Geospatial Data Management Final Project

Simone Laudani

Scenario

Geospatial data points logging during my visit in Berlin

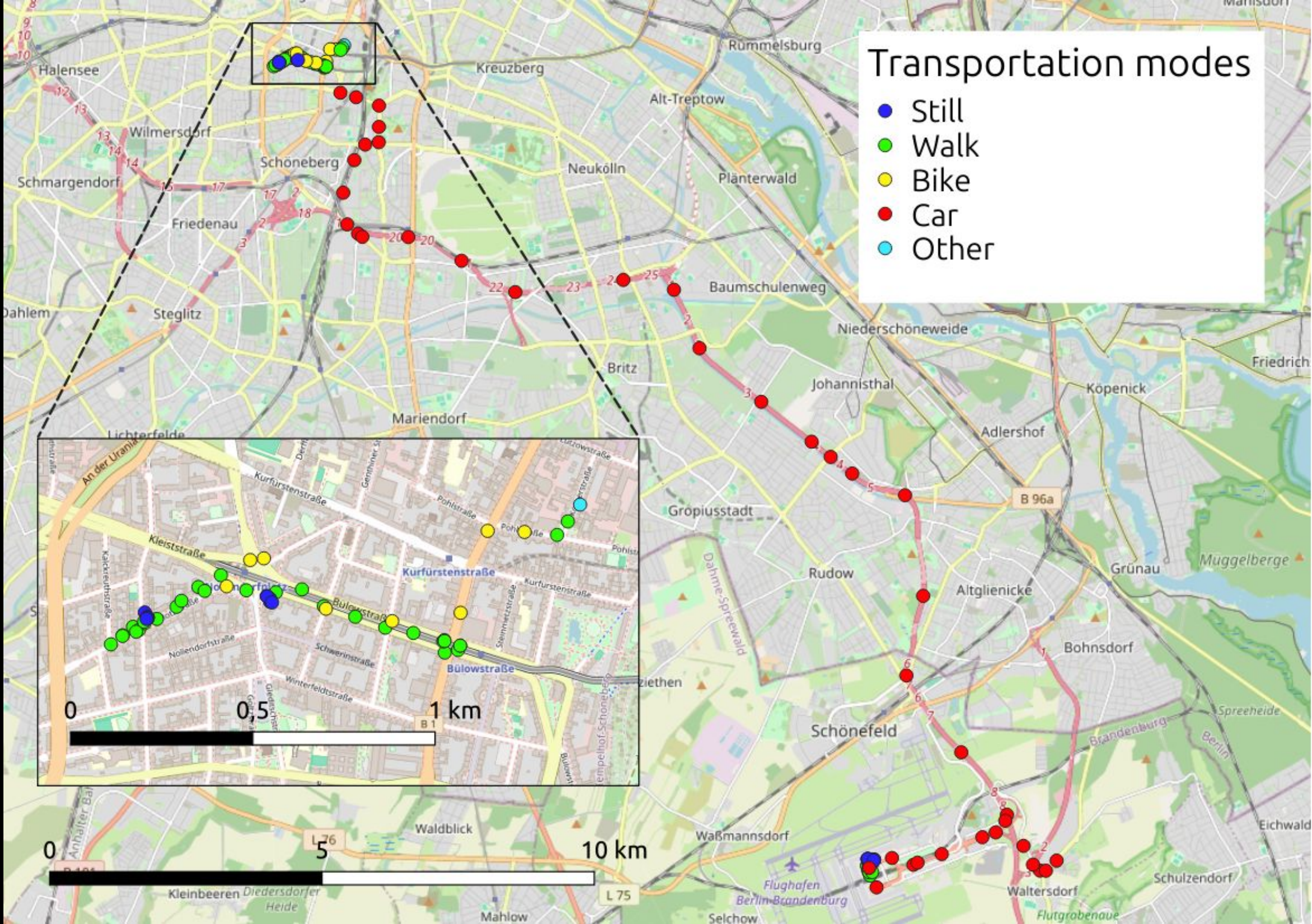
Using the phone app: GPSLogger

https://gpslogger.app/

Data scheme

- Geometry with reference system: EPSG:4326 WGS 84
- Time: UTC + timezone
- Transportation mean: still, walk, bike, car

Interval of acquisition set to 1 minute


Transportation modes
● Still
● Walk
● Bike
● Car
● Other
0
5
10 km
0
0,5
1 km
Halensee
Wilmersdorf
Schmargendorf
Schöneberg
Friedenau
Kreuzberg
Rummelsburg
Alt-Treptow
Neukölln
Plänterwald
Baumschulenweg
Niederschöneweide
Steglitz
Dahlem
Britz
Johannisthal
Köpenick
Friedrich
Mariendorf
Lichterfelde
Adlershof
Gropiusstadt
Rudow
Altglienicke
Grünau
Müggelberge
Bohnsdorf
Schönefeld
Waldblick
Waßmannsdorf
Flughafen Berlin-Brandenburg
Waltersdorf
Schulzendorf
Eichwald
Kleinbeeren
Diedersdorfer Heide
Mahlow
Selchow
Dahme-Spreewald
Spreeheide
Brandenburg
Berlin
Flutgrabenaue
B 96a
L 76
L 75
Kleiststraße
Kurfürstenstraße
Bülowstraße
Nollendorfstraße
Schwerinstraße
Winterfeldtstraße
Pohlstraße
An der Urania
Genthiner Str.
Steinmetzstraße
Lützowstraße
B 1
Tempelhof-Schöneberg

Objective

To infer transportation means for each data point,and getting the lowest possible error

Measuring the error ratio:

