



Evaluation of Automated Driving by Large-Scale Piloting on European Roads – The L3Pilot Project

Christian Rösener¹(✉), Adrian Zlocki², Hendrik Weber¹,
and Johannes Hiller¹

¹ Institut für Kraftfahrzeuge (ika), RWTH Aachen University, Steinbachstraße 7,
Aachen, Germany
{christian.roesener, hendrik.weber, johannes.hiller}
@ika.rwth-aachen.de

² fka Forschungsgesellschaft Kraftfahrwesen mbH Aachen, Steinbachstraße 7,
Aachen, Germany
adrian.zlocki@fka.de

Abstract. Automated driving is still the main focus of research activities in the European automotive industry. The European Commission’s flagship project “L3Pilot” is testing the viability of automated driving as a safe and efficient means of transportation and additionally exploring and promoting new service concepts to provide inclusive mobility. Therefore, L3Pilot will create a standardized Europe-wide piloting environment, coordinating activities across the piloting community in order to acquire data from 100 vehicles with 1000 drivers operating in 10 countries in Europe. Concluding, the project will evaluate the automated driving functions (ADFs), calculate the impacts and provide a cost-benefit analysis.

The evaluation of all ADFs is challenging and requires a sophisticated methodology. This paper describes the overall evaluation approach and in particular the methods for technical- and impact assessment that are taken in the project.

Keywords: L3Pilot · Pilot · Evaluation methodology · Technical assessment · Safety impact assessment

1 Introduction

Automated driving technology has matured to a level motivating an extensive phase of road tests, which can answer the key questions on safety, security, interaction, and the societal benefit before market introduction. A large-scale pilot provides appropriate assessment of the impacts of automated driving. “What is happening inside and outside the vehicle?” and “How can vehicle security be ensured?” are two questions the L3Pilot is focused on, as well as the evaluation of the societal impact and emerging business models. L3Pilot is a European research project funded by the European Commission with €36 million of funding. The project started in September 2017 and has a duration of 48 months.

The core of the project is a large-scale pilot in which 100 automated vehicles will be operated on European roads collecting subjective and objective data, which is used to derive predictions on the social, economic and ecological impact of automated driving. Many partners of the consortium can rely valuable experience from predecessor projects like AdaptIVe [1], DRIVE C2X [1] and euroFOT [3], which form an optimal basis for the implementation of the experiment from a technological point of view as well as from a methodological one.

The overall objective of the L3Pilot project is to test and study the viability of automated driving as a safe and efficient means of transportation and to explore and promote new service concepts to provide inclusive mobility. In order to achieve this objective, a standardised Europe-wide piloting community will be created within which the piloting activities will be coordinated and harmonised. By this means, it will be possible to pilot, test and evaluate ADFs and connected automation. Furthermore, efforts will be made to innovate and promote ADFs for market introduction and wider awareness.

Overall, the project is expected to have impact on various areas. Both technical and methodological knowledge will be generated, which allow to derive requirements for function design and support simulative testing. Furthermore, an understanding of the societal impacts of automated driving will be achieved. These concern possible gains in road safety, reductions of emissions and influences on infrastructure, jobs, the economy and healthcare. The business impact will consist of guidelines defining a common basis for system design as well as validation. User data collected from the pilot will serve to explore possible business cases for market introduction of automated driving. A deployment roadmap will give an overview of necessary actions that need to be undertaken by various parties involved in automated driving in order to deploy automated driving on European roads smoothly.

2 Evaluation Methodology

Tests on public roads in real traffic are essential for the evaluation of SAE L3 and SAE L4 [4] systems since all relevant aspects of an ADF are addressed. In addition, tests on public roads offer a high complexity and variety of driving situations. Large-scale testing efforts in public traffic can ensure that the situations in which a system is tested represent all relevant driving situations. The required distance to be driven in order to guarantee a safe performance of an ADF has been predicted as high as ten million kilometres [5]. For tests conducted within the design and development stage at the manufacturer the collected data is kept confidential, especially data on critical situations involving automated vehicles. Due to this lack of data, it is not possible to draw conclusions of the overall impact that the introduction of the ADF will have. Even if data were shared, an efficient evaluation of data from multiple players would be problematic since data formats are not harmonised.

Within L3Pilot, data for evaluation will be collected from different European manufacturers, vehicles and automation functions. Although not all data is shared on all detail levels the evaluation possibilities of the available data will exceed all known data

sources up to today. This will allow L3Pilot to provide new insights for the market introduction of L3 automated driving.

