Technische Universität Berlin

Scientists who stare at data

Dipl.-Ing. Jan Grüner | MathWorks Automotive Conference 2019

About

TU Berlin // FVB

- Chair of Naturalistic Driving Oberservation for Energetic Optimisation and Accident Avoidance
- Focus on Electromobility

www.fvb.tu-berlin.de

Myself

- Research Assistant
- Matlab since 2009
  - Incl. several years of teaching Matlab
- PHD Thesis: “Usage behavior of hybrid vehicles”

jan.gruener@tu-berlin.de

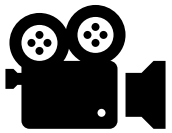
Matlab Projects

- IDCB (Create individual driving cycles from user data)
- Database Toolbox (wrapper functions for MySQL Database Connector)
- AMPERE (Usage behavior of hybrid vehicles)
- WeatherDB (Weather DB / API to make use of data provided by the DWD)

We are living in the information processing age

We record everything & everywhere

- Cameras
- Cars
- Smart-X (Phones, Homes)
- IoT

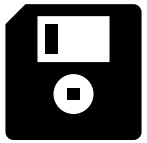


We can transmit data from A to B

- Worldwide
- Instantly / Fast
- Everything is connected

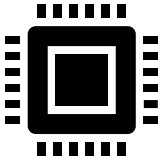
We can store huge amounts of data

- Fast storage
