



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/151182/>

Version: Presentation

---

**Conference or Workshop Item:**

Innamaa, S, Merat, N, Louw, T et al. (Accepted: 2019) Methodological challenges related to real-world automated driving pilots. In: ITS World Congress 2019, 21-25 Oct 2019, Singapore. (Unpublished)

---

This is an author produced version of a conference presentation, accepted for presentation at the the 26th World Congress on Intelligent Transport Systems 2019.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

## Methodological challenges related to real-world automated driving pilots

Satu Innamaa<sup>1\*</sup>, Natasha Merat<sup>2</sup>, Tyron Louw<sup>2</sup>, Barbara Metz<sup>3</sup>, Thomas Streubel<sup>4</sup>,  
Christian Rösener<sup>5</sup>

1. VTT Technical Research Centre of Finland Ltd., Finland, \*satu.innamaa@vtt.fi
2. Institute for Transport Studies, University of Leeds, UK
3. WIVW, Germany
4. Chalmers University of Technology, Sweden
5. ika, RWTH Aachen University, Germany

### Abstract

This paper discusses the methodological challenges related to automated driving (AD) pilots in the real world, providing an overview of some of the solutions offered by the L3Pilot project. Although the overall methodology defined for Field Operational Tests (FOTs) has been developed quite extensively for driver support systems, our efforts in the L3Pilot project show that the evaluation process can be adapted to suit the needs of AD pilot projects, as long as some caveats related to the pilot nature of AD studies are acknowledged. The AD pilots currently in place around the world provide important insights into the impacts of AD on their users, other road users and the society at large. However, as these systems mature, large-scale FOTs will be needed as (closer to) ex-post evaluation, to verify the assessed impacts. This paper outlines the challenges and offers some solutions for working towards that goal.

### Keywords:

Real-world study, automated driving pilot, methodology

### Introduction

Ensuring that automated driving systems are safe, eco-efficient, and used appropriately, will not be achieved simply by applying more technology to vehicles and infrastructure. As the ultimate beneficiaries, users (i.e. both individual drivers, and society as a whole), their preferences, and factors that influence their acceptance will need to be incorporated early into the design stage [1], and any follow-on evaluation of automated driving systems. Thus, user behaviour and acceptance are key factors for ensuring the successful deployment and use of Automated Driving Functions (ADFs) in vehicles, which, in addition to other factors, such as overcoming the legal, ethical, security, and technical feasibility challenges, must be considered when assessing the full implementation of these novel systems.

In a step towards the introduction of automated cars in Europe, the L3Pilot project is preparing to conduct large-scale testing and piloting of automated driving with developed SAE Level 3 (L3)

## Methodological challenges related to real-world automated driving pilots

functionality for passenger cars, exposed to a range of users in mixed-traffic environments, along different road networks on open roads. The L3Pilot project, coordinated by Volkswagen, has 34 partners including 13 OEMs, with a total of 68 M€ budget and 48 months in duration [2]. Extensive on-road testing of SAE level 3 (L3) vehicles is vital to ensure sufficient ADF operating performance, and to allow an assessment of user interaction, understanding and acceptance of the system. A large and diverse sample of drivers needs to be involved in this work, to ensure an inclusive and effective piloting and evaluation of ADFs.

Since these L3 ADFs are still being developed by OEMs, testing involves pilot field studies, rather than Field Operational Tests (FOTs). While FOTs - a method, which has been developed and utilised for a number of European projects (listed e.g. by FOT Net in [3]), are historically used to examine user and technical performance of production vehicles and market-ready driver support systems and services, there are a number of fundamental differences between FOTs and pilot studies, which affects interpretation of the results. Essentially, a pilot project or study as defined by the Concise Oxford Thesaurus [4] is “*an experimental, exploratory, test, preliminary, trial or try out investigation*”. Thabane et al. [5] defined pilot experiment as “*an action... to test novel practices or technologies*” where the main characteristic is “*to be implemented on a smaller scale than that of the ultimate objective*”.

The objective of this paper is to discuss the methodological challenges related to automated driving pilots in the real world, providing an overview of some of the solutions offered by the L3Pilot project, to address these.

### **Comparing FOTs and On-Road Pilots**

The FESTA (Field opERational teSt supporT Action, 2007-2008) project developed a comprehensive guidance on the evaluation and delivery of driver assistance systems and functions, using an FOT methodology, to ensure that the systems delivered real-world benefits.