Error = number incorrect predictions / total points number

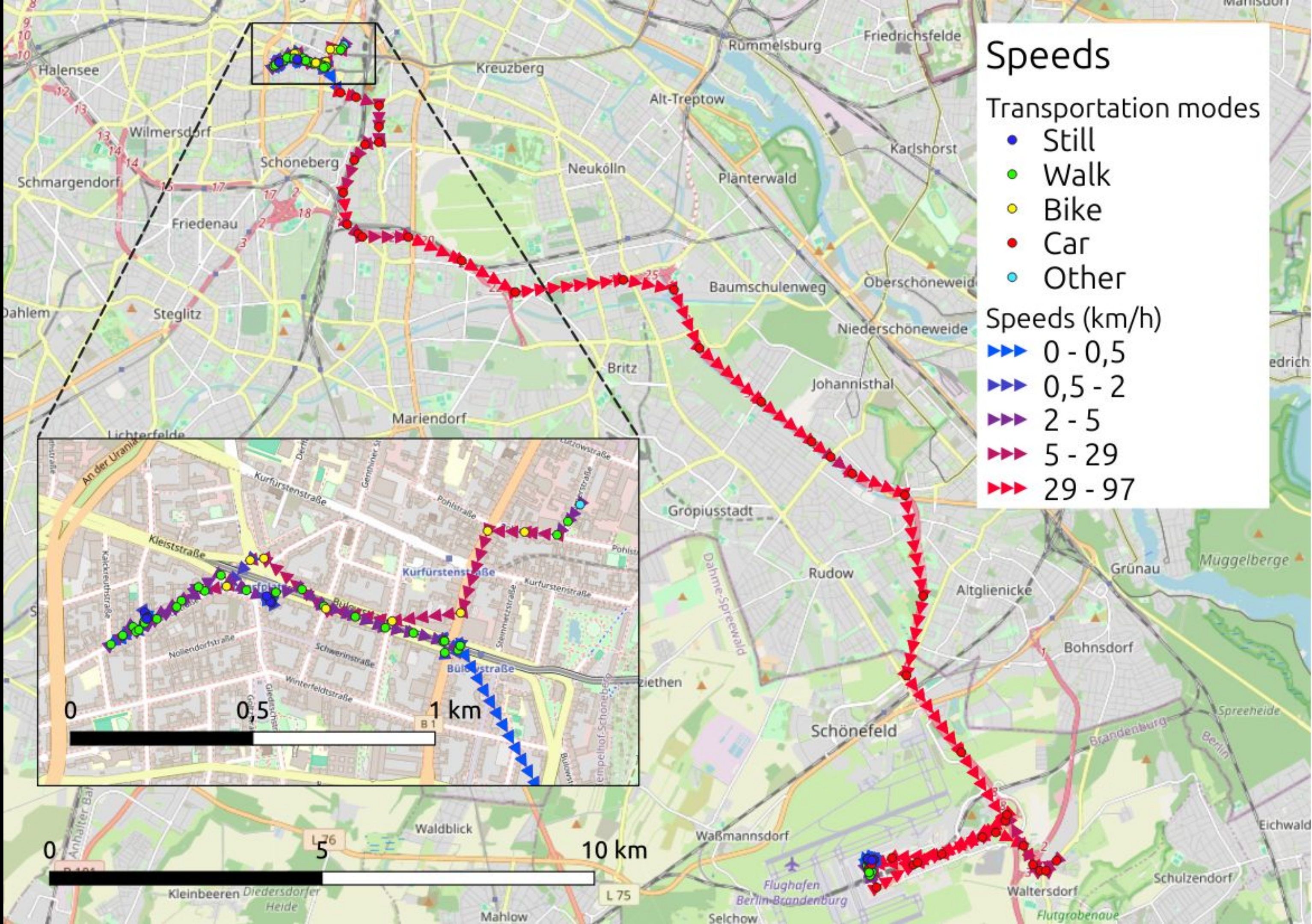
Data preprocessing

Create a new table with selected attributes:

