Automatic segmentation and classification of movement trajectories for transportation modes

A thesis submitted in partial fulfilment of the requirements for the degree Master of Science in Geomatics

Filip Biljecki

2 July 2010

Supervisors: Peter van Oosterom and Hugo Ledoux
Department of GIS Technology, OTB

TU Delft Delft University of Technology

Objective

$$(x_1, y_1, z_1, t_1)$$

$$(x_2, y_2, z_2, t_2)$$

$$\vdots$$

$$(x_n, y_n, z_n, t_n)$$

$$\Downarrow$$

Transportation mode

Objective (2)

$$\left.\begin{matrix}(x_1, y_1, z_1, t_1)\\ \vdots \\ (x_i, y_i, z_i, t_i)\end{matrix}\right\}\ \text{1st transportation mode}$$

$$\left.\begin{matrix}(x_{i+1}, y_{i+1}, z_{i+1}, t_{i+1})\\ \vdots \\ (x_j, y_j, z_j, t_j)\end{matrix}\right\}\ \text{2nd transportation mode}$$

$$\vdots$$

$$\left.\begin{matrix}(x_k, y_k, z_k, t_k)\\ \vdots \\ (x_u, y_u, z_u, t_u)\end{matrix}\right\}\ \text{n-th transportation mode}$$