- Large storage

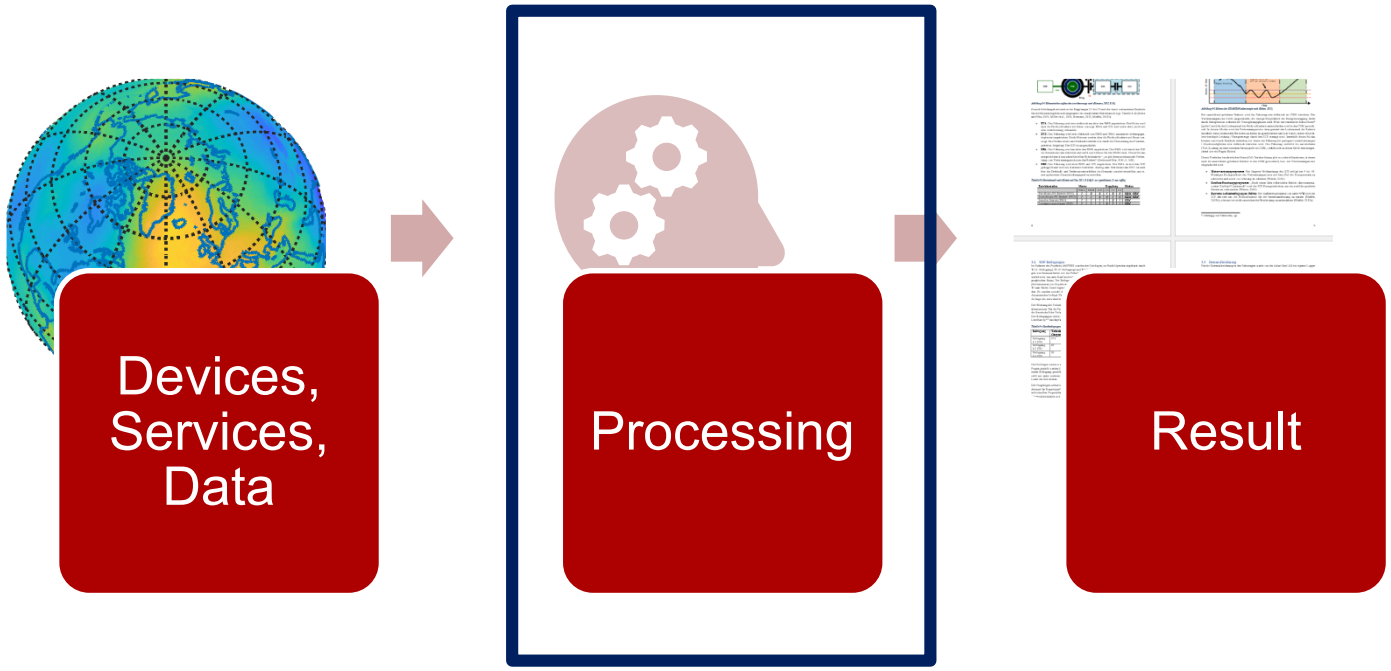


We have the computational power to analyze it

- Server
- Cluster
- Databases
- Software



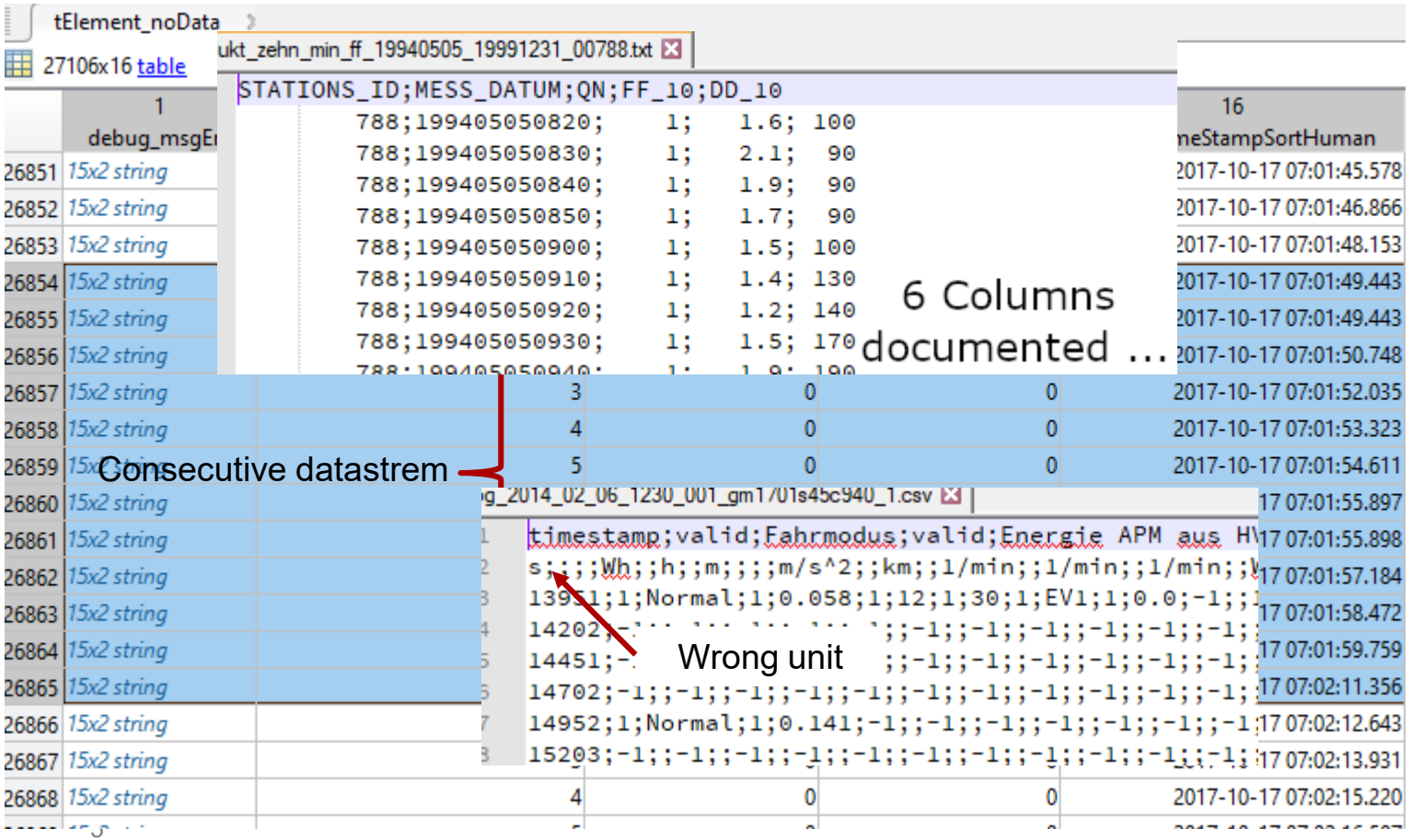
The challenge is no longer getting data, it‘s processing it



(and making sense of)

- With endless possibilities comes complexity
- Complexity creates chaos
  - Data formats
  - Competing standards
  - Methods
- A huge amount of information

In tech / data we (blindly) trust



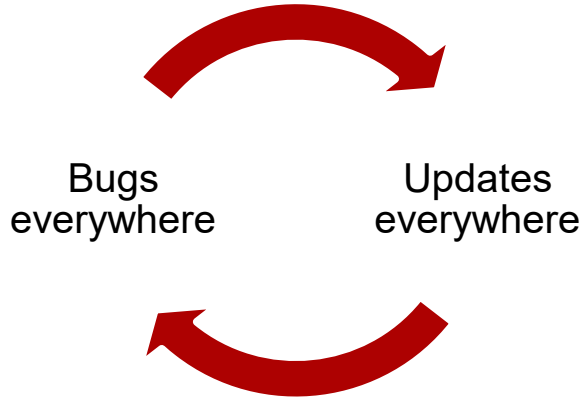
The pessimistic view on hardware / services

- Sensors are cheap, the device isn‘t
- Updates?
- Security?
- Testing?
- Documentation?
- Support?
- May change at any time

Question the data and your work (and everyone else)

The situation

- Assumptions ➔ Can become false over time
- Documentation ➔ Outdated (best case)
- Human error ➔ nobody is prefect



- (Variable-)Names will not change over time
- The documentation is complete

"

- The RMS value is inserted into a circular list of 401 en
  - No event is fired until this list is initially fi
  - TBD should this list be persiste it after each reboot?

