

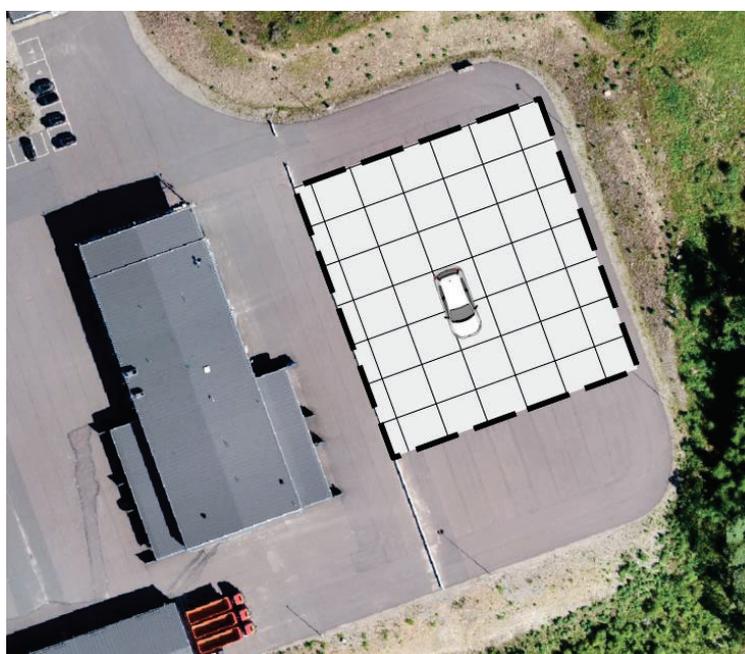
PreMAT

Predictive Monitoring for Autonomous Trucks

DNr: 2021-05052

Concluding report

2025-11-02



Project within **FFI - Road Safety and Automated Vehicles**

Authors: Fredrik Bergquist, Amritpal Singh, Cecilia Frank, Johan Florbäck,
Samuel Scheidegger, Samuel Thorén, Sebastian Lindholm, Yuri Tarakanov,
Partners: AstaZero, Asymptotic, Axis, Einride Autonomous Technologies AB, NCC,
Viscando, Volvo Autonomous Solutions

1 Content

1 Content	2
2 Summary	4
3 Introduction.....	4
3.1 Background	4
3.2 Project Objectives.....	4
3.3 Project boundaries	5
4 Research Questions.....	5
5 Methodology.....	5
5.1 Project pre-requisites	5
5.2 Research tasks	6
5.3 Functional definitions	6
5.4 Safety requirements.....	7
6 Automated Monitoring System	7
7 The monitoring functions.....	8
7.1 Dynamic geofence	8
7.1.1 Offline phase	9
7.1.2 Online phase	10
7.1.3 Learning-based minimal-intervention geofence	10
7.1.4 Results	11
7.1.5 Discussion	12
7.2 Anomaly Detection	13
7.2.1 Literature review.....	13
7.2.1.1 Designing a streaming algorithm for outlier detection	14
7.2.1.2 Isolation Forest Algorithm for Streaming Data using Sliding Window	14
7.2.1.3 Online and Unsupervised Anomaly Detection of Streaming Data Using an Array of Sliding Windows and Probability Density-based Descriptors.....	15
7.2.1.4 Micro-Cluster-Based Continuous Outlier Detection	15
7.2.2 Solution Architecture and Implementation	16
7.2.2.1 Overview	17
7.2.2.2 Base Model.....	17
7.2.2.3 Accumulative Model	17
7.2.2.4 Evaluation	18
7.2.3 Results.....	18

7.2.4	Discussion	19
7.3	Pedestrian Detection	19
7.3.1	Development of sensor technology	20
7.3.1.1	Real-time detection of presence and movements of pedestrians	20
7.3.1.1.1	Sensor system overview.....	20
7.3.1.1.2	Sensor installation at AstaZero FLXZone.....	21
7.3.1.1.3	Coordinate system alignment	21
7.3.1.1.4	Data visualization	22
7.3.1.1.5	Installation of 3D&AI sensors and data collection at Generic Site	22
7.3.1.1.6	Ground truth-less methodologies for improvement of tracking.....	23
7.3.1.1.7	Data collection in public road traffic for improvement of risk detection... 23	
7.3.1.1.8	Viscando tools for ground truth-less analysis of object tracking	25
7.3.1.1.9	Conflict Analysis tool.....	25
7.3.1.1.10	Traffic Analytics Tool.....	26
7.3.1.1.11	Evaluation and improvement of pedestrian tracking algorithms.....	27
7.3.2	Result from the demonstration of Pedestrian detection.....	28
8	Safety requirements.....	29
8.1	Safety Analysis of AMS	29
8.1.1	Dynamic geofence	30
8.1.1.1	Dynamic geofence fault tolerance time interval.....	30
8.2	Pedestrian detection.....	31
8.2.1.1	Pedestrian detection fault tolerance time interval	32
8.3	System level safety requirements	32
8.3.1	Safety architecture	33
9	Results	34
9.1	Dynamic geofence	34
9.2	Anomaly detection.....	34
9.3	Pedestrian detection.....	34
9.4	Conclusion.....	35
10	Recommendations for future steps	35
10.1	Dynamic geofence	35
10.2	Anomaly detection.....	36
10.3	Pedestrian detection.....	36
11	References.....	36

2 Summary

The project objective was to investigate how the total risk level during testing of automated vehicles under development can be reduced, without affecting the vehicle's own system.

Key research questions:

- How can we guarantee that an autonomous vehicle never exits a predefined physical space?
- How can we prevent that a pedestrian is hit by a test vehicle in the test area?
- What are the requirements on a surveillance system to reach an adequate safety level?

A successful outcome of the research would be an outline for an automated surveillance system that enables testing of automated vehicles in early development phases as well as high testing uptime being less dependent on human supervision.

Three surveillance functions were investigated:

- Dynamic geofence
- Pedestrian detection
- Anomaly detection

The first two were proven as concepts, including a safety requirement analysis. Within the Dynamic geofence concept, two research papers were published. The Anomaly detection did not reach a proof of concept but indicated a good potential to do so.

3 Introduction

3.1 Background

The ability of driverless systems is continuously expanding. The number of complex situations that the vehicles can manage increases steadily, as well as the driving velocity.

This growing capability also drives an increased need for testing. As the testing is conducted on systems under development, these may not yet have full and reliable functionality. The general countermeasure is to use human supervision as a safeguard, to intervene if something goes wrong.

But as the operational design domain for driverless systems extends, the human ability to act as a guarantee that the system behaves safely decreases. Testing of automated systems even at 50 km/h carries a high risk, despite human monitoring. In some cases, the vehicles are even developed without a driver cab, making human supervision onboard the vehicle impossible.

Consequently, a reliable supervising safety system which is independent of the vehicle, would be most valuable for testing of automated vehicles.

3.2 Project Objectives

The project objective was to investigate how the total risk level during testing of automated vehicles under development can be reduced, without affecting the vehicle's own system.

Such system would enable safer testing at both high speed and complex situations, as well as driverless testing. It would also have the potential to increase testing uptime and allow around-the-clock testing.

The central question for the project was: “How can we guarantee that an autonomous vehicle never exits a predefined physical space?”.

3.3 Project boundaries

The intention of the planned supervising system, the *Automated Monitoring System* (AMS), is to be independent of the test vehicle. However, for the system to fulfil its purpose, it needs to be able to stop the vehicle. Hence, the responsibility of the system is limited to send a stop signal, receivable by the vehicle, when judged needed. The test vehicle would then be responsible for acting on that signal.

The target for the investigated surveillance system is to operate in confined areas where the surrounding environment can be controlled or monitored, such as truck loading areas, quarries or test tracks. Hence, requirements for applying the system in other areas, e.g. public roads, are not considered.

4 Research Questions

The research questions in focus for the project were:

- How can we guarantee that an autonomous vehicle never exits a predefined physical space?
- How to prevent that a pedestrian is hit by a test vehicle in the test area?
- What are the requirements on the data feeding to the system to reach an adequate safety level?
- What is the level of reliability achieved by the backward reachable set method for verifying that a vehicle can reach a safe stop from its current state?
- What is the effectiveness and limitations of data-driven anomaly detection for safety in semi-controlled environments?

5 Methodology

5.1 Project pre-requisites

The purpose of the project was to investigate means to lower the risk in testing of automated vehicles. The project was defined with some pre-requisites, assumed to be good starting points and relevant to the research questions at hand. These were, in turn, based on two main risks identified by the project partners representing the industry:

- Risk that the test vehicle would drive outside the test area.
- Risk of injuring pedestrians inside the test area.

Thus, the project wanted to investigate three selected monitoring functions: two functions to prevent the vehicle to exit the test area, and one to prevent injuries on pedestrians. Each function would result in an emergency stop of the vehicle.

- Backward reachable set method: a “dynamic” version of geofencing.
- Anomaly detection – using machine learning algorithms to identify unexpected and potentially hazardous vehicle movements.
- Pedestrian detection using sensors.

The project also envisioned to prototype an *Automated monitoring system (AMS)*, on which the selected surveillance functions would be realized and evaluated. The AMS should be stand-alone and independent of the monitored vehicle.

5.2 Research tasks

The three functions were to be investigated through two different approaches. The Dynamic geofence and the Anomaly detection was novelty applications of partly existing concepts and algorithms. In these cases, thorough literature studies were conducted to build knowledge about opportunities and limitations of each method. Following, prototype implementations and desktop tests, before evaluation tests would be done on the prototype system *AMS*, at the AstaZero test track.

The third function, *Pedestrian detection*, was based on existing technology which was intended to be tested and improved in the specific test track environment. By making use of the everyday ongoing activities on the test track, the sensors would be exposed to situations very authentic to the intended surveillance environment of the *AMS*.

5.3 Functional definitions

As a guide for the work, the classical V-model was used, visualized in figure 1. The model was not intended to be used to a full extent, but as a mean to structure the work.

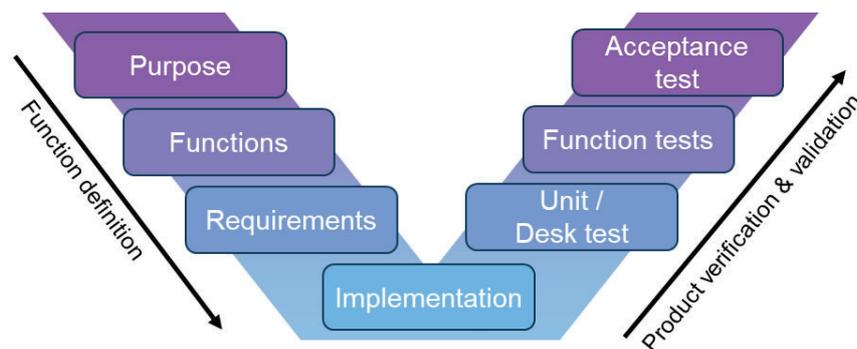


Figure 1. V-model basic principles

In the general practice all functions are defined, or described, at each level with different granularity; the overall *Purpose* is broken down step by step to the requirements needed for the function implementation. In this case, as displayed in figure 2, only the top levels were known, and the bottom levels was the subject of the project research work.

