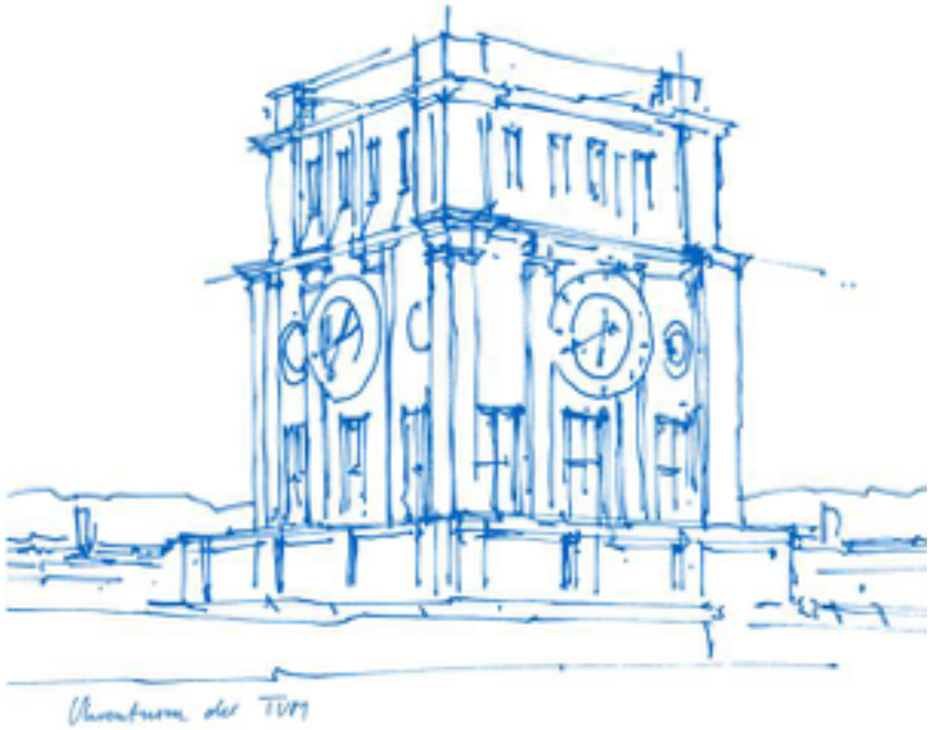
ENERGETIC MAP DATA IMPUTATION: A MACHINE LEARNING APPROACH

Master Thesis - Final Presentation
Mandy Nagy

Technical University of Munich
Department of Informatics
Chair of Connected Mobility

Supervisors:
Leonardo Tonetto (Chair of Connected Mobility, TUM)
Tobias Straub, Maxim Sidorov (BMW Group)

Garching, 14.01.2020



CONTENT

1

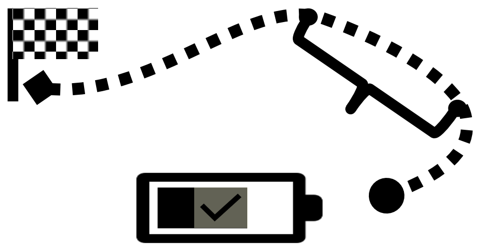
MOTIVATION

ELECTRO-MOBILITY CONTEXT

VDI nachrichten "E-mobilität: Deutsche haben Angst vor zu wenig Reichweite", 31.11.2019

