Estimating traffic flows from vehicle trajectories based on sparse mobile phone geolocation data

Roos Teeuwen & Jorge Gil, *Chalmers University of Technology, Gothenburg, Sweden*
roos.teeuwen@chalmers.se, jorge.gil@chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY

AIM

Large-scale empirical traffic flow data are useful in spatial planning, transport research, and data-driven decision making, but existing data, collected using traffic sensors or counts, lack spatio-temporal coverage and granularity. Mobile phone geolocation data are promising for capturing traffic flows at scale given their size, granularity, and coverage, but their potential for such cases remains unexplored.

We investigate how traffic flows with high spatial and temporal coverage and granularity can be estimated from vehicle trajectories based on sparse mobile phone geolocation data.

DATA

Raw mobile phone GNSS data from the aggregator company *Pickwell,* covering Sweden whole of 2024.

- Stockholm: 1292 million data points; from 3.2 million devices.
- Gothenburg: 670 million data points; from 1.4 million devices.

Ground truth from traffic counts in Stockholm and Gothenburg, covering a combination of municipal and national roads, from different years.

Road network from transport authority.

Trajectories	*Stockholm*	*Gothemburg*
Records	85,008,151	45,917,250
Avg total length	902,544661	932,78
Avg total duration	113,645941	110,527622
Avg points	4,387886	4,35075777
Avg dist. points	241,874996	254,596403

How well do estimates capture the ground truth?

Mostly **strongly correlated** (rho, p< 0.01)

- **0.707**, Stockholm's 2014 municipal (n=9744)
- **0.687**, Stockholm's 2024 municipal (n=366)
- **0.716**, Gothenburg's 2023 municipal (n=176)
- **0.490**, Gothenburg's 2023 national (n=245)

APPROACH

Developed a data processing methodology, implemented it for two major cities in Sweden, and compared the outcomes to ground truth data.

Case study cities
1. Stockholm
2. Gothenburg

DATA & PREPROCESSING

Raw geolocation traces
Sweden, 2024

Road network data
Sweden, 2025

Ground truth data
- Stockholm 2024
- Stockholm 2014
- Gothenburg 2023 municipal roads
- Gothenburg 2023 national roads

- **Clip** case study cities
- **Categorise** point types
- **Filter** on point types
(exclude stationary, dangles, errors)

Geolocation points (2 cities)

- **Clip** case study cities
- **Split** long road segments
- **Modify** data structure

Road network (2 cities)

- **Filter** data

Ground truth data sets (4x)

ESTIMATING FLOWS

- **Group** trajectory points
- **Clean** groups
- **Link** into trajectories

Trajectory - points with same device ID, between stays, s>3 km/h and/or d > 100m

Group trajectories between short stays (t<120s)
Split sudden jumps (d>1000m, t>120s, s>250 km/h)

Trajectories (2 cities)

- **Map-match** trajectories
- **Count** trajectories per road segment

***LCSS map-matching algorithm** from mappymatch Python library.*

Similarity score parameter 0.95 and other parameters with default values

Iterations, varying in amount and time frames of input data

Flows (2 cities)