Example

- GPX file in Google Earth

Applications

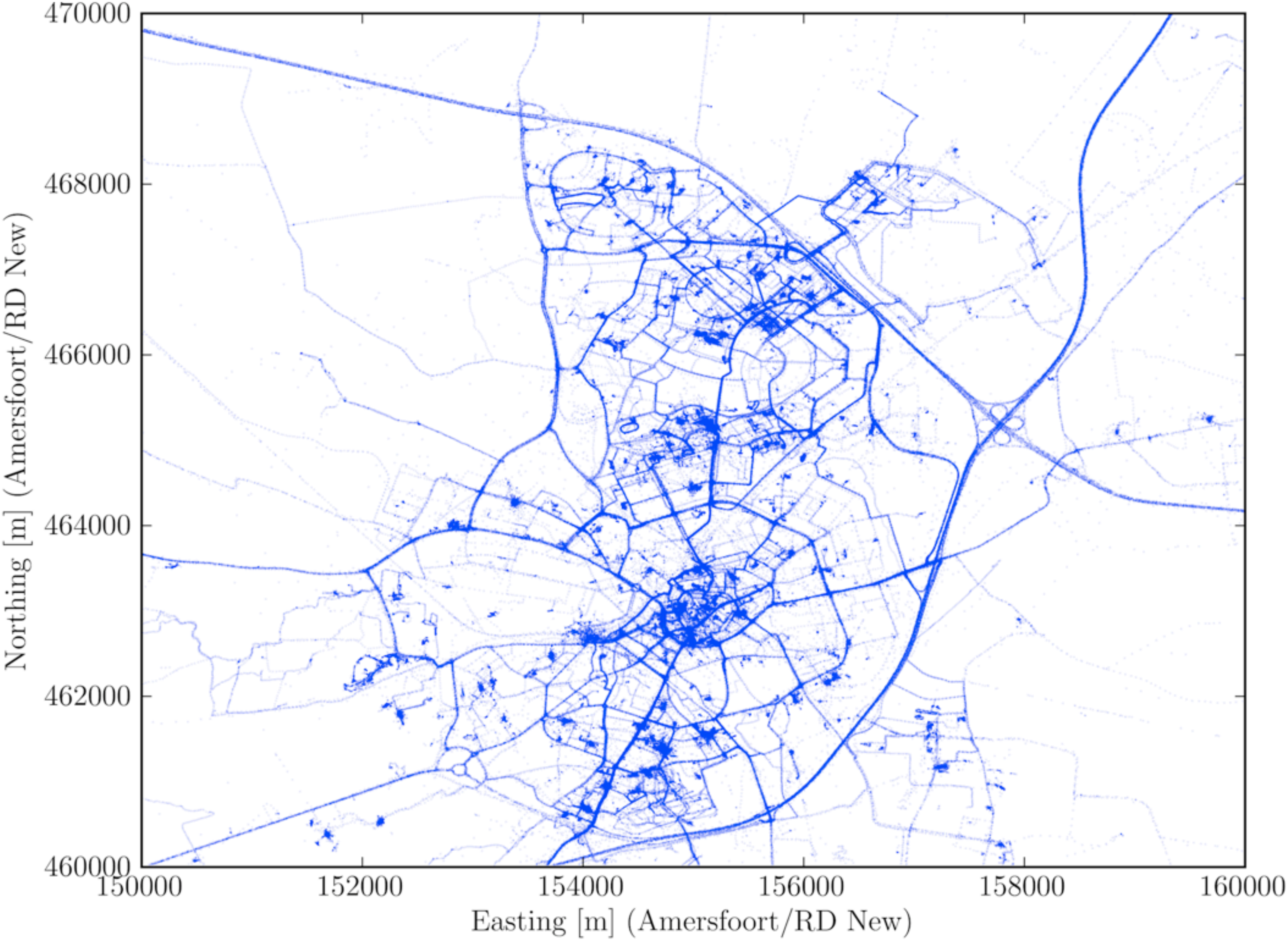
Delft
University of
Technology

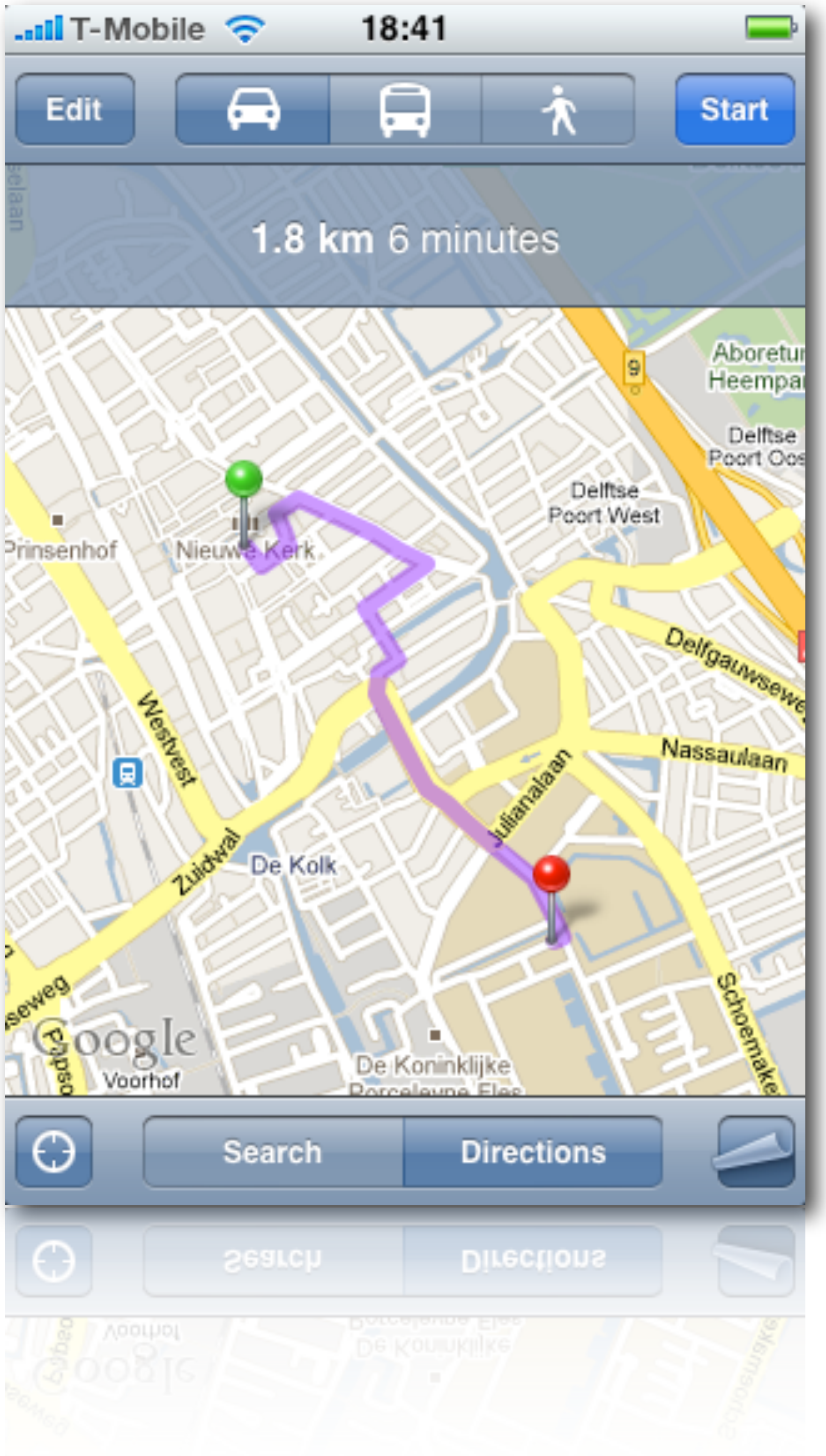
Applications

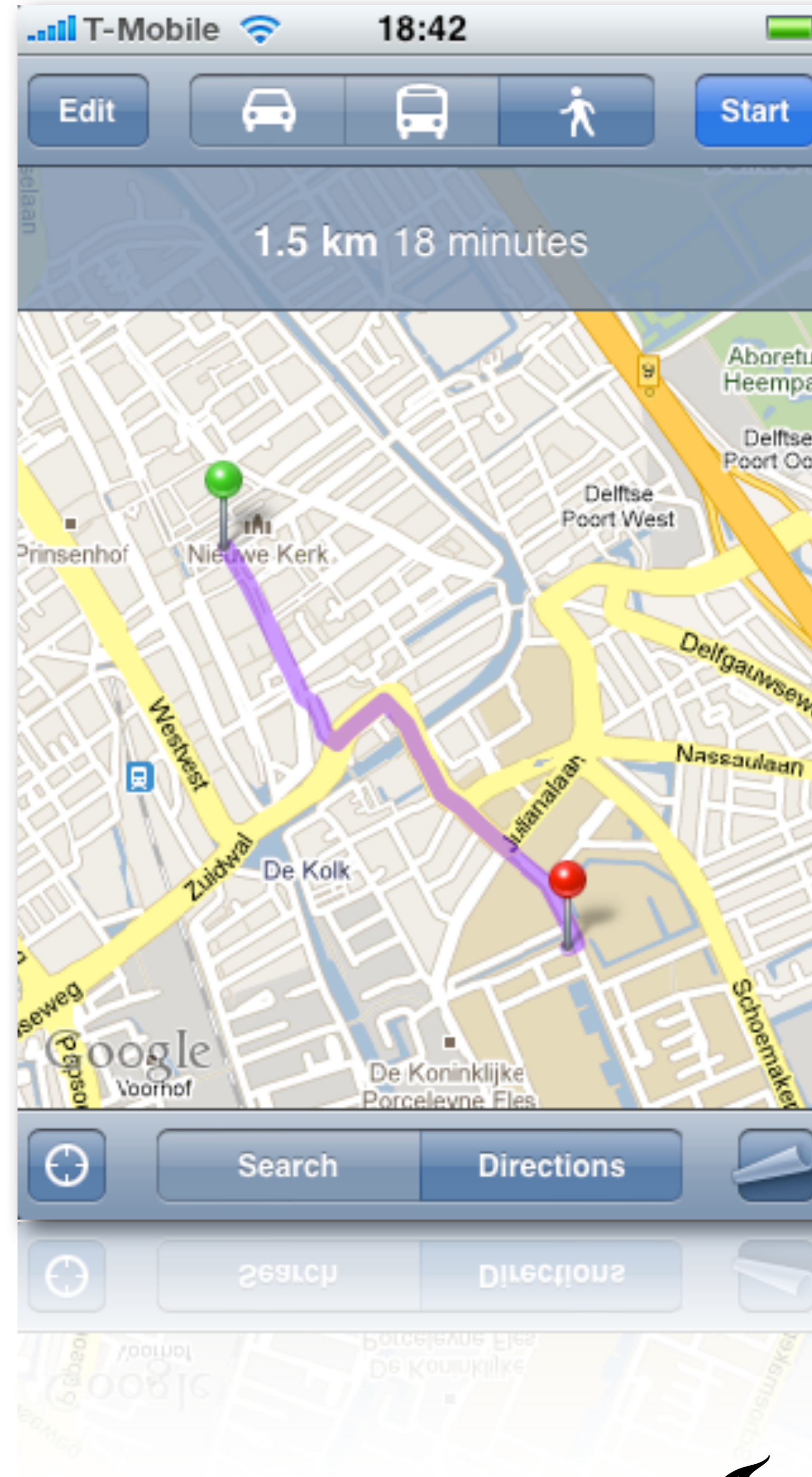
- Travel behaviour studies

Applications

- Travel behaviour studies
- Datasets:
    - Department of Urban and Regional Development, OTB
    - Department of Urbanism, Faculty of Architecture
    - 17.6 M points from 1369 individuals
    - 539000 km


Northing [m] (Amersfoort/RD New)
Easting [m] (Amersfoort/RD New)
470000
468000
466000
464000
462000
460000
150000
152000
154000
156000
158000
160000