Forbes "Here's Why Car Shoppers Are Still Avoiding Evs", 19.09.2019

The Atlantic "Why electric vehicles make drivers anxious?", 27.06.2019



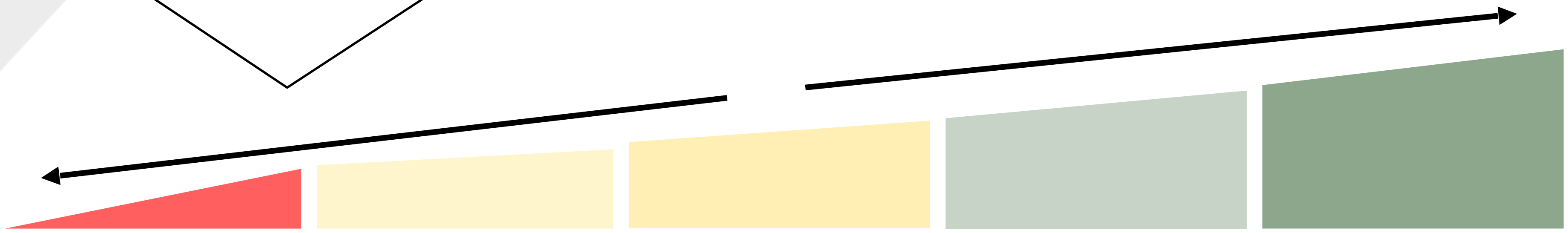
reliable energy demand prediction model

Impediments [1,2]:

- range anxiety
- inconvenient charging
- higher price

Ease:

- reliable range assistant
- specific route planning
- proactive driving support

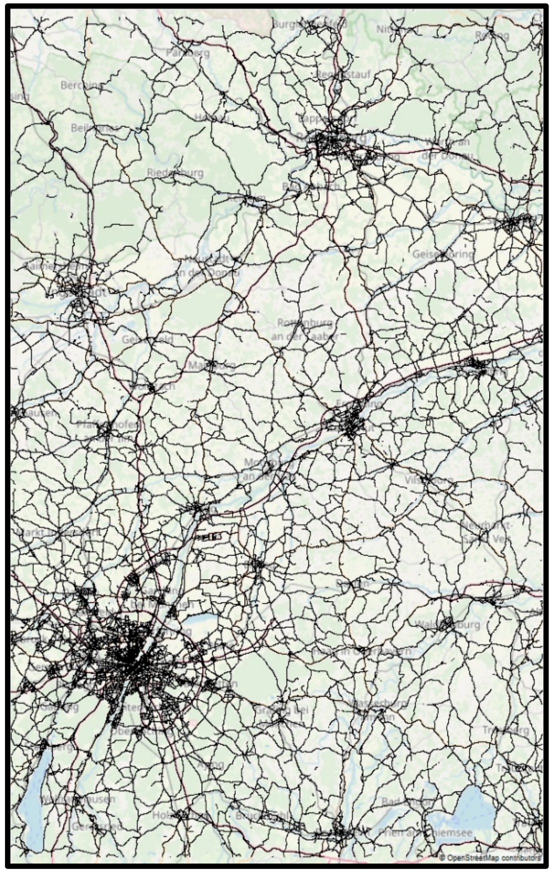


Electric vehicle acceptance level among customers

[1] Yan, Q.; Qin, G.; Zhang, M.; Xiao, B. Research on Real Purchasing Behavior Analysis of Electric Cars in Beijing Based on Structural Equation Modeling and Multinomial Logit Model. Sustainability 2019, 11, 5870.
[2] Hübner, Y.; Blythe, P.T.; Higgins, C.A.; Hill, G.A.; Neaimeh, M., Eds. Use of its to overcome barriers to the introduction of electric vehicles in the North East of England, 2012.

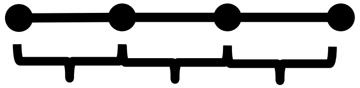
DRIVING PROFILE

vehicle sensory data	spatio-temporal segmentation	Driving Profile Map Attribute Distributions (DPMADs)

