Method for Technical & Traffic Evaluation

L3Pilot Final Event

Hendrik Weber

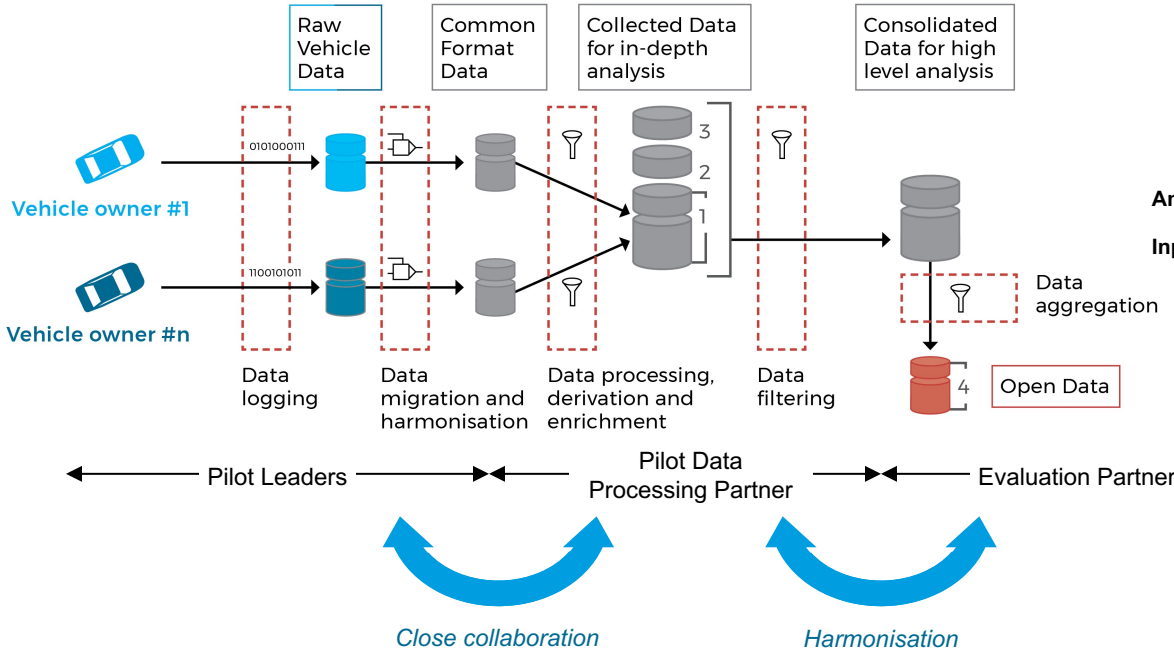
ika RWTH Aachen University

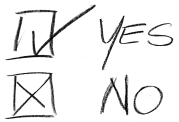
Data Flow for Evaluation

Vehicle Data (e.g. CAN)

Videos (external & internal)

User Questionnaires





Answer research questions

Input data to Impact Assessment

Challenges for data evaluation

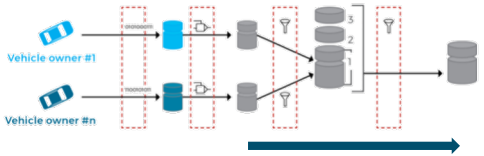
Challenge:

- Analysis needs to include data coming from various pilot sites, various ADF implementations
- Evaluation methods and tool should not allow benchmarking between the different pilot sites or re-engineering of systems

Chosen approach:

- Anonymous upload of performance indicators derived from **vehicle data**
  - Performance indicators linked to individual research questions
  - Only necessary meta data
- Unaggregated **questionnaire data** (without free text)
  - Minimum amount of meta data (e.g. to tell apart data urban and motorway ADF)

Basic approach
Analysis based on driving scenarios



Driving Scenario: *Short period of driving defined by main driving task or triggering event*

- Instances of Driving Scenarios are the basic unit for analysis
- Allow harmonized analysis and merging of data on level of driving scenarios