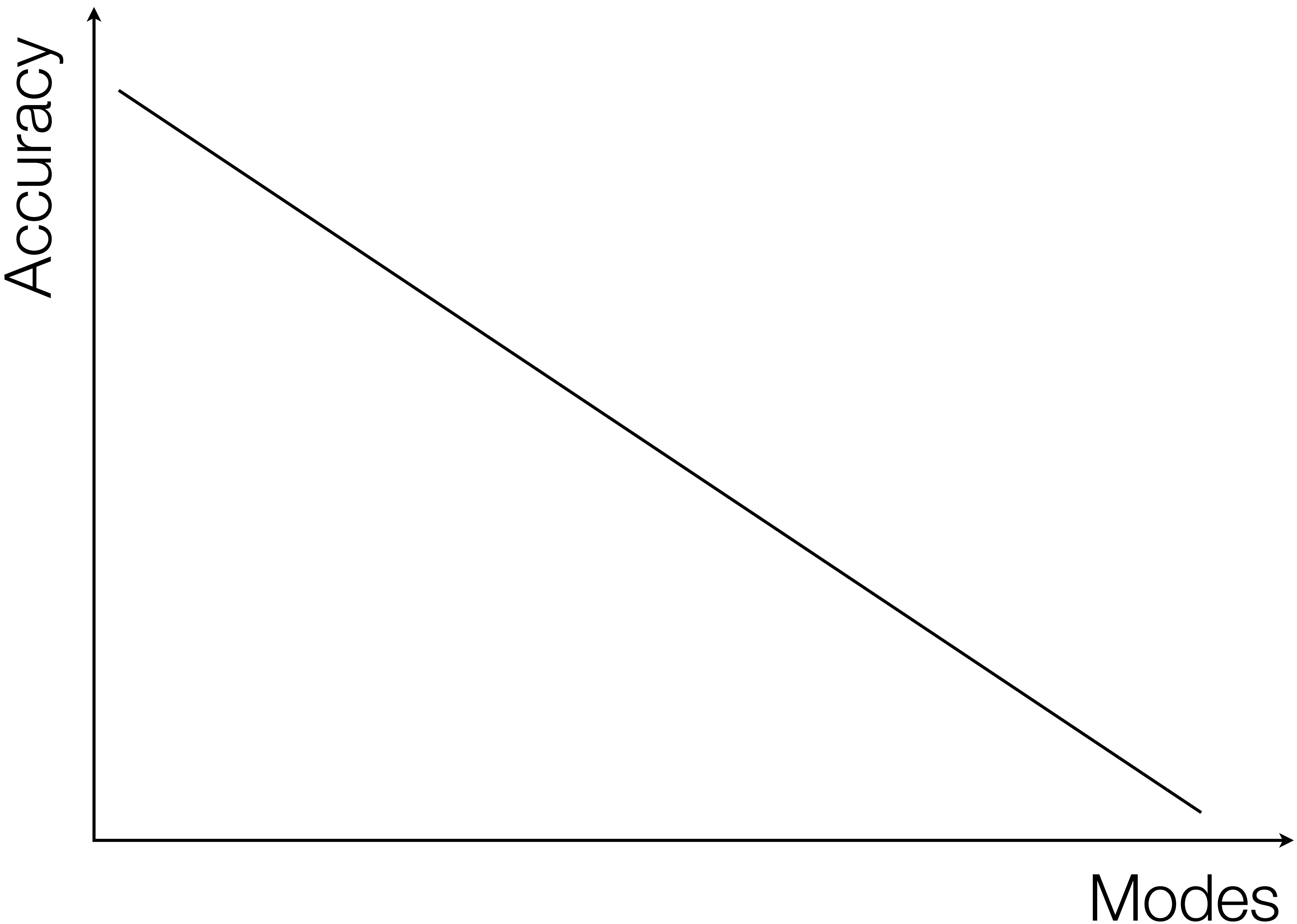




TU Delft Delft University of Technology

Existing solutions

- Deterministic solutions
- Speed
- A few modes (average: 4.5), dissimilar in behaviour
- Problem with data gaps not solved
- No segmentation


Accuracy
Modes

Overview

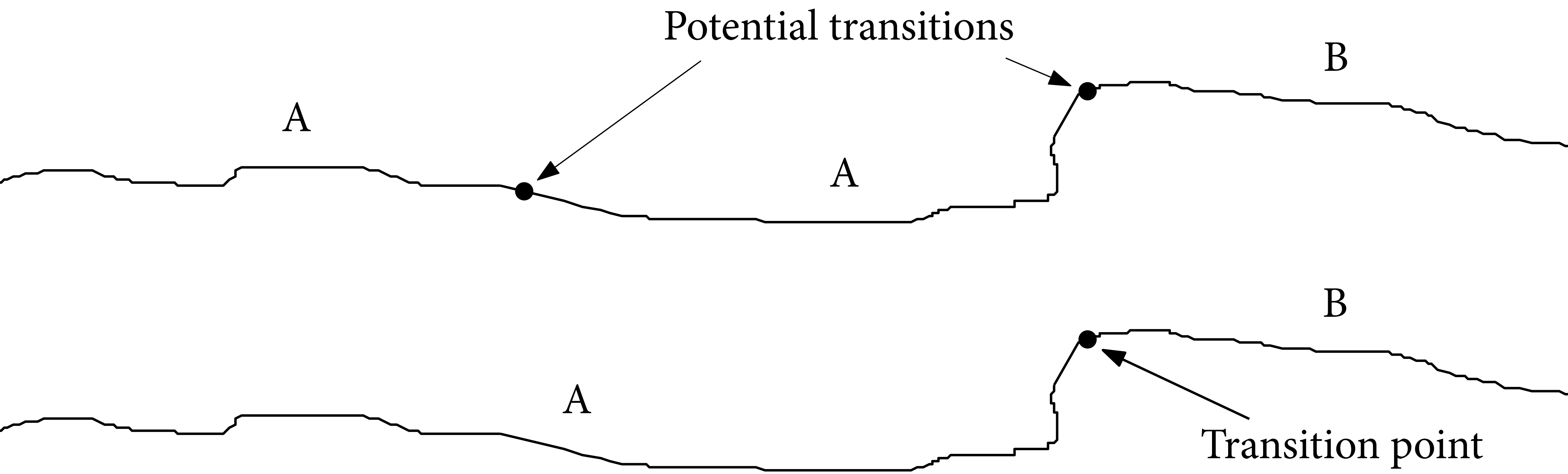
- Segmentation with sensitive thresholds
- Classification system inspired by fuzzy expert systems and strongly supported by geo-information
- Developed experimental software (Python + PostgreSQL/PostGIS)
- 10 modes: walking, bicycle, car, bus, tram, train, underground, sailing boat, ferry, aircraft

TU Delft Delft University of Technology

Segmentation

- Detection of stops or shortages in the data

Segmentation (2)