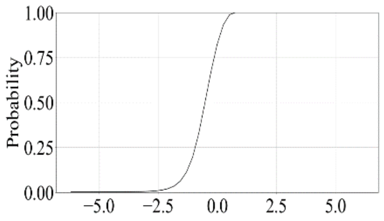
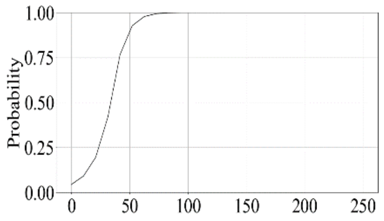
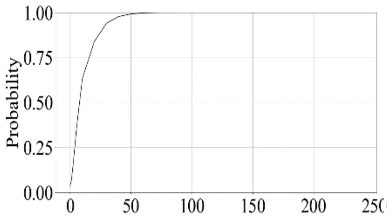
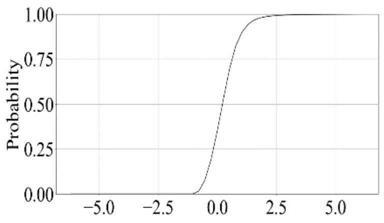
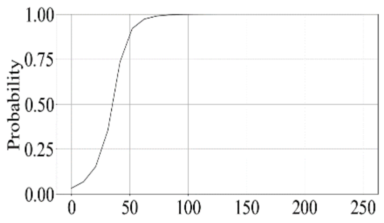


all weekdays

every 30 min

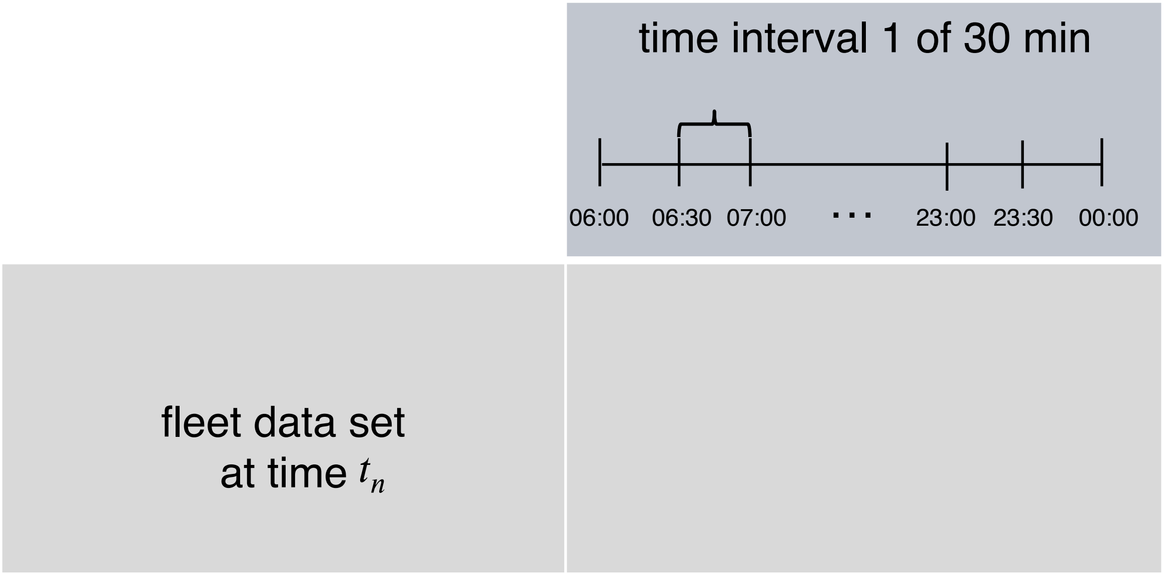


$$E_{Route,\ an\ Batterieklemme} = \sum_{Start}^{Ziel}\left[\frac{1}{\eta^{sgn(F_{Rad})}}\cdot\int_{x_n}^{x_{n+1}} \begin{pmatrix}\boldsymbol{a}(\boldsymbol{s})\\ \boldsymbol{v^2}(\boldsymbol{s})\\ cos(\alpha(s))\\ sin(\alpha(s))\end{pmatrix}^T \begin{pmatrix} m\cdot e\\ \frac{\rho\cdot c_x\cdot A}{2}\\ f_R\cdot m\cdot g\\ m\cdot g\end{pmatrix} ds + \frac{\overbrace{\Delta s_n}^{\Delta t_n}}{\boldsymbol{\bar{v}_n}}\cdot P_{NV,n}\right]_n$$

