI could happen in the future and in which systems/functions/vehicles?
- what are the proposals and how to use recommendations of PRISSMA other WP's: WP1 for AI evaluation & metrics, WP2 for virtual tests, WP6 for safety recommendations, WP8 for link with ecosystem and outside of the Project

1.2 CONTEXT and STATE OF THE ART

1.2.1 Main trends for AI deployment in vehicles, systems, subsystems

AI will arrive gradually in all vehicle functions, first perception, then route planning, trajectory, and control, Driver Monit, IHM, maneuvers like automated minimum risk maneuvers (MRM).

One paper discussed in pre-regulatory work of the GRVA automotive regulation group concluded that AI is necessary for automated vehicles because human driving behaviour and best practices are not precise/quantitative requirements, not programmable for an automate, but can be learned by AI system. It could arrive on premium-automated vehicles in 3 years.

Experts do not see on-board « live » learning in vehicles in the short/medium term, as this would lead to changes in vehicle behaviour that are impossible to validate. OEM process is to validate and freeze a software for a certain time, generally one or three years (batch learning process).

An IA system is not deterministic, does not meet a specification; It is a black box that can only be validated statistically:

- On potentially dangerous scenarios
- In relation to requirements/criteria/metrics which remain to be defined (data and learning, development, outcomes and safety)

There is therefore a need for new metrics, pre-critical scenarios, other than the world of ADAS.

A catalogue of critical scenarios will be known/learned by AI! And will not offer rare scenarios.

The AI only masters what it has learned (Operational Design Domain (ODD)), so we will need tests of robustness (edge case), very numerous & expensive therefore if possible virtual.

But to be able to perform these virtual tests , the OEM models of sensors, fusion, vehicle decision, vehicle control, actuators commands…will be required, and also a huge computing capacity ; Hence, the current projects and attempts to communalize the means of simulation by subcontracting them , and also opportunity for the regulations to require that the OEM model be provided for certification and that the data that led to the training be shown. Or at minimum that the Certification Technical Service provides the OEM secret randomized scenarios (corner case defined with the OEM at the limit of its ODD), for OEM testing them in SIL-HIL-VIL.

The LNE experts clearly see for the future the approval of components made by LNE and the AI-based vehicle approval made by UTAC.

A predictive model will be needed for type approval if no OEM model is available, for three objectives.
- for many simulations for safety virtual verifications
- for some approval physicals tests of robustness verification (identify the edge cases to be tested) and verification of the correlation tests/simulation
- for explanatory-interpretability (understand-explain the black box).

So, a lot of testing will be needed to build a simplified predictive model by predictive modelling.

For the approval, an audit will also be necessary (of the database and learning, validations,).

Both will be approved: AI and each AI-based vehicle (according to AI act)
One example of this is the Cyber Security and SW/OTA approval process.

Today GRVA regulation group discussions are not very advanced and target to evaluate if existing or soon existing regulations could be sufficient to verify AI-based vehicles safety: Complex systems safety audit annexes, EU AI act, UN-ECE software update and cyber regulations (UN-ECE R155 & 156 regulations). These regulations mainly require audits (of AI and software development, validations, production, reparations, data management, and safety assessment for robustness and black box assessment) but not additional & standardized tests to evaluate vehicles on testing tracks.

Another big problem with AI right now is perception, very hard to work out. It is also very difficult to specify an ODD in perception (examples: objects, sunset truck, pedestrian morphology, weather characteristics). Therefore, it is very difficult to make a specification of perception and to validate the perception function (OEM needs & type approval).

Therefore, it is very difficult to assess the reliability of a perception subsystem (allocation of requirements for reliability, which is necessary for the safety of operation).

AI will arrive in a few years into the vehicles decision systems, because on-screen learning over thousands of kilometers of filmed driving becomes possible.

For Predictive and explanatory models, OEMs need and work on it and are in much better position than the Technical Approval Service to have or build them.

The Technical Service of approval must however be competent (as in Safety or cyber), or even offer it to small OEMs (via the projects and programs French or European as TEF).

These principles have been taken into account during all our activities and reflections on PRISSMA WP3.

1.2.2 Adaptation of tests according to vehicle ODD and PRISSMA WP8 inputs

According to the requirements of all autonomous vehicle regulations (ALKS, ADS, draft of the Arreté francais autonomous urban shuttles), the OEM will have to declare to the customers and to the type of approval authority its ODD (Operational Design Domain).

For example, an OEM will declare that its autonomous driving functionality are safe and operational for speeds of not more than 30 km/h.

This constant of the ODD is therefore an important input for the approval tests of automated vehicles: these limits are the limits on which the AI based vehicle will be tested, verified and type approved.

As will be seen in the chapters below, while remaining within the budget and time constraints of WP3 PRISSMA, we have tried to find vehicles with different driving systems with AI and different ODD, as varied as possible, in order to solidly build our proposals to adapt type approval tests for all kind of AI-based vehicles.

1.2.3 Adaptation of tests according to OEM homologation safety audit and first WP6 inputs

PRISSMA WP6 constructs and adapts the safety audit of the vehicle type approval.

There is consensus on PRISSMA that the approval process of an AI based vehicle should begin with this functional safety audit, which will provide first inputs and themes and priorities for the approval tests & verifications.

We summarize this by the **diagram below** that was one of the conclusions of the WP3 + WP4 meeting of 16/09/2022:

WP6 audit (vehicle safety weak points / validations, ODD & vehicle limits)

⇩

WP2 virtual approval tests for dangerous/complex scenarios

WP3 physical approval tests for ODD limits scenarios & critical scenarios

⇩

WP4 physical approval tests on real open roads for real tests & verifications

1.2.4 Adaptation of tests according to needs/complementarity with virtual tests/open road tests (WP1 & WP2 & WP4 inputs).

As above written, Regulation and type approval tests could change and allow some virtual tests in a few years, allowing complementarity and using tests developed in PRISSMA WP2 (virtual tests) and in PRISSMA WP4 (open road tests), depending on several dimensions:

The first dimension is the hazardous, feasible, or expensive nature of physical testing, which simulation enables to avoid.

The second dimension is the digital model availability for virtual testing: will the manufacturer provide executable software models?

The 3rd dimension is the time available for approval, as virtual tests sometimes require more time of preparation than closed track tests, and as open road tests allow a lot of representatives and relevant tests (but not dangerous nor critical) in a short time.

1.2.5 Inputs from WP1 for methods and metrics to evaluate IA repeatability, robustness, and overfitting.

WP1 and in particular its deliverable L1.4 of October 2022 aim to provide an overview of the state of the art and recommendations on methods and metrics to evaluate systems based on AI.

The pages 57-59 of the PRISSMA deliverable L1.4, **reproduced below**, recommend three types of validation tests: perturbations, robustness, uncertainty:

4.4.1.4 A proposal for a validation protocol

The evaluation and validation protocol that we propose should allow to have clear and precise answers on the following questions:

1. How trustworthy are the uncertainty estimates of our model under perturbations ?

2. How robust are the prediction of our model under perturbations?
3. How do uncertainty and accuracy of different methods co-vary under perturbations

Concretely, we previously described corruptions and perturbations proposed, and ideally would like the model predictions to become more uncertain with increased shift, as far as shift degrades accuracy. This is usually called "covariate shift". Hereafter, we start by selecting a subset of perturbations ,following state of the art results, allowing model evaluation and validation with reduced cost. Next we explain decision process.

1. Data perturbations

   (a) Data-set shift: We propose the following shift for autonomous driving system:

   - Time of day / Lighting
   - Geographical location (City vs suburban)
   - Changing conditions (Weather / Construction)

   They may be simulated using domain adaptation technique [55] that has emerged as a new learning technique to address the lack of massive amounts of labeled data by using labeled data in one or more relevant source domains to execute new tasks in a target domain. In our context, we propose the following validation condition.

   (b) Adversarial perturbations

   (c) General corruptions

   (d) OOD samples

2. Robustness validation : In general, there are two different approaches one can take to evaluate the robustness of a neural network: attempt to prove a lower bound, or construct attacks that demonstrate an upper bound. The former approach, while sound, is substantially more difficult to implement in practice, and all attempts have required approximations. On the other hand, attacks used in the the latter approach are not sufficiently strong and fail often, the upper bound may not be useful. Moreover, as seen before, there exist different types of adversarial attacks and defenses for machine learning algorithms which makes assessing the robustness of an algorithm a laborious task. thus, there is an intrinsic bias in these adversarial attacks and defenses to make to further complicate matters. For instance an evaluation process must avoid a model dependence behavior, insufficient evaluation, a perturbation dependent results. This requires a model agnostic adversarial robustness assessment. In [56], authors have recently observed that dual synchronised attacks based on L_0 and L_∞ distance-norms allow a good robustness assessment on several neural network architectures. Moreover, their results suggest that L_1 and L_2 metrics alone are not sufficient to avoid spurious adversarial samples and it is better to combine dual norms (1 and ∞) to construct an upper bound on the robustness of the model.

3. Uncertainty validation : Naturally, we expect the accuracy of a model to degrade as it predicts on increasingly shifted data, and ideally this reduction in accuracy would coincide with increased forecaster entropy. A model that was well-calibrated on the training and validation distributions would ideally remain so on shifted data. On the completely OOD data, one would expect the predictive distributions to be of high entropy. Essentially, we would like the predictions to indicate that a model "knows what it does not know" due to the inputs straying away from the training data distribution.

First Recommendation of 'perturbation in black box' robustness tests are difficult and expensive for physical testing.

To evaluate the robustness of a system with IA, this deliverable L1.4 recommends on page 54 of “adversarial attacks in black box”, in order to see what the reaction of the system is and then gradually to adjust the attack to the system.

These attacks (or perturbations) with misleading data/configurations and at the limit of the system ODD are possible in the machine learning phase of an AI but also in the operational phase; For very famous example, AI based sensor vision were attacked by road signalization panels with little black rectangles.

The work of WP5 (cybersecurity) aims to protect the database for AI learning because to know this database is very helpful to attack it in operational phase.

Attacks (or disturbances) of corruption are also recommended, that is with data/configurations for use unavoidable & normal but misleading because at AI limits, like weather limits (fog, snow, cold, vibrations or movements decreasing image quality). The AI based system can be weak on these limit conditions because it made very learning on them.

But, according to the meeting with the AI expert and leader of WP1, Rémi Regnier, on 7/9/2022, these attacks are easy and relevant in virtual tests but difficult and expensive in physical tests because a step-by-step process is necessary to identify the system limits.

Second Recommendation of robustness tests:

This type of testing seems to be very suitable for closed-track testing by UTAC, according to the same meeting with the AI expert and leader of WP1, Rémi Regnier.

So, we will see further we developed many new robustness tests through the 2023 POC in UTAC testing tracks with three different, representative and intelligent vehicles.

Third Recommendation for uncertainty testing (repeatable/stable or chaotic system)

Again, these tests seem suitable for closed track UTAC tests, according to the meeting with AI expert and WP1 leader Rémi Regnier.

These tests assess the uncertainties of the system due to the different dispersions/ margin of error of its components (sensors, position & RTK ...) and the propagation of uncertainty in the neuronal networks of the AI system.

The system will be validated repeatable and stable if for very close inputs, there are very close results, otherwise it will be labelled chaotic.

So, we will see further we developed new repeatability tests through the 2023 POC in UTAC testing tracks with three different, representative and intelligent vehicles.

1.3 FUNCTIONS TO TEST IN HOMOLOGATION AND AI-BASED Functions

Vehicle type approval doesn't address systems, components, subsystems but only the whole vehicle performance, verifying that functions are safe (verification of compliance with regulation and regulation type approval tests): braking, steering, automated functions like ACC (automated cruise control), AEBS (automated emergency braking system), ESF (emergency steering function), ALKS (automated lane keeping system), ADS (automated driving system), AVP (automated Valet Parking), …

Therefore, UTAC WP3 worked only on the whole vehicle testing and evaluation,

In addition, defined which intelligent vehicles to use for 2023 POC tests and for building new tests, protocols, metrics and requirements.

For that during all PRISSMA project we had discussions and meetings with UTAC experts, PRISSMA experts, and OEM's experts.

The exchanges with the UTAC experts made it possible to have the inputs and first visions of new intelligent ADAS functions and of regulatory and consumerist works (Euro NCAP).

The discussions with PRISSMA and OEM's experts provided a technological vision of progress of AI and of AI possible applications for automotive industry, they also provide options and ideas of solutions to evaluate and type approve AI based vehicles.

1.3.1 UTAC Expert inputs: New Intelligent Speed Control Functions

UTAC is taking part in all regulation groups and the trend is clear: the priority for these groups is to challenge anticipation and prevention. The regulation DCAS working group plan to build in 2024 a new regulation, called DCAS (entry in force in 2025), but requirements for these first version will mainly address driver monitoring and a L2+ function (hands-off eyes-on driving assistance). UTAC experts think we will have to wait 2024, 2025, or more for an amendment of this DCAS regulation to see the first requirements and evaluations for prevention and anticipation functions.

As it is often the case, Euro NCAP is the precursor and incentive for new driving intelligent functions that will improve safety. These new functions, called Speed Limit Information Functions (SLIF) and Speed Limit Control Functions (SLC), do arrive in the future Euro NCAP safety assessments, which are still unofficial and are being discussed in the Euro NCAP WGs (in which UTAC participates).

In SAS 2023 protocol, The Euro NCAP has introduced bonus points in its vehicle evaluations if such functions of driving can manage (with an alert to the driver or with an automated speed reduction) the situations in the **figure below**, called features:

Systems that are able to properly identify road features where a speed, lower than the legal speed limit, is more appropriate and/or advised or the vehicle should come to a stop can attract points based on the number of road features. These road features are listed in the table below and example traffic signs of a limited number of countries are specified in Appendix I.

ROAD FEATURES	Points	Required Action
Curves*	2	Show and start reducing to appropriate speed
Roundabouts*	2	Show and start reducing to appropriate speed
Junctions*	1	Show and start reducing to appropriate speed
Traffic Lights	2	Warning only
Stop Signs	1	Warning only
Yield Signs	1	Warning only
No Entry	1	Warning only
TOTAL	**10**	

* Only eligible for scoring when linked to ISA and/or i-ACC

Euro NCAP is well aware of nature/numbers/root causes of road accidents in the main European countries and is convinced on the well-known fact (and widely shared by the French authorities in charge of road safety, DSR and ONISER) that excessive speed is the main cause of road accidents.

Euro NCAP will increase in 2026 protocols the challenging and rating of these intelligent speed control functions.

The French Working Group ‘rating level 3 Euro NCAP’, led by UTAC and attended by French manufacturers (Stellantis, Renault, Valeo) has the same vision and confirms (**see figure below**) that the best automated vehicle is the one that avoids emergency maneuvers through the use of intelligence and anticipation, as the human driver know to drive:

At the LNE Forum for Evaluation of AI on 24 November 2021, during the round table UTAC confirms this vision with the following example: « an intelligent vehicle should not have to choose between crashing an old woman crossing or a baby running on the road, il should be able to anticipate and to avoid this critical situation ».

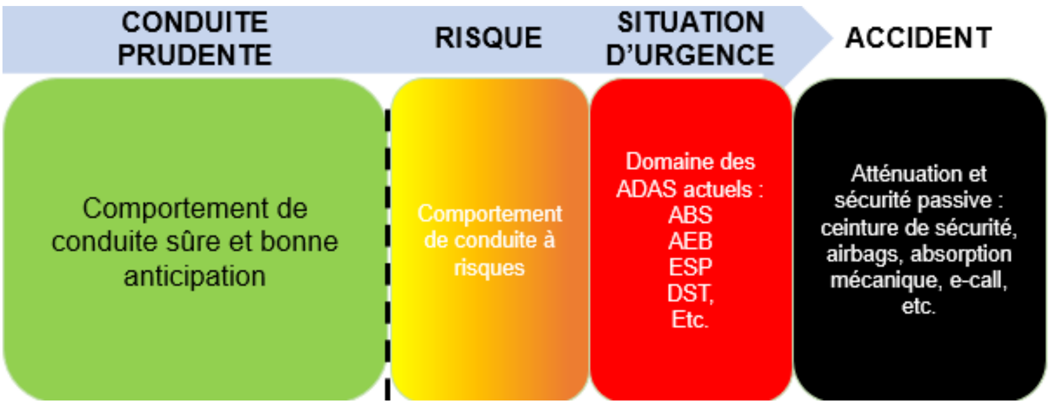
1.3.2 The proactive ACC and the 'safe speed' developed by the French start-up NEXYAD

Among all the manufacturers contacted for the UTAC tests of the POC of PRISSMA WP3, NEXYAD is clearly the most skilled about anticipation functions, working for more than 10 years on AI based intelligent automated driving functions.

NEXYAD has gained experience since the 2000s through 12 collaborative research programs with road safety and infrastructure experts from 19 countries, and NEXYAD has developed a new driving functionality that estimates road risk and therefore adapted and relevant safe speed. This relevant speed can be lower than authorized speed limit!

NEXYAD calculates in real time the level of risk of the situation , according to the context (infrastructure, traffic, presence of vulnerable persons, ...), according to speed, configuration of the road, signaling, visibility, proximity, ... and according to more than 5000 rules built by IA from a database of road accidents built during its experiment on 12 collaborative research programs with road safety and infrastructure experts during 10 years and in 19 countries.

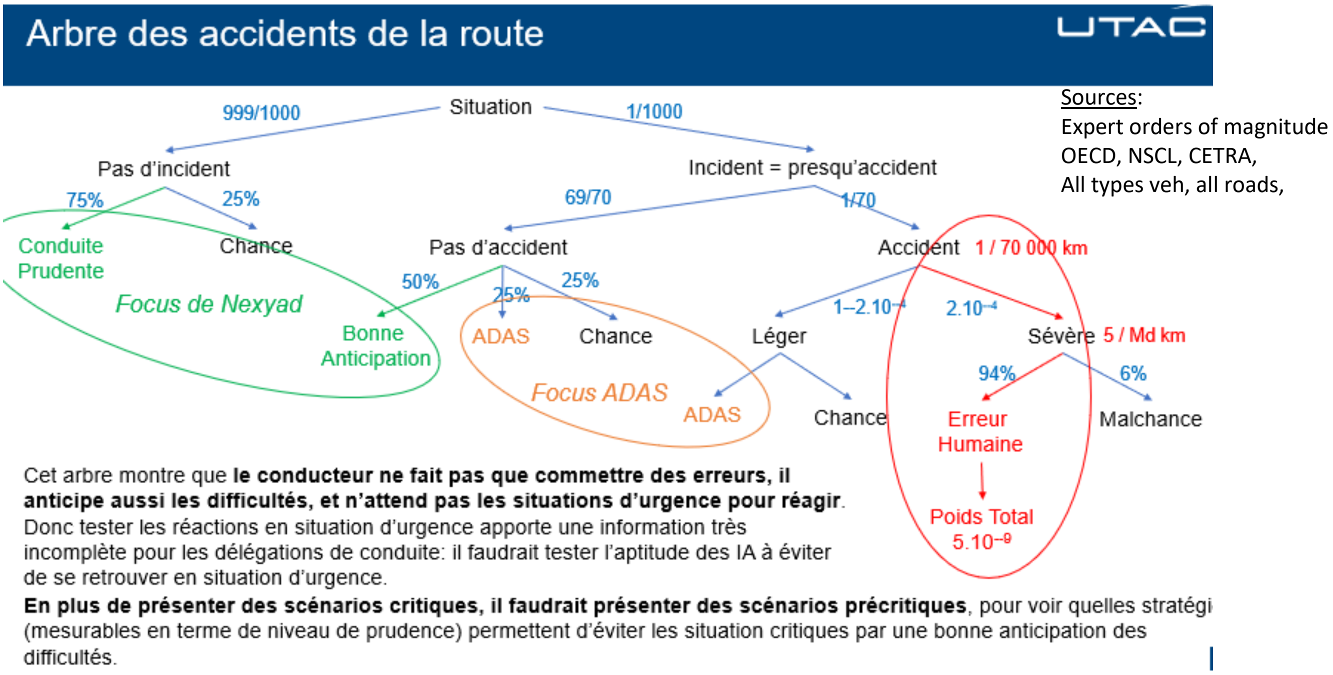
The level of risk calculated by NEXYAD, called safety score, is illustrated with the **figure below**, and the common-sense principle that « to drive safe with a low risk, you have to stay in green situations and not be close to red high-risk situation »:



NEXYAD also agrees with the principle and consensus (Euro NCAP, French authorities DSR and ONISER), that speed is the main cause of road accidents.

NEXYAD has studied accidents extensively (over 10 years and in 19 countries) and estimates that the prudent driving and anticipation, at the very beginning of the chain, represent 99.9% of the behaviours observed. Risk behaviours that do not have consequences are absent from the statistics, but they do exist and sometimes lead to emergencies, which are rare but focus all the attention of manufacturers and OEMs. Fatal accidents are even rarer (around five deaths per billion km in the OECD).

Thus, NEXYAD is very advanced in accident analysis, and is almost the only one to have a numerical analysis (in probability) not of accidents but of the near-accidents : See **figure below**, the road accidents and near-accidents tree, in which NEXYAD estimates that for one accident there have been 69 near-accidents, that is potential accidents that have been avoided through good driver reactions ; NEXYAD used these 69 potential accidents to build its anticipative system, which have to anticipate them, like a good and prudent human driver :



Thus, NEXYAD has developed (and patented) two new functions of intelligent driving, and is in discussions with many French, German and Japanese manufacturers to market them: these driving functions use the estimation of the risk of NEXYAD and the consequently relevant safe speed to have to minimise the risk and stay in the green zone of driving risks (previous figure).

The risk is estimated according to the road map (arrival on a steep curve, a tight crossroads...) and also what the vehicle sensors see (vehicle poorly parked, crowded crossroads, low field of vision

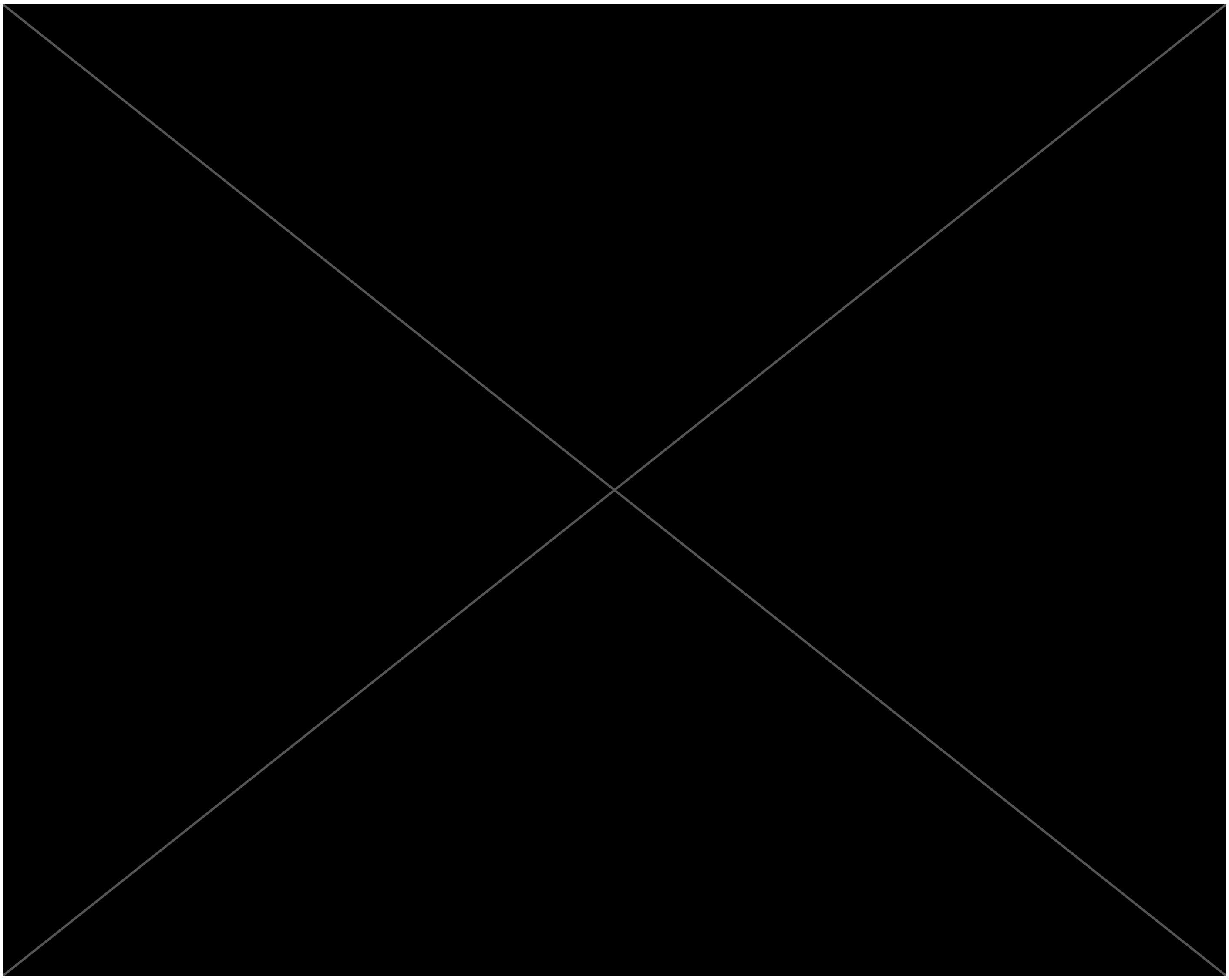
NEXYAD's two innovative and intelligent proactive driving functions are:

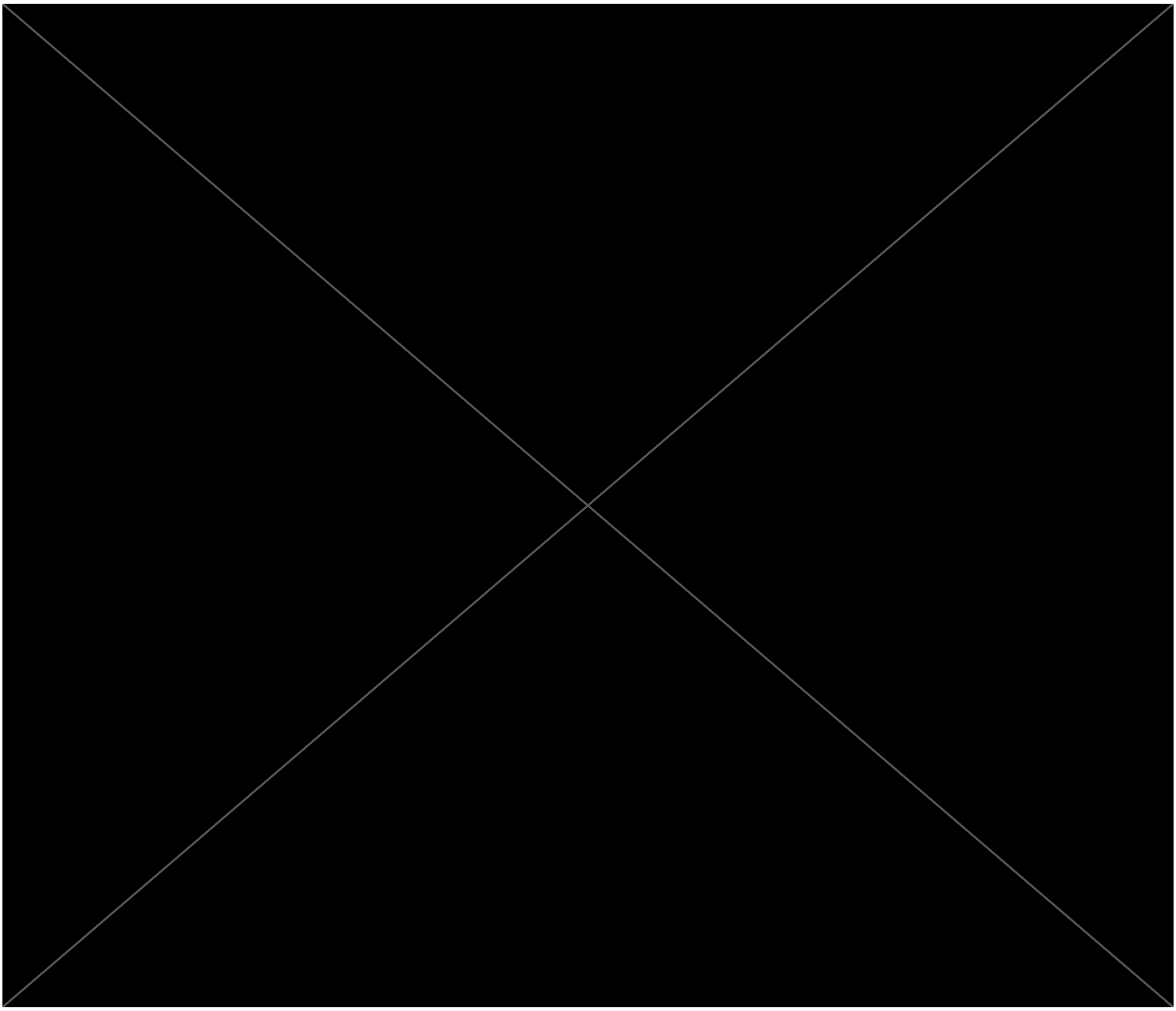
- **A safety assistant (named "safety coach") who alerts** the driver when his driving behaviour is no longer prudent (risk to high) in relation to the driving context (accident reduction estimated by NEXYAD of at least 25%).
- **An intelligent and proactive ACC** that automatically regulates the vehicle speed according to the driving context (up to 75% accident reduction according to NEXYAD)
- NB the difference in the result between 25% reduction of accident in alert mode compared to 75% of the intelligent ACC mode is explained by the fact that the driver may not immediately and always take into account the warnings and not slow down.

These two new functions are being implemented on a prototype vehicle, the NEXYAD DREAMOTOR1, **see photos below**, which is therefore one of the most advanced prototypes in the world (On PRISSMA there is no French actor among vehicle manufacturers and it's difficult to know their skills and developments on these very upstream and very competitive subjects). NEXYAD is part of the French industrial research and development group of the Regions Normandy/Ile de France, called NEXTMOVE (previously MOVEO), which supported and facilitated these innovative projects.

This prototype vehicle was rapidly tested in September 2022 by the team PRISSMA of the UTAC, on a one-hour circuit, allowing identifying interesting scenarios for PRISSMA tests in 2023; This prototype was not fully operational in March or April 2023, so we had to wail July 2023 to test it in PRISSMA 2023 UTAC POC & tests.

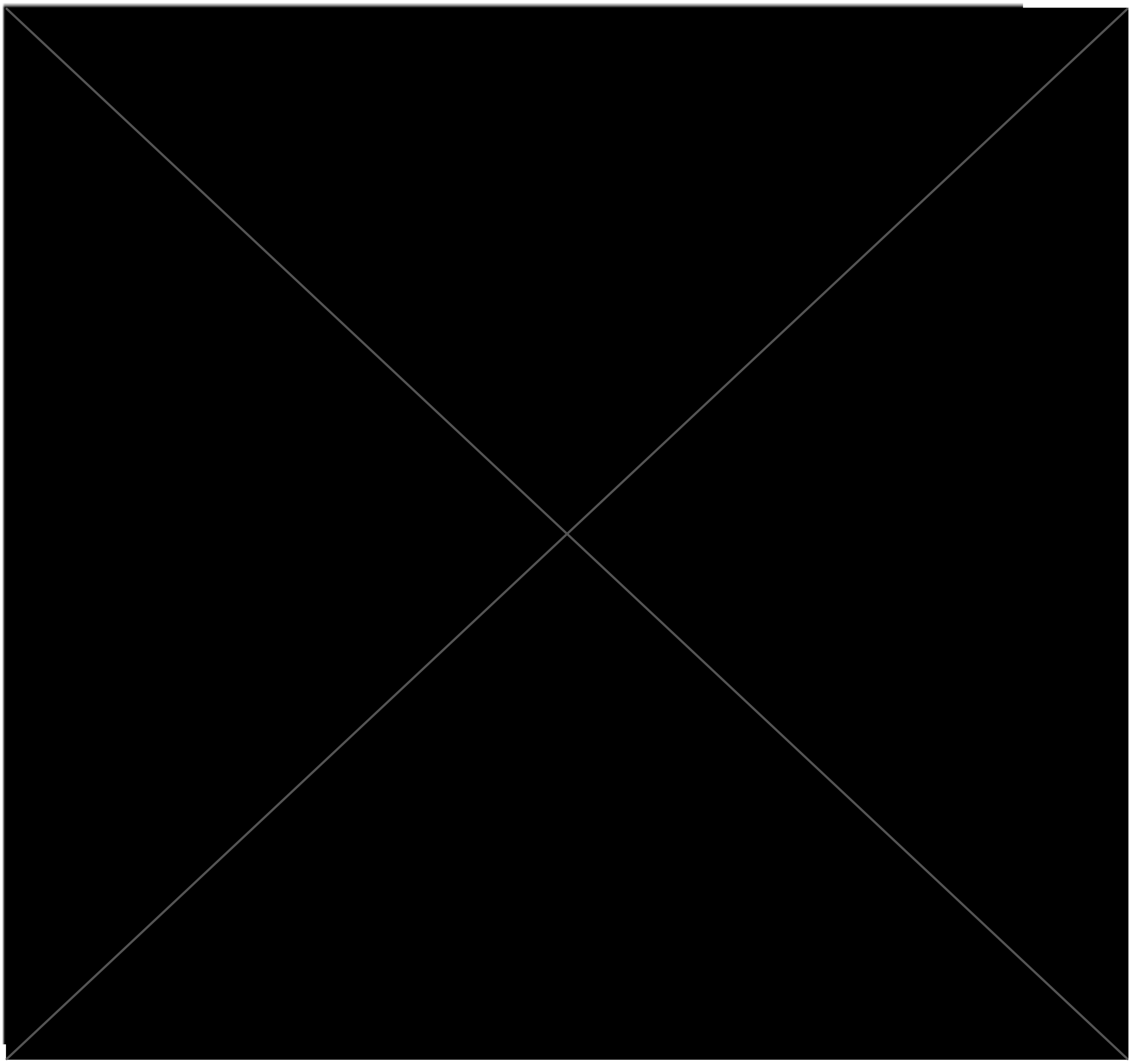
To find the best intelligent vehicle to test in UTAC WP3, our position was clear: either NEXYAD prototype is ready for WP3 POC 2023 tests, or we have to find another interesting up-to-date intelligent vehicle, that means to rent an up-to-date level 2 commercialized vehicle like a Tesla or a Mercedes.





[redacted]

[redacted]



1.3.4 Other AI-based vehicles available and testable for UTAC PRISSMA WP3 tests & POC :

According to that the objective of WP3 part UTAC is to prepare for the adaptation of the approval test to AI based vehicles, our needs were ideally to make these tests on different vehicles with a maximum of AI on board, not only like today on camera sensors.

Rent an up to date automatized vehicle. like **the New level3 Mercedes Class S**, **photo opposite,** recently approved for German motorways was therefore an interesting opportunity. However, the vehicle has still not yet been approved for the French motorways and no cooperation exist with the constructor to enable the driving function (ALKS) in France, and to have access to the results internal to the vehicles and its computers and functions. Another point is that today the Knowledge of technology content, AI, and performance of this new vehicle is very limited.
So, this opportunity seems a little premature even for the second part in 2024 of the UTAC WP3 tests.

The opportunity of **2 VEDECOM Easy Mile autonomous shuttles** have also been studied. These two shuttles, **see opposite figure**, were experimented on open roads more than 1 year on a circuit linking a bus stop and the VEDECOM site in Versailles-Satory.
But here too, this approach has not been taken, because it is from the older generation of shuttles, with few AI available and which have only been validated with the manufacturer only on a few predefined and fixed paths.
Therefore, it should also be necessary to benefit from a wide collaboration with the manufacturer before being able to make any test.

We also came to the same negative conclusion for the opportunities of testing:

- the ‘old’ **shuttle ARMA NAVYA from the UGE**, proposed by UGE for UTAC tests
- **the autonomous Renault ZOE of INRIA** in Grenoble, which is technically very interesting but not available because already used for WP3 TRANSPOLIS POC in Lyon.

1.3.5 German commercialized vehicles with intelligent predictive ADAS

Most Volkswagen vehicles (Golf 8, Arteon 2017, ID3, Nouvelle Polo,) propose a predictive ACC, capable to read (with on-board cameras) agglomeration entrance/exit speed limitation signs and (with road HD maps) strong curves (of roads, roundabouts,), and also capable to automatically adapt its speed through the ACC function.

This is a basic predictive driving feature, but without many AI and machine learning nor driving risk evaluation as NEXYAD proposes.

Here below is a good summary of this functionality of the Golf 8, found on commercial advertising:

“Adaptive Cruise Control ACC helps you to maintain a previously set maximum speed and a predefined distance to the vehicle ahead. In conjunction with a navigation system, ACC is enhanced by predictive cruise control and a cornering assist function.
ACC can adapt the vehicle speed to the applicable speed restrictions and course of the road (bends, roundabouts, etc.)”.

That’s the same for SEAT vehicles (which is a Volkswagen Group brand) .

Régulateur de vitesse adaptatif prédictif (ACC)

Jusqu’à maintenant le système ACC permettait d’adapter la vitesse de la voiture à celle des véhicules de devant, grâce aux radars avant. Avec la Nouvelle SEAT Leon, ce système dote le véhicule de nouveaux éléments prédictifs qui permettent au conducteur d’adapter sa vitesse de circulation en fonction de la route et des données GPS fournies par le système de navigation, ce qui lui permet également de corriger la vitesse en fonction du tracé de la route, virages, ronds-points, carrefours, limites de vitesse ou zone de travaux. De plus, grâce aux informations fournies par la caméra montée à l’avant et à la reconnaissance des panneaux de signalisation, le système peut ajuster la vitesse du véhicule lorsque les limites changent.

The Audi brand also offers regulation of Predictive Speed (on A4, Q3, Q7...), **see here below**:

► Audi A4 ► Aides au conduct... ► Systèmes d'aide ... ◄Assistant d'eff...

Valable pour les véhicules avec Audi adaptive cruise control et assistant d efficience prédictif

Fig. 121 Combiné d instruments : régulation prédictive

Description

La régulation prédictive 1) utilise les informations cartographiques du système de navigation pour adopter une conduite par anticipation. Pour l identification des panneaux de signalisation, le système tient également compte des informations reçues par la détection des panneaux* identifiés par caméra lien ►. Le système freine de manière autonome en cas de limitation de vitesse ou en amont de changements dans le tracé de la route (virages, intersections ou ronds-points), puis accélère de nouveau votre véhicule pour atteindre la vitesse mémorisée.

The users of these predictive ACCs testify on the internet to the effectiveness of this function, and are satisfied with the flexibility of the speed regulator which automatically detects and adapts the vehicle speed before entering on a roundabout or on a strong curve.

However, there are also many dissatisfied people who say that they no longer use this function (by disabling it) because it regularly generates false alarms (false positives) and sharp slowdowns or even sharp brakes when there is no risk, just because they read speed limitations signs from others close roads or from incorrect roadmaps datas.

that is also why many trucks have now this function: MERCEDES, DAF Trucks...

The predictive ACC function is also proposed in the after-sales (second assembles), but in this case, it is rather the community of users that indicates the zones where it is necessary to slow down.

The French or Japanese manufacturers do not offer any ACC predictive, but announcement could be imminent from STELLANTIS and NEXYAD.

VW GOLF 8 predictive ACC

We finally chose the GOLF 8 for PRISSMA UTAC WP3 tests in 2023 after preliminary tests.

These tests confirmed that VW Golf 8 and its predictive ACC is an up-to-date and intelligent function, representative of today best intelligent commercialized functions, and so interesting to test in the UTAC POC in order to estimate its performance and define new tests related to repeatability, robustness, anticipation & overfitting verification of intelligent functions.

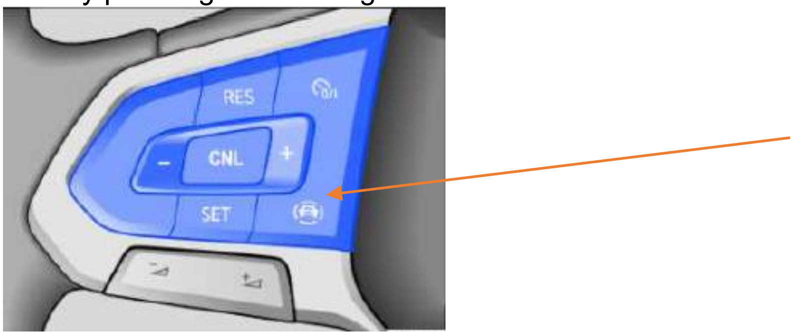
Golf 8 predictive ACC (called Travel Assist) description:

The system combines two driver assistance functions, Adaptive Cruise Control (ACC) for longitudinal assist and Lane Assist for lateral assist.

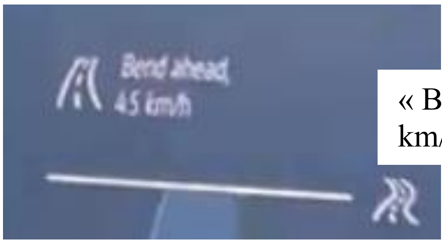
This function is activated by a button on the multifunction steering wheel, which therefore triggers longitudinal speed assist and lateral position assist. For safety reasons, the driver must keep his hands on the steering wheel for the guidance to be effective.

To this longitudinal speed guidance can be added an anticipation function. The system calculates the position of the Golf based on GPS and route data from the navigation system and must adapt the speed in advance to the approach of bends, roundabouts, crossings, speed limit zones etc…
At the same time, it uses the traffic sign recognition system via the front camera and must adapt the speed as soon as a limitation is detected.

The "Travel Assist" function is activated by pressing the steering wheel button:

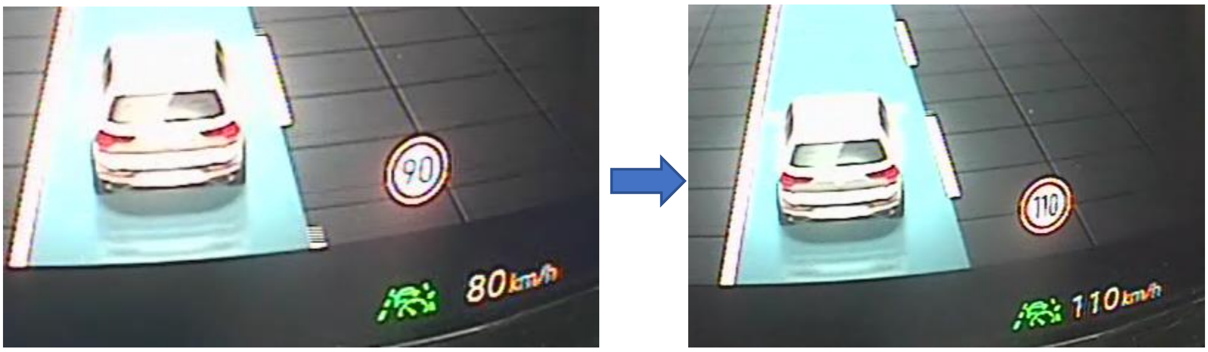


A reasonable speed instruction is given to the system at the start depending on the environment. Pre-tests of several minutes were carried out on UTAC 3 types of tracks (roads, city, and highway) with activation of "Travel assist". During driving, the so-called "anticipative" feature could be observed in different places, approaching a bend or a dangerous curve by this type of message on the Dashboard:



« Bend ahead, 45 km/h »

This message was followed by an automatic speed adaptation by braking the vehicle, at the speed recommended by the message.
The adaptive function by reading the speed limit signs could also be observed, for example when passing a traffic sign 110kph:

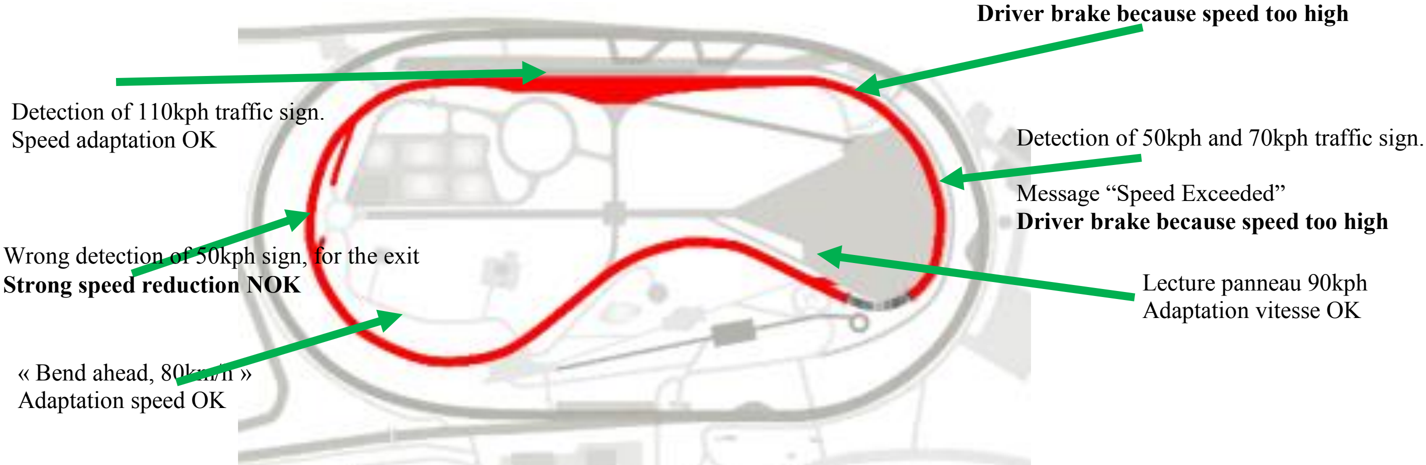


This message was also followed by an automatic speed adaptation by braking the vehicle, at the speed read on the traffic sign.

Main interesting observations:

- Highway tracks:
  Interesting track with several adaptive reactions of the system:

Detection of 90kph traffic sign.
Message "Speed Exceeded"
Driver brake because speed too high



- Road tracks:

Some speed adaptations in curves:

Appearance of an additional pictogram when driving on the road circuit, dangerous turn sign before almost every bend.

- City tracks:

No anticipation was observed on this track.