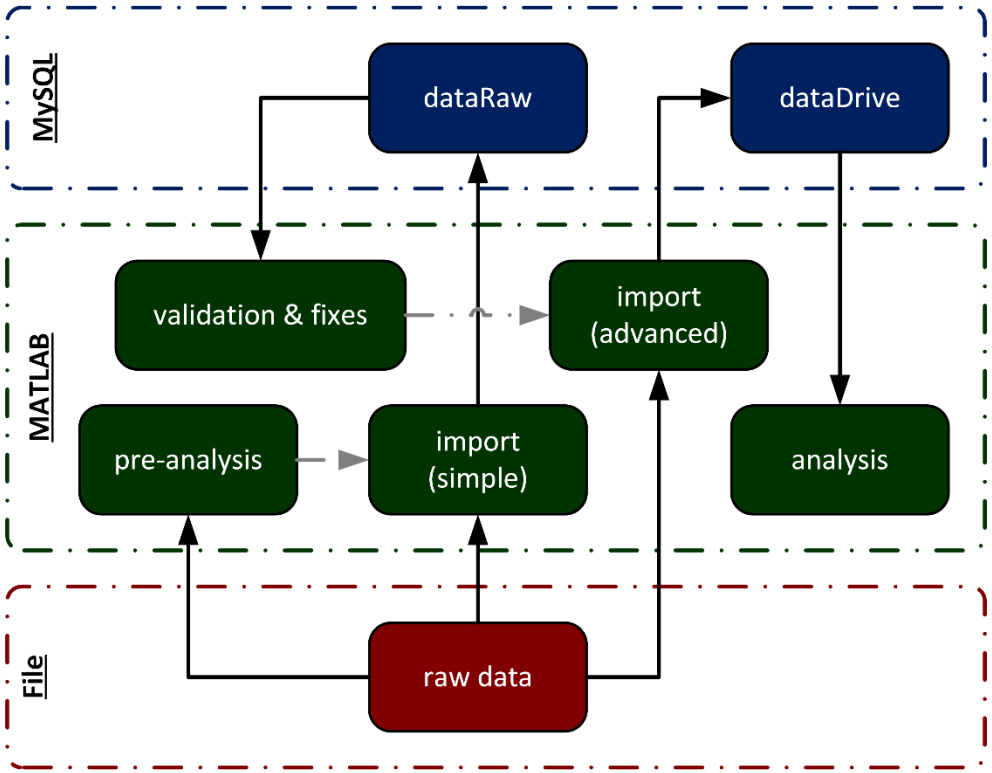
Final document

"

Rules of data processing

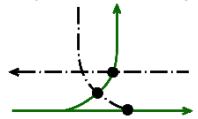
- Assume the data is broken
- Mistakes will happen
- Re-check your data / existing processes regularly

Workflow (an example)



Why Matlab

- University (Everything is DIY)
  - You want it, you build it
  - You built it, you run / manage it
- Maintenance / teaching experience
  - Easy to learn / to understand (for students)
  - Avoid multiple languages
    - Complexity
    - Knowledge required
- Development
  - Easy to visualize data along the way
  - Ready to use toolboxes (or create your own)



Import (a simple) (CSV)

GetMD5()

Pre-Analysis
- Column names, units, limits
- Create overall mapping architecture

Challenges
- Remove duplicates
- Data format

- Missing columns
- Changing column names
- No fixed column order
- Changing data units

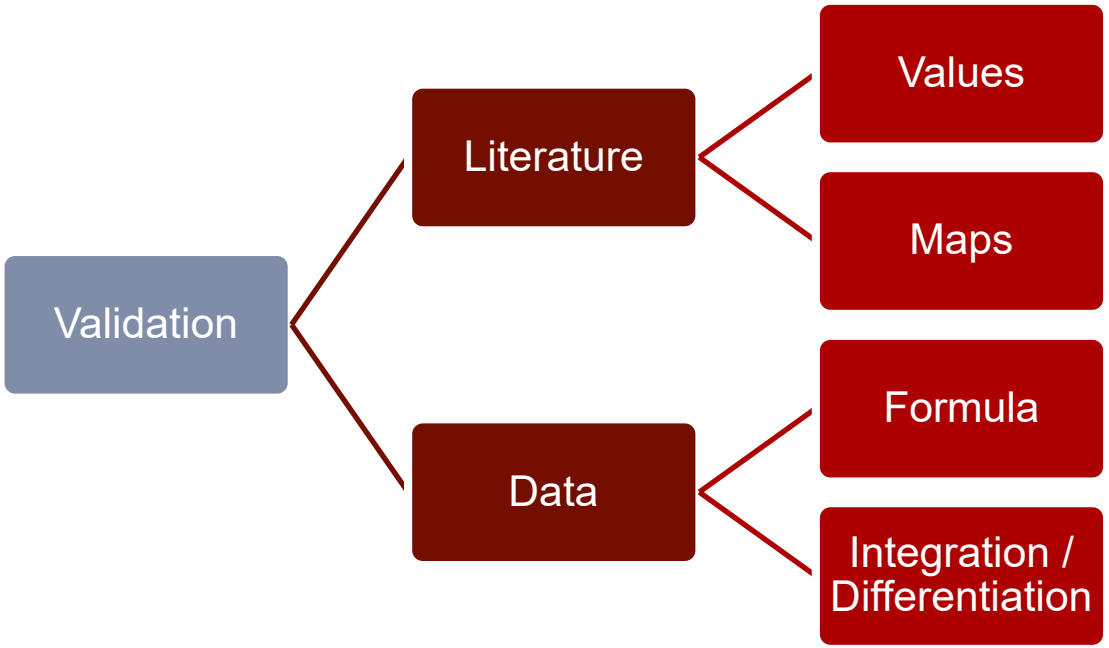
Order, names, units (SI)

Import
- Read header
- Map data
- Process

Storage
- Set datatypes
- Store in database

MATLAB to MySQL
➔ MySQL Database Connector

Validation & Fixes



Literature

- Not always correct either
- Contradicting sources

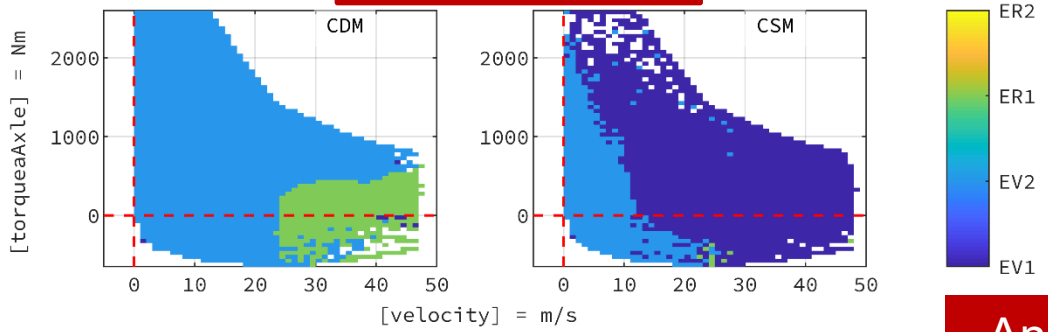
Knowledge is always required!