EVALUATION

- **Sample** evaluation points

Sample-point flows

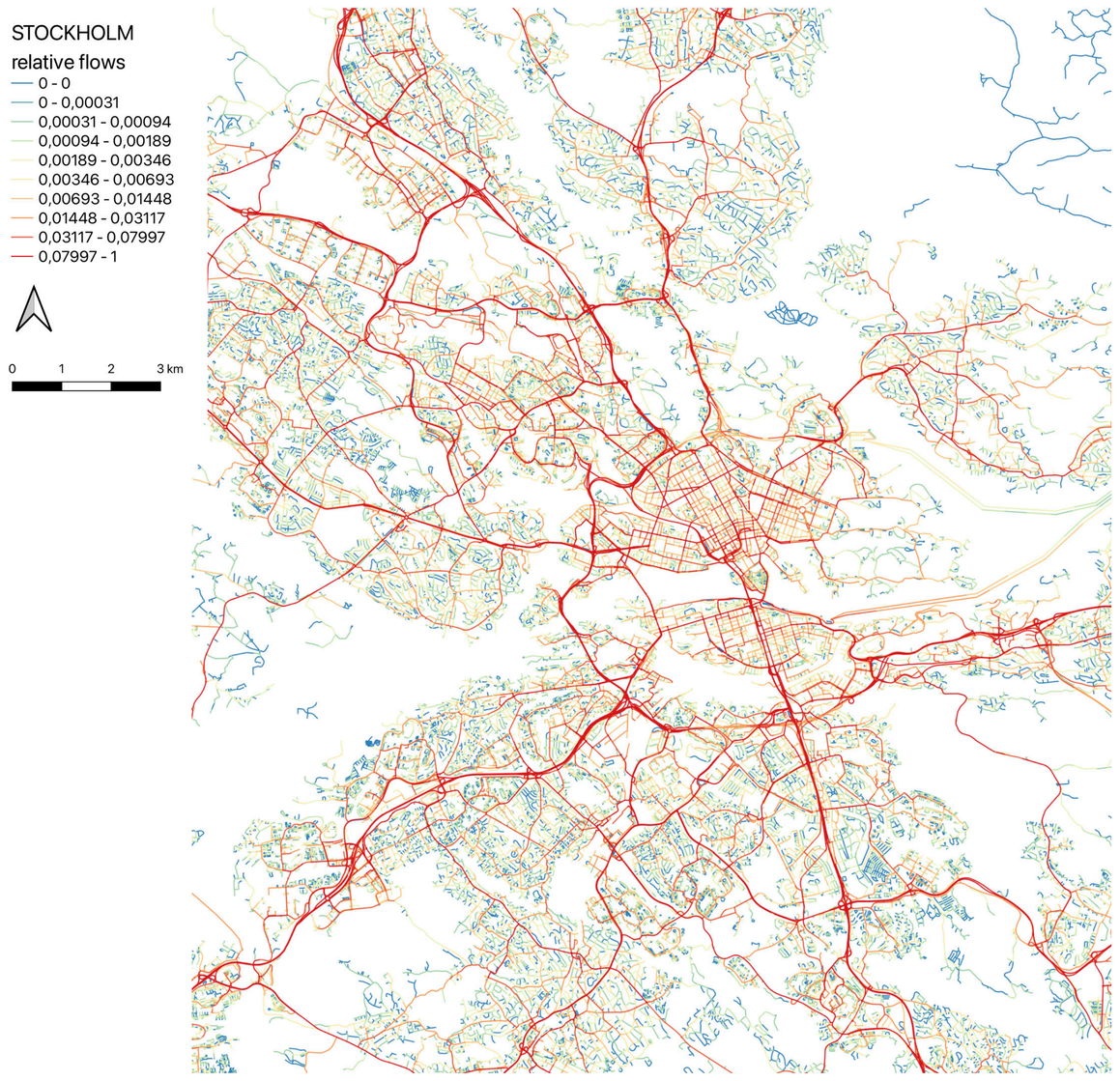
- Quantitative evaluation: correlation and coverage per iteration

Deviations

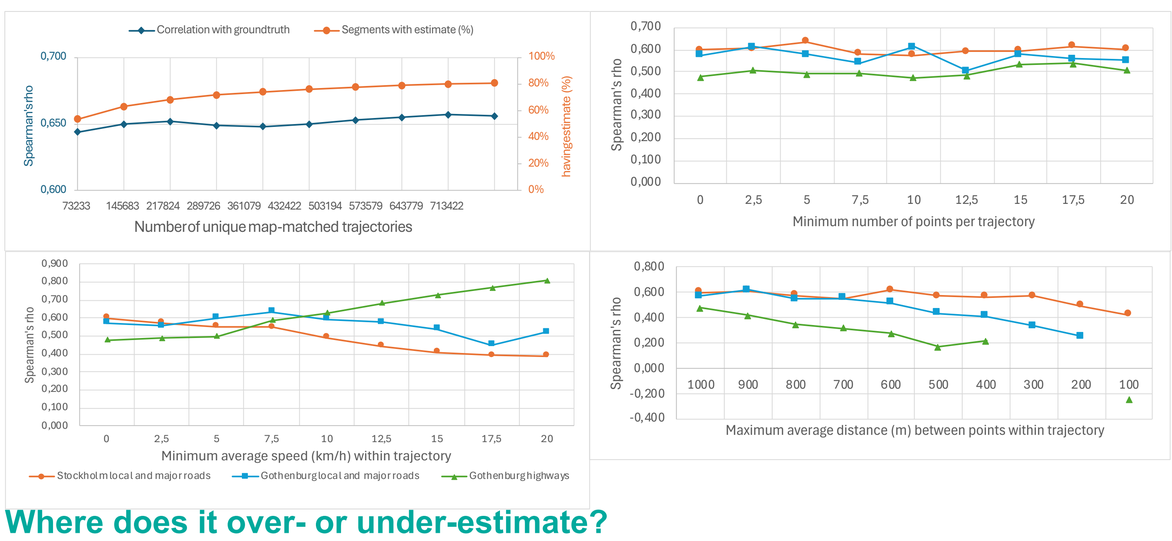
- **Qualitative** evaluation: road types and areas