The methodology applied in L3Pilot follows the guidelines defined in the FESTA handbook [6]. This handbook gives a well-elaborated roadmap for measures to be taken in preparation of a field operational test and during its deployment as well as under what aspects the gathered data should be evaluated. Based on these guidelines four key phases of the L3Pilot project can be identified as shown in Fig. 1. A distinction is made between the stages “PREPARE” (i), “DRIVE” (ii), “EVALUATE” (iii) and legal aspects & cyber-security (iv).

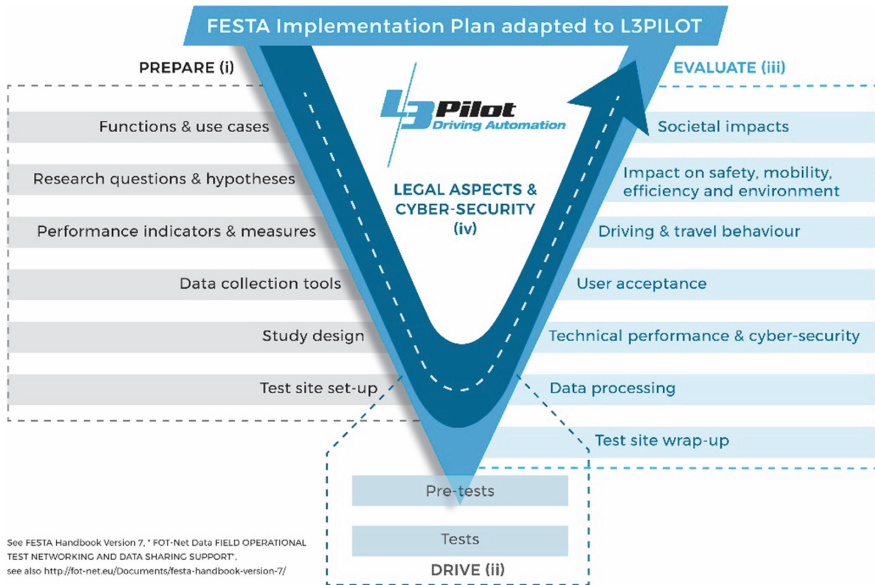


Fig. 1. Structure of L3Pilot project in accordance with FESTA guidelines.

During the “PREPARE” stage, research questions and the respective hypotheses are defined to assess the use cases of the project. In L3Pilot, the use cases are traffic jam, motorway (including traffic jam), parking and urban automation. Data collection tools are developed that are capable to analyse the derived performance indicators (PIs). In order to make sure that data to answer all research questions is collected, a harmonized study design is developed for all pilot sites. Afterwards, the subjective and objective data is collected during the “DRIVE” phase. This data is assessed in the “EVALUATE” phase. Similar to previous projects like PreVAL [7] and Adaptive [8], the data is assessed in four different areas. These are technical & traffic-, user-, impact- and socio-economic evaluation. As illustrated in Fig. 2, the entire evaluation is based on real-world data collected during the pilot. While technical & traffic evaluation and

user evaluation are in-depth analyses of the pilot data, the impact evaluation is based on aggregated and thus de-identified results of this analysis. In consequence, the impact in terms of traffic safety and efficiency will be derived based on real-world driving data. In the following, the evaluation approach is presented following the example of technical & traffic- as well as impact evaluation.

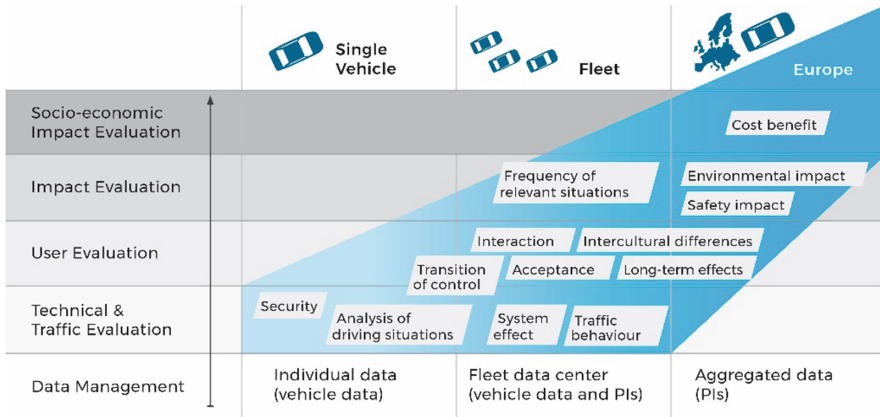


Fig. 2. Fields of evaluation

3 Technical- & Traffic Assessment

The ADFs are evaluated with regard to technical & traffic aspects based on the objective data collected in the pilot in a *scenario-based* manner. The analysis is carried out on single vehicle data. For the technical & traffic evaluation the data logged in a single vehicle (CAN-data, GPS, videos) is analysed stepwise. First, relevant driving scenarios are automatically detected. The performance indicators (PIs) are calculated for each identified driving scenario. In the last step, the derived PIs are interpreted in order to answer the defined research questions and hypotheses. Technical & traffic evaluation cover the following areas:

- What is the system’s technical performance?
- What is the impact on the own driving behaviour?
- What is the impact of ADF on the interaction with other road users?
- What is the impact on the behaviour of other traffic participants?