1.3.6 Conclusion for UTAC PRISSMA WP3 POC and tests:

We identified what are.
- the state of the art and the main trends for AI-based vehicles and functions
- the inputs of PRISSMA other WP's

We made large review of potential vehicle to test, we made preliminary tests, and we finally chose three interesting and representative AI-based vehicles:
- ZOE NEXYAD,
- VALEO Drive4U,
- Volkswagen Golf 8

We also made pre-test of these three vehicles to identify and chose which scenario are managed or not by these intelligent functions.

We developed through 2023 UTAC tests new scenarios and protocols in order to test and evaluate IA repeatability, robustness, anticipation and overfitting.

As we prepare homologation tests, we chose to develop critical and difficult scenario tests and protocols, usual and relevant for type approve tests.

2. TESTS and PROTOCOLS

2.1 Preamble and functions to evaluate.

First, as explained in chapter 1.3, many AI-based functions have to be type approved (which means to verify compliance with regulation requirements & type approval tests):

- Braking (UN-ECE R13 regulations)
- Steering (UN-ECE R79 regulation)
- automated functions like ACC (automated cruise control), (no regulation today)
- automated Lane Keeping Warning / Alert / Centering functions (no regulation today)
- AEBS (automated emergency braking system) (UN-ECE R152 regulation)
- ESF (emergency steering function), (UN-ECE R79 regulation amendment)
- ALKS (automated lane keeping system), (UN-ECE R157 regulation)
- ADS (automated driving system), (EU ADS regulation)
- AVP (automated Valet Parking), (EU ADS regulation)
- Etc.

Some of these regulations requires only one to three type approval tests (breaking, steering), but some others require 20 to 40 type approval tests (AEBS, ALKS, ADS).

Secondly, as explained in chapter 1.2, according to the requirements of all autonomous vehicle regulations (ALKS, ADS, draft of the Arreté francais autonomous urban shuttles), the OEM will have to declare to the customers and to the type approval authority its ODD (Operational Design Domain). For example, an OEM will declare that its autonomous driving functionality are safe and operational for speeds of not more than 30 km/h.

In addition, the ODD limits will define the tests and the limits on which the AI based vehicle will be tested, verified and type approved.

For these two reasons, PRISSMA project duration and budget are not enough to investigate adaptation of all functions and all type approval tests for AI-based vehicles.

Therefore, we investigate the most important existing scenarios (and most frequent in these regulations and in road accidents), and how to adapt them for AI-based vehicle with potential safety weak points on repeatability, robustness, anticipation and overfitting.

We also develop new scenarios to evaluate anticipation and overfitting of AI-based vehicles.

So finally, we built a catalogue of 18 existing or new scenarios, to verify during vehicles or functions homologation that there are no weak points related repeatability, robustness, anticipation and overfitting.

These tests can be chosen when they are relevant for the considered vehicle/function/r.

These new tests and scenario are quite far ahead compared to on-going regulation or Euro NCAP discussions to evaluate AI-based vehicles, as discussed in chapter 1.2 CONTEXT and STATE OF THE ART:

Today GRVA regulation group discussions are not very advanced and target to evaluate if existing or soon existing regulations could be sufficient to verify AI-based vehicles safety: Complex systems safety audit annexes, EU AI act, UN-ECE software update and cyber regulations (UN-ECE R155 & 156 regulations). These regulations mainly require audits (of AI and software development, validations, production, reparations, data management, and safety assessment for robustness and black box assessment) but not additional & standardized tests to evaluate vehicles on testing tracks.

Therefore, our proposals of about 6 new scenarios to tests and new metrics to verify AI-based vehicles/functions (on repeatability, robustness, anticipation and overfitting) still remain to be presented and discussed to this GRVA regulation group.

All the following new tests are below described and specified, using EUNCAP (Euro NCAP) references and standards, AEB or SAS or AD protocol, depending on the functions to be tested.

ENCAP has been taken as a model because this is a well-known and mastered protocol; also, their requirements are stricter than Regulation Protocols and can reveal the weakness of an AI system.

Furthermore, most of the vehicles with ADAS and without AI can handle the Regulation Protocol quite easily, whereas the ENCAP Protocol can point out some weaknesses of the systems. The goal is to challenge the AI system with harder situations.

Below the specifications of ENCAP Protocol that are common to all our tests proposals:

	CPFA	CPNA	CPNCO	CBLA	
Section	7.2.1	7.2.2	7.2.3	7.3.4	
Type of test	AEB			AEB	FCW/ESS
VUT speed [km/h]	10-60			25-60	50-80
VUT direction	Forward			Forward	
Target speed [km/h]	8	5		15	20
Target direction	Coming from Farside	Coming from Nearside		Forward	
Impact location [%]	50	25,75*	50	50	25
Dummy Articulation	Yes – as per test speed				

	VUT	EPT	EBT
Speed	+ 1.0 km/h	± 0.2 km/h	± 0.5 km/h
Lateral deviation	0 ± 0.05 m	0 ± 0.05 m for crossing scenarios 0 ± 0.15 m for longitudinal scenarios	
Lateral velocity		0 ± 0.15 m/s	0 ± 0.15 m/s
Relative distance			
Yaw velocity (upto T_{STEER})	0 ± 1.0 °/s		
Steering wheel velocity (upto T_{STEER})	0 ± 15.0 °/s		

2.2 Critical scenarios and repeatability

The first category of testing is about Repeatability, the goal is to perform many repetitions of a given scenario, with the same conditions and verify if the performance is similar.

Today, on a vehicle equipped with classic ADAS systems (ex: AEB), we note that the performances are not always repeatable. Here are some examples of repeatability results on ENCAP scenarios:

Pedestrian scenarios:

CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	N	Avoidance
CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	N	Avoidance
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	Y	Impact
CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	Y	Impact
CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	Y	Impact
CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	N	Avoidance
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	N	Avoidance
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	Y	Impact
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	Y	Impact

Car to Car scenarios:

CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance
CCRs AEB	Day	20	0	-/-50	AEB	-	-	Y	Impact
CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance
CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance

Bicycle scenarios:

ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	N	Avoidance
ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	Y	Impact
ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	N	Avoidance

The goal is to see if the AI on the last ADAS system increases the performances or not, compared to a system without AI.

The following scenarios will be performed 10 times each.

2.2.1 CPNCO-50 (Car to Pedestrian Nearside Child Obstructed 50%)

This scenario refers to the ENCAP 2023 protocol:

7.2.3 Car-to-Pedestrian Nearside Child Obstructed

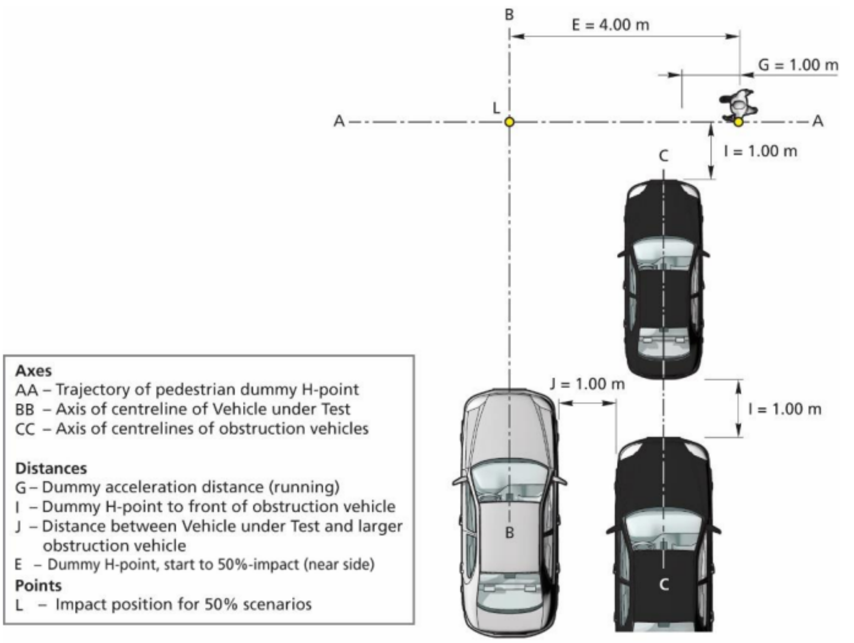


Figure 7-3: CPNCO-50 scenario, Running Child from Nearside from Obstruction

Car-to-Pedestrian Nearside Child Obstructed 50% (CPNCO-50) – a collision in which a vehicle travels forwards towards a child pedestrian crossing (5kph) its path running from behind and obstruction from the nearside and the frontal structure of the vehicle strikes the pedestrian at 50% of the vehicle's width when no braking action is applied.

2.2.2 CPFA-50 (Car to Pedestrian Far side Adult 50%)

This scenario refers to the ENCAP 2023 protocol:

7.2.1 Car-to-Pedestrian Farside Adult

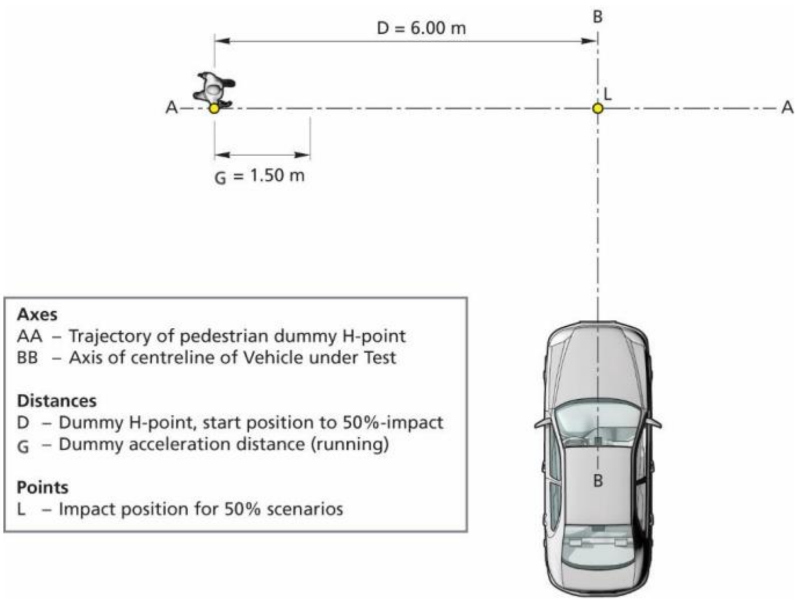


Figure 7-1: CPFA-50 scenario, Adult running from Farside

Car-to-Pedestrian Far side Adult 50% (CPFA-50) – a collision in which a vehicle travels forwards towards an adult pedestrian crossing (8kph) its path running from the far side, and the frontal structure of the vehicle strikes the pedestrian at 50% of the vehicle's width when no braking action is applied.

2.2.3 CBLA-50 (Car to Bicyclist Longitudinal Adult 50%)

This scenario refers to the ENCAP 2023 protocol:

7.3.4 Car-to-Bicyclist Longitudinal Adult

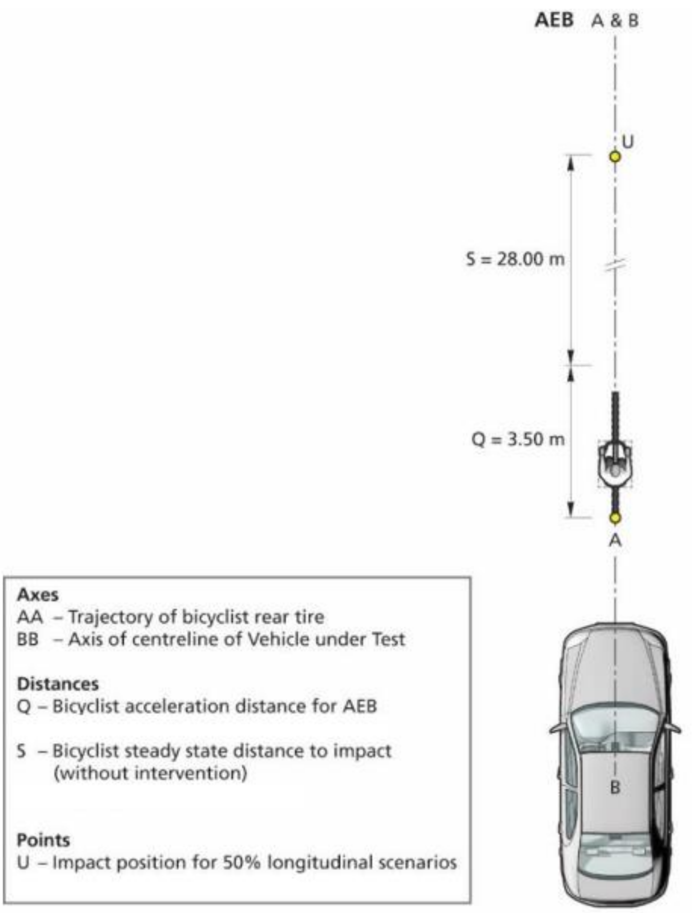


Figure 7-13: CBLA scenarios, Longitudinal Bicyclist

Car-to-Bicyclist Longitudinal Adult 50% (CBLA-50) – a collision in which a vehicle travels forwards towards a bicyclist cycling (15kph) in the same direction in front of the vehicle where the vehicle would strike the cyclist at 50% of the vehicle's width when no braking action is applied.

2.3 Critical scenarios and robustness

The second category of testing is about Robustness, the goal is to perform many variants of a given scenario and verify if the performance is similar.

For examples, we can change the speed of the target, the colour of the clothes…

2.3.1 CPNCO-50 (Car to Pedestrian Nearside Child Obstructed 50%)

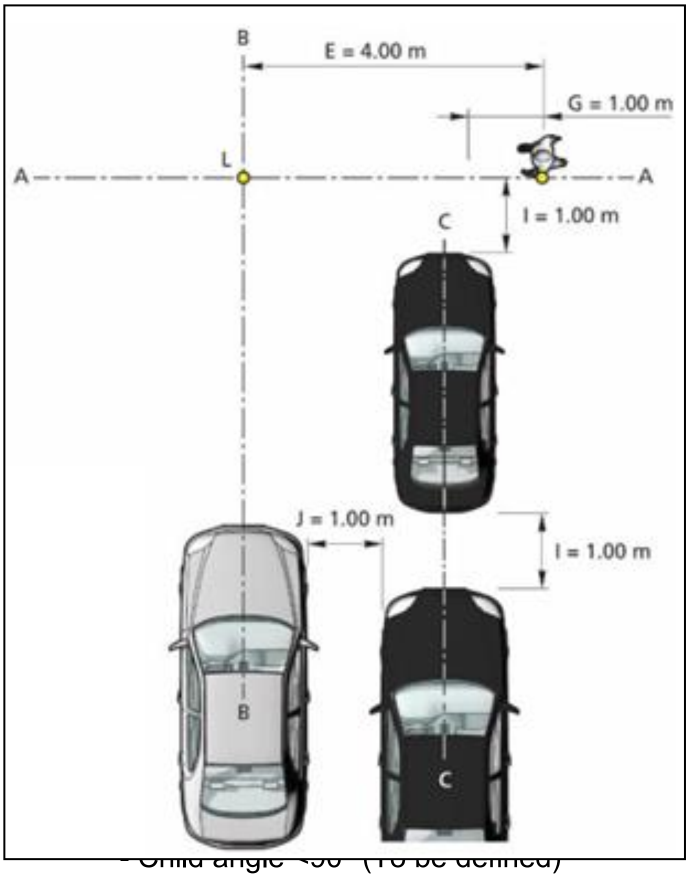
Same scenario as 2.2.1 with different alternative of it. If the obstruction is too harsh, it can be removed.

SPEED CHANGING:
- Child running at 6kph
- Child Start @3kph and accelerate @6kph

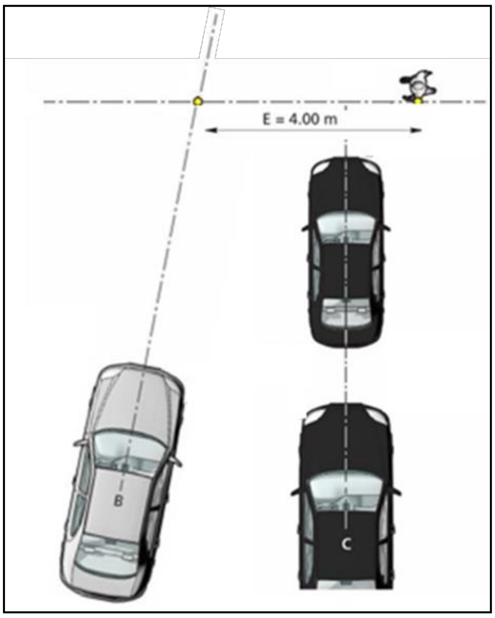
FORM CHANGING:
- Child wearing a backpack
- Child with a stuffed toy

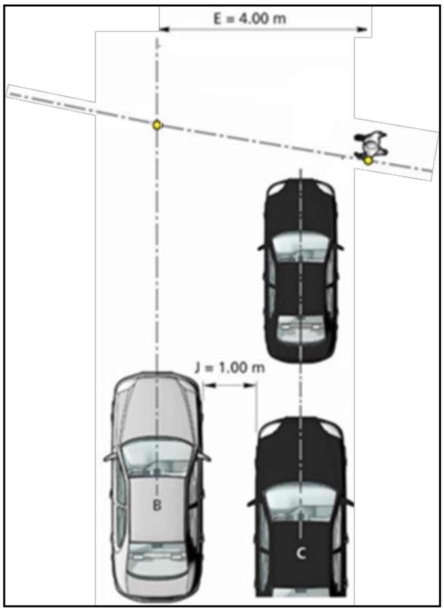
COLOR CHANGING:
- Child with yellow jacket

SURROUNDING CONDITIONS CHANGING:
- Strong Light in front of VUT



ANGLE CHANGING:
- VUT angle >90° (To be defined)
- Child angle <90° (To be defined)





2.3.2 CPFA-50 (Car to Pedestrian Farside Adult 50%)

Same scenario as 2.2.2 with different alternative of it:

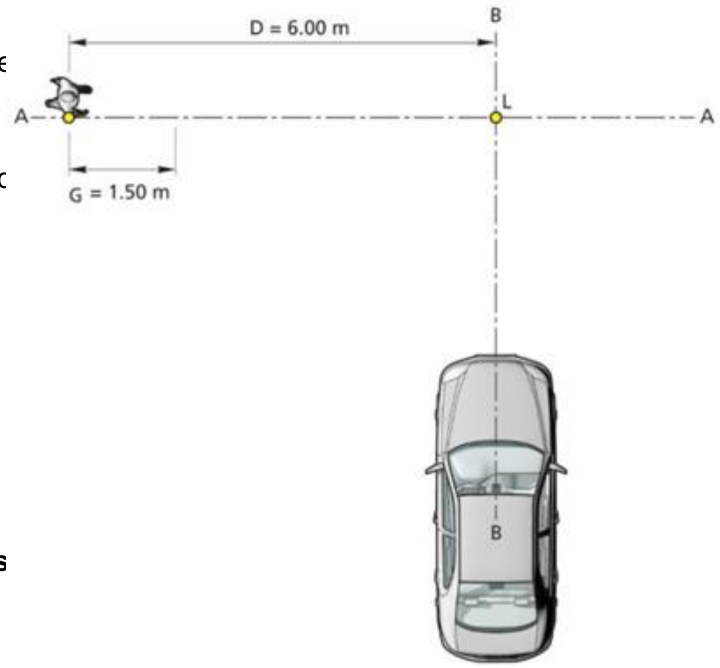
SPEED CHANGING:
- Pedestrian 5kph
- Pedestrian starts 5kph and acce

FORM CHANGING:
- Group of Pedestrian waiting anc
- Pedestrian with backpack

COLOR CHANGING:
- Pedestrian with yellow jacket

SURROUNDING CONDITIONS CHANGING:
- Strong Light in front of VUT

ANGLE CHANGING: (similar as
- VUT angle >90°
- Pedestrian angle <90°



2.3.3 CBLA-50 (Car to Bicyclist Longitudinal Adult 50%)

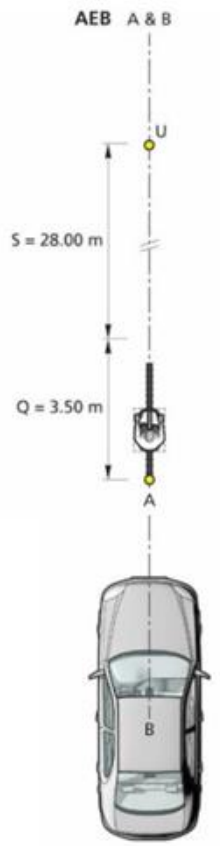
Same scenario as 2.2.3 with different alternative of it:

SPEED CHANGING:
- Bicyclist Starts 10kph and accelerate 20kph (to be confirmed at firs
- Bicyclist 25kph? (To be confirmed at first tests)

FORM CHANGING:
- Bike with cargo rack
- Adult with backpack

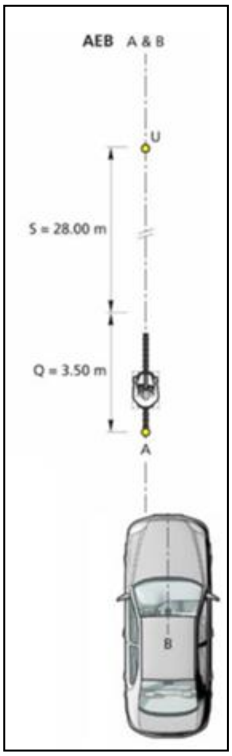
COLOR CHANGING:
- Adult with yellow jacket

SURROUNDING CONDITIONS CHANGING:
- Strong Light in front of VUT

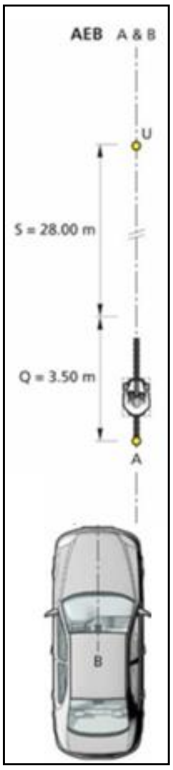


OVERLAP CHANGING:

- 75% (symmetry of 25% usual case)



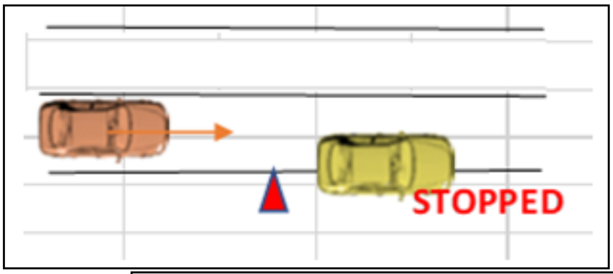
- 5% (bike close to road edge)



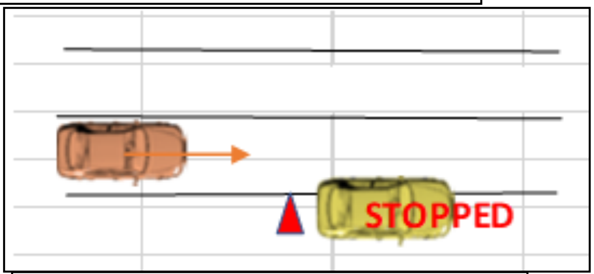
2.3.4 Stationary Car on Emergency Lane

A stationary car is stopped in an emergency lane, with a traffic sign (red triangle), different position of the stopped vehicle:

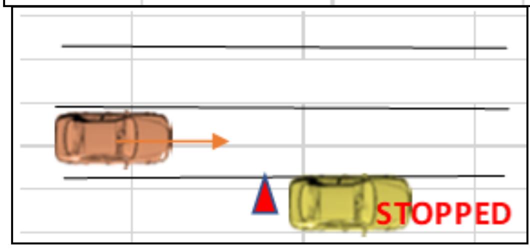
-50%:



- 25% (To be defined):



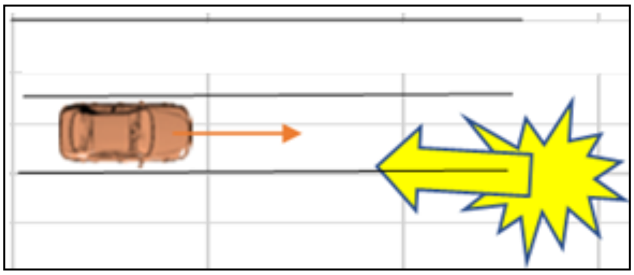
-0% (edge limit):



2.3.5 Stationary object or dazzling light on Highway

-Stationary Object:

- Dazzling light (difficult perception):
If possible, it will be performed at the exit of the highway Tunnel



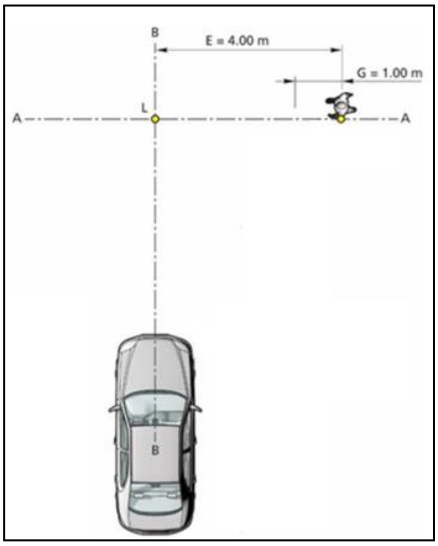
2.4 Pre-critical scenarios (anticipating avoiding AEB/critical maneuvers)

The third category of testing is about Anticipation, the goal is to perform some classic scenario by changing some conditions to see if the vehicle can anticipate a potential danger (without activation of AEB).
Each scenario will be repeated twice (two runs per scenario).

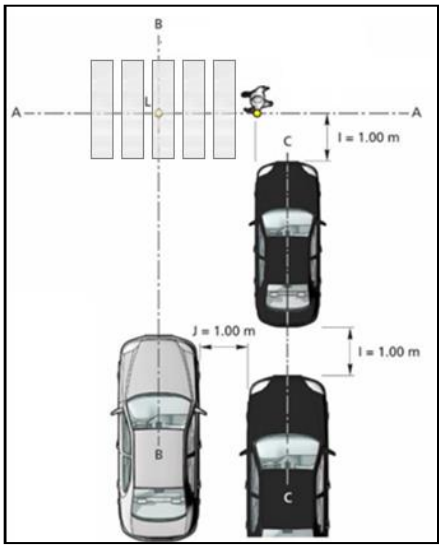
2.4.1 CPNCO-50 (Car to Pedestrian Nearside Child Obstructed 50%)

Same scenario as 4.2.1 with different alternative of it:

-Without Obstruction:



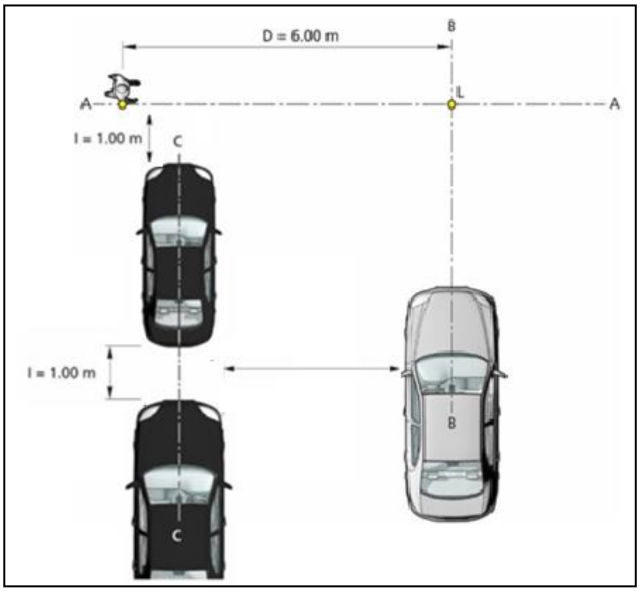
-Stationary Child (edge of pedestrian crossing):



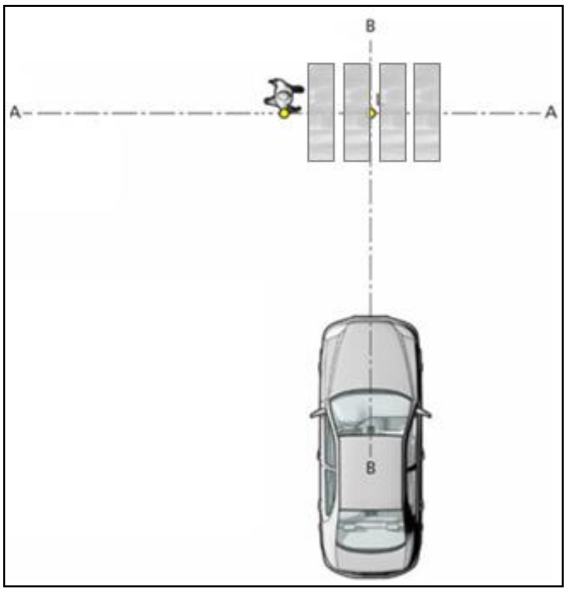
2.4.2 CPFA-50 (Car to Pedestrian Far side Adult 50%)

Same scenario as 4.2.2 with different alternative of it:

-With Obstruction:



-Stationary Adult (edge of pedestrian crossing):

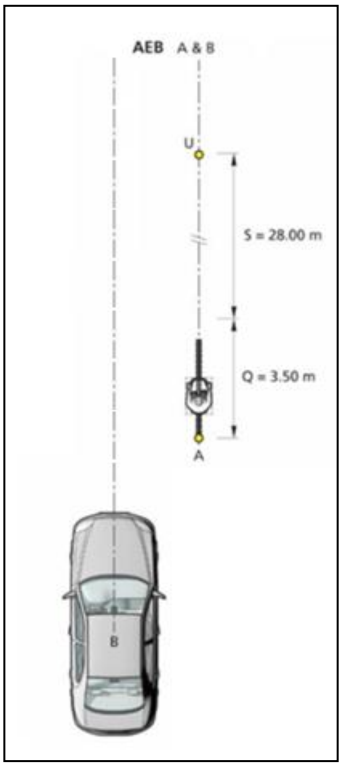


2.4.3 CBLA-50 (Car to Bicyclist Longitudinal Adult 50%)

Same scenario as 4.2.3 with different alternative of it:

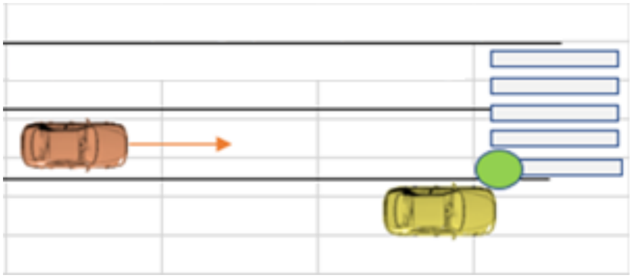
-With different bearing:

-Bicycle close to VUT path (cycling track):

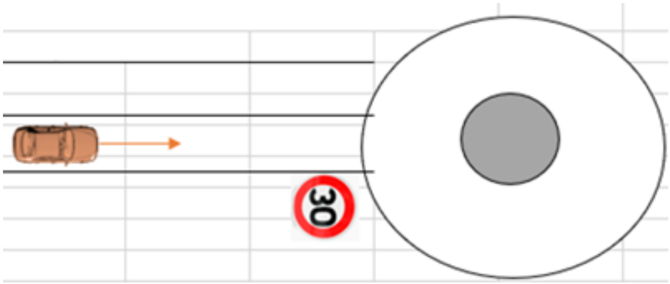


2. without Target

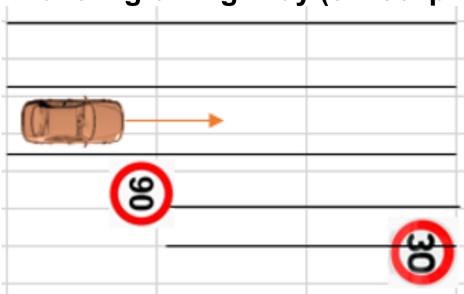
-Pedestrian Crossing, Green Traffic light and obstruction (potentially hidden pedestrian):



-Approach of strong curve (ex: roundabout) with late traffic sign:



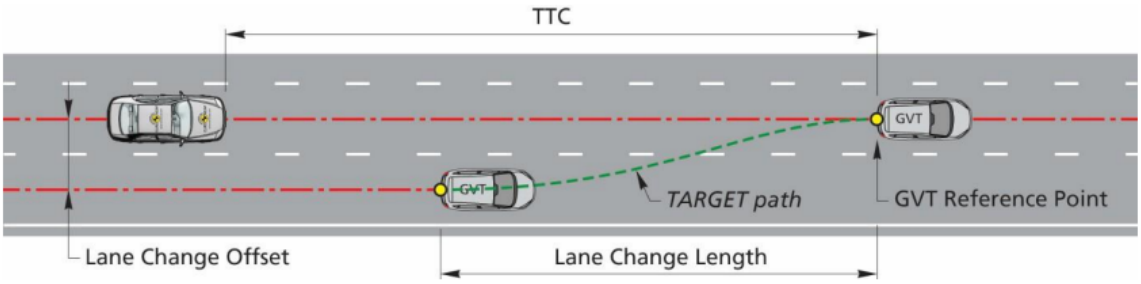
-Traveling on highway (ex: 90kph limited) and lower speed traffic sign visible (ex: exit):



2.4.5 Car to car

- Target Cut-in followed by a braking:

ACC CUT-IN	VUT	GVT	Lane Change Manoeuvre GVT		
			Lateral Acceleration	Change Length	Radius of turning segments
Cut-in					
Cut-in @ TTC = 0.00	50 km/h	10 km/h	0.5 m/s²	14.5 m	15 m
Cut-in @ TTC = 1.50	120 km/h	70 km/h	1.5 m/s²	60.0 m	250 m



The same configuration as ENCAP Highway Assist can be used for the Cut-in part.
This maneuver is followed by a braking of the target with a deceleration of 2m/s² or 6m/s² (same as ENCAP protocol).

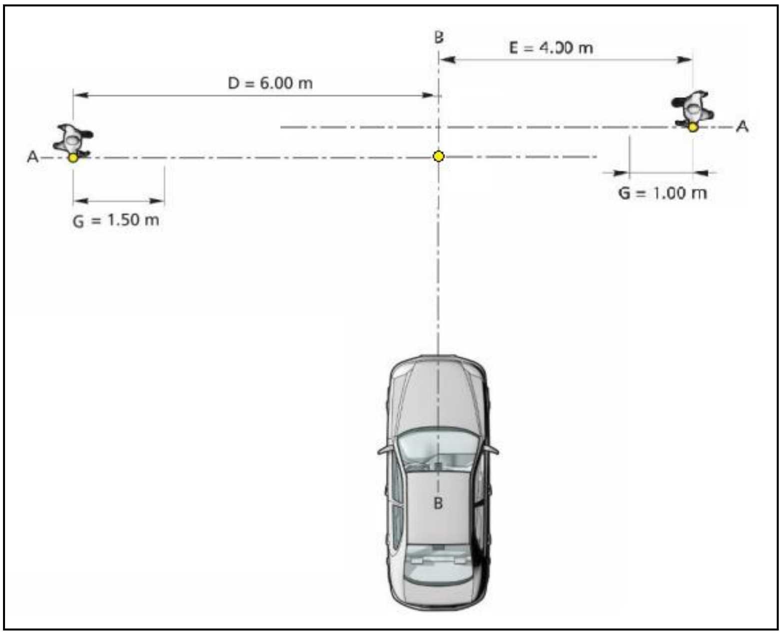
2.5 New random scenarios (to avoid the over-learning of AI (overfitting))

The last category of testing is about random situation, the goal is to perform some random scenario that (in theory) have never been met by the vehicle.

Each scenario will be repeated twice (two runs per scenario).

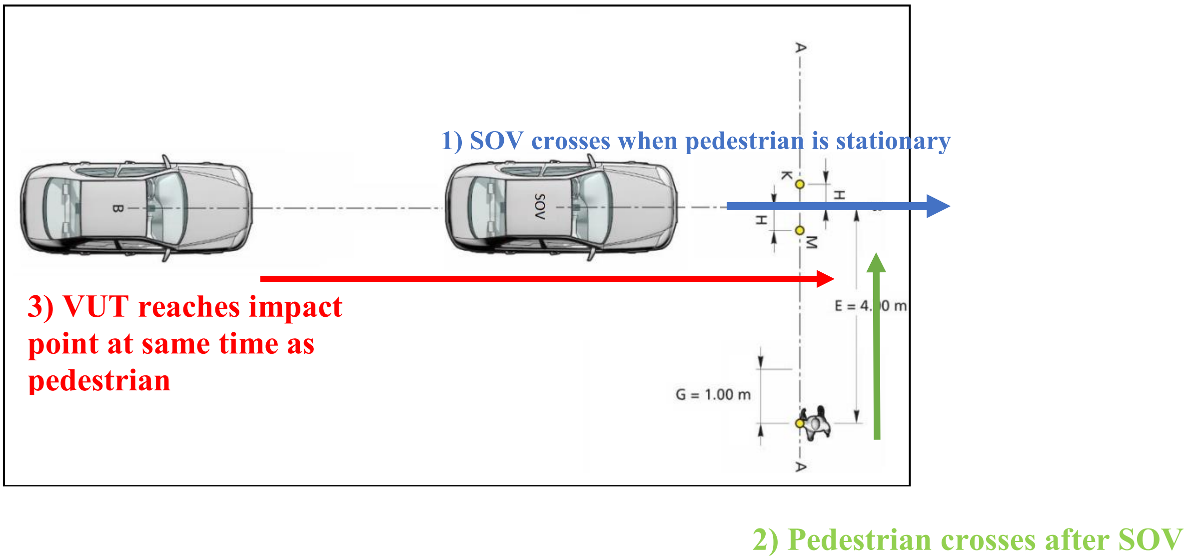
2.5.1 Pedestrians Crossing with two dummies:

Two crossing pedestrians, one from far side, one from nearside, synchronized or not.



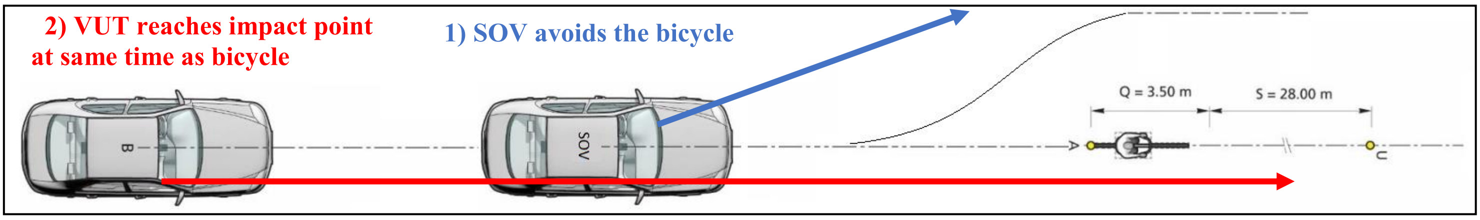
2.5.2 Crossing Pedestrian with VUT preceded by a vehicle:

The VUT follows an SOV (Secondary Other Vehicle) with a distance X, then a pedestrian (adult or child) crosses in front of the VUT. The distance between VUT and SOV depends on the ACC.



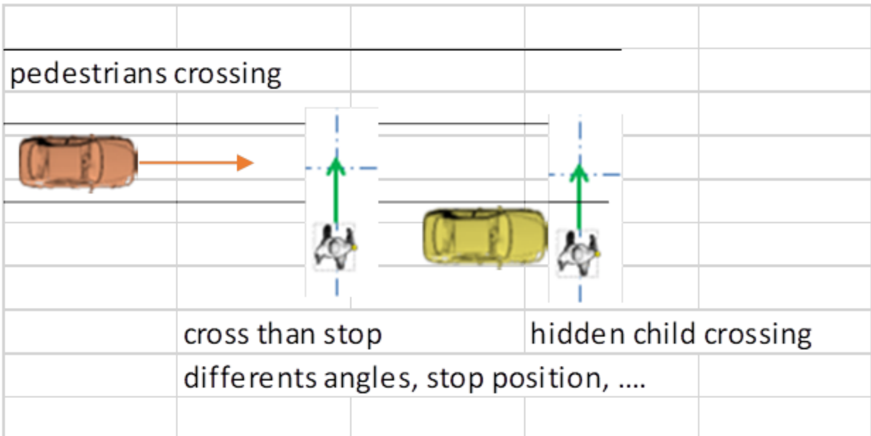
2.5.3 Longitudinal Bicyclist with VUT preceded by a vehicle

The VUT follows an SOV (Secondary Other Vehicle) with a distance to be defined in the same line as a bicycle. At X meters (depending on ACC) of the target, the SOV avoids the bicycle.



2.5.4 Crossing Pedestrian with two dummies, one stops before impact

This scenario is similar as the CPNCO, a second pedestrian is added and starts to cross the VUT path before the stationary vehicle, then stops before the impact.

