BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Spatio-temporal traffic flow forecasting on a city-wide sensor network

GIS Ostrava 2017

Felix Kunde, Alexander Hartenstein, Petra Sauer

About me

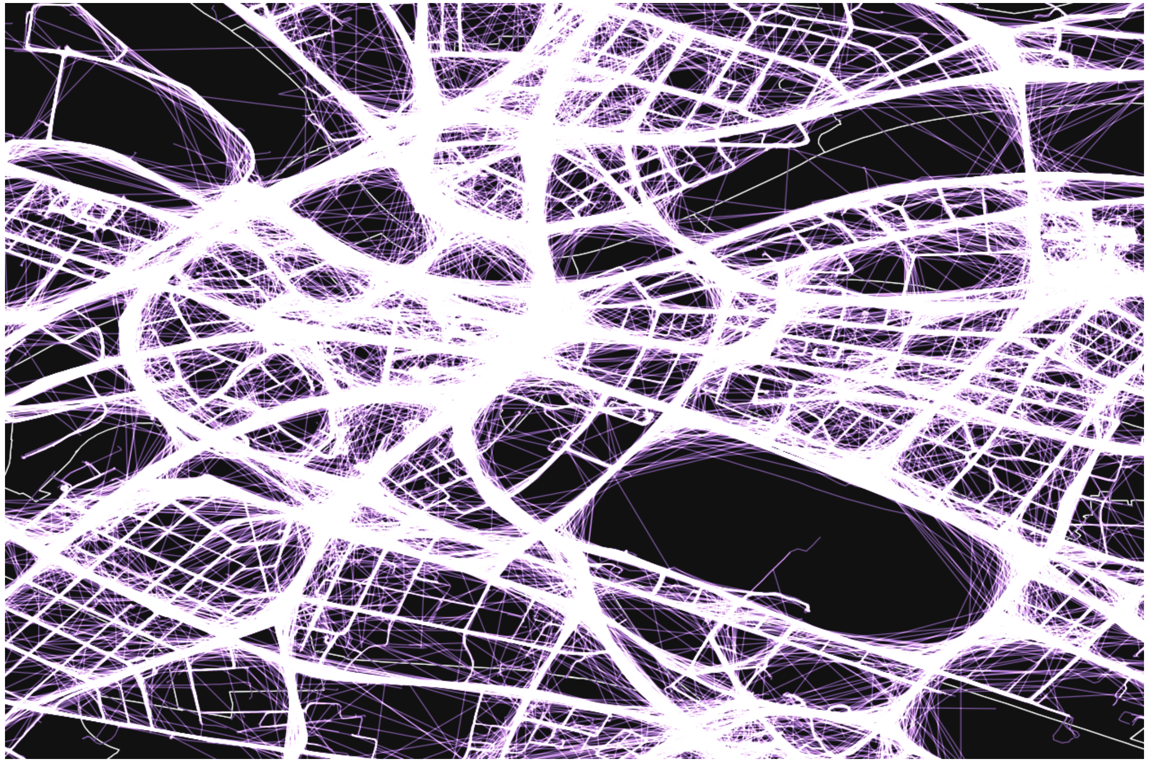
- Research assistant @BeuthHS
- Geography/GIS background
- Last 5yrs focus on spatial databases
- Core contributor to OS projects 3D City Database (CityGML) and pgMemento (Postgres versioning)
- Member of German Foss4g community
- Twitter: @FlxKu , GitHub: FxKu

Outline of this talk

- **Context:** Transportation research
- **Task:** Traffic forecasting
- **Problem:** Hardly used in the field
- **Question:** What works best?
- **My interest:** Spatio-temporal correlations
- **Methods:** Machine learning on time series
- **Results:** Where and when do we get a good prediction?
- **Discussion:** What can we tell / not tell from the results?

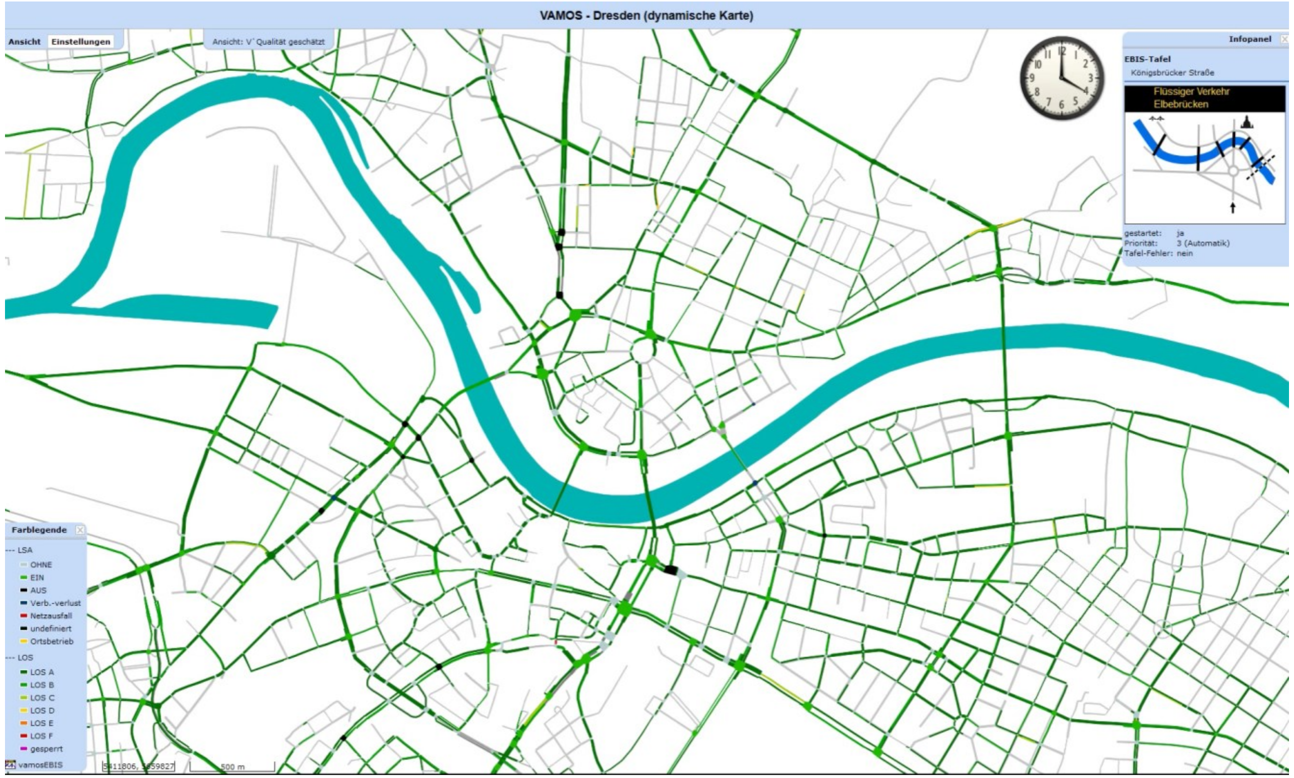
Context: ITS and Smart Mobility

- An Intelligent Transportation System (ITS) plays a crucial role in **optimizing the traffic flow** in a city
- Aims to recreate a most **realistic image of the current traffic** from **limited sample data** (static and floating sensors)
- Adopt **traffic theory to real-world** scenarios. Create programs to manage traffic flow (switching signs etc.)
- More data is always better (see Google, HERE) but usually it's **not available** to an ITS of the public sector