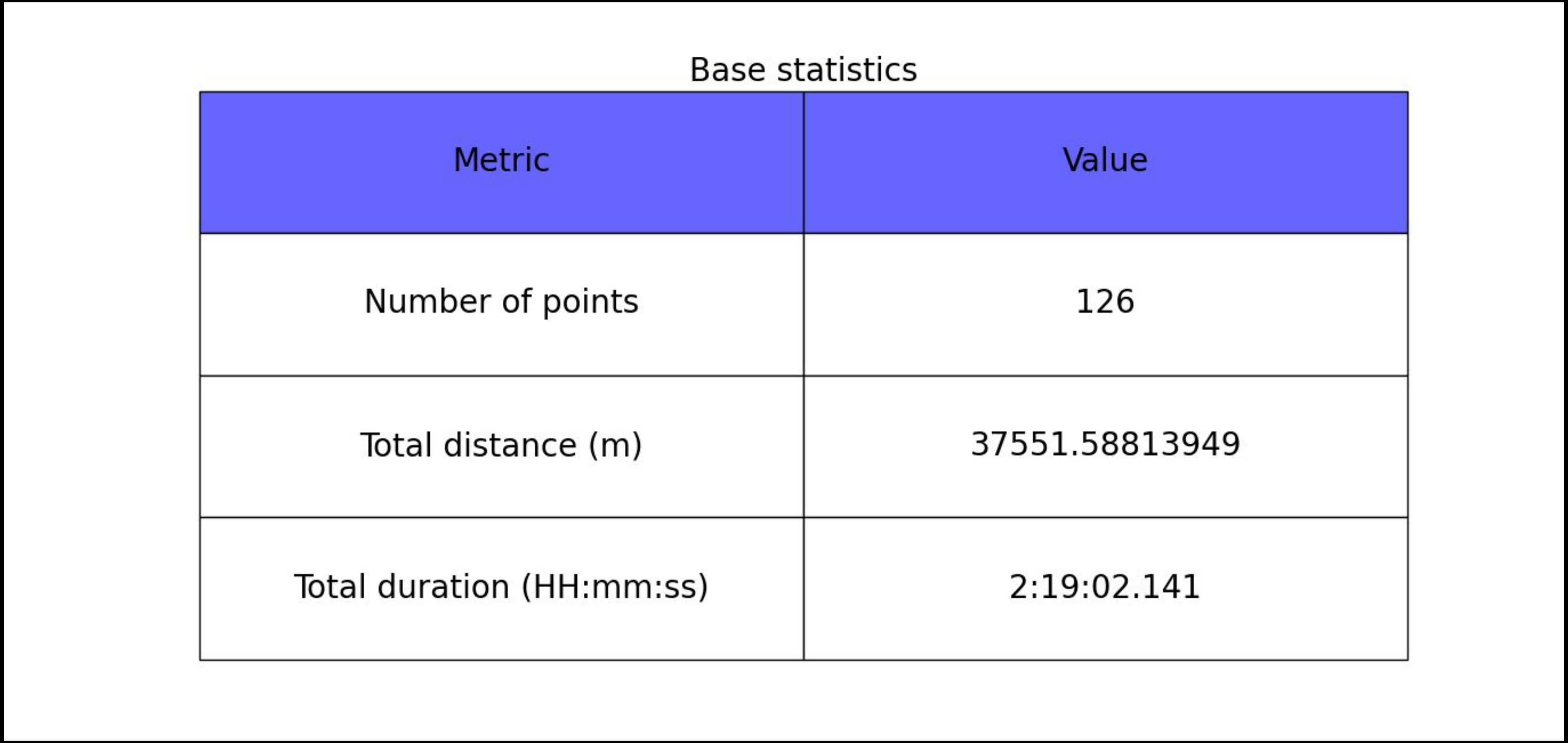
```
CREATE TABLE gps_20250520_processed AS
        (SELECT ogc_fid AS id, wkb_geometry, time, mode, src,
        LAG(wkb_geometry) OVER (ORDER BY time) AS prev_geom,
        LAG(time) OVER (ORDER BY time) AS prev_time
        FROM gps_20250520_original
        ORDER BY time);
```

Using *prev_geom* and *prev_time* fields to compute the spatial and temporal distances between two consecutive points