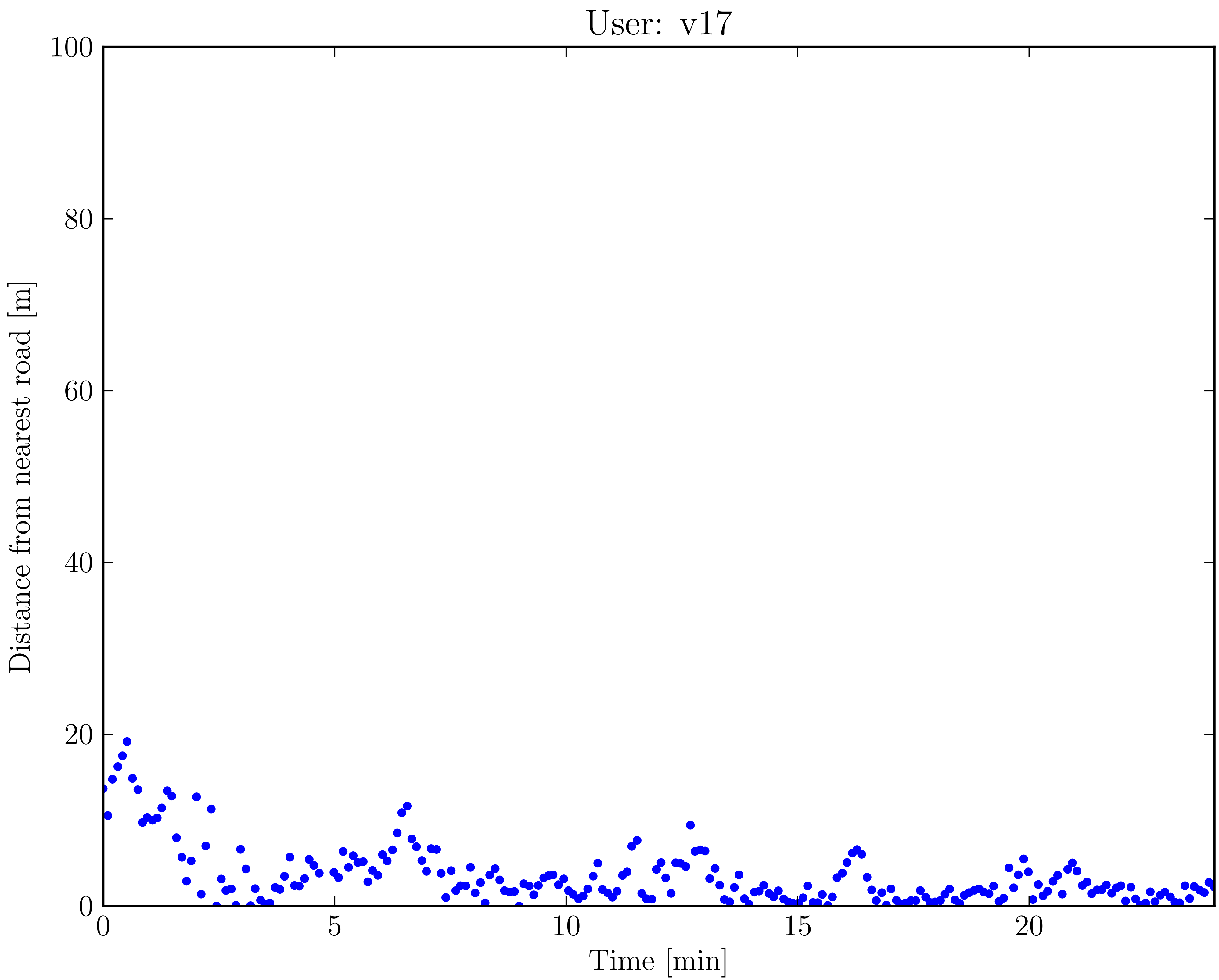
Segmentation (3)

- Sensitive thresholds
- Oversegmentation is better than undersegmentation

Classification solution

- Nine indicators: speed + geo-information
- Grouping similar modes
- Fuzziness
- Elimination of unlikely transportation modes

Proximities (from Openstreetmap data)



Grouping similar modes

1	Land	Sea	Air
2	Walk Bicycle Car/tram/bus Train Underground	Boat	Aircraft
3	Walk Bicycle Car Tram Bus Train Underground	Sailing boat Ferry	Aircraft

TU Delft Delft University of Technology

Classification (FES)

- Expert systems
- IF *e* is observed THEN *h* is true
- Certainty factors (CF): confidence of a claim
- Fuzzy logic

Classification (2)

Classification (2)

- IF (max. speed is 118 km/h) THEN (mode = car) WITH CF = 1.0

Classification (2)

- IF (max. speed is 118 km/h) THEN (mode = car) WITH CF = 1.0
- IF (max. speed is 138 km/h) THEN (mode = car) WITH CF = 0.6

Classification (2)

- IF (max. speed is 118 km/h) THEN (mode = car) WITH CF = 1.0
- IF (max. speed is 138 km/h) THEN (mode = car) WITH CF = 0.6

TU Delft Delft University of Technology

Classification (2)

- IF (max. speed is 118 km/h) THEN (mode = car) WITH CF = 1.0
- IF (max. speed is 138 km/h) THEN (mode = car) WITH CF = 0.6
- IF (max. speed is 138 km/h) THEN (mode = train) WITH CF = 0.8

Classification (2)

- IF (max. speed is 118 km/h) THEN (mode = car) WITH CF = 1.0
- IF (max. speed is 138 km/h) THEN (mode = car) WITH CF = 0.6
- IF (max. speed is 138 km/h) THEN (mode = train) WITH CF = 0.8
- IF (max. speed is 138 km/h) THEN (mode = {car, train, . . . }) WITH CF = {0.6, 0.8, ...}

Classification (2)

- IF (max. speed is 118 km/h) THEN (mode = car) WITH CF = 1.0
- IF (max. speed is 138 km/h) THEN (mode = car) WITH CF = 0.6
- IF (max. speed is 138 km/h) THEN (mode = train) WITH CF = 0.8
- IF (max. speed is 138 km/h) THEN (mode = {car, train, . . . }) WITH CF = {0.6, 0.8, ...}
- Assigning certainties with fuzzy variables (via empirically determined membership functions):

