Prediktions- och Scenariobaserad Trafikledning (POST)

- Clustering for Scenario Evaluation

David Gundlegård, Matej Cebecauer (KTH)

Erik Jenelius (KTH), Clas Rydergren (LiU), Rasmus Ringdahl (LiU), Joakim Ekström (LiU), Wilco Burghout (KTH)

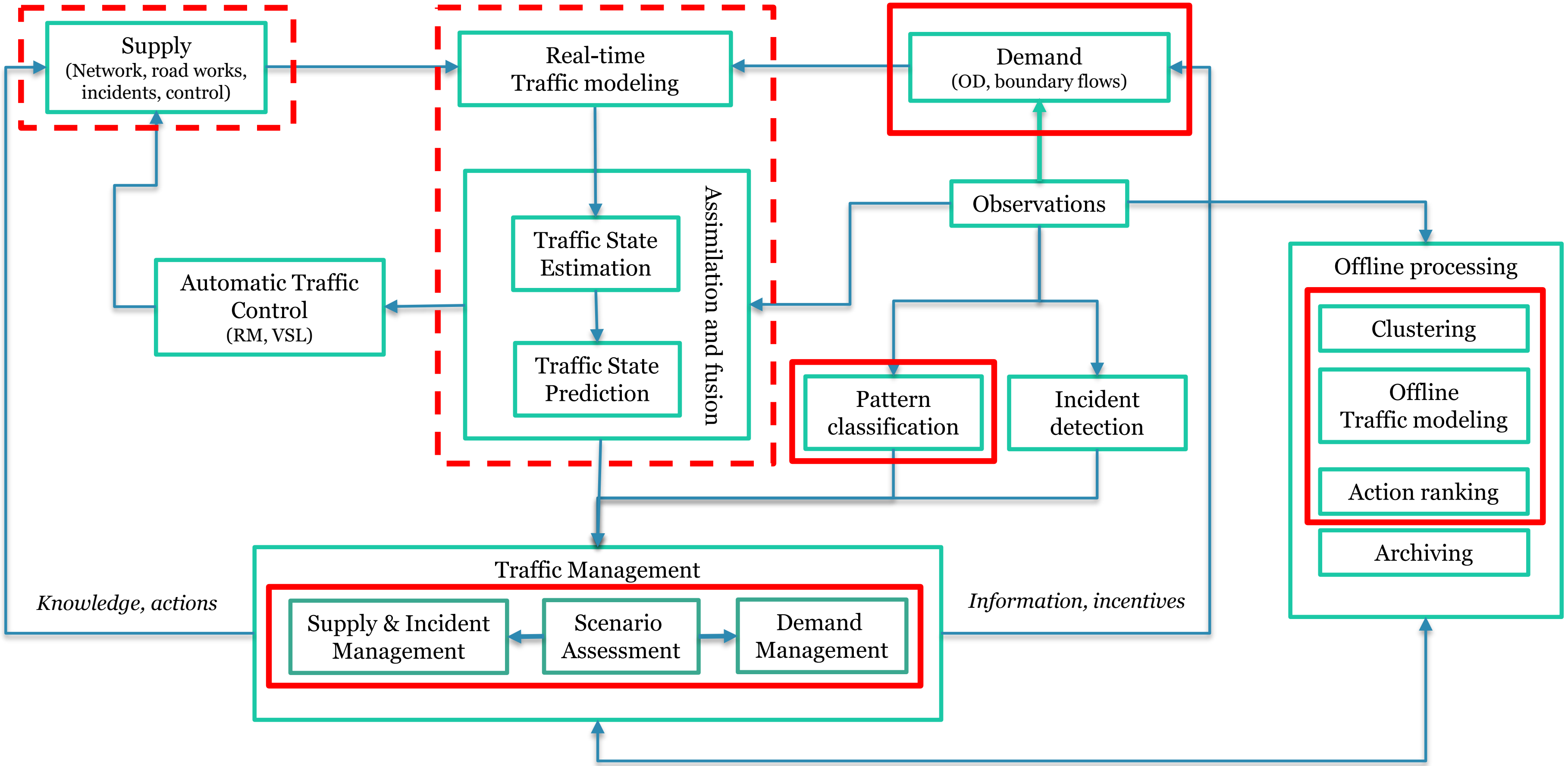
Trafik Stockholm, TrV STRESS

Sweco, UC Berkeley

Project Background

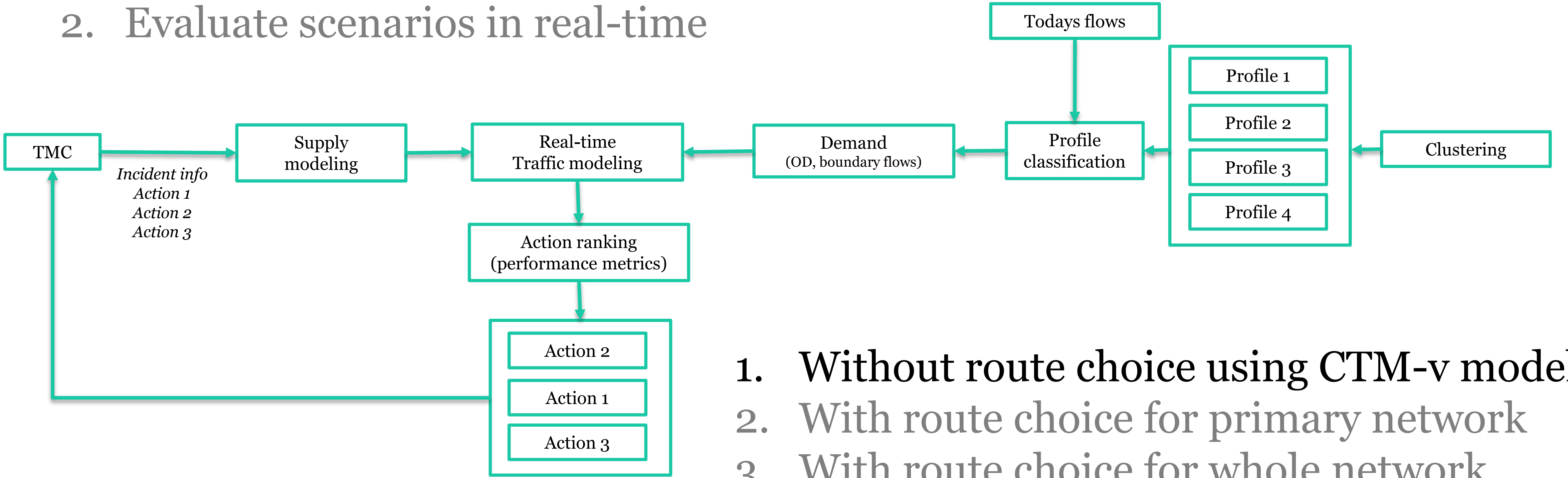
- Predict demand and route choice for scenario evaluation and action ranking
  - Offline processes for demand prediction and scenario evaluation
  - Online processes for classification of traffic situation and choice of control measure
- Example
  - Tow directly or after peak
  - Early information to travelers of severe incidents (i.e. do not use car)
- In this presentation focus on clustering methods for demand prediction

POST - Overview



Framework for Scenario-based Traffic Management

1. Offline för historical events
2. Evaluate scenarios in real-time



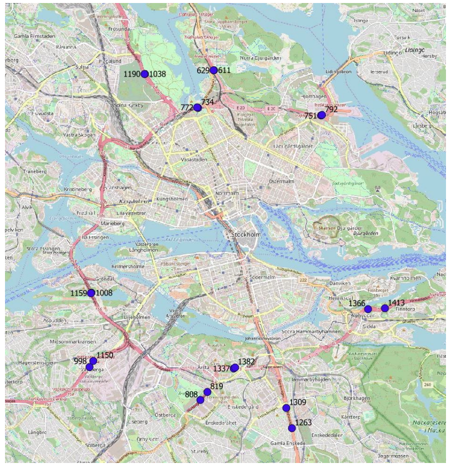
1. Without route choice using CTM-v model
2. With route choice for primary network
3. With route choice for whole network