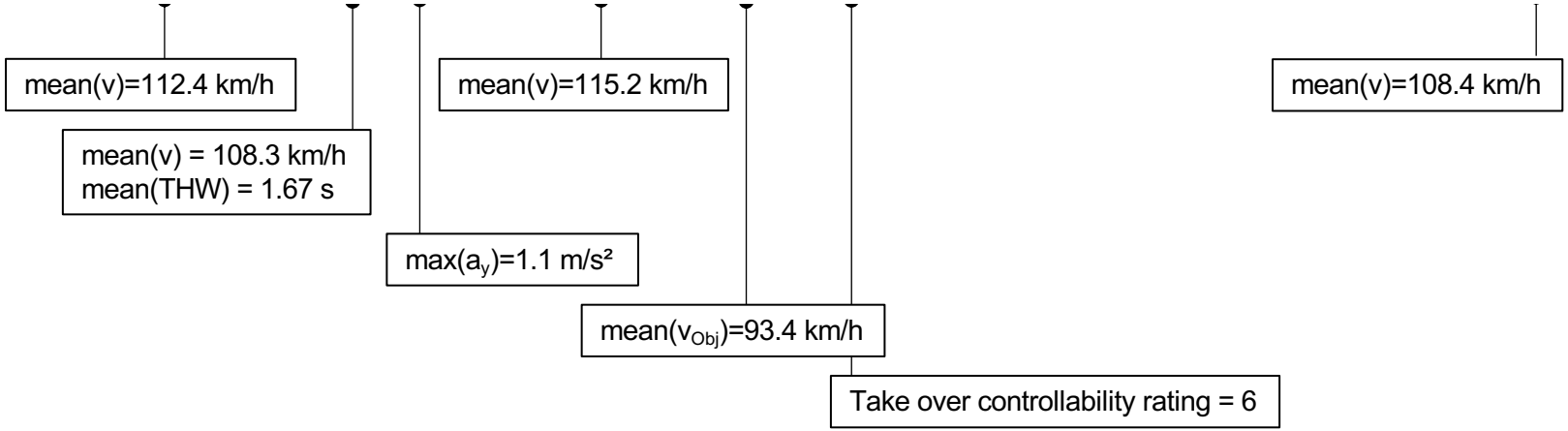
Performance Indicator: *Measure per instance of Driving Scenarios or per trip defined to answer research questions*

- Common scripts for identifying driving scenarios in driving data and calculating performance indicators

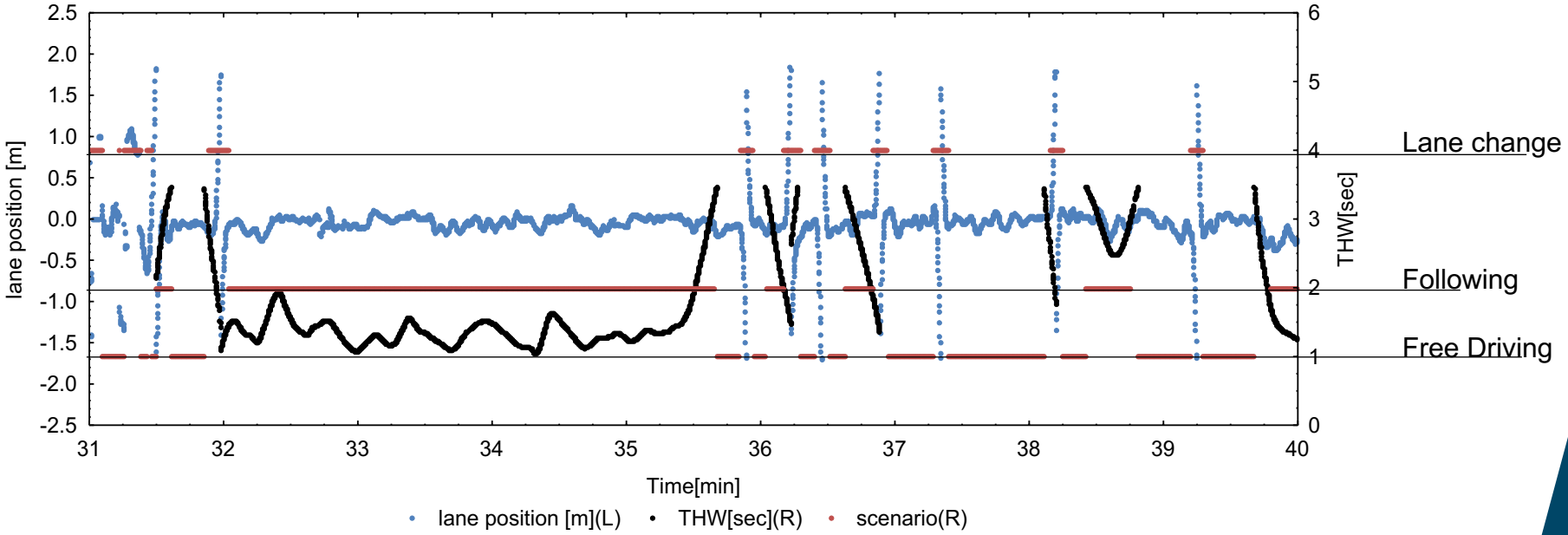
Motorway + Urban	Urban only
Free driving	Crossing
Approaching a lead object	Crossing with laterally moving object / VRU
Following a lead object	Turning (without conflict)
Driving in traffic jam	Turning with lead object
Lane change	Turning with laterally moving object / VRU
Cut-in	Overtaking of oncoming traffic