Data

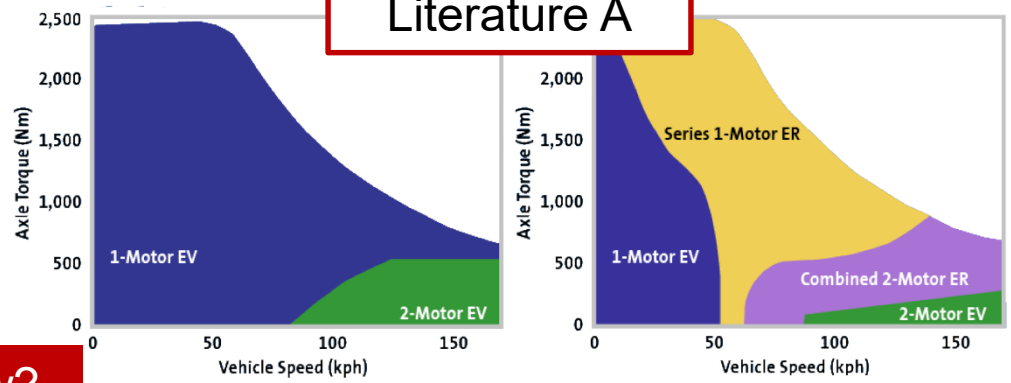
- Requires Formulas
  - Documentation?
  - Generally known?
- Signal can be estimated with “basic” math

Literature

Data

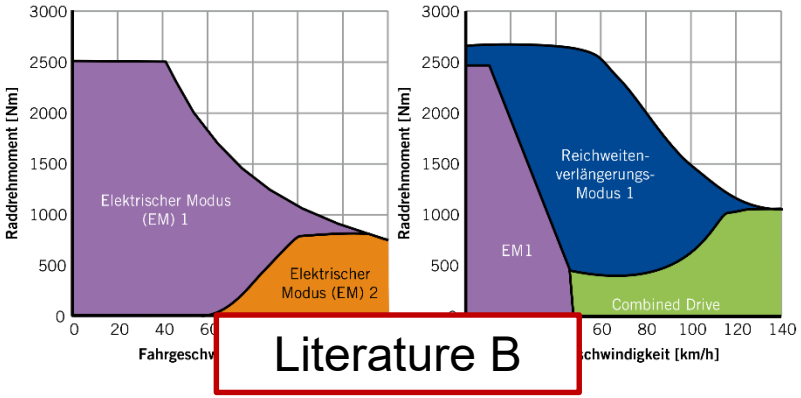


Literature A

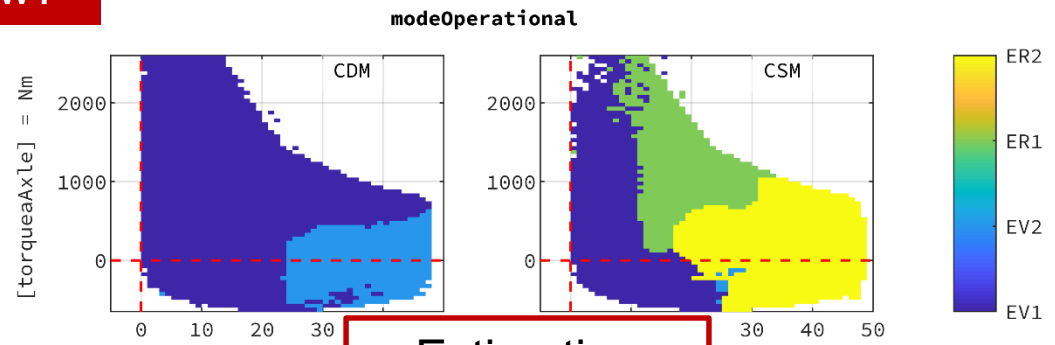


And now?

Literature B



Estimation



A: MATTHÉ, R., 2012c. Opel Ampera - Electric Propulsion and Energy Storage System, Elektrik / Elektronik in Hybrid- und Elektrofahrzeugen und elektrisches Energiemanagement, Deutschland
B: GREBE, U.D. und L.T. NITZ, 2011. Voltec – Das Antriebssystem für Chevrolet Volt und Opel Ampera. MTZ - Motortechnische Zeitschrift, 72(5), 342-351. ISSN 0024-8525

Literature

Relative Positive Acceleration (RPA)

- Descriptor for the dynamics of a (driving) cycle

#	Formula	Literature
(1)	$RPA = \frac{1}{x}\sum a_i^+ * v_i$	(Bratt und Ericsson, 2000, S. 5, The European Comission, 2016, S. 12)
(2)	$RPA = \frac{1}{x}\int v * a^+$	(Ericsson, 2000, S. 11)
(3)	$RPA = \prod_{k=1}^{end} v(k) * a(k)/mean(v)$	(Blanco-Rodriguez, Vagnoni und Holderbaum, 2016, S. 653)