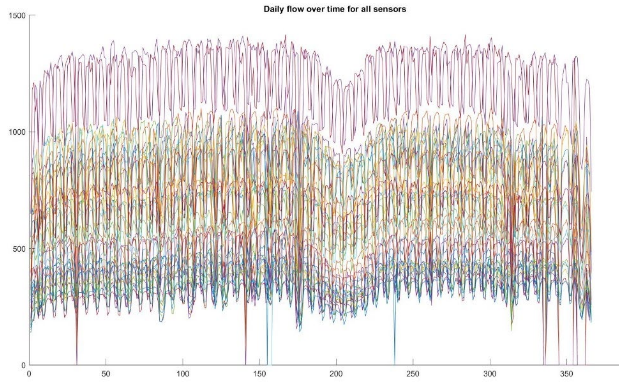
Demand Prediction

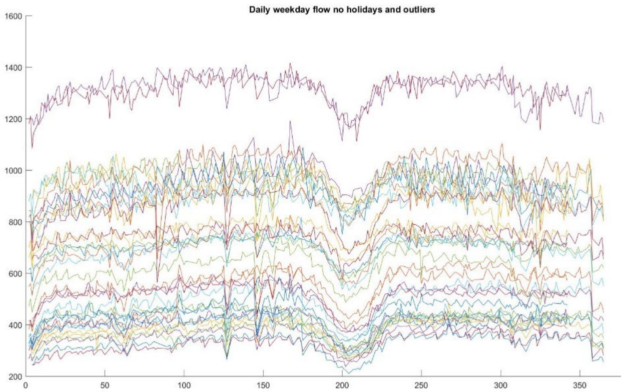
- To enable evaluation of traffic control measures we need to estimate (offline evaluation) and predict (online evaluation) demand
- Clustering + Classification VS Time series analysis
  - Important component to get an understanding of traffic patterns
  - With distinct traffic patterns we can determine how control measures perform for different scenarios patterns
- 1) Clustering
  - What type of data shall we cluster and at which aggregation level?
  - Which method shall we use for clustering?
  - How many clusters should we have?
- 2) How well does clustering-based demand prediction work?

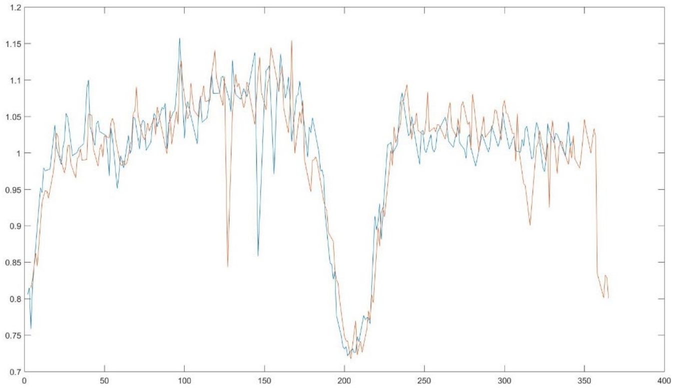
Datasets and Preprocessing

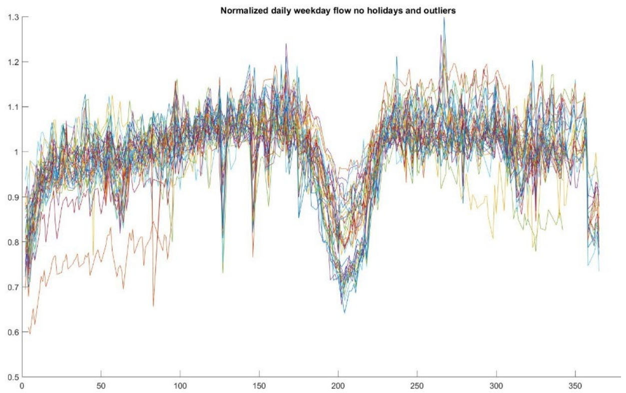
- Special days
  - Nyårsdagen
  - Trettondagen
  - Skärtorsdag
  - Långfredag
  - Påskafton
  - Påskdagen
  - Annandag påsk
  - Valborg
  - Kristi himmelsfärd
  - Klämdagar
  - Nationaldagen
  - Skolavslutning
  - Studenten
  - Midsommarafton
  - Midsommardagen
  - Midsommarsöndag
  - Julafton
  - Juldagen
  - Annandag jul
  - Nyårsafton
- Special periods
  - Januaridagar
  - Sportlov
  - Påsklov
  - Sommar/semester
  - Juli
  - Höstlov
  - Mellandagar











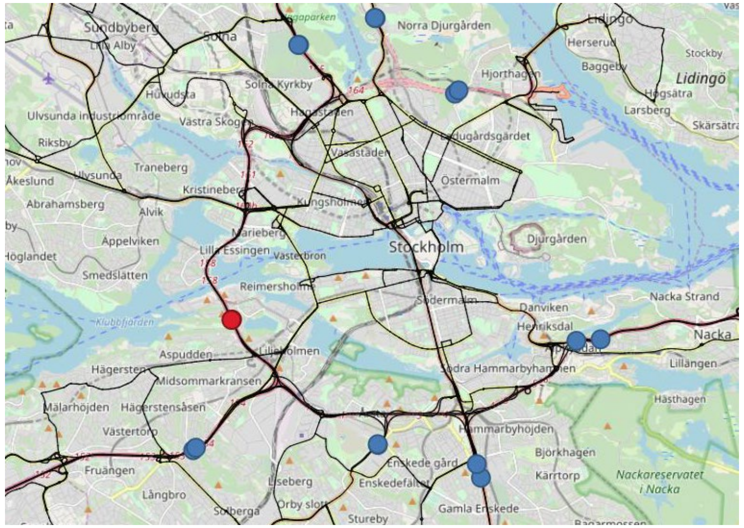
LINKÖPINGS UNIVERSITET

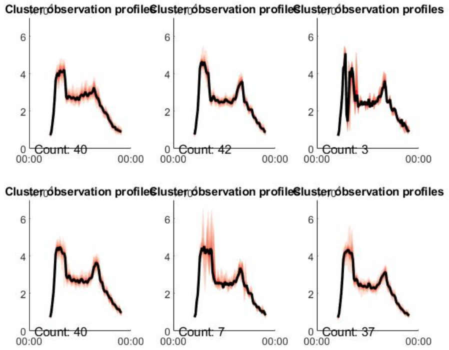
KTH VETENSKAP OCH KONST

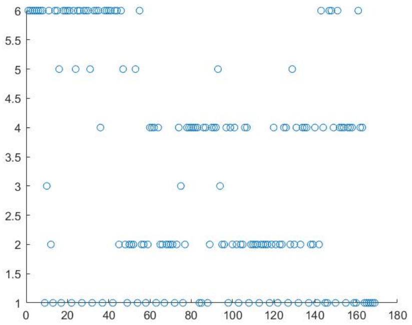
CTR

Datasets and Preprocessing

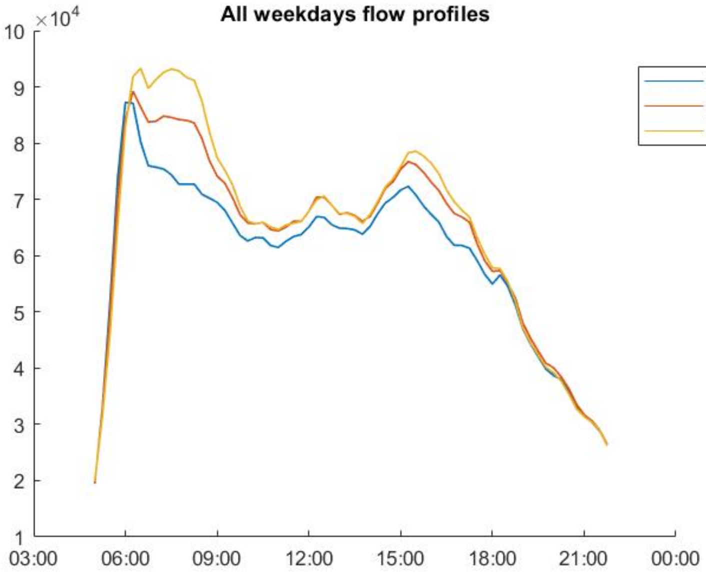
- 14 selected sensors on 7 major roads in Stockholm
- 2 sensors on each direction of Essingeleden
- Speed and flow aggregated to 15 minute intervals from 05-22
- Days with missing time intervals removed
- Special days, weekends and holiday periods removed
- Days with incidents are (still) included
- Remaining: 169 regular weekdays