And then get the speed between the current and the previous point

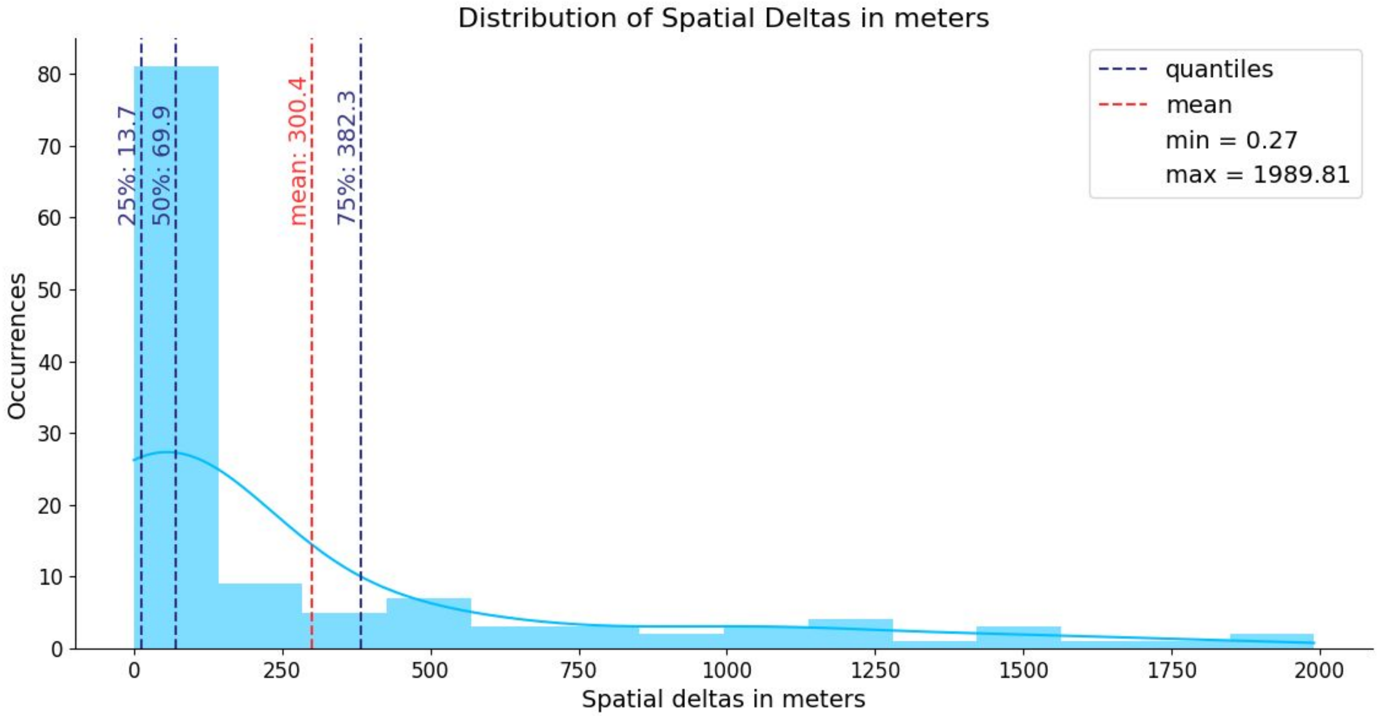

Speeds
Transportation modes
● Still
● Walk
● Bike
● Car
● Other
Speeds (km/h)
▶▶▶ 0 - 0,5
▶▶▶ 0,5 - 2
▶▶▶ 2 - 5
▶▶▶ 5 - 29
▶▶▶ 29 - 97
0
0,5
1 km
0
5
10 km
Halensee
Wilmersdorf
Schmargendorf
Schöneberg
Friedenau
Kreuzberg
Neukölln
Rummelsburg
Friedrichsfelde
Alt-Treptow
Plänterwald
Karlshorst
Baumschulenweg
Oberschöneweide
Niederschöneweide
Johannisthal
Britz
Dahlem
Steglitz
Mariendorf
Lichterfelde
Gropiusstadt
Rudow
Altglienicke
Grünau
Müggelberge
Bohnsdorf
Schönefeld
Waßmannsdorf
Waltersdorf
Schulzendorf
Eichwalde
Waldblick
Kleinbeeren
Diedersdorfer Heide
Mahlow
Selchow
Flughafen Berlin-Brandenburg
Flutgrabenaue
Spreeheide
Dahme-Spreewald
Brandenburg
Berlin
Ziethen
An der Urania
Kurfürstenstraße
Kleiststraße
Bülowstraße
Nollendorfstraße
Schwerinstraße
Winterfeldtstraße
Pohlstraße
Lützowstraße
Steinmetzstraße
Kalckreuthstraße
Tempelhof-Schöneberg