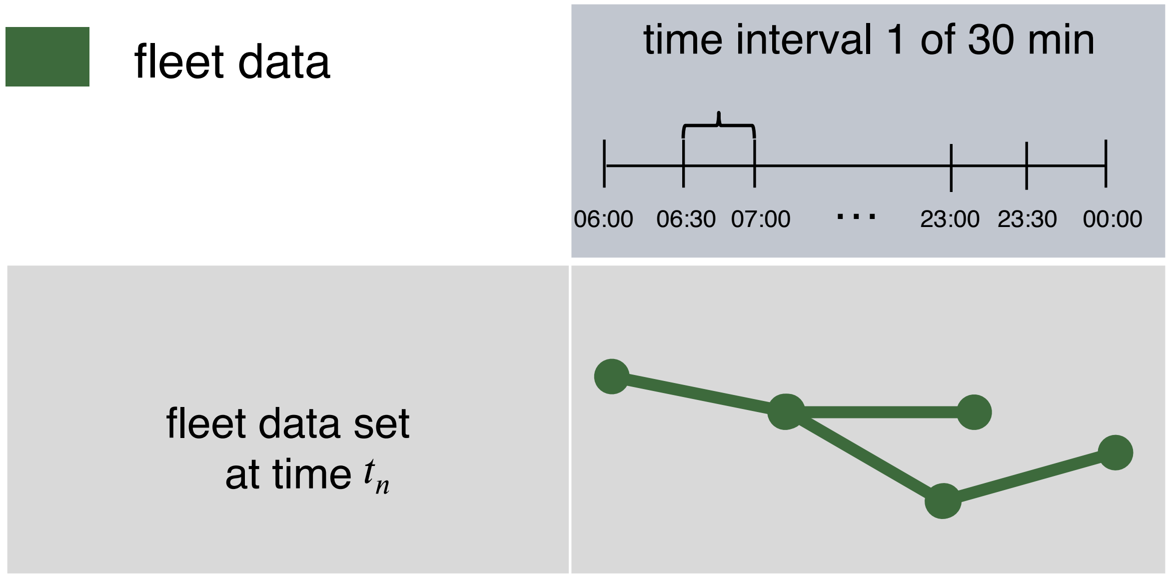
	Integral acceleration $\left(\frac{m}{s^2}\cdot\frac{m}{m}\right)$	Integral squared velocity $\left(\sqrt{\frac{km^2}{h^2}}\cdot\frac{m}{m}\right)$	Average velocity $\left(\frac{km}{h}\right)$
recuperation			
propulsion			