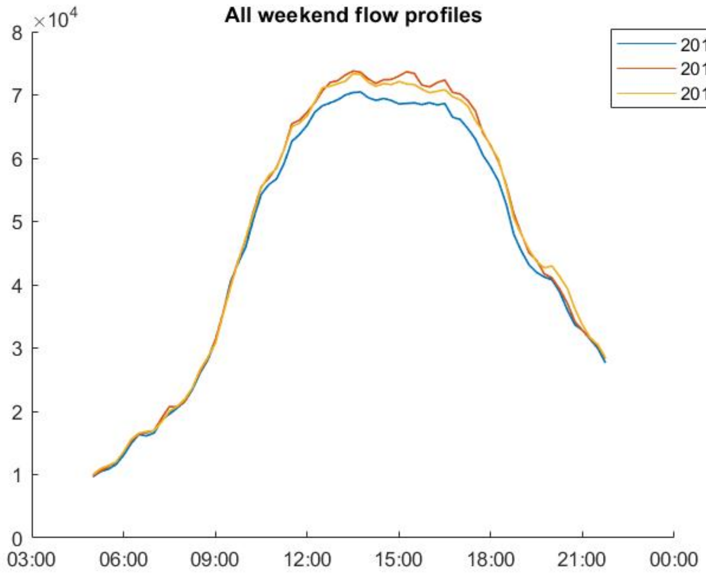


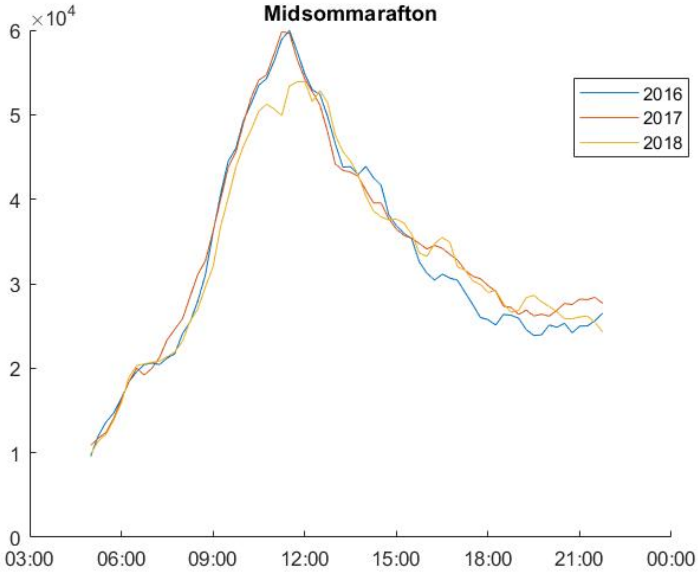


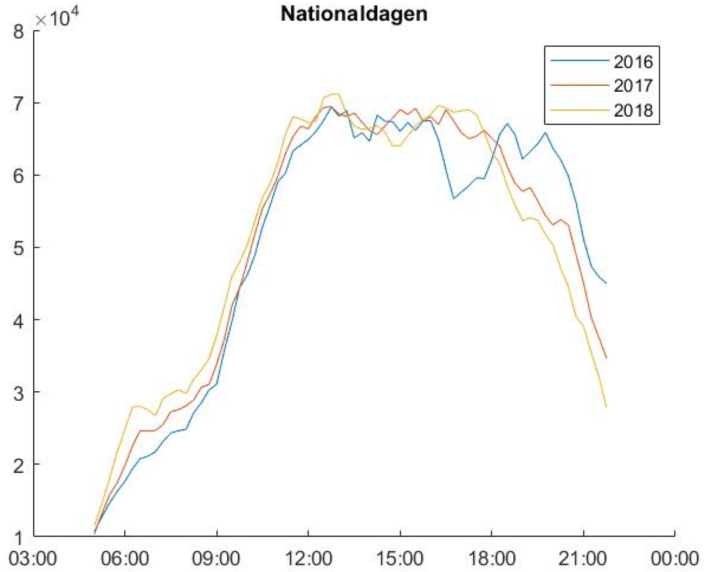


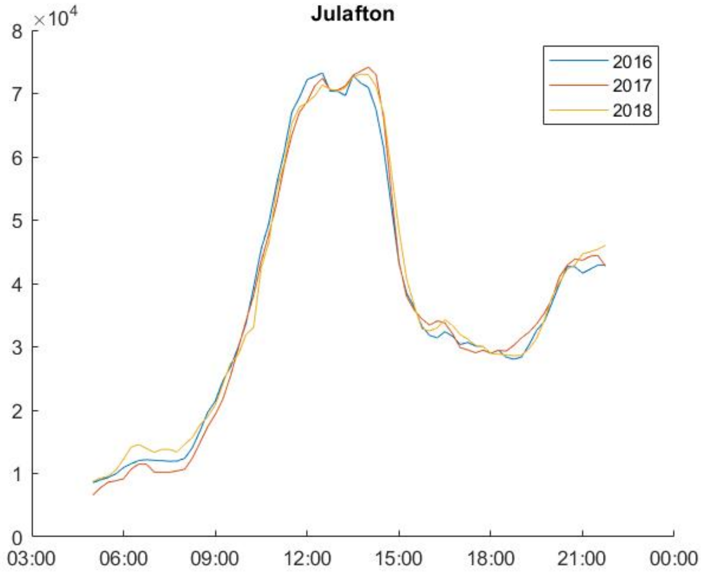
Special days, Weekends and Holidays

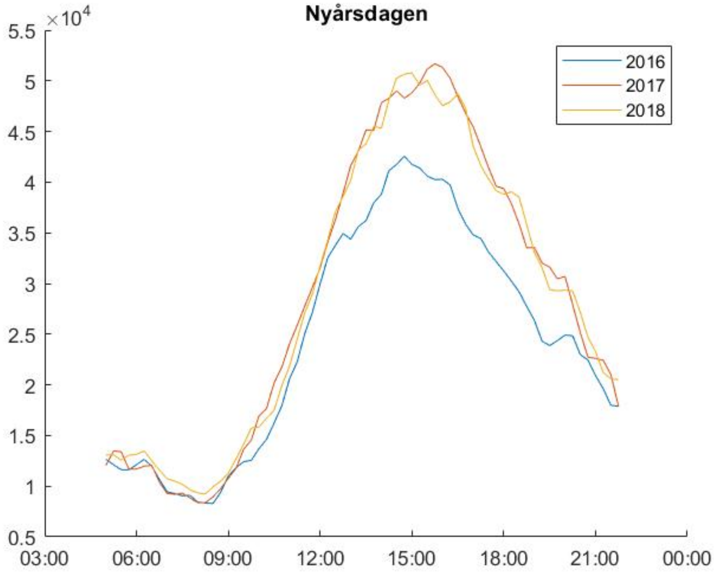






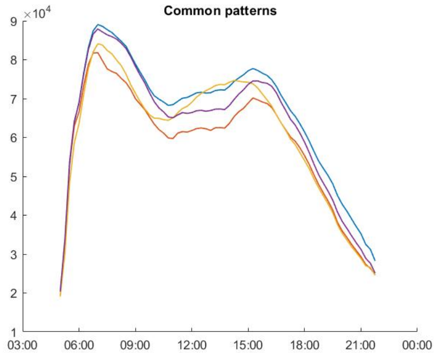


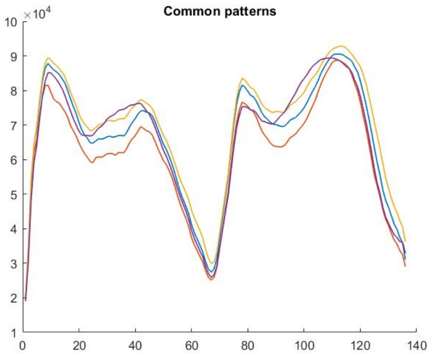


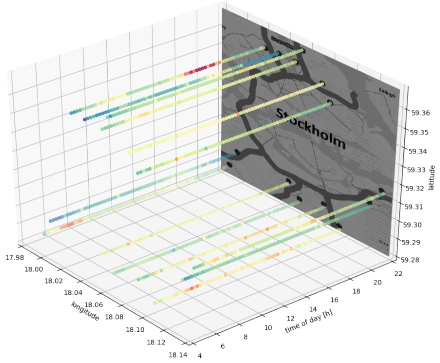


Clustering Methods

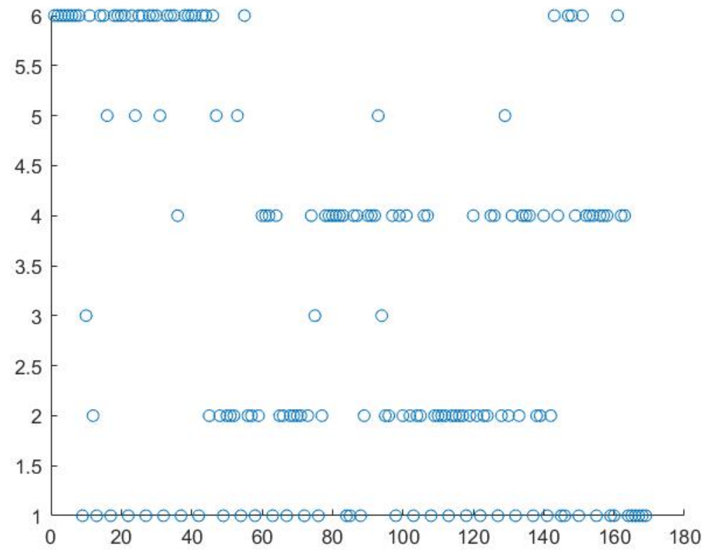
- Standard K-means per sensor (KmeanSensors)
  - Local pattern for each sensor
- K-means jointly for several sensors (KmeanVector)
  - Network effects, but no location clustering
- 3D speed maps (3Dmap)
  - Network effects and location included in clustering
- Median observation vector (MOV)
  - Hybrid between 3D speed map and joint K-means (2-dimensional)

