

# EE 495 Assignment 2: Probe Data Analysis for Road Slope

By: Christopher Tsai and Anuj Karnik

# Link Matching Methodology

1. Input a probe point of the form [latitude, longitude]
2. Create Pandas DataFrame that contains ['linkPVID', 'shapelInfo', 'shapeList'] columns from 'Partition6467LinkData.csv'
3. Add a column 'distFromLink' which contains the distance from the probe point to the nearest link
  - a. This is done by applying a Lambda function to the entries of the 'shapelInfo' column of the DataFrame
  - b. This Lambda function takes the minimum Great Circle distance (using haversine Python library function) of the probe point to all [lat, lon] pairs from 'shapelInfo' column of DataFrame
4. Now that we have the distance to nearest link, we identify it by acquiring the corresponding link's 'linkPVID'

# Link Matching Methodology

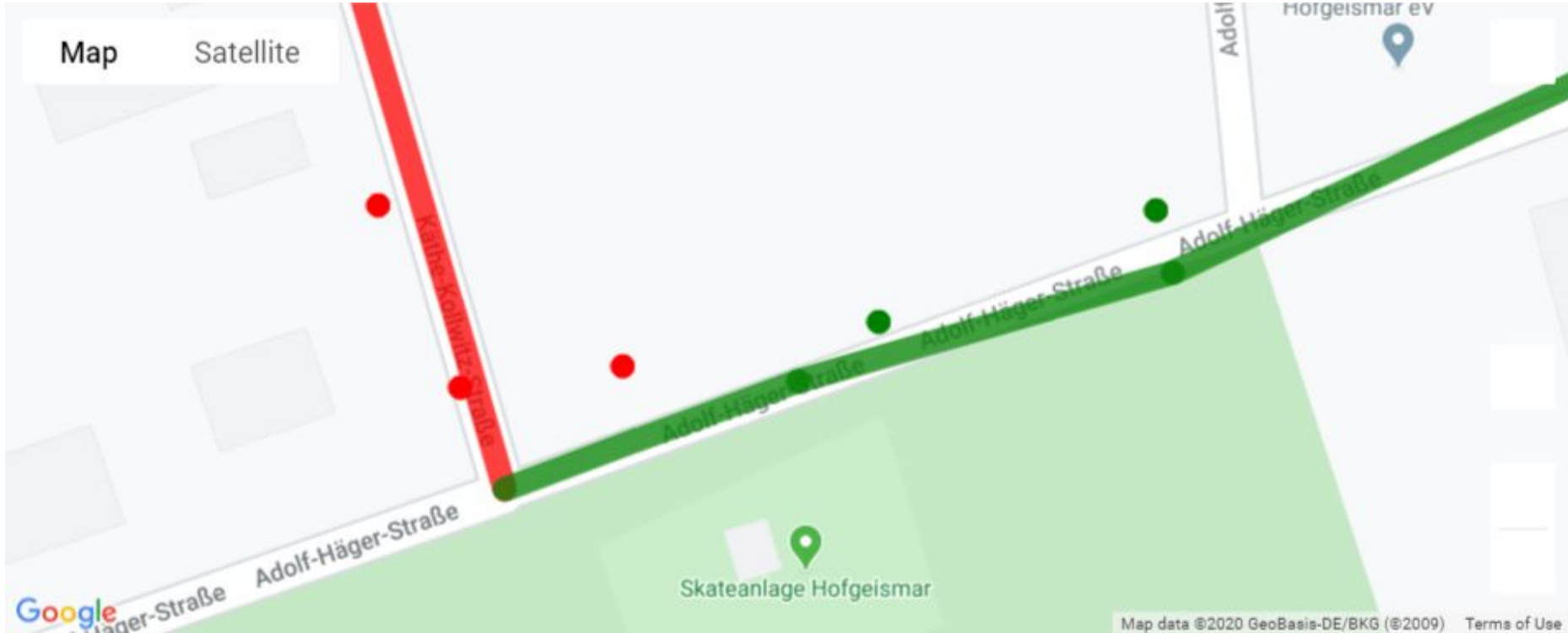
5. Our output file 'Partition6467MatchedPoints.csv' also requires us to also output 'distFromLink': the distance of each probe point to reference node of link. To do this:
  - a. Acquire corresponding index of the 'linkPVID' in DataFrame
  - b. Use this index to get first entry of 'shapelnfo' of each link, which is the ref node
  - c. Calculate Great Circle distance from probe point to ref node