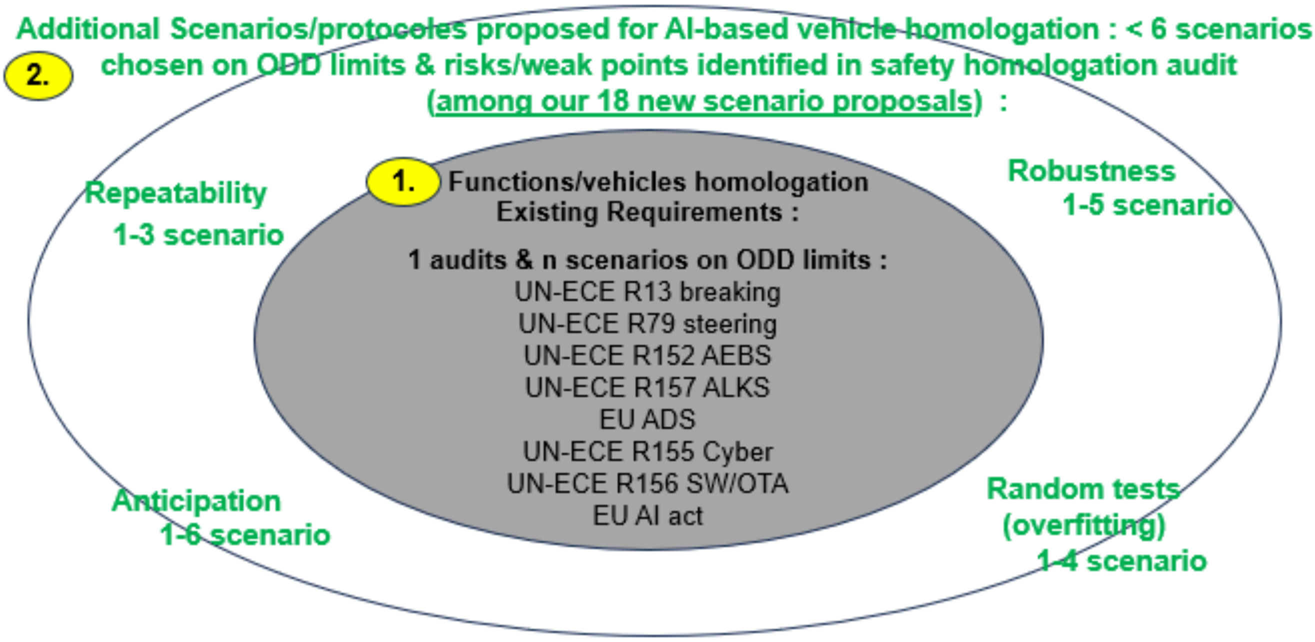


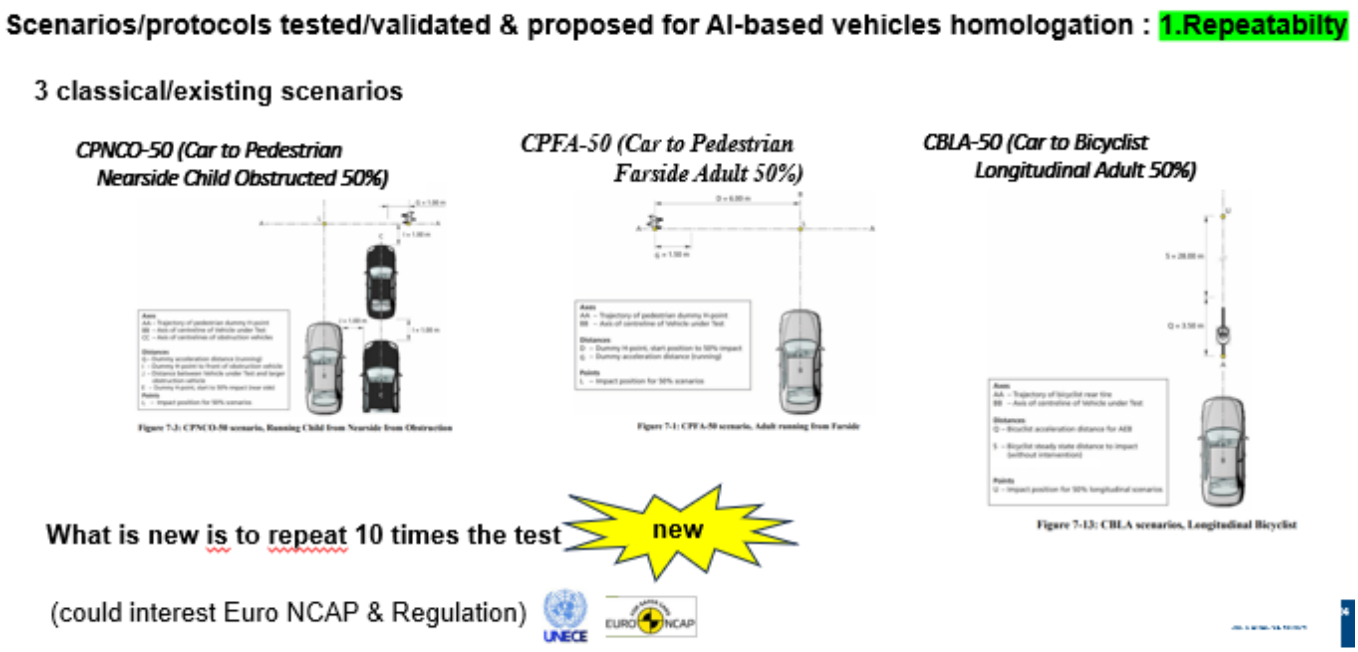
2.6 Summary of UTAC WP3 scenarios and tests built and validated as feasible & relevant to homologate AI-based vehicles:

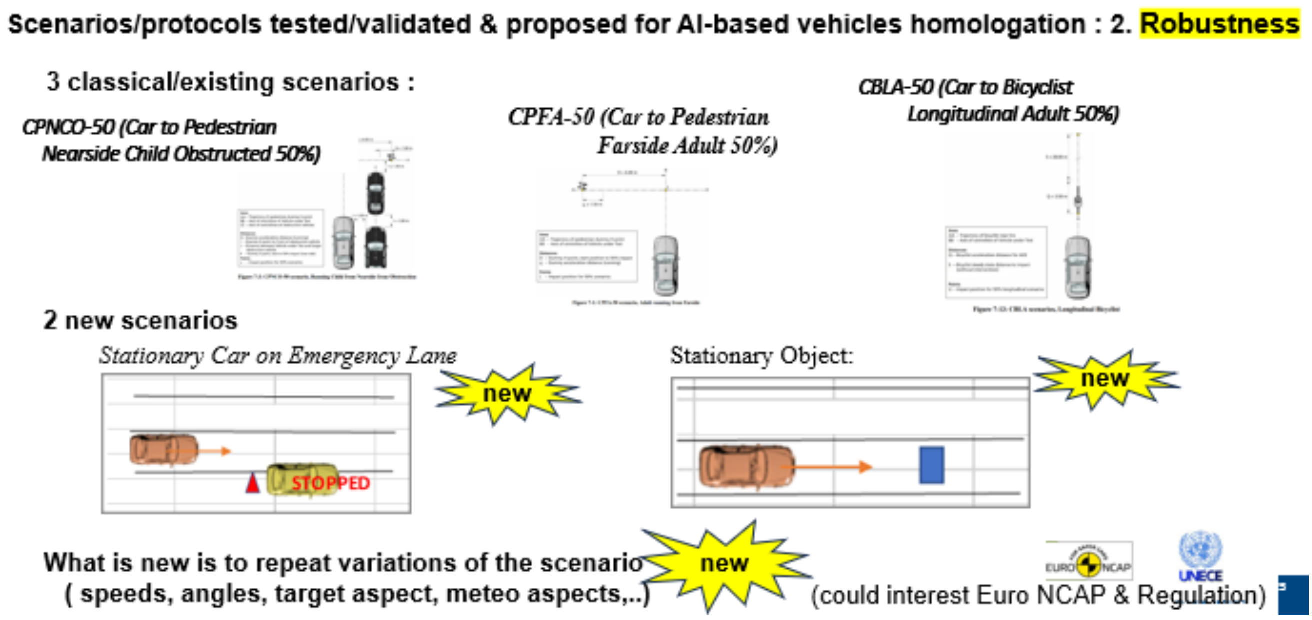
Finally, we can summarize our scenarios proposals with the five figures below related to

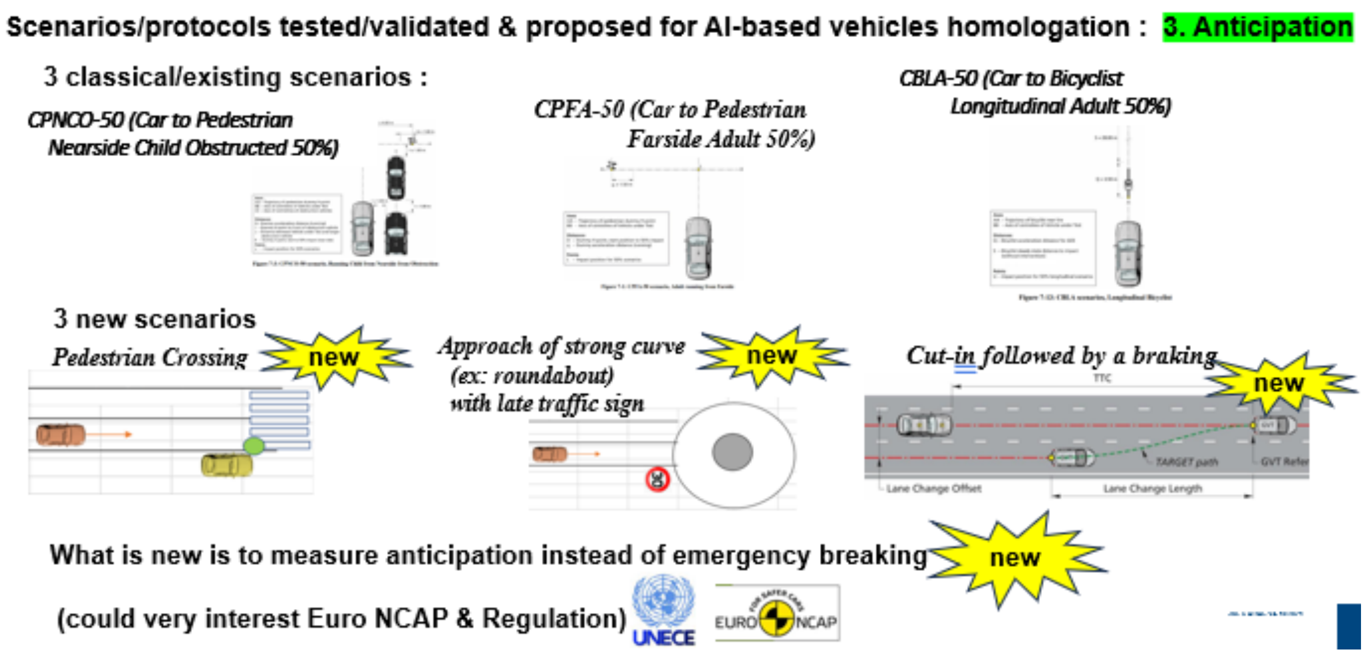
- The 4 axes of new tests/protocols proposed to complete today vehicle homologation
- Detail of the four axes proposed: repeatability, robustness, anticipation and overfitting.

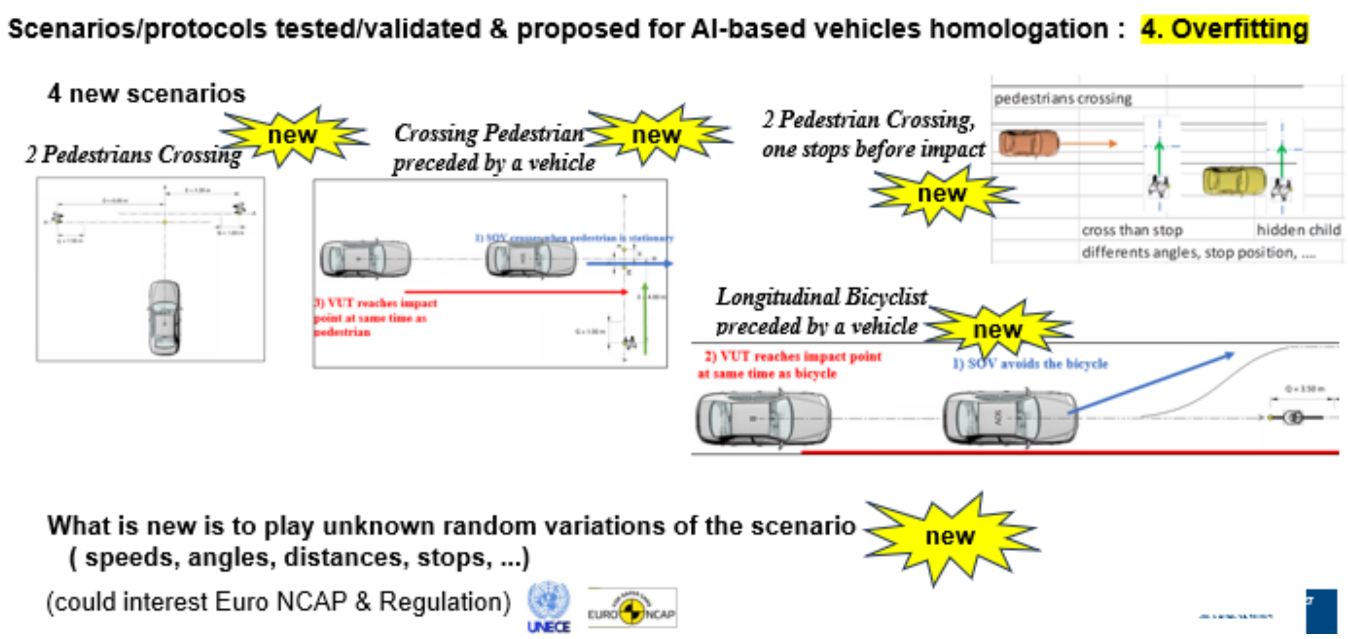
We finally have built, tested and validated as feasible 18 new scenarios and protocols, and what is also new is to repeat tests for repeatability, robustness, or to build random parameters of the tests for overfitting tests.






Scenarios/protocols tested/validated & proposed for AI-based vehicles homologation : 2. Robustness
3 classical/existing scenarios :
CPNCO-50 (Car to Pedestrian Nearside Child Obstructed 50%)
CPFA-50 (Car to Pedestrian Farside Adult 50%)
CBLA-50 (Car to Bicyclist Longitudinal Adult 50%)
2 new scenarios
Stationary Car on Emergency Lane
new
STOPPED
Stationary Object:
new
What is new is to repeat variations of the scenario
(speeds, angles, target aspect, meteo aspects,..)
new
EURO NCAP
UNECE
(could interest Euro NCAP & Regulation)


Scenarios/protocols tested/validated & proposed for AI-based vehicles homologation : 3. Anticipation
3 classical/existing scenarios :
CPNCO-50 (Car to Pedestrian Nearside Child Obstructed 50%)
CPFA-50 (Car to Pedestrian Farside Adult 50%)
CBLA-50 (Car to Bicyclist Longitudinal Adult 50%)
3 new scenarios
Pedestrian Crossing
new
Approach of strong curve (ex: roundabout) with late traffic sign
new
Cut-in followed by a braking
new
TTC
TARGET path
Lane Change Offset
Lane Change Length
What is new is to measure anticipation instead of emergency breaking
new
(could very interest Euro NCAP & Regulation)
UNECE
EURO NCAP



2.7 metrics and requirements

2.7.1 What are the today metrics and requirements for ADAS-AD homologation?

Homologation and regulations have to be very simple to guarantee safety verifications in reasonable duration and costs.

So today tests to type approve one new function are generally very few, one to 10 most of time, and 20 to 50 tests for the most complex functions like autonomous driving functions: AEBS, ALKS, ADS,
On many scenarios and configurations (speed, loading of the vehicle...).

With so few tests, metrics are also very simple, called KPI metrics. It is impossible to apply most WP1 recommendations and metrics, because they supposed to have thousands of test results, which is maybe possible in WP2 with simulation but impossible in WP3 with physical tests; Also remember that in today ADAS and AD regulations no virtual tests are allowed to replace physical tests on closed tracks. It is in discussion in GRVA but still not decided.

For example, most complex and recent intelligent functions like AEBS, ALKS, and ADS have the following “KPI” basic metrics, like for UN-ECE R152 AEBS regulation, the most deployed today:

- Do for each scenario and parameters (speed, loading of the vehicle,) 2 tests.
  If there is one unsuccessful test, do it a third time. If the third test is ok, the scenario is successful

- Do all tests for all categories of scenarios (scenarios with car target, pedestrian target, bicycle target,), the ratio of unsuccessful tests don’t have to be higher than
  - 10% for tests of car-to-car scenarios
  - 10% for tests of car to pedestrian scenarios
  - 20% for tests of car to bicycle scenarios.

As PRISSMA WP1 explains, maybe one day complex metrics based on very many tests will be required to type approve AI-Based vehicles, but as explained previously, Today GRVA regulation group discussions are not very advanced and target to evaluate if existing or soon existing regulations could be sufficient to verify AI-based vehicles safety: Complex systems safety audit annexes , EU AI act, UN-ECE software update and cyber regulations (UN-ECE R155 & 156 regulations). These regulations mainly require audits (of AI and software development, validations, production, reparations, data management, and safety assessment for robustness and black box assessment) but not additional & standardized tests to evaluate vehicles on testing tracks.

Therefore, our proposals of about 6 new scenarios to test and new metrics to verify AI-based vehicles/functions (on repeatability, robustness, anticipation and overfitting) are simple "KPI" basic metrics, and coherent with today homologation metrics.

In addition, these proposals remain to be presented and discussed to the regulation groups with States and OEMs and that could take a long time!

2.7.2 Repeatability metrics and requirements proposals

As previously explained, we propose simple "KPI" basic metrics, coherent with today homologation metrics.

First note that the state of the art for today repeatability is a difficult subject. Today performances references are quite rare and are changing every year! Euro NCAP has surely most information about that. Regularly tests campaigns are down to evaluate vehicles repeatability, but the results are not public and difficult to analyse.

Here are some results for AEBS repeatability as explained in 2.2 chapter:

Today, on a vehicle equipped with classic ADAS systems (ex: AEB), we note that the performances are not always repeatable. Here are some examples of repeatability results on ENCAP AEB (automated emergency breaking) scenarios:

Pedestrian scenarios:

CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	N	Avoidance
CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	N	Avoidance
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	Y	Impact
CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	Y	Impact
CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	Y	Impact
CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	N	Avoidance
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	N	Avoidance
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	Y	Impact
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	Y	Impact

Car to Car scenarios:

CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance
CCRs AEB	Day	20	0	-/-50	AEB	-	-	Y	Impact
CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance
CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance

Bicycle scenarios:

ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	N	Avoidance
ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	Y	Impact
ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	N	Avoidance

Metrics & requirements proposals:

So, such results give references, and authorities and technical services must build and regularly update the references.

The requirement is to:
- Choose 1 to 3 scenarios, among our new scenarios proposals for repeatability, and coherent with vehicle ODD/weak points.
- Do these scenarios/protocols and verify during these tests that the AI-based vehicle performance is still acceptable and not significantly lower compared to these references = average performance of vehicles with few or without AI.

2.7.3 Robustness metrics and requirements proposals

As previously explained, we propose simple “KPI” basic metrics, coherent with today homologation metrics.

First note that the state of the art for today robustness is a difficult subject, because this is a really new subject: all regulations and Euro NCAP protocols are tested on good and nominal conditions, clear weather, dry tracks, no rain no dazzling sun, such more difficult conditions will arrive in these official tests but not before many years. In Euro NCAP roadmaps is planned in 2026 for first tests and in 2029 for “ADAS-AD performances in adverse weather/lights rating”.

So here again today performances references are quite rare and are changing every year! Some Institutions like AAA give regularly many tests’ results and information about that.

Metrics & requirements proposals:

So, Authorities and technical services must build and regularly update performances references (as long there is no standard nor regulation).

The requirement is to see during homologation tests if the AI-based vehicle performance is

- Acceptable and not significantly lower compared to theses references = average performance of vehicles with few or without AI.
- Safe and coherent with manufacturer declarations related to vehicle notice and particularly the vehicle ODD (operational design domain) and limits:

Choose 1 to 3 scenarios (among our new scenarios proposals for robustness) and 2 repetitions tests have to be done with some condition's variations (Target aspect, weather aspects,) related to ODD limits / weak points identified during homologation safety audit).

- If one of the 2 tests is KO, make a third test and verify it is OK

If there are enough test results (through homologation tests),

- verify global performance (< 10% KO for C2C & C2V, 20% for C2B)
  (C2C: test Car to Car, C2P: test Car to Pedestrian, C2B: test Car to Bicycle)

2.7.4 Anticipation metrics and requirements proposals

As explained before in 1.3 and 2.7.1 chapters, Anticipation capability is not yet required in regulation and Euro NCAP is the precursor and incentive for new driving intelligent functions that will improve safety. These new functions, called Speed Limit Information Functions (SLIF) and Speed Limit Control Functions (SLC) by Euro NCAP, do arrive in the future Euro NCAP safety assessments, which are still unofficial and are being discussed in the Euro NCAP WGs (in which UTAC participates).

A first step was down in Euro NCAP SAS 2023 protocol, in which Euro NCAP has introduced bonus points in its vehicle evaluations if such functions of driving can manage (with an alert to the driver or with an automated speed reduction) the situations in the **figure below**, called features:

Systems that are able to properly identify road features where a speed, lower than the legal speed limit, is more appropriate and/or advised or the vehicle should come to a stop can attract points based on the number of road features. These road features are listed in the table below and example traffic signs of a limited number of countries are specified in Appendix I.

ROAD FEATURES	Points	Required Action
Curves*	2	Show and start reducing to appropriate speed
Roundabouts*	2	Show and start reducing to appropriate speed
Junctions*	1	Show and start reducing to appropriate speed
Traffic Lights	2	Warning only
Stop Signs	1	Warning only
Yield Signs	1	Warning only
No Entry	1	Warning only
TOTAL	**10**	

* Only eligible for scoring when linked to ISA and/or i-ACC

Euro NCAP is well aware of nature/numbers/root causes of road accidents in the main European countries and is convinced on the well-known fact (and widely shared by the French authorities in charge of road safety, DSR and ONISER) that excessive speed is the main cause of road accidents.

Euro NCAP will increase in 2026 protocols the challenging and rating of these intelligent speed control functions.

So here again today performances references are very rare and will be changing every year.

As previously explained, we propose simple "KPI" basic metrics, coherent with today homologation metrics.

Metrics & requirements proposals:

Authorities and technical services must build and regularly update performances references (as long there is no standard nor regulation).

Choose 1 to 6 scenarios, among our new scenarios proposals for anticipation tests, and coherent with vehicle ODD and safety audit main points,

Verify during these tests that the AI-based vehicle has the two required reactions:

- an alert to the driver
- a smooth braking instead of emergency braking (value of maximum deceleration requirement proposal: 5ms-2)(this is the limit value of an AEBS in UN-R152 regulation)

2.7.5 *Random test (overfitting) metrics and requirements proposals*

As previously explained in 2.7.1, today GRVA regulation group discussions are not very advanced and target to evaluate if existing or soon existing regulations could be sufficient to verify AI-based vehicles safety: Complex systems safety audit annexes, EU AI act, UN-ECE software update and cyber regulations (UN-ECE R155 & 156 regulations). These regulations mainly require audits (of AI and software development, validations, production, reparations, data management, and safety assessment for robustness and black box assessment) but not additional & standardized tests to evaluate vehicles on testing tracks.

Overfitting tests and requirements are just an idea initially discussed in GRVA regulation group, but today neither Regulation nor Euro NCAP working groups work on such precise regulations with protocols and requirements for AI-based vehicles.

So, we build proposals of scenarios, tests and requirement, with simple “KPI” basic metrics, coherent with today homologation metrics; And these proposals still remain to be presented and discussed to the regulation groups with States and OEMs and that could take a long time!

Metrics & requirements proposals:

Authorities and technical services must build and regularly have to update references for performances (as long there is no standard nor regulation).

Choose 1 to 4 scenarios, among our new scenarios proposals for random tests, and coherent with vehicle ODD/weak points.

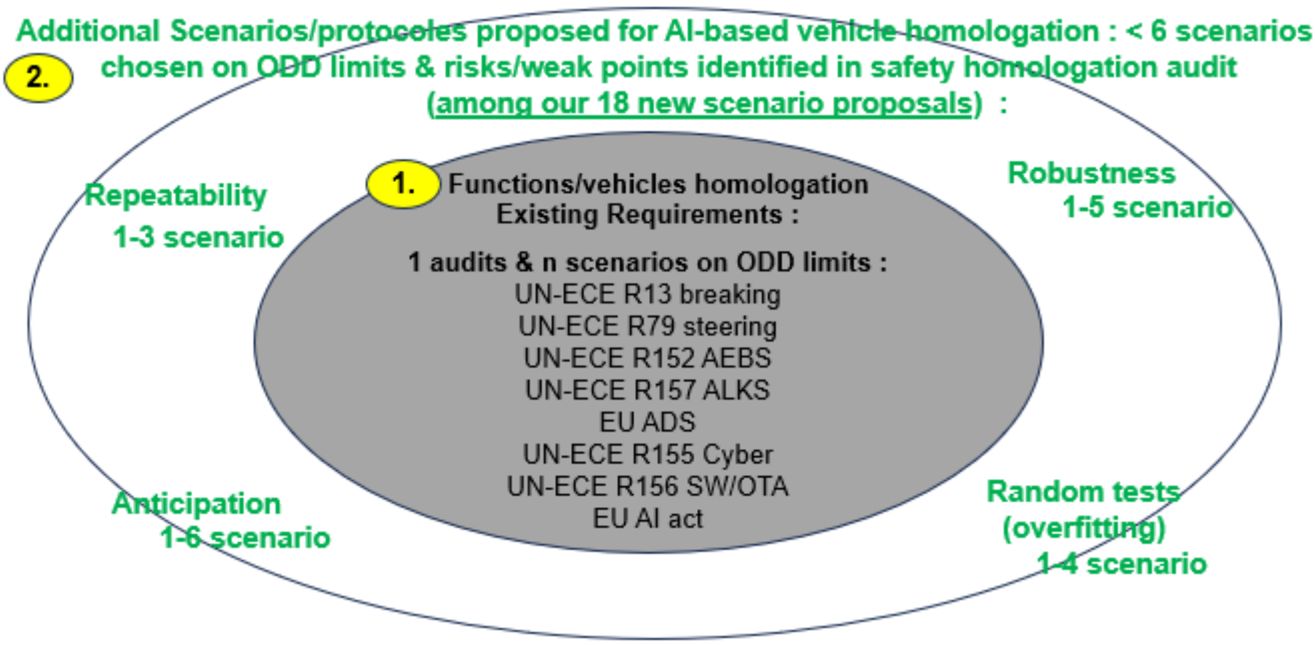
Do these scenarios/protocols and verify during these tests that the AI-based vehicle performance is still acceptable and not significantly lower compared to these references = average performance of vehicles with few or without AI.

2.7.6 Summary of metrics & requirements proposals:

Metrics & Requirements proposals for AI-based vehicles safety verification/homologation:

- Realize for type approval about 6 scenarios (among our 18 new scenario proposals), chosen on ODD limits & risks/weak points identified in safety homologation audit:
- Repeatability tests: 1-3 scenario, 10 repetitions, verify performance is acceptable and not significantly lower than average non-AI based vehicles (technical service or Euro NCAP have to build repeatability standards)
- Robustness: 1-3 scenarios, 2 repetitions with some parameter variations (for example at ODD limits/weak points): speeds, angles, target aspect, weather aspects… If one of the two tests is KO, make a third test and verify it is OK. If data is available through homologation tests, verify global performance (<10% KO for C2C & C2V, 20% for C2B). Note C2C: test Car to Car, C2P: test Car to Pedestrian, C2B: test Car to Bicycle
- Anticipation: 1-6 scenarios, verify alert to driver before a smooth breaking (instead of FCW & emergency breaking) (value of maximum deceleration requirement proposal: $5ms^{-2}$ = AEBS R152 definition)
- Random tests (overfitting): 1-4 scenarios, verify performance is acceptable & not significantly lower than average non-AI based vehicles (technical service or Euro NCAP have to build random tests standards)

These metrics and requirements are related to our additional scenario’s proposal: We propose a maximum of six additional scenario, to complete existing scenarios/tests of regulations. This is described in the figure below:



2.8 Evaluation and process description

2.8.1 Homologation process:

As previously explained, we cannot add to existing homologation tests too much new scenarios and tests to evaluate and homologate AI-based vehicles, so the proposed process is described on figure below:

1. Homologation safety audit

- Identification of vehicle safety weak points / validations
- Identification of vehicle ODD & vehicle limits

⇩

2. Definition & realization of maximum 6 more homologation scenarios to test (or to verify) to safety & AI-potential weak points:

- Virtual tests for dangerous/complex scenarios (verification done by the OEM)
- Physical approval tests for ODD limits scenarios & critical scenarios (realization)
- Open-roads physical approval tests real tests & verifications (realization)
- The six new scenarios can be chosen among the 18 new scenarios/protocols described in chapter 2.5 and summarized in chapter 2.6.

⇩

3. Conclusion and type approval of the AI-Based vehicle

2.8.2 Evaluation description:

We define the PASS/FAIL as:
- PASS: The system reacted and allowed to avoid the collision
- FAIL: The system didn't react OR reacted too late to avoid the collision

To go further in the analysis, we check the following values in the raw data (.txt file):
- **Maximum Speed (kph)** of the vehicle during the test

  For that, we use the channel named "Speed (kph)" and we check the maximum during the test.

- **Minimum distance (m)** between the vehicle and the Target

  This distance is 0 in case of Impact and in case of avoidance we use the channels named "Speed (kph)" and "Relative Longitudinal Distance (m)".
  First, we find the index where the vehicle stops, it means when "Speed (kph)" reaches 0 kph.
  Then, we check the "Relative Longitudinal Distance (m)" value at the same index.

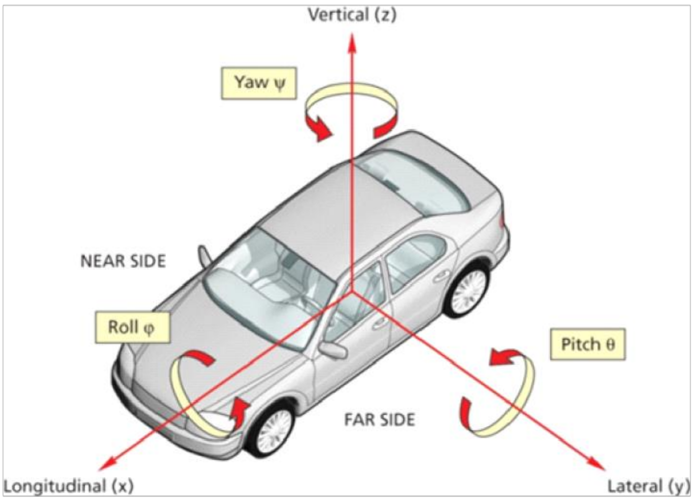
- **2.8.2 Vehicle Impact Speed (kph)** in case of impact

  This is the Vehicle Speed at the time of collision with the Target. We use the channels named "Speed (kph)" and "Relative Longitudinal Distance (m)".
  First, we find the index of the collision, it means where "Relative Longitudinal Distance (m)" reaches 0 m.
  Then, we check the "Speed (kph)" value at the same index.

- **Vehicle Speed (kph)** at driver avoidance in case of it.

  This is the Vehicle Speed at the time of driver avoidance (steering or braking). Depending on the action, we can find the index of the avoidance (huge variation) using "Yaw Velocity (°/s)" or "Forward Acceleration (m/s²)".
  Then we check the "Speed (kph)" value at the same index.

Reference data system



Example of post-processing & evaluation for repeatability (10 tests of one scenario) :

Scenario	Date	Time	Nbr of test	VUT Speed (kph)	Overlap	Success	VUT reaction	Anticipation	Comments	max speed (kph)	remaining distance (m)	impact speed (kph)	avoidance speed (kph)
CBLA	16/02/2023	10:00	1	30	50%	YES	YES	YES	ACC REGULATE, OVERLAP CLOSE TO 25%	28,35	5,3	0	0
CBLA	16/02/2023	10:05	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	32,21	3,93	0	0
CBLA	16/02/2023	10:11	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	31,55	4,06	0	0
CBLA	16/02/2023	10:15	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	29,15	4,48	0	0
CBLA	16/02/2023	10:19	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	29,06	4,46	0	0
CBLA	16/02/2023	10:22	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	28,02	4,79	0	0
CBLA	16/02/2023	10:25	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	28,3	4,82	0	0
CBLA	16/02/2023	10:28	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	32,36	3,6	0	0
CBLA	16/02/2023	10:32	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	30,46	4,14	0	0
CBLA	16/02/2023	10:35	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	28,09	4,99	0	0
CBLA	16/02/2023	10:39	1	30	50%	YES	YES	YES	ACC REGULATE, NO AEB	28,23	5,06	0	0

Example of post-processing & evaluation for robustness (variable parameters tests of one scenario) :

Scenario	Date	Time	Nbr of tests	ACC/AEB	VUT Speed (kph)	Overlap	Target Speed (kph)	camera	wipers with liquid	dummy accessories	Objec in the field of view	Success	Reaction	Anticipation	Comment	max speed (kph)	remaining distance (m)	impact speed (kph)	avoidance speed (kph)
CBLA	16/02/2023		1	AEB	30	50%	15	clean	NO	standard	NO	NO	YES	NO	AEB LATE, DRIVER BRAKE	29,42	0	0	27,61
CBLA	16/02/2023	12:20	1	ACC	30	50%	15	dirty	NO	standard	NO	YES	YES	YES	acc regulate	30,79	4,02	0	0
CBLA	16/02/2023	11:50	1	ACC	30	50%	15	clean	NO	Yellow jacket	NO	YES	YES	YES	acc regulate	28,22	4,91	0	0
CBLA	16/02/2023	11:56	1	ACC	30	50%	15	clean	NO	jacket + backpack	NO	YES	YES	YES	acc regulate	30,01	4,25	0	0
CBLA	16/02/2023	12:05	1	ACC	30	50%	15	clean	YES	standard	NO	YES	YES	YES	acc regulate	29,15	4,46	0	0
CBLA	16/02/2023	11:35	1	ACC	30	75%	15	clean	NO	standard	NO	YES	YES	YES	acc regulate	28,11	4,96	0	0
CBLA	16/02/2023	11:19	1	ACC	30	25%	15	clean	NO	standard	NO	YES	YES	YES	acc regulate	28,16	4,91	0	0
CBLA	16/02/2023	11:40	1	ACC	30	0%	15	clean	NO	standard	NO	YES	YES	YES	acc regulate	27,89	5,24	0	0
CBLA	16/02/2023	11:45	1	ACC	30	100%	15	clean	NO	standard	NO	YES	YES	YES	acc regulate	28,23	1,53	0	0
CBLA	16/02/2023	12:16	1	ACC	30	50%	15	clean	NO	standard	parked car	YES	YES	YES	acc regulate	27,78	4,3	0	0

Example of post-processing & evaluation for anticipation :

Scenario	Date	Time	Nbr of tests	VUT Speed (kph)	Success	Reaction	Anticipation
Highway driving (close and misleading traffic sign 50kph EXIT)	08/03/2023	11:07	1	ACC regulation	NOK	ACC	ACC
Highway driving (close and misleading traffic sign 50kph EXIT)	08/03/2023	11:13	1	ACC regulation	NOK	ACC	ACC

Comment	max speed before traffic sign (kph)	speed after 50kph traffic sign (kph)
Detection of the 50kph traffic sign (for EXIT) and speed adaptation (false positive)	85,26	48,97
Detection of the 50kph traffic sign (for EXIT) and speed adaptation (false positive)	85,74	48,46

Example of post-processing & evaluation for random test (overfitting) :

Scenario	Date	Nbr of test	VUT Speed (kph)	Overlap	Success	VUT reaction	Anticipation
Longitudinal Bicyclist with VUT preceded by a vehicle	17-févr	1	70	50%	YES	YES	YES

Comments	max speed (kph)	remaining distance (m)	impact speed (kph)	avoidance speed (kph)
ACC regulation on SOV then bike detection, then ACC regulation on bike	51,78	120,81	0	0

Data channels definitions

TIME INFORMATIONS		
Channel names	*Units*	*Comments*
Time	s	Time starts in the path
MP Time	s	GPS time of VUT
MP Time Tracker 1	s	GPS time of VRU or GST

VUT SPECIFIC INFORMATIONS		
Channel names	*Units*	*Comments*
Actual X (front axle)	m	X of the car (VUT) (at the bumper)

Actual Y (front axle)	m	Y of the car (VUT) (at the bumper)
Speed	kph	Absolute speed of the car (VUT)
Forward velocity	m/s	Forward speed of the car (VUT)
Lateral velocity	m/s	Lateral speed of the car (VUT)
Forward acceleration	m/s²	Forward acceleration of the car (VUT)
Lateral acceleration	m/s²	Lateral acceleration of the car (VUT)
Yaw angle	°	Yaw angle of the car (VUT)
Yaw velocity	°/s	Yaw velocity of the car (VUT)
Yaw acceleration	°/s²	Yaw acceleration of the car (VUT)

TARGET SPECIFIC INFORMATIONS		
Channel names	*Units*	*Comments*
Head tracker reference X position	m	Position of the VRU on X axis
Head tracker reference Y position	m	Position of the VRU on Y axis
Head tracker forward velocity	m/s	Speed of the VRU on its path
Head tracker forward acceleration	m/s²	Acceleration of the VRU on its path

RELATIVES VUT/TARGET SPECIFIC INFORMATIONS		
Channel names	*Units*	*Comments*
Time to Collision (longitudinal)	s	Remaining time before the VUT strikes the target, assuming that the VUT and the target would continue to travel with the speed it is travelling
Relative longitudinal distance	m	Difference between the longitudinal positions of the vehicle and the target
Relative lateral distance	m	Difference between the lateral positions of the vehicle and the target
Relative longitudinal velocity	m/s	Difference between the longitudinal speeds of the vehicle and the target
Relative lateral velocity	m/s	Difference between the lateral speeds of the vehicle and the target
Relative yaw	°	Difference between the yaw angles of the vehicle and the target

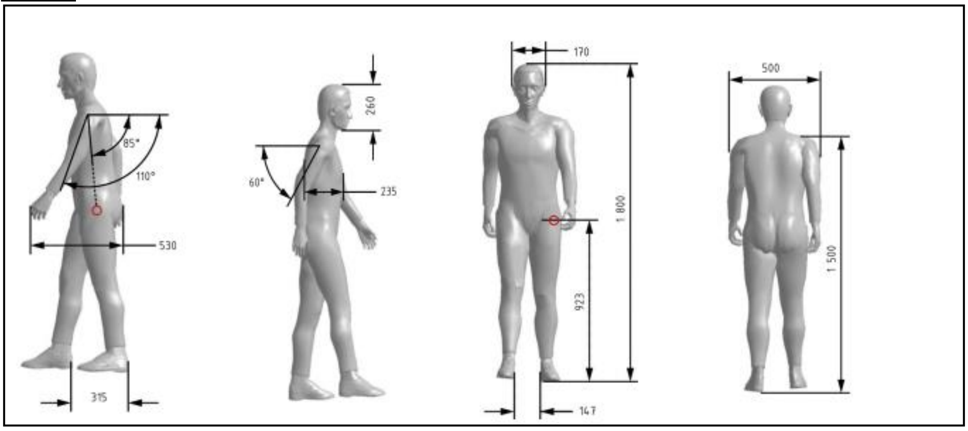
3. TESTING ENVIRONMENT AND EQUIPEMENTS

3.1 TARGETS

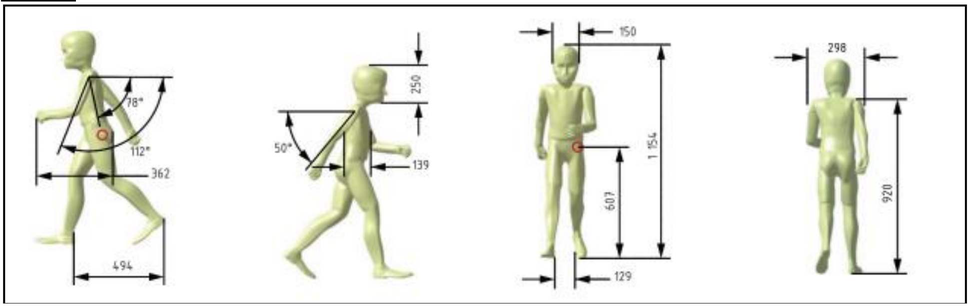
The used targets for the previous scenario are those defined by the **ISO 19206-2_2018** (Pedestrian) and the **ISO 19206-4_2020** (Bicycle). For the Robustness scenarios, the targets will be adapted.

3.1.1 ISO 19206-2_2018:

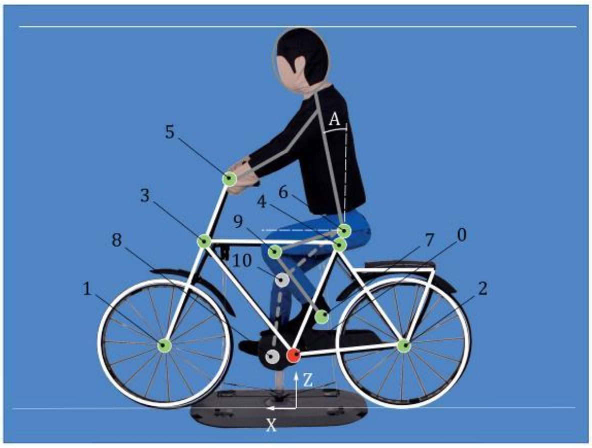
Adult:



Child:



3.1.2 ISO 19206-4_2020: ***Bicycle:***



Segment	X	Z	Tolerance	Unit
0 Centre of bottom bracket of BT bicycle	0	280	±10	mm
1 Centre axis front wheel	670	340	±10	mm
2 Centre axis rear wheel	-540	340	±10	mm
3 Front top frame	430	855	±10	mm
4 Rear top frame (upper range sloped top tube)	-215	860	±10	mm
4 Rear top frame (lower range sloped top tube)	-145	460	±10	mm
5 Handlebar	310	1 180	±10	mm
6 Saddle	-235	935	±10	mm
7 Lower edge left foot[a]	105	495	±20	mm
8 Lower edge right foot	80	200	±20	mm
9 Knee point, left [b]	150	860	±20	mm
10 Knee point, right	85	700	±20	mm
Total height (for 10° torso angle)	1 865		±20	mm
Total length	1 890		±20	mm
A Torso angle	10 and 30		±2	°

[a] Lowest point of shoe – centre line tibia.

[b] Knee point: rotation point of knee.

3.2 PROPULSION SYSTEMS

The propulsion systems used are in accordance with the TB029 of ENCAP.

	Supplier	Product	Version		Euro NCAP Test Targets			
					Global Vehicle Target (GVT)	Euro NCAP Pedestrian Target Adult (EPTa)	Euro NCAP Pedestrian Target Child (EPTc)	Euro NCAP Bicyclist Target (EBT)
				Supplier	ABD	4a	4a	4a
				Product	Soft Car 360	4activePA Adult	4activePA Child	4activeBS
				Version	DRI Rev G Feb 2020	v4v4	v3v3	v5v5
								
Propulsion Systems	ABD	GST100	V1.0 (P8503) & (P8328) with car panel					
	ABD	GST120	V1.0 (P12218)					
	ABD	SPT System	SPT20/SPT20s					
	ABD	LaunchPad 50 & Launchpad 60	V 1.0 (P9226) without extension					
			V 1.0 (P9226) with extension					
	ABD	LaunchPad 80	V 1.0 (P11000) without extension					
			V 1.0 (P11000) with extension					

3.3 VUT equipment:

3.3.1 Motion Measurement

MOTION PACK 1		
	*Manufacturer* Oxford Technical Solutions (OxTS)	
	Unit model TO BE DEFINED	
	Sensors Accelerometers (Servo) / Gyros (MEMS)	
	Data output rate 100 Hz	*Coupling method* GNSS / INS

3.3.2 Data Recording System

CONTROLLER

Manufacturer Antony Best Dynamics (ABD)	
Unit model XR Omni	
Sampling rate 100 Hz	
Analog input voltage ± 10 V	*A / D conversion* 16 bits

3.3.3 HMI Analysis

VIDEO VBOX

Manufacturer Racelogic	
Unit model	*Frame rate*

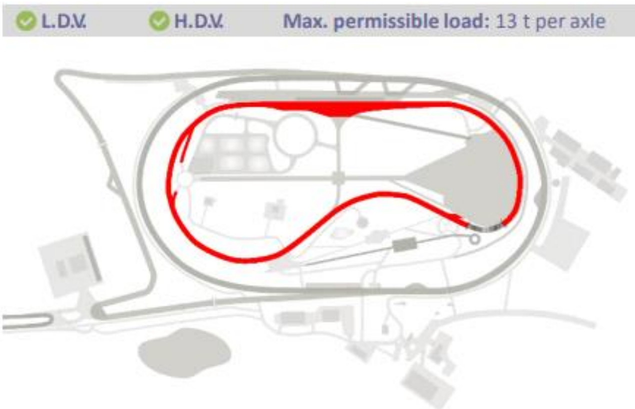
GOPRO

Manufacturer GoPro	
Unit model	*Video resolution*

3.4 TESTING TRACKS

UTAC different test tracks allow simulating different environments:

TEQMO Highway:

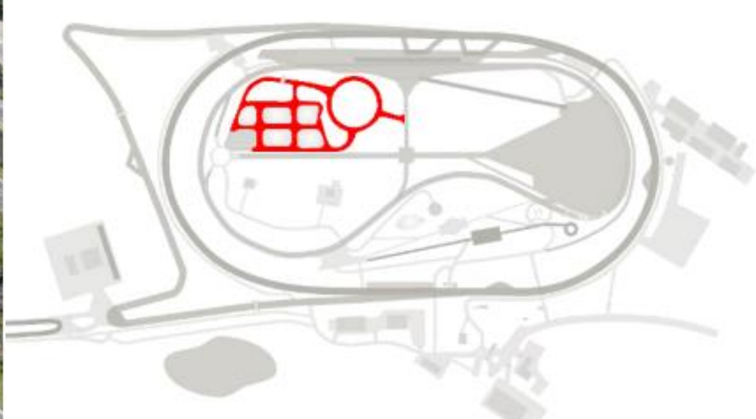


« Highway » circuit

Features
- Length: 2.2 km
- Width: 3.5 m
- 2 and 3 lanes with multiple entries & exits
- Tunnel and tollbooth
- Road marking & signaling

Services
- Platooning
- Adaptive cruise control
- Lane keeping
- Road sign recognition

TEQMO City:

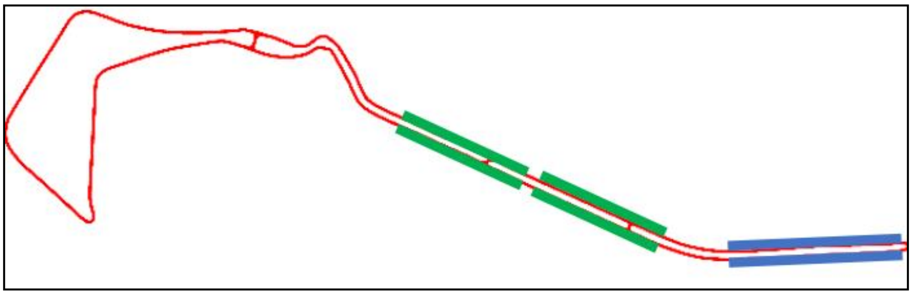


« City » zone

Features
- Area : 38,500 m^2
- 2 and 3 lanes 3.5 m wide
- Lighting & signaling
- 3 crossroads incl. 1 with roundabout and 1 with traffic lights, 7 « T » intersections
- 1 roundabout 50 m radius, 1 level crossing

Services
- Autonomous Shuttle
- Various projects
- Bus stations

Road Circuit Monthléry (MTY):



4. ROADMAP FOR 2024 POC TESTS

Our Need to verify/optimize our protocols/requirements are the following:

- Repeatability: no Need, our 2023 proposals for scenarios/protocols/requirements are fully OK

- Robustness: verify/optimize our proposals for scenarios/protocols/requirements:
  Test variations of parameters (speed, angles, target aspect, weather)
  Test / optimize / complete our requirements proposals

- Random tests: no Need, our 2023 proposals for scenarios/protocols/requirements are fully OK

- Anticipation: verify/optimize our proposals for scenarios/protocols/requirements:
  Test / optimize /complete our requirements proposals

The NEXYAD vehicle is clearly the vehicle with the most capabilities and the most AI inside. In addition, like during the 2023 tests, NEXYAD Engineers will really help us to build / define and measure the most relevant requirements and metrics

Therefore, we planed

- One week of tests
- With this NEXYAD ZOE intelligent vehicle,
- On 22nd to 26th January 2024 in UTAC testing tracks

Chapter 3: INRIA POC "Using Augmented Reality for AI-based ADAS certification (POC UAR)"

1. Introduction

1.1. Objectives of the document

The objective of this document is to present the protocol and the different components of the second test campaign that will be conducted in the framework of the POC proposed by Inria and Transpolis within the PRISSMA project. The tests will be conducted over five days at the beginning of 2024 (early February) at the Transpolis testing site.

1.2. Context

The goal of this proof of concept is to demonstrate Inria autonomous platform (automated Renault Zoe) to showcase the validation of its AI-based perception software stack using a scenario-based approach where dynamic virtual obstacles are injected in the sensor data by INRIA's Augmented Reality (AR) system. In particular, this AR framework includes a data fusion methodology that allows augmentation of LiDAR sensor data in real time. In this way, every element of the testing scenarios can be either real or virtual, offering a smooth transition from simulation to real testing. Thus, this POC has also the objective to prove how augmented reality can be a powerful tool to easily enrich testing scenarios in controlled environments and so make the evaluation and validation process cheaper and safer.

1.2.1. INRIA's autonomous vehicle

INRIA's autonomous platform is an automated Renault Zoe equipped with several sensors for localization and perception. Steering, throttle and brake commands are automated and enabling autonomous navigation through the embedded computer and the software navigation stack. Further description of the platform is provided in section 3.2.

1.3. Presentation of the functions to evaluate

1.3.1. CMCDOT

The CMCDOT framework [1] is a broad perception system, based on Bayesian filtering of dynamic occupancy grids, allowing parallel estimation of occupancy probabilities for each cell of a grid, inference of velocities, collision risk prediction and dynamic object segmentation. From various heterogeneous sensor data, ground form is estimated, instantaneous occupancy grids are generated and filtered using hybrid sampling methods (classic occupancy grids for static parts, particle sets for parts dynamics), into a Bayesian unified programming formalism. Based on this perception framework, navigation systems have been developed and integrated, allowing path finding-and-following, dynamic obstacle avoidance, localization, thus automation of various mobile robots. Also included are communication tools, allowing data fusion from infrastructure systems. The software is composed of ROS packages, which encapsulate the optimized core system on GPU Nvidia (Cuda), allowing real-time application on embedded boards (Tegra X2). First developed in an automotive setting, it is now exploited in other areas of mobile robotics and are particularly suited to highly dynamic and uncertain environment management. Thanks to an important engineering support over the years (notably thanks to IRT Nanoelec), this software has grown to be a core research and development tool of the team, an important technology demonstration and transfer vector, through maintained experimental platforms (most notably the automated Zoe) and associated research contracts and software licensing with industrial partners.