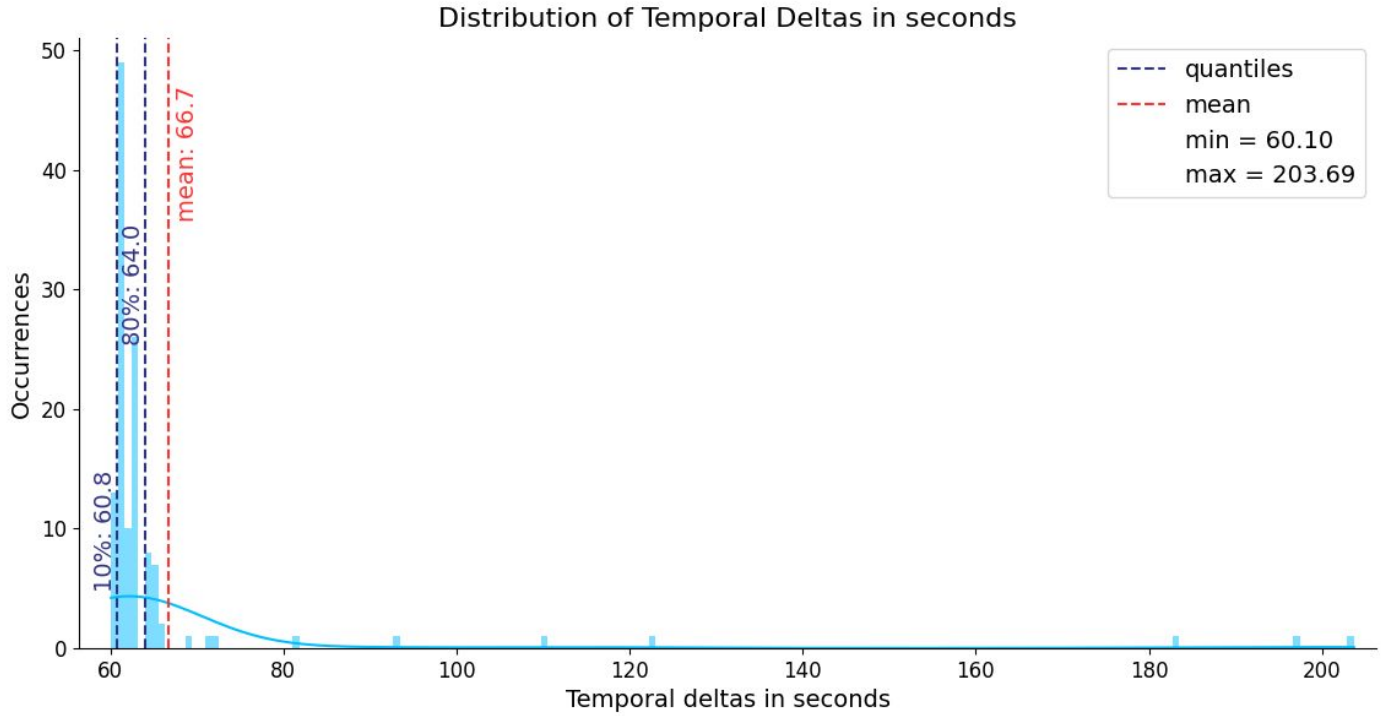
Statistics

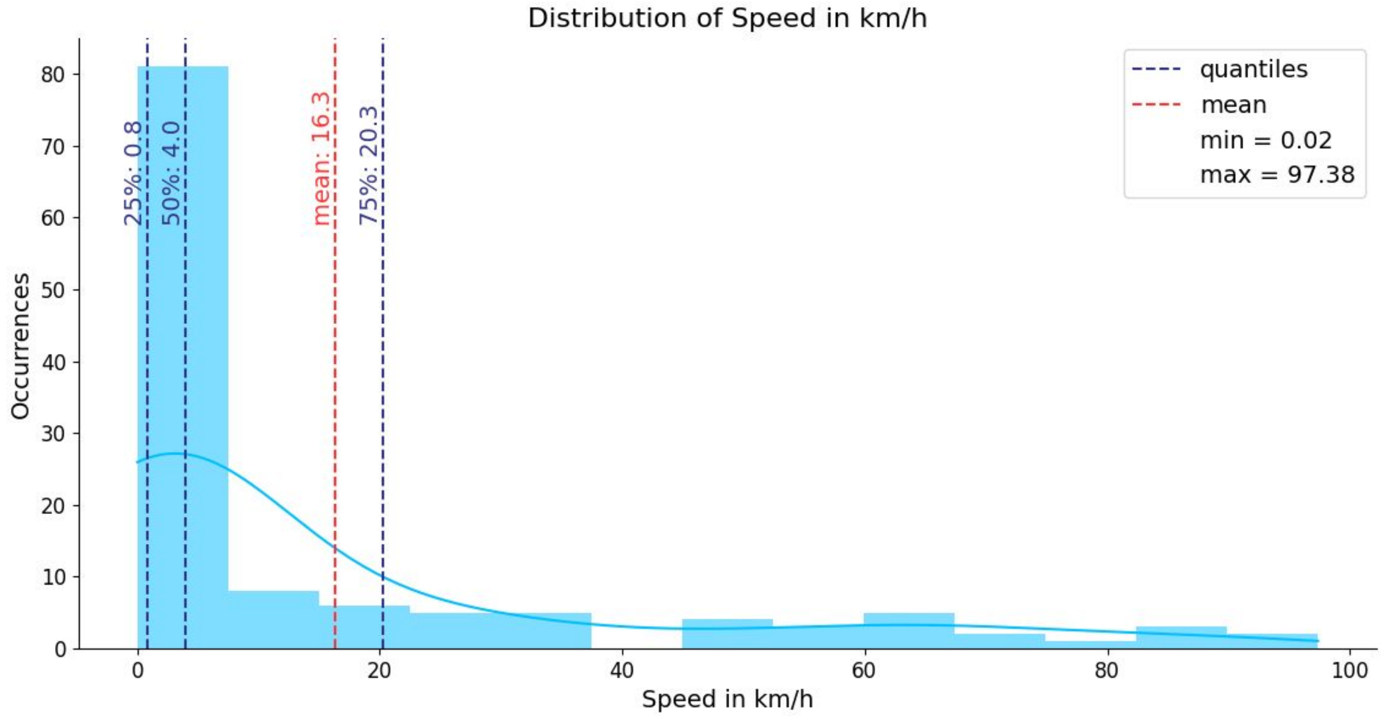


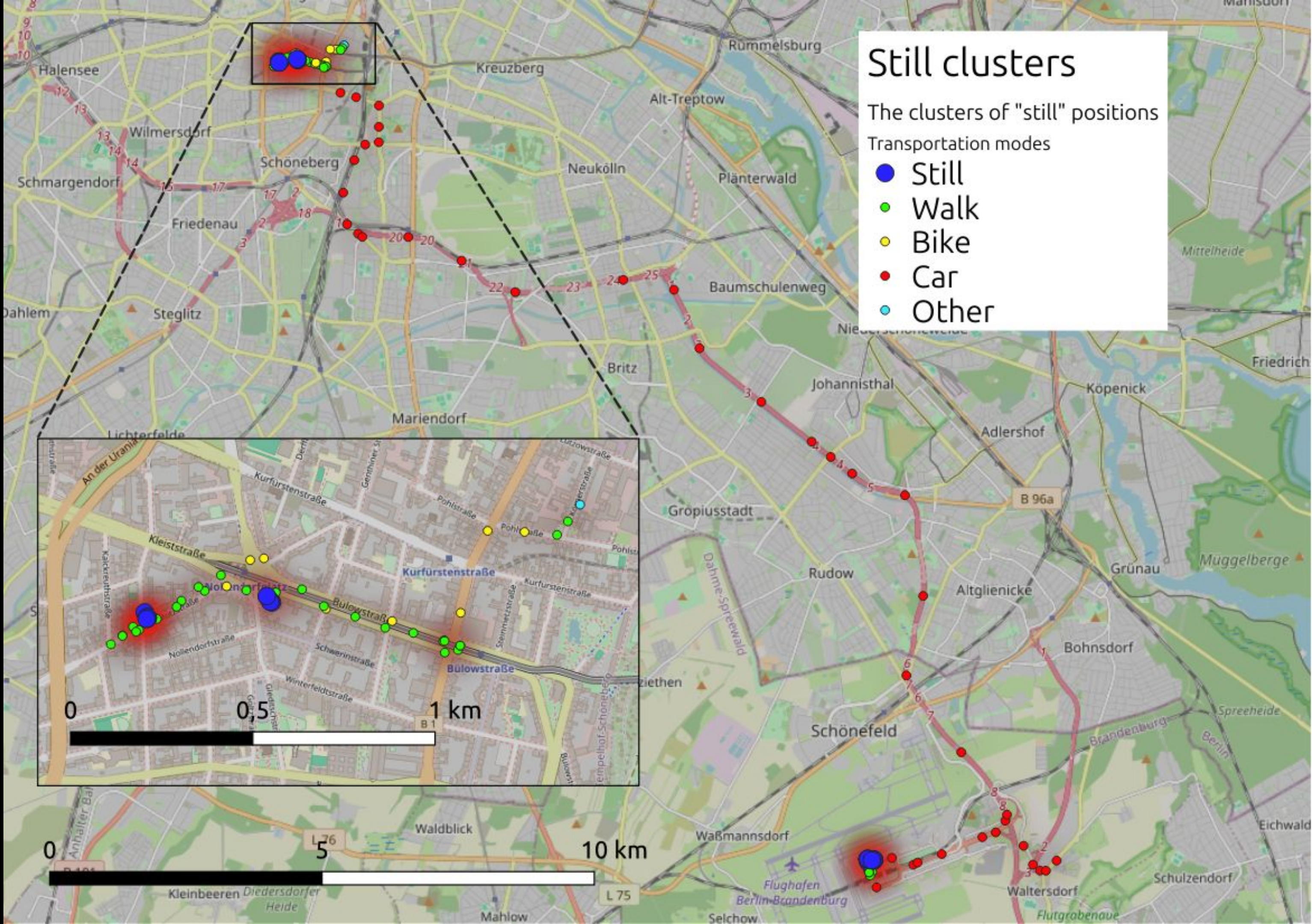
Base statistics

Metric	Value
Number of points	126
Total distance (m)	37551.58813949
Total duration (HH:mm:ss)	2:19:02.141


Distribution of Spatial Deltas in meters
quantiles
mean
min = 0.27
max = 1989.81
25%: 13.7
50%: 69.9
mean: 300.4
75%: 382.3
Occurrences
0
10
20
30
40
50
60
70
80
0
250
500
750
1000
1250
1500
1750
2000
Spatial deltas in meters


Distribution of Temporal Deltas in seconds
quantiles
mean
min = 60.10
max = 203.69
mean: 66.7
80%: 64.0
10%: 60.8
Occurrences
0
10
20
30
40
50
60
80
100
120
140
160
180
200
Temporal deltas in seconds


Distribution of Speed in km/h
quantiles
mean
min = 0.02
max = 97.38
25%: 0.8
50%: 4.0
mean: 16.3
75%: 20.3
Occurrences
0
10
20
30
40
50
60
70
80
Speed in km/h
0
20
40
60
80
100


Still clusters
The clusters of "still" positions
Transportation modes
● Still
● Walk
● Bike
● Car
● Other
Halensee
Wilmersdorf
Schmargendorf