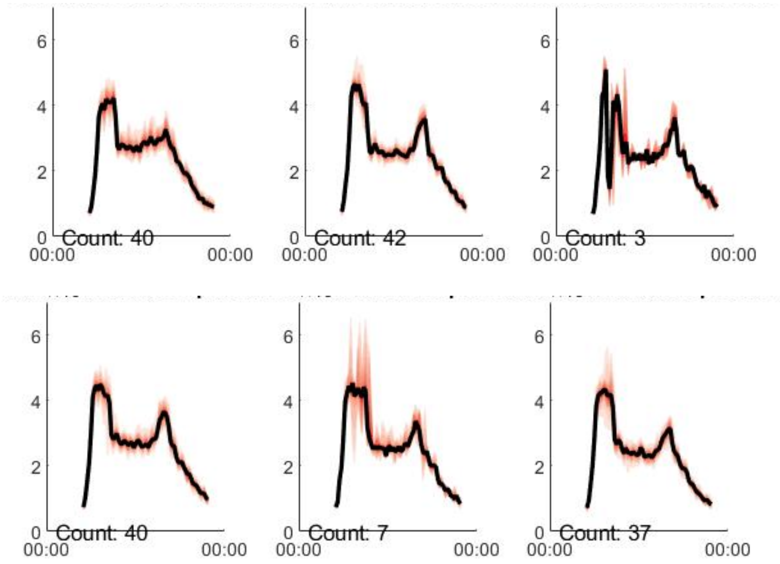


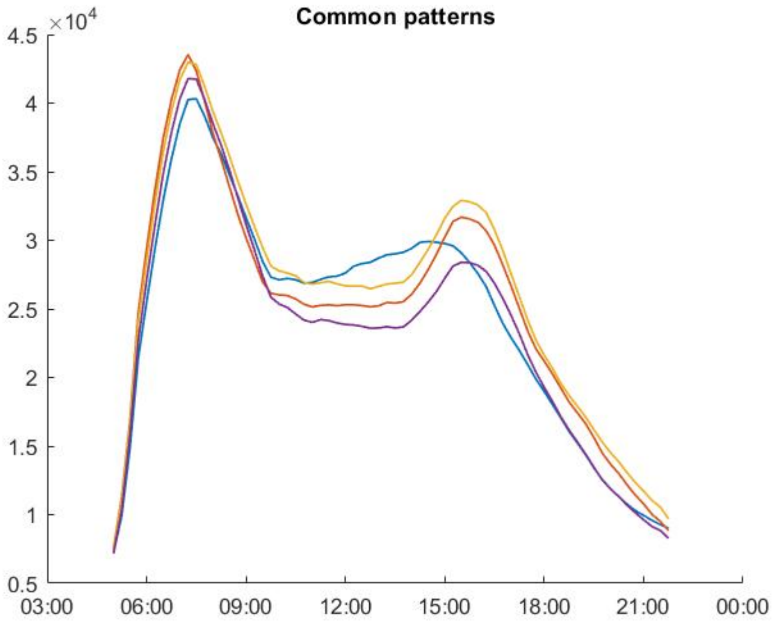


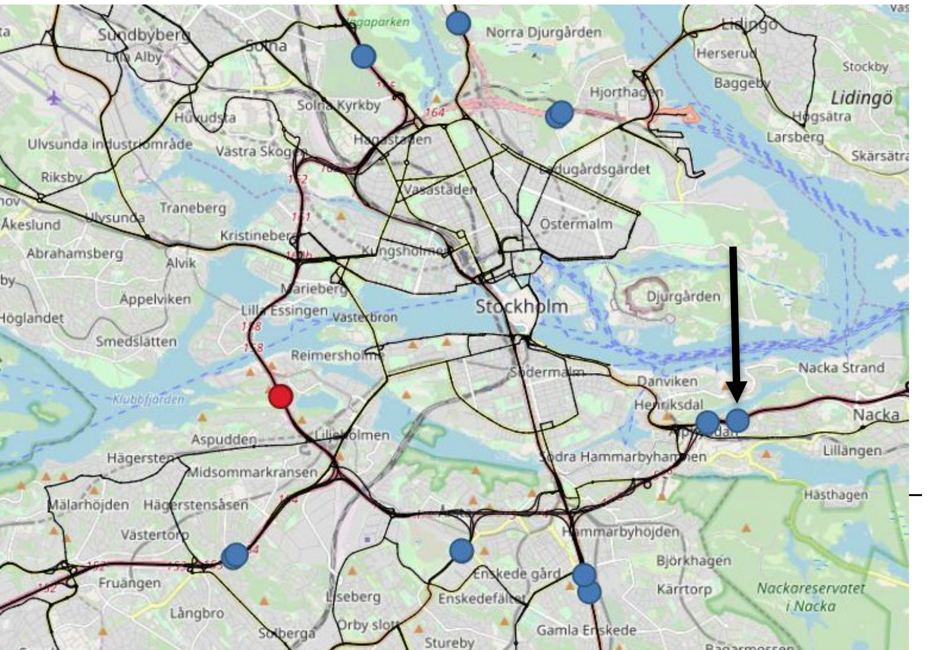


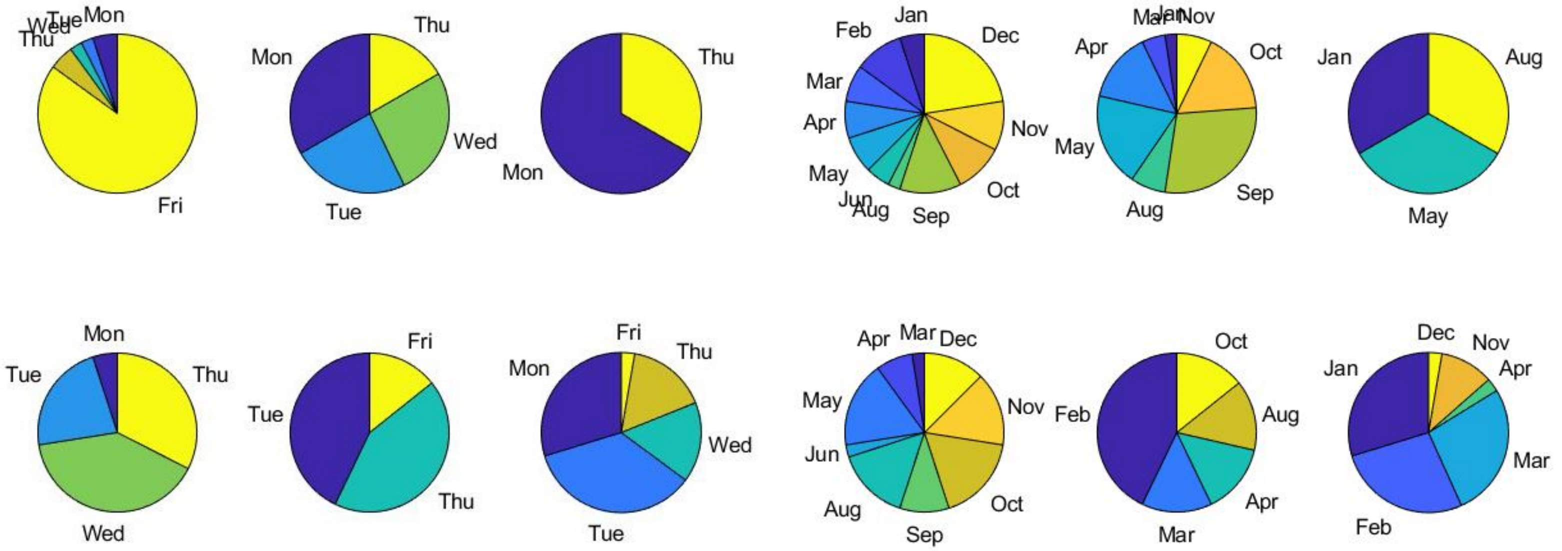
Clustering Results Example



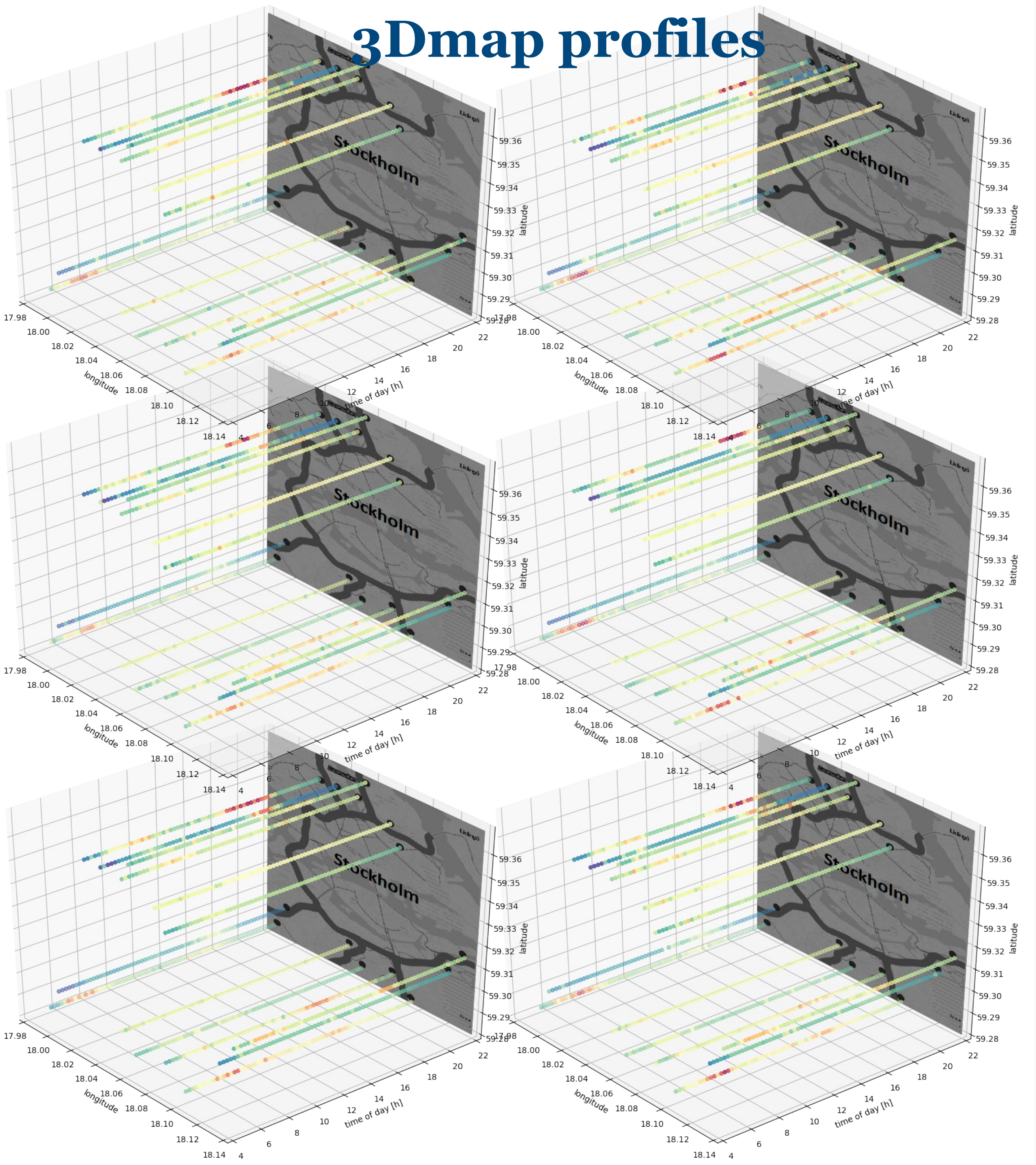




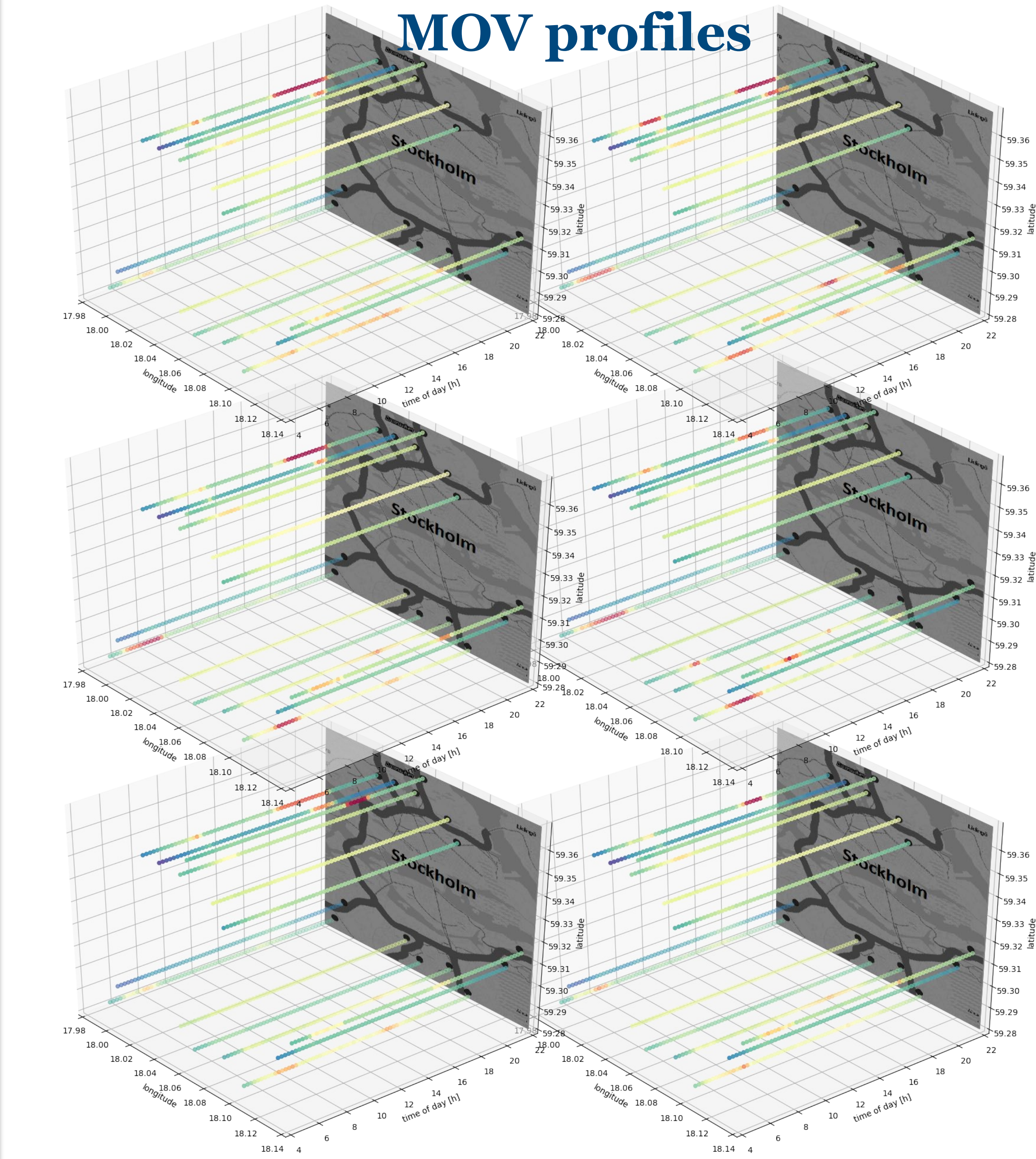




3Dmap profiles



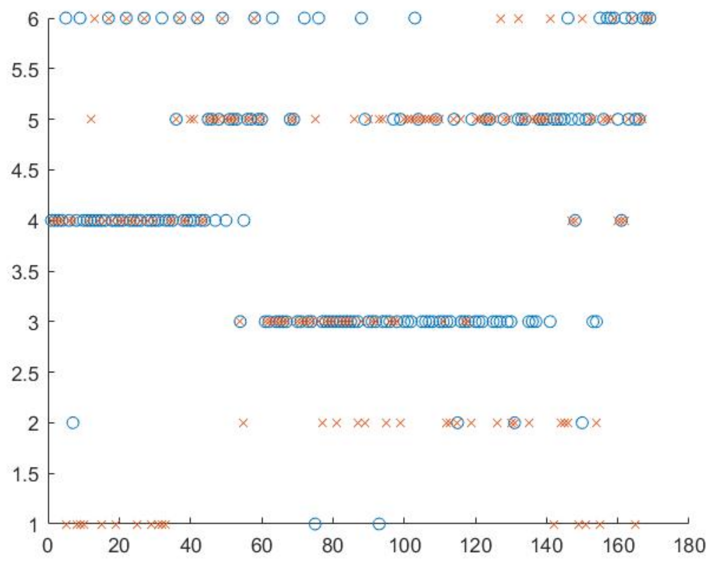
MOV profiles

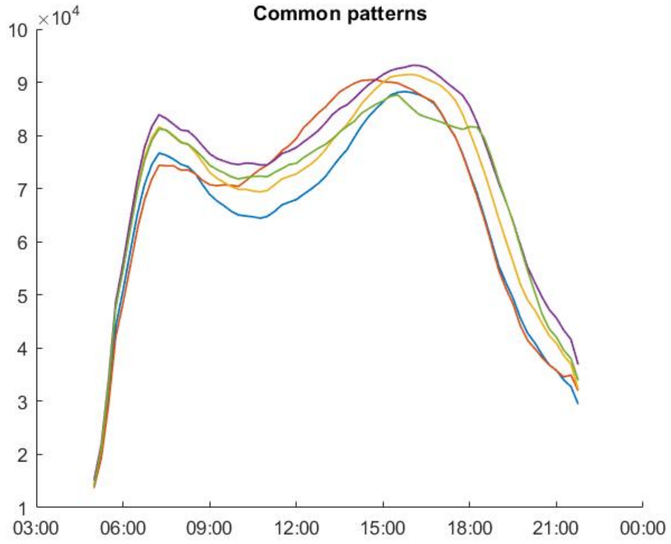


Clustering Similarity

- What matters?
  - Clustering method?
  - Clustering variables?
  - Sensor selection?
- Day clustering similarity
  - Optimum overlap
  - NMI and Rand's index
- Cluster time profile similarity

$$d_{ij} = 1 - \frac{\sum_k^N abs(v_{ik} - v_{jk})}{\sum_k^N max(v_{ik}, v_{jk})}$$





Day Clustering Similarity Results

Adjusted mutual information: 1 – the same / 0 – nothing in common

Speeds

	3Dmap_12	3Dmap_18	3Dmap_6	MOV_12	MOV_18	MOV_6	KmeanVector_12	KmeanVector_18	KmeanVector_6	KmeanSensors
3Dmap_12	1.00	0.41	0.32	0.21	0.26	0.16	0.16	0.25	0.15	0.18
3Dmap_18	0.41	1.00	0.28	0.30	0.36	0.19	0.25	0.33	0.18	0.21