- Assumptions make the difference
- Documentation

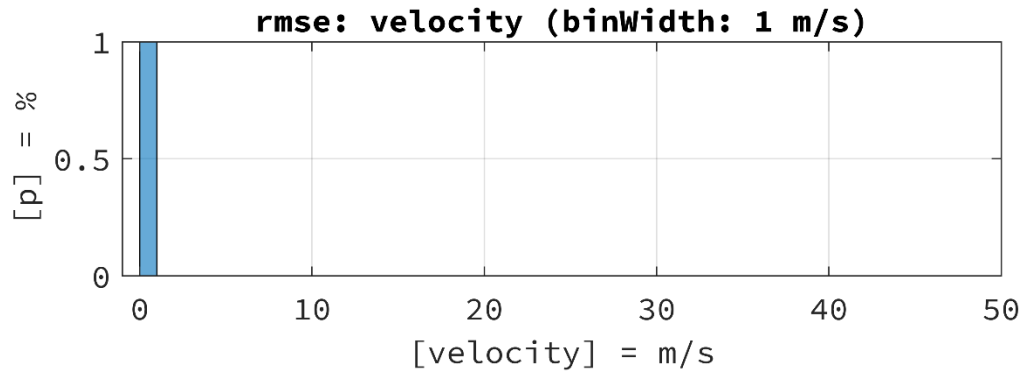
a : acceleration
v : velocity
x : distance

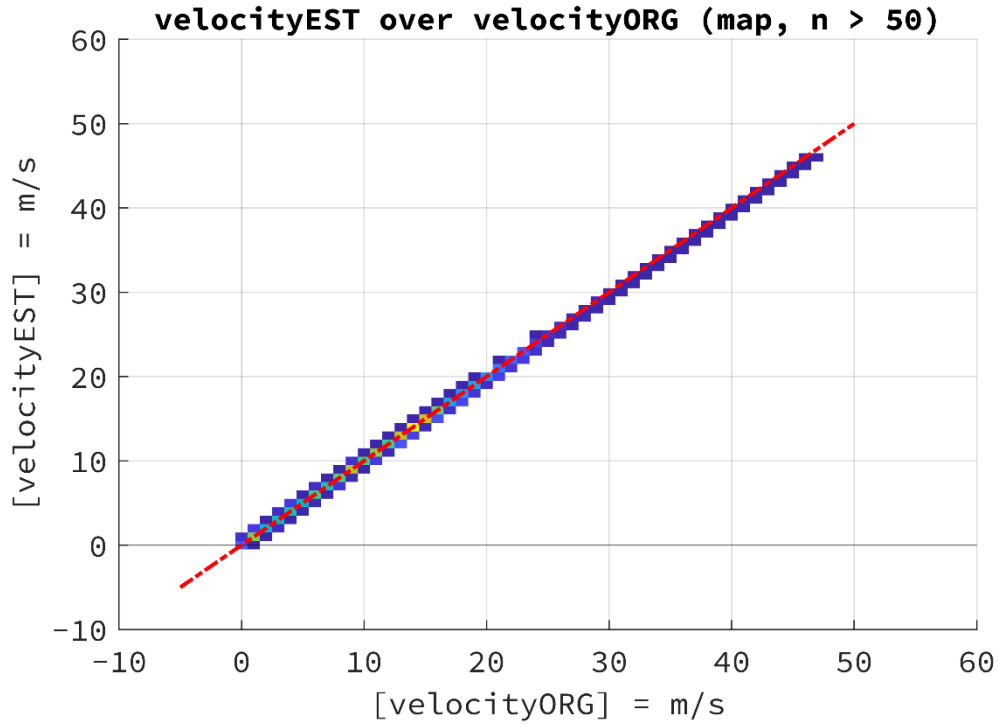
(1) BRATT, H. und E. ERICSSON, Hg., 2000. Measuring vehicle driving patterns – estimating the influence of different measuring intervals. Lund, Schweden.
(1) Comission Regulation (EU) 2016/646 [online]. In: Official Journal of the European Union [Zugriff am: 23. Februar 2019]. https://eur-lex.europa.eu/eli/reg/2016/646/oj
(2) ERICSSON, E., Hg., 2000. Driving pattern in urban areas - descriptive analysis and initial prediction model. Lund, Schweden: Lund Institute of Technology. Bulletin. 185.
(3) BLANCO-RODRIGUEZ, D., G. VAGNONI und B. HOLDERBAUM, 2016. EU6 C-Segment Diesel vehicles, a challenging segment to meet RDE and WLTP requirements. IFAC-PapersOnLine, 49(11), 649-656. ISSN 24058963.

Data

Velocity

- Formula
  - $velocity_{EST} = \frac{2*\pi*r_{wheel}*speed_{EMB}}{ratio_A*ratio_B}$
- (expected) Value range
  - 0 : 170 km/h
- Conditions
  - Validation only possible in a specific electric mode