$$\mathrm{CF} = f(e)$$

TU Delft Delft University of Technology

Membership functions

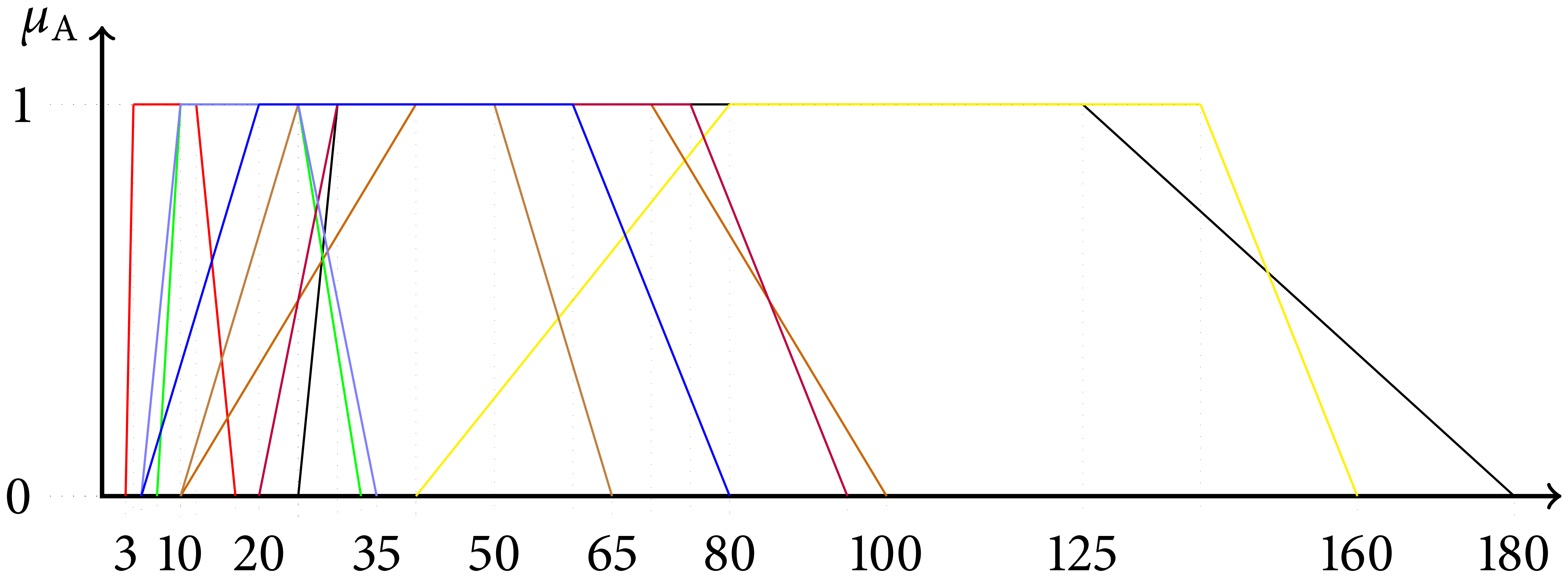
- One for each mode for each indicator

$$\mathrm{CF}_m^i = f_m^i(i)$$

$$\mathrm{CF}_{\mathrm{train}}^{\mathrm{max.speed}} = f_{\mathrm{train}}^{\mathrm{max.speed}}(\mathrm{max.speed})$$

TU Delft Delft University of Technology

μ_A
25
40
125
180



μ_A
1
0
3
10
20
35
50
65
80
100
125
160
180

Training

- Training data - manual trials, iterative process
- Extensibility with XML:

```
<indicator name="bus_proximity">
 <mode layer="3" name="bus">
  <values>0,0,10,30</values>
 </mode>
```

Chaining the results

- Result in nine CFs for each transportation mode

$$
\begin{array}{cccc}
\mathrm{CF}_1^1 = f_1^1(i_1) & \mathrm{CF}_1^2 = f_1^2(i_2) & \ldots & \mathrm{CF}_1^k = f_1^k(i_k) \\
\mathrm{CF}_2^1 = f_2^1(i_1) & \mathrm{CF}_2^2 = f_2^2(i_2) & \ldots & \mathrm{CF}_2^k = f_2^k(i_k) \\
\vdots & \vdots & \vdots & \vdots \\
\mathrm{CF}_n^1 = f_n^1(i_1) & \mathrm{CF}_n^2 = f_n^2(i_2) & \ldots & \mathrm{CF}_n^k = f_n^k(i_k)
\end{array}
$$

TU Delft Delft University of Technology

Chaining the results (2)

- CF[A ∩ B] = min(CF[A], CF[B])

$$
\begin{array}{cccccc}
\mathrm{CF}_1^1 = f_1^1(i_1) & \mathrm{CF}_1^2 = f_1^2(i_2) & \ldots & \mathrm{CF}_1^k = f_1^k(i_k) & \Rightarrow \mathrm{CF}_1 = \min(\mathrm{CF}_1^1, \ldots \mathrm{CF}_1^k) \\
\mathrm{CF}_2^1 = f_2^1(i_1) & \mathrm{CF}_2^2 = f_2^2(i_2) & \ldots & \mathrm{CF}_2^k = f_2^k(i_k) & \Rightarrow \mathrm{CF}_2 = \min(\mathrm{CF}_2^1, \ldots \mathrm{CF}_2^k) \\
\vdots & \vdots & \vdots & \vdots & \\
\mathrm{CF}_n^1 = f_n^1(i_1) & \mathrm{CF}_n^2 = f_n^2(i_2) & \ldots & \mathrm{CF}_n^k = f_n^k(i_k) & \Rightarrow \mathrm{CF}_n = \min(\mathrm{CF}_n^1, \ldots \mathrm{CF}_n^k)
\end{array}
$$

Chaining the results (3)