3Dmap_6	0.32	0.28	1.00	0.18	0.19	0.13	0.15	0.18	0.14	0.14
MOV_12	0.21	0.30	0.18	1.00	0.57	0.43	0.48	0.48	0.40	0.25
MOV_18	0.26	0.36	0.19	0.57	1.00	0.44	0.44	0.52	0.43	0.29
MOV_6	0.16	0.19	0.13	0.43	0.44	1.00	0.48	0.46	0.49	0.23
KmeanVector_12	0.16	0.25	0.15	0.48	0.44	0.48	1.00	0.56	0.63	0.27
KmeanVector_18	0.25	0.33	0.18	0.48	0.52	0.46	0.56	1.00	0.54	0.31
KmeanVector_6	0.15	0.18	0.14	0.40	0.43	0.49	0.63	0.54	1.00	0.24
KmeanSensors	0.18	0.22	0.14	0.27	0.31	0.24	0.28	0.32	0.25	1.00

Flows

	3Dmap_12	3Dmap_18	3Dmap_6	MOV_12	MOV_18	MOV_6	KmeanVector_12	KmeanVector_18	KmeanVector_6	KmeanSensorsl
3Dmap_12	1.00	0.68	0.63	0.55	0.48	0.55	0.63	0.63	0.56	0.45
3Dmap_18	0.68	1.00	0.58	0.55	0.52	0.54	0.62	0.64	0.56	0.47
3Dmap_6	0.63	0.58	1.00	0.57	0.48	0.75	0.71	0.63	0.72	0.43
MOV_12	0.55	0.55	0.57	1.00	0.64	0.55	0.59	0.60	0.60	0.43
MOV_18	0.48	0.52	0.48	0.64	1.00	0.48	0.57	0.59	0.51	0.41
MOV_6	0.55	0.54	0.75	0.55	0.48	1.00	0.68	0.65	0.73	0.42
KmeanVector_12	0.63	0.62	0.71	0.59	0.57	0.68	1.00	0.69	0.77	0.52
KmeanVector_18	0.63	0.64	0.63	0.60	0.59	0.65	0.69	1.00	0.65	0.51