# Link Matching Results (sample size = 500)

|     | sampleID | dateTime               | sourceCode | latitude  | longitude | altitude | speed | heading | linkPVID  | directionOfTravel | distFromRef | distFromLink |
|-----|----------|------------------------|------------|-----------|-----------|----------|-------|---------|-----------|-------------------|-------------|--------------|
| 0   | 3496     | 2009-06-12<br>06:12:49 | 13         | 51.496868 | 9.386022  | 200      | 23    | 339     | 62007637  | B                 | 35.125020   | 35.125020    |
| 1   | 3496     | 2009-06-12<br>06:12:54 | 13         | 51.496682 | 9.386157  | 200      | 10    | 129     | 62007637  | B                 | 12.429231   | 12.429231    |
| 2   | 3496     | 2009-06-12<br>06:12:59 | 13         | 51.496705 | 9.386422  | 201      | 21    | 60      | 62007637  | B                 | 19.238924   | 19.238924    |
| 3   | 3496     | 2009-06-12<br>06:13:04 | 13         | 51.496749 | 9.386840  | 201      | 0     | 360     | 567329767 | B                 | 46.237199   | 11.159886    |
| 4   | 3496     | 2009-06-12<br>06:13:09 | 13         | 51.496864 | 9.387294  | 199      | 0     | 360     | 567329767 | B                 | 80.145179   | 7.332394     |
| ... | ...      | ...                    | ...        | ...       | ...       | ...      | ...   | ...     | ...       | ...               | ...         | ...          |
| 495 | 4558     | 2009-06-12<br>08:34:34 | 13         | 53.073286 | 8.854174  | 45       | 55    | 289     | 811152825 | B                 | 153.829975  | 33.072664    |
| 496 | 4558     | 2009-06-12<br>08:34:39 | 13         | 53.073512 | 8.853083  | 45       | 56    | 289     | 811152826 | B                 | 110.157338  | 42.968140    |
| 497 | 4559     | 2009-06-12<br>08:18:15 | 13         | 53.071319 | 8.864019  | 49       | 0     | 164     | 51872166  | F                 | 42.000878   | 18.711507    |
| 498 | 4559     | 2009-06-12<br>08:18:21 | 13         | 53.071320 | 8.864019  | 49       | 0     | 156     | 51872166  | F                 | 42.096502   | 18.761808    |
| 499 | 4559     | 2009-06-12<br>08:18:26 | 13         | 53.071320 | 8.864018  | 49       | 0     | 168     | 51872166  | F                 | 42.128622   | 18.726213    |

500 rows × 12 columns

# Visualizing Probe Points and Links in gmaps



# Slope Calculation Methodology

1. Use Pandas function `.groupby()` to group Link Matching DataFrame rows to each 'linkPVID'
2. Acquire probe points that matched to each 'linkPVID'
3. Iterate through probe points, calculating distance between them and altitude difference between them
4. Calculate slope for each pair in degrees using  $\arctan(\text{changeInAltitude}/\text{distance})$
5. After iterating, take average of all slopes of each link and output it
6. Calculate error between this and the average of slopes per link of 'Partition6467LinkData.csv'

# Slope Results (sample size = 500)

|     | linkPVID  | slopesList  | averageSlope | groundTruthAverageSlope | absAverageSlopeError |
|-----|-----------|---|--------------|-------------------------|----------------------|
| 0   | 51865408  | [-0.014848655965818238, 0.0, 0.322861327837228... | -0.014849    | 0.000                   | 0.014849             |
| 1   | 51865408  | [-0.014848655965818238, 0.0, 0.322861327837228... | -0.007424    | 0.000                   | 0.007424             |
| 2   | 51865408  | [-0.014848655965818238, 0.0, 0.322861327837228... | 0.102671     | 0.000                   | 0.102671             |
| 3   | 51865408  | [-0.014848655965818238, 0.0, 0.322861327837228... | 0.077003     | 0.000                   | 0.077003             |
| 4   | 51866677  | [-0.04937019685324395, 1.108069157942699]         | -0.049370    | -0.014                  | 0.035370             |
| ... | ...       | ...   | ...          | ...                     | ...                  |
| 314 | 811768917 | [0.0, -1.103392124292288, 0.0, 0.0, 0.0, 0.0, ... | -0.183899    | 0.000                   | 0.183899             |
| 315 | 811768917 | [0.0, -1.103392124292288, 0.0, 0.0, 0.0, 0.0, ... | -0.157627    | 0.000                   | 0.157627             |
| 316 | 811768917 | [0.0, -1.103392124292288, 0.0, 0.0, 0.0, 0.0, ... | -0.137924    | 0.000                   | 0.137924             |
| 317 | 811768917 | [0.0, -1.103392124292288, 0.0, 0.0, 0.0, 0.0, ... | -0.122599    | 0.000                   | 0.122599             |
| 318 | 811768917 | [0.0, -1.103392124292288, 0.0, 0.0, 0.0, 0.0, ... | -0.110339    | 0.000                   | 0.110339             |

319 rows × 5 columns

Average absAverageSlopeError = 0.24347067152685037 degrees

# Results Discussion

- From visualizing the probe points and links in gmaps, we can observe that points are matched to corresponding link correctly for the most part
- Slope results turned out to be relatively similar to the theoretical slope values, but total average error of 0.24 degrees does seem high considering that the error is higher than some slope values themselves



# Problems and Improvements

- As mentioned before, slope accuracy can be improved, possibly by using a higher sample size or more robust algorithms
- The main reason such a small sample size was used (when compared to the total sample size) is that the code is very slow
- Other metrics can be combined with Great Circle distance to determine link matching