In this section, the evaluation methods for technical & traffic are introduced. A distinction is made between four different groups of ADFs: motorway including traffic jam, traffic jam, parking and urban. The purpose of this distinction is to group and aggregate the results of different ADFs.

Since the operational design domain of ADFs covers a high dimensional situation space including many different driving situations with lots of variations, the assessment

approach must ensure that data of the reference and the test object is available in a sufficient amount. Therefore, a holistic assessment approach covering as many different driving situations as possible is needed. The authors propose a scenario-based assessment approach based on real driving data of the test and the reference driving behaviour in accordance with [9]. By using this approach, the different test scenarios and variations of the test scenarios are generated stochastically by real-world traffic dynamics. As depicted in Fig. 3, the developed methodology first foresees a classification of test and reference driving behaviour data in relevant scenarios.

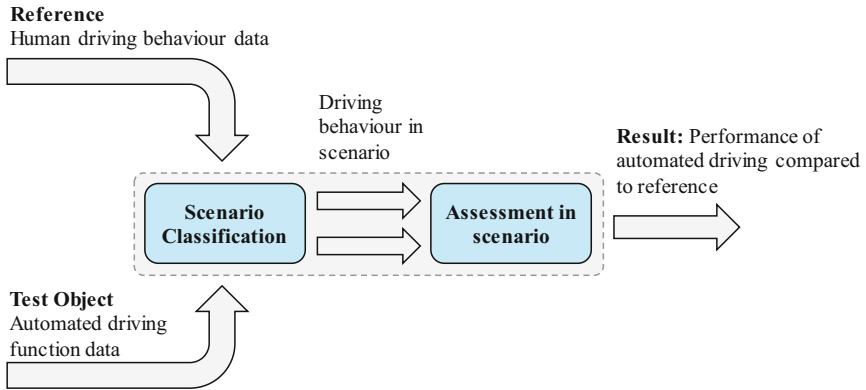


Fig. 3. Schematic view of method for technical & traffic assessment

Due to the diverse characteristics of traffic, the test approach must ensure that sufficient test and reference data is available. For this purpose, parts of the euroFOT database are considered [3] for estimating the mean frequencies of relevant driving scenarios. For calculating the minimal test distance for the occurrence of $k = 30$ driving scenarios which are necessary to assess the function, a cumulative Poisson distribution is assumed. Based on the mean distance necessary for the occurrence of a single event s_{ref} , the necessary distance is calculated for the occurrence of k events with a probability of $P = 95\%$. The basis for the calculation of the minimum distance is given with the following equation describing the Poisson distribution, where the probability for the occurrence of a driving situation is given by:

$$P = \sum \frac{\lambda^k}{k!} e^{-\lambda}$$

The expectancy value can be obtained by:

$$\lambda = \frac{S_k}{s_{ref}}$$

The resulting test distances are listed in Table 1.

Table 1. Estimated test distances.

Driving scenario	Test distance			
	k = 5	k = 10	k = 20	k = 30
Cut-in	250 km	350 km	600 km	800 km

After collecting the data in a reasonable amount, for both baseline and treatment data an enrichment is applied. In this step, all derived measures (DMs) related to dynamic objects in the environment of the ego vehicle are computed. Afterwards incidents and driving scenarios are detected based on defined thresholds. The detected incidents and driving scenarios are validated in the next step by video review according to the coding scheme presented in [10]. As a result, the distributions of PIs for baseline and treatment data are computed for each driving scenario. This is illustrated in Fig. 4 taking the example of the PI “time headway” for a “vehicle following” driving scenario.