2
PROBLEM STATEMENT & OBJECTIVES

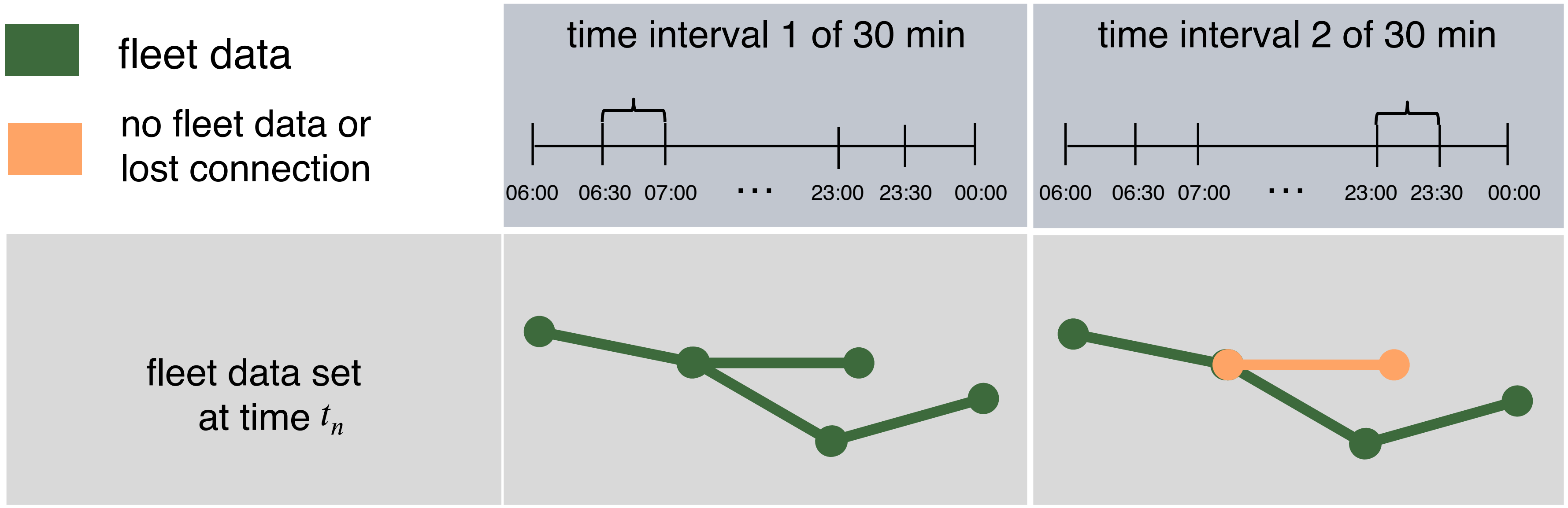
PROBLEM STATEMENT



PROBLEM STATEMENT



PROBLEM STATEMENT



Problem scenarios:

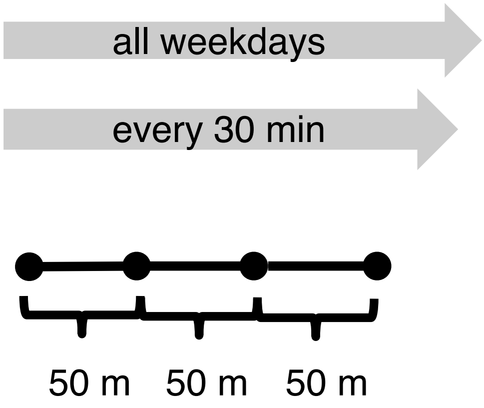
- Missing sensory data for spatio-temporal buckets
- Lost connection to the backend

PROBLEM STATEMENT

fleet data

Input

- fleet data

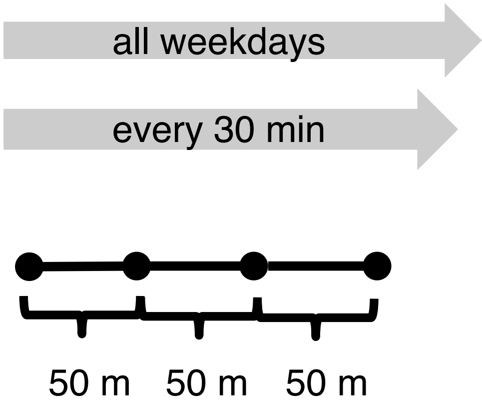


PROBLEM STATEMENT

fleet data

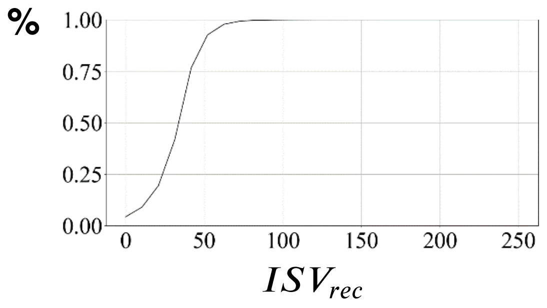
Input

- fleet data



Output

- 5 DPMADs
- a natural number n of a cluster representative (e.g. cluster #2)



PROBLEM STATEMENT

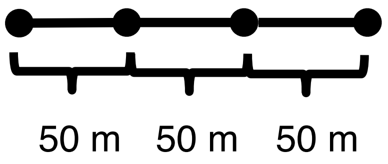
fleet data / connection

Input

?

all weekdays

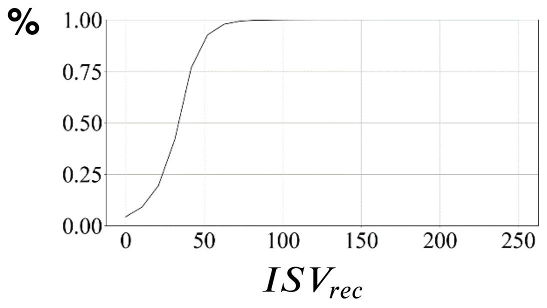
every 30 min



?