Schöneberg
Friedenau
Kreuzberg
Rummelsburg
Alt-Treptow
Neukölln
Plänterwald
Baumschulenweg
Niederschöneweide
Dahlem
Steglitz
Britz
Johannisthal
Köpenick
Friedrich
Adlershof
Mariendorf
Lichterfelde
Gropiusstadt
B 96a
Rudow
Altglienicke
Grünau
Müggelberge
Mittelheide
Bohnsdorf
Dahme-Spreewald
Schönefeld
Spreeheide
Brandenburg
Berlin
Waldblick
Waßmannsdorf
Flughafen Berlin-Brandenburg
Waltersdorf
Schulzendorf
Eichwald
Kleinbeeren
Diedersdorfer Heide
Mahlow
Selchow
Flutgrabenaue
L 76
L 75
Anhalter Bahn
An der Urania
Kurfürstenstraße
Kleiststraße
Nollendorfplatz
Bülowstraße
Pohlstraße
Nollendorfstraße
Schwerinstraße
Winterfeldtstraße
Steinmetzstraße
Kalckreuthstraße
Genthiner Str.
Lützowstraße
Kurmärkische Straße
Ziethen
B 1
0
0,5
1 km
0
5
10 km

Solving approach

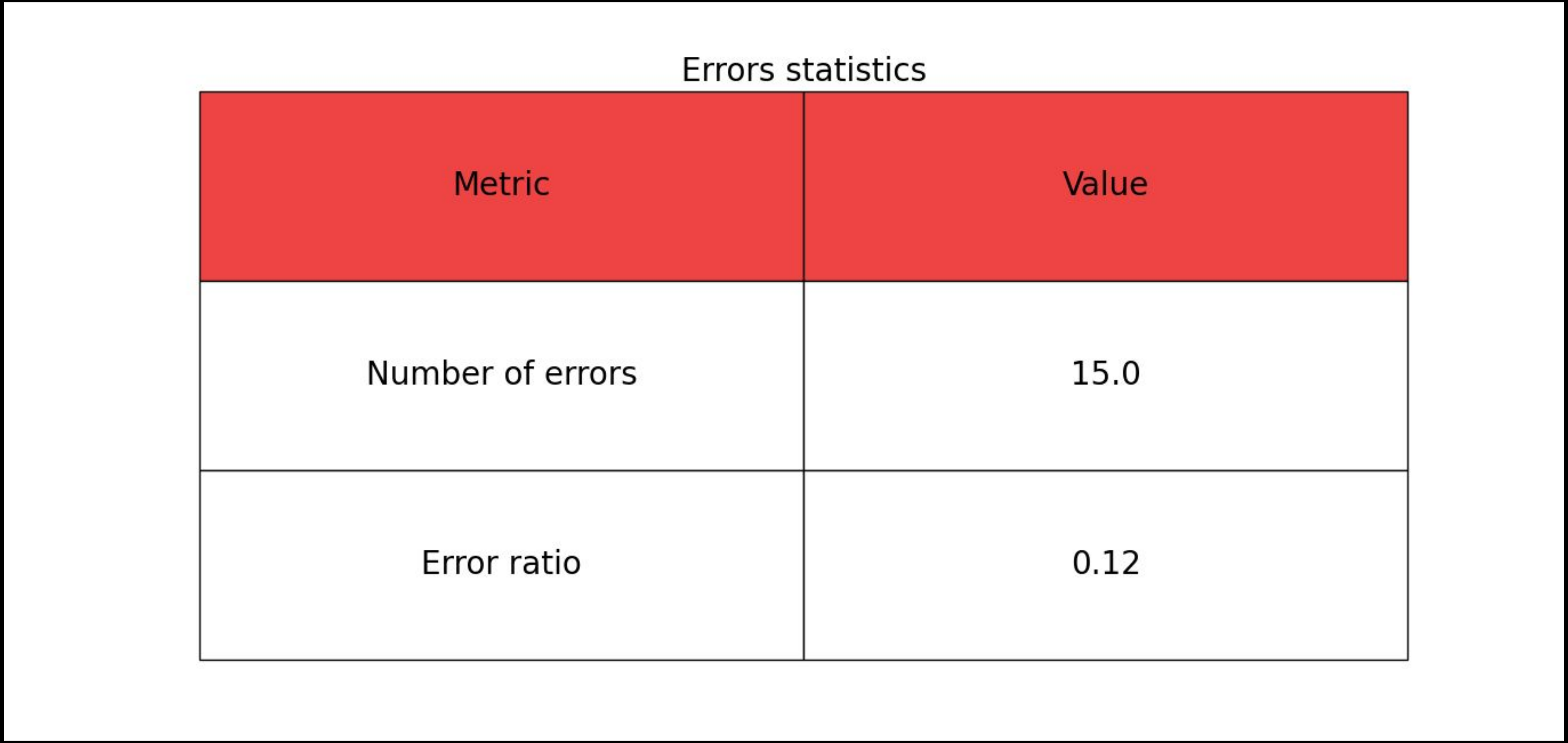
Python script to do a random search on range values:

```
thresholds = {
        'still-walk': np.arange(0, 5, .01),
        'walk-bike': np.arange(5, 10, .01),
        'bike-car': np.arange(10, 20, .01)
}
```