IF (max. speed is 55 km/h) THEN (mode = tram) WITH CF = 0.85

IF (proximity to tram network is 4933 m) THEN (mode = tram) WITH CF = 0

→ CF(tram) = min(0.85, 0) = 0

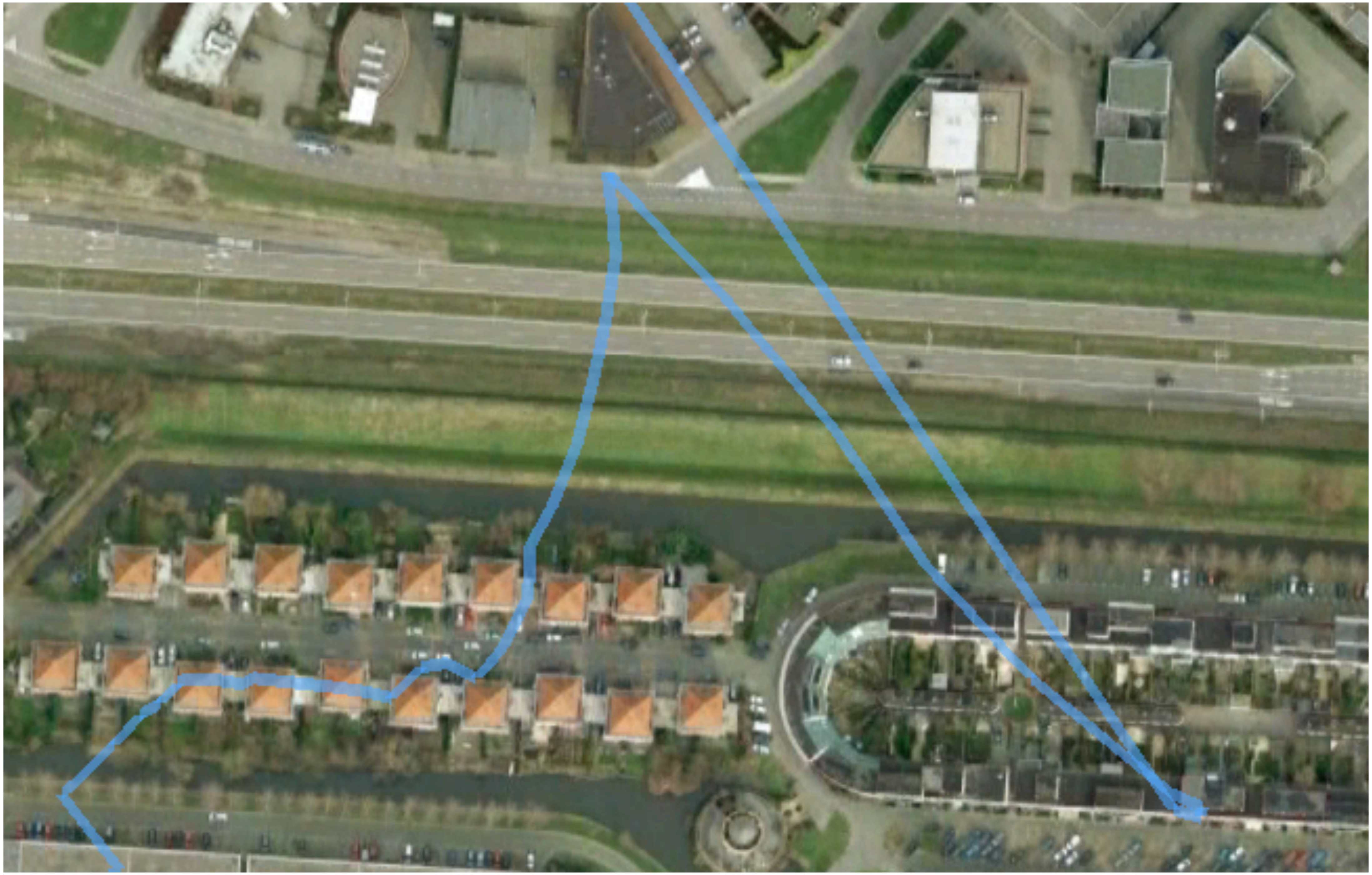
Demonstration

- Prototype report
- KML output

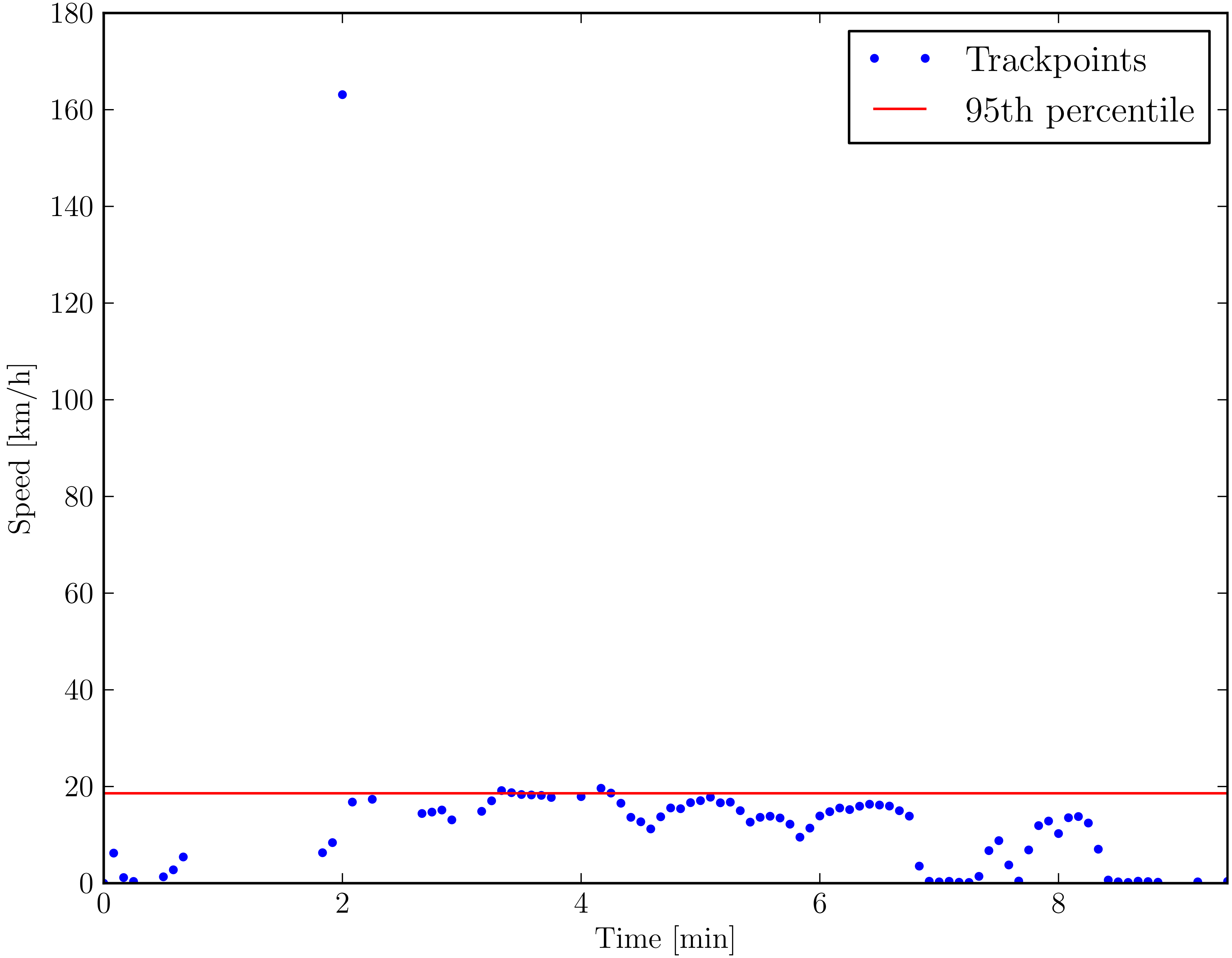
Solving specific problems

- GPS errors, noise
- Bus, tram, and cars in urban areas are similar in behaviour (speed, often stops, infrastructure-GPS errors)
- ‘Gaps’ in the data caused by signal shortages