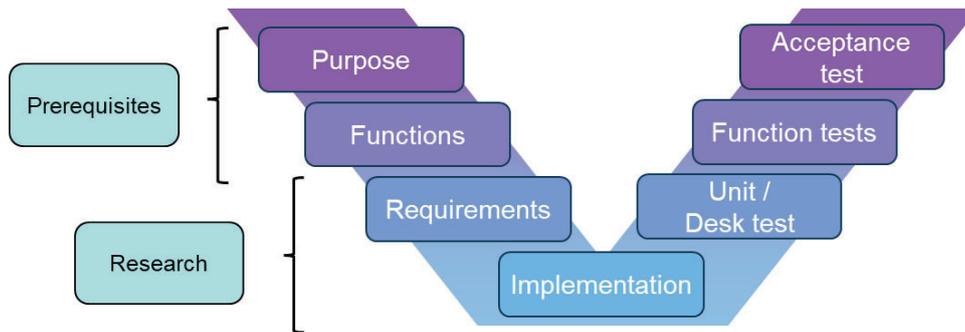


Figure 2. V-model as applied in the project

5.4 Safety requirements

The Automated monitoring system is intended to provide an increased level of safety. As part of the project, an analysis was made to establish what conditions needs to be met for the system functions to fulfil their purposes. The analysis was conducted through established methods advised by applicable standards.

6 Automated Monitoring System

Due to the planned functions, the prototype system needed to include components both onboard and offboard a test vehicle. Hence the design targeted a distributed and scalable system that could receive multiple inputs in a generic way, both wired and over the air. Designing the system this way would enable parallel development and evaluation of each function separately.

A RUT 955 modem was selected as a central computing unit. It would be located on the vehicle, providing both computing capacity, location (GNSS) and connectivity in one unit. Adapters were created to enable the communication between the external units and the modem.

The only system output would be an emergency stop signal to a vehicle interface; E-stop.

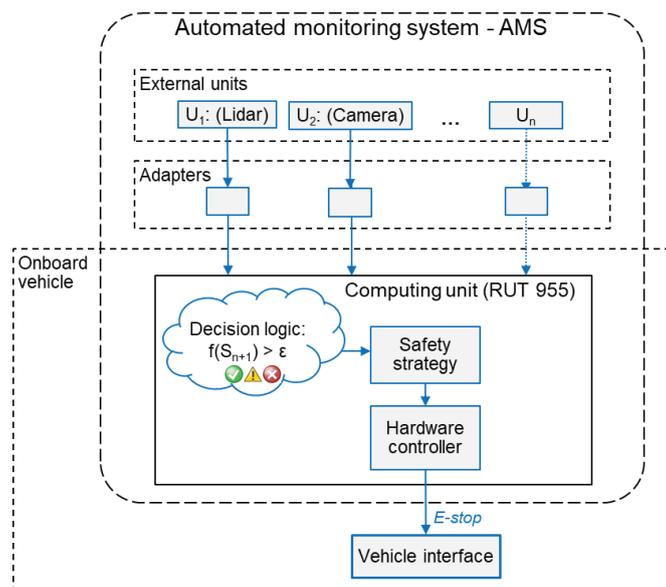


Figure 3 Automated monitoring system

7 The monitoring functions

The Automated monitoring system was defined to fulfil two purposes:

- Prevent a test vehicle from exiting a defined area.
- Protect pedestrians within a defined area.

The monitoring functions which would satisfy the purposes was:

- Dynamic geofence
- Anomaly detection
- Pedestrian detection

Following the pre-requisites, the purpose and functions are mapped into the V-model in figure 4. The project work for each function is further described in the following chapters.

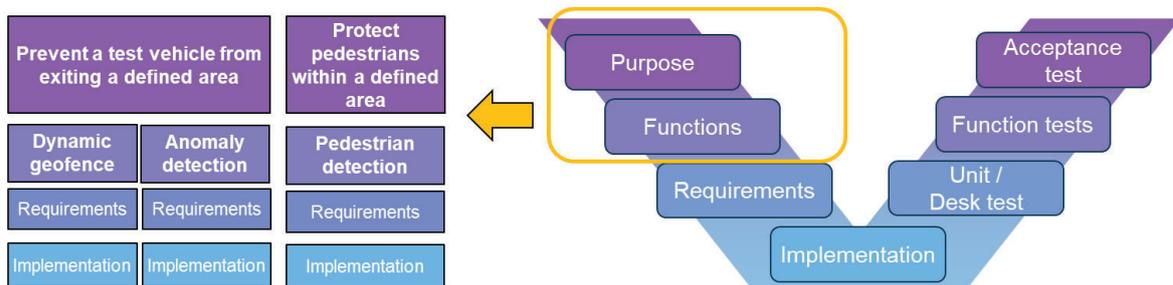


Figure 4. Purpose and functions mapped on the V-model

7.1 Dynamic geofence

Static geofencing is often used in autonomous vehicle testing but has the drawback that the vehicle does not stop until the geofence is crossed. In order to always keep a vehicle inside necessary borders, the geofence can be made smaller than the accessible test surface. But by doing so, the available test surface would not be fully used.

The project wanted to investigate if a dynamic geofence, based on a Backwards reachable set method, could fulfil the purpose of *Preventing a test vehicle from exiting a defined area* while making the whole area available for testing.

As opposed to only using the position as in the case for the static geofence, the dynamic geofence also utilises the velocity and orientation of the vehicle, allowing it to brake proactively before crossing the geofence.

The dynamic geofence is built using reachability analysis, more specifically *backward reachable tubes (BRT)*, and consists of an *offline phase* and an *online phase*.

7.1.1 Offline phase

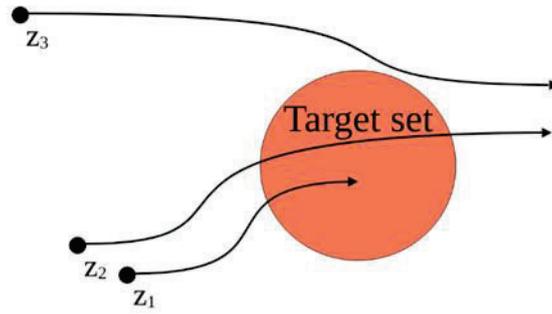


Figure 5. Three initial states z with their respective trajectories during time $[t, 0]$.

Figure 5 shows an example of a vehicle at three different states z , with the respective trajectories during time $[t, 0]$. If a state reaches the target set at any time in $[t, 0]$, the state belongs to the BRT. In figure 5, states z_1 and z_2 belong to the BRT, but not state z_3 . This means that if a vehicle is at either z_1 or z_2 , then the vehicle will exit the *safe set*, and the vehicle must therefore engage the emergency brake preventing an exit from the safe set.

When creating a dynamic geofence, we are interested in calculating the BRT for all the possible vehicle states. For the PreMAT project we chose the vehicle state as $z = (x, y, v, \theta)$, where x and y are the position, v is the velocity, and θ is the heading of the vehicle. The vehicle dynamics were then described as in equation 1, where a is the acceleration, ω is the turn rate, and U is the set of control inputs. Note that one may use other vehicle states and vehicle dynamic models, however the time complexity increases with the number of dimensions for when computing the BRT.

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{v} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \cos \theta \\ v \sin \theta \\ a \\ \omega \end{bmatrix}, \quad a, \omega \in \mathcal{U}$$

Equation 1. Vehicle dynamics equation

With the help of *OptimizedDP* (1), an open-source repository for reachability analysis, the BRT can be computed. In a real-world setting, there are a lot of trajectories that are possible in a timeframe $[t, 0]$ from each vehicle state z , as opposed to the example in figure 1. We are interested in the vehicle states where all the possible trajectories lead to the target set, meaning that the vehicle will exit the safe set no matter the control input when at these states. All the possible states leading to exiting the safe set are computed in the offline phase, meaning that no heavy calculations need to be done in the online phase, allowing for very fast decision making for whether to engage the emergency brake or not.

7.1.2 Online phase

The dynamic geofence was tested and evaluated in AstaZero's test vehicle, a Ford Mondeo Hybrid 2019 equipped with an Oxford Technical Solutions RT3000 v3 for measuring the vehicle state at one hundred times a second. The measured vehicle state was compared to the pre-computed BRT from the offline phase. If the measured vehicle state leads to exiting the safe set, then the emergency brake was engaged, otherwise the vehicle was allowed to continue driving. Figure 6 shows an example of what this might have looked like.

The figure shows a vehicle driving at two different velocities, 20 km/h and 30 km/h. As long as the vehicle is inside of the safe set, the green area, the vehicle is allowed to continue its driving. But as soon as the vehicle exits the safe set, it will lead to exiting the test area and the vehicle must engage the emergency brake. The green area will update in real time depending on the measured vehicle state, for instance if the velocity increases, then the green area will decrease.

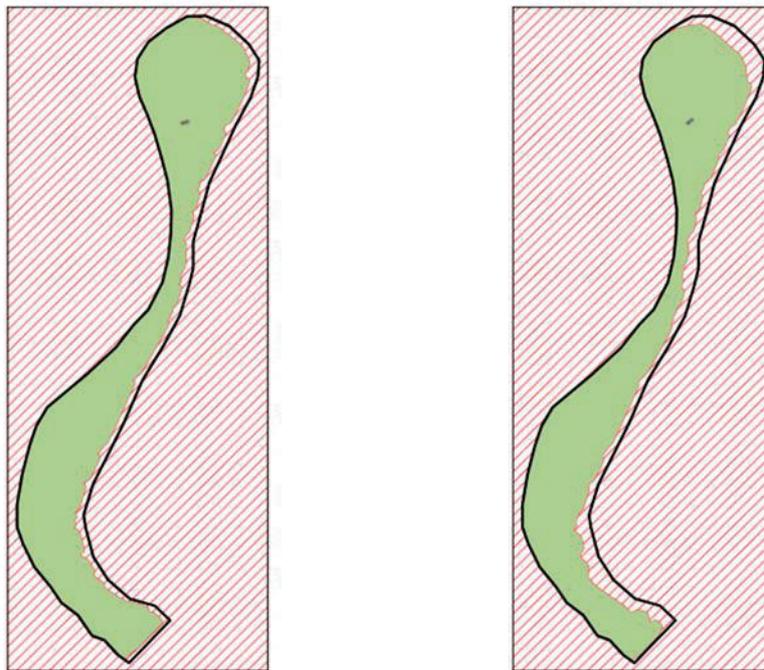


Figure 6. A vehicle with two different velocities and headings, showing a lower velocity to the left and a higher velocity to the right

7.1.3 Learning-based minimal-intervention geofence

As a complement to the emergency-stop (E-stop)-oriented dynamic geofence, we investigated an online minimal-intervention approach that aims to apply the smallest correction in steering and/or braking necessary to prevent a boundary breach, intervening only when needed. This method targets polygonal keep-in geofences and assumes short-horizon state estimation and bounded steering/braking actuation from the on-board motion controller.

Offline (model calibration). We construct a high-fidelity, physics-informed dynamics model that preserves a control-affine form, referred to as a Physics-Guided Control-Affine Residual Neural Network (PCARNN). The approach starts from an analytical bicycle model and learns structured residuals for the drift and control-effectiveness terms (e.g., tire nonlinearities,

load transfer), while keeping steering rate and longitudinal force as explicit inputs. This structure supports the linearization required by the safety controller. Calibration yields a versioned model package (physics parameters and residual weights) suitable for audit and deployment.

Online (predictive safety controller). The controller operates at the control-loop rate with a short, fixed preview horizon under zero-order-hold inputs, evaluating safety against a polygonal signed-distance barrier (SDF), positive inside. At runtime, a Discrete Control Barrier Function (DCBF) controller uses this short-horizon model preview to determine whether the current nominal inputs will keep the vehicle inside the geofence. The loop proceeds as:

1. Predict: use the PCARNN to preview the near-term state under the current inputs.
2. Check: evaluate barrier conditions on the previewed state relative to the fence.
3. Act (if needed): if a violation is predicted, solve a small quadratic program (QP) to compute a minimal correction in steering and/or braking that restores safety under actuator limits; otherwise pass the nominal inputs through. If no feasible minimal correction exists, the controller falls back to Automatic Emergency Braking (AEB).

The predictive layer proposes corrections; the on-board motion controller remains the command owner for actuation. If previews or state estimates are unavailable, the system reverts to the fail-safe policy (AEB).