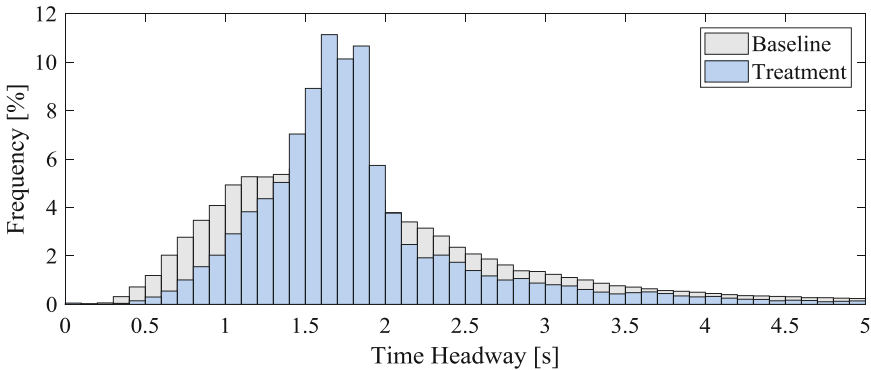


Fig. 4. Exemplary illustration of histogram of performance indicator “time headway” for baseline and treatment.

After classification of the relevant driving scenarios and derivation of the respective PIs, the earlier defined hypotheses are evaluated. For determining whether the behaviour of the ADF is within the range of normal driving behaviour, and furthermore to quantify the deviation from normal driving behaviour, an appropriate method has to be identified. In this case hypothesis testing cannot be used due to the large test samples obtained. Therefore, the usage of the quantitative measure ‘effect size’ is proposed in this approach, which is, according to [11], a simple way of quantifying the difference between two groups, that reveals many advantages over the use of tests of statistical significance alone. As depicted in [11], the effect size is a standardized mean difference between two groups and emphasizes the size of the difference rather than confounding this with sample size. The effect size d is

calculated in order to estimate the deviation of the behaviour of the ADF compared to human driving behaviour, see equation below:

$$d = \frac{\mu_{experimental} - \mu_{reference}}{\sqrt{\frac{\sigma_{experimental}^2 - \sigma_{reference}^2}{2}}}$$

4 Safety Impact Assessment

The safety impact assessment investigates the changes in accidents and injuries in road traffic due to automated driving. While in the past active safety systems were assessed based on a set of recorded accident scenarios obtained from human driving [12], this approach will not be sufficient concerning ADFs. ADFs – in contrast to active safety systems – continuously control the behavior of the vehicle. Due to this reason, it is possible that ADFs do not get involved in previously important accident scenarios any longer while other, for human driving less relevant accident scenarios, become more important.

However, it can be assumed that the relevant driving scenarios leading to certain accident scenarios will not change with automated driving. Their frequency and severity may rather change with automated driving [13]. For this reason, besides re-simulation of detailed accident scenarios for identifying the changes in severity due to automated driving, the changes in frequency of occurrence of relevant driving scenarios are investigated based on traffic simulations and the results of the technical & traffic evaluation from L3Pilot. The steps of the safety impact assessment are elaborated in the following:

- **Description of ADF & Identification of the effectiveness field**

Based on the operational design domain of the ADFs the target population of addressed accidents is identified in the accident statistics. For example, a Motorway-Chauffeur may address about 53% of all accidents on German motorways [13].

- **Changes in frequencies of driving scenarios**

Since ADFs operate continuously their engagement may lead to a change in the frequency of occurrence of certain driving scenarios, e.g. cut-in. This is investigated based on traffic simulations and the results of the technical & traffic evaluation of the pilot data.

- **Changes in severity of driving scenarios**

Within the driving scenarios which are relevant for the ADF its performance is compared with human driver reference performance.

- **Scaling-up of effectiveness to national target level**

Finally, the identified effectiveness fields in the accident statistics are used to scale-up the previously identified effects.

The overall approach for safety impact assessment incorporating the prediction of frequencies of driving scenarios is presented in Fig. 5.

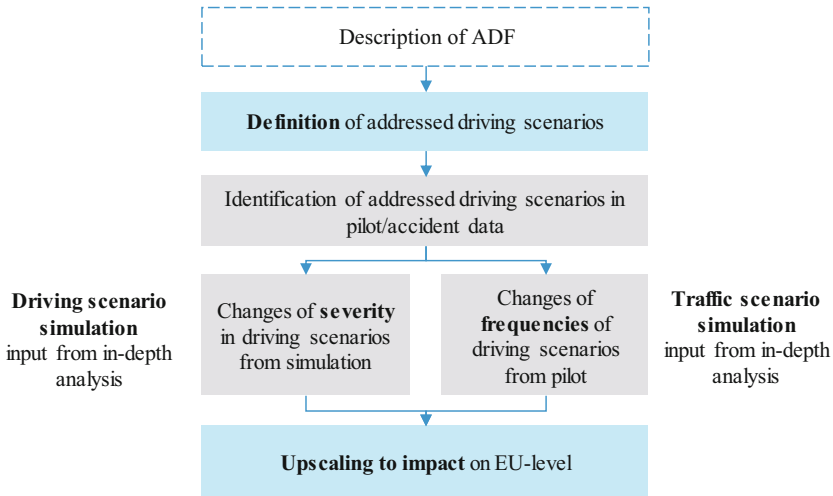


Fig. 5. Schematic of method for safety impact assessment based on [13].