1.3.2. Augmented reality

On the Gazebo simulator, the Inria CHROMA group has developed a virtual twin of its Renault Zoe experimental vehicle. This virtual twin generates the same outputs (sensors messages, localization) that the actual vehicle does, reacts to the same commands, and has a realistic kinematic and dynamic behavior. This allows testing software in Software-in-the-Loop and Hardware-in-the-Loop.
The Inria CHROMA group has also developed an Augmented Reality framework [4] for testing and validation of software on the Renault Zoe experimental vehicle. This framework provides a flexible way to introduce any virtual element in real time in the data of the LiDAR sensors of the vehicle. Our Augmented Reality accurately handles all possible occlusions between real and virtual elements. The representability of tests scenes generated by the augmented reality framework has been experimentally proven. It is then possible to easily and safely place the whole vehicle and all its software, from perception to control, in hybrid but realistic test scenes. This new testing methodology is intended to be a bridge between Vehicle-in-the-Loop and real-world testing.

2. Assessment protocol

2.1. Description of the function to be evaluated

2.1.1. Perception module

The perception module in Figure 1 relies on probabilistic occupancy grids and Bayesian fusion techniques to generate detailed and refined representations of occupancy and velocity. These representations can be effectively utilized in tasks such as planning, risk evaluation, and collision avoidance.

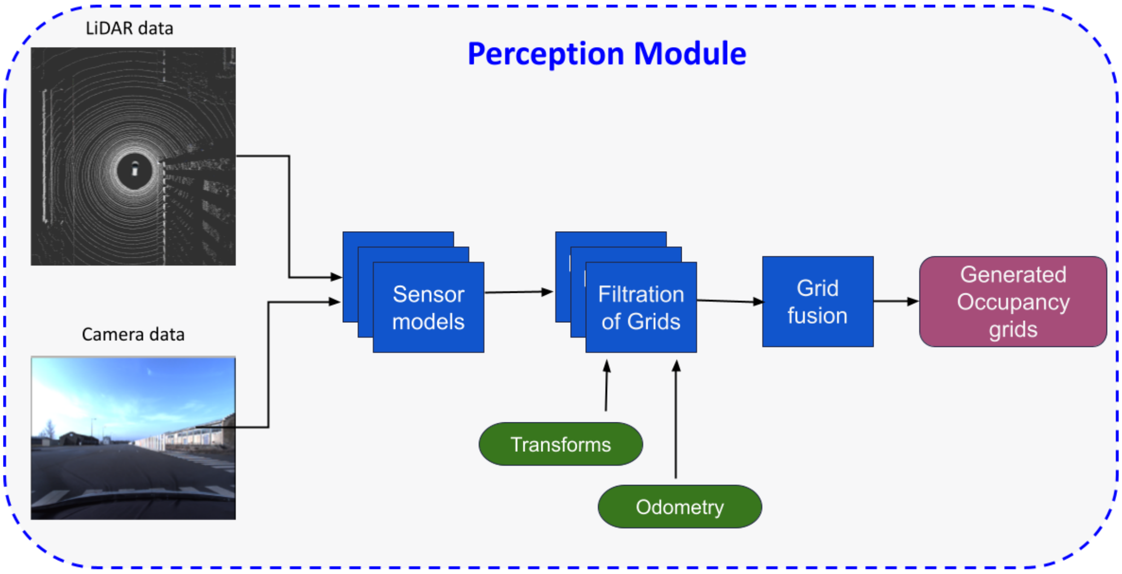


Figure 1: Overview of perception module: Generating detailed occupancy and velocity representations using probabilistic grids and Bayesian fusion for effective planning, risk evaluation, and collision avoidance.

In this POC, the perception module is represented by the CMCDOT framework, developed by Inria, which is a comprehensive method for tracking occupancy in dense environments. This approach draws inspiration from the Bayesian occupancy filter framework, incorporating abstract states and a conditional Monte Carlo technique to optimize velocity estimation and focus on relevant areas. The scene analysis encompasses static, dynamic, free, and unknown states, each associated with dedicated models. The method explicitly considers uncertainty and sensor coverage.

The CMCDOT modules takes the following inputs:

- LiDAR pointcloud data
- Observed occupancy grids or several grids from different sensors
- Odometry and localization of the ego-vehicle

As a result, the CMCDOT module generates the following output grids visualized in Figure 2:Figure 4

1) Instantaneous grid: Initially, a Bayesian model is defined for each sensor. By considering a specific sensor measurement, the sensor model calculates the probabilities of occupancy in the 2D space surrounding the robot. This results in instantaneously updated occupancy grids.
2) Filtered occupancy grid: The instantaneous occupancy for each sensor modality is filtered in both time and space. The CMCDOT occupancy filter, utilizing a Bayesian update model, performs local occupancy filtering while also tracking occupancy changes using a Monte Carlo approach. These yields filtered occupancy grids and velocity grids.
3) Velocity grid: This grid visually represents stationary elements (shown in white) as well as dynamic obstacles (represented by various colors). The intensity of each color indicates the obstacle's speed, while the color itself signifies its direction of motion.

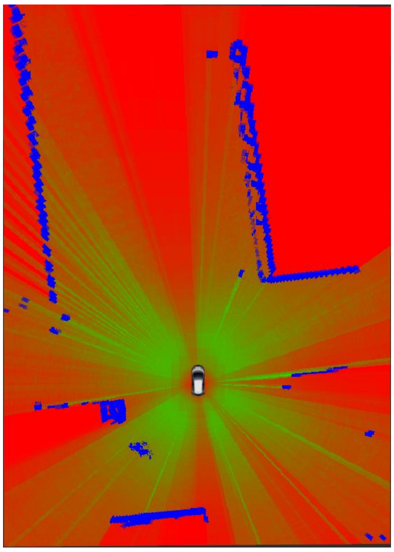


Figure 2: Instantaneous grid: Unknown (red), static and dynamic occupancy (blue), free space (green).

Figure 3: Filtered occupancy grid: Unknown (red), static occupancy (blue), dynamic occupancy (green), free space (black).

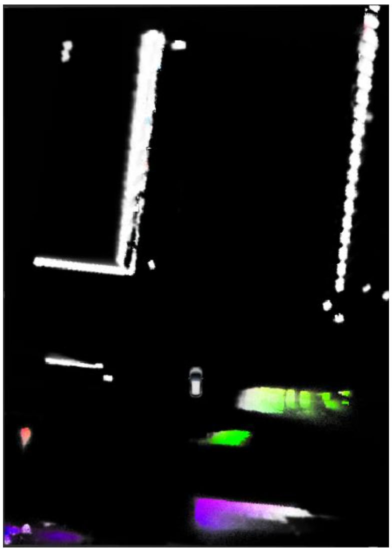


Figure 4: Velocity grid: Static occupancy (white) and dynamic occupancy (varied color intensities).

4) Prediction grid: This grid is an effective model used for predicting occupancy by incorporating essential input data, such as occupancy probabilities and estimated velocity in Figure 5. It projects each cell based on estimated velocity, enabling the representation of movement. To account for noise, cells are divided into particles with specific accelerations and angular velocities.

Acting as a probabilistic distribution, the prediction grid provides insights into future occupancies within a three-second time frame. It merges occupancy grids obtained from various sensor measurements, creating a unified representation that accumulates information over time. The velocity grid derived from LiDAR measurements is preserved as the most accurate estimation of motion.

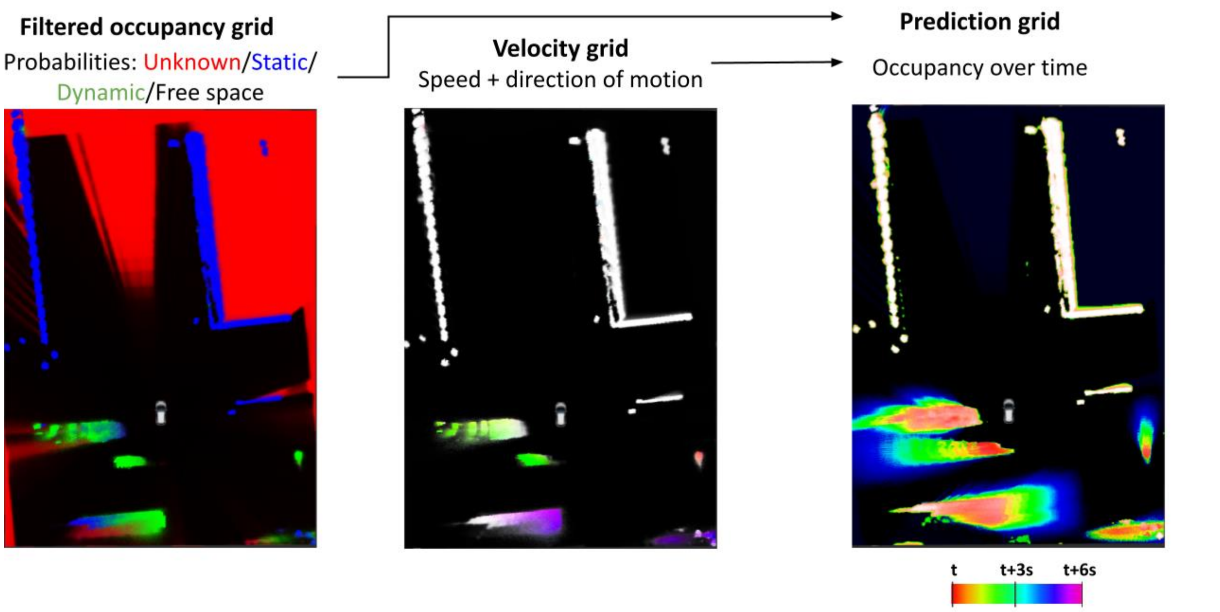


Figure 5: Prediction grid: Predicts occupancy by projecting cells with velocity, incorporating noise, merging sensor data. Offers insights into future occupancies, integrates path planning and obstacle avoidance for comprehensive understanding of dynamic environments.

In addition, the prediction grid enhances occupancy understanding by visualizing predictions over time. Static objects are represented in white, while moving objects are depicted with colors based on their estimated time of arrival. To ensure conservative behavior near moving objects, a large uncertainty is introduced during the prediction process, resulting in the creation of clouds of predicted occupancy. This accounts for potential variations and uncertainties associated with object movement.

2.1.2. Augmented reality

Our AR system consists of the four following modules:

- A virtual environment which contains a twin of the experimental vehicle
- A synchronization module which updates the position and state of the virtual twin
- A sensor emulation which generates outputs from the virtual sensors and integrates them in the actual sensors' outputs
- A visualization that helps testers to understand the AR scene.

Figure **6** proposes a schematic representation of the software framework. The periodic messages of the sensors of the real vehicle give rhythm to the virtual world. So, all modules must run in real time, their execution duration must be short compared to the period of the sensors. This is a heavy constraint on the design and implementation of the solution.

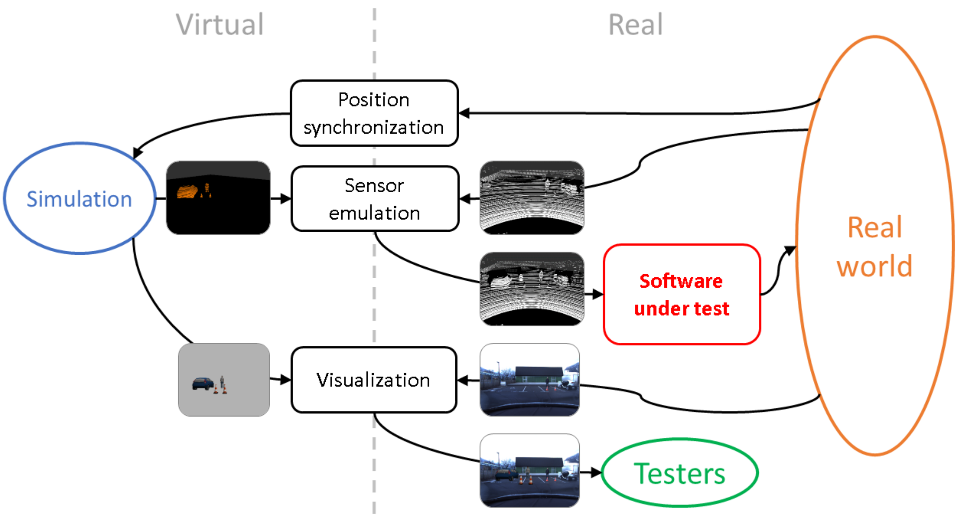


Figure 6: Augmented Reality (AR) framework.

We firstly generate a virtual environment, which is anchored to a real-world position with a reference in GPS coordinates. Then, the virtual environment contains only a virtual twin of the vehicle under test and the virtual elements that we want to add in augmented reality. There is no restriction on the virtual elements of the test scene. The scene can be as complex as required by the test and include any type of object; the only limits are the ones of the simulator. Apart from the virtual vehicle and the test elements, the virtual environment is empty. Our method does not need a background, a ground plane or any representation of the actual test site. This makes this method easy to deploy in a new place.

The absolute position of the vehicle under test must be constantly estimated by an accurate localization system. The estimated position is used to set the position of the virtual twin of the vehicle under test in the virtual environment. This straightforward synchronization gives a great flexibility. The AR system can be deployed without any installation. The virtual twin of the vehicle is equipped with a set of sensors that mimics the sensors of the actual vehicle. An accurate, realistic and real-time emulation of the sensors is needed. Although the framework is generic, for this POC we focus on LiDAR sensors. The emulated LiDARs must return the detection of the virtual objects under a point cloud format. The point clouds are then merged with those returned by each corresponding actual sensor. The merge process is a key component of the proposed AR framework: it must be real time despite the amount of data to process; it must consider a realistic sensor model; it must reproduce all occlusions between real and virtual world. For each sensor, the merge produces a new point cloud that represents the AR perception. It can then be sent to the software of the vehicle under test in place of the actual sensor's point cloud. Thanks to this, the use of AR is seamless for the software under test. For more technical details on how the LiDAR virtual and real data are merged in real time see [4].

The virtual twin of the vehicle is also equipped with a set of cameras that mimics the ones of the actual vehicle. Thanks to the simulator, the virtual cameras return images of the virtual objects. These images are then merged with those of each corresponding camera. For each camera, this produces a new image that represents the AR perception. It provides the testers with a convenient insight of the AR scene. If using a photo-realistic simulator and a realistic image merge function, this visualization can be used as AR for perception with cameras. However, a simulator with approximate graphics and a simple merge procedure suffices for the purpose of visualization.

2.2. Description of the scenarios or the evaluation database associated with the task

In the context of AV validation, it is not possible to test all the possible interactions an AV will have with other road users. To reduce the size of the scenario database, it is necessary to identify and select critical scenarios that are the most relevant to test the safety of an AV. In this POC we demonstrate this by using the work from project Surca [2, 3] that studied accident reports and road user behavior to identify the scenarios expected to be more challenging to an AV. These scenarios should be where efforts are made for AV validation. The goal of this POC is not to completely test all these scenarios but to demonstrate the feasibility of this methodology; therefore, we choose 4 scenarios that will be executed with Inria AV at Transpolis testing facility.

The scenario database is the same for the two evaluated functions. For this POC, 4 test scenarios were selected from the scenario database of the project Surca. From these critical interaction scenarios, we choose 3 scenarios involving vehicles only (Figure 9: and Figure 7:) [2] and 2 scenarios involving an AV and a pedestrian (Figure 8) [3].

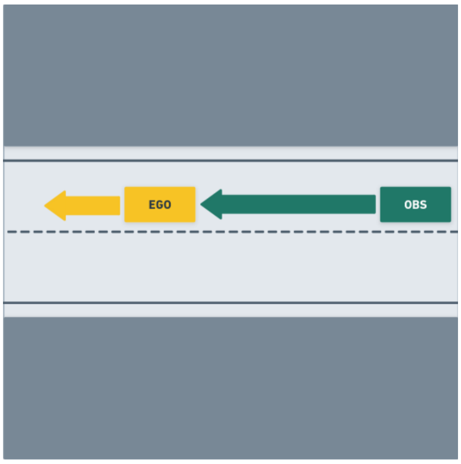


Figure 7: A vehicle is coming from behind the AV much faster or the AV has to perform an emergency braking. The obstacle vehicle will not have time to react and may collide with the AV.

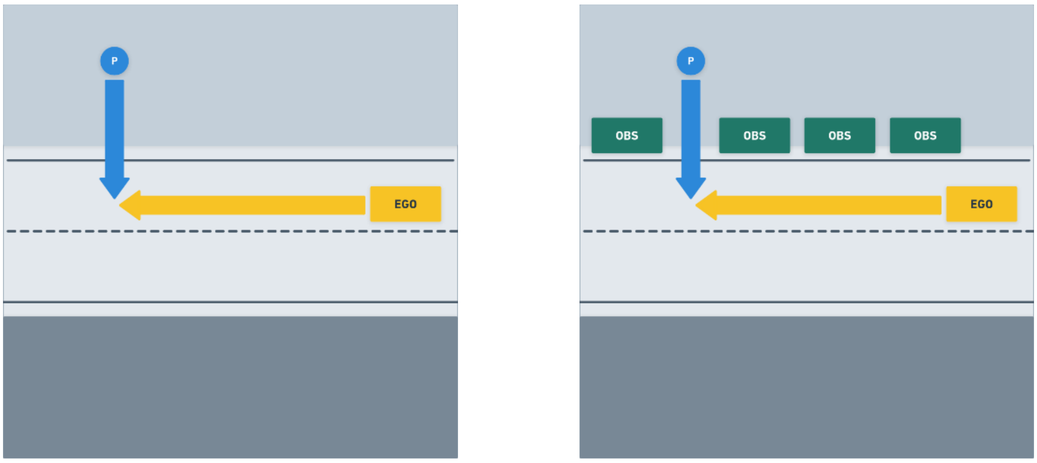


Figure 8: A pedestrians cross the road at the same moment the AV passes, leading to a possible collision. In left scenario, the AV has a clear visibility of the pedestrian while, in the right scenario, the pedestrian is occluded by parked vehicles.

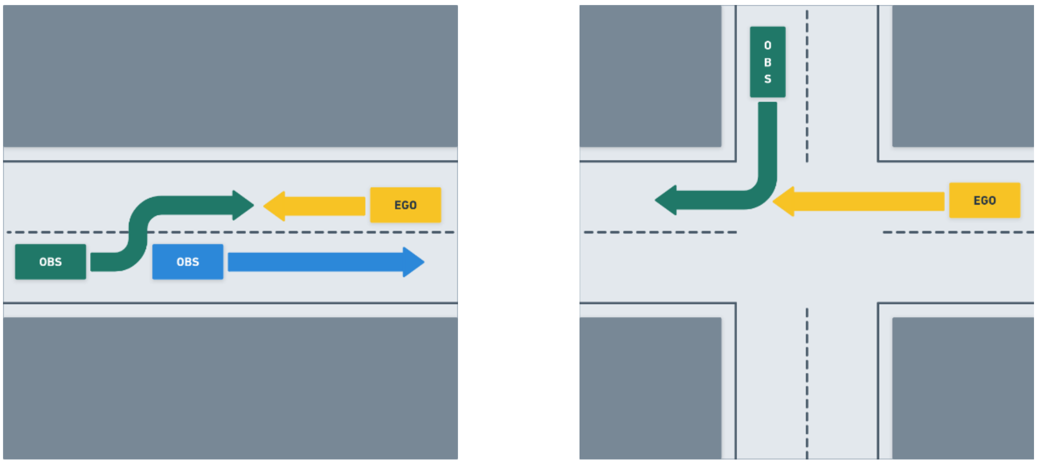


Figure 9: Two-vehicle scenarios. In left scenario, two vehicles drive in the opposite lane of the AV, the further vehicle (in green) tries to overtake the vehicle in front of it (in blue) possibly leading to a collision with the AV. In the right scenario, the AV crosses an intersection at the same time as another vehicle refusing the right-of-way, possibly leading to collision.

2.3. Requirements, metrics and criteria to consider

2.3.1. Augmented reality calibration

Calibration of the augmented reality tool is essential, particularly because the virtual environment relies on the simulator coordinate system, while the real environment uses the geographic coordinate system. The calibration involves finding the transform from one coordinate system to the other, for example it is used to compute the localization of the ZOE in the virtual environment matching the real localization of the ZOE. For experiments conducted at the Transpolis facility, a calibration landmark is centered around the main intersection in the urban area. The manual calibration process involves driving the car to this landmark point and adjusting the parameters to match the virtual location of the landmark with the actual location.

2.3.2. Ground truth of environment

To evaluate our results, we generate ground truth occupancy grids for the perception module (CMCDOT) using a satellite image of Transpolis facility as shown in Figure 10. The Renault Zoé is located on the satellite image using its geolocation during the test scenarios. An approximation of ground truth occupancy grid, fitting the dimensions of the CMCDOT grids, is cropped from the satellite image around the vehicle. However, it only contains static objects and the environment. By using the augmented reality framework, all dynamic objects of the tests are virtual actors. They are controlled by the simulator, therefore their state (position, orientation, speed, footprint) is known for each moment of the test. They are geolocalized on the ground truth satellite image and their footprint is drawn on it.

Figure 10: Satellite image of the Transpolis testing facility.

In the same way, we used the AR framework to merge a real environment with dynamic virtual actors during the test scenario; we generated a corresponding ground truth by merging a static ground truth of Transpolis facility with the ground truth data of the actors from the simulator as shown in Figure 11 and Figure 12.

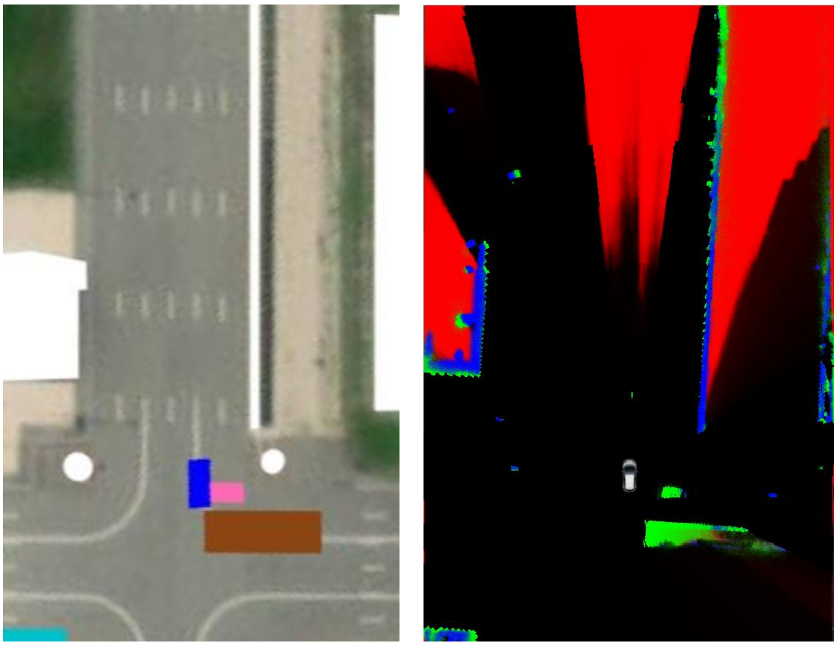


Figure 11: A bus approaches the ego-vehicle but stops and avoids collision. However, an occluded car from behind the approaching bus collides with the ego-vehicle. All dynamic objects in the scenario are virtual actors.

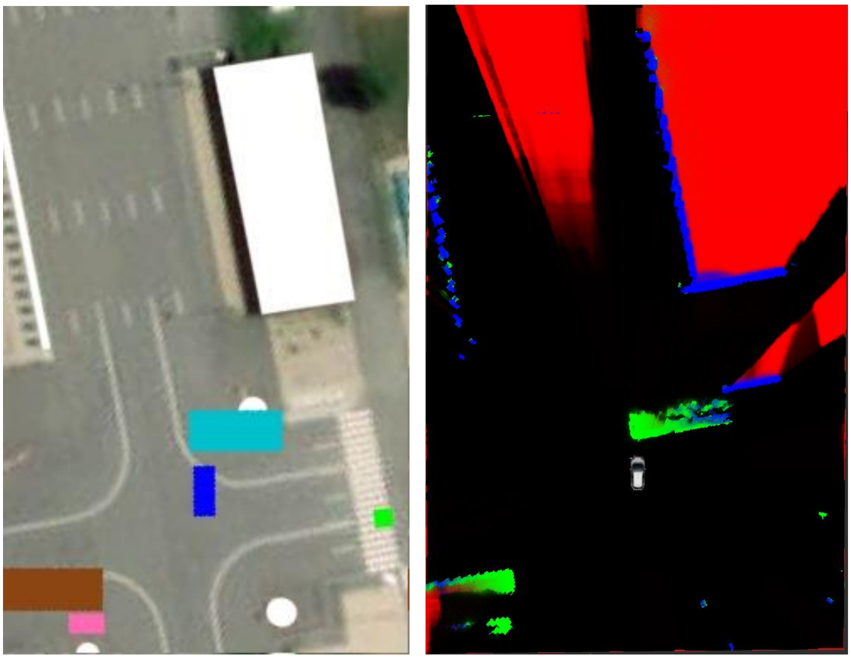


Figure 12: The ego-vehicle applies emergency brakes and avoids collision with a fire truck. All dynamic objects in the scenario are virtual actors.

2.3.3. Metrics

2.3.3.1. Navigation-based evaluation metric for probabilistic occupancy grids

In the context of autonomous driving, OGs are generated by a perception system based on raw sensor data and used for navigation tasks such as Automated Driving Systems (ADS), Advanced Driver-Assistance System (ADAS), and collision avoidance. Despite their im-

portance and broad use in autonomous driving, most existing approaches to evaluate the reliability of probabilistic OGs are based on general-purpose metrics derived from the computer vision literature. In [5], the authors proposed a new metric dedicated to probabilistic OGs that evaluates the similarity of two OGs: a Ground Truth (GT) of the environment and the inference of the environment made by the perception system. This metric is specifically designed to assess the similarity between OGs by considering the behaviour of an ego-vehicle navigating through the grids. The main postulate being that if a navigation algorithm generates similar trajectories using two OGs, the two are alike for navigation purposes.
The metric is computed using the following steps:

1. For each OG, we simulate a navigation algorithm to generate a shortest-path tree, composed of all the shortest paths from the AV position to every cell of the OG (some examples of paths are drawn in red in Figure 13: Evaluation).
2. Both OGs are transformed into cost grids using the cumulative costs of the paths from their respective shortest-path trees (e.g. the cost grids in Figure 13: Evaluation).
3. An intermediate distortion grid is computed by performing the cell-wise absolute error between both cost grids (e.g. the distortion grid in Figure 13: Evaluation).
4. The metric is evaluated by computing the MSE of the distortion grid weighed by the disjunctive probability of free occupancy on both grids (i.e. the probability of a cell to be free on either of the grids).

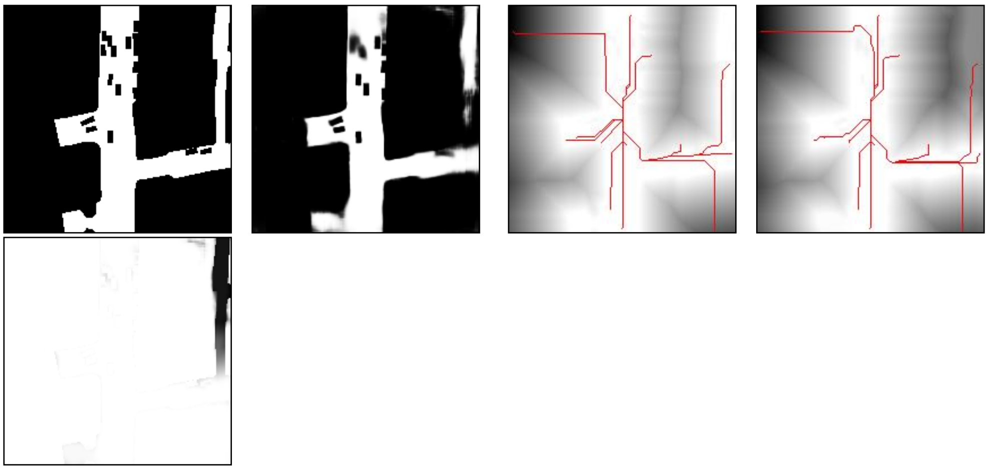


Figure 13: Evaluation process of the metric from [5]. From left to right, first and second images are the Ground Truth and its cost grid, third and fourth are the inference and its cost grid. Examples of paths are drawn in red on both cost grids. The fifth is the resulting distortion grid of the pixel-wise absolute error between both cost grids; it is also weighed by the disjunctive probability of free occupancy on the GT or the inference. The metric score is obtained by doing an MSE pooling the distortion grid, the example scene metric score is 41, 47. The driving scene (ground truth and sensor data used to generate the inference) is taken from Nuscenes dataset; the ego-vehicle is located at the center of the grids.

Simulating and comparing navigation behaviour instead of doing cell-wise comparison directly on the OGs gives this metric relevant properties. It can evaluate topological errors; it measures how occupancy errors on the inference changes the cost of the paths and how it affects the global topology of the OG. It puts emphasis on cells that are most crossed by paths since their occupancy is incorporated in more costs. In other words, these cells are topologically more important for navigation (e.g. areas closer to ego vehicle or bottlenecks). Furthermore, this metric is well fitted to evaluate uncertainty: paths tend to avoid uncertainty whenever possible or cross it otherwise, but in both cases, the navigation cost is increased.

2.4. Description of the tests

The tests are separated in two main parts: 1. execution of the scenarios and data recording; 2. metrics evaluation and data analysis.

2.4.1. Scenarios execution

The five scenarios will all be executed in the urban area of Transpolis facility (the precise locations are still to be decided). The topology of the environment and the road will have to correspond to the pictogram describing the scenarios. For each scenario, predefined trajectories are computed for every actor (the ZOE and the obstacles), they must correspond to the trajectories described by the corresponding pictogram of the scenario. During a scenario execution, obstacle actors are controlled to precisely follow their trajectories without considering possible interactions with the surrounding environment. The ZOE is ordered to follow its trajectory using its autonomous driving system and it should avoid possible collision with other actors. Obstacle actors' trajectories can be considered deterministic among the repetitive execution of a scenario, but the navigation stack of the ZOE is not deterministic. The behavior of the ZOE is expected to be different between two executions of the same scenario where ZOE's trajectory may lead to obstacle avoidance, emergency breaking or collision.
In combination with augmented reality, this POC will use real obstacles provided by Transpolis, called targets. Available targets at the facility are described in Section 3.3. It is not known now of this deliverable redaction which targets will be available. It is expected that the obstacle actors with trajectories possibly leading to a collision with the ZOE will be a real target. Other actors with less interactions or presenting a lower risk to the ZOE will be in augmented reality. This illustrates one of the possible ways augmented reality can enrich test scenarios: the actor with a critical interaction is a realistic obstacle while virtual "satellite" actors increase the complexity of the interactions by creating occlusions or blocking evasive trajectories. This test also provide data to evaluate the impact of virtual or real obstacles on AV validation.
Data will be recorded using the tool rosbag provided by the ROS framework. This tool allows to record timestamped ROS messages (messages sent between processes or for visualization on the ZOE) in bag file and then play the bag to publish the ROS messages in order and at corresponding timestamps, thus replaying the recorded scenario. The following data will be recorded:

- Sensors: LiDARs, camera, IMU, odometer, GPS coordinates.
- Scenario: obstacle trajectories, ZOE trajectory, scenario metadata.
- Perception: occupancy grid, velocity grid, risk grid.
- Navigation: global trajectory, local trajectory, throttle and acceleration commands.

2.4.2. Results analysis

Recorded data will be analyzed offline using the metric defined in Section 2.3.3. Two validation processes will be conducted in this POC: the CMCDOT and the augmented reality framework.

2.4.2.1. CMCDOT

Using the generated ground truths, we can use the metric introduced in [5] to assess the correspondence between the perception module's inferences and the ground truth. While conducting the experiments at Transpolis, we use the CMCDOT as perception module and all its occupancy grids are recorded, along with the necessary data for constructing the ground truth (obtained from the simulator). For each occupancy grid produced by the CMCDOT, we generate a corresponding ground truth with the same parameters as those applied to CMCDOT's occupancy grids as shown in Figure 11. Subsequently, we used the metric from [5] to evaluate the similarity between those grids.
By performing this procedure on a significant number of scenarios we can statistically evaluate the perception performance within the scenario context (i.e. crossing of an intersection in an urban area). As discussed in Section 2.3.3.1, the metric is suitable for a validation process: it

assesses the similarity by considering the behavior of the AV navigating through the grids, effectively evaluating how closely driving using the perception grid aligns with driving using the ground truth.

2.4.2.2. Augmented reality

Augmented reality replaces real actors of a scenario with virtual counterparts with the risk that these simulated actors are less realistic. We propose to evaluate the similarity between scenarios executed using augmented reality and real actors by evaluating the similarity of perception behavior. Scenarios with real obstacles can be replayed with only virtual actors by using the real actor's localization from the ground truth data, therefore, each scenario execution has it augmented reality counterpart to be compared with. This way, the scenario executions are synchronized enough to pair at each time step CMCODT occupancy grids from the scenario executions (an occupancy with real actors and an occupancy grid with virtual actors). The grids similarity is measured using the metric describe in 2.3.3.1. Occupancy grids from the real actor's scenario can be seen as ground truth for the inferences from scenario with virtual actors. This method evaluates the impact of using augmented reality on the validation process itself. The results should be used to evaluate how augmented reality can be integrated in a more general validation framework.

3. Presentation of environments and testing means

3.1. Transpolis

The tests are designed for Transpolis testing facility (Figure 14) and to be conducted in the City Area of the facility. It is a meticulously designed urban testing ground spanning 30 hectares. It features an intricate network of streets covering 12 kilometers, including two prominent boulevards with six lanes each. The area is divided into four sections, each presenting a unique layout with intersections, crossroads, and parking slots.

Figure 14: Transpolis, headquartered at Les Fromentaux, is a cutting-edge testing ground for future urban mobility, where vehicles and infrastructures undergo daily trials with advanced equipment and technologies.

To cater to diverse transportation needs, dedicated bus and cycle lanes have been incorporated. A ring road provides seamless access with three traffic lanes and four access lanes. The City Area boasts 40 real buildings, enabling connectivity testing in both line-of-sight and non-line-of-sight conditions.

The infrastructure is equipped with adjustable facilities like fiber optic cabinets, roving sidewalk configurations, EV charging stations, and a dynamic changing-message sign. Movable signs, traffic lights with GLOSA services and roundabouts cater to multifaceted testing requirements. Road markings, including luminescent lanes, provide precise guidance.

The driving environments include varied surfaces, vegetation, and sloping terrain for comprehensive assessment. Spanning 7000 square meters, the City Area serves as a parking facility and event area for flexible usage. This technologically equipped urban proving ground displays a commitment to advancing urban mobility.

3.2. Autonomous vehicle

The tests will be conducted with the INRIA's Renault Zoé autonomous vehicle, shown in Figure 15, which is equipped with a Velodyne HDL-64 on the top, 3 Ibeo Lux LiDARs on the front and 1 on the back, Spectra SP90 RTK Dual antenna GNSS, Xsens IMU providing vehicle velocity and orientation, a stereo camera and 2 IDS cameras. Data from LiDARS are fused and synchronized using the IBEO fusion box. These sensors are used as inputs for the navigation stack, it is composed of three modules: localization, perception, and navigation. Localization integrates data from various sensors through a Kalman filter, mainly relying on the centimetric RTK GPS. The perception module utilizes probabilistic occupancy grids within the CMCDOT framework [1], which is a Bayesian occupancy filter inferring information on occupancy probabilities, velocities, and collision risk with predicted obstacles. The navigation employs a model predictive control (MPC) with a predictive collision detector (PCD) to guide the vehicle collision freely through the environment. The MPC module forecasts the possible future trajectories of the vehicle from different command samples (throttle, brake and steering commands). The PCD module calculates the expected time to collision of each trajectory with perceived obstacles by predicting their expected behaviors. Ultimately, the command sample minimizing the collision risk is selected and sent to the embedded car controllers to drive the vehicle. The navigation stack offers robust localization, perception, pathfinding and dynamic obstacle avoidance that are well-suited for dynamic and unstructured environments.

Figure 15: Renault Zoe testing platform.

3.3. Target obstacles

3.3.1. Pedestrian

The 4activePA, the official Euro NCAP pedestrian target, meticulously designed to replicate human properties with unparalleled precision. This cutting-edge system comes in two variants, featuring a 50% adult male and a 7-year-old child as shown in Figure 16, allowing for comprehensive ADAS testing scenarios. Compliant with ISO 19206-2, Euro NCAP, CNCAP, JNCAP, and other industry standards, the 4activePA sets the benchmark for accuracy and reliability.

Figure 16: Target obstacle replica for pedestrian 4activePA-adult and 4activePA-child featuring 50% an adult male and a 7-year-old child.

The 4activePA excels with its robust construction, enabling testing in challenging conditions and ensuring durability. Its modular system allows quick spare parts replacement, minimizing downtime. This target's realistic response to Radar, Lidar, Camera, and IR-Systems is vital for testing and validating Advanced Driver Assistance Systems. Offering options like synchronized articulation, combined with compatibility with 4activeSB or 4activeFB-small, the 4activePA is the most efficient solution for ADAS testing. Meticulous attention to detail ensures a comprehensive evaluation of pedestrian detection and collision avoidance systems.

3.3.2. Cyclist

The 4activeBS, the official Euro NCAP bicyclist target meticulously designed to represent a 50% adult male on a standard average European utility bike as shown in Figure 17. This innovative target is exceptionally lightweight with a soft structure to safeguard the Vehicle Under Test (VUT) from damage during testing. Complying with ISO 19206-4, Euro NCAP, UN-ECE, and other industry standards, the 4activeBS ensures realistic properties in size, shape, and rotating wheels for comprehensive ADAS testing scenarios.

Figure 17: The 4active-BS-adult and 4active-BS-child models are the target obstacle replica for cyclists. The replica features both models with equal proportions.

Designed for testing under varied conditions, the 4activeBS features a robust and modular system, enabling an easy and fast exchange of spare parts. It's realistic response to Radar, Lidar, Camera, and IR-Systems makes it an indispensable tool for validating Advanced Driver Assistance Systems. The 4activeBS offers additional options such as different frame colors, and synchronized articulation of pedaling and arm signs. When combined with the 4activeSB or the driverless robotic platform 4activeFB-small, the 4activeBS stands as the most efficient ADAS testing solution, providing versatility and accuracy in assessing bicyclist detection and collision avoidance systems.

3.3.3. Scooter

The 4activeMC as shown in Figure 18, the CNCAP E-Scooter target designed for the upcoming CNCAP 2021 ADAS tests, representing the characteristics of an average Chinese Electric Scooter. Aligned with category L3e-A1 standards set by UNECE, this target complies with ISO/PWI 19206-5 and CNCAP 2021, offering realistic properties in size, shape, and microdoppler features crucial for comprehensive ADAS testing. The 4activeMC's extremely low weight and soft structure prevent damage to the Vehicle Under Test (VUT), enabling testing under rough conditions. Its robust construction and modular system facilitate quick spare parts replacement, ensuring an efficient and adaptable testing solution.

Figure 18: 4activeMC presents the e-scooter designed for the ADAS testing as an target obstacle.

Combined with the driverless robotic platform 4activeFB-small, the 4activeMC emerges as the best and most efficient ADAS testing solution for electric scooters. Additional optional features include synchronized movement, side leaning, and active lighting, providing a customizable testing environment for a thorough evaluation of Radar, Lidar, Camera, and IR-Systems. The 4activeMC is poised to play a pivotal role in advancing testing standards and ensuring the safety and efficacy of ADAS technologies in the dynamic landscape of electric scooters.

3.3.4. Portable belt system

The Soft Pedestrian Target (SPT) system shown in Figure 19, revolutionizes portable testing with its innovative belt propulsion mechanism. This system ensures precise and consistent replication of NCAP and custom-made test scenarios, offering unparalleled accuracy in evaluating vehicle safety features.

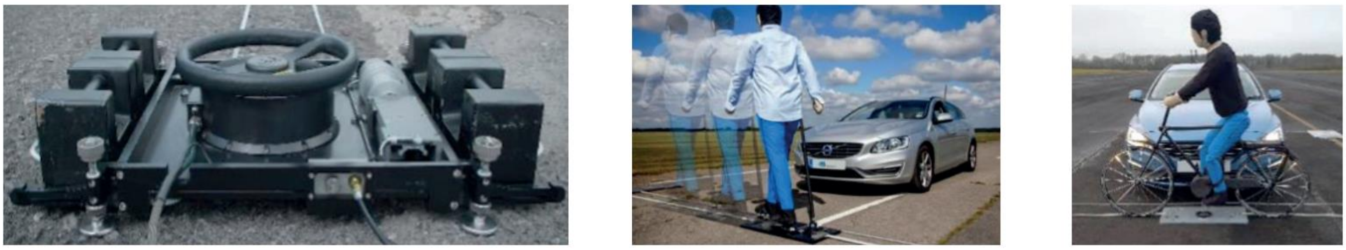


Figure 19: The SPT system consists of a belt propulsion mechanism, ideal for testing soft target obstacles like pedestrians, cyclists and scooters.

Utilizing a standard AB Dynamics controller and steering robot motor for the drive unit, the SPT system is seamlessly integrated with Robot Controller Software, shared across our diverse portfolio of track testing solutions. Customers have the flexibility to employ an existing steering robot as the drive motor or choose dedicated versions with built-in motors capable of reaching speeds up to 40 km/h. The SPT system's adaptability extends to its power options, accommodating a 12V car battery for standard operation or a mains power pack for enhanced performance, available in 115v and 230v versions. This versatility positions the SPT system as a dynamic and reliable solution for portable testing across various testing environments.

4. Roadmap

The main future steps of this POC are:

- Trajectories generation. Prior to the testing week at Transpolis site, the ego-vehicle and obstacles trajectories will be generated based on the scenario pictograms. An execution location in the testing field must be found for each scenario. Based on the scenario location, a target trajectory will be pre-generated for the AV and controlled trajectories for the other scenario actors.
- Experimental tests. During five days in early February (exact dates still to be decided) the five scenarios previously presented will be executed at Transpolis facility with the autonomous Zoé. Each scenario will be retreated several times to record a statistically adequate quantity of data.
- Generation of the ground truth data. For every scenario, generation of the occupancy grid ground truth of the CMCDOT inference using the satellite image of Transpolis facility and the actors' properties.
- Data analysis with the perception metric. Occupancy grids inference generated by the CMCDOT will be compared to their matching ground truth using the metric described in section 2.3.3.1. The statistical evaluation of these results will be used for the validation of the CMCDOT and of Augmented Reality as a testing and validation tool.

Chapter 4: CEREMA and LNE POC : “Simulation Tools Validation for AI-based ADAS certification (POC STV)”

1. Introduction

1.1. POC objectives

More and more intelligent systems on vehicles use AI (e.g., visual or mixed navigation, sign recognition, road tracking and obstacle detection). Certification up to SAE level 3 is now possible for vehicles featuring partially automated driving. The manufacturer must demonstrate that its vehicles ensure adequate safety conditions within their operational design domain (ODD), having conducted tests in diverse scenarios. This task concerns the first braking-related advanced driver assistance system (ADAS) that has been implemented as an ``Automatic Emergency Braking'' (AEB). The qualification of these systems requires verification in all kinds of scenarios, including, for example, considering degraded weather conditions. For cost and safety reasons, these qualification tests cannot be carried out in real conditions, as some tests may present risks or have frequencies of occurrence too low to allow the collection of large series of data. For this reason, sensor simulation tools and degraded weather conditions (physical, numerical or hybrid) must be implemented. These simulation tools can be real (like in the PAVIN Fog and Rain platform), purely virtual (integrating sensor models, as in LEIA 1 and with the Cerema fog model) or can combine the physical sensor with simulated inputs (as is done in LEIA 2). The purely virtual simulators can be physically based or empirical (mainly based on Beer-Lamber theory). The latter family of simulators is real-time but not the former. In the language of certification, which is now being established, we speak about X in the Loop (XiL) testing, with X representing the Software, the sensor (Hardware) or the entire Vehicle. In the HiL and ViL cases, we can imagine that the vehicle's real sensor is fooled by a screen system that makes the vehicle believe it's seeing things that don't exist. The advantage of not relying solely on software during simulation is that other disruptive elements can be considered during testing, such as sensor electronics, system response times, or vehicle dynamics in the case of ViL. These simulation tools need to be validated and qualified, as they may be used for certification. In particular, it is necessary to check these points:

- What scenarios should be considered to guarantee the results obtained on AI-based algorithms in the context of certification? In other words, what are the minimum scenario combinations to guarantee a given level of error and uncertainty during evaluation?
- The repeatability of an evaluation with the same tool: what is the uncertainty induced by the simulation tool on the evaluation?
- The reproducibility of a test from one tool to another: what are the differences in results between the different simulation tools (real or numerical)?

LNE and Cerema have different tools for AI systems evaluation at their disposal, that need to be qualified:

- Cerema’s PAVIN Fog and Rain platform for producing artificial fog and rain.
- Cerema’s K-HiL model that allows to add fog to real images in augmented reality mode.
- LNE’s LEIA 1 simulator to create fully digitally simulated images.
- LNE’s LEIA 2 to replay videos recorded and or fully simulated in front of a real camera, in order to address the HiL purpose.

The aim of this document is to present the protocol that will be used to verify the various objectives mentioned above. This protocol should enable simulation tools to be compared with each other, and to characterize scenarios that enable repeatable evaluation with a known level of error.

1.2. Context

As it is not possible to deal with every type of function at once, we have chosen to concentrate on the pedestrian detection function. This function is a priority, since it guarantees the safety of Vulnerable Road Users (VRUs). It is also already approved for the AEB function. Our three objectives are as follows:

First, how can we guarantee that we have tested a wide enough range of conditions? AI-based algorithms are black boxes, and it is, therefore, very difficult to find their boundary conditions. Indeed, the typology, position, and orientation of the pedestrian can influence the results of the algorithm. Similarly, the environment, disturbing objects, and occlusions can influence the detection. Beyond these geometric issues, weather conditions also have strong impacts, e.g., illumination, camera glare, fog, rain, and snow. Interest in this issue is recent in the field of autonomous vehicles and is the subject of numerous studies, but at present, the works listed in the literature only present cases and not a global solution.