This method and its evaluation, including the Containment F1 (CF1) metric introduced for minimal-intervention containment, are detailed in the companion paper, Y. Yu and S. Scheidegger, "PCARNN-DCBF: Minimal-Intervention Geofence Enforcement for Ground Vehicles" (2). Manuscript submitted for publication, IEEE Transactions on Intelligent Transportation Systems.

7.1.4 Results

The work on the dynamic geofence BRT resulted in the research paper *Model Predictive Geofencing for Vehicle Containment* which was presented at IEEE International Conference on Systems, Man, and Cybernetics (SMC) 2024 (3).

The BRT emergency stop was implemented in AMS and demonstrated at the AstaZero test track FLXZone. The concept was proven to work, with the best result at lower speed. Figure 7 shows how the vehicle has stopped just at the geofence. At higher speeds, the mitigation was more conservative with a bigger distance to the geofence. The reason was concluded to be not optimized vehicle and braking models.

This limitation in the physical tests contrasts with the project's parallel work on a minimal-intervention geofence. That work, detailed in the PCARNN-DCBF paper, was specifically tested in challenging, high-speed simulated scenarios involving both straight-line driving and sharp turns, where it demonstrated high containment performance .

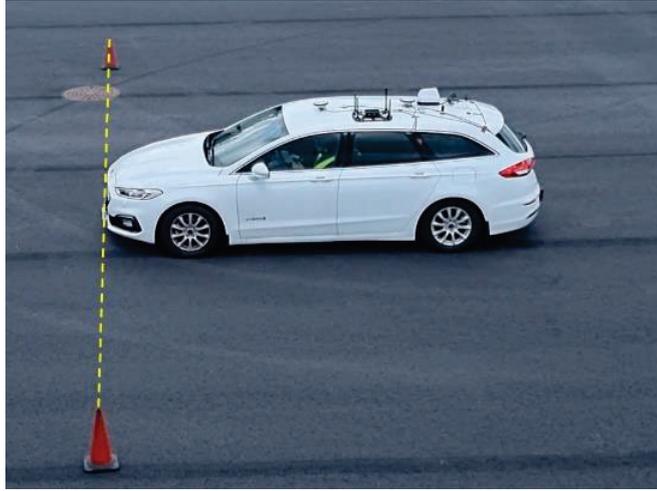


Figure 7. Demonstration of BRT at AstaZero test track

The minimal-intervention geofence work resulted in the research paper, Y. Yu and S. Scheidegger, “PCARNN–DCBF: Minimal-Intervention Geofence Enforcement for Ground Vehicles” (2). Manuscript submitted for publication, IEEE Transactions on Intelligent Transportation Systems.

This concept was tested in simulation using the PCARNN-DCBF method, achieving a Containment F1 (CF1) of 0.92–0.94 (CF1: containment precision/recall harmonic mean), with a low unnecessary-intervention rate (FPR) of 5.0–6.6% (false-positive interventions) and a positive median containment margin (MCD) of 0.35–0.47 m (median positive margin to boundary).

These results were enabled by PCARNN. This hybrid model combines the stability of a known physics-based bicycle model with a neural network that learns to correct for complex, unmodeled "residuals" (i.e., errors) like tire nonlinearities and load transfer. An Automatic Emergency Braking (AEB) fail-safe was retained for infeasible cases. Compared to a brake-only baseline, the approach reduced unnecessary hard stops while maintaining boundary invariance under the same actuation limits.

These results indicate that a minimal-intervention layer can preserve full test-area usage without shrinking fences, provided short-horizon state estimation and bounded actuation are available. Full methodology and evaluation details are provided in the companion paper (2).

7.1.5 Discussion

The BRT research paper (3) discusses more in depth about the method used, results and setup from real-world testing, and discussion about those results.

The evaluated BRT would use emergency brake as the only mitigation manoeuvre. However, this might not always be the most optimal or safe solution. For instance, braking hard when driving a heavy-duty vehicle with a trailer attached would carry a high accident risk.

For this reason, the project also explored a complementary, learning-based minimal-intervention geofence that predicts short-horizon motion and applies the smallest steering and/or braking correction needed to avoid leaving the area, with an emergency brake as fallback.

As detailed in the companion paper, this PCARNN-DCBF method is explicitly designed to be minimal by solving a small quadratic program to adjust brake and steering only when needed, thereby preserving continuity of motion rather than defaulting to a hard stop. The brake-centric dynamic geofence could use steering or progressive braking, or a combination of both, to keep the vehicle in the safe set.

The project saw an improvement potential if it would be possible to generate the safe sets in real time. One big benefit of the pre-calculated sets is that they enable a fast mitigation response during testing. This however makes the safety function static. A combined approach, where precomputed sets provide the hard safety envelope and a minimal-intervention layer offers small corrections inside that envelope, would be a valuable future improvement.

7.2 Anomaly Detection

When developing autonomous systems, the unexpected is always a risk. Foreseen error modes are assumed to be handled by dedicated systems with individually created control actions. Unforeseen incidents in hardware, software or environment may compromise components and systems in a way that reduces the ability to control the vehicle or in ways that changes the general risk profile. Worst case scenarios include complete loss of braking or steering ability of the vehicle and resulting situations such as collisions or explosions. In practice less dramatic but still potentially hazardous scenarios can be expected with increased exposure.

To reduce the overall risk, an early-warning system capable of detection abnormal behaviour in systems and subsystems, like how an experienced human driver can make observations during driving, would be potentially useful. A project pre-requisite was to investigate how an anomaly detection system could increase the safety level during automated vehicle testing.

While the main task for the project was to find means to prevent test vehicles to drive outside a test area, the anomaly detection was considered to have a potential stretching beyond that goal.

The hypothesis was that vehicles used in e.g. quarries would repeatedly drive along the same routes, day after day. Hence, the vehicle internal control signals would over time create a pattern, corresponding to those driving routes. If something unexpected would happen, it would also manifest as an unexpected event in the signal pattern.

A big challenge to manage is the diversity of anomalies that can occur but also identifying their respective root causes. What they have in common is that they need to be detected in near real-time and that they, by definition, are “new” in the sense that we can’t expect any existing definition or example of them previously occurring.

For this reason, the project collected a large amount of real testing CAN signal data from the project partners engaged in development of automated vehicles. The data was used as a base for evaluating and hopefully finding an algorithm that could detect unexpected events.

7.2.1 Literature review

For this part of the project a literature review was performed to identify leading unsupervised anomaly-detection algorithms specifically designed for streaming contexts, where data continuously arrive and models must both bound resource usage and adapt to evolving

distributions. In streaming scenarios, a sliding-window framework is often adopted: only the most recent N observations are retained in memory, and each new point triggers an update of the anomaly-scoring model. To write a robust anomaly detection algorithm, the following properties of time-series data needs to be handled:

1. Masquerading:
 - a. Outliers may hide within existing distributions
 - b. Outlier may be hard to identify
2. Uncertainty
 - a. Data is vulnerable to external events (noise)
3. Concept Drift
 - a. Distribution of data points is not fixed, and may change over time

We focus on four papers that tries to address the following, yet different in their definition of “anomaly” and their computational trade-offs.

7.2.1.1 *Designing a streaming algorithm for outlier detection*

Yu, Shi and Santoro (2019) propose an algorithm (4) which applies a sliding-window of size W together with a kernel function to maintain an online density estimate for the most recent data. Rather than re-computing the Kernel Density Estimate (KDE) from scratch, this method uses a binned statistical summary via fixed histograms, along with a retrospective correction step and a forgetting factor to adapt to concept drift while bounding memory and compute.

Because it blends KDE with summarization, the algorithm requires setting (a) the window length W , which trades off drift sensitivity versus stability, (b) the kernel bandwidth h , which controls smoothing, (c) the number of histogram bins B for statistical summaries, and (d) the forgetting factor α , which down-weights older windows.

To tackle the masquerading problem, the authors extend the classic *Local Outlier Factor* (LOF) (5) into a streaming variant. This algorithm maintains incremental summaries of k -nearest neighbour distances within the sliding window and applies a forgetting factor and statistical summaries to update local density ratios in real time. By preserving only summary statistics rather than full distance matrices the algorithm delivers timely LOF scores.

The algorithm requires a few user-defined parameters, namely: the sliding window size W , the neighbour count k , the forgetting factor α , so that older densities fade appropriately.

7.2.1.2 *Isolation Forest Algorithm for Streaming Data using Sliding Window*

Ding and Fei (2013) propose an online variant of the Isolation Forest (iForest) tailored for sliding-window streams (6). At each time t , the algorithm maintains a fixed-size window W_t of the N most recent points. Upon arrival of a new observation x_t , they insert x_t , evict the oldest point, and then draw multiple small sub-samples from W_t to grow a forest of isolation trees. Unlike the normal, batch version, of iForest, whose sub-samples come from the full dataset once, the proposed algorithm repeatedly resamples the sliding window, ensuring that model structure reflects the most current distribution. Anomaly scores are computed as the average path length $h(x_t)$ across all trees: shorter paths (i.e. fewer splits needed to isolate x_t) indicate higher likelihood of being anomalous.

To balance responsiveness to concept drift against statistical stability, the algorithm relies entirely on the natural forgetting induced by the sliding window; there is no separate drift-detector. Consequently, the window size N is the key tuning parameter: too small a window

yields noisy, jumpy models, while too large cause the algorithm to react sluggishly to distributional shifts. In addition, the algorithm require that users must choose the number of trees M and the sub-sample size Ψ . A small Ψ speeds up updates but produces higher-variance scores, while a large Ψ smooths out noise at the cost of slower model rebuilding.

7.2.1.3 *Online and Unsupervised Anomaly Detection of Streaming Data Using an Array of Sliding Windows and Probability Density-based Descriptors*

Zhang et al. (2021) introduce an approach based on Probability Density-based Descriptors (PDDs) that augments sliding-window schemes with explicit density tracking (7). They first partition each main window of size W into L non-overlapping sub-windows. Within each sub-window, a multivariate Kernel Density Estimate (KDE) captures the local probability density, which is then projected via Principal Component Analysis (PCA) into a low-dimensional descriptor vector d_i . Stacking these L vectors yields a descriptor sequence representing the window's evolving density structure.

For anomaly detection, the method computes the distance between the latest descriptor and a running reference (e.g., the mean of the previous K descriptors). If this distance exceeds a threshold derived from historical variation, the current observation is flagged as an outlier. By recalculating KDE only on fixed-size sub-windows and summarizing via PCA, PDDs keep per-point cost down to while providing explicit sensitivity to gradual density shifts—contextual drifts that mere window-sliding might blur.

The algorithm requires a few user-defined parameters, namely: the number of sub-windows L , the KDE bandwidth h , the PCA output dimension r , and the length K of the descriptor buffer. A small L or coarse bandwidth h may miss fine-grained shifts, while overly fine settings overreact to noise.

7.2.1.4 *Micro-Cluster-Based Continuous Outlier Detection*

Kontaki et al. (2011) propose a sliding-window (8), distance-based detector in which the algorithm first carves the most recent N observations into dense “micro-clusters” of radius $R/2$. Each micro-cluster holds at least $k+1$ points, so every member automatically has k neighbours within distance R and can be declared an inlier without further checks. Points that aren't yet absorbed into any cluster form the “unsafe set,” and only these candidates require explicit neighbour counting to determine outlier status.

As the data stream advances, the algorithm maintains these structures dynamically. When a new observation arrives, it is immediately absorbed into any existing micro-cluster whose centre lies within $R/2$; otherwise, it joins the unsafe set. Conversely, if an unsafe point later accumulates $k+1$ neighbours within distance R , it and its neighbours are promoted into a new micro-cluster. Likewise, as the window slides forward and old points expire, any cluster that falls below $k+1$ members dissolves, returning its remaining points to the unsafe set. This continuous promotion and dissolution process ensures that only a small boundary of points ever incurs the cost of full neighbour searches, while dense regions remain safely tucked away in clusters.