Output

- 5 DPMADs
- a natural number n of a cluster representative (e.g. cluster #2)



SOLUTION APPROACH

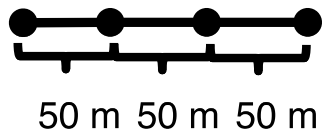
Input

Map features:

- slope
- curvature
- functional road class
- speed limit

all weekdays

every 30 min



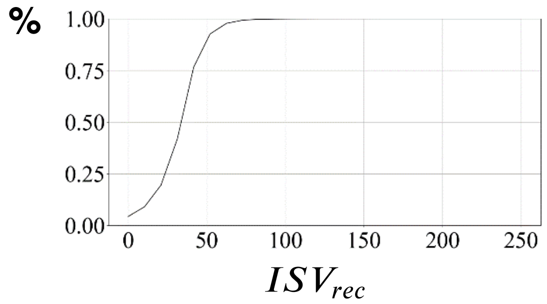
data imputation

supervised Machine Learning (ML) methods:

- Regression
- Classification

Output

- 5 DPMADs
- a natural number n of a cluster representative (e.g. cluster #2)



fleet data / connection

3

METHODOLOGY

EXPERIMENTAL SETUP

Data Set	Traces(km)	Traces (count)
training	3.503.958	95% of Munich
test (Munich)	556.135	5% of Munich
test (Leipzig)	554.366	similar to Munich

feature extraction

- slope, curvature: {min, max, average}

outlier detection

- Inter Quartile Range (IQR) method
- Q1 - 1.5·IQR < filtered output < Q3 + 1.5·IQR
- reconsider the structure of DPMADs

clustering

- addressed in previous Master Thesis [3]
- robustness of the model: DPMADs based on 3 or more measurements

regression

- Linear Regression algorithm
- main parameters:
  fit-intercept = true, normalise = false

classification

- Decision Tree algorithm
- main parameters:
  max-depth = 10, splitter="best"

comparison based on problem and test scenarios

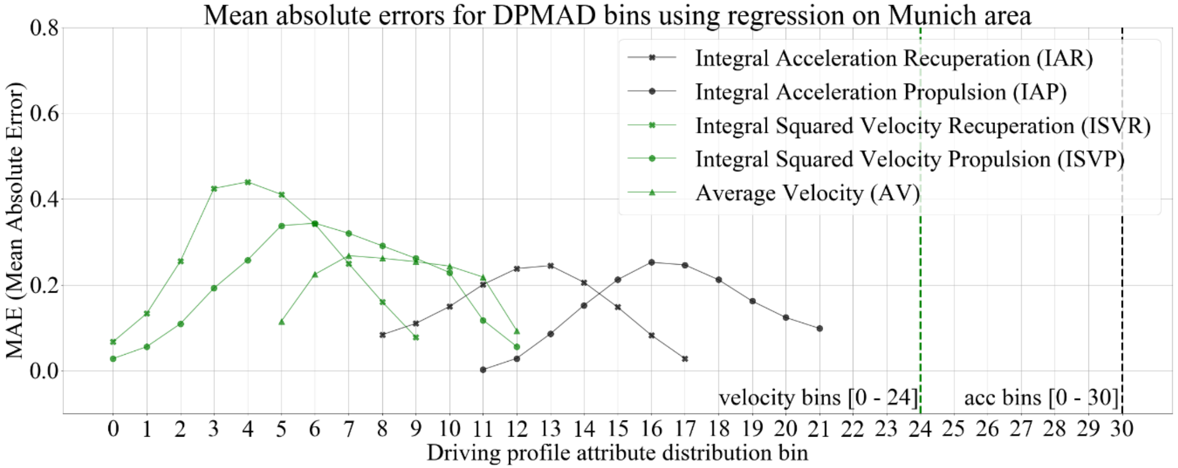
[3] Martin Kiener, Master Thesis: Clustering of Fleet Data for Energy Prediction, May 2019