KmeanVector_6	0.56	0.56	0.72	0.60	0.51	0.73	0.77	0.65	1.00	0.47
KmeanSensors	0.45	0.47	0.43	0.43	0.41	0.42	0.52	0.51	0.47	1.00

Clustering Profile Similarity Results

$$ps_{IJ} = \frac{1}{|I|}\min\sum_{i}^{I}\sum_{j}^{J} d_{ij}x_{ij}$$

$$\sum_{j}^{J} x_{ij} = 1\,, \qquad \text{for } i \in I$$

$$\sum_{i}^{I} x_{ij} \leq 1\,, \qquad \text{for j} \in J$$

$$x_{ij} \in \{0,1\}$$

I,J – set of profiles

1 – the same profiles

Speed

J/I	3Dmap_12	3Dmap_18	3Dmap_6	MOV_12	MOV_18	MOV_6	KmeanVector_12	KmeanVector_18	KmeanVector_6
3Dmap_12	1.00	0.95	0.96	0.96	0.95	0.96	0.94	0.92	0.96
3Dmap_18	0.95	1.00	0.96	0.96	0.95	0.96	0.95	0.93	0.96
3Dmap_6	0.96	0.96	1.00	0.95	0.94	0.95	0.92	0.90	0.95
MOV_12	0.96	0.96	0.95	1.00	0.96	0.97	0.94	0.92	0.97
MOV_18	0.95	0.95	0.94	0.96	1.00	0.97	0.95	0.93	0.97
MOV_6	0.96	0.96	0.95	0.97	0.97	1.00	0.92	0.90	0.96
KmeanVector_12	0.94	0.95	0.92	0.94	0.95	0.92	1.00	0.91	0.94
KmeanVector_18	0.92	0.93	0.90	0.92	0.93	0.90	0.91	1.00	0.94
KmeanVector_6	0.96	0.96	0.95	0.97	0.97	0.96	0.94	0.94	1.00