The algorithm requires the users to set the radius R and neighbour threshold k . A small R can leave most points “unsafe,” while a large R risks masking tight, localized anomalies.

Likewise, setting k too high makes the detector overly strict, catching only the most extreme outliers; too low and it floods with false alarms.

7.2.2 Solution Architecture and Implementation

Due to the broad diversity of anomalies, the data-driven algorithm architecture was chosen to be designed in a way that provides flexibility and options in model selection and parametrization. The proposed architecture is composed of four principal components: a pre-processing module, a base model, an accumulative model, and an evaluation framework, see Figure 8.

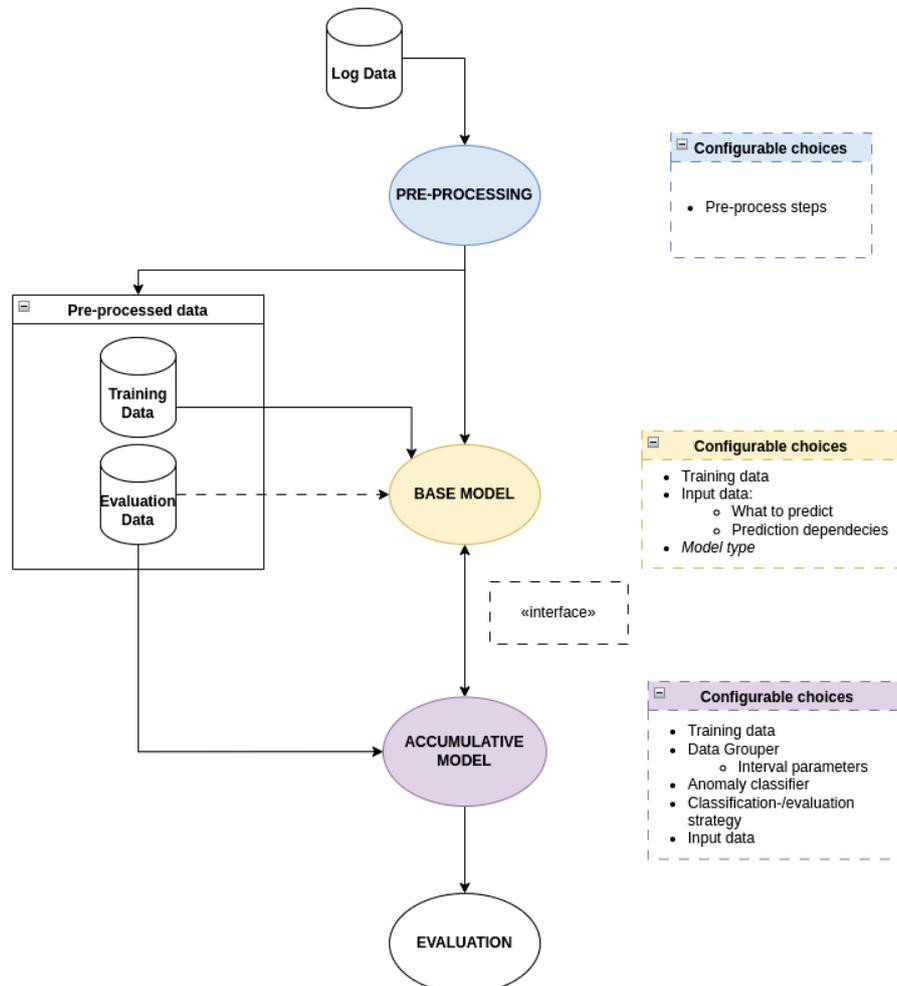


Figure 8. Solution Architecture

This architecture allows for combining of various implementations of each individual component (including the algorithms from the literature study, but also endless more) in a way that enables accurate and comparable detection of very different types of anomalies. This, in turn, allows for multiple anomaly detection algorithms to run in parallel, sharing common elements, and finally report a combined list of detections ordered by estimated severity.

This section will account for the core components, namely the base model, the accumulative model, and the evaluation component, providing a description of their functionality and interconnection within the overall architecture as well as a short description of what was implemented within the project.

7.2.2.1 Overview

The overall approach involves a model predicting the probability distributions of input data and a model detecting anomalies. During anomaly detection, the predicted probabilities are aggregated into distinct regions, each of which is then evaluated by a suitable anomaly classification algorithm.

7.2.2.2 Base Model

The base model serves as the predictive component within the overall architecture and is formulated as a supervised learning model. It outputs probability distributions, enabling the representation of the uncertainty in its predictions. The probabilistic framework facilitates the integration of the subsequent accumulative model, which leverages the probability estimates for classification of anomalies. The model architecture is designed with a modular structure, where the characteristics of the dataset and problem definition may determine what architecture to use.

The base models implemented were all chosen to be either a normal distribution fitted to (single) raw values, or as statistical (auto-) regression models coupled with a normal distribution fitted to respective residuals. Depending on anomaly detection configuration different input- / output- variables were manually chosen accordingly, for auto-regressive cases signal value from previous timestamp and delta time was included. Accuracy of these models (and variable selections) was evaluated using adjusted R-squared and negative log likelihood. This setup represents perhaps the simplest and most well-understood statistical modelling available and was chosen as an intuitive baseline for gaining insight in the connected end product.

7.2.2.3 Accumulative Model

The accumulative model comprises three principal sub-components: a data-grouping functionality, a comparison strategy and an anomaly classifier. The data-group strategy can be configured in diverse ways, each with its respective parameters. The comparison strategy mainly describes what a group should be compared against, baseline for short. And the anomaly classifier can represent various anomaly classification algorithms all determining if the group is anomalous compared to the selected baseline and each presenting how anomalous they are in form of an “anomaly score” that can be compared between configurations. Given this flexibility, the total number of potential configurations is considerable resulting in a wide adaptability. The appropriate combination of these building blocks is determined by the type of the anomalies of interest and the operational definition of “anomaly” within the given context.

For the scope of the project two main data-grouping methods were implemented: Time and Location. Time grouping means that windows of time with variable length is used to assign each measurement to a group. Location grouping means that measurements that fall within (squarely) binned areas are grouped. Comparison strategies were chosen to be either comparing each group to all other groups or to compare each group only to the corresponding group in the training set of the base models where the second strategy is only reasonable to apply to location grouping.

In addition, three different anomaly classifiers were implemented and evaluated. The first one is based on the chi-squared test, commonly used to compare distributions, where a

lower estimated probability of the two sets (evaluation group and baseline) coming from the same original distribution is directly proportional to the inverse of the anomaly score. The second classifier is based on quantile values and selects a number of quantile points in the baseline and evaluates the difference between the fraction of points that fall between them in the baseline set and the fraction in the group.

The average size of these differences is proportional to the reported anomaly score. The final classifier evaluates overall probability (summed negative log likelihood) in the group and the baseline respectively and the relative probability is converted into an anomaly score.

7.2.2.4 Evaluation

The evaluation component constitutes the final stage of the architecture and is responsible for assessing the overall performance and validity of the proposed anomaly detection framework. By employing appropriate performance metrics, this component evaluates the integrated output of the base model and accumulative model. This stage also provides the basis for model refinement and thus iterative improvements.

Each anomaly detector was subjected to the evaluation procedure, which was conducted by three main steps: labelling, metric computation, and finally performance and visualisation scoring.

As for the labelling, each grouped frame generated by the accumulative model was assigned a binary label indicating whether it represented an anomaly or a normal group. A group was labelled anomalous if it contained at least one detected anomaly instance, and as normal otherwise. Thereafter the True Positive Rate (TPR) and False Positive Rate (FPR) were calculated for each distinct anomaly scoring threshold. These values were then presented and visualised in the form of a Receiver Operator Characteristic (ROC) Curve together with the corresponding Area Under the Curve (AUC) metric. The AUC provides a single scalar measure, where a value of 0.5 indicates performance to random guessing, while an AUC of 1.0 denotes perfect classification performance.

In addition to the ROC Curve and AUC, several supplementary metrics including Classification Accuracy, Classification Error, F1 Score, and Youden's J statistic were computed and evaluated to provide further assessment of the anomaly detector's effectiveness. These additional measures offer complementary perspectives on the trade-offs between sensitivity, specificity, and overall predictive performance.

7.2.3 Results

Supported by the project partner's, Volvo Autonomous Solutions and Einride, vehicle tests, a large amount of vehicle signal data could be collected. This was also done together with partner NCC. This data could in turn be used as real-life vehicle data sets to evaluate and test different approaches of anomaly detection algorithm combinations.

The biggest challenge was to tune the algorithm parameters to achieve a reasonable level of indication, i.e. where neither false negatives nor false positives was too dominant. The second challenge was to find "correct" anomalies. It turned out to be easy to identify deviating patterns, but more difficult to establish confidence that the anomaly was originating from an interesting event. The assumed cause for this was that when the available

data was recorded, the occurring anomalies were of such small magnitude that detection was not possible to do with certainty.

Even though it was not possible to conclude that the anomaly detection system would be the confident tool, the research showed interesting results. The most significant finding was that such a system would require not just one model, but rather a suite of models with different configurations to adequately capture different forms of anomalous behaviour within the data. Furthermore, as the system did find anomalies, even though small ones with low confidence, this indicated that there is a potential that can be developed. However, a robust anomaly detection almost certainly requires insightful human supervision and expertise during development and evaluation to ensure both high enough quality in accuracy and precision, and trust and understanding in the capability of the system.

7.2.4 Discussion

The proposed approach resulted in a flexible solution architecture that enables multiple complementary models and tailored parameterizations. An important insight from the obtained result is the importance of flexibility in the structure.

As mentioned above, the notion of an anomaly is inherently context-dependent and may be defined in multiple ways depending on the specific operational scenario and detection objectives. Consequently, no single model configuration suffices, rather a suite of models with varying parameterizations would be required to adequately capture different forms of anomalous behaviour within the data.

This flexibility allows for future scenarios where a supervisor of the autonomous vehicles dynamically updates and adjusts the suite of anomaly detectors until a satisfactory level of safety is achieved.

7.3 Pedestrian Detection

During the development phase of automated vehicles, the vehicle ability to detect humans or other obstacles may be limited or even non-existing. This entails a risk for severe injury of persons who could enter an area where such vehicles are being tested.

A common practice to prevent accidents is to use physical barriers and working routines to keep pedestrians outside the test area. In real life however, mistakes are made, and persons may sometimes enter restricted zones. It is also not always feasible to arrange completely fenced areas to keep humans on the outside.

For this reason, the project wanted to investigate how a sensor-based system can be used as a safeguard at such test areas.

To enable a physical proof of concept of the *Pedestrian detection* functionality, the project integrated the partner's sensors to the AMS. The sensors would be mounted on fixed positions at the AstaZero test track FLXZone, where they would cover an area considered to be a forbidden zone for humans. The sensors provided input to the Automated monitoring system about human presence and sent an alert signal when a person entered the forbidden zone. The AMS would in turn send an emergency stop command to the test vehicle, to mitigate that a person is located within the test area.

Having the sensors installed at the test site, it also provided a valuable environment for the sensor developing partners (Axis and Viscando) to evaluate the functionality in the same

environment as was intended for AMS, as well as providing development improvement opportunities of the sensors.

7.3.1 Development of sensor technology

7.3.1.1 *Real-time detection of presence and movements of pedestrians*

7.3.1.1.1 Sensor system overview

To ensure safety and security for both test vehicles and personnel during tests on test tracks and autonomous operations in confined areas, it is imperative to detect risky movements of vulnerable road users. Presence and movements of pedestrians in test/production areas may lead to close encounters and accidents – and in many cases require immediate stop of tests of production activities with autonomous vehicles.

While it is relatively simple to just detect the presence of pedestrians in the area using cameras and AI, accurate movement tracking and prediction are required to distinguish between the normal movements of pedestrians within “allowed” areas and risky situations, like movement towards the test area during active testing.