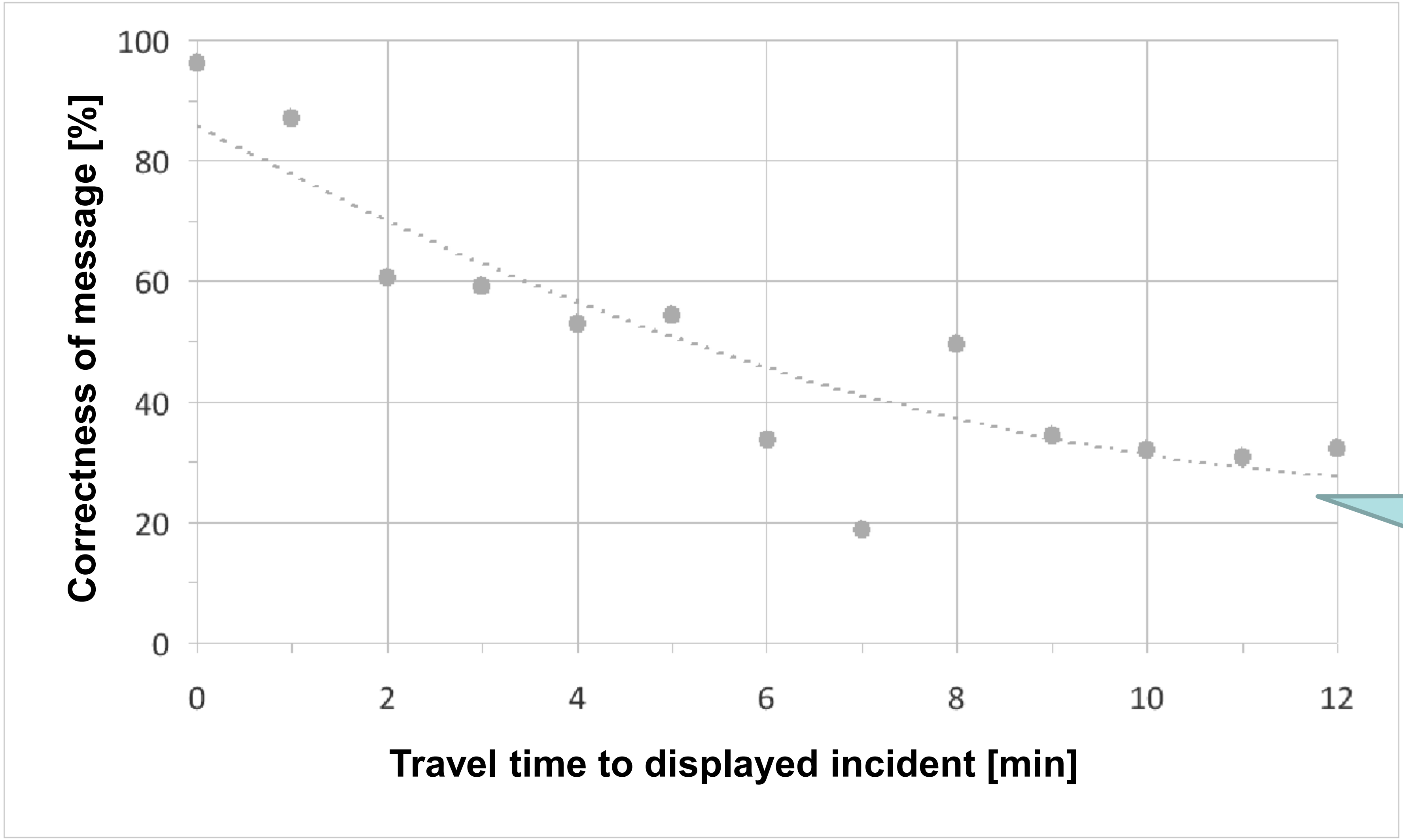


Problem: Adoption

- Sophisticated forecasting methods are not used most ITS (in Germany)
- Data aggregation and fusion is only done for monitoring reasons
- Trigger rule-based systems which are based empirical studies



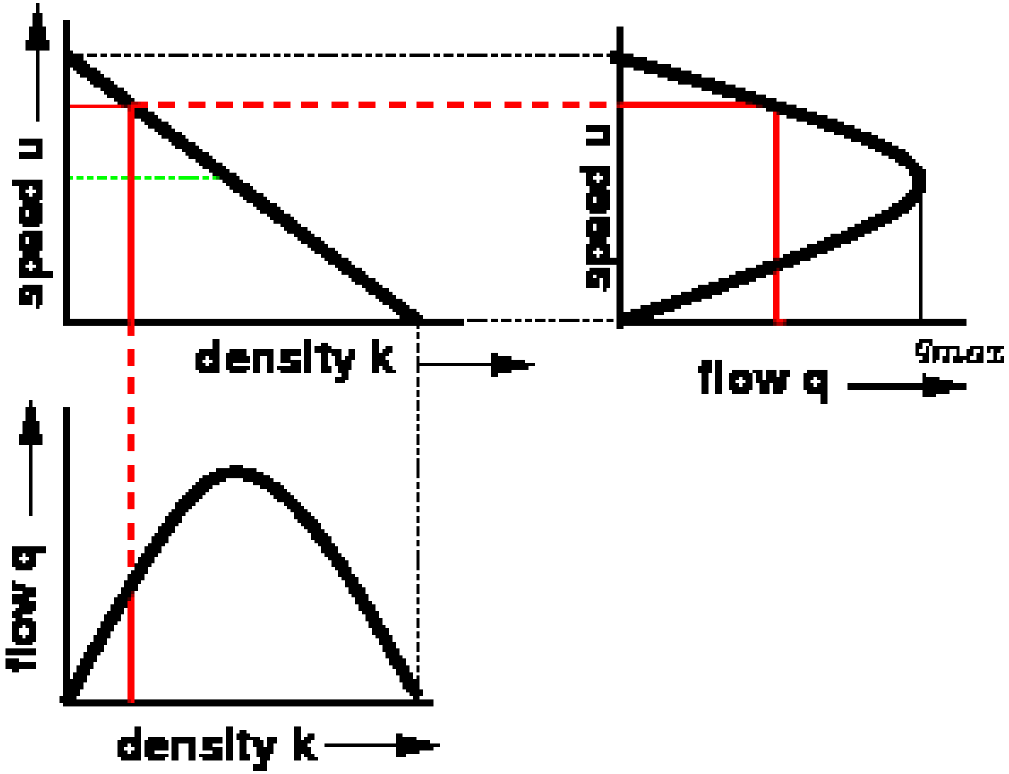
Problem: Forecasting in the field



Great uncertainty in traffic messages!