Theoretical Example of Scenario-based Assessment

- Segmenting the drive into instances of the defined driving scenarios



Identifying Scenarios in Real World Data



Performance indicators

- Most indicators used for describing **behaviour of the ADF** are derived per driving scenario instance and describe continuous vehicle signals e.g.
  - Mean speed: $m(v)$
  - Mean Time Headway: $m(THW)$
  - Std. deviation of lat. Acceleration: $sd(a_y)$
  - Minimum longitudinal acceleration: $\min(ax)$
- Indicators describing the **functionality of the ADF** are derived per trip for trip sections within ODD e.g.
  - Frequency of Take-over requests: $f(TOR)$
  - % of time automated driving function is available
- Indicators describing the **frequency of events** – per hour / per km
  - Frequency of driving scenarios: $f(Cut\ In)$
  - Frequency of incidents: $f(TTC\ <\ 1.75\ s)$

Statistical Testing

What is the impact of ADF on the accuracy of driving?

How to answer the defined research questions?

- **Non-parametrical tests** are used (mostly Mann-Whitney-U-test)
  - One procedure can be used for all parameters, it is not necessary to use different procedures or tests based on characteristics of data
  - One test per parameter and situational combination which tests for differences between experimental conditions (baseline vs. ADF active).
  - Impact of other factors (e.g. scenario type, situational factors) is not considered directly but addressed in multiple tests.

Additional Information

- Additionally, reporting of **effect size** and **change in percent**
  - Because of the large sample sizes (especially for scenario based indicators) almost everything will turn out significant
  - Effect size & descriptive information are necessary to understand the relevance of the effects
- Calculation of effect size (Cohen's D)
  - $D = \frac{\mu_{ADF} - \mu_{BL}}{\sigma}$ with σ estimated as $s = \sqrt{\frac{(n_{ADF} - 1)s_{ADF}^2 + (n_{BL} - 1)s_{BL}^2}{n_{ADF} + n_{BL} - 2}}$
- Calculation of change in percent
  - For Performance Indicator X: $\text{Change} = \frac{X_{ADF} - X_{BL}}{X_{BL}}$

Chunking of Long Scenarios

- Scenario instances will have different durations – especially *Following lead vehicle* and *Uninfluenced driving*
- This affects our ability to correctly answer Research Questions
  - Performance indicators are influenced by scenario duration
  - Scenario durations are different for BL and ADF

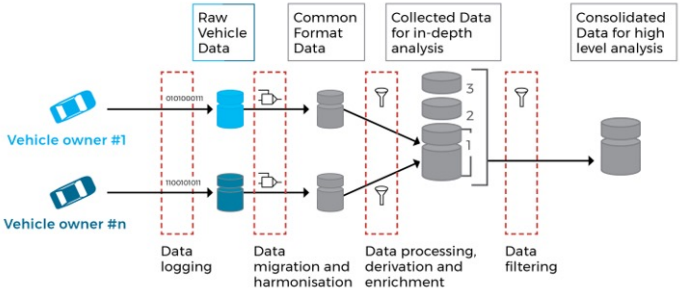
  } *Potential confounding factor!*
- To mitigate this confounding factor, data are chunked to 10 s intervals for Uninfluenced Driving and for Following lead vehicle
- After this, the standard non-parametric test can be applied

Lessons learnt from the evaluation

- L3Pilot Common data format (CDF) was a key enabler for a harmonised evaluation

  → *Available at https://github.com/l3pilot/l3pilot-cdf*

- Sharing the code for the toolchain and a collaborative testing create consistency in data evaluation across pilot sites



- Checks to ensure data format and quality are appropriate
  - Ensure that no unit conversion errors are present
  - Check that scenarios are detected correctly in different countries & environments
- Problems you notice too late, require a complete update of database contents

  → *We had to make 5 re-uploads until we got to our final dataset.*

Thank you for your kind attention.

Hendrik Weber
hendrik.weber@ika.rwth-aachen.de

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723051.