In PREMAT project, two test areas were in scope. In the first one, at FLXZone test track, the test control room was located next to the test track. A walking path around the control room was separated from the test track by the barrier. Movements of personnel within the walking path should not be considered as risky as persons on the test track, or on their way to go around or jump over the barrier.

Personal integrity, GDPR and secrecy requirements need to be fulfilled when the sensors record the environment where both workers and pre-production vehicles are present. This puts strict limitations on processing of video data. As regulations forbid storage or transmission of video data, the processing must be performed immediately on the device.

Viscando is a start-up specializing on measurements, analytics and prediction of road user movements in a broad range traffic environments, spanning from public roads to yard management and factory intralogistics. Viscando’s own developed 3D&AI based sensors detect, classify and track all kinds of road users. The sensors are based on stereovision principles, where two spatially separated cameras take synchronous images and produce the so-called depth image from these. The depth image is similar to Lidar point cloud with colour information in every detected point. This point cloud is segmented to separate traffic actors, which are consequently classified as vehicles, trucks, pedestrians, cyclists and other road users using both visual appearance and 3D shape. Advanced tracking algorithms associate consecutive detections to individual actors, calculating their movement dynamics and predicting their future movements.

Viscando is actively pursuing innovative safety applications based on real-time knowledge and predictions of movement dynamics of interacting road users. Examples of such applications are accident risk detection, adaptive traffic lights, dynamic warnings, and infrastructure support to connected vehicles [ref: FFI projects *Future 5G Ride* (9), *I2Connect* (10)].

Viscando's deployment of sensors and movement tracking/conflict detection at FLXZone and Generic Site areas at AstaZero had two-fold goals:

1. To contribute to test supervision system by providing presence and movement data of pedestrians and vehicles on and around the test track in real time
2. To better understand the role of infrastructure sensors and data requirements in testing and operations of autonomous vehicles in confined areas

7.3.1.1.2 Sensor installation at AstaZero FLXZone

Viscando deployed two OTUS3D sensors at FLXZone test area as shown on figure 9.

The sensor covered two sides and a corner of the control room facing the intersection area of the FLXZone track, where tests by customers of AstaZero were performed.

The installation was done on 2022-11-22, and the data was collected during the following time intervals:

- between 2022-11-23 and 2023-12-18
- between 2024-06-03 and 2024-06-05
- between 2024-06-17 and 2024-09-05



Figure 9. Sensor installation at AstaZero FLXZone

Figure 9: Installed Viscando sensors (A, C) and corresponding sensor views (B, D) at FLXZone test area. The barriers visible in the sensor views were set up after sensor installation. Sensor view images were taken during the sensor calibration at a time when no pedestrians or vehicles were present in the field of view, with a written permission from AstaZero. Inset in (A) shows the installed sensor positions on the bird-view image of the control room.

7.3.1.1.3 Coordinate system alignment

The coordinate systems of Viscando sensors were aligned with the internal coordinate system at FLXZone used by AstaZero. This was done by matching the positions of several well visible points on the test track (such as edges of road markings) in Viscando sensor coordinate systems with respective positions in AstaZeros coordinate system. In this way,

the positions and directions of actors transmitted by Viscando were aligned with both the area map, 3D scanned model of the area, and the positions reported by in-vehicle GNSS systems. This allowed to skip coordinate transformation steps in the supervisor application.

7.3.1.1.4 Data visualization

The movement data was visualized in the Track Analytics Tool developed by Viscando within PreMAT project. The tool is described in detail in section 7.3.1.1.10. Two examples of tracks of pedestrians and vehicles are shown in figure 10.

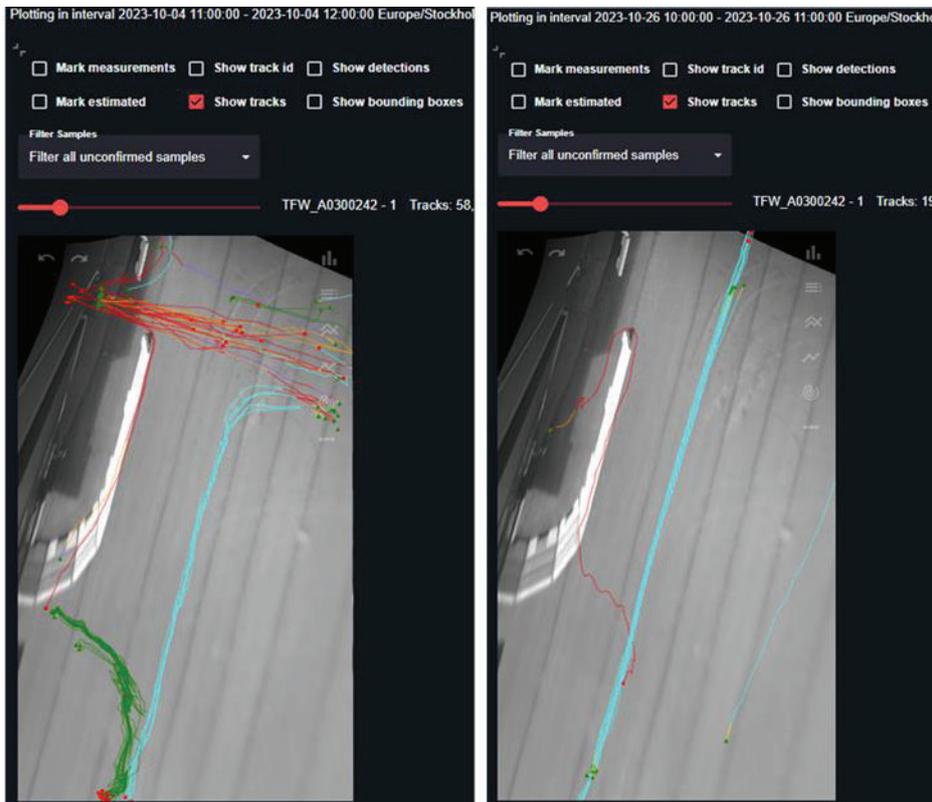


Figure 10. Movement data visualization

Figure 10: Examples of movement trajectories visualized in Viscando Track Viewer tool. These trajectories represent all movements within 1 minute at different dates and times when the intensive testing activities were conducted. The movements of personnel (red lines), regular vehicles (cyan lines), and heavy-duty vehicles (green lines) are clearly seen.

7.3.1.1.5 Installation of 3D&AI sensors and data collection at Generic Site

While the FLXZone test area is suitable for testing of on-asphalt use cases such as public roads, yard management or ports, it is less adapted for testing offroad confined area applications such as mining. To extend the infrastructure sensor support of the test supervision to the mining sites, Viscando sensors were moved to the Generic Site, a simulated mining site at AstaZero. There Viscando sensors performed detection and tracking of the movements of pedestrians and vehicles and transmitted the movement information to the Test Supervisor in real time, in the same way as in the previous use case.

The sensors were set up on 2024-09-05 and were in use to the project end. The installation area and fields of view of installed sensors are shown in figure 11.

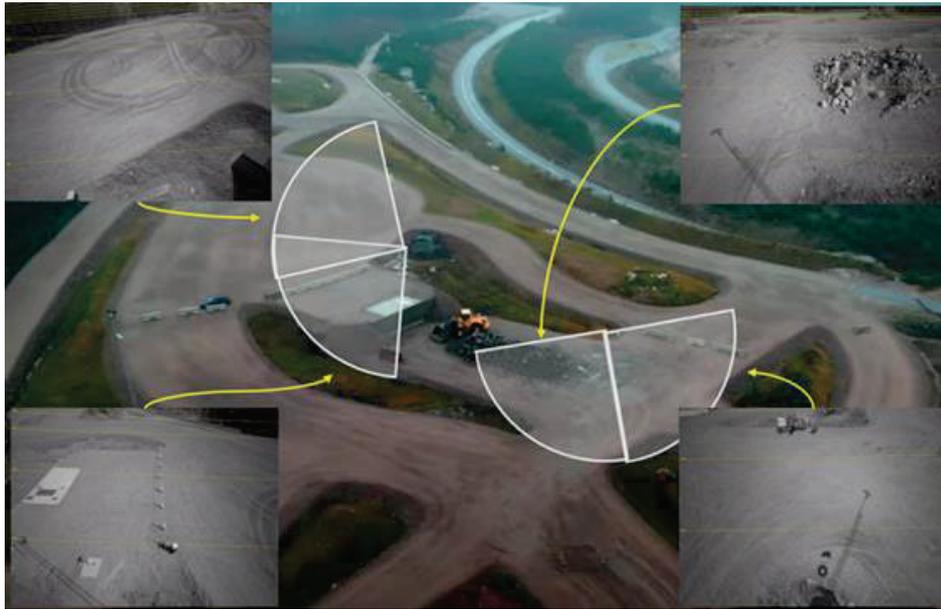


Figure 11. Installation points and fields of view for Viscando sensors installed at Generic Site area at AstaZero.

7.3.1.1.6 Ground truth-less methodologies for improvement of tracking

Reliability of sensor systems is crucial in safety-critical applications. Ideally, a large amount of accurate ground truth data (e.g. high-precision GNSS or high-resolution Lidar data) is collected and the sensor outputs are statistically compared with this ground truth. Another option is to collect large amounts of video, which is then manually annotated to yield ground truth.

Both these options were not feasible in PREMAT project, because of the following limitations:

- Video data collection was not allowed due to GDPR and secrecy regulations
- Portable positioning units for pedestrians were not available in the project
- Most of the time, there was very little or no traffic around the control room
- No risky behaviours were reported during the measurement period

Therefore, Viscando used other approach to assess and improve the pedestrian detection and tracking performance in risky traffic situations.

7.3.1.1.7 Data collection in public road traffic for improvement of risk detection

Public roads in urban areas are the environment where pedestrians and vehicle drivers interact, and where risky behaviours, rule violations and conflicts are common. To enable data-driven evaluation and improvement of pedestrian detection and tracking, Viscando collected traffic movement data in several locations which exhibited intensive traffic and interaction between pedestrians and vehicle drivers:

1. Hallituskatu street, Lappeenranta, Finland, 2023-09-04 to 2023-09-15. The measurement was performed near a school zone and public transport stop. Pedestrian interactions and conflicts with both vehicles and cyclists were in scope. Figure 12 shows the tracks of different road users in this measurement, as well as a map of detected conflicts. The number of observed pedestrians were around 312 per day.

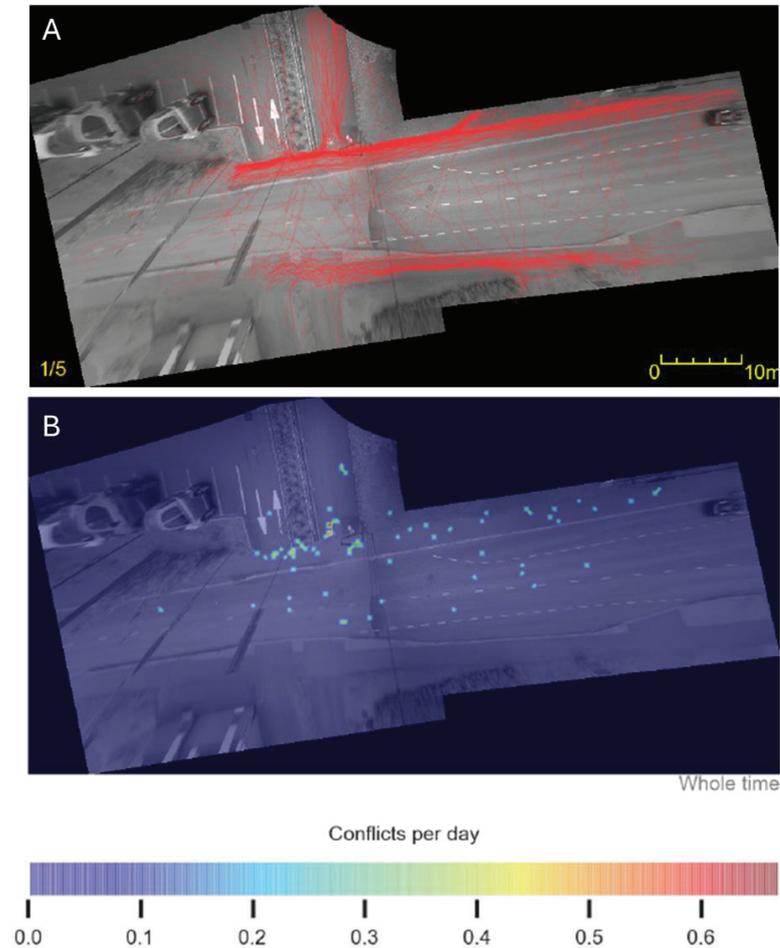


Figure 12. Pedestrian trajectories (A) and detected conflict points (B) in Viscando measurement in Lappeenranta.