Pape, S.; Körner, M. (2016): Verkehrslageprognose unter Berücksichtigung der dynamischen Kapazitäten an LSA-abhängigen Knotenpunkten zur qualitativen Aufwertung der Verkehrslageinformation im Verkehrsmanagementsystem VAMOS

- Macroscopic properties
  - **Speed** = Avg. speed of cars per time per location / trajectory
  - **Traffic flow** = Number of cars per time per location / trajectory
  - Max traffic flow = **Road capacity** / lane / hour
  - **Traffic density** = Flow / Speed (simplified)

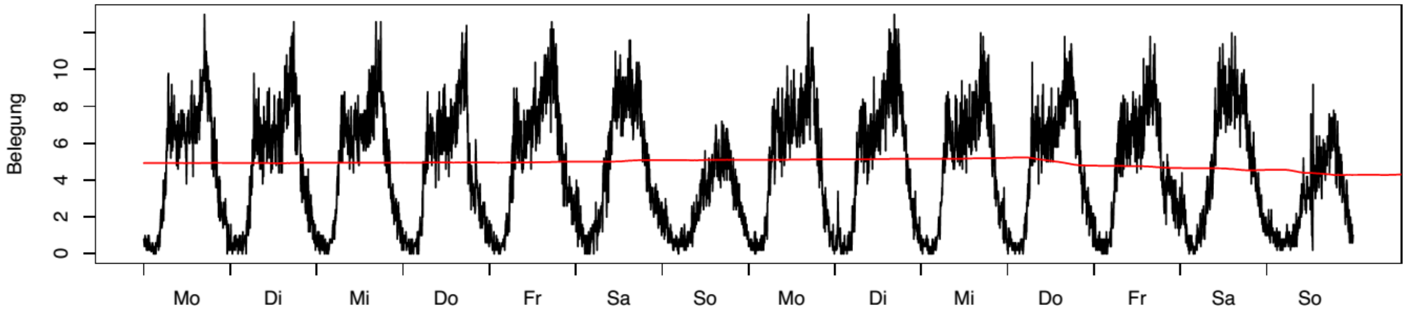


Source: https://www.civil.iitb.ac.in/tvm/1111_nptel/512_FundRel/plain/plain.html

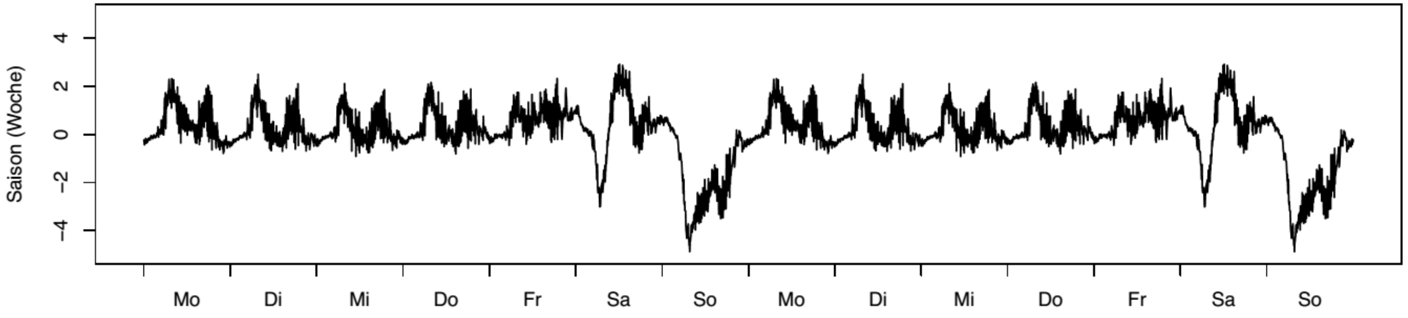
Treiber, M.; Kesting, A. (2013): Traffic Flow Dynamics – Data, Models and Simulation

Task: Traffic Forecasting – Dependencies

- Temporal dependencies
  - Many time-series analysis methods require **stationarity** = no trend or seasonal dependency (rarely given but can be reached)
  - **Autocorrelation**
  - How to deal with **missing data**? How does data imputation affect the **statistical characteristics** of the time series?



Measured occupancy (with MA 5 minutes) and global trend



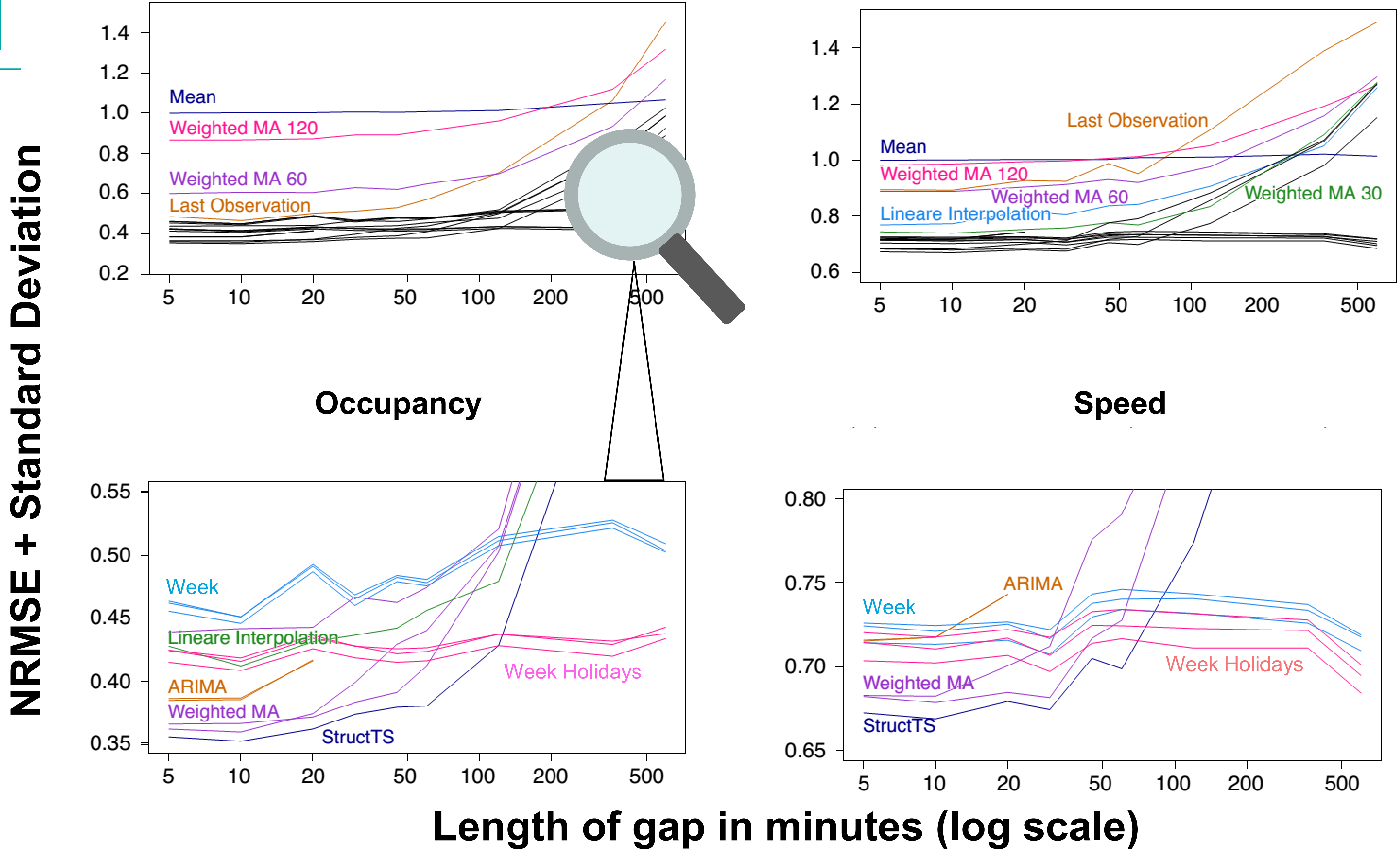
Extracted weekly season

Waterloo, S. (2017): Analyse und Bereinigung unvollständiger und fehlerhafter Messreihen von Verkehrssensoren. Bachelor Thesis.

