Estimating macroscopic fundamental diagrams from mobile phone data

June 01, 2017

Thierry Derrmann, PhD candidate

SnT – Interdisciplinary Centre for Security, Reliability and Trust

Introduction (1)

Intuition:

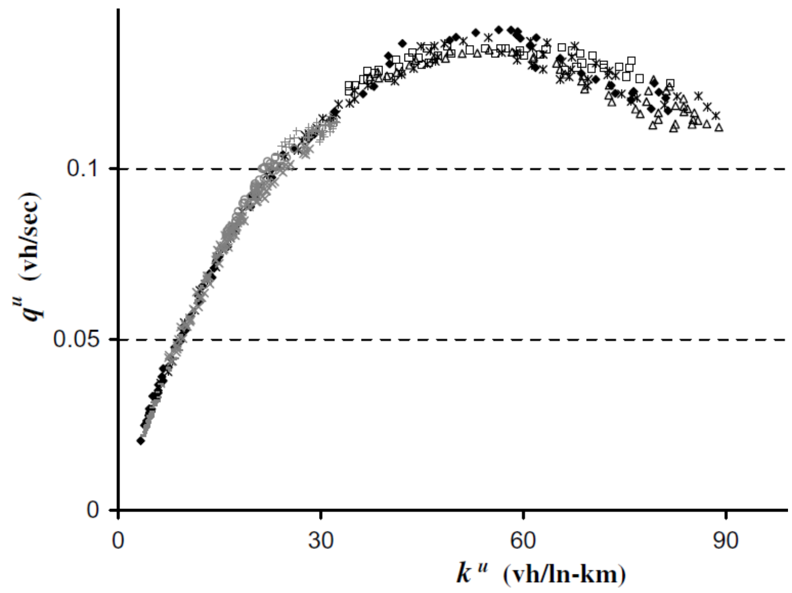
- Human mobility is reflected in mobile phone networks
- People enter and leave antenna coverage areas as they move
- We should be able to describe traffic states based on the data the mobile network generates

→ Sounds a lot like…

Introduction (2)

Macroscopic Fundamental Diagram

Characteristic relation between flows leaving an area and the vehicle density within it.



Yokohama MFD, Geroliminis et al.

Introduction (3)

Fo research, mobile network data comes in 2 main types:

- Billing Data:
  - Typically **Call Detail Records**
  - Low resolution, 1 entry per call / text message / data session

- Signaling Data:
  - **Handovers**: active connection is transferred between network cells

    High spatial resolution, low temporal coverage
  - **Location Area** updates: unconnected phone moves between cell groups

    Low spatial resolution, high temporal coverage

Motivation

Why do we want to estimate **MFDs from mobile network data**?

- The infrastructure is there (**zero cost**)
- **Wide coverage**, even in developing countries
  (often better than loop detector coverage)
- No **privacy** concern
  (works with aggregated data)

→ Simulation and Real Data Studies

Simulation Study (1)

Goal:

See if 'mobile network MFDs' emerge from mobile network data in a controlled environment

Idea:

Simulate mobile network and cars moving around, connecting to the network
evaluate mobile network MFDs and see if they can be used for velocity estimation

Solution: Veins-LTE + LuST scenario + handovers = **LuST-LTE**

Publication:

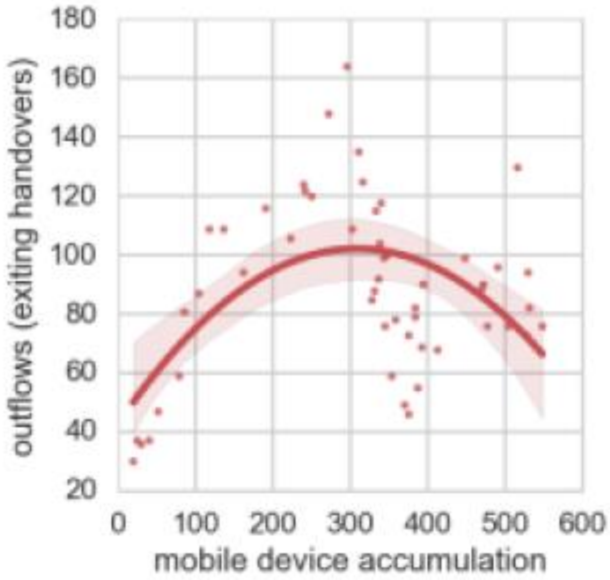
'Poster: LuST-LTE: A simulation package for pervasive vehicular connectivity',T Derrmann, S Faye, R Frank, T Engel, IEEE Vehicular Networking Conference (VNC) 2016