Updating the values on the data points table and comparing to the ground truth to find the minimal error ratio

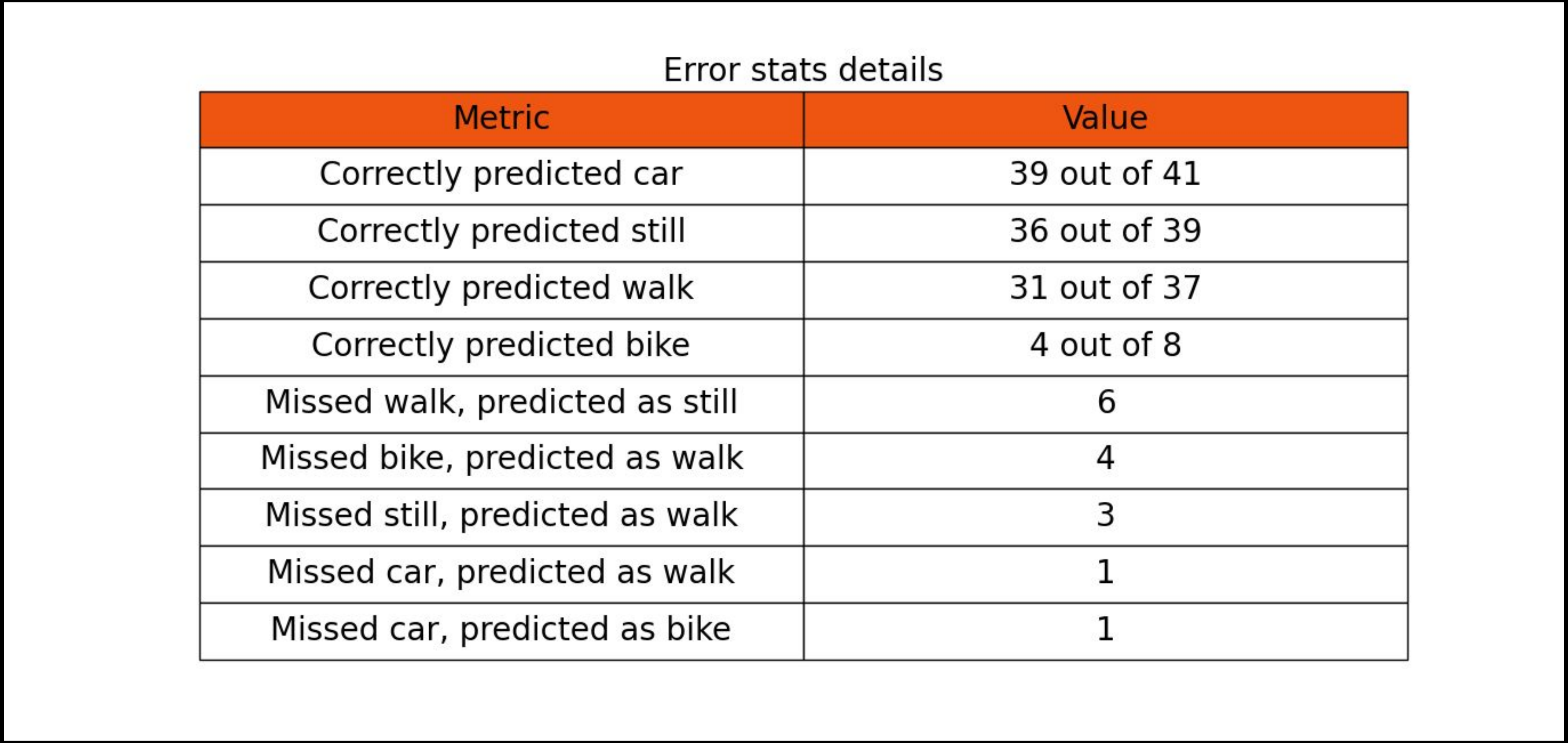
https://github.com/simone-05/gdm_project/blob/main/random_search_with_csv.py