Flow

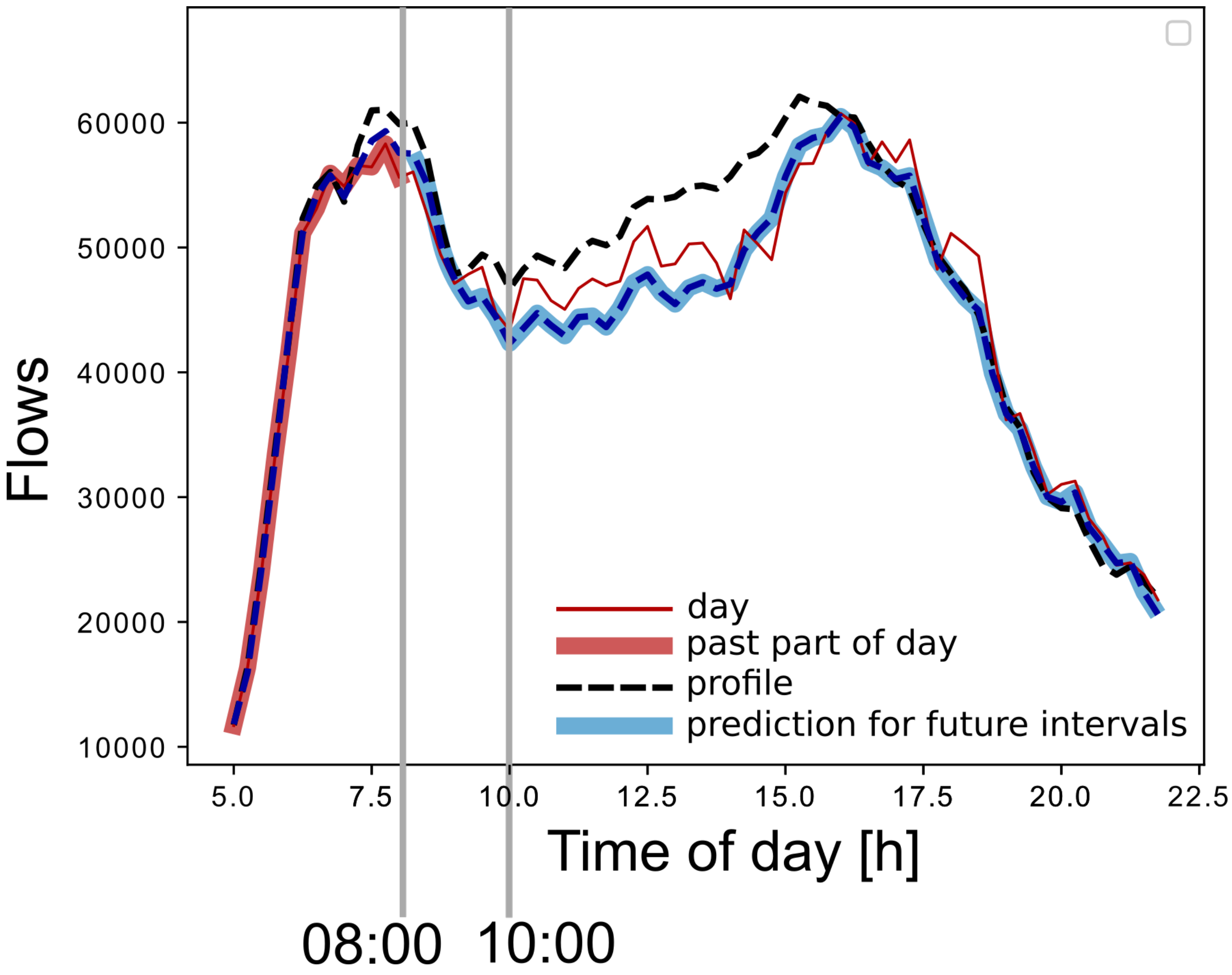
J/I	3Dmap_12	3Dmap_18	3Dmap_6	MOV_12	MOV_18	MOV_6	KmeanVector_12	KmeanVector_18	KmeanVector_6
3Dmap_12	1.00	0.87	0.88	0.89	0.88	0.90	0.84	0.80	0.89
3Dmap_18	0.87	1.00	0.89	0.90	0.89	0.91	0.87	0.83	0.90
3Dmap_6	0.88	0.89	1.00	0.88	0.87	0.89	0.80	0.73	0.88
MOV_12	0.89	0.90	0.88	1.00	0.91	0.93	0.86	0.81	0.92
MOV_18	0.88	0.89	0.87	0.91	1.00	0.93	0.89	0.85	0.93
MOV_6	0.90	0.91	0.89	0.93	0.93	1.00	0.81	0.73	0.91
KmeanVector_12	0.84	0.87	0.80	0.86	0.89	0.81	1.00	0.77	0.87
KmeanVector_18	0.80	0.83	0.73	0.81	0.85	0.73	0.77	1.00	0.86
KmeanVector_6	0.89	0.90	0.88	0.92	0.93	0.91	0.87	0.86	1.00

Demand Prediction

- Prediction for 2 hours (8 x 15 minute time intervals) into the future at the 08:00 and 15:00
- Match the current day with closest profile considering all past intervals to the 08:00 and 15:00

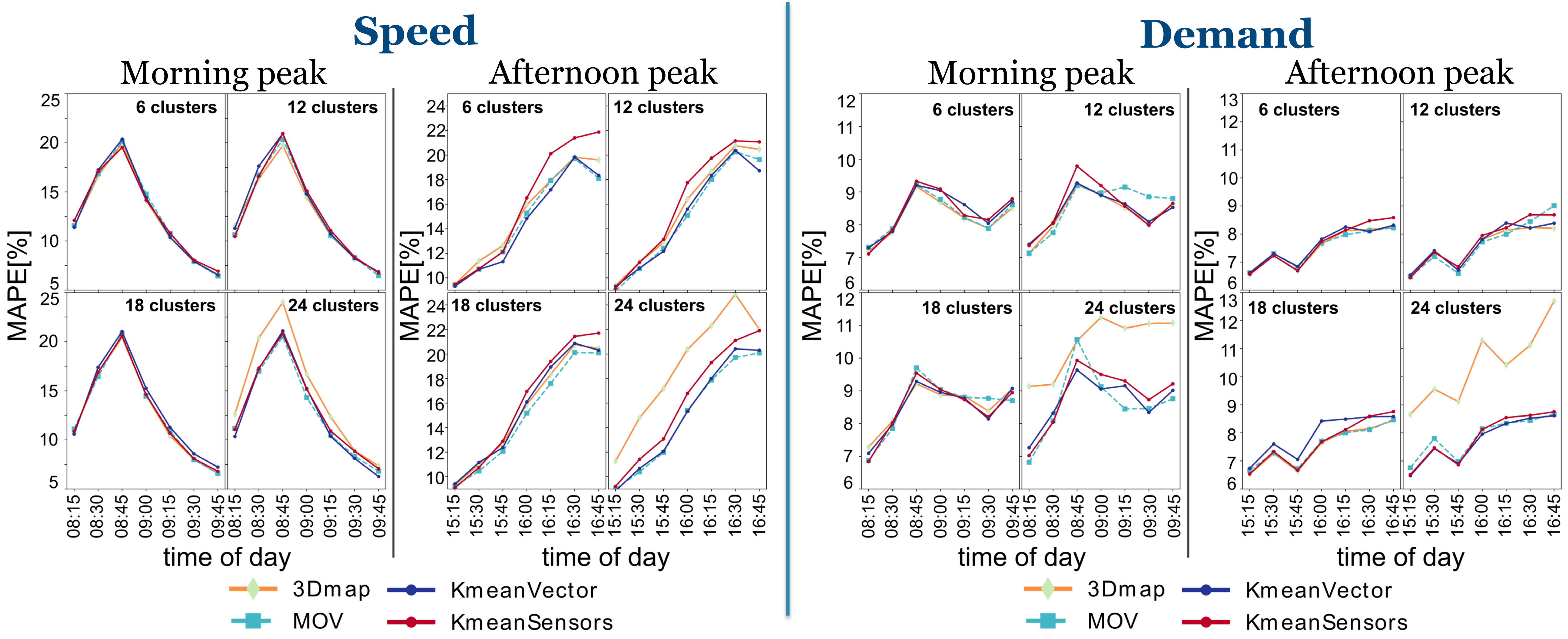
$$\min_i \sum_{k}^{K} abs(v_{ik} - c_k)$$