Even if all the conditions required for successful validation have been identified, it is impossible to reproduce them all in real-world conditions. For this, one solution is to use numerical simulation. Many numerical simulators dedicated to autonomous vehicles exist. Most offer variants regarding pedestrians, environments, or weather, but only a few are calibrated against real-world conditions, to our knowledge. The second question is: how can we validate the realism and representativeness of a digital simulator? Will the behavior of artificial intelligence be the same in front of different simulators? To address more exhaustive scenarios, the data can be partially or totally simulated, so X-in-the-loop simulators appear to allow using augmented reality mechanisms. These are simulation tools of this type that we propose to test in this protocol (K-HiL model and LEIA 1).

Beyond numerical simulation, real simulation methods are used to simulate adverse weather conditions. This is the case with the PAVIN fog and rain platform, which can reproduce adverse weather conditions on demand. This platform is calibrated from a meteorological point of view (calibration of intensities, drop size, and velocity). A real physical test must be qualified from a repeatability point of view. In the same way, the repeatability of virtual simulators is closely linked to the determinism of the simulator algorithms. Several sources can affect the determinism of the simulation, they can be classified into two categories, first are due to the algorithms of simulator itself such as randomness, and stochastic processes and the others are due to the hardware and operating system, which hosts the simulator such as floating-point arithmetic or parallelism and concurrency between processes. This is essential in the context of certification tests, where test laboratories are often qualified and audited, making repeatability tests and uncertainty measurements mandatory. Can this type of platform guarantee the repeatability of tests, as well as a standard deviation on the results obtained with AI?

To answer these questions, this protocol introduces a new pedestrian database, focusing on weather (clear weather and fog) and an associated evaluation method of detection tools. That database comprises real data, gathered in clear weather and artificial fog conditions within the PAVIN fog and rain platform, and numerically simulated data (using the digital twin), executed in HiL mode, from a simplistic model prevalently used in most numerical simulators outlined in existing literature. Both real and simulated data are annotated with 2D pedestrian detection bounding boxes.

1.3. Methodology overview

The aim of the present study is to characterize and evaluate the protocols and simulation tools enabling AI algorithm certification, including degraded weather conditions (fog). The evaluation of the used proving AI-based algorithm is outside the scope of this study. The proposed method is therefore as follows. First, an AI-based algorithm, which is applied to intelligent vehicles and representative of the state of the art, is chosen. This algorithm will be used as a proving algorithm for the qualification of the simulation tools. A metric applied to this algorithm will be chosen. Then identical datasets will be prepared using the different simulation tools available.

These datasets will have to include adapted scenarios to evaluate the identified proving algorithm. Indeed, they will include data in clear weather and foggy conditions, but also repeated scenarios to verify repeatability. Finally, the proving algorithm and the associated metric will be applied to all the datasets. A comparison of the scores obtained for each dataset will allow verifying reproducibility from one simulation tool to another. At the same time, this method will make it possible to discuss the repeatability of tests with a single simulation tool, and the minimum protocols to be put in place to guarantee error-free evaluation of pedestrian detectors. The following sections present in detail the protocol, the tests carried out and the databases obtained, the metrics implemented, and the simulation tools involved.

2. Evaluation protocol

2.1. Function to be assessed

As already mentioned, there is a very wide variety of algorithms using AI for autonomous vehicles. Then we have chosen to limit ourselves to the example of pedestrian detection. As a reminder, the aim of the present study is to characterize and evaluate the protocols and simulation tools that are used for the certification, and not the pedestrian detection algorithm itself, which only serves as proving algorithm.
Concerning the pedestrian detection algorithm, the third version of YOLO detection algorithm [12], which stands for "You Only Look Ones", was chosen in this analysis. It is indeed a very common algorithm in the literature on object detection. Moreover, it is very easy to handle. The library of objects available in this version contains 80 items. The algorithm requires two main parameters: the confidence threshold (a value between 0 and 1) of the labeling and the object to label in the images. Only the class "person" is labeled in this study and the confidence threshold chosen is explained in the following section. A frame can get multiple detections with different levels of confidence even though only one pedestrian is walking in the scene into our database.

2.2. Database

In this study, we want to compare the following simulation tools: the PAVIN Fog and Rain platform, the K-HiL fog model, the LEIA 1 fully digital simulator (digital twin), the LEIA 2 simulator to better address HiL purpose. To this end, we propose to acquire the same data for these different simulation tools. We will then compare the results obtained in the different cases. We will also try to measure the uncertainty for some of them.

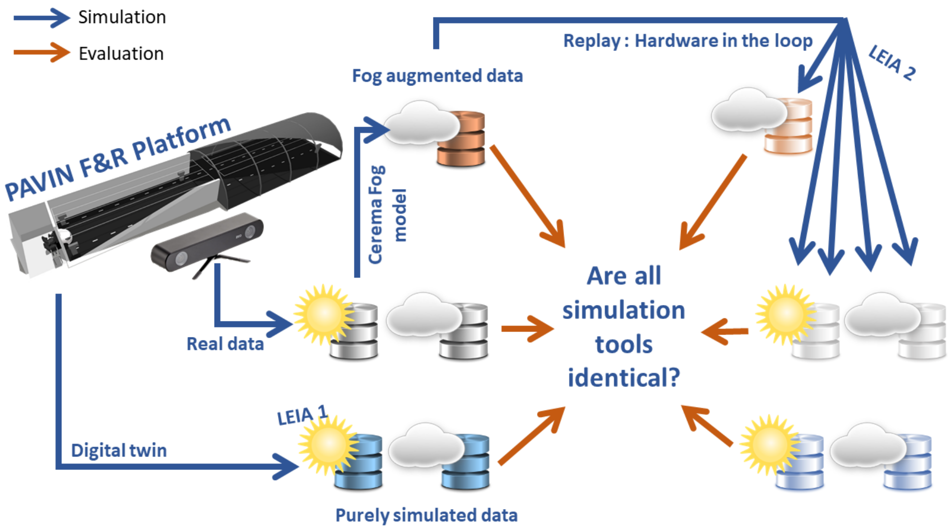


Figure 21: Presentation of the protocol used to obtain the various databases to be compared as part of the POC.

Figure 21 shows the different databases that will be compared. As will be described in Part 2.4, each is a complete database containing pedestrians walking in a road scene. Our proving algorithm (AI pedestrian detector) will be applied on each of these databases. The aim is to compare whether the scores obtained on each of them are similar.
The process used to obtain these different databases is complex and needs to be described in detail. Indeed, some of the simulation tools used enable real data to be augmented (Cerema's K-HiL model), while others enable data to be replayed to add the Hardware in the Loop aspect (LNE's LEIA 2). Figure 21 shows a schematic diagram of how each database is created.
First of all, a first database is created in clear weather and foggy conditions on the PAVIN platform. To achieve this, a camera records real pedestrians making their way through a road scene (dark gray base). The PAVIN platform can reproduce artificial fog conditions on demand. It is described in greater detail in Part 3.1. The scenarios played out on the platform and the sensors used for testing are described in Part 2.4. After the actual tests, a database with real images acquired in clear weather and fog conditions is available (dark gray base).
Next, fog is added to the real data acquired in clear weather using the K-HiL model (dark orange base). This model, described in more detail in Part 3.2, enables fog to be simulated digitally over an image acquired by a real camera. Once the model has been applied, a second database with digitally simulated fog is available.
Thanks to a digital twin of the platform (3D model), the same scenarios are reproduced independently in the LEIA 1 simulator (dark blue base). The LEIA 1 simulator is described in detail in section 3.3. It enables the same data to be created in a virtual world (full 3D simulation). This makes it possible to obtain a third database with clear weather and fog conditions.
From these three simulation tools, one real database (dark gray), one SiL database (dark blue) and one HiL database (dark orange) are obtained. To better address HiL simulations, we use the LEIA 2 simulator, described in Part 3.4. This simulator enables us to replay a database in front of the real camera, in order to obtain images from the real camera, as if it had filmed the scene itself. This is important in the context of vehicle evaluation, as it enables the entire processing chain to be included in the evaluations (sensor, electronics, cables, central processing unit, etc.). The LEIA 2 simulator is therefore used to replay the PAVIN, LEIA 1 and K-HiL databases (dark colors) in front of the camera, resulting in 3 new databases taking HiL into account (bright colors).
In the end, there are 6 databases from various simulation tools to compare. Each of these six variants contains identical clear weather and fog conditions. These six variants are named and summarized in Table 1. In particular, the variants replayed in LEIA 2 have the same name with an *.

Table 1 : Nomenclature and description of the databases used in the POC.

Variant name	Location / acquisition method	Type	Peds num.	Total number of videos
PAVIN	A real camera records pedestrian on the PAVIN platform. The platform can reproduce clear weather or fog.	Real	100	3 weather conditions * 100 pedestrians * 2 sequences = 600
K-HiL	Camera data from the PAVIN database (clear weather) is reused. Using the K-HiL simulator, digital fog is added to the images.	HiL	100	2 weather conditions * 100 pedestrians * 2 sequences = 400
LEIA	The platform's digital twin is used to recreate scenarios in an entirely virtual world, thanks to the LEIA 1 simulator.	SiL	36	3 weather conditions * 36 pedestrians * 2 sequences = 216
PAVIN*	PAVIN database replay into LEIA 2.	HiL	100	3 weather conditions * 100 pedestrians * 2 sequences = 600

K-HiL*	K-HiL database replay into LEIA 2.	HiL	100	2 weather conditions * 100 pedestrians * 2 sequences = 400
LEIA*	LEIA database replay into LEIA 2.	HiL	36	3 weather conditions * 36 pedestrians * 2 sequences = 216

Table 1 shows the nomenclature of each database. It also shows the volume of data in each database. As described in Part 2.4.2, the weather conditions chosen include clear weather, and two fog conditions (medium fog and dense fog). In addition, for each pedestrian/weather combination, the route was replayed and recorded twice, so that repeatability measurements could be made. As a result, a total of 2,432 video sequences are available. Each sequence lasts around one minute, so there are around 40 hours of real, partially or fully simulated videos in the final database.
The first objective of the study is therefore to compare the similarity of the 6 variants. For this purpose, a metric is defined in the next section. The second objective of the study is to measure the repeatability and uncertainty of a pedestrian detector evaluation. To this end, each of the databases will be randomly split into sub-sections. The metric will then be applied to each part as shown in Figure 22. This will also be analyzed. The general structure of the tests has been described, the precise description of the tests carried out is given in section 2.4, while the presentation of the simulation tools used is proposed in section 3. The following section presents the metrics used for the evaluation.

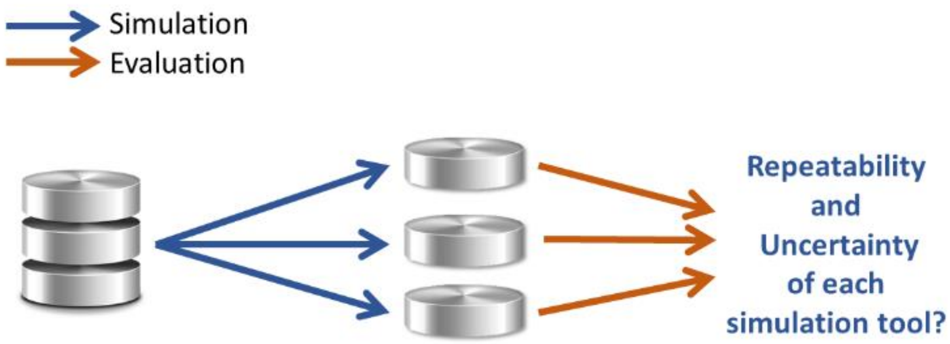


Figure 22: Method used to check repeatability and uncertainty. One of the variants is divided into subgroups, and then the score is measured on each subgroup.

2.3. Requirements, metrics and criteria to consider

As explained above, our approach to comparing and qualifying physical and numerical test equipment is based on analysis of the results obtained by a detection algorithm, rather than on analysis of the raw images themselves. To do so, it is, therefore, necessary to have a proving pedestrian detection algorithm (AI based), a database labeled with ground truth, and a typical evaluation metric dedicated for detection algorithms. In this paper, we have chosen to use the so-called AUC score.
Concerning the pedestrian detection algorithm, the YOLO algorithms was chosen in this analysis. As a reminder, the objective is not the evaluation of the YOLO algorithm itself but to use a popular object detection algorithm, as proving algorithm, to evaluate the main characteristics of the database, and to compare numerical and physical artificial fogs.
As an example, the different detections obtained by the YOLO algorithm, for different levels of confidence, from 0.3 to 1, on two images from the database, are presented in Figure 23. An image can obtain multiple detections with different levels of confidence even though only one pedestrian is walking in the scene in our database. The image on the left of figure 23 shows the 9 YOLO labels with two labels far from the pedestrian present in the scene, yet for one of

them a confidence value greater than 0.5. The image on the right shows labels well-centered on the pedestrian, but with a high variability of the confidence value ranging from less than 0.5 to more than 0.9.

Figure 23: Example of YOLO detections on two clear weather images with different pedestrians. Colors: Green is for confidence > 0.9, yellow is for 0.9 > confidence > 0.7, orange is for 0.7 > confidence > 0.5, Red is for 0.5 > confidence > 0.3.

In object detection, a widely used metric to evaluate the validity of a detection is the intersection over union (i.e., IOU) between bounding boxes. The intersection is calculated following the figure 24. The higher the IOU, the better the algorithm's detection.

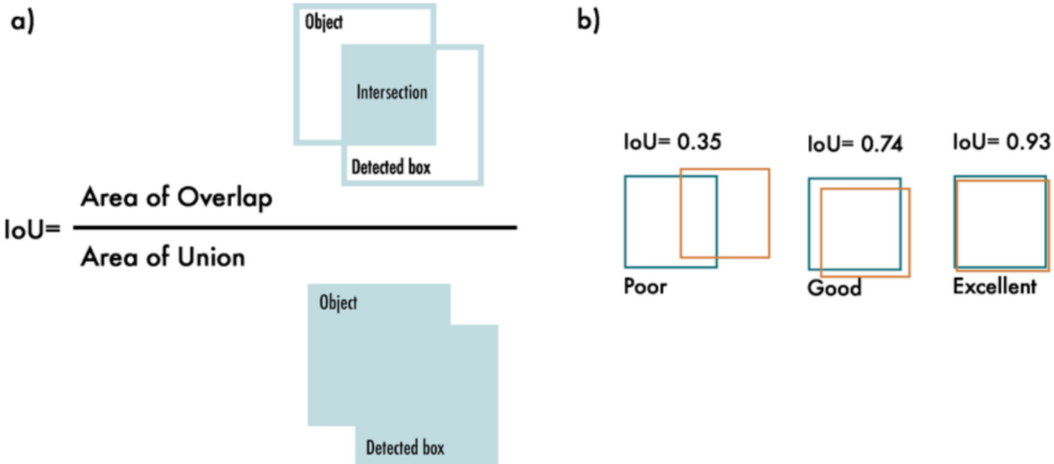


Figure 24: Intersection over Union (IoU). a) The IoU is calculated by dividing the intersection of the two boxes by the union of the boxes: b) examples of three different IoU values for different box locations [13].

The precision--recall curve is then calculated based on the results of intersection over union values. The curve shows the trade-off between precision and recall for different confidence threshold values from the YOLO algorithm. Precision is the fraction of relevant instances among the retrieved instances. Recall is the fraction of relevant instances that were retrieved. Then, the AUC score is calculated. A large AUC value represents both high recall and high precision. A high precision value indicates a low false positive rate (good confidence value but no ground truth label), and a high recall value indicates a low false negative rate (low confidence value but ground truth has a label). The AUC score is between 0 and 1. The higher the AUC score, the better the algorithm.

The evaluation method and detection tools just presented are applied to the 6 databases at our disposal. Detailed characteristics of the test carried out to create each database are presented in the next section.

2.4. Description of tests

2.4.1. Sensors

As described above, the aim is to use an AI-based pedestrian detector for cameras to validate simulation tools. The stereo camera ZED2i (See Figure 25) from StereoLab [14] has been chosen and purchased by Cerema for the data acquisition. Indeed, the latter will allow the testing of monocular detection and recognition algorithms (by taking only one channel) but also stereoscopic. This will allow proposing a database in agreement with the literature. Cerema will also make acquisitions in parallel with a thermal camera (Xenics). This will allow labeling the images of the ZED2i camera in dense fog conditions, thanks to a preliminary geometrical calibration. In fact, the pedestrian is almost invisible on the ZED camera in dense fog, which makes labelling very complicated. The different instruments were positioned at the beginning of the greenhouse (See Figure 28). Finally, meteorological conditions are recorded by the PAVIN platform's usual sensors.

Figure 25 : StereoLab's ZED2i camera.

2.4.2. Meteorological conditions

The objective of the scenarios defined for this study is to collect videos containing pedestrians moving in a scene subjected to various weather conditions (clear weather and two types of fog) and seasons using clothing representative of summer or winter.

Fog is characterized in meteorology by the Meteorological Optical Range (MOR), also called visibility, and noted as *V* (WorldMeteorologicalOrganization2009). MOR, expressed in meters, corresponds to the distance at which the human eye no longer perceives contrast on a calibrated white-and-black target. The smaller the MOR, the denser the fog. It is considered that there is the presence of fog for a MOR below 1000 m in meteorology (WorldMeteorologicalOrganization2009) and below 400 m in road context (Afnor 1989).

The three types of weather conditions chosen are:

- Clear weather (CW): it allows to have a reference scene without disturbances due to the presence of fog.
- Medium fog (MF): the visibility is of 23 m allowing to modify the general aspect of the objects of the scene by leaving detectable all the elements of the visible scene.
- Heavy fog (HF): the visibility is of 10 m allowing elements of the background to disappear for stereo camera but not for thermal camera.

These MOR values were chosen to obtain critical fog conditions. Thus, it is certain that these conditions will challenge the proving detection algorithm. Subsequently, the scores obtained by the latter will drop down, which will allow us to check whether the scores are similar for physical fog and numerically simulated fog. Figure 26 shows an example of the images obtained for the three weather conditions of the real data.

Figure 26: Three weather conditions for a daytime configuration of the scene with (from left to right): clear weather (CW), medium fog (MF), and dense fog (DF).

2.4.3. Scene

To recreate a realistic environment, an urban scene with different elements was created in the PAVIN Fog and Rain platform: a Renault Megane vehicle, trees, a wooden picnic table, different traffic signs, ground marking strips, and orange traffic cones, as well as four calibrated targets (a large black and a large grey (50 x 50 cm), and a small white and a small black (30 x 30 cm). A 3D model (digital twin) with all the elements of this scene is also available with the dataset.

Figure 27: Daytime scene of the PAVIN platform for the PRISSMA tests.

For each trial, the pedestrians follow the same path through the platform and repeat it twice, consecutively, to ensure repeatability. Following the different colored lines in Figure 28, the path allows the pedestrian to be presented from the front (paths 4 and 7), the back (path 1) and the side (path 2, path 3, path 5, and path 6), in relation to the camera position (the red star in Figure 28). In addition to walking at a moderate pace, the pedestrians also find themselves sitting on the bench at the picnic table.

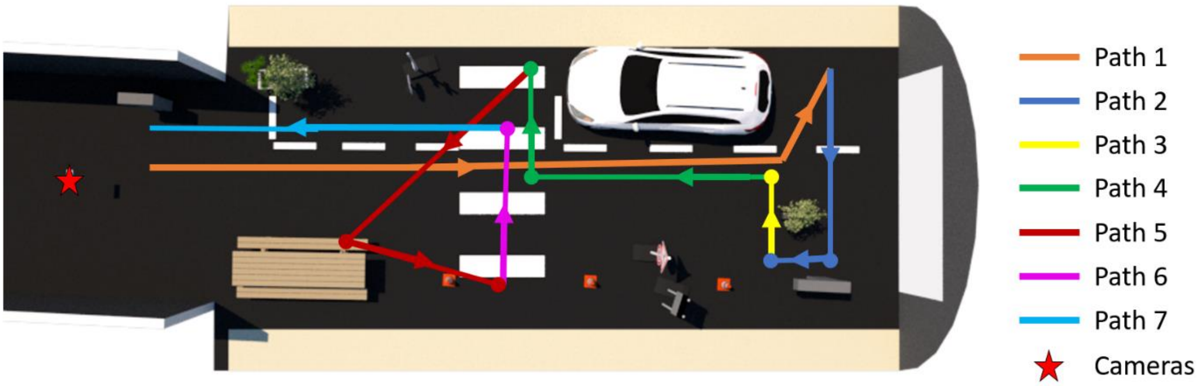


Figure 28: Path of the pedestrians during the tests following the colored lines and arrow directions.

2.4.4. Pedestrian

The databases set up contain both real data and numerically simulated data. In this way, the pedestrians present in both types of databases are described successively.
First, real pedestrians are described. To be representative of a wide variety of pedestrians, different characteristics have been made variable to form the batch of 100 different pedestrians (Figure 29) such as:

- Clothing: 50% of the clothing is representative of summer weather and 50% of winter weather.
- Accessories: a selection of pedestrians carry accessories with different sizes.
- Gender: 60% of the pedestrians are male and 40% are female.

To add a seasonality in the scene (summer/winter), the pedestrians have been dressed in clothes characteristic of high or low temperatures such as: hats, caps, shorts, pants, coats, ... and as much as possible, a variability of the color of the clothes has been respected (bright colors, dark or light colors). Wigs have also been used to increase the number of female pedestrians. Different sizes of accessories have been used in the tests. The objective is to have an impact on the overall silhouette of the pedestrian in an attempt to fool the detection algorithm. Considering the accessories worn by the pedestrian is crucial to guarantee his safety. An object worn by the pedestrian that would not be detected by the detection algorithm of an autonomous vehicle could endanger the pedestrian.
The data can be classified into four sub-lists:

- Small: for small accessories, such as a small backpack, a helmet, a plant, etc.
- Large: for large accessories, such as a large cardboard box, a snowboard, an open umbrella, etc.
- No accessories: when the pedestrian is not wearing any accessory or the accessory does not alter the pedestrian's overall silhouette (e.g., a headlamp, a yellow fluorescent vest, a cell phone).
- All: all pedestrians, regardless of the accessory sizes.

Table 2 shows the distribution of the number of pedestrians by the accessory size category and a thumbnail of the 100 pedestrians in the PAVIN database is shown in Figure 29.

Table 2 : Number of pedestrians per sub-list of accessory sizes.

Accessory Size	Number of Pedestrians
Small	25
Large	33
No Accessories	42
All	100

Figure 29: Thumbnail of the 100 pedestrians of the PAVIN database.

On the simulation side (LEIA database), numerous pedestrians are also available. The 4DV library offer a wide choice of human in terms of gender, ethnicity, age and type of clothing.

Figure 30 summarizes the humans used in the simulation. The appearance of the human can also be modified to change the hair or clothing color. In 4DV each human can be set with 3 appearances which means we have 36 different humans.

Figure 30: Thumbnail of the 36 pedestrians in LEIA database.

2.4.5. Labelling

The labeling consists of tracing a 2D box containing the pedestrian and the accessory that he or she is carrying during the measurement, which clearly has an influence on the bounded box boundaries. The goal is to define the area that the vehicle should be able to detect and avoid. In the case of the LEIA database, labeling is automatic, as the images are fully numerically simulated. Conversely, in the case of the PAVIN database, manual labeling is required.

The PAVIN database contains data from two cameras (visible stereoscopic and thermal). Images from the ZED 2I stereo camera were used to labelize the clear weather and medium fog images. The images from the dense fog test conditions were labeled using the thermal camera images. Indeed, as can be seen in Figure 31 the pedestrian is barely detectable on the right ZED 2I visible image when located at crosswalk level and even invisible on the left ZED 2I visible image when he is at the end of the platform. In both cases, the pedestrian is easily detectable on the thermal images. A geometric and temporal calibration is used to labelize the pedestrians on the images of the ZED 2I camera and the thermal camera, as shown in Figure 31.

The protocol is therefore described in detail in this section. It is based on the following simulation tools: the PAVIN Fog and Rain platform, the K-HiL fog model, the LEIA1 digital simulator and the LEIA2 HiL replay simulator. These various simulation tools are described in the next section.

Figure 31. Examples of the field of view of the ZED 2I stereoscopic camera (top images) with a synchronized field of view of the thermal camera inside the red rectangles (bottom images).

3. Presentation of test environments and simulation tools

3.1. Real Simulated Fog: PAVIN Fog and Rain Platform

The PAVIN database is recorded in the PAVIN fog and rain platform [15]. It allows the production of various and reproducible fog and rain conditions. The PAVIN fog and rain platform is a facility situated in Clermont-Ferrand (France). The platform dimensions are as follows: 30 m long, 5.5 m wide, and 2.20 m high. Its dimensions allow the reproduction of an urban scene, and, thanks to a removable greenhouse, it is also possible to reproduce day or night conditions on this platform [15]. Figure 32 shows a scheme of the platform. Only the “Day and Night area” on the upper part of Figure 32 has been used to create the urban scene. This part of the platform is 18 m long and approximately 8 m wide.

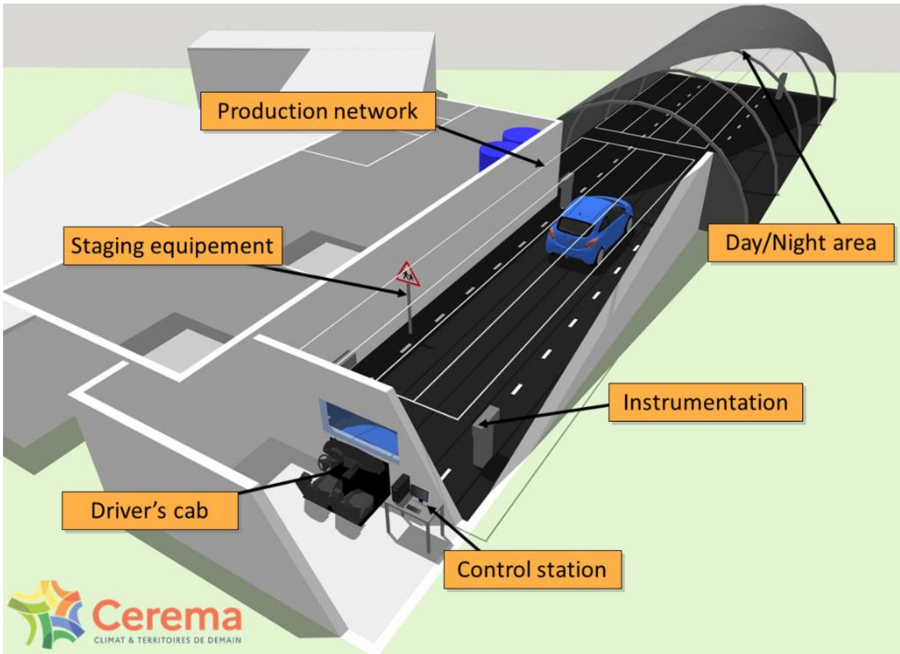


Figure 32: PAVIN fog and rain platform scheme.

3.2. Numerically Simulated Fog: K-HiL Model

The numerically simulated fog on a clear weather visible image is obtained by applying a loss of contrast. The most popular method to simulate fog is to use the visibility attenuation theory of Koschmieder, defined a century ago [16]. This theory makes it possible to determine the luminance of a black object on a sky background by an attenuation of the visibility due to the extinction of the medium between the object and the observer. According to the Koschmieder law, the visibility V (in m) is related to the extinction coefficient β_{ext} (in m^{-1}), if we consider that the minimum contrast identifiable by an observer is 0.05 (i.e., 5% [17]).

$$V = \frac{-\ln(0.05)}{\beta_{ext}}$$

The transmittance of a pixel at position *(x,y)* in the scene is a relation between the distance $d_{x,y}$ from a target to the observer and the extinction coefficient β_{ext} of the medium (in m^{-1}):

$$t_{x,y} = \exp(-\beta_{ext} d_{x,y})$$

Based on the attenuation law of Beer-Lambert, the object luminance $L_{x,y}$ of a pixel *(x,y)* at a distance of $d_{x,y}$ with intrinsic luminance of $L_{0;x,y}$ and L_s, being the luminance of the air light, can be described by the following relation:

$$L_{x,y} = L_{0;x,y} \exp(-\beta_{ext} d_{x,y}) + L_s(1 - \exp(-\beta_{ext} d_{x,y}))$$

The depth $d_{x,y}$ from the observer to the target is used to obtain the right estimation of the transmission map, which makes it an important parameter for an accurate simulation of adverse weather on camera images. The equation requires three main parameters: the MOR value (V), the background luminance (L_s), and the depth of objects in the images ($d_{x,y}$). The depth can be extracted from the stereoscopic camera images. The visibility values depend on the artificial fog parameters from the tests. It was explained in section 2.4.2. Finally, the background luminance is considered as the mean luminance of 10% of the brightest pixels of the image [18].

Figure 33 shows an example of an image acquired in clear weather (a.) with a pedestrian crossing the crosswalk, a depth image from the stereoscopic camera (c.), an image acquired with the same pedestrian characteristics under artificial fog conditions (b.), and an image on which the fog has been numerically simulated (d.).

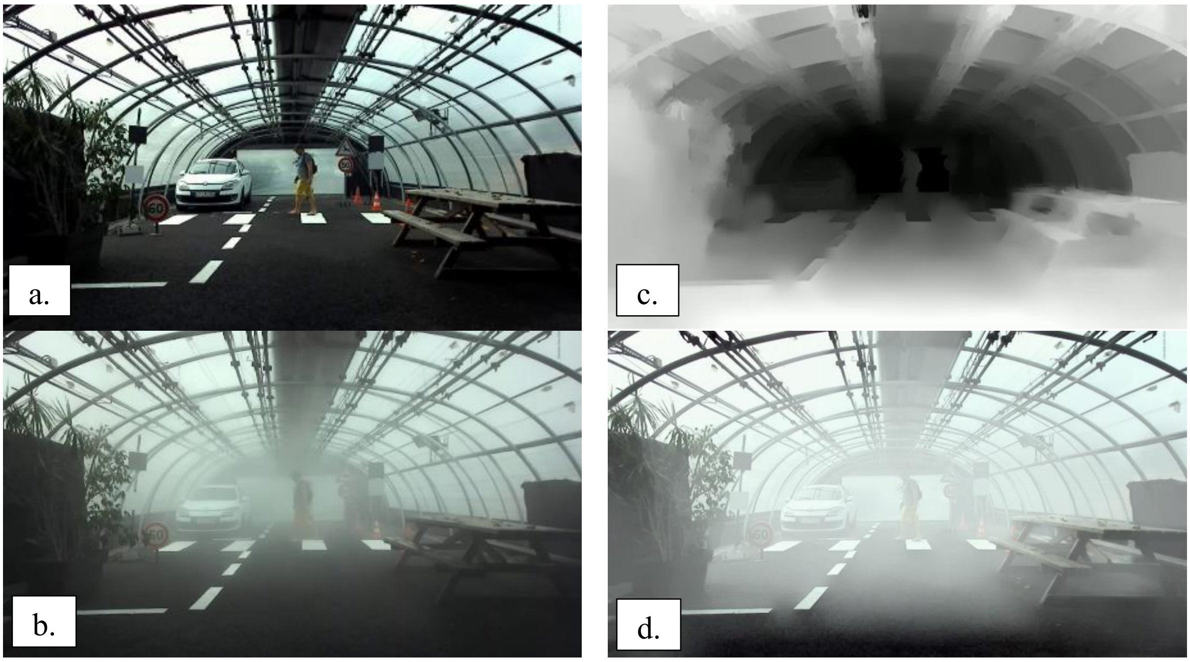


Figure 33: Example of an image with clear weather (a.), real fog (b.), and numerically simulated fog (d.). The depth map is shown as an illustration of the stereoscopic outputs (c.).

3.3. Fully numerical simulation: LEIA 1

The simulation tests are performed by LNE using 4DVirtualiz (4DV), which is a digital twin software devoted to robotics and the automotive field. This simulator allows creating scenarios from scratch using the items included in the library of the simulator or by importing our 3D models of building and vegetable ... etc. In this work, we use the 3D model of the PAVIN produced by Cerema. This model is designed in SketchUp and should be converted to fbx format to be supported by 4DV. Here, Blender, a free tool, is used to achieve this task.

Figure 34: PAVIN 3D model in Blender.

Once the fbx file has been imported into the 4DV simulator (see figure 1), the scenario is configured by specifying the time of day (day or night), the weather conditions (clear or foggy), the humans (their appearance and their route), and the cameras used to retrieve images. Two cameras are used here, the first being ZED2 to retrieve RGB images and a semantic camera to obtain the segmented images which are then used for the annotation step.

Figure 35: PAVIN 3D model in 4DV.

In 4DV simulator, the light of sun is simulated by directional light at infinity and the only parameter to set is the time of day in the 4DV by setting the simulator clock such as 12 pm for daytime and 8 pm for nighttime.
The weather conditions can also be specified in the 4DV simulator; however the visual rendering of the fog is very poor as shown in figure 36. In fact, the fog intensity increases slowly with the distance and fog haze is practically non-existent. The fog intensity in 4DV simulator can be set only with one parameter by choosing two values "weak fog" or "strong fog". The fog model is not described in detail in the simulator documentation but the 4DV simulator company informed us that they use exponential squared fog that takes as an argument the distance from the viewpoint. In the literature, the exponential squared fog model depends on two parameters, which are the distance from the viewpoint and an arbitrary fog density that can range from 0.0 to 1.0. If this last parameter is not set correctly, it can result in poor rendering. To overcome this problem, smoke is used in addition to fog to enhance the visual rendering as shown in figure 37. The smoke is set to zero speed, it does not move and there is no smoke ripple as shown in Figure. In this study, we define two smoke intensities to obtain weak and strong fog. In collaboration with Cerema, the smoke intensity and fog intensity values are set in such a way as to ensure visual acceptability. In particular, the smoke parameters have been set from 0.1 to 0.8. The team then visually compared which smoke levels most closely resembled the image obtained by the ZED camera in the PAVIN platform for a given visibility level. In fine, the settings selected were as described in the following Table.

Fog rendering	Fog intensity in 4DV	Smoke intensity in 4DV
Medium fog	weak fog	0.15
Heavy fog	strong fog	0.30

Table 3: Fog and smoke intensities values

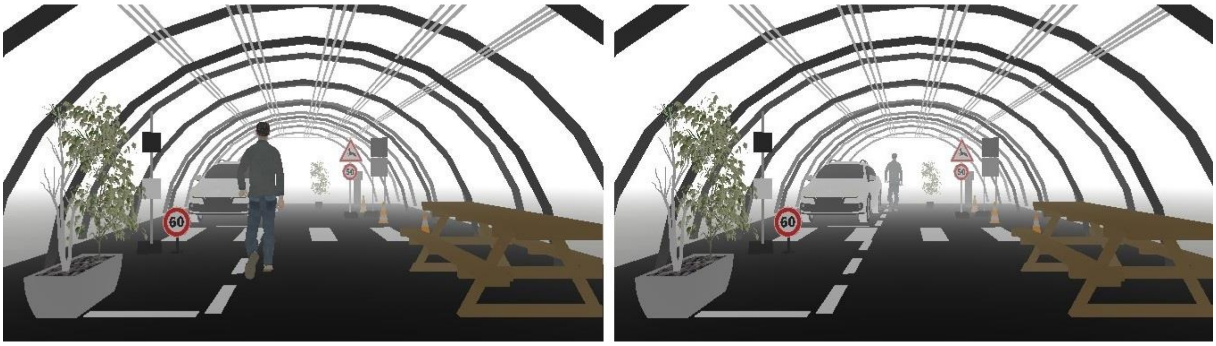


Figure 36: Original fog of 4DV.

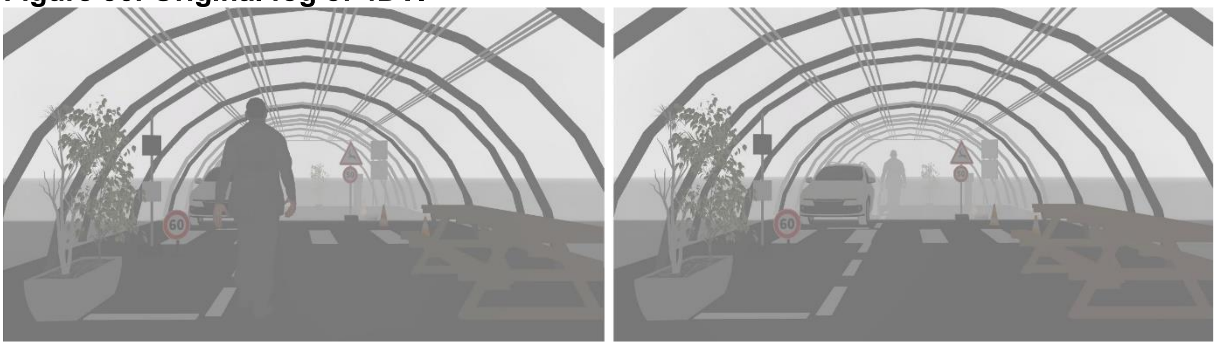


Figure 37: Fog using smoke of 4DV.

4DV simulator includes ZED2i camera model, which is close to that used by Cerema for test campaigns. Some of ZED2 parameters can be set as focus, zoom and frequency. A semantic camera is also used to segment images retrieved by ZED2, facilitating the annotation process. The semantic camera is positioned as the same location as ZED2 and should have the same zoom and frequency configuration as the ZED2.

4DV simulator offers automated test management, which means you specify the variable parameter in the scenario and let 4DV handle them automatically, for example, with 2 parameters of weather conditions (clear and foggy) and 2 parameters for time of day (daytime and nighttime), 4DV will combine all these parameters and execute 4 tests.

Once the tests have been executed and images retrieved, the annotation process should be carried out to have the ground truth related to the position of the human in images. This ground truth will then be used to compare it with the output of the detection algorithm. 4DV includes tool to achieve this task by using the semantic images retrieved by semantic camera to output JSON file containing the bounding box of human detection for each single image.

3.4. HiL replay simulation tool: LEIA 2

To narrow the gap between simulation and physical tests, we propose to use "mixed tests". In this type of test, the physical device is evaluated using synthetic data presented in a physical manner. The LNE is currently developing a platform of mixed test called (LEIA 2). This platform consists of a conical projection screen installed and high-resolution images will be seamlessly projected, surrounding the tested system in a 300° virtual reality. For this project, a single high-performance projector will be used on a flat projection surface in a dark room as shown in figure 38. This is not a problem as we are testing a single camera that does not cover a wide angle.

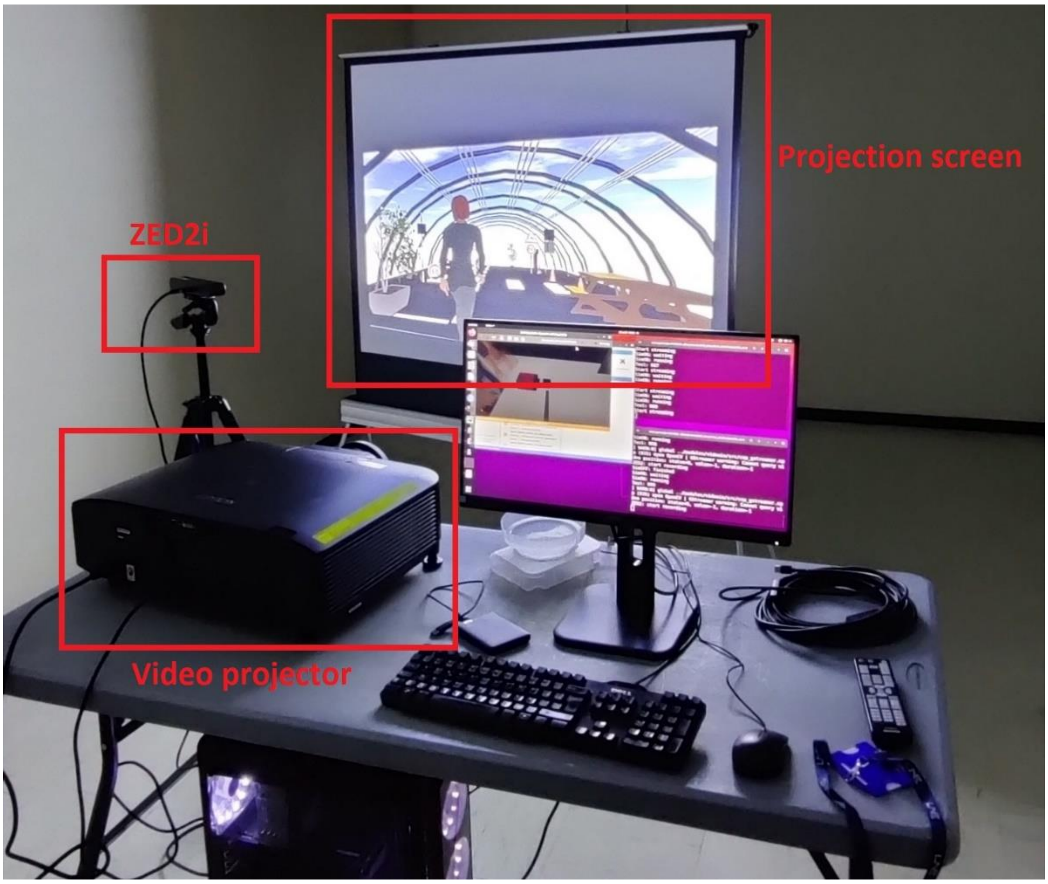


Figure 38: Darkroom of mixed tests

In this study, images retrieved from real simulated fog, numerically simulated fog and fully numerical simulation are projected onto the screen. To guarantee the high quality of recovered images and avoid any light disturbance, mixed tests are carried out in a darkroom. The tests are run automatically and manipulated using python scripts developed during this task. To illustrate, the figure X.X shows a hybrid test image.

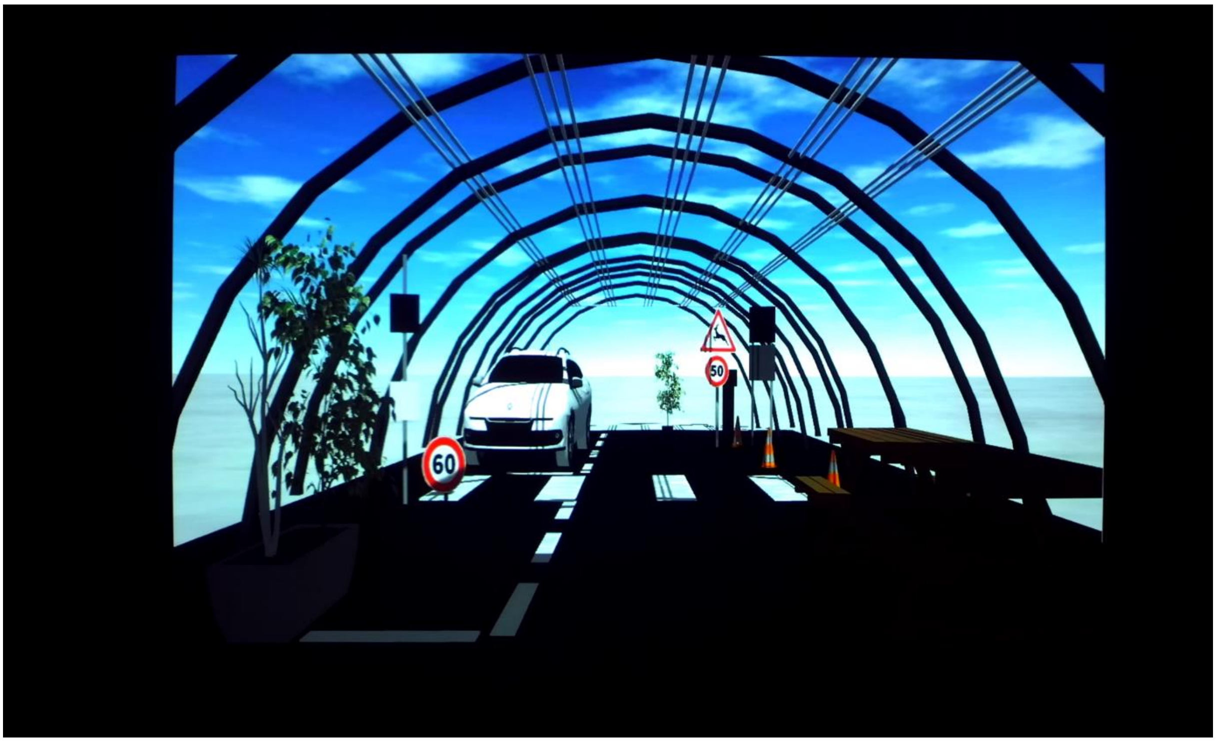


Figure 39: ZED2i image of hybrid test

Once the mixed tests have been completed, the images obtained are annotated on the basis of the previous annotation according to the type of tests. For example, for the fully numerical simulation, the JSON file generated by the 4DV simulator during the simulation tests. To compute the bounding box of mixed test images, the bounding box of the simulation test images are used. As shown in figure 39, the bounding box of the simulation test image has parameters, $(B1, B\_h, B\_w)$, and the idea is to calculate the new parameters $(B1_{ZED}, B\_h_{ZED}, B\_w_{ZED})$ of the bounding box of the ZED image (see figure 40). To do this, we can use the following expression:

$$h_{ratio} = \frac{h_{ZED}}{h_{4DV}}$$

$$w_{ratio} = \frac{w_{ZED}}{w_{4DV}}$$

$$B\_h_{ZED} = B_h * h_{ratio}$$

$$B_{w_{ZED}} = B\_w * w_{ratio}$$

Where $B1_{ZED} = [shift\_w + B1\_w * w_{ratio},\ \ shift\_h + B1\_h * h_{ratio}]$ and $B1 = [B1\_w, B1\_h]$.
As ZED2i camera retrieve the entire scene containing the 4DV image projection, including the black edge $(shift\_w, shift\_h)$, we need to remove this black edge from the ZED image to find the correct ZED image bounding box.