- **Mean average error (MAE)** for imputing estimated values in gaps of 5 minutes
- **Conservation of standard deviation** using different estimation methods in R (0 worst, 1 best)

Method	Occupancy (MAE)	Occupancy (SD)	Speed (MAE)	Speed (SD)
StructTS (na.kalman)	**0,84**	0,95	**5,38 km/h**	0,66
Average week incl. holidays	0,95	0,93	5,64 km/h	0,58
Weighted Moving Average (25 min before and after gap)	**0,86**	0,93	**5,49 km/h**	0,66
ARIMA	0,90	0,97	5,71 km/h	0,65
Linear Interpolation	0,98	0,97	6,08 km/h	**0,85**
Last Observation Carried Foreward	**1,13**	**1,0**	**7,00 km/h**	**0,97**
Overall Mean	**3,32**	**0**	**10,84 km/h**	**0**

Waterloo, S. (2017): Analyse und Bereinigung unvollständiger und fehlerhafter Messreihen von Verkehrssensoren. Bachelor Thesis.



Waterloo, S. (2017): Analyse und Bereinigung unvollständiger und fehlerhafter Messreihen von Verkehrssensoren. Bachelor Thesis.

Task: Traffic Forecasting – Dependencies

- Spatial dependencies
  - **Spatio-temporal autocorrelation** in time series at different locations (e.g. STARIMA)
  - General assumption: Neighbours are having a higher impact on the target than distant detectors, but **how to select neighbours**? (geometrical, physical, mathematical)

Pfeifer, P.E. & Deutsch, S.J. (1981): Seaosonal Space-Time ARIMA Modeling.

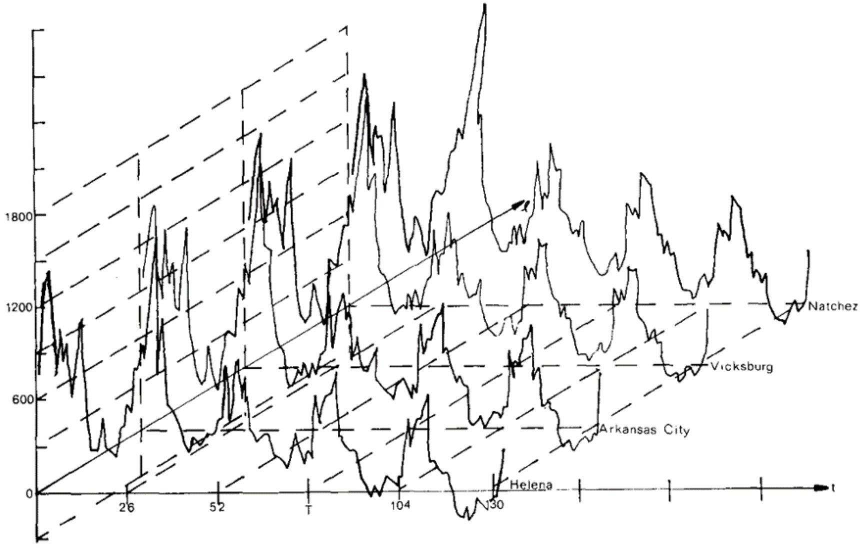
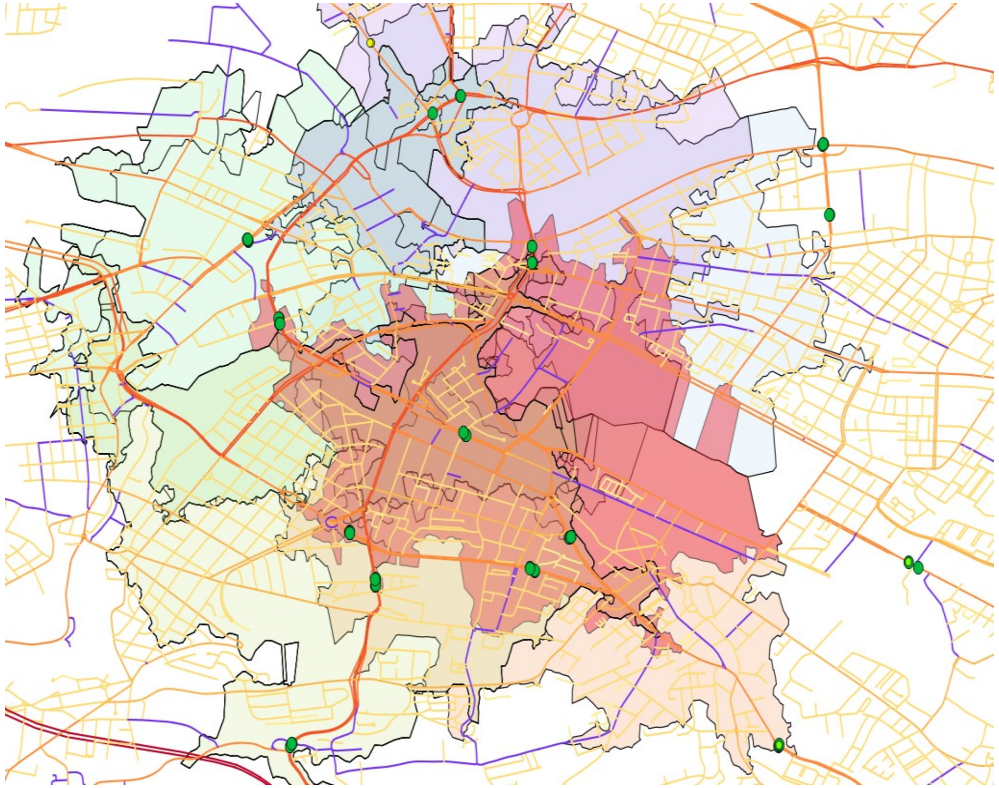


FIG. 7. Past Observations and STARIMA $(2_{11}, 0, 0) \times (0, 1, 1_0)_{26}$ Forecasts Made from $T = 79$

- Consider underlying **road network**
  - Travel times for trajectories
  - **Temporal dynamics** of neighbourhoods (see Cheng et al. 2014)

- **Spatial Weight Matrix** specifying existence and strength of spatial neighbours

	s1	s2	s3	s4	s5	s6	s7	s8	...
s1	1	1	0	0	1	1	0	0	...
s2	1	1	1	0	0	0	0	0	...
s3	0	1	1	0	0	0	0	0	...
s4	0	0	1	1	0	1	0	1	...
s5	1	0	0	0	1	0	1	0	...
s6	1	1	0	1	0	1	0	1	...
s7	0	0	0	0	1	0	1	0	...
s8	0	0	0	1	0	1	0	1	...
...	...	...	...	...	...	...	...	...	...