Results



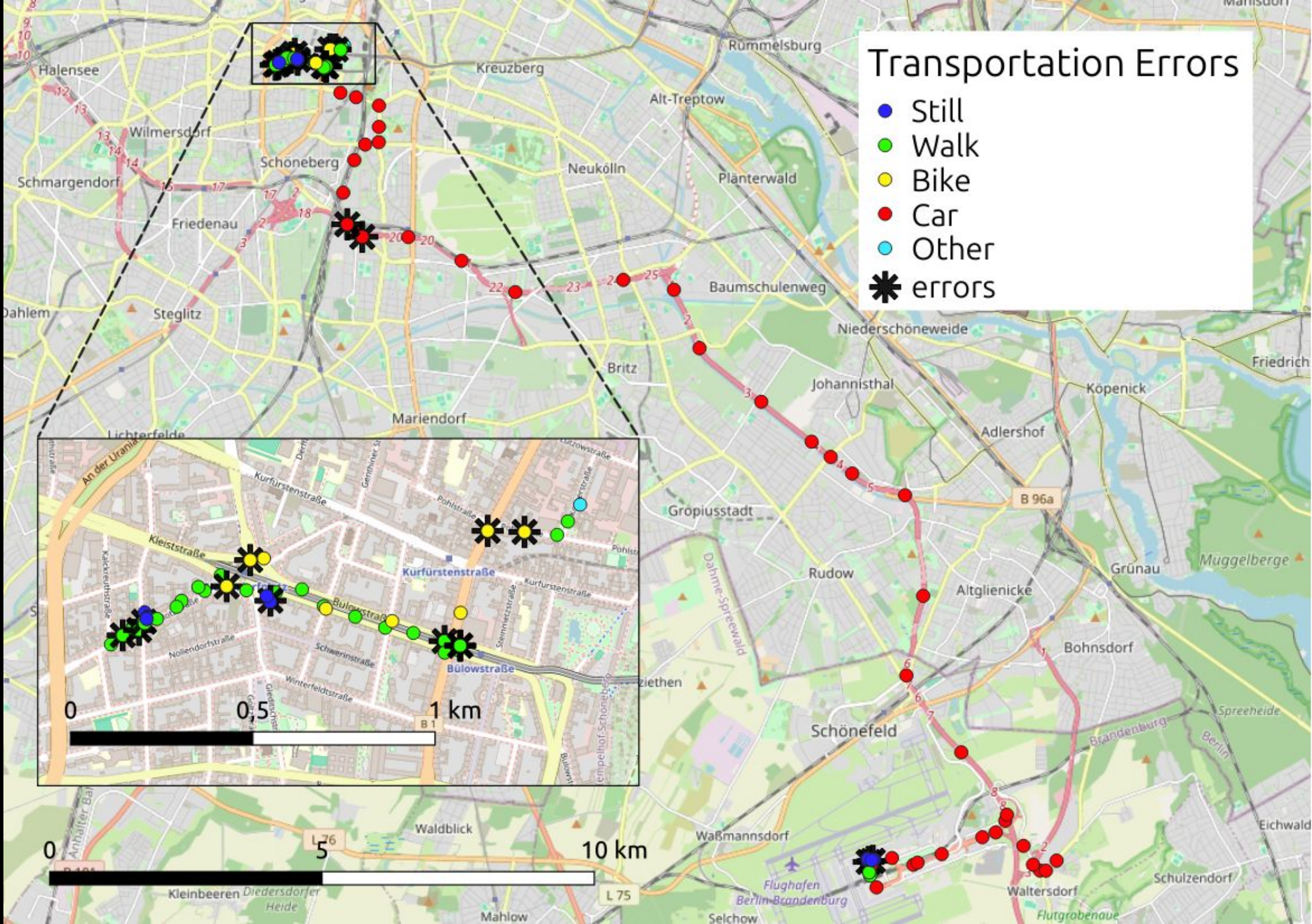
Errors statistics

Metric	Value
Number of errors	15.0
Error ratio	0.12



Error stats details

Metric	Value
Correctly predicted car	39 out of 41
Correctly predicted still	36 out of 39
Correctly predicted walk	31 out of 37
Correctly predicted bike	4 out of 8
Missed walk, predicted as still	6
Missed bike, predicted as walk	4
Missed still, predicted as walk	3
Missed car, predicted as walk	1
Missed car, predicted as bike	1


Transportation Errors
● Still
● Walk
● Bike
● Car
● Other
✱ errors
0
0,5
1 km
0
5
10 km
Halensee
Wilmersdorf
Schmargendorf
Schöneberg
Friedenau
Steglitz
Dahlem
Kreuzberg
Rummelsburg
Alt-Treptow
Neukölln
Plänterwald
Baumschulenweg
Niederschöneweide
Britz
Johannisthal
Köpenick
Adlershof
Gropiusstadt
Rudow
Altglienicke
Grünau
Müggelberge
Bohnsdorf
Schönefeld
Waßmannsdorf
Waldblick
Flughafen Berlin-Brandenburg
Waltersdorf
Schulzendorf
Eichwalde
Kleinbeeren
Diedersdorfer Heide
Mahlow
Selchow
Mariendorf
Lichterfelde
Kurfürstenstraße
Kleiststraße
Bülowstraße
Nollendorfstraße
Schwerinstraße
Winterfeldtstraße
Pohlstraße
An der Urania

End