The FESTA Handbook [6], defines a “Field Operational Test” as: “*a study undertaken to evaluate a function, or functions, under normal operating conditions in road traffic environments typically encountered by the participants using study design so as to identify real-world effects and benefits*”. Here, “*Normal operating conditions*” implies that the participants use the platforms during their daily routines, that data logging works autonomously, and that the participants do not receive special instructions about how and where to drive. Except for some specific occasions, there is no experimenter in the vehicle, and typically, the study period extends over at least a number of weeks. In order to set up an experiment like this, the technology readiness level of the tested function must be high enough to allow the use of naïve subjects in real traffic, without supervision.

More generally, FOTs are large-scale user tests in terms of number of test participants and duration of the tests (commonly between a few months and two years). During this testing period, questionnaires,

## Methodological challenges related to real-world automated driving pilots

direct measurements and video observations are used to identify how the system potentially changes the participants' driving and travelling behaviour. FOTs also study the effect of driving behaviour on other road users, and wider impacts on the transport system and society. [6]

When designing the experimental procedures for an on-road pilot study, one must understand the difference between FOTs of (nearly) market-ready products and the piloting of systems that are at an earlier technology readiness level, such as those under investigation in the L3Pilot project. Therefore, due to safety and ethical reasons, for this type of automated driving pilot, it is only realistic to conduct tests first with designated and trained *safety drivers*<sup>1</sup>, rather than with members of the public. This procedure is very different to FOTs, where *ordinary drivers*<sup>2</sup> use the system as part of their daily lives. Thus, such a pilot study produces partly indicative estimates of impacts, and one must make assumptions about the eventual use of market-ready versions, where an FOT provides more direct proof of impacts from the field measurements.

The goal of the L3Pilot project is to demonstrate and assess the functionality and operation of Level 3 ADFs of passenger cars in real or close-to-real use contexts and environments. The project provides a great opportunity for large-scale on-road testing of automation which is not yet available in the mass market. The engagement of a large number of different OEMs, and the implementation of various ADFs in different environments and different parts of Europe enable a broader view of the potential impacts of automation than an evaluation based on a single trial. However, the pilot nature of these tests will bring some practical limitations to the possibilities regarding use and conclusions related to real-world implementations, or expected impact. To generate valid results on the impacts of the ADFs, principles used to collect the evaluation data, and any ensuing conclusions need to be considered carefully. The move towards automated driving will change the mobility ecosystem, and therefore, some new innovative assessment approaches may be needed.

### **Methods used for evaluating vehicle systems**

The purpose of an evaluation of a technical system, such as automated driving is “*to assess its status in terms of original or current expectations and to chart its future direction*” [7]. These expectations are related to the users, but also composed requirements by the developer and legislative authority. To assess the performance and benefits of systems, a comparison with the status quo (or other baseline) is also advisable. In terms of automated driving, this would, for instance, mean evaluating any changes in terms of attributes such as driver and road safety, mobility patterns and environmental impact, when compared to the status quo. However, any evaluation requires data as input, and depending on which aspect is to be evaluated, there are different evaluation methods, where both quantitative and qualitative methods

---

<sup>1</sup> ‘*Safety driver*’ is used here on specially trained, highly skilled drivers who are in charge of the vehicle during the test ride.

<sup>2</sup> ‘*Ordinary driver*’ is used here for the general public, i.e. naïve participants in terms of ADF use.

## Methodological challenges related to real-world automated driving pilots

can be used for the evaluation of systems.

In contrast to *verification*, which is testing a system against its specification, *technical evaluation*, which means “*tests or studies conducted to investigate and determine the technical suitability of... a system for the intended objectives*” [8], can be used to capture the performance of the automated system, in contrast to the human controlled vehicle. This requires a reasonable number of hours of both automated and manual driving data. On the other hand, *user evaluation* can capture the expectations, preferences, needs, etc. of the user towards the ADFs, and also identify real interactions between systems and the driver. Crucially, this kind of evaluation goes beyond any OEM-specific testing regime, and is important for ensuring the needs and limitations of different groups of drivers. Data collection methods for such user evaluation include, but are not limited to, interviews, questionnaires, and video-based/researcher observations [9].

For evaluating certain aspects of automated driving, (quasi)-experimental studies are often conducted in simulators or on test tracks, mainly due to ethical, practical, or safety reasons, and when the system is early in the development cycle. These more controlled and repeatable setups also allow response to specific research questions, by limiting the number of dependent variables. Since a real-world pilot is conducted on-road and in a mixed traffic environment, this control is given up in favour of the bigger picture, and to confirm or discard issues and benefits identified in individual, more controlled studies.