4DV image

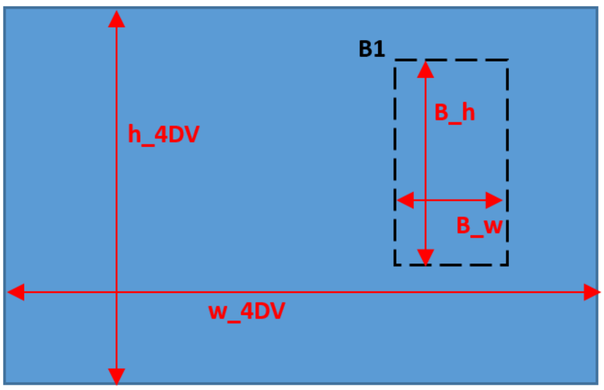


Figure 40: Bounding box on 4DV image

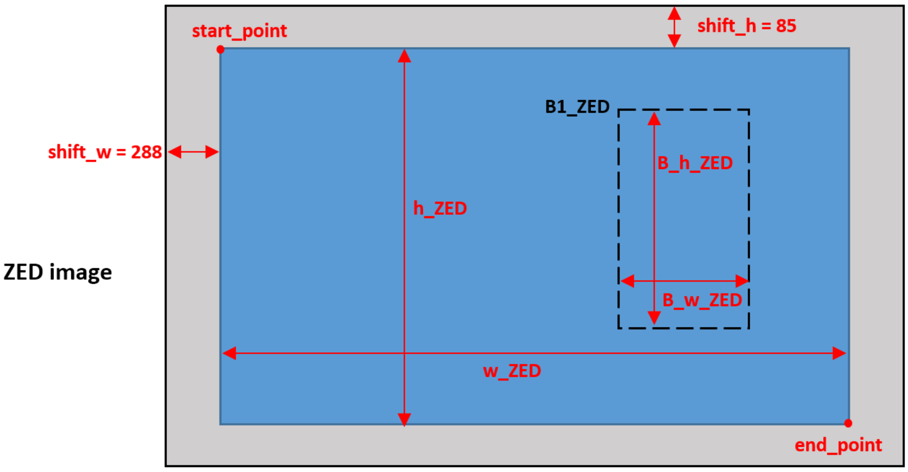


Figure 40: Bounding box on ZED image

4. Roadmap

Table 3 shows the past and future roadmap for carrying out the analyses for the POC presented in this protocol. The work presented in this protocol represents a year and a half's work, with tasks ranging from manipulation to simulation and implementation of an AI-based algorithm. The table is containing two columns, showing what was present in the first POC, and the added value of the second POC in terms of testing. POC 1 mainly enabled the PAVIN and LEIA1 databases to be set up, but labelling was incomplete. POC 2 enabled us to complete the labeling process, enabling us to carry out a full analysis of the database, and to add digitally simulated and replayed tests in LEIA 2 to take account of HiL aspects. The protocol defined in POC 1 has remained unchanged, so that the different simulation solutions can be compared. The results obtained following this protocol on POC 1 and POC 2 will be presented in deliverable D3.6.

Table 4: Roadmap for the POC "STV".

Date	Action	POC1	POC2
September 2022	Drafting the test protocol.	x	
October 2022	Testing and recording of the PAVIN database in the PAVIN Fog and Rain platform.	x	
October 2022 - May 2023	Annotation	x	x
February 2023	Replay of PAVIN base on LEIA 2 -> PAVIN*.	x	
May 2023	Creation of the LEIA database thanks to simulation on LEIA1.	x	
April 2023	Creating the K-HiL database with the K-Hil simulator.		x
September 2023	Replaying the two new databases on LEIA 2 -> LEIA* and K-HiL*.		x

November 2023	Drafting of final analysis protocol (D3.3).		x
December 2023	Application of YOLO and calculation of scores on all databases.		x
January – February 2024	Data and results analysis.	x	x
March - April 2024	Drafting of final deliverable on results (D3.6).		x

Chapter 5: IGN and VALEO POC

1. Introduction

1.1. Document goal

This document describes the test that will take place at the beginning of 2024 on the UTAC site at Montlhéry.

The POC consists of a determination of a Valeo autonomous vehicle trajectography by an independent system designed by IGN.

1.2. Presentation of functions to be assessed

Ground system designed by IGN gives trajectography positions of a vehicle at a frequency of 10 Hz. These positions are associated with precision indicators.

2. Evaluation protocol

2.1. State of the art: possibly relevant standards

EN 16803-1	Space - Use of GNSS-based positioning for road Intelligent Transport Systems (ITS) - Part 1: Definitions and system engineering procedures for the establishment and assessment of performances
ISO/IEC 18305	Information technology – Real time locating systems – Test and evaluation of localization and tracking systems

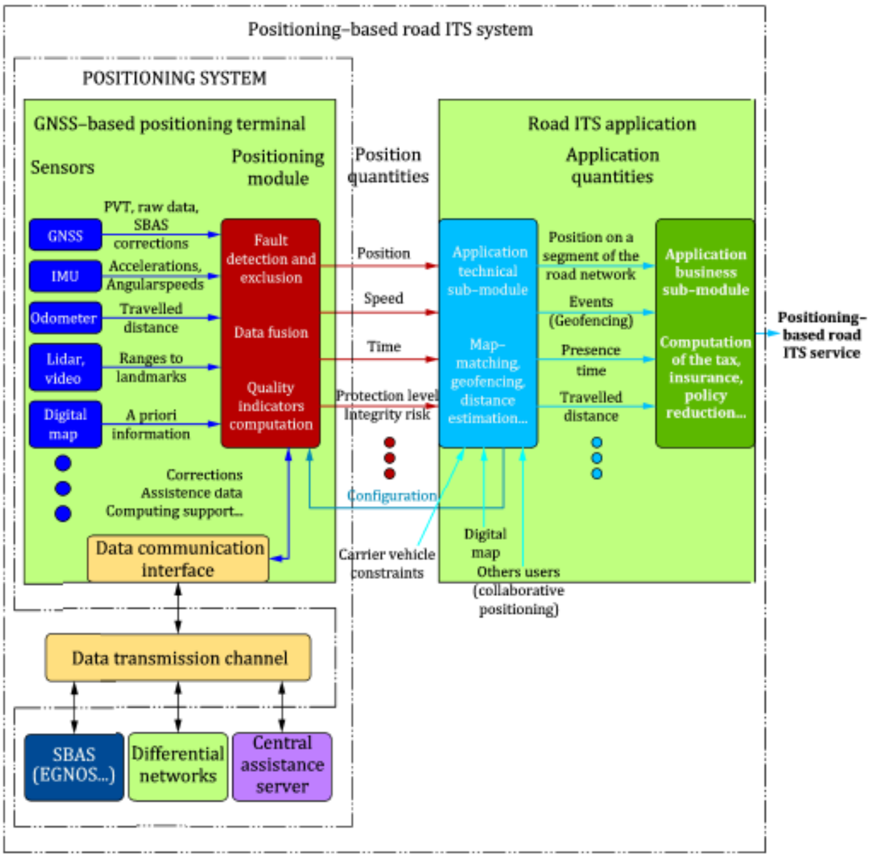


Figure 20 Generic architecture of a road ITS system (EN 16803-1)

2.2. Performance characteristics & metrics and criteria to consider

Performance of a localization system for AV can be assessed according to different characteristics (according EN 16803-1 or the Required Navigation Performance from the International Civil Aviation Organization):

- Accuracy
- Integrity
- Availability
- Continuity
- Timing performance (timestamp resolution, output latency, rate stability or time to first fix).

For this work package/proof-of-concept, a focus will be done on the **accuracy** of the absolute position returned by the localization system.

Such accuracy is defined as the *" closeness of the agreement between the Positioning State estimated by the Positioning System and the truth"* in "ISO 5725-1:1994: Accuracy (trueness and precision) of measurement methods and results; Part 1: General principles and definitions".

The other performance characteristics may be assessed in further steps.

To quantify the accuracy, the EN 16803-1 provides the following inventory of metrics

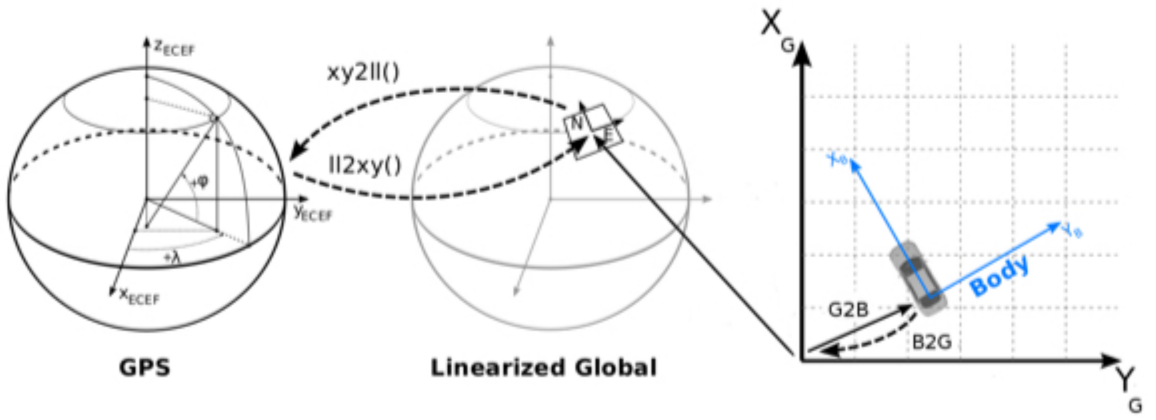
Table 1 — Accuracy metrics summary

Output	Component	Accuracy metric
Position	3D	3D Position Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of 3D position errors.
	Horizontal	Horizontal Position Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of horizontal position errors.
	East	East Position Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of position errors along the East-west direction
	North	North Position Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of position errors along the North-south direction.
	Along track	Along Track Position Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of Along Track position errors.
	Cross track	Cross Position Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of Cross-Track position errors.
	Vertical	Vertical Position Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of Vertical position errors.
Velocity	3D	3D Velocity Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of 3D Velocity errors.
	Horizontal	Horizontal Velocity Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of horizontal Velocity errors.
	East	East Velocity Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of Velocity errors along the East-west direction.
	North	North Velocity Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of Velocity errors along the North-south direction.
	Along track	Along Velocity Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of Along Track Velocity errors.
	Cross track	Cross Velocity Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of Cross-Track Velocity errors.
	Vertical	Vertical Velocity Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of the absolute values of Vertical Velocity errors.
Speed	-	Speed Accuracy is defined as the set of three statistical values given by the 50th. 75th and 95th percentiles of the cumulative distribution of Speed errors.

2.3. Methods of post processing

During the performance assessment test (on a test track), the localization system for AV under test and the ground truth system (infrastructure-based) are required to return both an absolute position in the same spatial reference frame with consistent timestamps.
The following process is proposed to to analyze the accuracy of such a localization system under test.

Step 1: Coordinate transform
Convert the geodetic coordinates (latitude/longitude/altitude) into a local tangent plane coordates, such as ENU "east-north-up" (x/y/z) coordinates for example.



(source : documentation related to "Ford AV dataset")

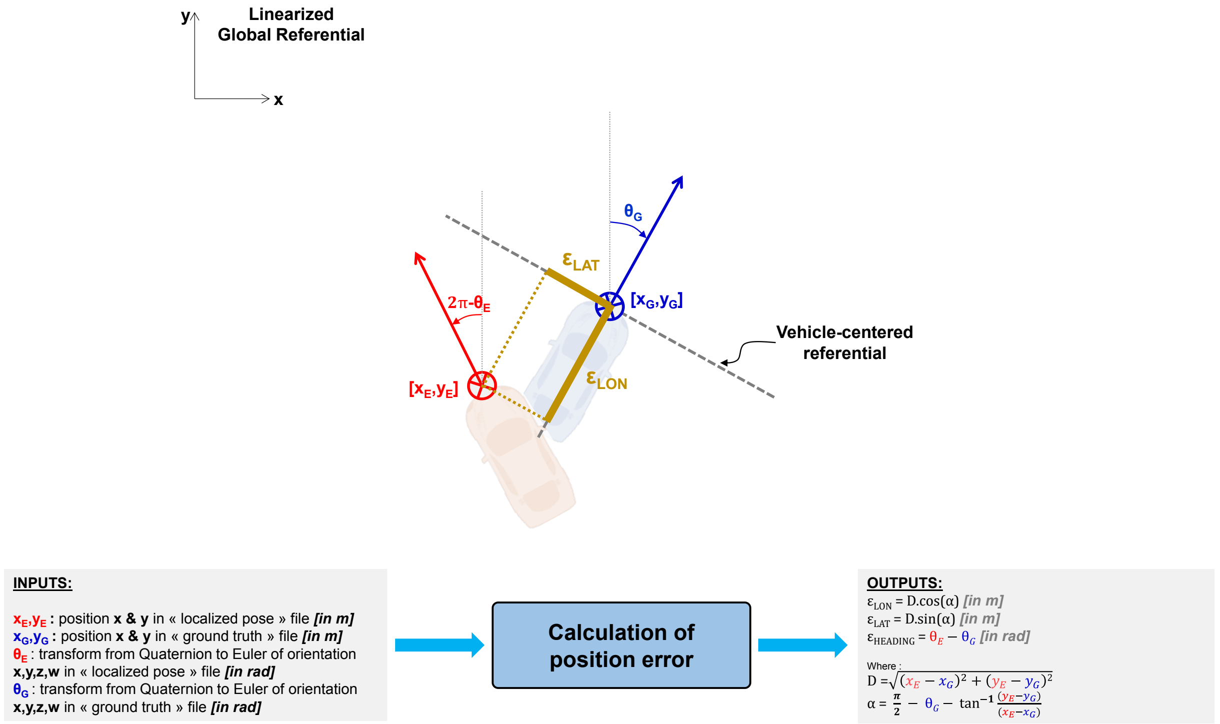
Step 2 : Calculate the positioning & heading error

At each epoch, calculate the lateral and longitudinal error by comparing the "estimated" position (returned by the system under test), and the "true position" returned by the ground truth system.

The vertical error is not considered here.

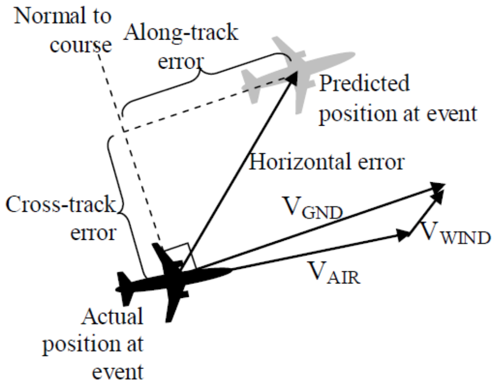
A possible method is illustrated below.

At each epoch, the 2D position error (lateral & longitudinal) is calculated by projecting the "estimated" position on the axis of a (right-handed) reference frame, whose center and orientation correspond to the true vehicle position & heading at that epoch.



This method refers to the "along track" and "cross track" position accuracy metrics that was previously mentioned. Such method, by calculating the error at each epoch in comparison to the true vehicle trajectory, is considered as the most relevant for the automated driving use case (among all accuracy metrics listed before).

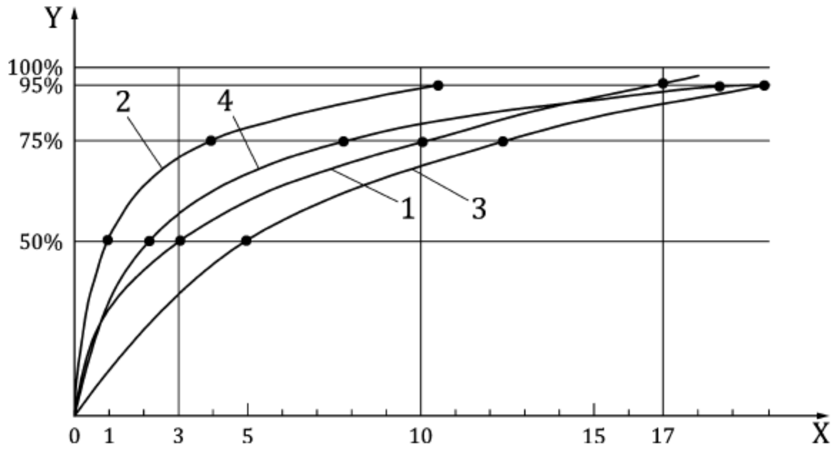
The article from the aviation community *"Assessing Trajectory Prediction Performance – Metrics Definition" (Mondolini et al., 2005)* provides an explicite definition of such "along track" and "cross track" errors, that can be easily translated for the automotive industry.



"<u>Along-track error:</u> Measures the difference in the position of the predicted location of the flight and the actual location of the flight, projected onto the actual course at the time of a specified event."
"<u>Cross-track error:</u> Measures the difference in the position of the predicted location of the flight and the actual location of the flight, projected onto a vector perpendicular to the actual course at a specified event"

Step 3 : Characterize the accuracy of the localization system for AV
For both lateral & longitudinal errors, calculate the 50th, 75th and 95th percentiles of the cumulative distribution of errors.
The EN 16803-1 indicates that the cumulative distribution function of a real-valued random variable X is the function given by: $F_X(x)=P(X\leq x)$.

An example of result is provided below (from the EN 16803-1 again).



Key

X horizontal position error (m)

Y cumulative percentages

1 reference

2 better accuracy

3 worse accuracy

4 unclear

The EN 16803-1 recommends to use such accuracy metrics rather than mean & standard deviation of the error distribution, that may not be relevant if the error distribution does not belong to a "well characterized family of statistical distributions such as Gaussian family of distributions".

In comparison, the UN R144 regulation assesses the accuracy of AECS (Accident Emergency Call System) positioning system by using the horizontal position accuracy metrics, rather than the "along track" & "cross track" position accuracy metrics.
The method to calculate the horizontal position accuracy according to UN R144 is presented below:

$$\Pi = \sqrt{dB^2(m) + dL^2(m)} + 2 \cdot \sqrt{\sigma_B^2(m) + \sigma_L^2(m)} \quad (6)$$

Where:

$$\Delta B(j) = B(j) - B_{truej}, \quad (1)$$

$$dB = \frac{1}{N} \cdot \sum_{j=1}^{N} \Delta B(j) \quad (2)$$

Where:

B_{truej} is the actual value of B coordinate in "j" time moment, in arc seconds;

B(j) is the determined value of B coordinate in "j" time moment, by the GNSS receiver, arc seconds;

N is the amount of GGA (RMC) messages, received during the test of GNSS receiver.

Similarly calculate the systematic inaccuracy of L (longitude) coordinate.

$$\sigma_B = \sqrt{\frac{\sum_{j=1}^{N} (\Delta B(j) - dB)^2}{N-1}}, \quad (3)$$

(same calculation of σ_L standard deviation for L (longitude) coordinate

Step 4 : Validate that the accuracy complies with system requirements
The performance requirement will depend on the Automated Driving System manufacturer/supplier, and may be provided by the latter to validate that the localization system for AV satisfies the target performance requirements.

3. Presentation of environments and test means

3.1. System presentation

3.1.1. Objectives

The UTAC POC measurement system is designed to estimate the trajectory of an autonomous vehicle moving around and next to a given roundabout at various speeds between 15 km/h and 50 km/h.
The used method is independent of the vehicle navigation system and does not modify the vehicle except for some targets stuck to it.

The expected trajectory is a set of positions and orientations of the vehicle body in spatial reference frame with a precise timing. These points should represent the position of the vehicle with a frequency of a few Hz and with a three-dimensional geometric precision of around 2 cm.

3.1.2. Principles

The proposed method is based on photogrammetric tracking of targets on the vehicle by stationnary cameras precisely positioned, oriented and synchronized.

3.1.3. Constraints

To be able to estimate the vehicle position and orientation at each epoch, at least 3 targets must be seen simultaneously at each epoch from at least 2 different positions. More are needed in order to be able to estimate the final precision.

The targets on the acquired pictures must be large enough for automatic detection and identifier decoding, which implies that the physical targets must be relatively large, and the cameras concentrated around one roundabout, to have a maximum of usable pictures.

The system is not designed to be operational in rainy weather. Moreover, visibility conditions have to be sufficient. Thus, the acquisition must take place without thick fog and only during the day.

3.1.4. Output

The vehicle body positions and orientations are estimated in 3D with a precision for each value estimated.

3.2. Sensors

3.2.1. Cameras

The cameras are industrial models, having:

- sufficient resolution (around 16 Mpx)
- sufficient maximum acquisition frequency (around 10 Hz)
- fast Ethernet connection to transfer pictures
- panchromatic sensor for better geometrical accuracy
- possibility to save pictures without lossy compression
- external triggering input
- SDK for PC-camera communication

For the planned POC, the selected model is BFS-PGE-161S7 from FLIR, a camera with a 16 MP resolution sensor.

Fig 40. BFS-PGE-161S7 cameras

One computer is used to configure and save the images of several cameras. Using a dedicated 4 ports gigabit Ethernet PCI card, it is possible to reach 10 Hz acquisition with 3 cameras configured in lossless compressed mode per PC.

A group, composed of one PC, several cameras and one GEOSTIX (c.f. 3.2.2.) is called an "acquisition pole". Poles have no wire connection between each other, to simplify on-field setting across a road

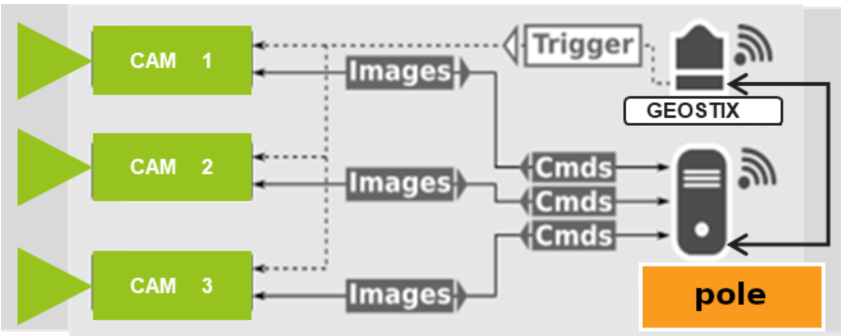


Fig 41. One acquisition pole

The PC software is developed using the camera constructor SDK. The acquisition process is as follows:

- Load settings_json file
- Create the data_path "location_date_hours_minutes"
- Network config check
- Force the IP addressing of the cameras to have the same interface subnet.
- Launch the acquisition process:
  - Create 3 threads to start capturing images (three cameras):
    - For each thread:
      1) display the device info
      2) init the camera
      3) set the packet size
      4) reset camera cycle time
      5) enable Chunkdata
      6) enable the image compression
      7) start camera acquisition:
         - set the acquisition mode
         - configure exposition time, trigger delay in camera settings
         - begin image acquisition
         - waiting for the synchronization pulse
    - if synchronization:
      - capture an image
      - get chunkdata
      - if image is incomplete, reject it.
      - get width, height, timestamp, and if is compressed, get ratio, CRC
      - set file name and save the raw image.

    - Save some image data info in a data structure, for future decompression.

3.2.2. Timing

Timing has two main objectives:

- to have a good synchronization between all cameras (an order of magnitude better than exposure time)
- to have a correct timing reference to be able to compare photogrammetric trajectory with other trajectories.

GNSS timing is used to have a precise and universal time reference. It is quite difficult to synchronize many cameras without wiring them all together. We use GEOSTIX, GNSS sensors designed at IGN and industrialized by Geobsys. Those sensors are versatile via extensions. For camera synchronization, a new layer was developed to send captured signals precisely on sub-multiples of GNSS round seconds. Several GEOSTIX can be programmed to start triggering cameras at the same time to ensure that the pictures indexes of all cameras are consistent. GNSS times of triggering signals will be recorded for precise image timing.

The cameras can be configured to add a delay between the trigger signal and the exposure. This is adjusted in coordination with exposure duration to make sure that the instant at the middle of exposure time is the same for all cameras, regardless of exposure duration. The actual delay had been precisely compared to the requested one, and its precision is better than 50 µs (corresponding to a 0.5 mm displacement of the car at 36 km/h, which is negligible in this experiment).

The GEOSTIX modules also synchronize the computers real time clocks for information.

Fig 42. GEOSTIX GNSS receiver

Image synchronization has been heavily tested to synchronization was validated:

- multiple cameras on one computer with wired sync
- cameras on different computers with GEOSTIX sync
- camera trigger delay

As a result, we selected the best camera data settings and maximal acquisition frequency and validated synchronization precision.

3.2.3. Topometry

Topometry have several objectives:

1. get (stationnary) camera positions and orientations
1. get relative positions of vehicle targets
2. provide vehicle trajectography in a common spatial reference frame

Device, observation figure are planned to meet the needs.

3.2.3.1. Cameras positions and orientations

The first objective is to compute an *a priori* solution for the optical center positions of all 12 stationnary cameras. Cameras must be mounted on specially designed supports, as shown below (Fig 43a. and 43b.), to collocate the optical center with topometric reflectors center, which will make the topometric phase more efficient and accurate for the localization of camera centers (a few mm).

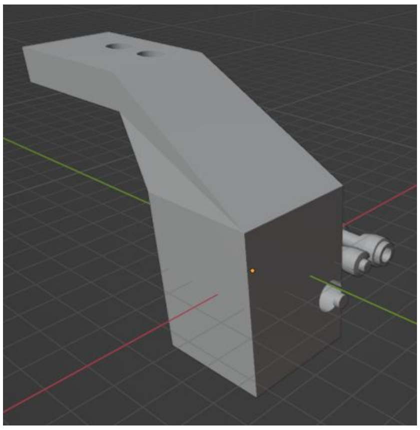


Fig 43a. Camera support (left), 8 mm objective

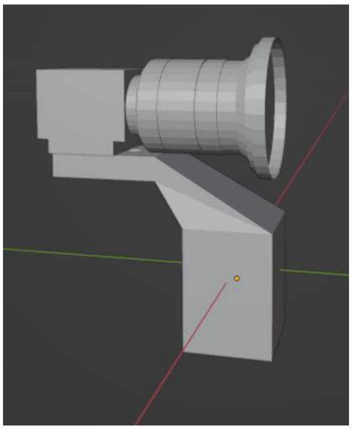


Fig 43b. Camera on its support (center)

Fig 43c. Topometric reflector (right)

The second objective is to compute positions of Ground Control Points (GCP) for photogrammetric camera orientations. It is done by using 2 types of points: sphere and targets (Fig 44.). Natural points (road mark, road signs, etc.) can also be used depending on the test track configuration.

The sphere has the advantage of being visible from all points of view, but it is more difficult to observe with a topographical instrument. An adapter allows replacing a sphere by a topometric reflector during measurements to determine accurately the sphere center.

The target has the advantage of being automatically detected in the camera images. It is also very simple to determine in topometry.

Fig 44. Sphere (left) and black/white target (right)

The observations are angular and distance measurements with a total station (see Fig 49). Several stations are carried out around the points to determine. A least squares global adjustment of all these observations is led to compute coordinates and estimate precision with redundancy.

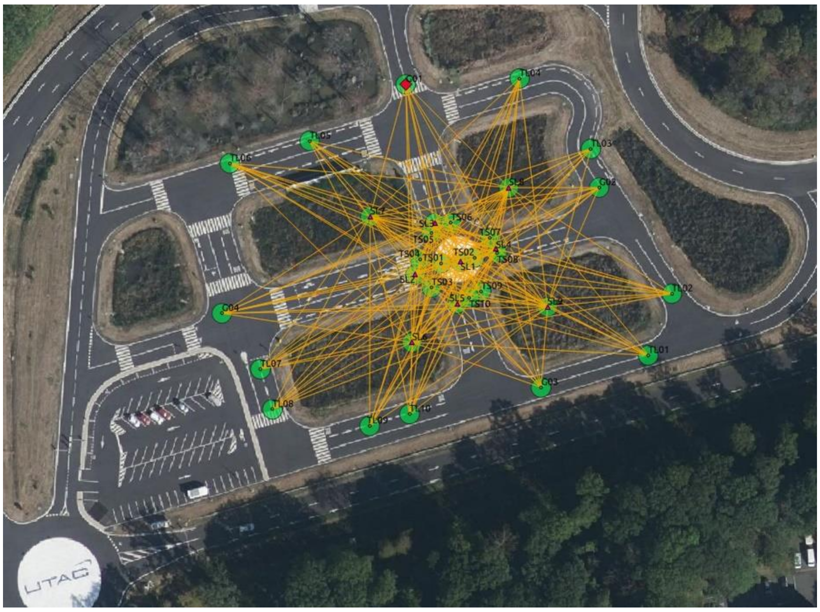


Fig 45. Simulation of topometric observation figure at UTAC

The simulation (Fig 45.) allows us to ascertain that all points (camera positions included) of the UTAC POC will be known with a relative uncertainty between 0.4 mm and 1.0 mm.

3.2.3.2. Relative position of vehicle targets

During topometric observations, targets on the vehicle are measured and integrated to the global computation. A set of scans with static laser scanner (Fig 46a.) could be done and added to the data model to improve redundancy and **reliability** of the computation.

Fig 46a. Laser scanner Leica RTC 360 LT

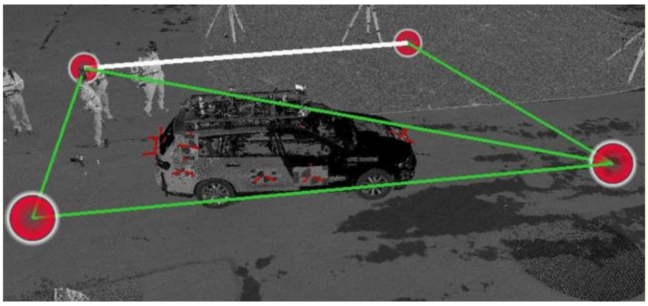


Fig 46b. Visualization of laser scanner point clouds and position for the test in Valeo site

The laser scanner model that will be used for the POC is a Laser Leica RTC 360 LT. The two main characteristics of this device are:

- resolution: 3 mm at 10 m
- precision: 1.9 mm at 10 m.

For the acquisition, laser scanner positions will be as close as possible to the vehicle to be able to make 2 sides visible of the vehicle. Thus, the distance between the laser scanner station and visible targets on the vehicle will be between 1 and 5 m. So, the resolution at the distance will be between 0.3 mm and 1.5 mm.

The precision at a distance between 1 and 5 m will be about 1.9 mm. Indeed, the precision decreases with the distance but there is a constant part and under 10 m, the constant part takes the top.

Pointing precision, mainly due to resolution, and laser scanner precision combines into quadratic sum since they are independent. The final precision of a target measured in a laser scanner sub-frame is between 1.9 mm and 2.4 mm.

The precision is poorer than topometric precision. But it is still useful to use it for 3 reasons:

- it improves reliability
- it is possible to fix target indexing ambiguities at the office
- any point visible on the vehicle can be determined at the office

A vehicle target frame is extracted from the computation. We call this frame “rigid block”.

It is needed for 3 purposes:

- predict target positions for the target's detection second pass (see section 3.3.4.2.)
- bring more redundancy of vehicle positions with pictures
- access to the vehicle reference frame given by the positioning system of the vehicle.

The vehicle frame is a set of target coordinates in a common reference frame. For example, the Table 5. below gives the target coordinates in an arbitrary local frame for the test at Valeo (Creteil).

Point	X	Y	Z	Sig X	Sig Y	Sig Z
XS02	0,2222	0,5486	0,4880	0,0002	0,0002	0,0004
XS03	0,3739	0,5473	0,4861	0,0002	0,0002	0,0004
M05	0,7274	0,5084	0,4868	0,0002	0,0002	0,0004
XS00	1,1459	0,0474	0,7415	0,0002	0,0002	0,0004
XS01	1,0875	-0,0273	0,8606	0,0002	0,0002	0,0004
XL11	1,2441	-1,3519	0,2980	0,0002	0,0003	0,0004
M04	1,1799	-2,1601	0,2995	0,0002	0,0003	0,0004
L08	1,1236	-2,7217	0,2428	0,0002	0,0003	0,0004
M06	1,2825	-0,0383	0,4210	0,0003	0,0003	0,0004
M07	-0,3825	0,0385	0,4590	0,0002	0,0004	0,0004
L09	-0,2162	-0,0738	0,8367	0,0002	0,0004	0,0004
L07	-0,5699	-1,2769	0,3465	0,0002	0,0004	0,0004
XL10	-0,6473	-2,0788	0,3013	0,0002	0,0004	0,0004
L06	0,1440	-3,9106	0,5146	0,0006	0,0008	0,0004

Table 5. Example of rigid block coordinates

This rigid block is given with reliable accuracy indicators thanks to a very good knowledge of topometric observation content.

For the test on the Valeo site at Créteil, coordinates of targets on vehicle were determined with 0.2 mm of uncertainty. This very good knowledge of targets' relative positions will improve the quality of the final trajectography by combining with images observations.

3.2.3.3. Absolute reference

Absolute positioning of the targets on the test track is performed with 4 GNSS receivers associated to geodetic antennas. All antennas must be calibrated to know its exact phase center position depending on the satellite position.

Positions are computed with (network) post-processed static solutions, with permanent GNSS network (RGP) base stations located at less than 30 km. This provides an absolute precision of about 5 mm (resp. 10 mm) in planimetric (resp height) coordinates.

The post processing is led with the scientific software Bernese to ensure mastering of precision.

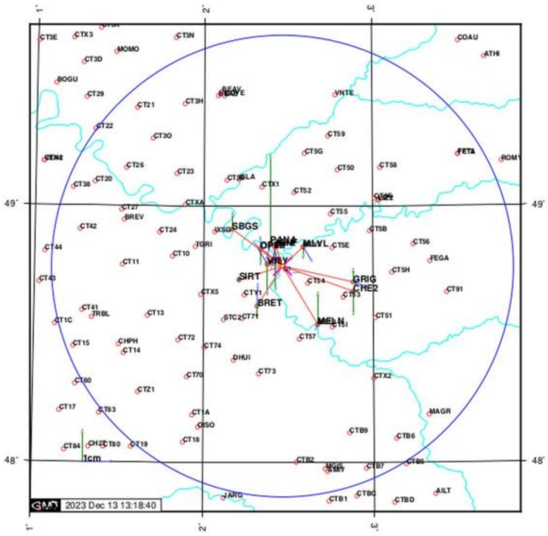


Fig 47. Illustration of baseline computation in Bernese for a GNSS point of the test in Valeo site

GNSS very short baselines are then computed with Leica Infinity software. They are exported with compete covariance variance matrix.

Then, the coordinates and GNSS baseline observations are integrated to the topographic computation to perform a simultaneous block adjustment with IGN Comp3D software. The optimum link between topometric and GNSS observations is done thanks to a simultaneous rigid observation device.

Fig 48. Rigid observation device

Thus, GNSS coordinates of one GNSS point are integrated with constraints, with precision of GNSS computation, in topometric computation. These constraints determine the position of the figure. To make sure of this position, the 4 GNSS points will be constrained alternately and results will be compared.

GNSS baselines as for them determine scale factor and orientation of the frame. The scale factor given by GNSS baselines is less precise than scale factor given by total station distancemeter. So, reference frame scale factor is given by the latter.

Orientation in RGF93 reference frame given by GNSS baselines has a precision of about 0.0010 gon according to simulation results.

To improve reliability of this orientation and test new methods, we will use a gyrotheodolite (Fig 49). Our model is a Gyromat 3000 coupled with a Leica total station TS16.

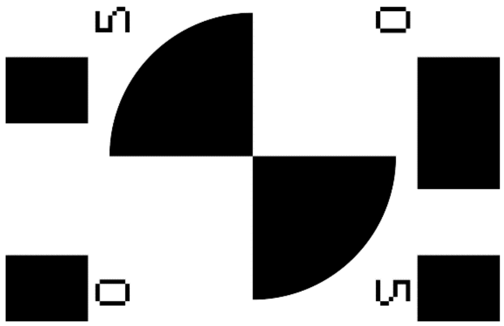
Fig 49. Gyromat 3000 coupled with a total station Leica TS16

Azimuths given by gyrotheodolite has a precision of about 0.0015 gon, *i.e.* less precise than GNSS baselines. But, this orientation is fully independent of GNSS baselines orientation, so it improves significantly orientation reliablity.

Finally, according to simulations, absolute precision of points in RGF93 reference frame will be about 1 cm for the planned POC at UTAC.

3.2.4. Car equipment

The car is equipped with a set of two-dimensional targets (Fig 50).



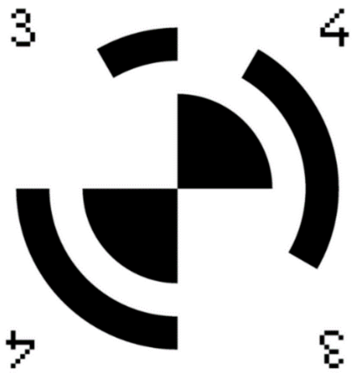


Fig 50. The candidate designs selected for the coded target

3.3. Measurement system and results

3.3.1. Calibration

Calibration of cameras is done in two steps, with MicMac software, developed at IGN.

3.3.1.1. Estimate the distortion of the optical lens

Calibration is done separately for each pair of sensor/objective, on a textured calibration polygon.

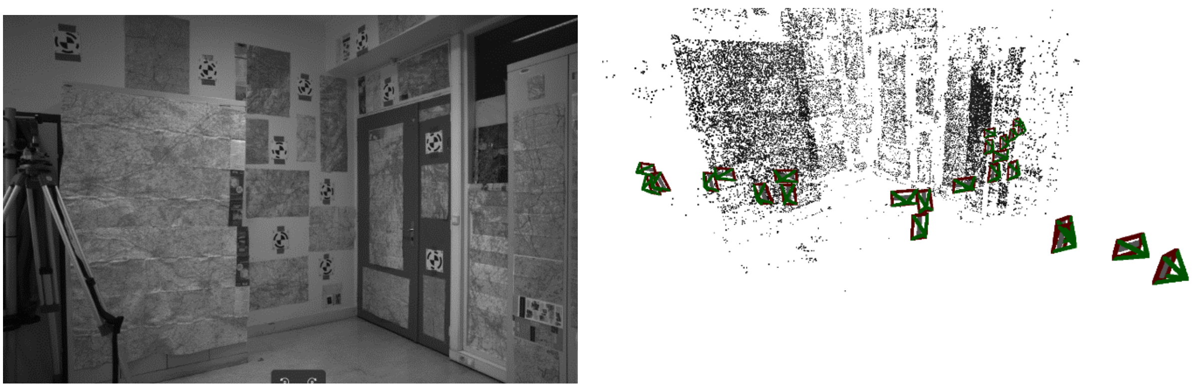


Fig 51. Textured area for calibration (left) an bundle adjustment result (right)

Calibration is performed by taking a set of images with various orientations (to ensure that the calibration is homogenous on the whole image, as illustrated on Fig 52).

Typical residual value of calibration is at 0.70 px, which corresponds to a 0.013° angle with the 8 mm objective (1 cm at 50 m) and 0.008° with the 12 mm objective (7 mm at 50 m).

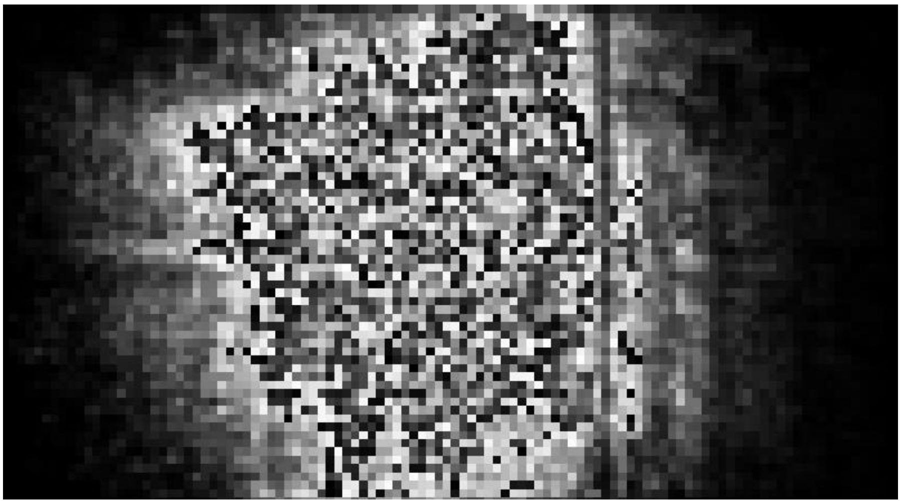


Fig 52. Distribution of tie points in the calibration process. Brighter areas represent higher densities of tie points, therefore better estimation of lens distortions

Tests have been performed to assess the duration of validity of a calibration.

As illustrated below on Fig 53, calibration seems to be holding for at least a few days, with only a few 1/100th of pixels between on-the-fly calibration and using a pre-calibrated camera.

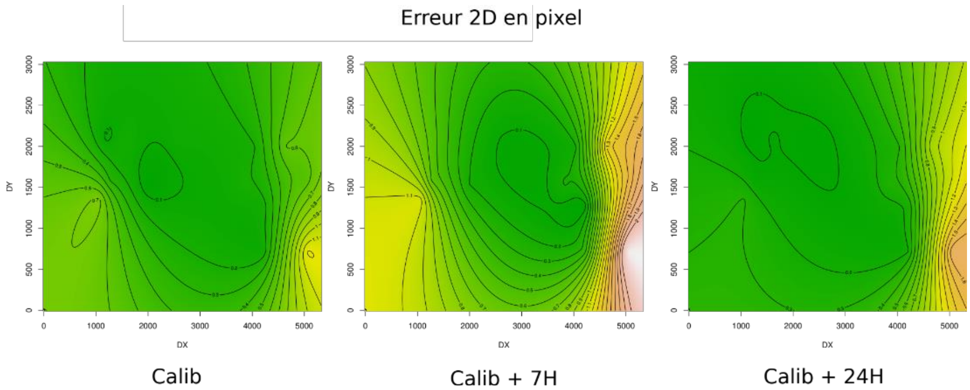


Fig 53. Distorsion errors just after calibration (left), and a few hours later on the same day (right)

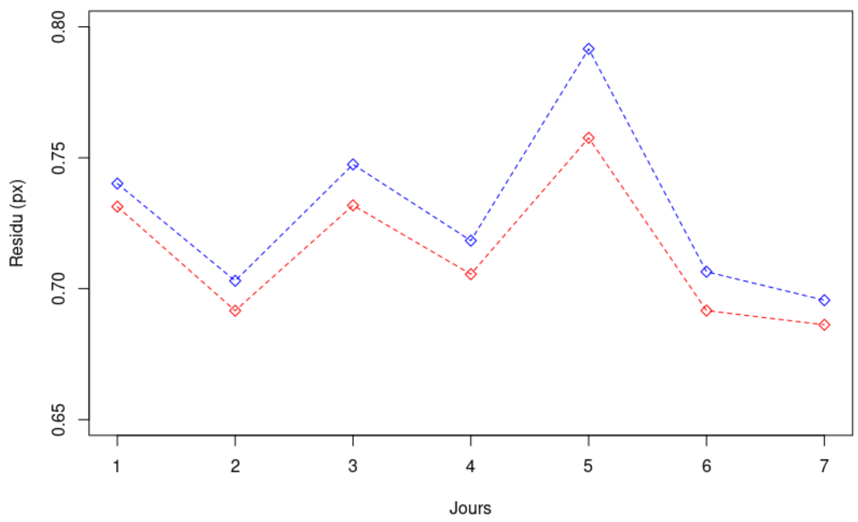


Fig 54. Number of days after calibration vs residual value of bundle adjustment on the same polygon (in px), revealing that calibration of cameras can easily be done up to one week before the experimentation

All these tests enable to conclude that with the chosen model of camera, calibration need not be done on-site just before the experiment POC. In our case, it is planned to be done with one week before the experiment, which saves a lot of time for the POC set up.

3.3.1.2. Estimate the position of optical center

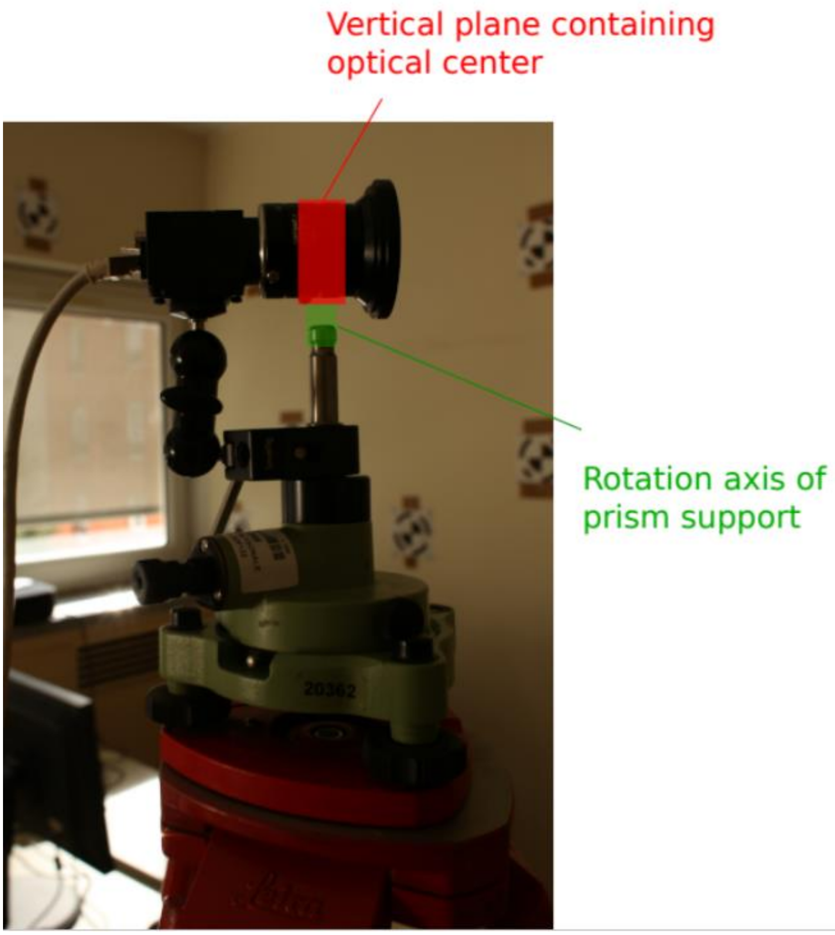


Fig 55. Experimental set up to estimate optical center depth

This is done also for each pair of camera sensor/lens. Optical axes are assumed to be coincidental with the cylindrical symmetry axis of cameras, which leaves only one parameter to be estimated: the depth of optical center position.

Using a calibration polygon composed of 35 coded targets (topometric accuracy ~ 0.1 to 0.2 mm on each axis), disposed on the 4 walls of a square room: camera is placed in the center, and is being rotated, taking picture in stop-and-go. For high accuracy, 100 pictures are captured but experimental comparisons showed that the solution almost converges with 8 pictures (one in each 45° octant).

Automatic target detection is performed on each frame, and orientation of the camera is performed with photogrammetric spatial resection for each captured image. Of course, this is done with pre-calibrated camera (see section 3.3.1.3 above).