Task: Traffic Forecasting

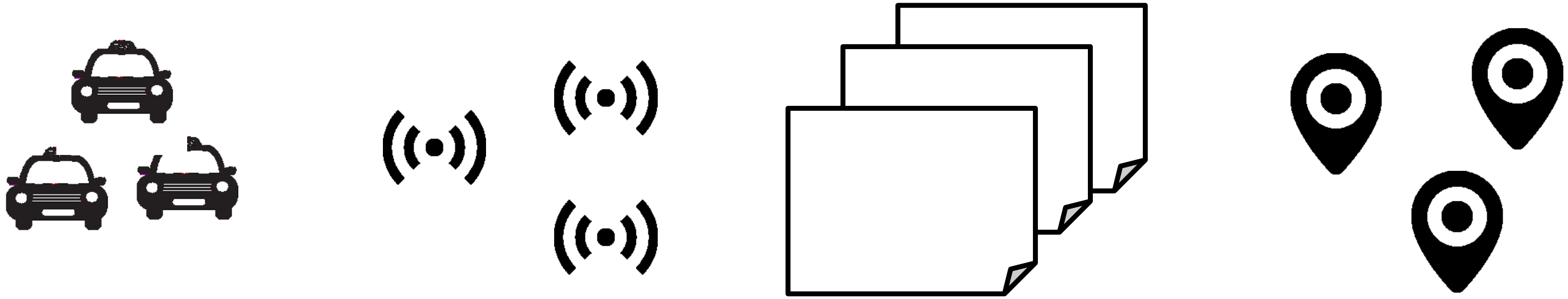
- **Trivial**: Take current value, seasonal average, random walk etc.
- **Simple**: Exponential Smoothing, Moving Average (MA)
- **Autoregressive**: ARIMA, SARIMA, VARMA etc.
- **kNN**: Past scenarios closest to recent one
- **Graphical models:** Decision Trees, Bayesian Networks, Markov Chains
- **Machine Learning**: Support Vector Machines (SVM), Artificial Neural Networks (ANN)

- What works best?

Research questions

- In the papers, usually, **only one method is engineered exhaustively** and only **compared to rather simple baseline methods** from other algorithmic families
- Non parametric approaches (SVM, ANN etc.) are regarded as superior in predicting **non-linear traffic patterns**. How big is the gap to parametric models?
- Simplified scenarios (freeway setting, low number of sensors, fixed spatial dependencies)
- **Performance and scalability** is often neglected. Would it work with >1000 sensors?
- Do we need a **combination of different methods**?

- What data do we have?



Floating Car Data from Dresden (1 year)

Sensor Data from Dresden (8 years)

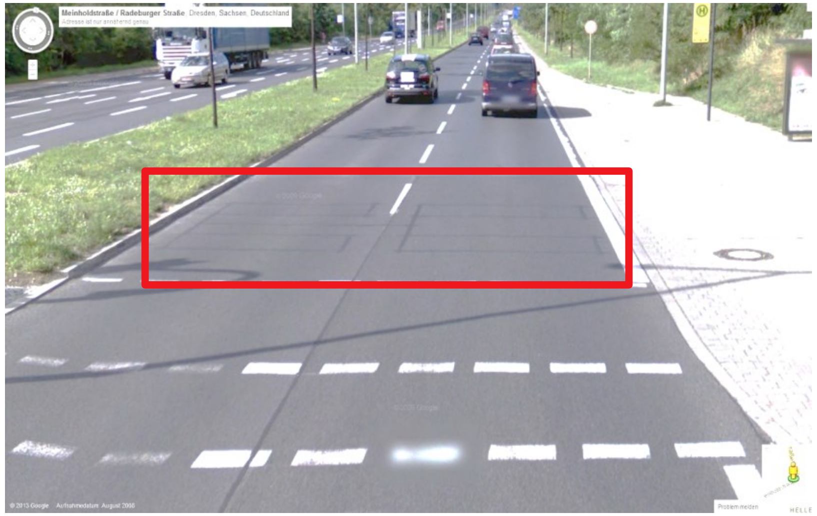
Protocol Data from Dresden (1 year)