2. Varberg, Sweden, Crossing Östra Hamnvägen – Södra Hamnvägen. The focus of the measurement was the movement patterns and interactions of pedestrians, cyclists and vehicles in at T-shaped intersection.
3. F11/Heuvelstraat, Boechout, Belgium, 2023-10-17 to 2023-10-19. The focus in the measurement is a road section with pedestrian/bicycle crossing close to railroad crossing.
4. Kinderwelzijnstraat, Kalmthout, Belgium, 2024-10-17 to 2024-10-19. The focus in the measurement is a road section with pedestrian/bicycle crossing close to railroad crossing.

Conflicts between pedestrians and vehicles were identified in track data using the surrogate safety metric Post-Encroachment Time (PET). Besides the road safety assessment for the benefit of road and city authorities, the conflict analysis demonstrated and improved the

ability of Viscando sensors to correctly detect the risky interactions, which is the required functionality in the PREMAT project.

7.3.1.1.8 Viscando tools for ground truth-less analysis of object tracking

Ground truth was not available when recording general traffic on public roads. However, two factors enabled analysis and improvement of the movement and risk detection:

1. possibility to record anonymized point cloud sequences, which were consequently analysed using **Conflict Analysis tool**
2. large traffic volumes allowing for statistical evaluation in **Track Analytics tool**

7.3.1.1.9 Conflict Analysis tool

Viscando has implemented a methodology to record “anonymized video”, which is in fact not a video but a sequence of sparse point clouds that preserves only the rough shape and movement of a road user. Such video allows for manual annotation of the road user types and movements but has no features that might contain any personal information. Therefore, it is compliant with GDPR.

When a conflict or other event of interest was detected in trajectory data, the anonymized point cloud could be manually inspected to confirm that (i) the detected movements were correct, and (ii) a risky situation or conflict took place. Based on the manual inspection of the conflict cases, the assessment of the movement tracking and conflict detection accuracy can be performed. Moreover, errors are analysed to identify improvement areas in the movement detection algorithms.

A *conflict Analysis tool* was designed based on this methodology. A screenshot of this tool, showing the anonymized point cloud videos of a conflict situation from different viewpoints, as well as a top-view plot of the movement trajectories, is shown in figure 13.

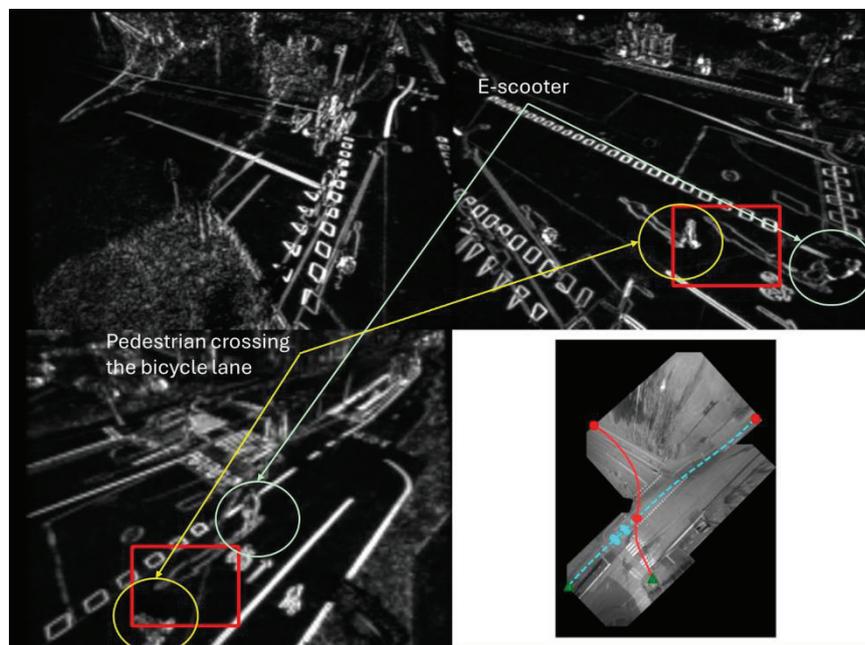


Figure 13. Screenshot from Conflict Analysis tool

Figure 13: An example of conflict in the conflict analysis tool developed by Viscando. The point cloud videos are shown from different angles, together with a top-view of the trajectories (bottom right corner). In this example, a pedestrian crosses both the road and the

bicycle lane outside of the designated pedestrian crossing and has a close encounter with an e-scooter.

7.3.1.1.10 Traffic Analytics Tool

The second tool that Viscando has developed is the Traffic Analytics Tool. This web-based, interactive tool allows loading, visualizing and analysing large datasets of traffic movement trajectories. A systematic inspection of movement patterns, speed profiles and long-term statistics in the datasets provides a comprehensive traffic analytics and insights for road authorities, municipalities and public organizations – but at the same time, allows for identifying potential deficiencies in detection and tracking of road users:

- Tracking issues such as anomalous movement patterns, tracks appearing and disappearing inside the field of view, “ghost” tracks that cannot correspond to any real objects.
- Systematic errors stemming from illumination (e.g. errors appearing at nighttime or during low sun)
- Apparent errors in object classification

The tool allows for multiple data views adapted to track data, such as

- 2D top views which can be filtered by e.g. time, object type, track ID, and track length
- Overlay with the map of the area
- 1D time-series plots of position, velocity and acceleration
- Time-statistics visualizations, for example traffic flow, distributions by track ID and other parameters.
- Side-by-side comparison of different versions of track data, which is extremely useful for evaluation of improvements in track data processing and filtering.

Figure 14 shows examples of views available in the Traffic Analytics Tool, using a dataset representing a busy intersection with different road user types.

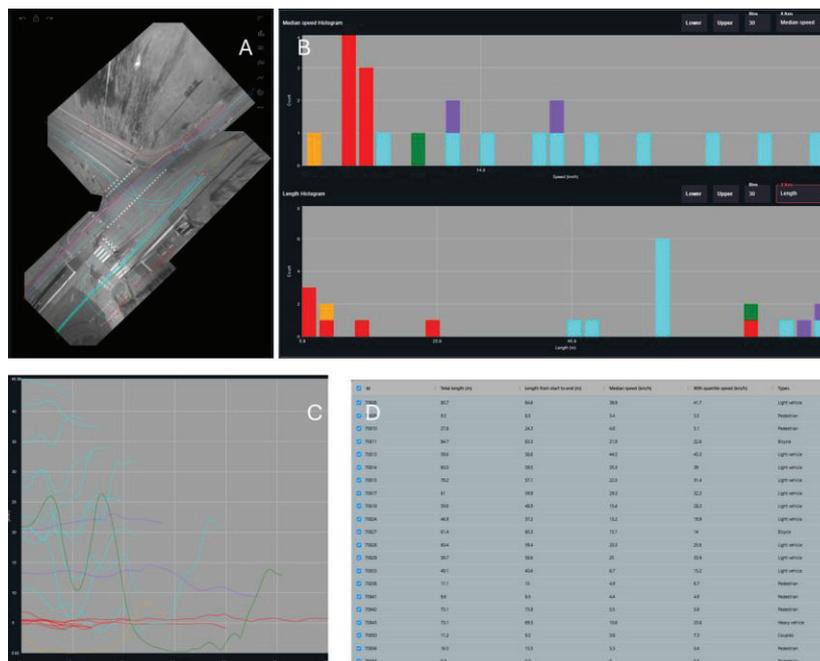


Figure 14. Examples of views available in the Track Analytics Tool

Figure 14: Examples of views available in the Track Analytics Tool: Track top-view with the map overlay (A); average speed and track length statistics (B); speed profiles (C) and track property list (D).

7.3.1.1.11 Evaluation and improvement of pedestrian tracking algorithms

Using the Conflict Analysis and Track Analytics tools with the datasets described above lead to the following achievements in PreMAT project:

1. Identifying the challenges and errors in object tracking, especially concerning pedestrian trajectories in risk and conflict situations
2. Improving the tracking and filtering algorithms to achieve higher performance and reliability of the movement and risk detection

The following list presents the main identified challenges, improvement strategies and effects on Viscando technology that has been achieved through Viscando work in PreMAT project:

1. **Identified challenge:** Pedestrian tracking stability is limited by the distance from the sensor. At distances over 25 meters from the sensor, the rate of loss of pedestrian tracks is quickly increasing, and over 25 meters, pedestrians are tracked only sporadically.

Improvement strategy: multi-sensor setups, such that the risk area is within the range of stable detection for at least one sensor.

Effects on Viscando technology: improved planning of sensor setups in use cases and locations where detection of pedestrian motion and conflicts is important.

2. **Identified challenge:** Accurate tracking of standing pedestrians is challenging. Pedestrians tend to slightly move even when standing still, which may confuse the tracker. Typically, the resulting pedestrian track is moving sporadically around the true position, showing “movement noise”. This may lead to overestimated risks when the tracker erroneously estimates that the pedestrian starts going towards the risk area due to movement noise.

Improvement strategy: Improved tracking algorithm which better estimates the low velocities and ideally detects stand-still.

Effects on Viscando technology: The tracking was modified to improve the position and velocity estimate at low velocities, which reduced the movement noise for non-moving pedestrians and reduced the false risk warnings.

3. **Identified challenge:** Pedestrians are often obscured behind common infrastructure elements, such as traffic signs or trees. This leads to interrupted pedestrian tracks, especially in cases when pedestrians change direction or speed. During the time required to re-initialize the pedestrian track, their movement is uncertain, and the potential risk cannot be reliably evaluated.

Improvement strategies: (i) enhanced detection and tracking which takes the obscuring objects and the hidden space behind them into account, and (ii) sensor placement avoiding occlusions in risk areas.

Effect on Viscando technology: (i) Detection and tracking algorithms were improved to preserve the tracks of hidden pedestrians. (ii) Planning process of measurements and sensor placement was refined to avoid the occlusions in the conflict zones.

4. **Identified challenge:** Pedestrians exiting the buildings or vehicles are detected with some delay. 3D segmentation of pedestrians in these situations was proven challenging until the pedestrians moved a sufficient distance (typically 0.5-1 meter) from the door. This delays the initialization of the pedestrian track, which, in certain situations, may lead to missed risk warning.

Improvement strategies: (i) in case of buildings, masking of the building helps to easier segment out and detect an exiting person; (ii) DNN based object detection algorithms can be utilized to detect the presence of pedestrians even when they are partially obscured.

Effect on Viscando technology: Masking methods have been introduced to improve detections of pedestrians exiting the buildings. Preliminary results show earlier and more robust detection of pedestrians, although more rigorous evaluation is required. Early detection of pedestrians exiting vehicles will be addressed in future research and development.