Contributions:
1. Method
2. Flow data
3. Recommendations

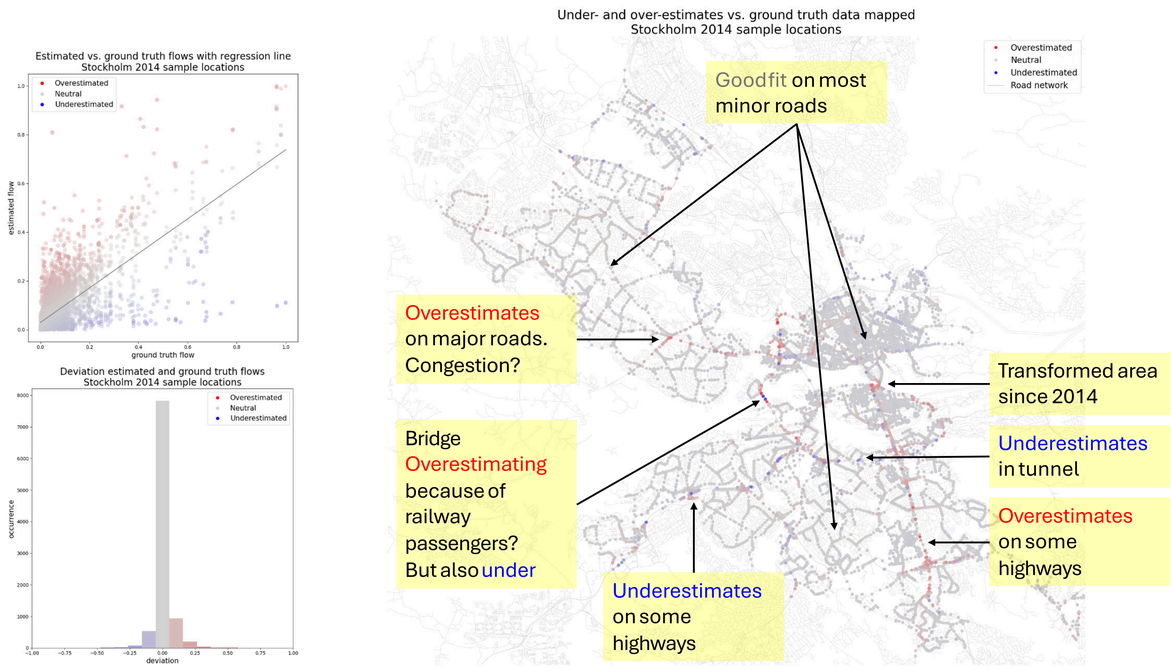
RESULTS



How do input trajectories affect these correlations?



Where does it over- or under-estimate?



KEY TAKE-AWAYS

1. Methodology works on sparse mobile phone GNSS data
2. Small sample of 1 million trajectories sufficient for decent results
3. Mobile phone data does not work in tunnels
4. Correlations are strong and can be improved: increase data size, use road categories, break road segments for improved matching

Repositories:
https://github.com/rflteeuwen/flowsense_trafficflows
https://zenodo.org/records/16794871

References: J. C. Herrera, D. B. Work, R. Herring, et al., "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment," Transp. Res. Part C: Emerg. Technol. 18, 568 (2010); C. Bongiorno, Y. Zhou, M. Kryven, et al., "Vector-based pedestrian navigation in cities," Nat. Comput. Sci. 1, 678 (2021); L. Zhu, J. R. Holden, and J. D. Gonder, "Trajectory Segmentation Map-Matching Approach for Large-Scale, High-Resolution GPS Data," Transp. Res. Rec. J. Transp. Res. Board 2645, 67 (2017); G. Boeing, "Modeling and Analyzing Urban Networks and Amenities With OSMnx," Geogr. Anal. p. gean.70009 (2025).