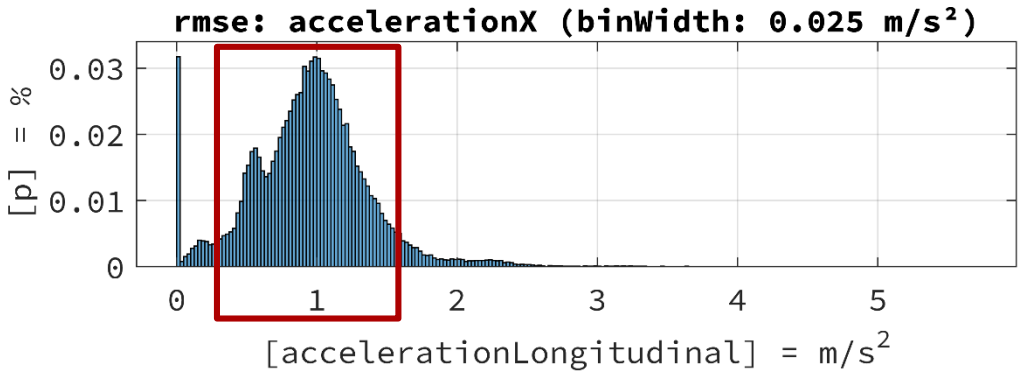


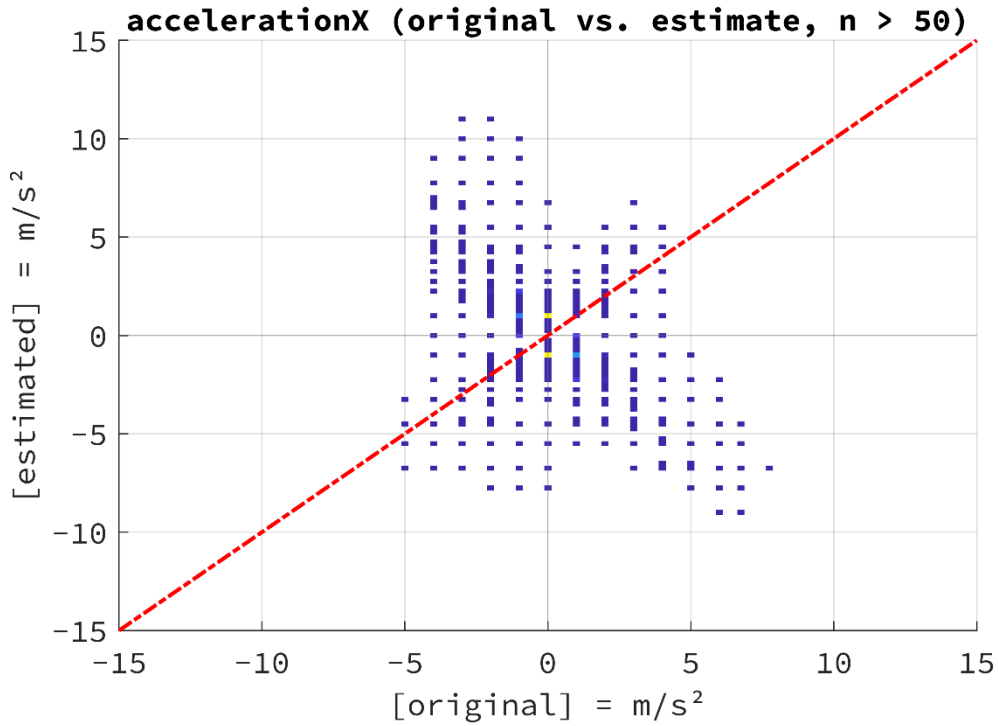


Data

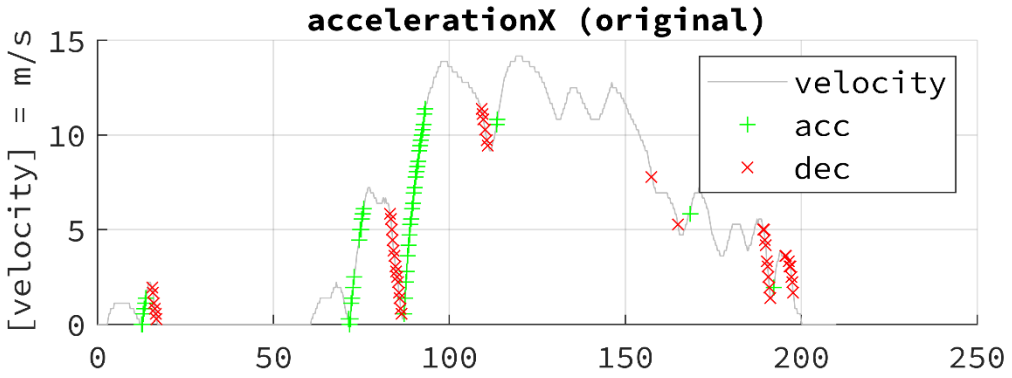
Acceleration

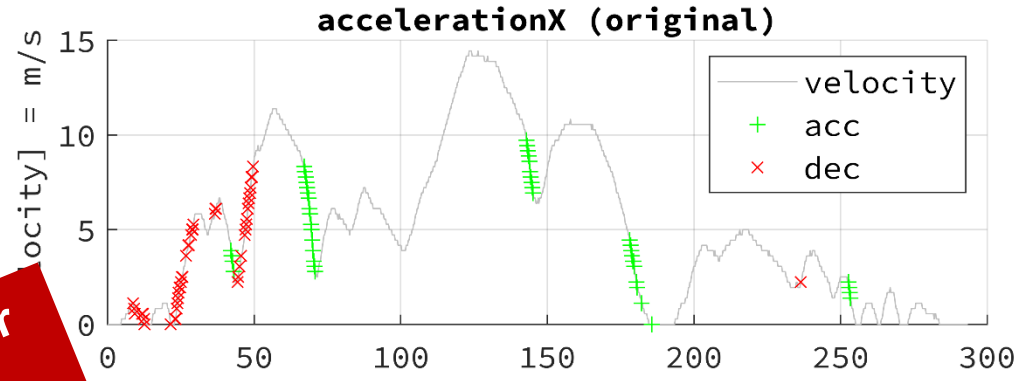
- Formula
  - $acceleration_{EST} = \frac{\partial velocity}{\partial time}$
- (expected) Value range
  - -10 m/s² : +10 m/s²