Overall, the tools and methods developed within PREMAT project allowed Viscando to improve pedestrian movement detection rate from estimated 75-80% to 95-98% and reduce the rate of false conflict detections from estimated 30-50% to 5-10%. The drastically decreased rate of false conflict detections allows for more robust real-time conflict warning with fewer false alarms.

It has not been possible to collect accurate reference data in relevant scenarios to make an accurate quantitative evaluation of the performance improvement and achieved level. Viscando has an ambition to set up the ground truth measurements for pedestrians at AstaZero and carry through accuracy measurements in future research collaborations.

7.3.2 Result from the demonstration of Pedestrian detection

The sensors mounted at the AstaZero test track FLXZone was integrated into the AMS system, marked in green in figure 15.

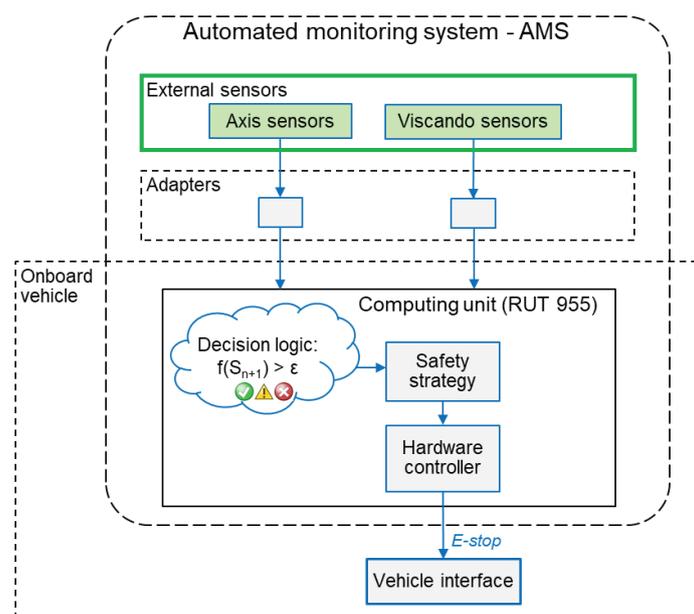


Figure 15. Sensor integration with AMS

The demonstration was successful in having the test vehicle stopped the moment a person entered the defined forbidden zone. Axis noted that some containers, placed inside the test zone, caused occasional false positive, but *only* when there were humans present in the forbidden area. Hence, no false positives causing an emergency brake event.

The demonstration was visited by representatives from the autonomous vehicle industry who showed great interest and indicated that the function would be a valuable safety asset in their operations.

8 Safety requirements

To determine if a monitoring system such as the AMS reaches a sufficient level of safety it is recommended to make an assessment according to established methods outlined in applicable safety standards. Which standards are appropriate is determined by the nature of the system, how it is intended to be used and during what conditions. Each safety system needs to be assessed according to the conditions valid in that case.

The standards are in general not trivial to follow as a beginner. Achieving a complete and satisfying analysis requires an experienced person, skilled in applying the methods. Hence, the intention of the project is to demonstrate what kind of results a safety analysis can give rather than explaining in detail how it was done.

8.1 Safety Analysis of AMS

For the system at hand, the AMS, the risk analysis conducted was based on **ISO 19014-1** *Earth moving machinery – Functional safety*. ISO 19014-1 lays out a methodology for determining the performance level required for safety related parts of control systems on earth moving machines.

The standard says that a Machine Control System Safety Analysis (MCSSA, as of **ISO 13849:2015** *Safety of machinery — Safety-related parts of control systems*) should identify hazards and determine how much risk reduction is needed. This reduction is expressed as a Machine Performance Level (MPL), which is calculated using a risk assessment as defined in **ISO 12100**. ISO 19014-1 therefore describes how to:

- identify hazards associated with earth moving machinery control systems
- determine the required performance level of safety control systems to mitigate those hazards
- classify safety related parts of the control system (e.g., sensors and actuators).

To assess the necessary AMS safety measures and response times for a vehicle operating in environments with potential pedestrian interactions (confined area in the context of dynamic geofencing and pedestrian detection functions), analysis is conducted to understand the relationship between:

- Vehicle speed
- Braking distance (estimated)
- Road surface (via friction coefficient)
- Pedestrian movement
- Reaction and stopping time

The analysis focuses particularly on low-speed conditions dictated by application use-case (≤ 40 km/h). This can be further extended to moderate and high speeds.

8.1.1 Dynamic geofence

The dynamic geofence BRT is based on a chain of events, displayed in **Fel! Hittar inte referenskälla.16**, starting with detecting the vehicle position, speed and driving direction.

In the prototype system, GNSS/RTK was used as the positioning source. The system accuracy can be quantified using the following formula:

$$\text{Dynamic Geofence accuracy} = \text{GNSS / RTK Accuracy} + \text{Propagation Delay} \times \text{Vehicle Speed (Distance Travelled Before Decision Execution)}$$

Taking into consideration the worst-case scenario, a detection time of 20ms have been calculated. Thus, at a maximum vehicle speed of 40km/h (11,11 m/s) it will result in:

$$0.002 * 11,11 \text{ m/s} = 0.13 \text{ m} = 13 \text{ cm}$$

Adding the accuracy of the position system will result in an allowed tolerance of **20 cm** as worst case scenario.

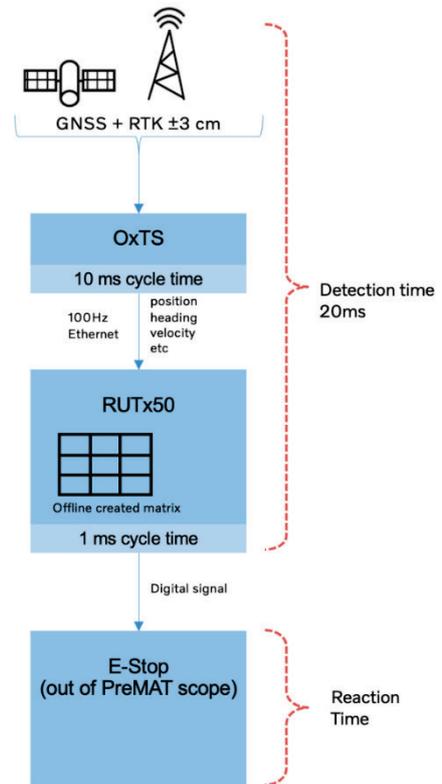


Figure 1616. Event chain in Dynamic geofence BRT

8.1.1.1 Dynamic geofence fault tolerance time interval

The chain of events is broken down to a *fault tolerance time interval*, displayed on a timeline in figure 17.

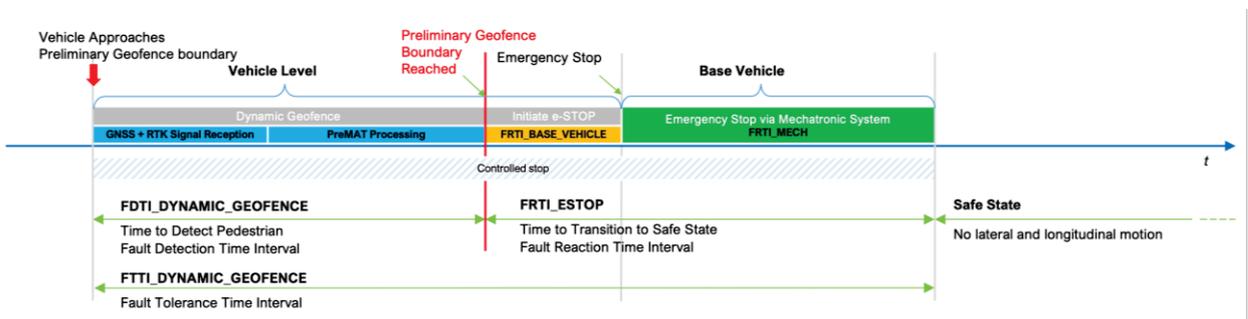


Figure 1717. Fault tolerance time interval, Dynamic geofence BRT

Each section of the timeline is described in detail in table 1.

Parameter	Value	Description
FTTI_DYNAMIC_GEOFENCE	2640 ms	Fault tolerance time interval allowed for dynamic geofence at system level. It is defined as minimum time-span from the occurrence of an event to a possible occurrence of a hazardous event, if the safety mechanisms are not activated. <i>This is the time defined for the vehicle to exit dynamic geofence if no action is taken. Within this time the vehicle needs to detect, react and stop before the outside geofence boundary is reached.</i>
└ FDTI_DYNAMIC_GEOFENCE	20 ms	Allocated fault detection time to calculate that the vehicle might exit the geofence.
└ GNSS + RTK	10 ms	This represents the estimated time allocated to the positioning unit to sample GNSS + RTK signals.
└ PreMAT Processing	10 ms	The time needed to process the information for Dynamic Geofence.
└ FRTI_ESTOP	2620 ms	This represents the fault reaction time interval at vehicle level: information was received that a ESTOP was requested until the complete stop of the vehicle.
└ FRTI_BASE_VEHICLE	100 ms	The monitoring function will transition to safe state which will take time to propagate to the base vehicle.
└ FRTI_MECH <i>Braking time</i>	2520 ms	Time allocated for the base vehicle to slow down to a complete stop. <i>This time is dependent of vehicle speed, braking capabilities and actual friction coefficient.</i>

Table 1. Dynamic geofence BRT: timeline breakdown

8.2 Pedestrian detection

The Pedestrian detection is based on a chain of events, starting with detecting a pedestrian in a hazardous location. To make the assessment, some assumptions was needed to be made:

- Minimum distance from geofence to vehicle test area is 5 m, meaning that the pedestrian needs to walk 5 m before it enters danger zone on.
- The road width at its lowest cross-section is ≥ 5 m.
- Pedestrian intention is to cross the test area.
- Pedestrian is walking and not running, average walking speed is 1.4 m/s

The critical reaction window for the complete safety function, including surveillance-based pedestrian detection and vehicle actuation, is:

$$5 \text{ m} / 1.4 \text{ m/s} \approx 3.6 \text{ sec}$$

This value represents the maximum time available from the moment a pedestrian begins moving toward the danger zone until they cross into it. Therefore, any fault detection, decision-making, and vehicle reaction must occur within this 3.6 second timeframe to ensure safe intervention.

As established previous section, the vehicle — traveling at an initial speed of 40 km/h — requires approximately 2.52 seconds to reach a full stop when emergency braking is applied.

This means that:

- The braking phase alone consumes more than half the available reaction window.
- Only about 1.08 s remain for the system to detect the pedestrian, communicate the stop signal, and initiate braking.

Given this narrow margin, the system’s overall performance — particularly the pedestrian detection latency and signal transmission reliability — must be extremely robust and predictable. Any delay in detection or signal processing could result in the vehicle entering a conflict zone before a full stop can be achieved, thus endanger pedestrian safety.

8.2.1.1 Pedestrian detection fault tolerance time interval

The chain of events is broken down to a *fault tolerance time interval*, displayed on a timeline in figure 18.

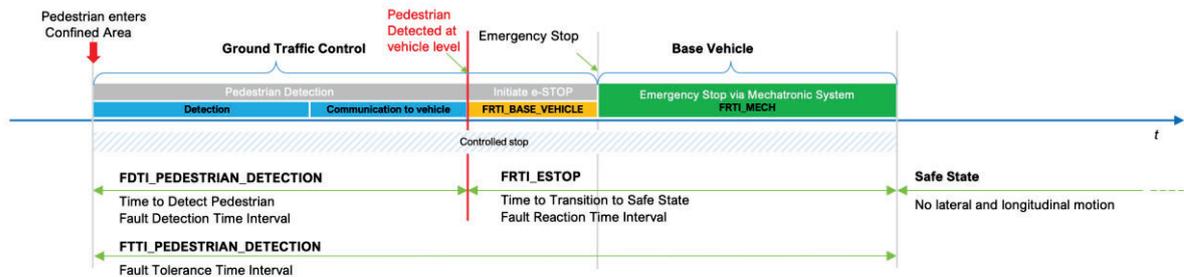


Figure 1818. Fault tolerance time interval, Pedestrian detection

Each section of the timeline is described in detail in table 2.