If the optical center is co-located with the device rotation axis, all spatially resected centers coincide in a unique point (up to few 10ths of mm of error). This is of course not the case at first attempt with a new camera. In this case, a circle is fitted on the estimated centers (Fig 56), and the radius is an estimate of the distance between the optical center and the rotation axis of the device. A second loop of image capture enables confirming the position of the optical center.

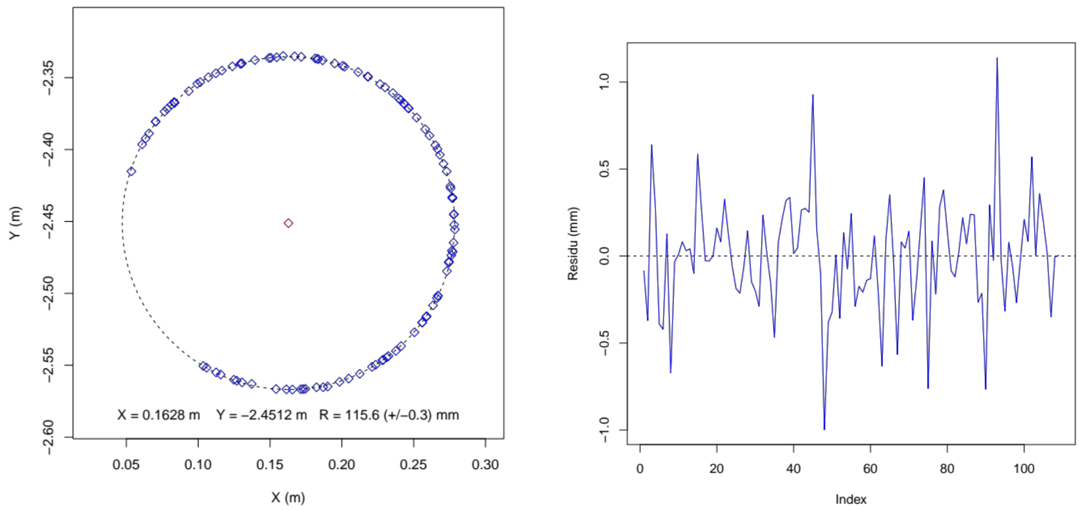


Fig 56. Fitted circle on estimated optical center positions (rotation center is depicted in red) and residual values

With this method, an accurate position of the optical center (at most 1 mm) is estimated in about 5 minutes of experimentation.

Different experiments revealed that (with the degree of accuracy needed) the position of the optical center does not significantly depend on the sensor, the aperture (see Fig 57 as an example) and the focus.

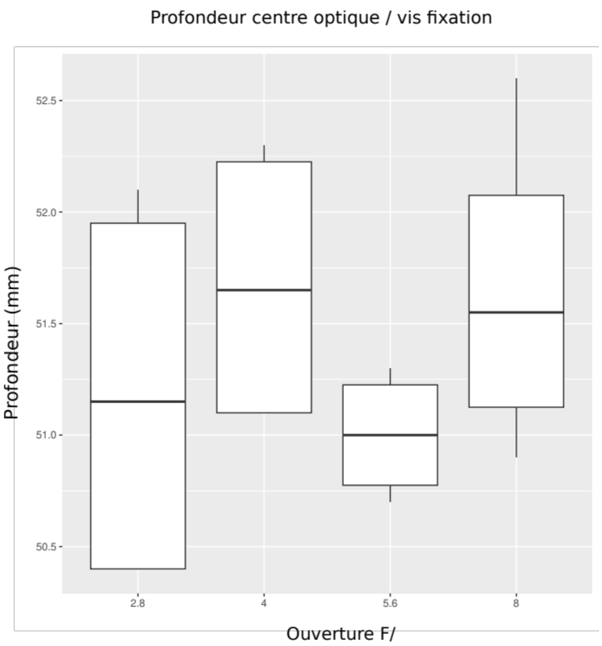


Fig 57. Depth of optical centers (on the optical axis, referenced to an arbitrary point) for different apertures

It was tested also that optical center position does not depend on camera models either (Fig 58).

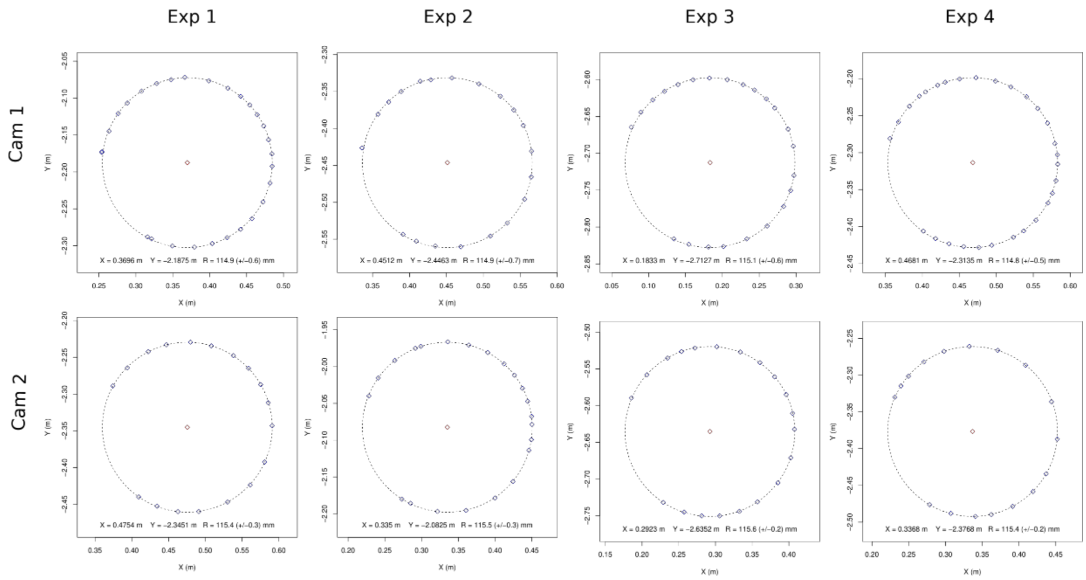


Fig 58. Optical center positions for 2 different cameras and 4 different experiments, revealing that determination is accurate, reproducible, and does not depend on individual camera models.

3.3.2. Positioning each element

To choose positions of all items (cameras, targets on vehicle, black & white target and spheres), different parameters must be taken into account:

- camera field of view.
- image resolution of the vehicle depending on the distance between vehicle and camera.

To design the observation topometric figure, estimate precision of different points, estimate observation time and topometric devices required, simulation must be done before the observation.

3.3.3. Acquisition

The main steps of the acquisition (after topometry) are:

- set the correct exposure duration for each camera
- set the starting time for GEOSTIX triggering signals
- have the vehicle follow its trajectory
- **stop the acquisition**

The data can then be pre-processed:

- decompress the pictures
- **register acquisition time from GEOSTIX into pictures**

3.3.4. Computation

3.3.4.1. Topometry

Topometry computation is done using Comp3D software, an IGN least squares adjustment software.

All observations described in this paragraph about topometry are included to this computation. Our data are point coordinates and associated precision in the French legal reference frame, RGF93.

3.3.4.2. Target detection

With a few images per second, and multiple cameras aiming at the vehicle, the number of targets to pinpoint on the images may be considerable (up to a thousand per second). Besides, for precise localization of the vehicle on each frame, it is important that the target pointing on images is done with a sub-pixel accuracy. This constraint does not allow for a manual detection of targets in any reasonable amount of time.

For this reason, target design has been conceived in order to enable automatic detection and decoding of targets on a vehicle. This detection must be fast (at most a few seconds per image) and accurate enough to guarantee that bundle from the camera to each target is precise enough to reach a 2 cm accuracy at 50 m range.

3.3.4.2.1. Algorithm design

The main steps of the detection target algorithm are as follow:

1) **Preliminary filter:** a set of three filters are applied to identify candidate pixel for target centers. These filters are designed to have high response on areas sharing similar properties with the butterfly pattern (binary, 180° symmetry and radial symmetry). They are fast and applied in a cascading strategy, enabling to eliminate the maximal number of false candidates on the first pass. At the end of this step, target center is usually known with at most 1-pixel error.

2) **Butterfly edge detection:** on each candidate selected at step 1, a radial search and circular mean is performed to extract directions of the cross pattern of the butterfly. Note that, as distance to the target is not known beforehand, the expected size of target in the image is still unknown. Therefore, cross pattern directions are searched on a conservatively small area, and then may not be very accurate (a few degrees off in not uncommon). For this reason, step 2 will be executed one more time when target size in image is accurately determined.

3) **Circular edge detection:** cross pattern detected at step 2 is used to sample a set of radial semi-lines starting from the initial guess of target center position. Black to white transitions are searched on each of this set of lines to extract the two circular edges of the butterfly (typically sampled with at least 10 to 20 pixels per edge).

4) **Ellipse fitting:** the projection of the circular edge of butterfly being an ellipse in the camera projective space, least squares ellipse fitting is used to estimate the full butterfly circular edge. Two strategies have been tested, depending on whether the ellipse is constrained on the initial guess of center. Experimentation has shown that constraining the center does not improve significantly the ellipse geometric quality. Therefore, it was decided to leave the center unconstrained, which in turn pays off in robustness since it saves one degree of freedom, which enables an unbiased comparison of the estimated ellipse center with its initial guess. The difference between initial guess center and ellipse center can be used as a proxy for target detection standard deviation (in pixels). Typical values of differences are below 0.5 pixels.

5) **Image rectification:** knowing the elliptical projection of the butterfly circular edge is not enough to estimate the rectified images. At least 3 points need to be determined. This is done by computing the 4 intersections of cross pattern directions estimated at step 2 with the fitted ellipse. These 4 points are then used as input in the least squares adjustment of the affinity transformation between 3D space and image space. Note that for close target (with the 8 mm objectives, that means closer than 5 m), perspective deformation cannot be neglected, and affinity may not be sufficient, and should theoretically be replaced by an homography. While this option is still open, preliminary tests showed that on real test track, most (if not all) of detected target will be further than this 5 m critical limit, and in this case, 3-point affinity estimation seems more efficient as it leaves again 1 degree of freedom for robustness. Examples of rectified images are provided below (Fig 19 and 20).

6) Target decoding: target bit positions are estimated by extrapolation (provided that target is flat enough). Tolerance may be added to make up for (to some extent) the curvature of the vehicle. Experimental tests showed that up to approximately 1 m curvature radius is acceptable for correct decoding of the largest size (XL – 60 x 40 cm). Decoding is also made more robust with the help of an error detection code. Note that 10 bits are available for 15 targets. Since target codes should be invariant to 180° rotation, in fact only an effective number of 5 bits are available for the code, leaving 2^5 = 32 configurations. With 15 targets, this provides another security control to avoid false positive detections.

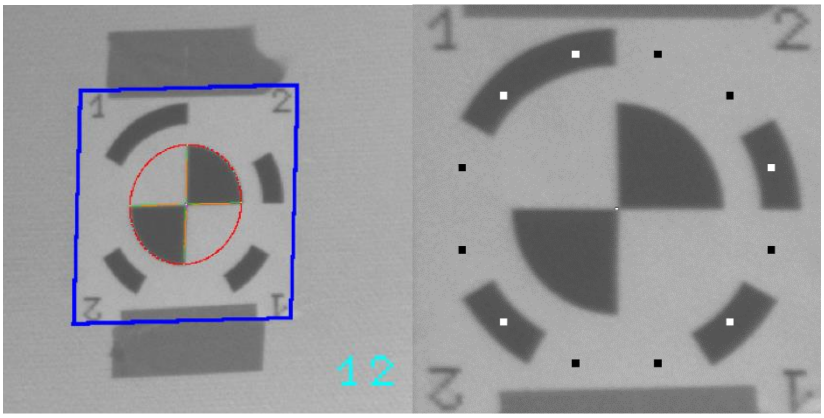


Fig 59. Example of rectification process for a circular pattern target. Bit decoding is searched on squared areas (black for bit '0' – white background and white for bit '1' – black background)

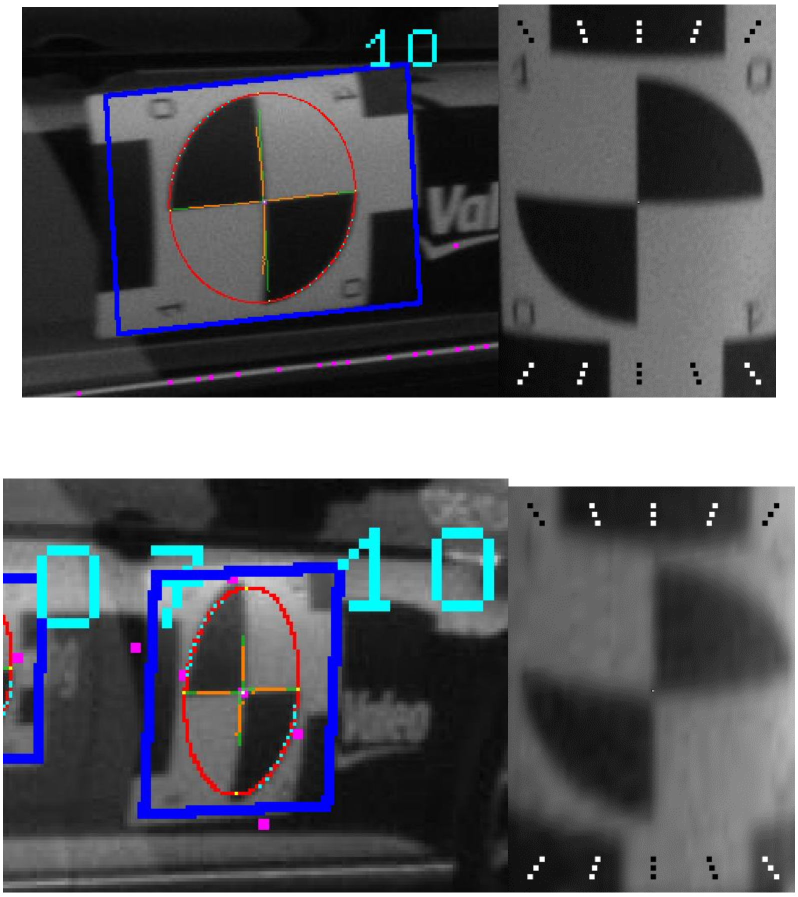


Fig 60. Example of rectification process for a rectangular pattern target on a vehicle. Again, bit decoding is searched on squared areas (black for bit '0' – white background and white for bit '1' – black background). Tolerance is depicted by 3 searching positions for each bit (with 5% spacing) to offset partially the vehicle body curvature. The algorithm is design to perform detection under up to 60° angle in line of sight

7) **Output:** for each camera, each frame and each detected target, position (in pixel floating point values) and target code name are registered in an output file. An estimate of the standard deviation of center location is also provided in the file.

3.3.4.2.2. Tests and results

To date, detection capabilities have been tested:

- by **simulations**, with a detection ratio of 80 % for targets with pixel size in the image over 15 px (size of the butterfly pattern) and with moderate incidence angle (below 45°) between target plane normal vector and line of sight from camera. This detection rate also includes the correct decoding of target. For the largest target model (XL size: 40 x 60 cm), the butterfly is 32 cm wide: for cameras equipped with 8 mm lens (0.02° angle per pixel), this corresponds to 20 px. Hence this target should be detected and correctly decoded with a success rate of 80 % up to a theoretical maximum distance of **60 m** for the 8 mm lens and a **95 m** for the 12 mm lens. All the other target detection performance can be assessed proportionally given the size of their butterfly pattern:

Center	**Dimensions**	**Number**	**Distance***
XL	40 x 60 cm	5	60
L	30 x 45 cm	4	40
M	22 x 33 cm	4	35
S	22 x 22 cm	4	35
XS	18 x 24 cm	4	25

*Maximal theoretical distance to get a 80 % rate of detection and correct decoding with the 8 mm camera lens

Table 6. Model, sizes, number and effective distance of detection/decoding of targets

These distance values are to be considered in relation with the typical ranges on UTAC test track: 10 to 80 m range between cameras and vehicle.

Based on simulation results, the accuracy of target center detection is constant for all targets above 20 pixels. The root mean square error between estimated and true center location is around **0.05 px**. With 8 mm lens camera, this corresponds to an angle of 0.001° (1 mm at 50 m range).

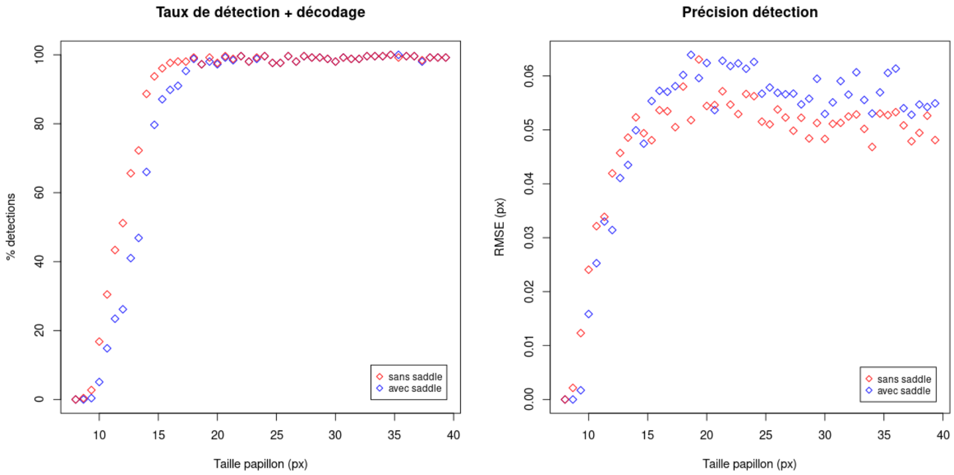


Fig 61. Detection + decoding rate (left) and center localization accuracy (right)

The detection algorithm has been optimized and validated based on simulations (Fig 62). Target are placed with random positions, sizes and orientations on a typical image, and noise is also introduced to simulate a real camera sensor. This method enables to get a real ground truth for target centers (in pixel coordinates), and to optimize the algorithms on many configurations.

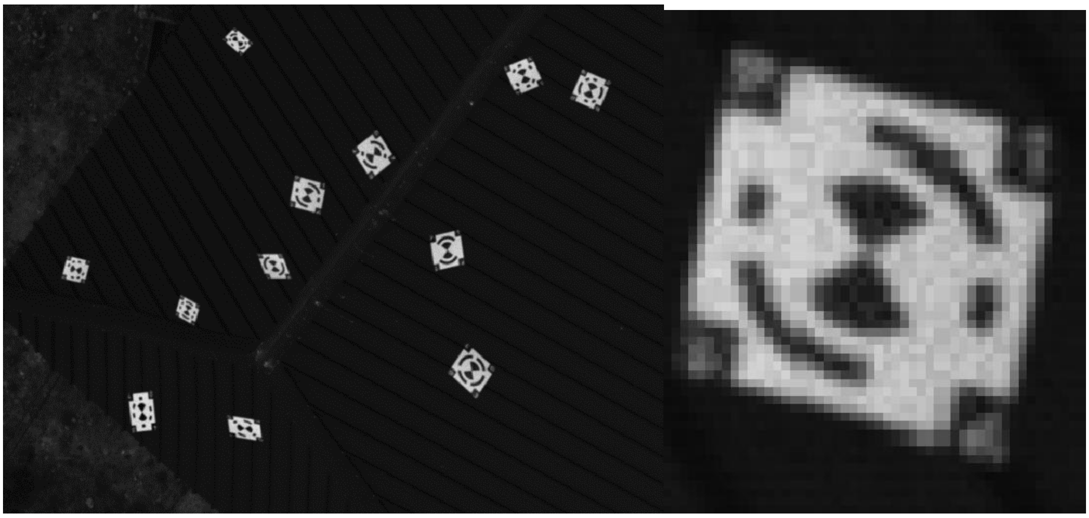


Fig 62. Left: simulation of (circular pattern) targets on an aerial image for target detection and recognition algorithm development, tuning and optimization. Right: simulation of sensor noise.

The relevance of simulation has been confirmed with real experimentation showing approximately similar performances.

- with **real experimentations,** decoding of XL (40 x 60 cm) has been tested with a ratio of 90% within at least a **50 m range**, for moderate angle of incidence (<45°) of targets respectively to the line of sight of the camera. Further test with ground control points showed that the target center localization accuracy is below **0.15 px** (compared to 0.37 px with manual detection). Using automatic detection of targets provided about 5 times more accurate 3D photogrammetric intersections (0.6 mm vs 3 mm error) on an experimentation conducted with ground control points calibration polygon.

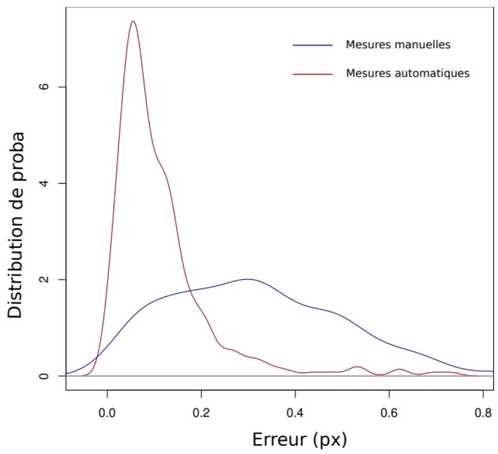


Fig 63. Comparison of distributions of manual (blue) and automatic (red) target detection errors on center localization (in pixels)

Simulations on Valeo test track showed that on two camera objectives (8 and 12 mm), on average 81 % of visible targets are detected with 1.5% of false positive detections. All visible targets are detected on about 55% of images.

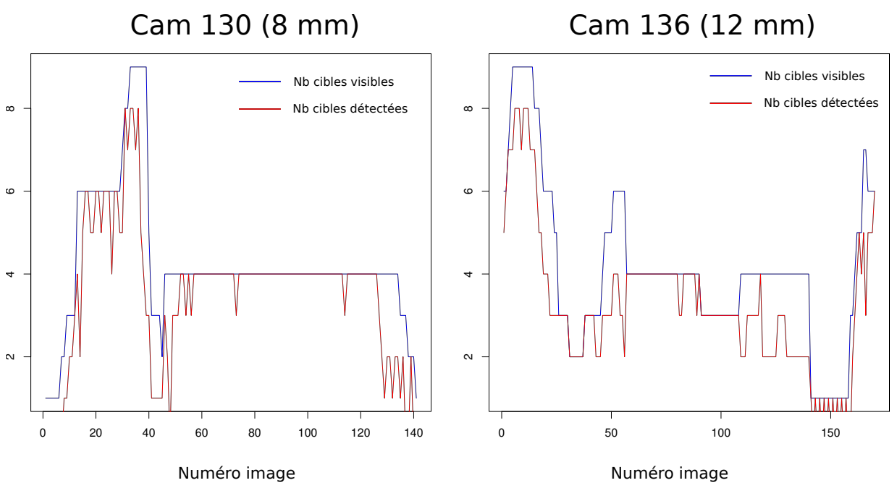


Fig 64. Detection rates for the two models of objectives: number of visible targets in blue; number of detected and correctly decoded targets in red

This test has been conducted at 5.30 PM in winter (end of November), in challenging night conditions and with long exposition time (5000 us). Better performances are expected in nominal conditions.

Camera-vehicle distance and target size seem to be fairly secondary criteria in detection (at least, up to around 50 m range). Curvature of the vehicle body and the angle of view (difficult beyond 60°) are the most important parameters affecting detection/decoding success rate.

3.3.4.2.3. Detection strategy

Detection will be executed iteratively, in a two-pass strategy. On the first pass, detection is performed with conservative parameters, to exclude as much as possible, the risk of false positive detection (requiring human intervention). With this mode, it is assumed that the strict minimum number of targets detected on each camera frame would be at least two on each side at more than 50 m and one on the back at a minimum of 35 m. Having 12 cameras disposed 2 sets of 2 common targets should be visible simultaneously on 2 times 2 pairs of cameras. This enables (with redundancy) to estimate an initial approximate position of 4 targets, each of them being visible on 2 to 4 cameras. The 4 estimated targets can be used to compute a solution for the whole rigid block of 15 targets. Knowing very accurate positions and orientations of each target at each time step, non-detected target positions can be predicted on each frame. This will be used as a prior solution for a second pass of the target detection process. Two strategies are being considered:

(1) Detection pipeline is applied only on the predicted target positions. Target found on these locations are decoded and if the retrieved code is matching (with a tolerance of 1 or 2 in terms of Hamming distance), the newly detected target is added to the preliminary set detected at first pass.

(2) The full detection pipeline is applied on the whole image with much less conservative parameters (low tolerances on fitted curves, center coincidence, affinity residual values…). Any target found close within the estimated tolerance of a predicted target is decoded and compared to the expected code. If the codes are matching (again with a tolerance of 1 or 2 in terms of Hamming distance) the newly detected target is added to the preliminary set detected at first pass.

Since it enables to crop the images and apply target detection only on candidate locations, the former is more efficient in terms of computation time. The latter is faster to develop and integrate in the full pipeline, as it need not modify the detection algorithm.

In both the above strategies, targets detected at second pass can be underweighted, and integrated iteratively in the global least squares adjustment to avoid false observations.

In any case, the system does not require that all targets be detected at every time step. A minimum of three targets is necessary; all other targets are used for increase in precision and robustness. Based on preliminary tests conducted at Valeo test track, it is assumed that 60 to

80% of visible targets will be detected at first pass. 80 to 90% of visible targets may be detected at the end of second pass.

3.3.4.3. Trajectory computation

3.3.4.3.1. Camera orientations

Experimentation (at lab and on Valeo test track) showed that camera heat is causing a slight and progressive deformation in the camera support (about 10 times the order of magnitude of the camera accuracy). About 20 minutes are required before stabilization. A solution would be to burn about 30 minutes of data before having the vehicle run on the track. However, this duration is dependent on conditions and image frequency.

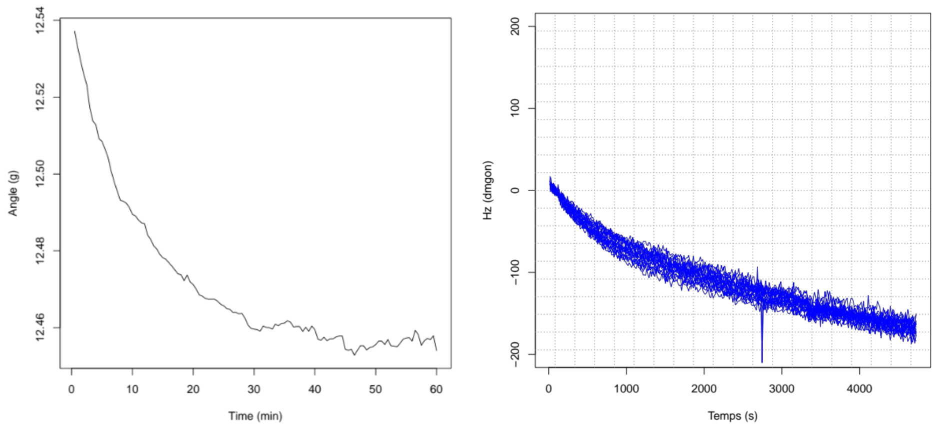


Fig 65. Angular variations of camera line of sight with heating

Besides, other elements may contribute to challenge the stability of cameras (wind, tensions on cables, tripod stability, etc.). See Fig 66 as an example.

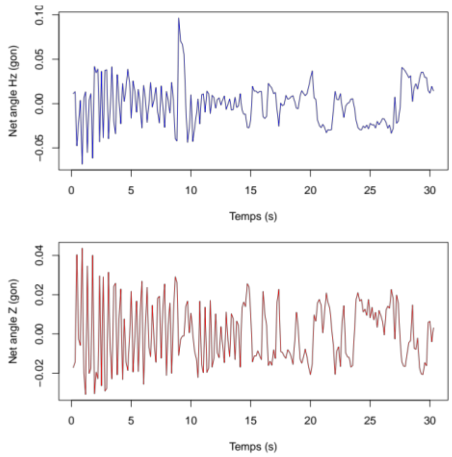


Fig 66. Angular variations on camera line of sight with cable motions

To overcome this aspect, camera is oriented on a frame-by-frame basis. Manual detection of Ground Control Points (GCPs) (spheres, non-coded target and natural points) is performed on the first (and possibly last) frame of the image sequence.

At each time step, correlation is performed about manually detected targets to track the positions of GCPs on each frame. This will enable us to get the most accurate external orientation of cameras at each time step.

Note that since the vehicle might be moving in front of spheres and non-coded targets, the minimal number of GCPs to get external orientation (or at least accurate enough external orientation) is not always fulfilled. In general, four GCPs are needed. In case only three GCPs are visible on an image, 6 observations are available, and 1 of the 7 unknown parameters must be known beforehand. This is done by constraining the optical center (3 parameters) of each camera on the position of the prism used during topometry step (see section 3.2.3). In this case, only 4 unknown parameters are left to be estimated with spatial resection, which could theoretically be done with only 2 visible GCPs. Moreover, even in the case where the critical number of 4 GCPs is reached in an image, prior knowledge of the optical center position will provide much better estimate of the external orientation.

Since all detections would be performed based on the manual detections, it is important that these latter are performed with extreme precision. To what extent a manual detection of a sphere in a image can be done manually is still being investigated.

3.3.4.3.2. Global adjustment

IGN's topometric computation software Comp3D is used for the global least-squares adjustment.

For each epoch, this adjustment is performed with the following measurements:

- Ground targets georeferenced coordinates (from initial topometric computation)

• Camera's center georeferenced approximate coordinates (from initial topometric computation)

• Coded targets coordinate in car frame (from initial topometric computation)

• Ground targets image coordinates (manually determined on the first epoch and then adjusted for every other epoch via correlation, as we know that their image displacement will be small)

• Car coded targets image coordinates (from automatic detection)

Each of these values comes with its precision.

The residuals of measurements are checked to detect any error, especially in coded targets identifiers. If the computation succeeds, the car sub-frame position and orientation is estimated. It can then be used to add the coded targets where id decoding failed to the computation to improve the final precision.

At least three car targets must be seen to estimate the 6 degrees of freedom of the car. With the number of cameras and targets, there will be a good amount of redundancy when the car is in the tracking area.

This redundancy will help to get a good estimation of the final precision.

3.3.5. Simulation

To be able to estimate the final trajectory precision and the best cameras and targets distribution, a simulation environment has been set up. A digital twin of the measurement area has been created within the 3D computer graphics software Blender, with:

- they considered area re-created to scale from digital surface model and orthoimage
- ground targets for cameras orientation with their expected accuracy (via a topometric simulation c.f. 2.3)
- a 3D model of the considered vehicle (with its trajectory and speed)
- camera characteristics (resolution, sensor size, focal length, focus distance, aperture, exposure time)

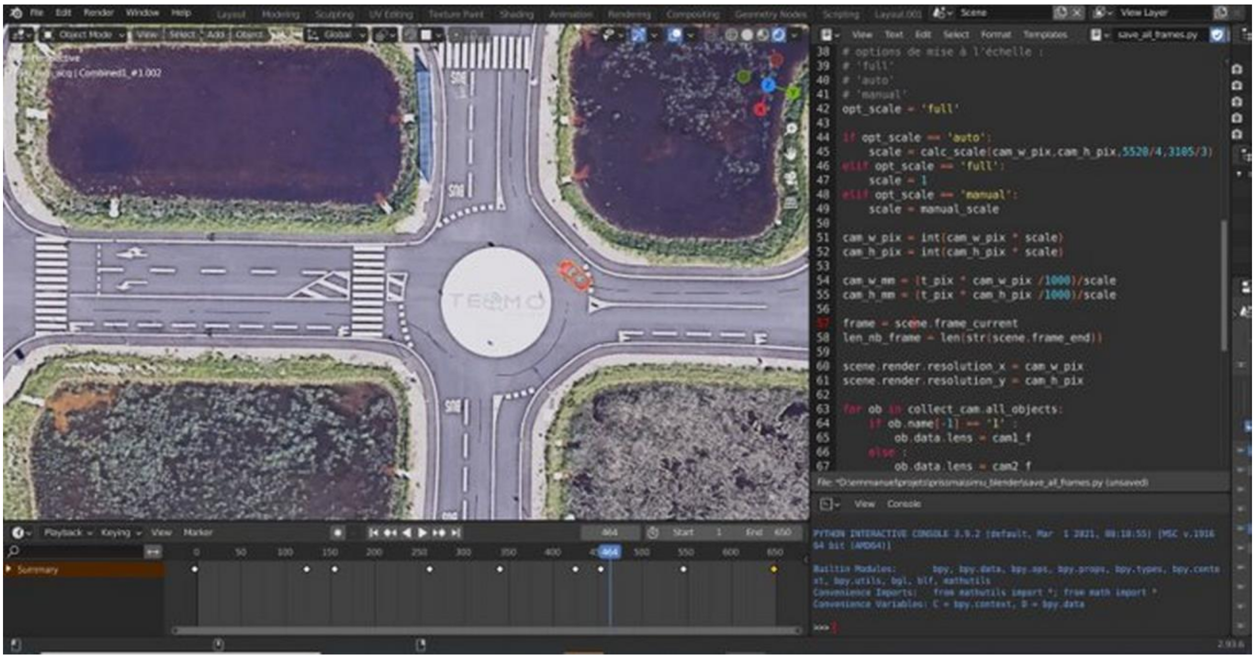


Fig 67. Simulation in Blender

The Blender simulation helped to find optimal target size and coding characteristics: the renders demonstrated their size in the images and their deformations due to the body car curvatures.

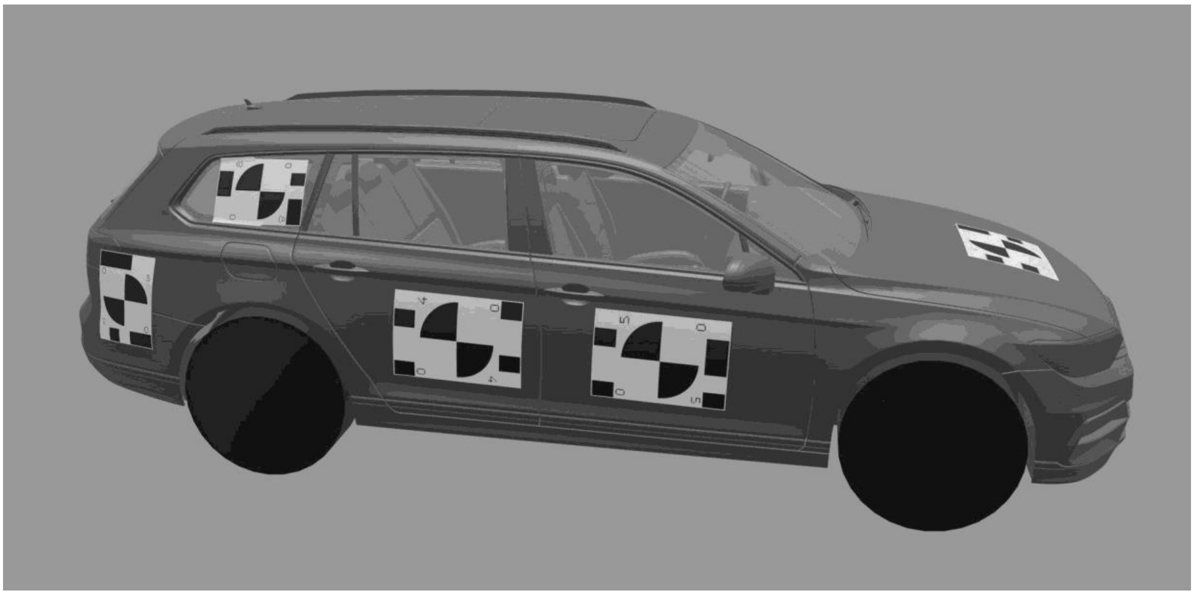


Fig 68. Targets repartition in simulation

Many cameras and targets distributions were tested. All the computation operations processes can be run:

- images generation
- targets detection
- global adjustment
- **comparison with simulated truth**

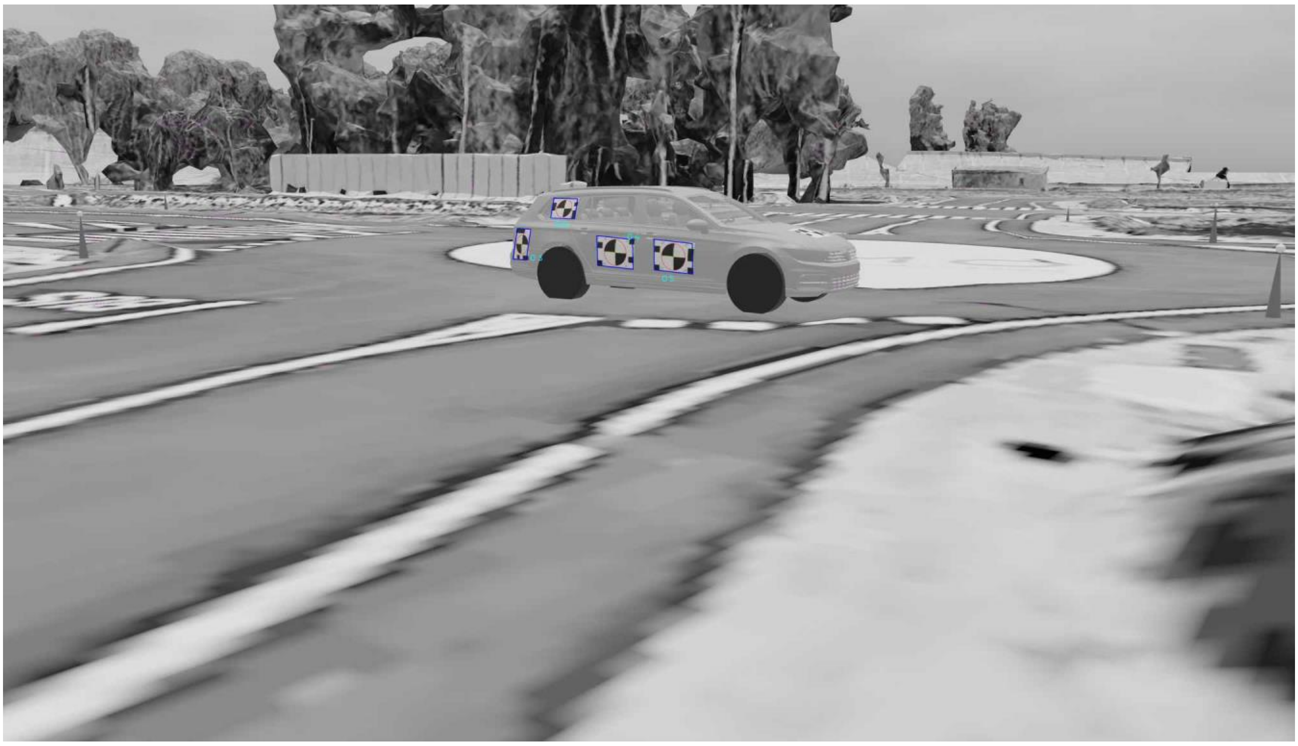


Fig 69. Simulated picture

For global adjustment, upper bound of uncertainty and random errors was added:

- 1.5 px to 3 px for ground targets detection (radial increase to simulate an imperfect lens calibration)

- 2 mm for ground targets georeferenced coordinates

- 2 mm for car targets in car frame

This confirmed that the aimed precision is reachable with our method:

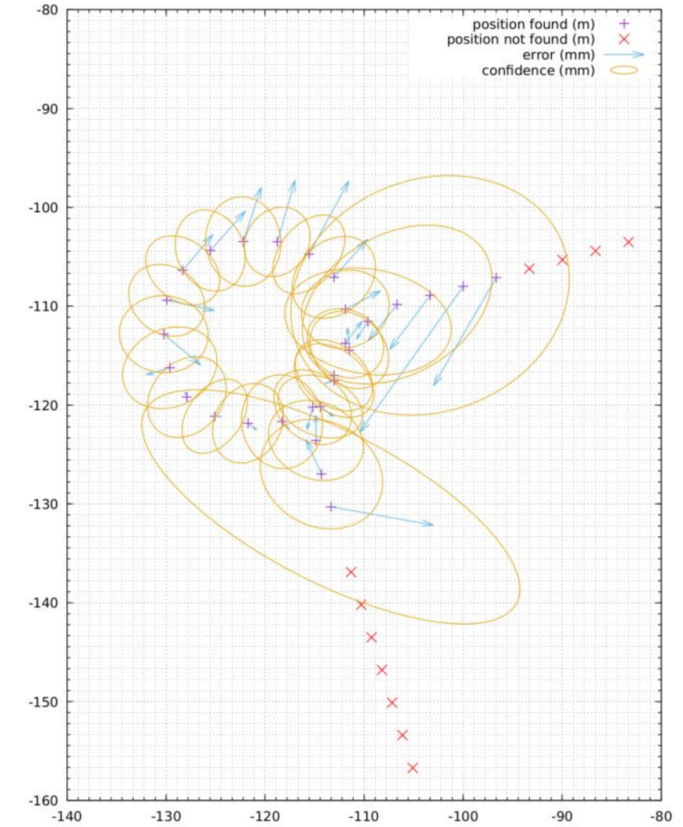


Fig 70. Simulation errors

Inside the roundabout, errors compared to simulation truth are about 5 mm and confidence is coherent with errors.

It also helped to choose an efficient organization, to limit the number of cameras and targets while having redundancy for precision and safety.

3.3.6. Output data

The photogrammetric process produces car sub-frame 3D position and orientation in a common georeferenced ground frame. The orientation part can be limited to heading to simplify trajectory comparison.

Another trajectory is the direct positioning and heading given by the two GEOSTIX on the moving car.

As both trajectories are expressed in a common frame (GNSS georeferencing and timing), and accompanied with their respective accuracy, they can be compared and validated.

The topometric process also provides different point coordinates and precisions that can be useful for analysis vehicle trajectography.

- Vehicle coded targets and other details on car body coordinates can be used to transform trajectography in vehicle navigation system reference frame. This step will be done by Valeo.
- RGF93 coordinates of all points measured in topometry will be provided with uncertainty. This can be useful to check if the HD map created by the vehicle before is correct by comparison between points coordinates in the vehicle point cloud with coordinates given by topometry.

4. Roadmap

Under the hypothesis of a final POC test between 22nd and 26th of January, the road map for IGN is composed of the following steps:

[01] Processing Valeo POC dataset: **three trajectories of 1 to 3 minutes and three cameras.**

[02] **Purchase of 3D printing material and printing of 11 remaining camera supports (prism adaptors)**

[03] End of processing pipeline**: based on the results provided by the Valeo POC dataset**

[04] **Preliminary visit on UTAC test site**

[05] **End of simulations: defining precise, optimal and final dispositions of targets on vehicle, cameras and GCPs on the test site.**

[06] Validation of timestamping of images

[07] **Synchronization (+ validation) between 2 poles, and validation of RJ45 and BNC cables**

[08] **Purchase of RJ-45 and BNC cables for the 4 poles**

[09] Test and synchronization between four poles

[10] Transfer from local internet network to Wi-Fi

[11] Configuration of master pole **to make easier on-site last minute adaptations (exposure time...)**

[12] **Validation of complete synchronization system**

[13] **Development of tools to** improve ergonomic of acquisition system **and be able to detect any problem of image capture during final POC**

[14] **Development of** module for second pass of target detection on vehicle**: for missing targets**

[15] **Calibration of vertical offsets for prism/sphere adaptors**

[16] **Development of a protocol for accurate « manual » acquisition of GCPs on first image on each camera sequence and validation of protocol accuracy**

[17] Development of a script for automatic detection of GCPs **on all images of sequence (based on correlation and acquisition on initial image)**

[18] **Final preparation POC at IGN for validation of accuracy (1/2 scale POC with 2 poles and 2 cameras per pole): comparison of different strategies for trajectory computation and validation final method**

[19] Unit test of Gyro **and test of measurement integration in Comp3D software**

[20] Logistic step: **investigating solution to make the 4 poles easier to deal with on site**

[21] End of topometric step simulation: **validation of accuracy of GCPs, camera positions and absolute georeferencing. Simulation will be also used as a « operation map » for topometric step**

[22] Calibration of Gyro **from Guilands Park, 2 km North of IGN: 1/2 day before/after final POC**

[23] Calibration of 12 cameras**: lens distortion and optical center dep**

[24] **Final POC at UTAC**

[25] **Processing UTAC POC dataset**

[26] Writing final deliverables

#	Task	Time*	2023		2024			
			S51	S52	S01	S02	S03	S04
[01]	Processing Valeo POC dataset	5						
[02]	Printing of 11 remaining camera supports	1						
[03]	End of processing pipeline	5						
[04]	Preliminary visit on UTAC test site	1						
[05]	End of photogrammetric simulations	2						
[06]	Validation of timestamping of images	1						

#	Task	Time*	2023		2024			
[07]	**Synchronization between 2 poles**	1						
[08]	**Purchase of RJ-45 and BNC cables**	0						
[09]	**Test and synchronization between 4 poles**	2						
[10]	**Transfer from local internet network to Wi-Fi**	1						
[11]	**Configuration of master pole**	1						
[12]	**Validation of complete synchronization system**	2						
[13]	**Improve ergonomic of acquisition system**	2						
[14]	**Module for second-pass of target detection**	2						
[15]	**Calibration of vertical offsets**	1						
[16]	**Protocol for « manual » acquisition of GCPs**	1						
[17]	**Script for automatic detection of GCPs**	2						
[18]	**Final POC at IGN for validation of accuracy**	3						
[19]	**Unit test of Gyro and adjustment integration**	2						
[20]	**Logistic step**	1						
[21]	**End of topometric step simulation**	1						
[22]	**Calibration of Gyro**	1						
[23]	**Calibration of 12 cameras**	1						
[24]	**Final POC at UTAC**	1						
[25]	**Processing UTAC POC dataset**	10	Until 2 weeks before deadline					

#	Task	Time*	2023	2024
[26]	**Writing final deliverables**	5	Until deadline	

*** Estimated number of working days on the task**

Chapter 6: TRANSPOLIS POC for WP3: Crossing a traffic light intersection.

1. Introduction

TRANSPOLIS participation was first based on two objectives:

- To support INRIA trial needs to develop their technologies,
- To apply PRISSMA methodologies on an autonomous shuttle POC.

The work to achieve the first objective is going on as expected. However, the progress to reach the second objective was significantly delayed due to the lack of an AI brick in the autonomous shuttle for conducting the tests. Moreover, TRANSPOLIS has invested more time than expected on the requirements and scenario generation working group of WP1.