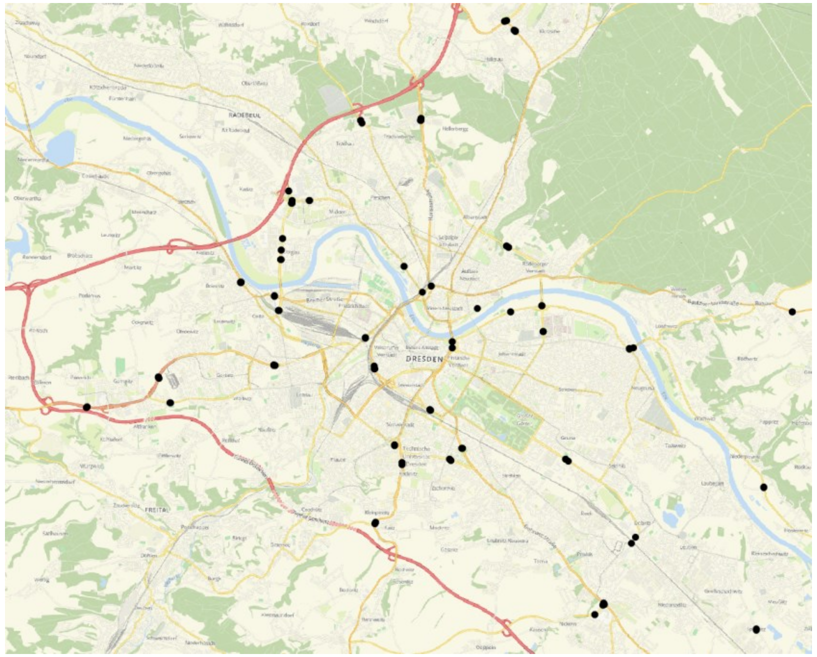
GPS-Tracking Data from ExCELL

Not present here, but you can check:
https://magda-beuthhs.github.io/magdablog/

Experimental Setup: Data selection

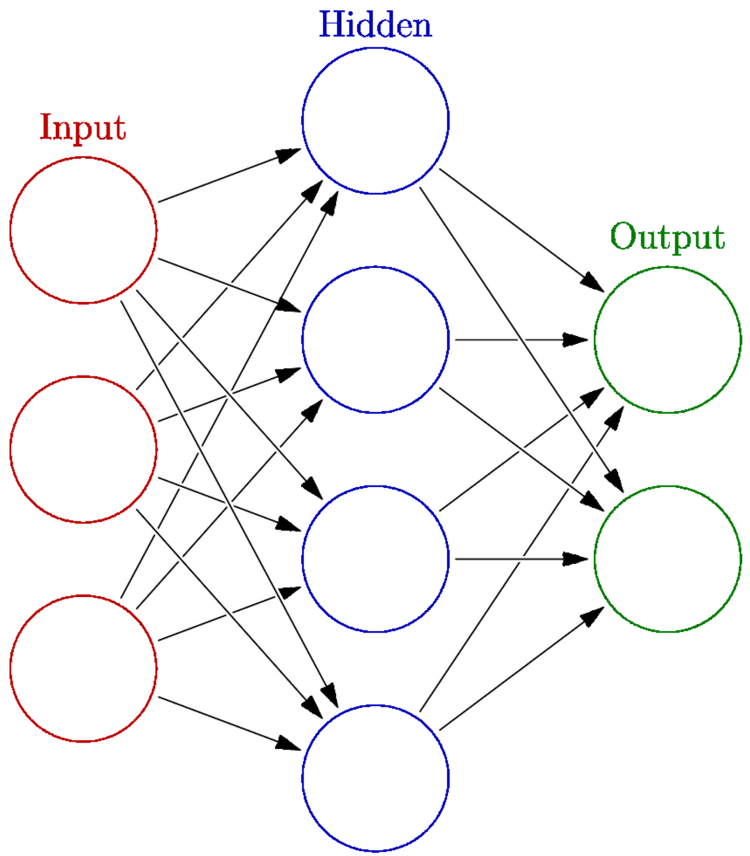
- Take data from **double inductive loops**
  - Installed on main roads with distance to waiting queues at intersections
  - Can capture the speed properly
  - Good distribution across the city
- Task: **Predict future values** (occupancy, speed) at the sensor locations (5, 10, 15, 30, 45 minutes offset)
- **Different inputs** to check spatial relevance for predictions:
  - Only historic values of the target
  - Historic values from the target's neighbours
  - Historic values from the whole data set





Experimatental Setup: FFNN

- **Data**: Occupancy of July 2015, tested against August and September 2015
- **ANN Type**: Feed Forward (FFNN)
- **Layers**: 1 input, 1 hidden, 1 output
- **Activation Function**: Sigmoid
- **Optimization**: Stochastic Gradient Descent
- **Loss Function**: RMSE
- **Other Hyperparameters**:
  - n Neurons = 59 (= number of sensors taken)
  - Learning rate = 0.01
  - Batch Size = 20
  - Iterations = 10000



Experimatental Setup: FFNN

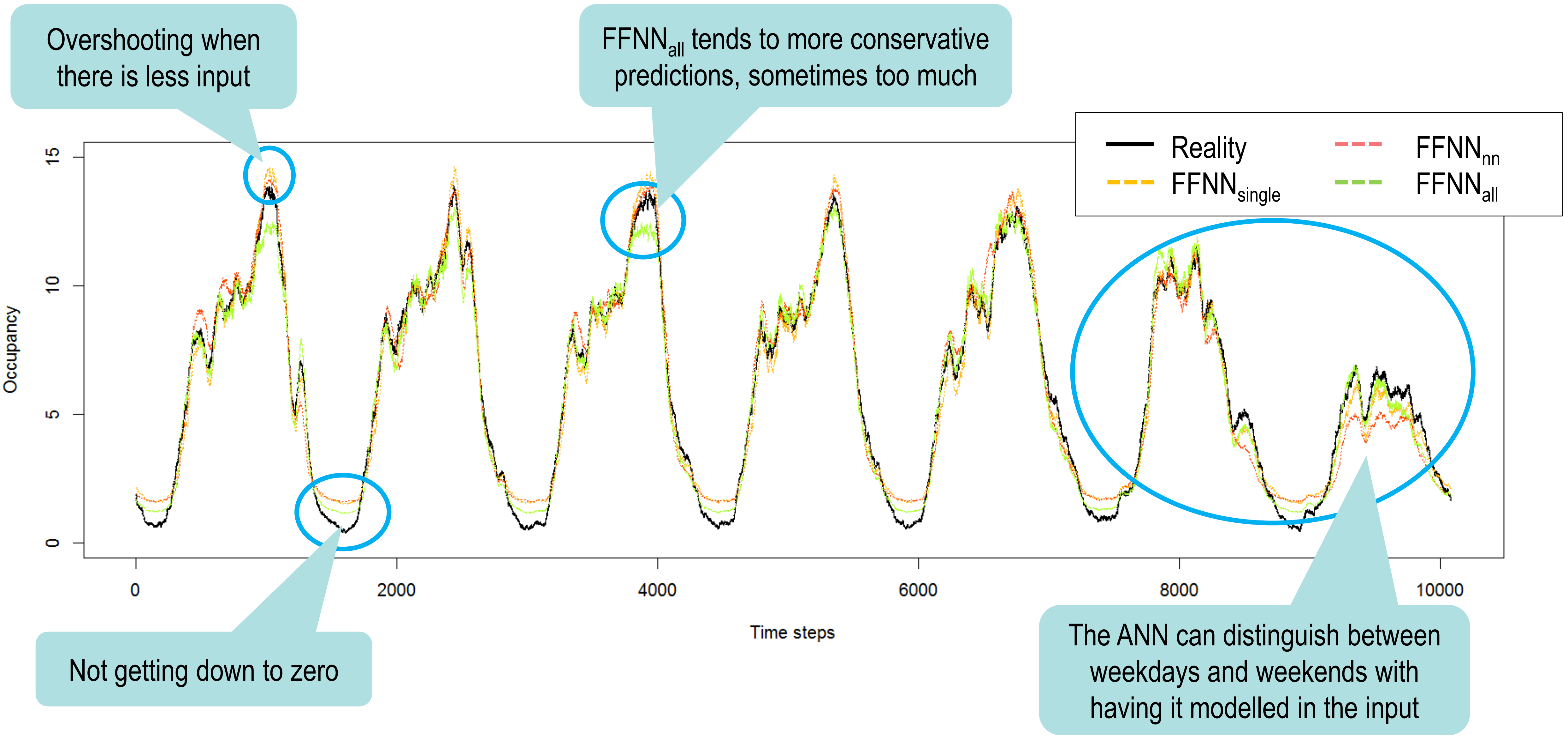
- $\text{FFNN}_{\text{single}}$ fed only **with target** sensor
- FFNN_{NN} fed only **with neighbours** of target sensor
- $\text{FFNN}_{\text{NN+}}$ fed **with neighbours incl. target** itself
- FFNN_{all} fed with values from **all sensors**

$$\begin{bmatrix} x_{s_1,t_0} & \cdots & x_{s_n,t_0} \\ \vdots & \ddots & \vdots \\ x_{s_1,t_n} & \cdots & x_{s_n,t_n} \end{bmatrix} \begin{bmatrix} x_{s_1,t_{0+offset}} & \cdots & x_{s_n,t_{0+offset}} \\ \vdots & \ddots & \vdots \\ x_{s_1,t_{n+offset}} & \cdots & x_{s_n,t_{n+offset}} \end{bmatrix}$$

Input Matrix *Target Matrix*

- Testing environment:
  - **Hadoop Cluster** for SVMs (Apache Spark, LibSVM)
  - **CUDA node** for ANNs (Tensor Flow, DeepLearning4J)
  - **In-Memory DB** node for storing sensor data (Exasol DB)