The assessment of societal impact of automated driving is another area of interest. Impact assessment can be divided into ex-ante and ex-post assessments. *Ex-ante impact assessment* analyses the likely effects of an intervention, like the introduction of automated vehicles into our transport system, and the reasoning behind them, before this intervention takes place. *Ex-post impact assessment* is an exercise, which aims to assess the outcomes and relevance of the intervention in the light of its initial objectives, and expected effects, but also indirect impacts, including consideration of undesired implications. This assessment is based, as far as possible, on empirical information that has been collected and critically analysed. An impact assessment in an open-road pilot of automated driving lands between these two. It is an ex-ante assessment but including evidence of the impacts from the field tests. Thus, the results are more valid than assessments without this evidence.

### **Methodological challenges and solutions**

This paper identifies methodological challenges that need to be considered, when considering the different phases and types of evaluation, including: the set-up of experiments, as well as assessing impacts on driving behaviour, user experience and acceptance, and societal factors. The description of these challenges is followed by proposed solutions, as offered by the L3Pilot project.

#### *Set-up of the experiments*

The field experiments are designed to provide the data needed in order to answer the research questions

## Methodological challenges related to real-world automated driving pilots

set for the project. A list of relevant research questions are defined for a pilot project, just as they would be formulated for an FOT. They are shaped around the focus of the project; description of the tested systems; and theories of the related impact areas. However, the prioritisation and selection of research questions cannot be based on these factors alone, but must also consider the practical possibilities of each pilot site and vehicles used for each study. In L3Pilot, feasibility of research questions were also checked against possibilities for data provision (sensors, logging, features of test rides, experimental design) as well as the role and type of participants (drivers).

Since ADFs are still prototypes, special measures regarding safety are necessary for the open-road pilot tests. This is why, for example, it not possible to study the interactions and behaviour of ordinary drivers, using the vehicles during their daily routines. Instead, such testing requires the inclusion of safety drivers and additional observers, to monitor system and driver performance. However, to ensure a more representative sample of drivers and overcome this limitation, the L3Pilot project uses a number of special safety concepts, including:

- Equipping some vehicles with driving-school-style pedals, allowing interjection by a trained observer in the passenger seat, where necessary.
- Placing an ordinary driver in the passenger seat, to observe the safety driver's interactions with the system and vehicle, providing some impression of the system.
- Following some work with the safety driver, some pilot studies are then progressed to using ordinary drivers, with increasing maturity of ADFs.

### *Assessing the impacts on driving behaviour*

When assessing the impact of the ADFs on driving behaviour or dynamics (like observed differences between the ADF and human driver regarding e.g. car-following behaviour, or speed and headway distributions), it is important to consider the maturity of the system, and whether it offers a representative and realistic driving scenario. Also, in order obtain permission for testing ADFs on an open road, public authorities require that it is safe to do so. The evaluation of the details of driving dynamics of automated vehicles can also be sensitive to car manufacturers. Therefore, it is also important to ensure that vehicle telemetry data is kept confidential, and any information that is shared within the project, or to the general public, cannot allow manufactures to be ranked or compared, compromising the competitiveness of OEMs at this crucial development stage. A further challenge for developing a broad view of driving behaviour impacts of ADFs is that each result may only be limited to certain test routes, within certain speed ranges, and in certain weather conditions, as specified by the function's Operational Design Domain (ODD) [10]. Such pilot studies are also obliged to adhere to the OEM rules and national regulations drawn up for testing of automated driving on open roads and in controlled environments, such as the mandatory use of safety drivers in some situations. Thus, the above specifics limit the possibility of addressing all of the research questions set by the evaluation team.

## Methodological challenges related to real-world automated driving pilots

To address these limitations, the following solutions have been drawn up by the L3Pilot project:

1. Project partners have agreed on a common data format and produced a common methodology and analysis toolkit, to be used by different analysts, at different pilot sites, with data from different ADFs.
2. The most detailed data from each pilot site is only handled by one or two research partners, ensuring that the distribution of commercially sensitive data is controlled.
3. A sophisticated data sharing process designed to anonymize the data coming from each pilot site is used. While the aspects concerning technical evaluation are investigated with detailed data coming from the field tests, the high-level analysis of societal impacts are assessed based on *aggregated data* that can be anonymized. To obtain the aggregated data, the results from individual pilot sites are merged.
4. The results from several pilot sites will be merged in such a way that sensitive information is protected. In practice, the public outcomes for a particular ADF can be presented only if it was piloted at more than one site. A fundamental requirement for the merging is, that in addition to protection of the privacy of manufacturers, the outcome is meaningful for those who utilise the results, either as such or in the following steps of evaluation. Thus, the results must give insights of the impacts of automated driving without compromising this privacy. These same principles are applied also for presenting the results of user experience and acceptance evaluation. This merging across different sites also ensures that the results of L3Pilot do not show the impact of single (OEM)-specific ADFs, but the averaged impact that can be expected if such systems are introduced to the road.