- Clustering of weekdays 2017
- Prediction of weekdays 2018



Demand Prediction Results

$$MAPE = \frac{100\%}{n} \sum^{n} \frac{observed - predicted}{observer}$$



LINKÖPINGS UNIVERSITET

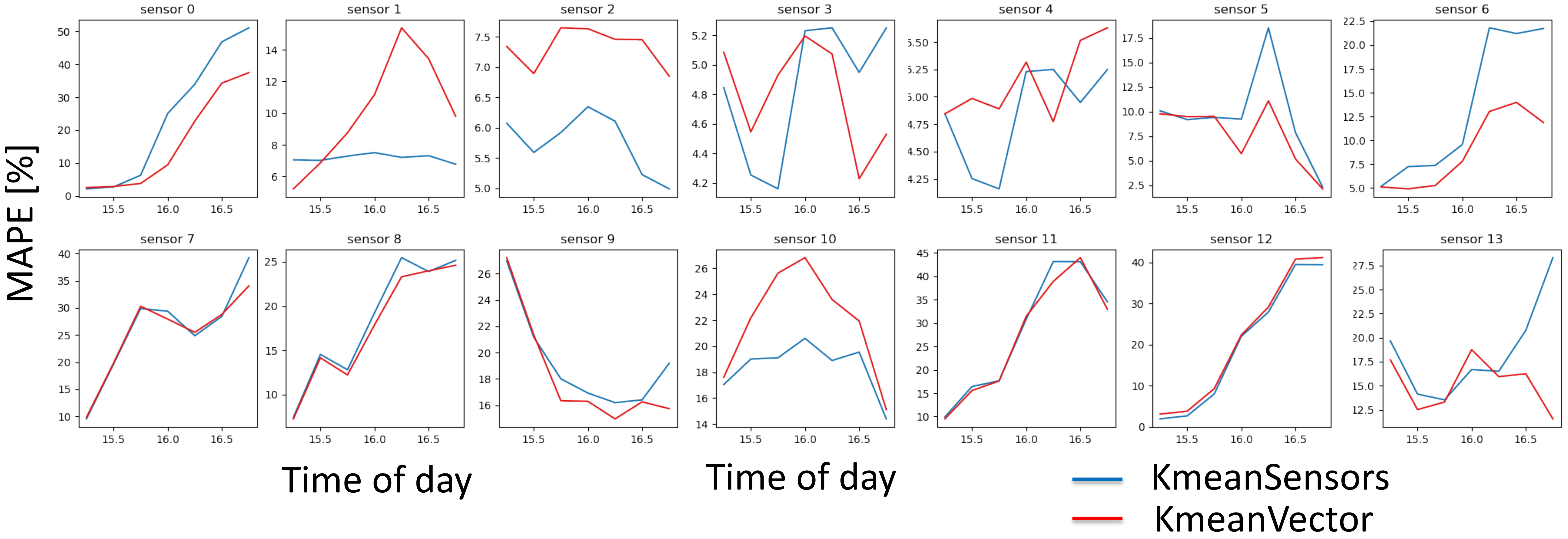
KTH VETENSKAP OCH KONST

CTR

Demand Prediction Results

$$MAPE = \frac{100\%}{n}\sum^{n}\frac{observed - predicted}{observer}$$

KmeanSensors VS KmeanVectors sensor by sensor, speeds and afternoon peak, 6 clusters



Demand Prediction Based on Speed/Travel times?

- Increasing amount of network-wide probe data
- Can we find network-wide demand patterns using travel time / speed data?
  - Cluster link speeds or OD matrix?
- Do we need network-wide clustering?
  - Large-scale actions that affects route choice

Normalized mutual information:
1 – the same / 0 – nothing in common

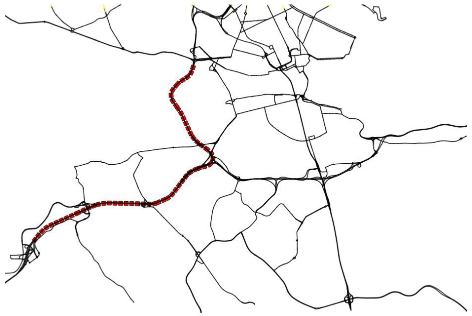
Clustering similarity between flows and speeds clusters [NMI]

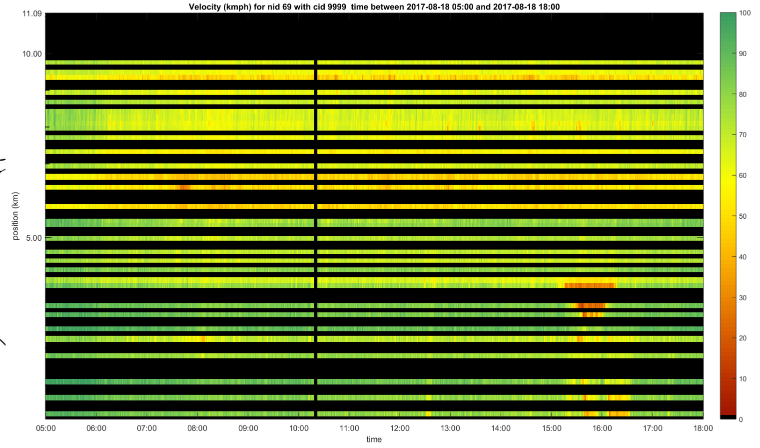
	Number of clusters			
Method	6	12	18	24
3Dmap	0.17	0.23	0.35	0.52
MOV	0.32	0.38	0.46	0.50
KmeanVector	0.41	0.42	0.44	0.54
KmeanSensors	0.27	0.31	0.36	0.42

Scenario Evaluation Example

- 25/8 2017: Malfunctioning bridge splice (”broskarv”) on Essingeleden
  - How to manage lane blocking?
  - What information should be given to travellers and when?

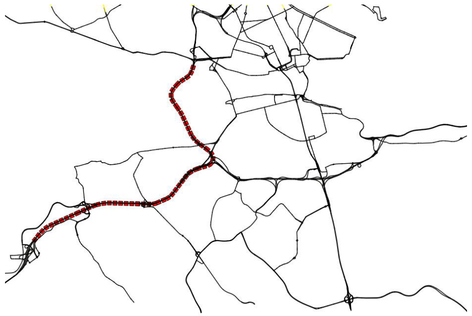
E4N 18/8 2017



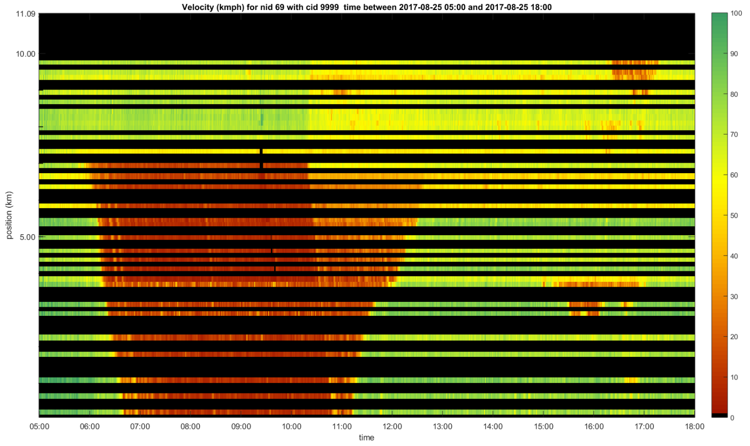


Scenario Evaluation Example

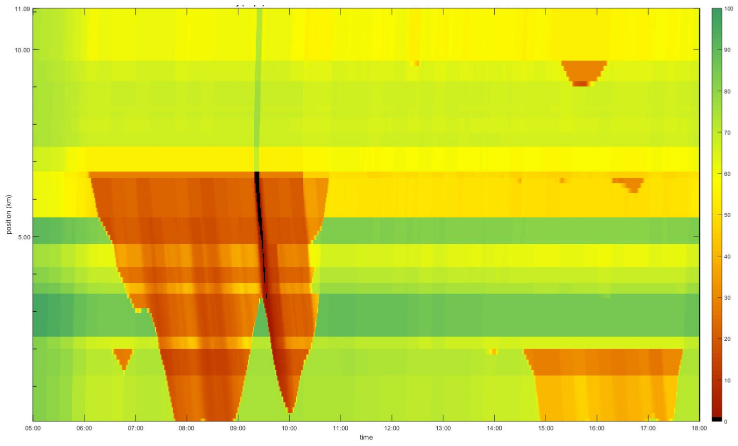
- 25/8 2017: Malfunctioning bridge splice (”broskarv”) on Essingeleden
  - How to manage lane blocking?
  - What information should be given to travellers and when?



E4N 25/8 2017



E4N 25/8 2017 Model



Conclusions and Future Work

- Is clustering useful for scenario evaluation?
  - Yes, to determine typical days
- Does clustering method/variables matter?
  - Yes, especially for day clustering
  - However, small variations for weekdays and profiles quite similar
- Does cluster-based prediction work?
  - Yes, and tentatively quite little differences between clustering methods
- More work needed
  - To evaluate clustering effects on action ranking
  - To understand network-wide clustering and relationship between speed/flow-based clustering