As a result, TRANSPOLIS decided to set-up a late POC with a reduced scope. This POC focuses on the requirements, safety analyses and set-up of validation plans of an automated shuttle crossing an intersection equipped with traffic lights. This POC shall produce results for the WP3 and WP5.
TRANSPOLIS considers this work within PRISSMA project as a first step toward the validation and type-approval of this kind of system.

2. Framework

For an automated shuttle or any L4 automated vehicle, the ability to cross safely an intersection equipped with traffic lights is an important requirement to allow deployments in urban areas.
This work does not aim at defining a full ODD and requirements of a vehicle or an intersection.
The objectives are to:

- Focus on the validation work of the full system of vehicle + intersection (including AIs),
- To analyse the type-approval and safety demonstration legal background for this system of systems,
- To define tests scenarios,
- To specify test equipment and procedures.

This document presents only a first version of this work that may evolve and be updated in the deliverable L3.6. TRANSPOLIS goals in PRISSMA project is not to validate a system but to develop test protocols and methods. Consequently, the tests that will be carried out in the task3.4 (following this first step) will evaluate test methods.

3. Legal background

In Europe, vehicle or vehicle subsystems shall be type-approved by an approval authority of one of the State Member to be marketed. The verification and type-approval tests are carried out by a technical service. The European regulation 2018/858 [6] defines the type-approval functioning for the main types of vehicles.
The regulation for Automated Driving System (ADS) is the regulation 2022/1426 [7]. It means that the vehicle manufacturer shall present a documentation and a system to be evaluated by a technical service to be type-approved according to this regulation.
The documentation shall present all the safety analyses, requirements, and validations leaded by the manufacturer of the ADS.
It shall include the definition of the ODD. If the crossing of an intersection using a V2X system is included in the ODD, it means that the vehicle manufacturer will impose requirements on road equipment, outside of the vehicle. The documentation shall specify clearly the remote requirements expected for the road equipment.
Manufacturer documentation shall present also all the tests run to validate the ADS of the vehicle working with the specified road equipment.
ADS are usually systems of systems; as a consequence, all components of the ADS, including the AI bricks, shall have been validated and the documentation of the manufacturer shall present this work.
It is important to notice that the regulation 2022/1426 [7] is strongly related to the NATM definition of UNECE WP29 [8]. The NATM defines test pillars: a/ scenarios, b/ virtual testing, c/ track testing, d/ real world testing, e/ Audit, f/ In service monitoring as shown on 72.
However, the definition of which test shall be carried out in simulation and which test on track is not yet clear and it is one of the technical problems PRISSMA project is working on. This POC will also help to progress on this subject by studying a concrete case.

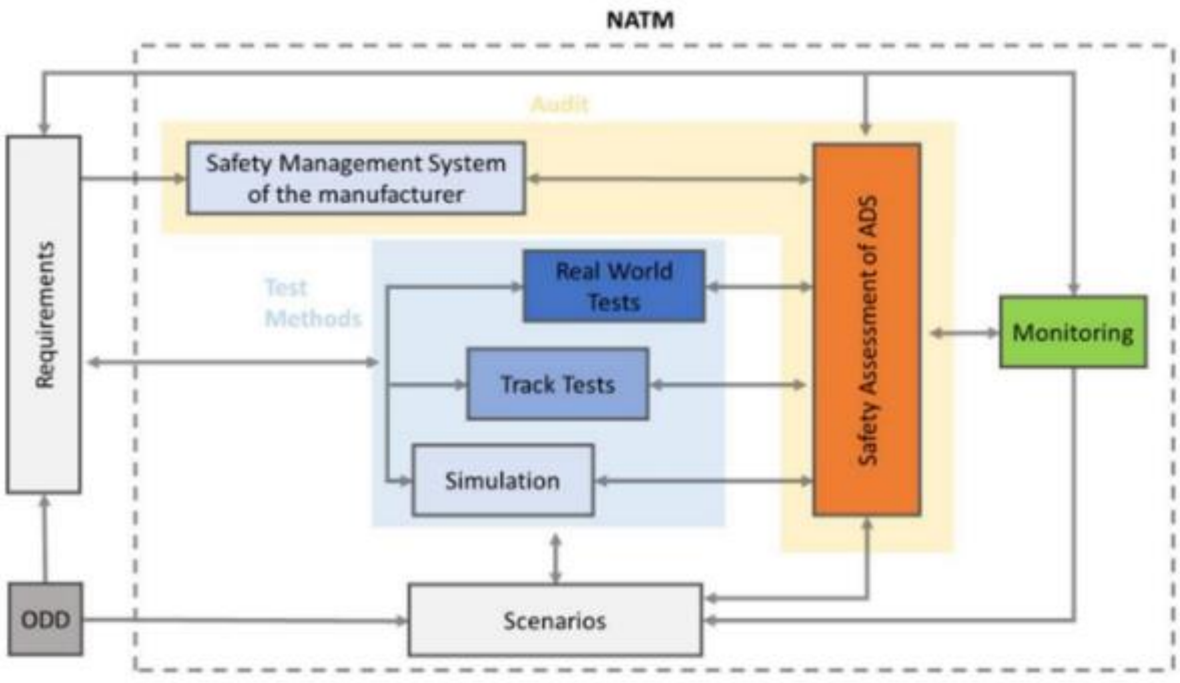


Figure 71: Relationship between NATM pillars and safety requirements [8]

In France, to deploy an automated shuttle service, the decree n°2021-873 [9] defines the stages of safety demonstrations and missions on OQA (qualified bodies).
The deployment on an automated shuttle service is treated as a system: an Automated Road Transport System (ARTS), since it includes remote equipment on the pathway (including the RSU and traffic light controller), a safety management system, a supervision system, etc.
PRISSMA deliverable 1.5 [10] presents these stages in paragraph 3.1.3.
For an ARTS deployment, the validation of the functioning of the shuttle crossing an intersection controlled with traffic lights will be presented in the documentations: DCST, DPS and DS and reviewed by OQAs from the following domains:

1. Functional safety of embedded systems
2. Functional safety of connectivity and positioning devices: for the V2X system (RSU)
3. Cybersecurity
4. Infrastructures and road equipment safety: for the traffic light controller
5. Safety of road behaviour of the vehicles: for the vehicle strategies
7. Global evaluation of the system safety.

The work of deployment of an ARTS also implies some validation tests that shall be presented in the different documentations listed above.

The test scenarios defined in this document can be used for type-approval or ARTS validation according to the stage of the project.

4. POC specification

This POC focuses on traffic light intersections and the ability of an L4 automated vehicle to cross safety such an intersection with a “going straight” manoeuvre.

4.1 Global functioning description

4.1.1 @the vehicle level

The vehicle is equipped with an On-Board Unit (OBU) receiving and emitting standardized ITS-G5 messages:

- Emitted messages: CAM
- Received messages: SPATEM

The vehicle is equipped with a vision-based system able to detect the position and the state of a traffic light. This system is made of a camera and an AI brick to detect the position and the state of the traffic light.
To the experiments, an Openpilot and comma 3X system will be implemented.
A fusion of the information from the OBU and the camera ensure the ability of the vehicle to detect the traffic light status.
A decision algorithm can also be implemented using some AI. This brick will not be developed during the project. However, a work to initiate its specifications and validation could be further develop in the next deliverable.

The camera is also used to detect other vehicles on the way. The path of the vehicle is always the same and go straight through the intersection.

4.1.2 @the intersection level

For the purpose of this reduced POC, this work will concentrate on a cross intersection and the automated vehicle manoeuvre will be to cross straight the intersection.
The intersection is equipped with:

- A Traffic Light Controller (TLC) of Lacroix city brand – this controller is coded in the DIASER language and can be programmed using Lacroix city WinTraffy software.
- A roadside unit sending standardized ITS-G5 messages:
  - SPATEM: for traffic light phases and timing
  - MAPEM: for traffic lights locations
- Classical French Road markings with longitudinal street markings and pedestrian cross-ways. Figure 72 presents the configuration of the intersection.

The TLC is set with a fix and repetitive cycle.

4.2 Environmental conditions

- The shuttle shall operate from 6:00 am to 22:00 PM – all year long.
- The shuttle shall work under heavy run.
- The shuttle shall not work in foggy situation when the visibility is reduced to 20m.
- The shuttle shall not work under snow or on icy road.

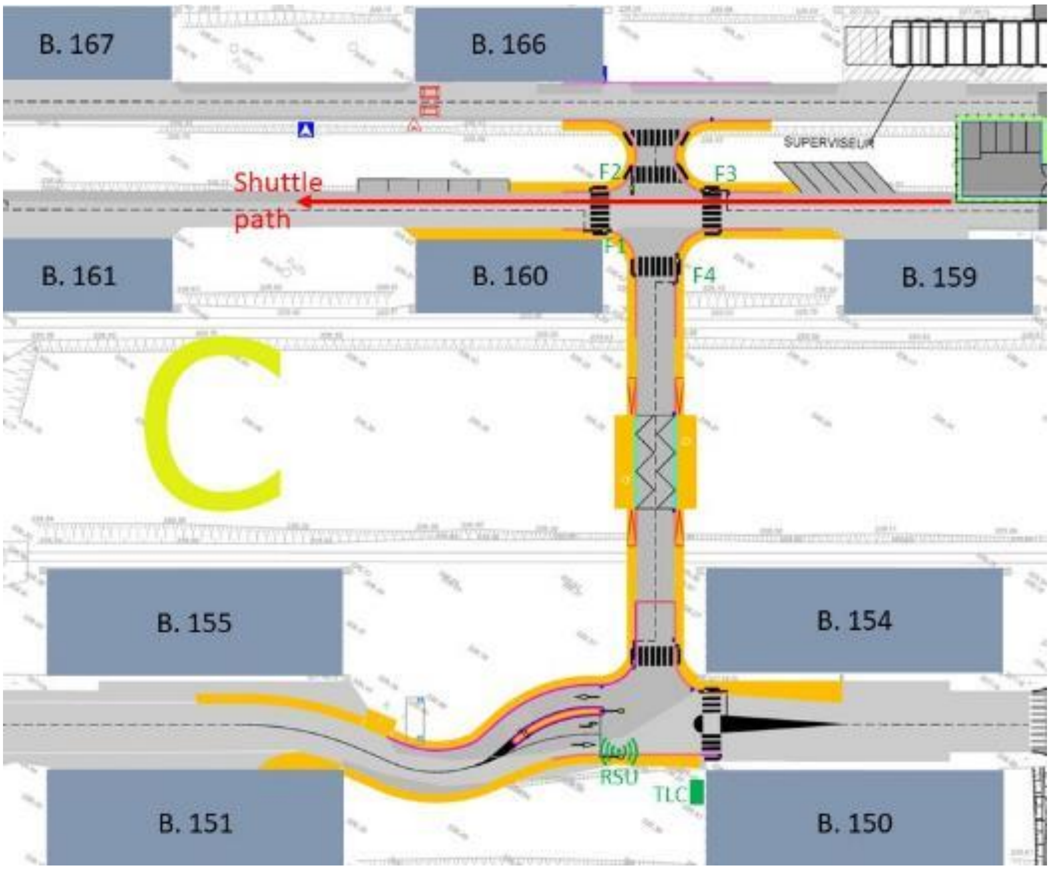


Figure 721 : Intersection configuration in TRANSPOLIS city area C. Scale: Building 150 is 50m long. RSU: roadside unit. TLC: Traffic light controller

5. Test scenarios

5.1 Introduction

Tests scenarios can be defined according to requirements, ODD, safety analyses, functional safety analyses, accidentology databases, etc. Scenario generation is further presented in PRISSMA L1.5 [10]. This deliverable also presents the project work about the requirements.
According to the requirements and the stage of the project, tests can be carried out in simulation, on tracks or on open roads. The scenarios defined here are mainly focused on what can be done on tracks.
Following the work in WP1, the test scenario list given below is defined based on requirements and ODD definition. The method presented in Annex A of deliverable L1.5 [10] for nominal scenarios based on the ODD is applied. Figure 73 shows the three stages of this method.
Since the POC focuses on an intersection and one vehicle manoeuvre, the first branches of the ODD are specified by the POC itself. Then, at stage 2, traffic conditions and other road users' behaviour shall be considered. Here, only a vehicle running ahead of the EGO vehicle is taken into account. For a real and complete validation plan, vulnerable road users such as pedestrian are also to be considered but the detection of these other users in not in the objective of this test plan.
The last stage introduces the masks and the environmental conditions. All the basic tests are to be carried out in normal daylight conditions. The environmental conditions are to be crossed and added to the basic scenarios if they should have an effect. This “addition” of layers in the test scenarios is given in the paragraph 5.5.

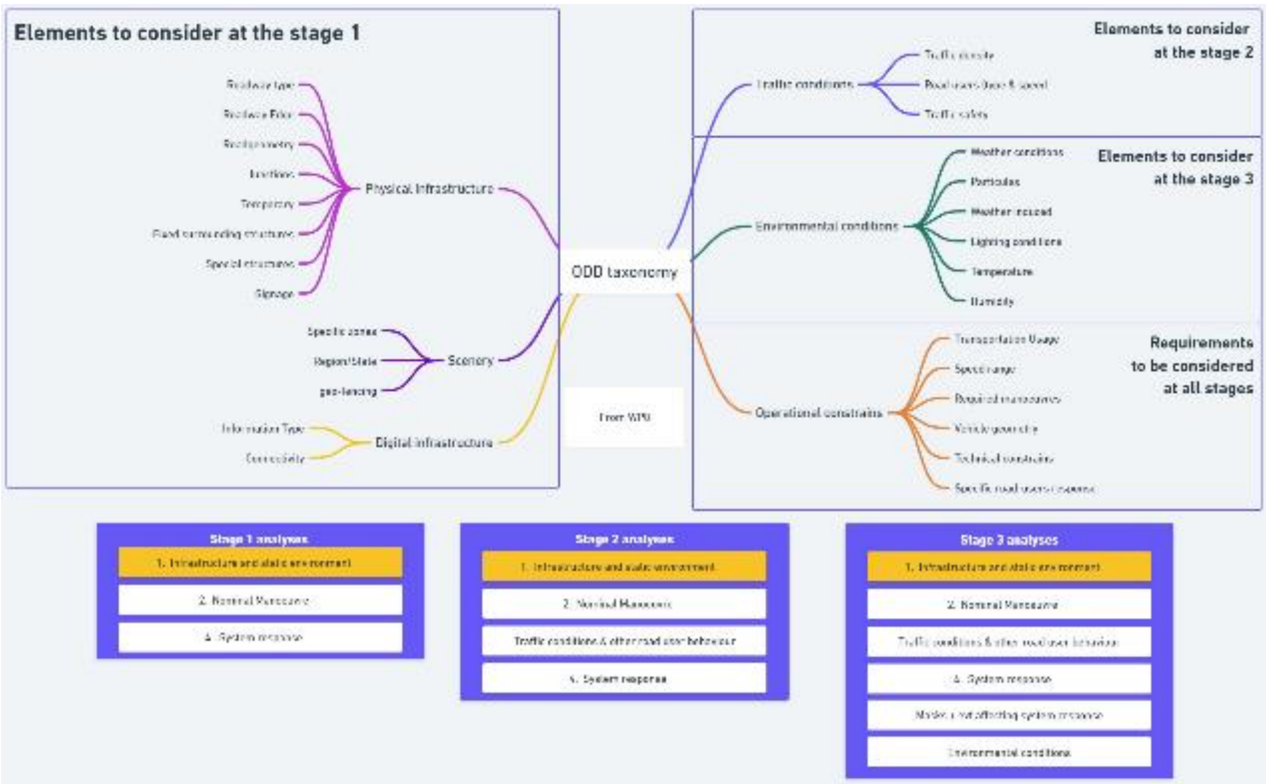


Figure 722 : Scenario generation from ODD and requirements [10]

Some classical failure scenarios are also proposed.

5.2 Vehicle systems test scenarios (V)

5.2.1 Validation of the vision by camera (AI) of the Traffic lights:

These tests must be carried out by the system manufacturer and be presented in the type-approval documentation.

Scenario V-camera-S.x: Testing the ability of the camera to locate and detect the correct status of a traffic light whatever is status in a **s**tationary position. X for the iterations with respect to the position.
Scenario V-camera-D.x: Testing the ability of the camera to locate and detect the correct status of a traffic light whatever is its status on a dynamic trajectory. x- for the iteration with respect to the TLC plan.
Scenario V-camera-S-Rep.x: Testing the ability of the camera to locate and detect the correct status of a traffic light whatever is status in a **s**tationary position with multiple repetitions to assess the repeatability of the system. X for the iterations with respect to the position.
Scenario V-camera-D-Rob.x: Testing the ability of the camera to locate and detect the correct status of a traffic light whatever is status, random tests for **robustness** validation, speed & position variations. x for iteration or this test that shall be repeated several times (to be defined)
Scenario V-camera-S.alldaylong: Testing the ability of the camera to detect the correct status of a traffic light on the path for on complete day (time of the OD)
Scenario V-camera-Latency: Evaluating the latency of the camera to detect the correct status of a traffic light
Scenario V-camera-position: Evaluating the accuracy of the camera to locate the traffic light

These tests scenarios can be run on a proving ground or in simulation at a validation stage.
The documentation may also present some information about the set of data used for the training of the AI.

The reliability of this component shall be validated on real road runs or using SIL or HIL with a large set of data different from the data used for the training.

The environmental conditions have an important influence on the performance of this system. It is important to run all these tests in all environmental conditions and to define the "boundaries" of the conditions in which the system performances are validated. These boundaries are significant results for the qualification of the ODD limits.
Environmental conditions at listed paragraph 5.5.

5.2.2 Validation of OBU information (SPATEM)

Scenario V-OBU-S: Testing the ability of the OBU to receive SPATEM and to extract the precise traffic light status and remaining time in a static position.
Scenario V-OBU-Latency-S: Evaluating the latency of the ITS-G5 messages and OBU information in a static position.
Scenario V-OBU-Latency-D: Evaluating the latency of the ITS-G5 messages and OBU information on a dynamic test.
Scenario V-OBU-D: Testing the ability of the OBU to receive SPATEM and to extract the precise traffic light status and remaining time on the shuttle path (Coverage). This test shall be repeated several times (number to be define) for assessing the robustness.

5.2.3 Validation of information fusion (F) from the camera and the UEV information

Scenario V-F-S: Testing that OBU and camera information are consistent in a static position.
Scenario V-F-D: Testing that OBU and camera information are consistent while running on the pathway.
Scenario V-F-D-robustness.x: Testing that OBU and camera information are consistent while running on the pathway and repeating this test several times.

N.B. repeating the exact configuration of a dynamic test is complex for such de complicated system. Repeatability should be assessed with the robustness tests.

These tests can also be crossed with the environmental condition modalities.

5.2.4 Validation of the decision brick (D)

Scenario V-D-cycle.x: Testing the vehicle dynamic behaviour with iterations to cover all the traffic light cycle and analyse all situation.
Scenario V-D-repeatability.x: Testing the vehicle dynamic behaviour repeatability by starting the vehicle run always at the same traffic light cycle time. X for several iterations.
Scenario V-D-robustness.x: Testing the vehicle dynamic behaviour robustness by starting the vehicle run at one or two strategic times of the traffic light cycle that implies a shirt in the vehicle strategy (stopping at red light or running before it turns red). X for several iterations around the strategic points.

These tests can also be crossed with the environmental condition modalities.

5.3 Road equipment tests (R)

Traffic light controller is a standard product with many safety certifications. Testing its functioning is not an objective.
These tests shall be carried out during an ARTS deployment preparation, or by a technical service on its tracks to be able to run type-approval tests.
RSU and its message conformity with the Specifications given by the vehicule or ADS manufacturer shall be verified. Usually, ETSI specifications are the basis [11].

Scenario RSU-SPATEM: Testing the validity of the signal phase and timing in the SPATEM sent by the RSU.
Scenario RSU-coverage: Testing the good reception of the RSU messages on the ways of the intersection.

5.4 Full system tests (FS)

5.4.1 Nominal scenarios (N)

Scenario FS-N-1.x: Testing the ability of the vehicle to cross the intersection whatever the status of the traffic light – x gives the iteration of the test in the TLC plan. X for the iteration according to TLC cycle.

Scenario FS-N-2.x: Testing the ability of the vehicle to cross the intersection whatever the status of the traffic light, following a slow car.
Scenario FS-N-3.x: Testing the ability of the vehicle to cross the intersection whatever the status of the traffic light, following a slow truck (hiding the traffic light for the camera).

5.4.2 Failure scenarios (F)

Scenario F-camera-off: Testing the behaviour of the system / vehicle if the camera is electrically switched off.
Scenario F-camera-dirt: Testing the behaviour of the system / vehicle if the windscreen in front of the camera is dirty.
Scenario F-OBU-off: Testing the behaviour of the system / vehicle if the OBU is electrically switched off.
Scenario F-light-ooo: Testing the behaviour of the system / vehicle if the current traffic light bulb is out of order.
Scenario F-RSU-off: Testing the behaviour of the system / vehicle if the OBU is electrically switched off.
Scenario F-TLC-default: Testing the behaviour of the system / vehicle if the TLC is in default mode (flashing yellow light)
Scenario F-TLC-off: Testing the behaviour of the system / vehicle if the TLC is electrically switched off.

5.5 Environmental conditions

According to the stage of the project, type approval (vehicle), or deployment of an ARTS, the environmental conditions of the ODD or OD shall be analysed to define metrics and boundaries. In this POC, the environmental conditions are specified in the paragraph 4.2. The following tests are proposed:

Scenario XX – Night: Testing the system, at night without urban lighting or with urban light according to the ODD or OD.
Scenario XX – Rain : Testing the system during a rainy day. The rain conditions shall be characterised by a metrics.
Scenario XX – Night - Rain: Testing the system, at night without urban lighting in rainy conditions

The pathway is heading northwest, so grazing light conditions are not considered in this POC but it can be a test modality.

The system is not supposed to operate in foggy conditions, meaning that it shall have a mean to control it ODD boundary. Additional scenarios to verify the ability of the system to detect that it is operating inside the boundaries of its ODD shall be define and tested.

6. Test equipment

6.1 The tracks

Tests will be carried out in TRANSPOLIS urban area in les FROMENTAUX site. The ground network and track configuration are shown on Figure 74. The shuttle path of the POC is presented in yellow.

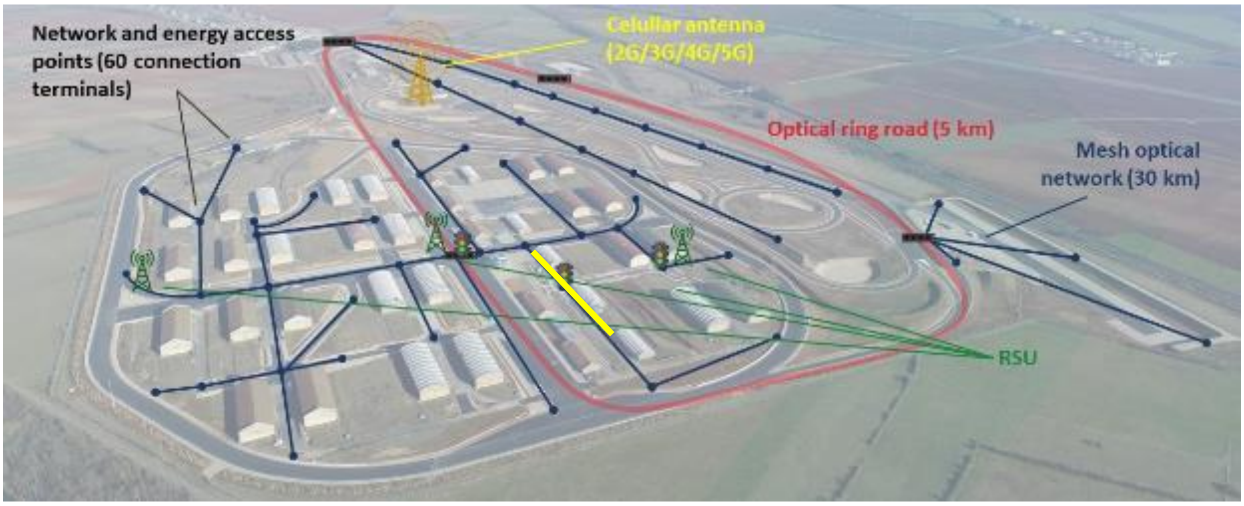


Figure 723 : Les Fromentaux proving ground of TRANSPOLIS and its ground network.

Figure 724 : Picture of the POC intersection. F3 in the front plan (see Figure 721)

6.2 The intersection equipment

As presented Figure 721, the intersection is equipped with full and classical French traffic light system:

- A traffic light Controller (TLC) Lacroix traffic TRAFFY
  - Aluminium street cabinet 800, 1250 / 800 / 420, RAL 1015
  - General electrical protection 32A 300 mA
  - CPU GPS
  - Traffic lights cards
  - Command agent for 2 positions, Lacroix city
- Four Alumix traffic lights equipped with Equinoxe LEDs
- Four Aluminium posts
- Four R12 pedestrian signals.

The controller is programmed using a Winfraffy software. This software is installed on a Windows computer connected to the controller and allows to define all the intersection configuration and to set traffic lights cycles and programs. All traffic light system specifications are given in Annex.

An ITS-G5 Roadside Unit of Lacroix city manufacturer is connected with ethernet to the TLC. THE RSU logs (csv of wireshark captures) will be used for verifications and validations. The specifications of the RSU are presented in annex.

6.3 The vehicle POC

The equipped vehicule is a 2018 FORD FOCUS.

The AI brick is Openpilot, installed on a comma 3X device connected to the vehicle CAN with a specific FORD hardness.

(a) (b)

Figure 25 : (a) COMMA 3X display – Installation inside the Ford Focus

Comma 3x technical specifications:

- CAMERAS
  - Three 1080p cameras w/ 140 dB of dynamic range: dual-cam 360° vision and a narrow cam to see far-away objects
- PROCESSOR
  - Qualcomm Snapdragon 845
- CAN FD ENABLED
  - Supports CAN FD vehicles without extra hardware
- STORAGE
  - 128GB built in storage
- CONNECTIVITY
  - LTE
  - Wi-Fi
  - High-Precision GPS
- NIGHT-VISION
  - IR LEDs for interior night-vision monitoring

- DISPLAY
  - 2160x1080 Beautiful OLED display
- PORTS
  - OBD-C port (USB-C w/ CAN)
  - USB 3.1 Gen 2 port

A Lacroix city On-Board Unit is also installed in the vehicle. The specification of this OBU is presented in Annex 9.3. The OBU logs will be used for verification and validation.
It communicates in Wi-Fi with a computer that is also connected with Ethernet to the Comma device.

6.4 The vehicle equipment for the tests

To validate the functioning of the systems, the position of the vehicle shall be recorded; for that, the equipment presented in Table 7.

Information and event synchronisation shall also be evaluated by filming HMI. For that, a video synchronisation system for four cameras can be used (see **Erreur ! Source du renvoi introuvable.**7)

The environmental conditions shall also be recorded using equipment such as a luxmeter, rain gauge or other devices to be specified.

Table 7 : Motion recording system

OxTS RT3003 XG	SyncOmni
	
Inertial measurement Unit and GPS with dual antenna	**Controller**
Performance (accuracy): Position: Glonass L1,L2 0.01m Speed : 0.05km/h RMS Roll/Pitch: 0.03° Cap : 0.1° Angular speed : 0.01°/s Acceleration: 0.01%/ Taux 100m/s² Frequency : 100Hz	Functionality: Differential Measurement in real time (up to 16 targets) Synchronisation of data CAN output
ABD Camera	

	Specifications: 1 Camera Logitech C920 1080p 30fps Synchronised videos with motion data

Figure 726 : PDRIVE system (tests and training no longer produced)

7. Conclusion

This work is a first step to evaluate the possibility to validate an ADS including one or two AI bricks. Again, the objective of this work is not to validate a product but to show the ability of TRANSPOLIS to validate a potential POC including its AI bricks.
Protocol details will be defined and tested. Further reflexion will be carried out about the completeness of the scenarios and considered requirements. The results of these investigations will be presented in PRISSMA deliverable L3.6.
No cybersecurity scenario is presented in this document. They will be defined and tested within the framework of WP5.

8. Acronyms

AI: Artificial Intelligence
CAM : Cooperative Awareness Message
MAPEM : MAP Extended Message
OBU: On Board Unit
RSU: Road Side Unit
SPATEM : Signal Phase and Timing Extended Message
TLC: Traffic Light Controller

9. Annex

9.1 Traffic light system specifications

Traffic LACROIX

1 - Introduction

TRAFFY® est un contrôleur de feux de signalisation tricolore de dernière génération, étudié et fabriqué par LACROIX Traffic. Modulaire et évolutif, il est capable de répondre à tous types de besoins de gestion du trafic et d'optimisation des plans de déplacement urbains. De part l'intégration de nouveaux composants de très faible consommation et RoHS, la nouvelle version de TRAFFY® apporte une solution fiable, maîtrisée et durable tout en conservant une compatibilité avec les générations précédentes. Les fonctions de régulation, des plus simples au plus complexes, peuvent être traitées par TRAFFY®. Depuis la traversée piéton jusqu'à la gestion des tramways et BHNS, une solution unique sur le marché.

2 - Normes applicables

NF P 99-100	Caractéristiques complémentaires des sécurités fonctionnelles d'usage
NF P 99-105	Caractéristiques fonctionnelles d'usage
NF P 99-110	Echanges de données par liaisons fil-à-fil avec des organes externes
NF P 99-022	Norme contenant les essais fonctionnels à passer pour être homologué
NF EN 12675	Exigences de sécurité fonctionnelle
NF P 99-071	Protocole de communication DIASER
NF P 99-000	Terminologie, définitions de termes employés en régulation du trafic
NF P 99-050	Principe de maintenance des contrôleurs de carrefours à feux
NF P 99-060	Conditions de mise en œuvre des équipements

3 - Marquage CF

Le contrôleur TRAFFY® est marqué **CF** conformément à l'arrêté du 18/06/2003, relatif à l'attestation de conformité des contrôleurs : **TRAFFY 3 CF08**

4 - Caractéristiques Techniques

Générale	- *Tension de fonctionnement : 230Vac, 50 Hz* - *Température de fonctionnement : -25 à+75 °C*

4.1 – Racks TRAFFY®

Rack de base de dimensions
H=295mm - L=240mm - P=210mm

1 - Introduction

La lanterne ALUMIX a été pensée design, écologique et pratique afin d'être intégrée dans tous les types de paysages urbains. Son profil en aluminium recyclable présente une très bonne protection thermique et mécanique assurant une longévité du produit tout en respectant l'environnement.
Grâce à ses possibilités de personnalisation aux couleurs de votre ville (RAL ou Futura), l'ensemble de la gamme donnera une identité propre à votre agglomération. Ses fixations arrière non visibles renforcent le design et offrent un réglage en hauteur pratique pour l'installation notamment en rénovation. Les inserts équipés de diodes "haut flux", démontables individuellement et d'encombrement très réduit, contribuent à la faible épaisseur du profil.
Le feu ALUMIX est conforme aux exigences de la norme française NFP99-200 et Européennes NF EN 12368.

2 - Normes applicables

NF EN 12368	Équipement de régulation du trafic Têtes de feux.
NFP 99-200	Régulation de trafic routier - Signaux lumineux de circulation - Caractéristiques techniques.
NF EN 50293	Compatibilité électromagnétique - Systèmes de circulation routière - Norme de produit.
NF EN 60598-1	Luminaires Sécurité - Prescription générale et essais.
NF EN 60529	Degrés de protection procurés par les enveloppes.

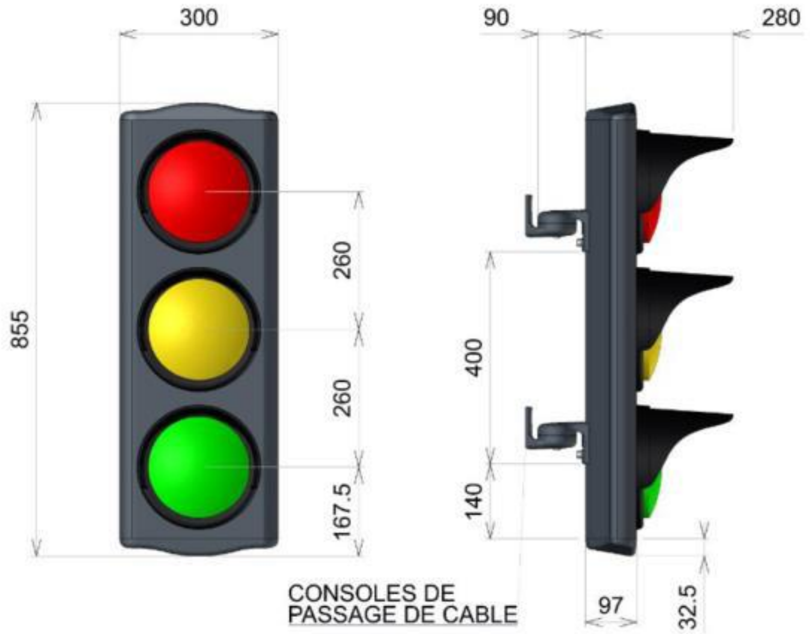
3 - Marquage CE

Équipé du kit à diodes Équinoxe, le feu ALUMIX bénéficie du marquage CE - Certificat n°1826 – CPD – 120303 suivant NF EN 12368-2006.

CE

4 - Caractéristiques techniques

Dimensionnelle



Caractéristiques mécaniques

Matières - Corps - Fixation et flasque - Couronne et visière	- *Profilé d'aluminium peint* - *Fonte d'aluminium peinte* - *Polycarbonate noir texturé stabilisé aux UV*
Poids	*10.6 Kg*
Résistance vibratoire	*Conforme à la norme NF EN 12368*

1 - Introduction

Les kits Equinoxe à foyer centraux équipés de diodes haut flux, augmentent les performances optiques, la fiabilité du produit tout en réduisant la puissance consommée. Le kit Equinoxe associé à la lanterne ALUMIX donnera une identité propre à votre agglomération tout en offrant une sécurité accru.
Le feu ALUMIX équipé du kit à diodes Equinoxe est conforme aux exigences de la norme française NFP99-200 et européenne NF EN 12368.

2 - Normes applicables

NF EN 12368	Équipement de régulation du trafic Têtes de feux.
NFP 99-200	Régulation de trafic routier - Signaux lumineux de circulation - Caractéristiques techniques.
NF EN 50293	Compatibilité électromagnétique - Systèmes de circulation routière - Norme de produit.
NF EN 60598-1	Luminaires Sécurité - Prescription générale et essais.
NF EN 60529	Degrés de protection procurés par les enveloppes.

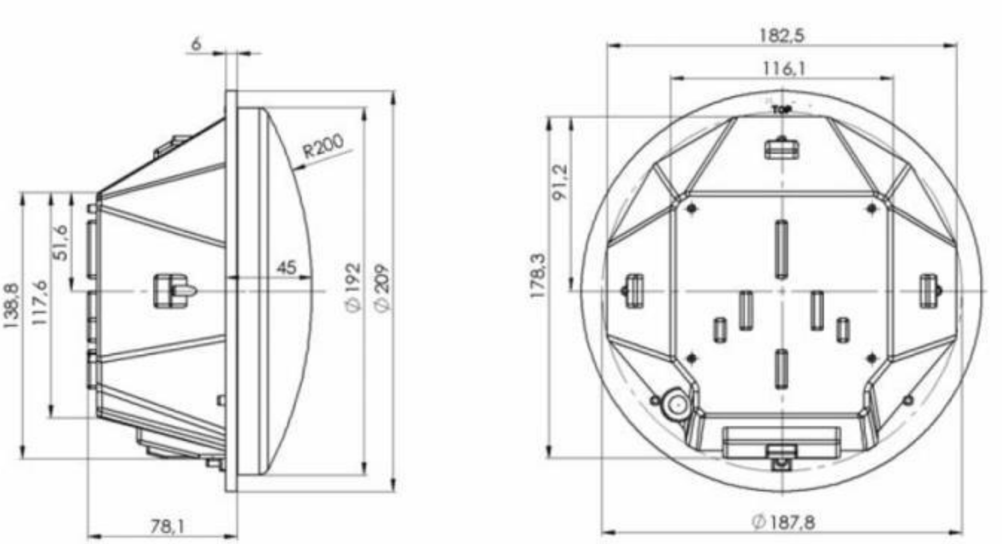
3 - Marquage CE

Équipé du kit à diodes Équinoxe, le feu ALUMIX bénéficie du marquage CE - Certificat n°1826 – CPD – 120303 suivant NF EN 12368-2006.

CE

4 - Caractéristiques techniques

Dimensionnelle 200mm



Caractéristiques mécaniques

Matières - Corps - Lentille	- *Polycarbonate opaque* - *Polycarbonate incolore fumée stabilisé aux UV*
Poids	*0.7 Kg*
Résistance vibratoire	*Conforme à la norme NF EN 12368*

9.2 Road Side Unit specifications

https://www.lacroix-city.com/wp-content/uploads/sites/7/2019/10/0322_LACROIX-CITY_V2X_Station_outdoor_Brochure_EN-1.pdf

V2I Station

C-ITS ROAD SIDE UNIT - **OUTDOOR**

V2I Station

C-ITS ROAD SIDE UNIT - **OUTDOOR**

LACROIX's V2I station is an interoperable and cybersecure solution to connect infrastructure and vehicles.. The V2I Station operates on 5.8 / 5.9GHz bands according to US or European standards (WAVE 1609 / EN 302 571). The unit is designed to ensure permanent and rugged use along the roads, while ensuring technological scalability. A wide range of interfaces is available to communicate with sensors for advanced vehicle perception and existing traffic lights.

The embedded software includes a Web HMI and API, as well as all application / communication stacks required to communicate with the vehicles and the traffic management centers.

Delivered with a Bluetooth traffic sensor, the unit can be used to measure travel times, detect traffic jams or determine Origins / Destinations.

BENEFITS

- → Reduced risks of accidents
- → Improved quality of services for road users
- → Less traffic congestion
- → Time savings during commissioning and maintenance
- → Cybersecure data exchanges
- → Interoperable solution that fully interfaces with existing systems.

APPLICATIONS

- Connected Highways
- Dynamic Lighting
- Automated Intersection Crossing
- Roadworks warning
- Connected VMS

FONCTIONS

- Traffic light priority
- Hazardous locations notifications
- Bluetooth travel time
- V2X vehicles data agregation

COOPERATIVE APPLICATIONS VEHICLE-INFRASTRUCTURE

TECHNICAL CHARACTERISTICS

- Dimensions: 10,6 x 11,8 x 2,7 in (270 x 270 x 62 mm)
- Weight : 3,30 lbs
- Temperatures : -58°F / +158°F (-40°C / +70°C)
- IP67

POWER

- Consumption : 6W
- PoE 802.3af / 12V-48V DC

ANTENNAS

- V2X x2 omnidirectional 5.8Ghz 7dBi
- Bluetooth omnidirectional 3dBi
- Integrated GPS / Cellular Antennas
- Connector V2X type N
- V2X Directional Antenna (option)

INTERFACES

- V2X and C-V2X
- Bluetooth
- GNSS
- Ethernet
- 3G/4G
- I / O extension module (option)

9.3 On-Board Unit specifications

https://www.lacroix-city.com/wp-content/uploads/sites/7/2019/10/0322_LACROIX-CITY_V2X_V2V_Unit_Brochure-EN-1.pdf

V2V Unit

C-ITS ON BOARD UNIT
FOR CONNECTED VEHICLES

V2V Unit

C-ITS ON BOARD UNIT FOR CONNECTED VEHICLES

LACROIX 's V2V Unit is a communication unit suited to provide a V2X connectivity to all service vehicles. Thanks to an embedded Wifi access point, data is displayed on portable devices (Tablet, smartphone) and Alerts can warn the driver. Manual V2X Alarms can also be raised through the display and all the field data is sent to management centers through the 3G/4G link of the device.

BENEFITS

- → Reduced risks of accidents
- → Improved quality of services for road users
- → Less traffic congestion
- → Time savings during commissioning and maintenance
- → Cybersecure data exchanges
- → Interoperable solution that fully interfaces with existing systems.

APPLICATIONS

- Priority for public transport (PrioV2X)
- Priority for emergency vehicles
- Roadworks warning for automated vehicles
- Connected road patrol
- In-vehicle signage

FUNCTIONS

- Traffic data collection
- Real-time vehicle location
- Hazardous locations notifications
- V2X and C-V2X radio communication
- Third Party integration through API

COOPERATIVE VEHICLE TO VEHICLE AND VEHICLE TO INFRASTRUCTURE APPLICATIONS

ENVIRONMENTAL

- Dimensions : 158*174*31 (mm)
- Weight : 500g
- Temperature : -40°C / +70°C

ANTENNAS

- Multiband V2X / cellular / GPS,
- Magnetic base
- Seperate Wifi antenna

HARDWARE INTERFACES

- DSRC V2X double channel
- C-V2X 5G
- GNSS
- Cellular 3G/4G
- Wifi (option)
- Bluetooth (option)

CONCLUSION

In conclusion, this deliverable stands as an important part of the PRISSMA project, showing how traditional track testing can be included in the use of AI-enabled vehicles, and how it can be complemented with new approaches to bench testing or the addition of new technologies. The five evaluation protocol (for each POC) presented herein offer a wide range of possibilities for track and bench testing in this context, while retaining a common structure. They show not only how to generalize conventional and regulatory track testing, but also how to integrate the simulation approach. Being able to couple real-world testing with simulation is a real break-through compared with traditional vehicle homologation. If simulation is involved, the protocol developed in Deliverable 2.7 must also be applied for this specific part. However, because of the versatility of the test facilities, and in contrast to what was done in Deliverable 2.7 for the simulation part, the choice was made here to keep only a common structure for the test protocols and to emphasize the declination in the POCs rather than the reverse.

The inclusion of use cases through five Proofs of Concept demonstrates the practical application of the protocol common structure but also showcases the adaptability and versatility of the PRISSMA methodologies across various scenarios. Each POC exemplifies the protocol's effectiveness in assessing the AI's performance, ensuring its robustness, safety, reliability under diverse applications and with regard to several technologies that could be tested within this work package 3. Deliverable 3.4 will shed further light on the methods used to validate these tools.

Moving forward, this deliverable serves as a springboard for continued refinement, optimization, and expansion of the classical test tracks. The collaborative efforts involved in its development reflect our commitment to ensuring the safety, efficiency, and advancement of autonomous vehicles.

REFERENCES

1. Lukas Rummelhard, Amaury Nègre, Christian Laugier. Conditional Monte Carlo Dense Occupancy Tracker. *18th IEEE International Conference on Intelligent Transportation Systems*, 2015, Las Palmas, Spain.
2. V. Judalet, L. Désiré, G. Saint Pierre, N. De Rus, G. Thiolon. Livrable 4.1: Choix des scénarios d'interactions VA / VL. *Projet Surca*, financé par la FSR et la DSR, 15 p. 2020.
3. T. Serre, V. Battaglia, C. Chauvel. Livrable 6.1: Choix des scénarios d'interactions VL/piétons et cyclistes. *Projet Surca*, financé par la FSR et la DSR, 15 p. 2020.
4. Thomas Genevois, Jean-Baptiste Horel, Alessandro Renzaglia, Christian Laugier. Augmented Reality on LiDAR data: Going beyond Vehicle-in-the-Loop for Automotive Software Validation. *IV 2022 - 33rd IEEE Intelligent Vehicles Symposium IV*, 2022, Aachen, Germany.
5. Jean-Baptiste Horel, Robin Baruffa, Lukas Rummelhard, Alessandro Renzaglia, Christian Laugier. A Navigation-Based Evaluation Metric for Probabilistic Occupancy Grids: Pathfinding Cost Mean Squared Error. *ITCS 2023 - 26th IEEE International Conference on Intelligent Transportation Systems*, IEEE, 2023, Bilbao, Spain.

6. European Parlement, Regulation (EU) 2018/858 - the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles,, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R0858, 2018.

7. Commission Européenne, «Réglement d'exécution (UE) 2022/1426 établissant des règles relatives à l'application du règlement (UE) 2019/2144 du Parlement européen et du Conseil en ce qui concerne les procédures uniformes et les spécifications techniques,» 5 aout 2022 pour la réception par type des systèmes de conduite automatisée (ADS) des véhicules entièrement automatisés - 5 août 2022. [En ligne]. Available: https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32022R1426.

8. UNECE, «New Assessment/Test Method for Automated Driving (NATM) - Master Document,» juin ECE/TRANS/WP.29/2021/61. [En ligne]. Available: https://unece.org/sites/default/files/2021-04/ECE-TRANS-WP29-2021-61e.pdf.

9. Ministère de la transition écologique, Décret n° 2021-873 du 29 juin 2021 portant application de l'ordonnance n° 2021-443 du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation, https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043729532, 2021.

10. PRISSMA project, [L1.5] TESTS AND AUDIT REQUIREMENTS - FINAL REPORT, 2023.

11. ETSI, TS 103 301 V1.2.1 Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Facilities layer protocols and communication requirements for infrastructure services, https://www.etsi.org/deliver/etsi_ts/103300_103399/103301/01.02.01_60/ts_103301v010201p.pdf, 2018.

12. Joseph Redmon, Ali Farhadi. Yolov3: An Incremental Improvement. https://doi.org/10.48550/arXiv.1804.02767, 2018.
13. J. Terven and D. Cordova-Esparza, “A comprehensive review of yolo: From yolov1 to yolov8 and beyond,” arXiv preprint arXiv:2304.00501, 2023.
14. https://www.stereolabs.com/zed-2i/ Stereolabs Website
15. Sébastien Liandrat, Pierre Duthon, Frédéric Bernardin, Amine Ben Daoued, Jean-Luc Bicard, A review of Cerema PAVIN fog & rain platform: from past and back to the future, ITS WORLD Congress 2022
16. H. koschmieder, Theorie der horizontalen Sichtweite, Beitrage zur Physik der freien Atmosphare 1924
17. Il Gordon, Standard errors of heritabilities based on perennial observations, with application to Yorkshire fog grass, Euphytica, Springer, 1979
18. Charlotte Segonne and Pierre Duthon, Qualification of the PAVIN Fog and Rain Platform and Its Digital Twin for the Evaluation of a Pedestrian Detector in Fog, *J. Imaging* 2023, *9*(10), 211; https://doi.org/10.3390/jimaging9100211