Results: One exemplary week of August



Results: MAEs of different methods

Table 1. MAE for different prediction horizons for one sensor

FFNN type	t5	t10	t15	t30	t45
$FFNN_{single}$	0,5042	0,5743	0,6679	0,9698	1,2970
$FFNN_{NN}$	0,6255	0,6660	0,7248	0,9831	1,0299
$FFNN_{NN+}$	1,7177	1,7157	1,7236	1,7813	1,8928
$FFNN_{all}$	0,4975	0,4933	0,5262	0,7474	1,0299
$mFFNN_{single}$	0,4324	0,5213	0,6137	0,9360	1,2771
$mFFNN_{NN}$	0,5616	0,6080	0,6667	0,9231	1,2519
$mFFNN_{NN+}$	1,6420	1,6543	1,6862	1,8126	1,9612
$mFFNN_{all}$	**0,3998**	**0,4086**	**0,4521**	**0,6405**	**0,8684**

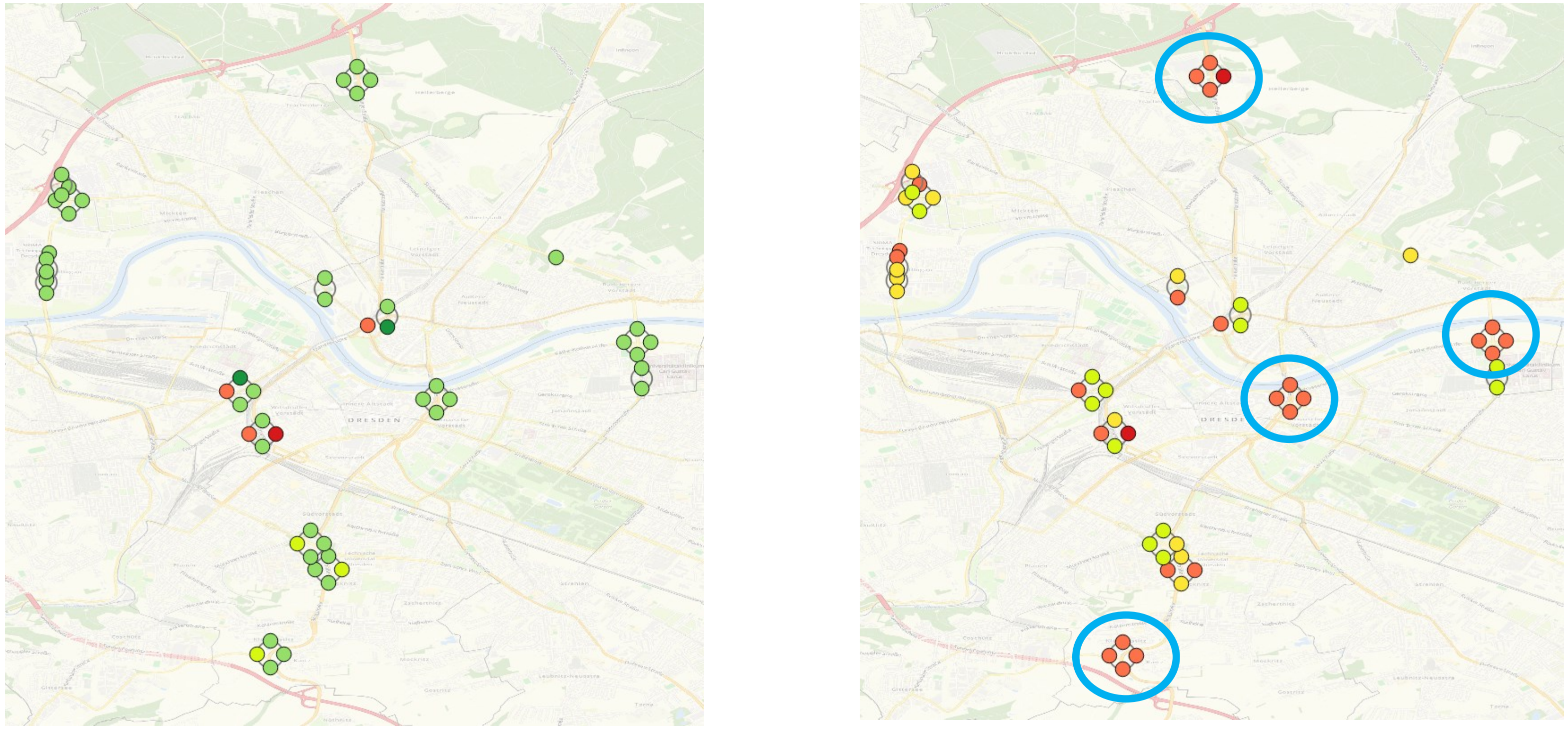
Here, we modelled a sequence into the FFNN input matrix

- Having information about neighbours did **not improve** the predictions compared to $FFNN_{single}$
- For **short-term predictions** the most simplest FFNN is fine enough
- With all sensors (and sequence) the FFNN gives the best predictions