Noise



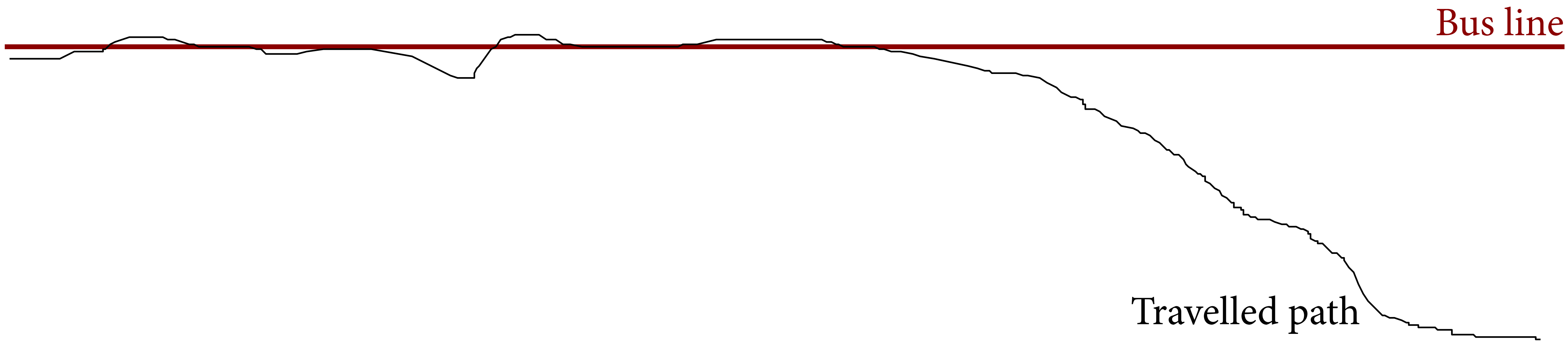
Noise (2)



Trackpoints
95th percentile
Speed [km/h]
Time [min]
0
20
40
60
80
100
120
140
160
180
0
2
4
6
8

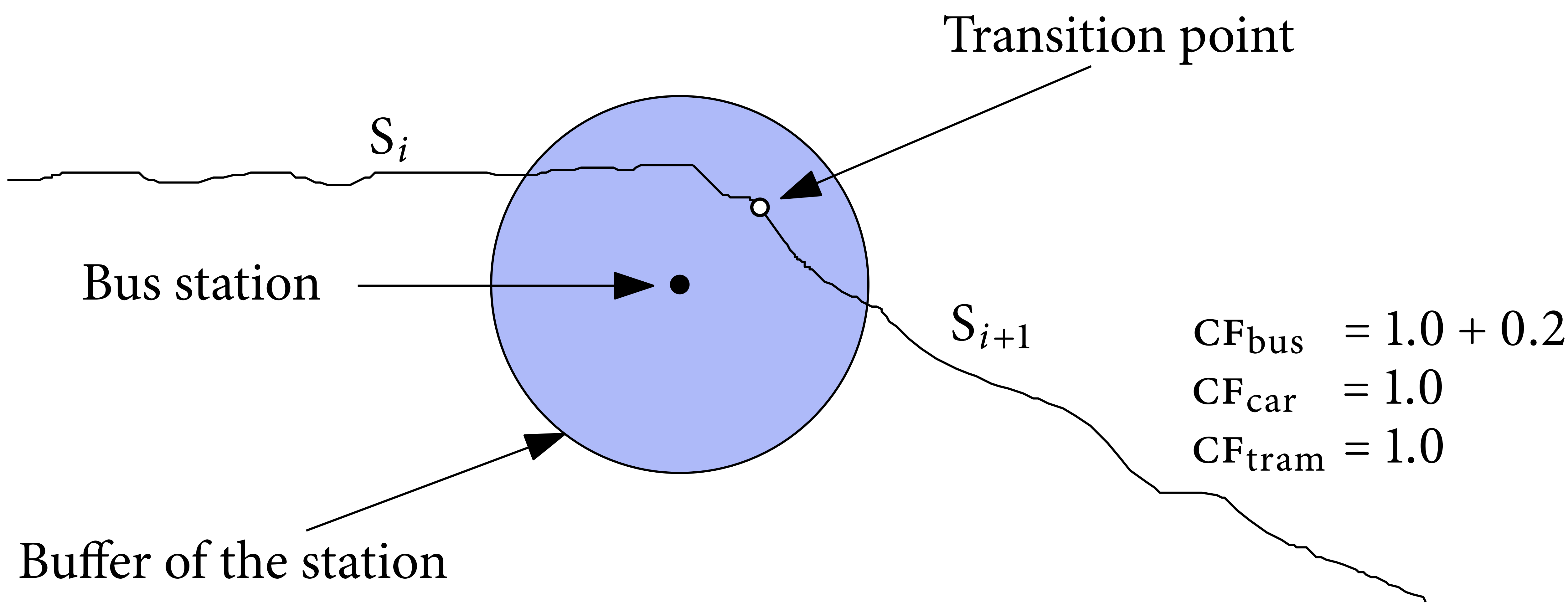
Distinguishing car, bus, and tram

- Infrastructure



Distinguishing car, bus, and tram

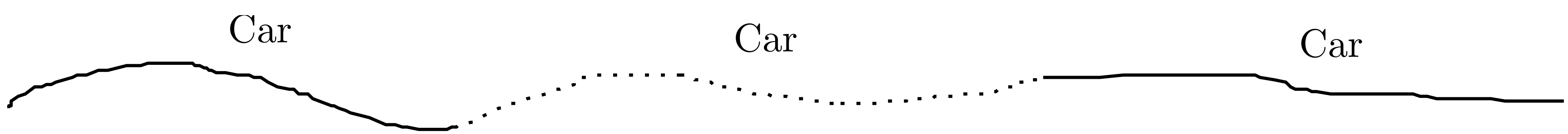
- Assisted by the location of nearest bus/tram stops



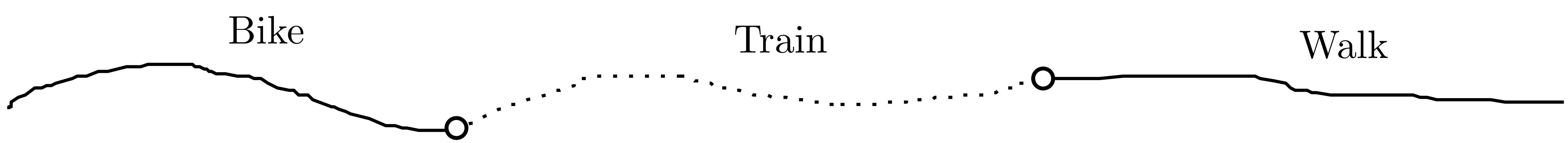
$$CF_{bus} = 1.0 + 0.2$$
$$CF_{car} = 1.0$$
$$CF_{tram} = 1.0$$