Parameter	Value	Description
FTTI_PEDESTRIAN_DETECTION	3600 ms	Fault tolerance time interval allowed at for pedestrian detection at system level. It is defined as minimum time-span from the occurrence of an event to a possible occurrence of a hazardous event, if the safety mechanisms are not activated. <i>This is the time it takes the pedestrian to enter a danger zone. Within this time the vehicle needs to detect, react and stop before the pedestrian is exposed.</i>
└ FDTI_PEDESTRIAN_DETECTION	1080 ms	Allocated fault detection time to detect pedestrians in the confined area. It includes multiple iterations of detection and the time needed to broadcast the information to the vehicle.
└ Detection	980 ms	This represents the estimated time allocated to sensor processing until a clear status is provided if a pedestrian is detected or not.
└ Communication	100 ms	The time needed to provide the information to the vehicle.
└ FRTI_ESTOP	2620 ms	This represents the fault reaction time interval at vehicle level since the information was received that a pedestrian was detected until the complete stop of the vehicle.
└ FRTI_BASE_VEHICLE	100 ms	The monitoring function will transition to safe state which will take time to propagate to the base vehicle.
└ FRTI_MECH Braking time	2520 ms	Time allocated for the base vehicle to slow down to a complete stop. <i>This time is dependent of vehicle speed, braking capabilities and actual friction coefficient.</i>

Table 2. Pedestrian detection timeline breakdown

8.3 System level safety requirements

In this chapter, a system analysis is summarized. This is not an exploration of the prototype system AMS, but a study of what would be needed for a surveillance system to fulfil the identified safety requirements.

Considering the environmental assumptions and the vehicle context, the following functions were identified during the MCSSA, table 3.

#	Description	PLr
1	The camera sensors shall detect a pedestrian in the confined area in less than 400 ms, with an accuracy of minimum 99%, and broadcast the information to the vehicle in less than 100 ms.	b
2	The vehicle shall engage emergency stop function when pedestrian is detected in the confined area in less than 100 ms.	b
3	The emergency stop function shall stop the vehicle in less than 3600 ms.	d
4	The e-stop shall be implemented as PLd	d

Table 3, Function identified in the MCSSA

8.3.1 Safety architecture

Category 2 is a safety architecture defined by ISO 13849-1 for safety-related parts of control systems (SRP/CS). It provides a balance between simple design and basic fault detection using periodic diagnostics. This category is commonly used for systems with low to medium risk levels, where safe system behaviour can be ensured through testable functions. Considering the intended functionality and MCSSA result this is suited for PLb safety functions, figure 19.

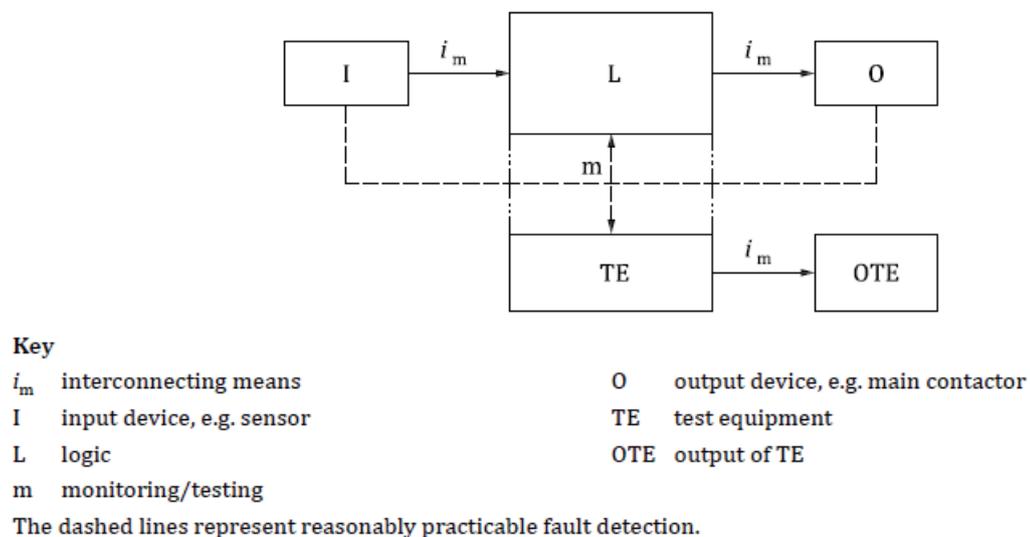


Figure 1919. ISO 13849-1:2023, Category 2, system design

Key aspects:

- Category 2 systems are based on a single-channel design.
- A diagnostic mechanism is included to detect faults, hereby either heart-beat or external watchdog.
- Diagnostics are not continuous, but must occur at defined intervals, at start-up or once per driving cycle
- If a fault is detected during diagnostics, the system must transit to a safe state.

9 Results

9.1 Dynamic geofence

The dynamic geofence was developed in two concepts. The first, *BRT*, was successfully implemented in the automated monitoring system AMS and demonstrated at AstaZero proving ground. Participants representing the industry showed big interest in the demonstration and work results, indicating that the function has the potential for industrial application. The research work and the method was presented at IEEE SMC 2024 and published (Model Predictive Geofence for Vehicle Containment).

An important finding is that a sharp brake manoeuvre like the emergency stop is not always the best mitigation. The second dynamic geofence concept, *minimal-intervention geofence*, has potential to expand the monitoring function to a more adaptive mitigation.

This concept was tested in simulation using the PCARNN-DCBF method, achieving a Containment F1 (CF1) of 0.92–0.94 (CF1: containment precision/recall harmonic mean), with a low unnecessary-intervention rate (FPR) of 5.0–6.6% (false-positive interventions) and a positive median containment margin (MCD) of 0.35–0.47 m (median positive margin to boundary). This was enabled by its hybrid Physics-Guided Control-Affine Residual Neural Network (PCARNN), which leverages a known physics model and corrects its errors with a neural network.

An Automatic Emergency Braking (AEB) fail-safe was retained for infeasible cases. Compared to a brake-only baseline, the approach reduced unnecessary hard stops while maintaining boundary invariance under the same actuation limits. These results indicate that a minimal-intervention layer can preserve full test-area usage without shrinking fences, provided short-horizon state estimation and bounded actuation are available. Full methodology and evaluation details are provided in the companion paper.

9.2 Anomaly detection

The proposed approach resulted in a flexible solution architecture that opens for multiple complementary models and tailored parameterizations. An important insight from the obtained result is the value of flexibility in the structure. The notion of an anomaly is inherently context-dependent and may be defined in multiple ways depending on the specific operational scenario and detection objectives. Consequently, no single model configuration suffices, rather a suite of models with varying parameterizations would be required to adequately capture different forms of anomalous behaviour within the data.

This flexibility allows for future scenarios where a supervisor of the autonomous vehicles dynamically updates and adjusts the suite of anomaly detectors until a satisfactory level of safety is achieved.

9.3 Pedestrian detection

The pedestrian detection concept was successfully implemented in AMS and proven to work in a demonstration at AstaZero proving ground. Participants representing the industry showed big interest also in this demonstration, indicating that the function could be a valuable tool for the automated vehicle development industry.

The concept development provided valuable learnings for the sensor developing partners, strengthening the Swedish engineering knowledge in the area. New tools supporting the development of further enhanced sensor performance was developed.

For the Pedestrian detection application, the sensor environment at the test track is fairly simple to manage compared to e.g. a crowded city street. This indicates that the sensors could have the desired capability to provide a sufficient safety level. However, the reliability is time consuming to prove since the number of events (objects to identify) is rather sparse which became an unforeseen issue.

9.4 Conclusion

The project set out to investigate means to reduce the overall risk during testing of automated vehicles, as a safety measure additional to human supervision. Ultimately, it was envisioned that a reliable supervisory system would make testing possible even without human presence.

Through the safety analysis it was shown that the investigated functions *Dynamic geofence* and *Pedestrian detection* has the potential to support this, given the right conditions. The two mentioned functions need to be further developed and tuned to match the safety requirements, but within the scope of the project they have been demonstrated as Proof of concepts.

Representatives from the autonomous vehicle industry participated during the demonstrations and confirmed that the developed functions are of high interest from their side. The Dynamic geofence has been presented at several conferences and was received by enthusiasm.

The participants of the project believe that the project results have a good potential of supporting the future development of automated vehicles by enabling testing in earlier phases and to a higher extent. If the safety requirements are fulfilled, even the safety driver could be removed from the vehicle.

10 Recommendations for future steps

The project work showed several promising results, but some research remains to reach a sufficient level of performance and reliability.

10.1 Dynamic geofence

Using only emergency stop as mitigation will probably not be suitable for many test cases. Instead, further development of adaptive mitigation strategies such as e.g. the *mMinimal-intervention geofence* is likely needed to include also the stopping phase in the safety concept.

Moreover, the pre-calculation of the BRT is quite time consuming, especially for larger and complex-shaped areas. To make the function more useful in practice, the calculation would benefit from being more efficient. When this project ended, there was a master thesis work initiated to address this issue.

10.2 Anomaly detection

While the proposed data-driven anomaly detection system establishes advantages in terms of flexibility and adaptability through its modular selection of models and parameters, there are areas that may need further research and development to determine its full potential.

First, it is recommended that future research systematically examine which model configurations offer the most promising performance for different types of anomalies and application scenarios. Second, it is advisable to explore the integration of the system with a supervisor present, enabling dynamic updates and refinements to the application of anomaly detection models until a satisfactory level of safety is achieved. Third, it would be beneficial to define and measure progress against a clear safety goal to be able to determine a definite benchmark for when the system is considered complete. Finally, it is recommended to evaluate the systems generalizability across diverse datasets, operating environments, and anomaly scenarios ensuring robustness and scalability.

10.3 Pedestrian detection

The pedestrian detection function is highly dependent on reliable sensor performance. Development of such sensors is constantly ongoing, increasing the capability

For the Pedestrian detection application, the sensor environment at the test track is fairly simple to manage compared to e.g. a crowded city street. This indicates that the sensors could have the desired capability to provide a sufficient safety level in that use case.

However, the reliability is time consuming to prove since the number of events (objects to identify) is rather sparse. It is recommended to continue the evaluation of sensors in the test track environment to ultimately reach a proven confidence of safety level in such conditions.

11 References

1. M. Bui, G. Giovanis, M. Chen, A. Shriraman. *OptimizedDP: An Efficient, User-friendly Library For Optimal Control and Dynamic Programming*. 2022.
2. Scheidegger *PCARNN-DCBF: Minimal-intervention geofence enforcement for ground vehicles*. 2025.
3. <https://ri.diva-portal.org/smash/get/diva2:1892390/FULLTEXT01>.
4. Kangqing Yu, Wei Shi, Nicola *Designing a Streaming Algorithm for Outlier Detection in Data Mining - An* 2019.
5. M. Breunig, H. Kriegel, R. Ng, J. Sander. *LOF: Identifying Density-Based Local Outliers*. 2000.
6. Z. Ding, M. Fei. *An Anomaly Detection Approach Based on Isolation Forest Algorithm for Streaming Data using Sliding Window*. 2013.
7. L. Zhang, J. Zhao, W. Li. *Online and Unsupervised Anomaly Detection for Streaming Data*. 2021.
8. M. Kontaki, A. Gounaris, A. Papadopoulos och K. Tsihlias, Y. Manolopoulos. *Continuous monitoring of distance-based outliers over data streams*. 2011.
9. <https://www.vinnova.se/p/future-5g-ride/>
10. <https://www.vinnova.se/p/intelligenta-interaktiva-och-uppkopplade-nasta-generation-realtids-forarstodssystem-i2connect/>