5 Summary and Outlook

The overall objective of the L3Pilot project is to test and study the viability of automated driving as a safe and efficient means of transportation by a standardised Europe-wide piloting community will be created. Within this community the piloting activities will be coordinated and harmonised.

This paper describes the evaluation approach that is applied in the European research project L3Pilot following the example of technical & traffic assessment and safety impact assessment. A major challenge is the variety of ADFs ranging from motorway- to urban automation functions. In order to cope with this variety, a scenario-based assessment approach is established that is generating the results for each driving scenario instead of for each ADF. Afterwards, these can be aggregated for the analyzed ADFs.

Next, the assessment of the impact in terms of traffic safety poses challenges due to the large situation space addressed by complex ADFs. Although 100 automated vehicles will be assessed within the L3Pilot project, the penetration rate of automated vehicles will be too low to draw conclusions on overall traffic safety. To cover the entire situation space and to be able to draw conclusions on high penetration rates of automated vehicles a safety impact assessment will be performed. To ensure its validity, it is based on the aggregated results of the in-depth analysis within L3Pilot and is complemented by traffic simulations of automated vehicles.

Acknowledgment. The research leading to these results has received funding from the European Commission Horizon 2020 research programme under the project L3Pilot, grant agreement number 723051. Responsibility for the information and views set out in this publication lies entirely with the authors. The authors would like to thank all partners within L3Pilot for their cooperation and valuable contribution.

References

1. Knapp, A., Fahrenkrog, F., Zlocki, A.: Automated Driving Applications and Technologies for Intelligent Vehicles – Adaptive, 23rd Aachen Colloquium Automobile and Engine Technology (2014)
2. Stahlmann, R., Festag, A., Tomatis, A., Radusch, I., Fischer, F.: Starting European field tests for Car-2-X communication: the DRIVE C2X framework. In: 18th ITS World Congress and Exhibition (2001)
3. Benmimoun, M., Fahrenkrog, F., Zlocki, A., Eckstein, L.: Incident detection based on vehicle CAN-data within the large scale field operational test “euroFOT”. In: 22nd Enhanced Safety of Vehicles Conference (ESV 2011). Washington, DC, USA (2011)
4. N.N. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems, SAE Standard J3016_101401 SAE International (2018)
5. Winner, H.: Quo vadis, FAS? In: Handbuch Fahrerassistenzsysteme. Springer Fachmedien Wiesbaden (2015)
6. FESTA Consortium: “FESTA Handbook” FESTA Support Action, Field opErational teSt supporT Action. D6.4 (2008)
7. Scholliers, J., Hendirks, F., Aust, M.L., et al.: Project final report and recommen-dation for future assessments. PREVAL project, Deliverable DL IP D12/D16.4 (2008)
8. Fahrenkrog, F., Rösener, C., Zlocki, A., et al.: Technical evaluation and impact assessment of automated driving. In: Road Vehicle Automation 3. Springer Lecture Notes in Mobility (2016)
9. Rösener, C., Fahrenkrog, F., Uhlig, A., et al.: A scenario-based assessment approach for automated driving by using time series classification of human-driving behaviour. In: IEEE Intelligent Transportation Systems Conference (ITSC), Rio de Janeiro (2016)
10. N.N. Researcher’s Dictionary for Video Reduction Data, SHRP 2, Virginia Tech Transportation Institute, Virginia, USA (2015)
11. Coe, R.: It’s the effect size, stupid: what effect size is and why it is important. In: Annual Conference of the British Educational Research Association, pp. 12–14. University of Exeter, UK (2002)
12. Noort, M.v., Bakri, T., Fahrenkrog, F.: SIMPATO - the safety impact assessment tool of interactive. In: 16th International IEEE Annual Conference on Intelligent Transportation Systems. The Hague (2013)
13. Rösener, C., Hennecke, F., Sauerbier, J., et al.: A Traffic-based Method for Safety Impact Assessment of Road Vehicle Automation, 12. In: Uni-DAS Workshop Fahrerassistenzsysteme und automatisiertes Fahren, Walting, Germany (2018)