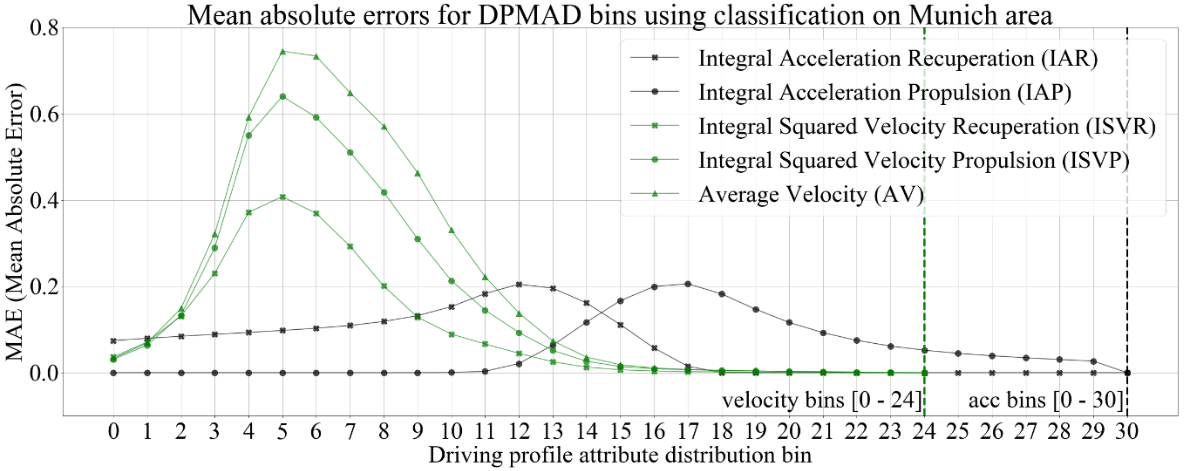
EVALUATION SCENARIOS

test scenario \ problem scenario	Lost Backend Connection	Missing Data
Leipzig area	regression ⟷ classification	
Munich area	regression ⟷ classification	regression ⟷ classification