### *Assessing of user experience and acceptance*

There are a number of aspects to consider when aiming to extrapolate the user experience and acceptance in an automated driving pilot to the real world. These can be divided into three main areas, namely, the maturity of the system, the test environment used, and the type of driver.

1. As mentioned in the previous section, drivers will be exposed to prototype Human Machine Interfaces and automated driving control systems, which are still under development, potentially resulting in occasionally unpleasant driving and interaction experiences. This will affect the user experience and thus acceptance. For example, a development system that is prone to errors is likely to elicit different acceptance ratings compared to a market-ready system.
2. The nature of the pilot test means that participants will not be able to use the systems in their daily lives. Moreover, the test environments will likely range from free driving on a designated route, to performing specific manoeuvres in a controlled environment. Therefore, the fidelity of the test environment may elicit different behaviours, uses, and perceptions than those in a potential daily life environment.
3. While ordinary drivers should be used as participants in every instance possible, the prototypical nature of the systems means that, for safety and legal reasons, in many instances there will be the requisite of participants either being trained safety drivers or being recruited from the OEM

## Methodological challenges related to real-world automated driving pilots

workforce. When using OEM employees who may not be familiar with the system, they are often required to be accompanied by safety drivers. Therefore, the perceptions and behaviour of the safety driver are likely to be influenced by their special training, but also their presence in the vehicle may influence the behaviour and perceptions of an ordinary driver. Careful consideration must therefore be given regarding instructions to the safety driver, but ethical matters must also be well-thought-out, e.g. in terms of collision avoidance strategies.

The above factors limit the extent to which one can make firm and long-lasting conclusions about driver behaviour during interactions with, and perceptions of ADFs, in the real world. However, while controlled test rides in prototypical systems with safety drivers on board are not the same as driving on your own, during your daily commute, the behaviour and opinions of those who have participated in the test rides arguably have some value, and may be more valid, than those who have had no such physical experience, or indeed that of designers of the systems, who are traditionally used for such evaluations. Therefore, it is possible to answer many of the user- and acceptance-related research questions, so long as caveats are included when results and recommendations are presented. However, the nature of the system, environment and driver type used in an on-road pilot study means that, ultimately, there are some aspects that cannot be addressed.

L3Pilot will also use a number of additional approaches to answer user- and acceptance-related questions that cannot be addressed with questionnaires or via vehicle-based videos captured during the test drives. First, interviews or focus group discussions will target specific questions for which answers cannot be obtained directly from video or questionnaire data. For example, these can be used to probe insights of safety drivers regarding interactions with the system. Second, driving simulators will be employed to assess safety-critical aspects that are either ethically or practically not possible to examine in field experiments. For example, fatigue-related performance impairments during resumption of control from automated driving. Finally, a large-scale, multi-wave, international survey will be used to tackle questions related to market trends that field experiments cannot address. These include aspects of willingness to pay for a system, which may not be relevant to a company's test participants.

### *Assessing of societal impacts*

For FOTs, the impact analysts can assume that the field measurements represent true changes in driving behaviour. The implications of these the changes on factors such as driving dynamics and travel behaviour, and their effect on traffic safety, transport network efficiency, environment and mobility could then be scaled up to make estimations about the entire traffic flow and vehicle fleet levels, and up further to understand expected impacts on national or EU levels. In automated driving pilots, a challenge (and source of uncertainty) is that you have one additional level of assessment, involving the assessment of any differences between piloted versions of prototypes, and their market-ready version, relevant to assessing their higher penetration rates for investigating societal impacts.