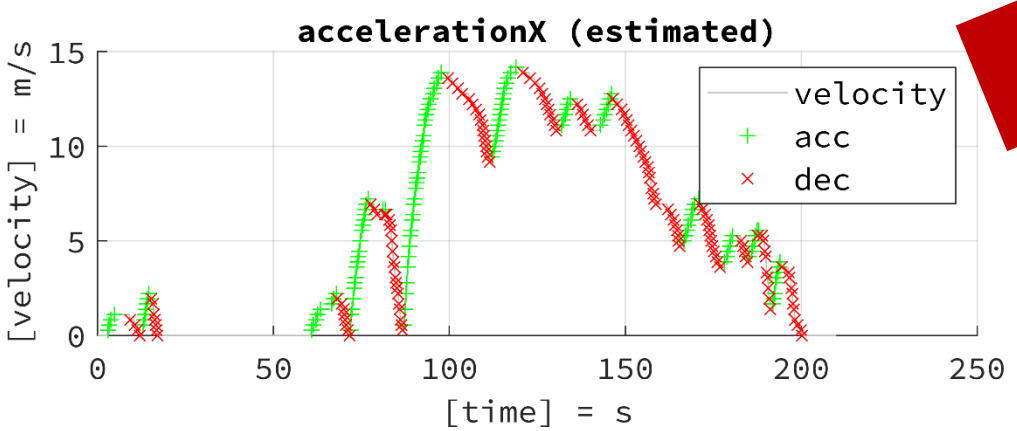


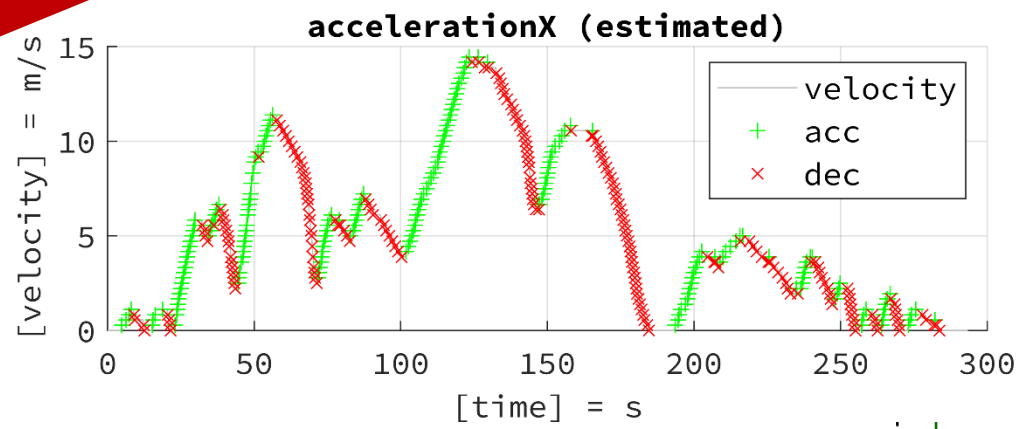


Data







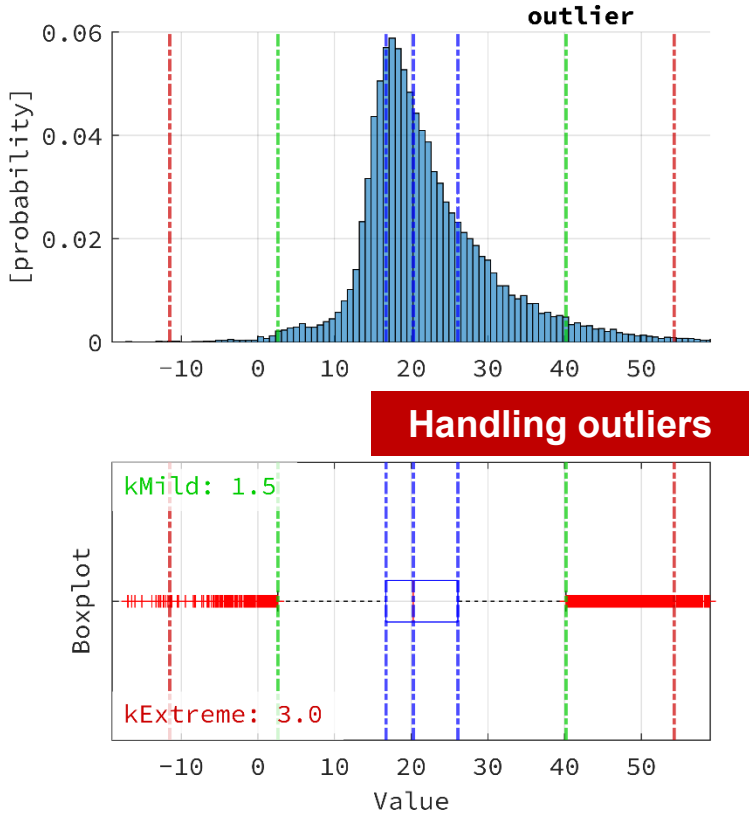


Affects other values

(fixed) Data

Steps (so far)

- Look at all data
  - RMSE values for identification
  - Original vs. Estimate histograms
  - Literature (if available)
- Look at individual data
  - Plot individual trips
- Correct / Fix
  - Your code
  - Assumptions