4
RESULTS

MACHINE LEARNING PERSPECTIVE



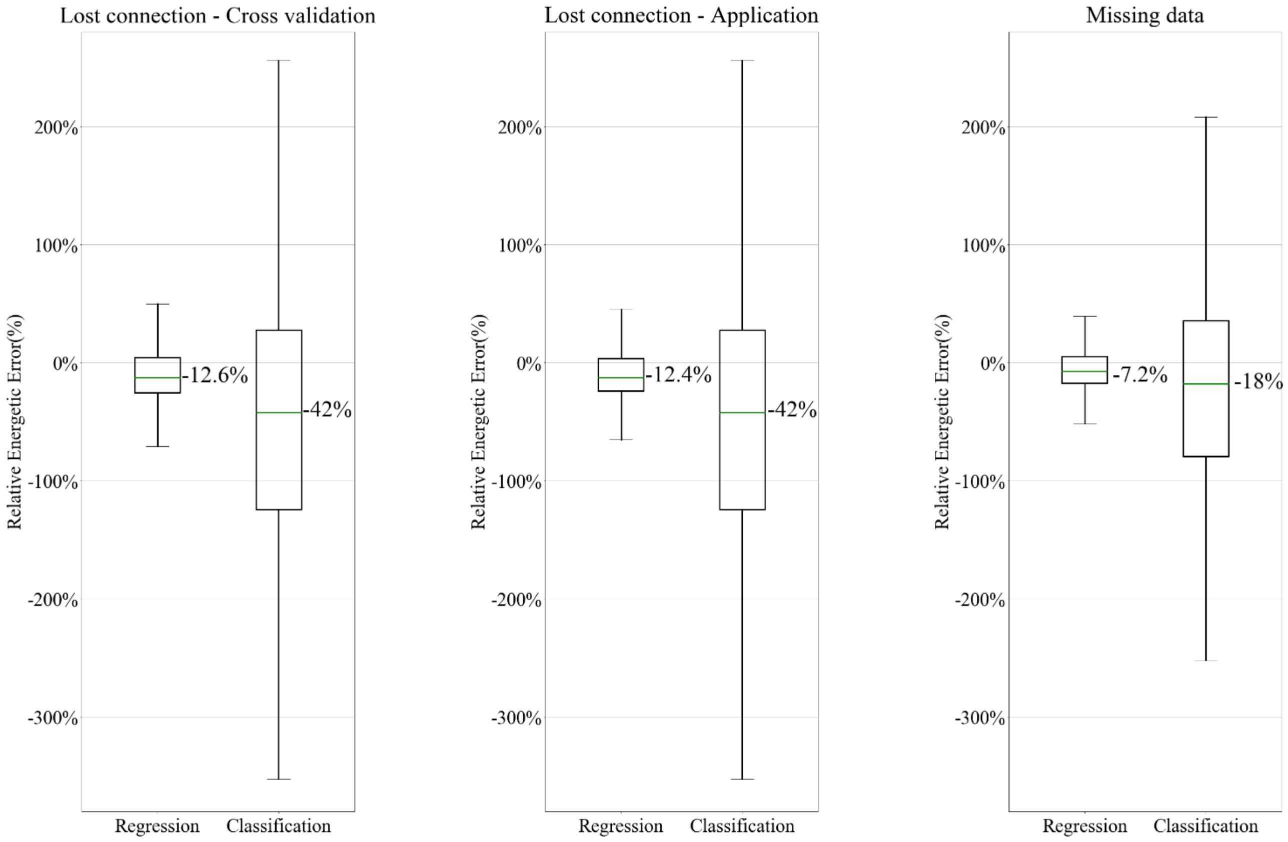


Evaluation region	Munich area		Leipzig area	
	recuperation	propulsion	recuperation	propulsion
Integral Acceleration (IA)	Clsf (-1.5%)	Clsf (-2.3%)	similar (0.0%)	Clsf (-2.1%)
Integral Squared Velocity (ISV)	Clsf (-3.2%)	Regr (-10.7%)	Clsf (-1.3%)	Regr (-12.4%)
Average Velocity (AV)	Regr (-27.2%)		Regr (-28.9%)	

- Classification model: perform better in most cases
- Regression model: bigger performance advantages

} need problem-specific interpretability

ENERGETIC PERSPECTIVE



training region	Munich area	Munich area	Munich area
testing region	Leipzig area	Munich area	Munich area
data imputed	100%	100%	30%

Problem scenarios

- Lost connection:
  - the regression model better for both cross validation and application testing sets
  - no relevant overfitting
- Missing data:
  - closest to real world situation
  - regression model outcomes within the values in literature (4-8%) [4,5]

[4] Masikos, M.; Demestichas, K.; Adamopoulou, E.; Theologou, M. Mesoscopic forecasting of vehicular consumption using neural networks. Soft Computing 2015, 19, 145–156.
[5] Sarrafan,K.;Muttaqi,K.M.;Sutanto,D.;Town,G.E. AReal-Time Range Indicator for Evs Using Web-Based Environmental Data and Sensorless Estimation of Regenerative Braking Power. IEEE Transactions on Vehicular Technology 2018, 67, 4743–4756.

5
CONCLUSIONS & OUTLOOK

5 CONCLUSIONS & OUTLOOK

System Parameter	Outlook
input features	- include further available map features - real time features, e.g weather, traffic
feature engineering	- cross correlations - higher polynomial degree
machine learning algorithms	- more sophisticated algorithms, e.g Neural Networks, Support Vector Machines - further parametric optimization

Conclusions

- regression model can be deployed in the vehicle given the achieved performances:
    - ✓ lost connection scenario (worst case): 12.6% error
    - ✓ missing data scenario: 7.2% error comparable to related works
- reliable and precise energy prediction
- raise BEVs acceptance level

THANK YOU FOR YOUR ATTENTION!