## Methodological challenges related to real-world automated driving pilots

Another methodological challenge is the selection of a meaningful baseline condition for assessing the impacts of the ADFs. While it would be ideal, it is unrealistic to expect a baseline that does not include any automation, as current vehicles already include low-level automated driving and ADAS, such as (Adaptive) Cruise Control and Lane Keeping Systems, with (low) penetration rates varying between regions and fleets. Having a mix of automation levels as baseline sets high requirements for the evaluation, as you should then understand the differences between all the levels of automation included, and implications of their interaction. If a future dimension is added to the assessment, the influence of the other trends affecting mobility and transport (such as electrification, urbanization, (other) new mobility services and concepts, etc.) should be included. Therefore, not only are the baseline data sets we have rather redundant, but, given the nature of the changing mobility ecosystem (e.g. electric cars, MaaS, 5G, shared AVs) along with automation, it is difficult to extrapolate impacts of ADFs too far into the future.

The evaluation plan in L3Pilot include combining different best-practice solutions for different evaluation areas. For example, safety impact assessments is planned to be made based on different driving scenarios, their changes in severity and frequencies, and related impacts, using accident statistics for the scaling-up. Other assessments will be based on generalised higher-level impacts, combined with national- and EU-level statistics on transport, especially for areas and networks where detailed data is not available. The impact analysis is not done for the tested prototype versions. Instead, (future) market-ready functions are defined, and the extent to which their expected impact can be directly based on the results from the field data analysis is thoroughly checked, establishing where, and to what extent supplementing assumptions are needed. The descriptions of market-ready systems is made together with OEMs, to represent situations that are in line with penetration rate assumptions. In addition, it sets the evaluation on a more general level, not linking it with single solutions and locations.

## Conclusions

The aim of this paper was to discuss the challenges specific to automated driving pilots and address solutions found by the L3Pilot project.

The work in L3Pilot started with methodological approaches and methods developed and used in previous projects on Advanced Driver Assistant Systems (ADAS). Due to the nature of ADFs, the prototype systems tested in the project, and the need to ensure safety of the driver and all surrounding traffic throughout the tests, adaptation and further development of existing methods are needed and implemented. This method relates to the full chain of assessment, starting from technical assessment and evaluation of user related aspects, including adaptation of modelling approaches for impact assessment and ending at choosing the best set-up for socio-economic evaluation of the expected impact. The fact that data will be collected during on-road tests at various locations throughout Europe with varying versions of future ADFs provides the unique opportunity to estimate the future impact of system types (e.g. L3 motorway chauffeur) and not of single implementations. However, this also poses new

## Methodological challenges related to real-world automated driving pilots

challenges to the methodology, since procedures need to be developed that allow the merger of data across test sites. For this process, requirements by project partners (e.g. on confidentiality) need to be fulfilled, while also ensuring scientifically valid and meaningful results are achieved.

Although the overall methodology defined for FOTs has been developed for driver support systems, our efforts in the L3Pilot project show that the evaluation process can be adapted to suit the needs of automated driving pilot projects, as long as some caveats related to the pilot nature of automated driving studies are acknowledged.

The automated driving pilots currently on-going around the world provide important insights into the impacts of automated driving on their users, other road users and society. However, the overall conclusion is that as these systems mature, large-scale field operational tests will be needed as (closer to) ex-post evaluation, to verify the assessed impacts.

### Acknowledgements

The research leading to these results has received funding from the European Commission Horizon 2020 program 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 partners within L3Pilot for their cooperation and valuable contribution.

### References

1. Merat, N., Lee, J.D. (2012). *Preface to the Special Section on Human Factors and Automation in Vehicles: Designing Highly Automated Vehicles With the Driver in Mind*. Human Factors. 64(5), pp. 681-686
2. L3Pilot (2017). L3Pilot website: <https://l3pilot.eu>
3. FOT Net (2019). FOT Net Wiki: <http://wiki.fot-net.eu>
4. Waite, M. (2002) Concise Oxford Thesaurus. 2. Oxford, England: Oxford University Press.
5. Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L.P., Robson, R., Thabane, M., Giangregorio, L., Goldsmith, C.H. (2010). "A tutorial on pilot studies: the what, why and how". BMC Med Res Methodol. 10 (1). doi:10.1186/1471-2288-10-1
6. FOT-Net & CARTRE (2018). FESTA Handbook, version 7. <https://connectedautomateddriving.eu/wp-content/uploads/2019/01/FESTA-Handbook-Version-7.pdf>
7. The McGraw-Hill Companies Inc. (2003). McGraw-Hill Dictionary of Scientific & Technical Terms, 6E.
8. Business Dictionary (2019). <http://www.businessdictionary.com/> (Accessed 9 January 2019)

Methodological challenges related to real-world automated driving pilots

9. Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods: A practical guide for engineering and design*. (1st ed.) UK: Ashgate Publishing Limited.
10. SAE (2018). SAE J 3016-2018: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.