Discussion

MAE 0 – 0.50 | 0.50 – 0.75 | 0.75 – 1.00 | 1.00 – 1.25 | 1.25 – 1.50 | 1.50 – 3.00

$FFNN_{single}$



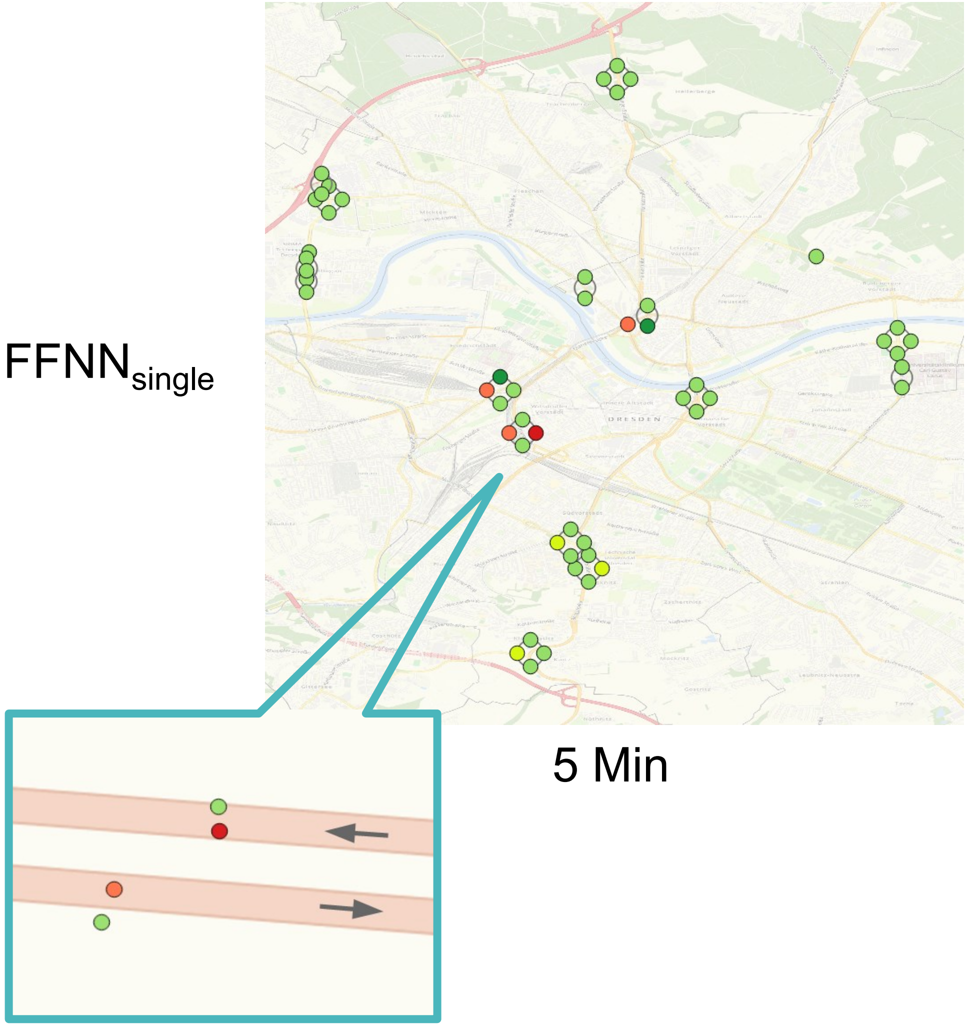
5 Min   45 Min

- Prediction error generally grows with temporal offset (expected)
- It is harder to predict future values (45 min) at locations with more regular traffic flow (main roads closes to highways and bridges)

Discussion

MAE ● 0 – 0.50 ● 0.50 – 0.75 ● 0.75 – 1.00 ● 1.00 – 1.25 ● 1.25 – 1.50 ● 1.50 – 3.00

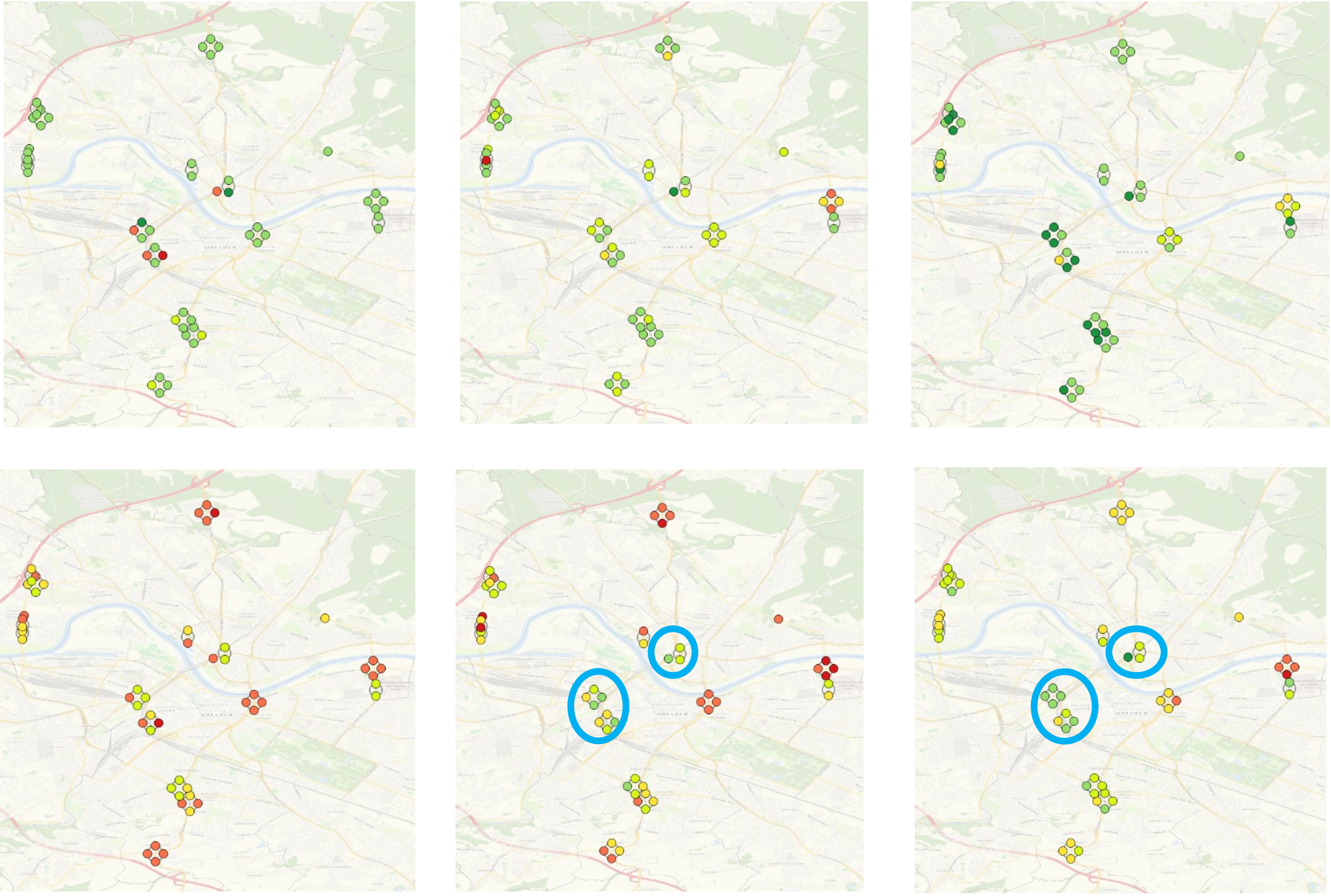
$\text{FFNN}_{\text{single}}$



5 Min

- When only regarding the target itself the error can be higher on the inner lane
- Higher irregularity of road occupancy on the “faster” lane

Discussion



- How is error distributed with other FFNN setups?
- As seen in the results for one sensor (table), adding more information to the inputs vectors, pays off for **long-term predictions**
- Again, higher errors can be seen at the same locations with some exceptions
- Sensors near intersections with potentially more irregular traffic work well with FFNN$_{NN}$ and FFNN$_{all}$

Discussion

- Traffic in Dresden is **very regular** – it is easy to **overfit**
- More **refinement** is necessary, e.g. introduce **more dimensions to the input vectors** (daytime, weekday, holiday, weather conditions etc.)
- Yet, not a strong focus on how good the predictions work at **predicting anomalies** in traffic. Requires a **labeling process** in advance. How much **variance** can be found?
- GIS and ITS specialists want to gain new insights about **spatio-temporal correlations**. Can we define **patterns** with a black box approach such as the ANN represents?
- How to install a **continuous learning** algorithm that adopts to changes in traffic

Future Work

- Extend the forecasting test to compare different approaches
  - **STARIMA** – *WIP*
  - **SVM** – *WIP*
  - Feed Forward Neural Network (**FFNN**) – *first results (see paper)*
  - Long Short-Term Neural Network (**LSTM**) – *WIP*

- Feature engineering to have a labeled input data set for traffic congestion patterns

- Test the developed models in the field against real-time data from the VAMOS ITS

Thank you for your attention!

Questions?

The work was supported by the Federal Ministry for Economic Affairs and Energy (BMWi) under grant agreement 01MD15001B (Project: ExCELL).

Gefördert durch:

Bundesministerium für Wirtschaft und Energie

aufgrund eines Beschlusses des Deutschen Bundestages

GIS Ostrava 2017

Felix Kunde

fkunde[at]beuth-hochschule.de