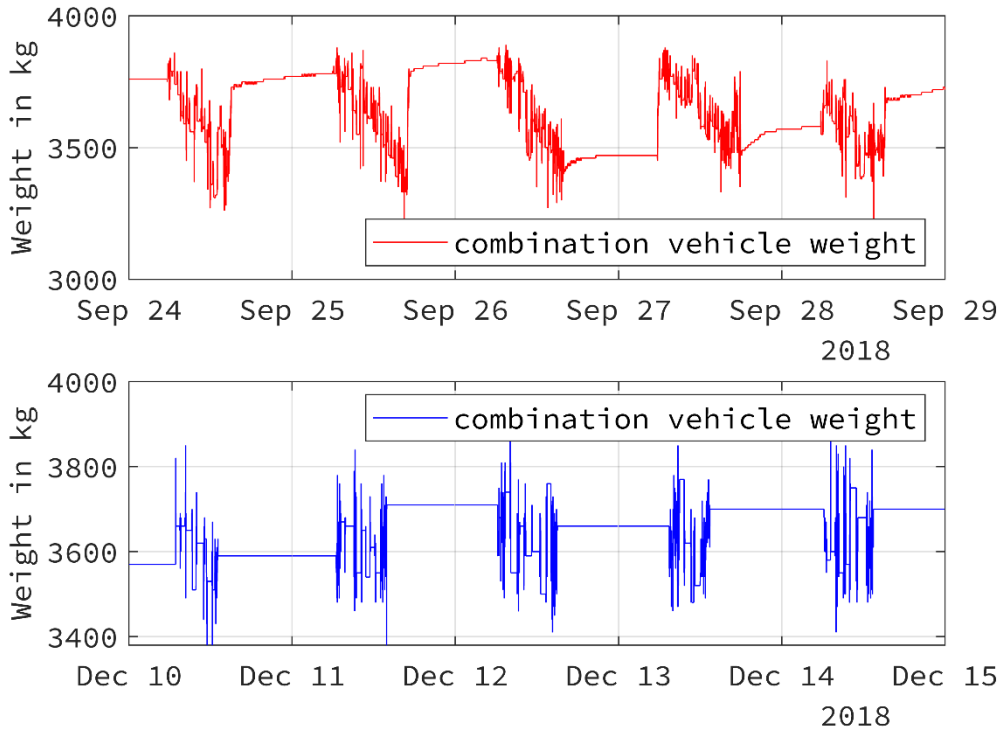


Quantile
Q_{50}: 20.3
Q_{25}: 16.7
Q_{75}: 26.1
D_{25}: 3.6
D_{75}: 5.8

Fence (Mild)
low: 2.6
high: 40.2
p: 4.2%
Fence (Extreme)
low: -11.5
high: 54.3
p: 1.0%

Info
n_{DP}: 42960
x_{Min}: -174.8
x_{Max}: 149.0
x-axis limited to Q50±4.0σ

(monitor) Data



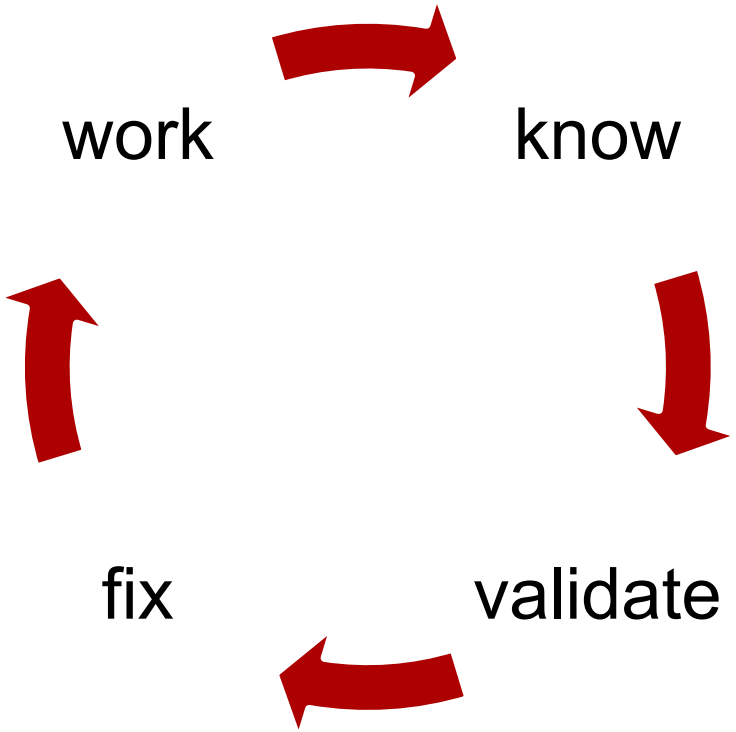
Bugs

- Sensor stopped working correctly for unknown reasons
- Bug reported in January (still not fixed)

Steps

- ?

Lessons learned



- There will be new issues
  - Re-check your (raw-)data & assumptions
- Make sure you are not the problem
  - Test your code
  - Review your code
- Make proper bug reports
  - Provide examples (figures if possible!)
  - Don’t expect the bugs to get fixed
- Documentation saves time
  - Use it whenever possible
  - Read it properly
  - Create your own (and keep it up to date)

Technische Universität Berlin

Lessons learned

- Nobody is perfect, neither is software
- Nothing is ever really finished

- Expect things to change / go wrong

- Look at your data and understand what you are looking at
- Know what your data is supposed to look like

- Never stop learning

- Decent hardware is a big help

Unexpected MATLAB lessons (over the years)

- read(datastore) doesn‘t always read the whole file
- webread() may not always return a variable (or produce an error)
- use wget() instead of mget()
- Different behaviour of the same function in different releases

- MATLAB and Linux work best from a terminal in software mode (best guess: buggy gfx driver)
- Never stop upgrading to the next release

Technische Universität Berlin

End