Missing time intervals

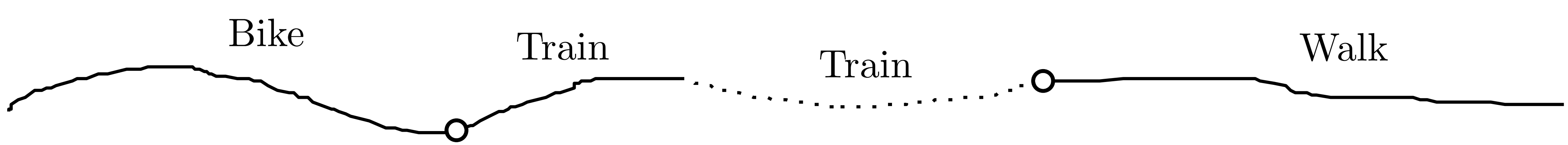
- Interpolation not possible



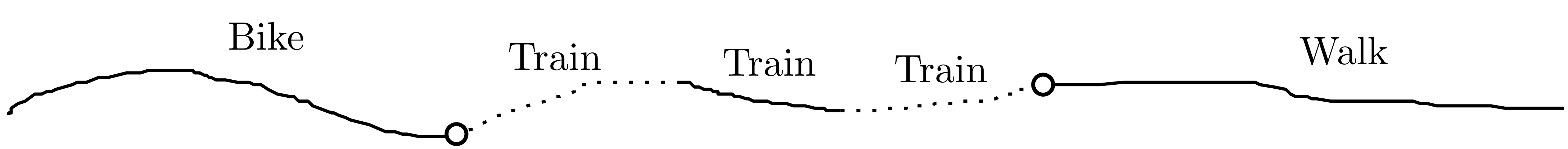
(a) "Regular" gaps where the mode was not changed.



(b) The whole segment done with another mode is not recorded.



(c) The first transition is recorded, but not the rest of the segment and the second transition.



(d) Neither of the transitions is detected, but a small fragment of the segment is recorded in between.

Missing time intervals (2)

- Example
- Attempted classification with geo-information and duration/distance of the gap
- Taking advantage of gaps for underground mode

Deriving additional mode-related information

- `Segment done by train.`
  `Departure station: Den Haag HS`
  `Arrival station: Delft`

- `Segment done by aircraft.`
  `Departure: Copenhagen, Denmark (CPH)`
  `Arrival:  Amsterdam, Netherlands (AMS)`
  `Carrier(s): Scandinavian Airlines System, KLM Royal Dutch Airlines`

TU Delft Delft University of Technology

Experiments

- A random subset of the dataset
- Available validated data (“ground truth”)
- Added a few tracks from abroad and specific situations

Experiments (results)

- Segmentation is precise and sensitive
- Very short segments are successfully detected and classified
- Long journeys (with a lot of observations), especially cars, are virtually always correctly classified
- Worldwide applicability (at least Europe)

Experiments (accuracy)

Quality of input data	Layers		
	1	**2**	**3**
Good GPS data	99.1%	94.5%	93.6%
Bad GPS data	99.0%	91.4%	89.2%
Total (all data)	99.0%	93.1%	91.6%

Experiments (problems)

- Very short trips with noisy points - human intervention not beneficial
- Attribution to water modes (in the Netherlands)
- Unusual behaviour (low or high speeds)
- Car, tram, bus - problems with incomplete data (combining them in one class - OK)

Conclusions

- Functional (extensible) prototype
- Geo-information is the key for solving this problem
- Openstreetmap is suitable for the classification
- Classification focused on removing unlikely classes

Conclusions (2)

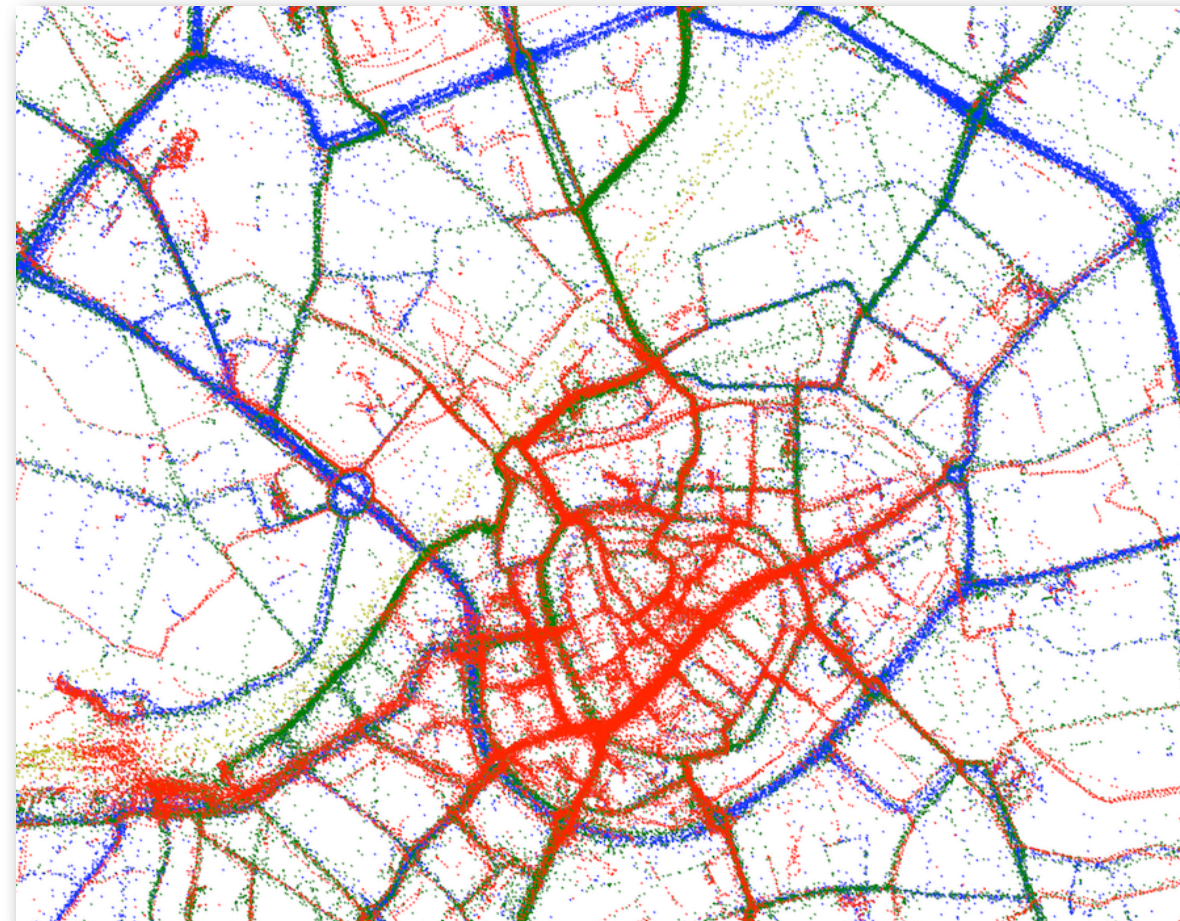
- More modes, higher accuracy, results with certainties
- Solving the gaps, coping with noise, classification of short segments
- Enriching the trajectories with more information

TU Delft Delft University of Technology

Future work

- Classification for trip purpose
- Established framework
- OSM data suitable as well

Questions?



TU Delft Delft University of Technology