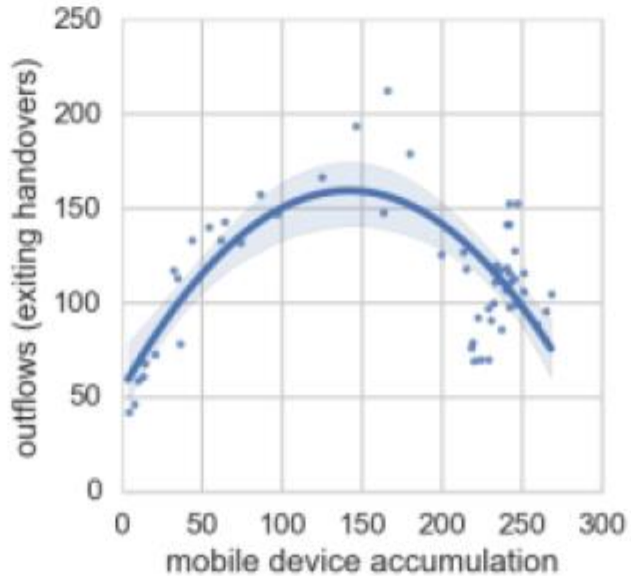
Simulation Study (2)

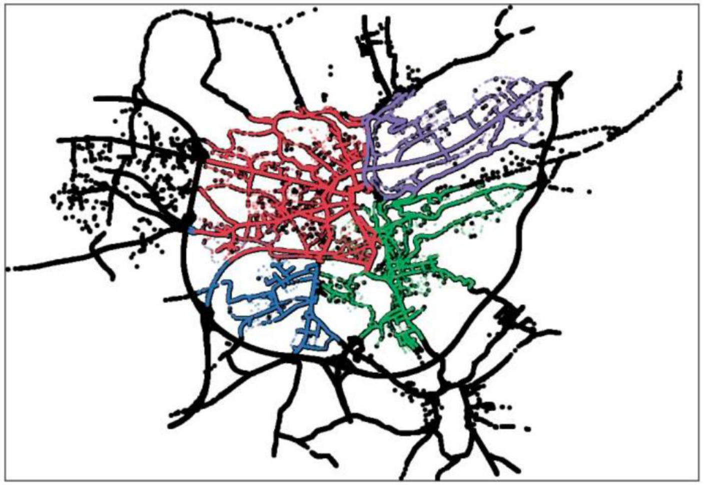
- **Half-day** simulation (midnight to noon)
- **Full LuST demand**, but lower re-routing rate
  → more congestion, so we can also see the ‘descending’ MFD phase
- Low mobile connection rate:
  3% of cars with **active** connection (data or call)
- Limiting Factor: **No pedestrians** and/or stationary users!!
- Clusters/MFD regions: drawn manually

Simulation Results (1)

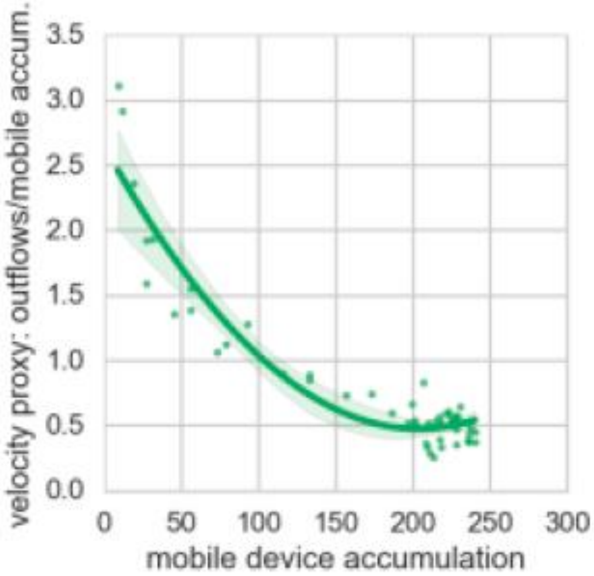
SnT
securityandtrust.lu

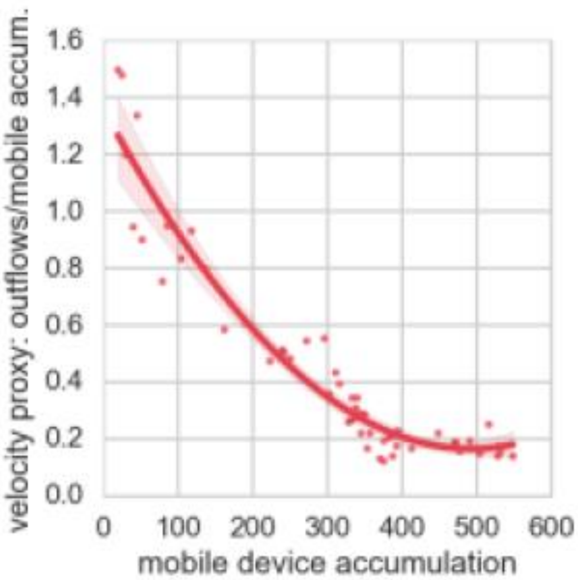


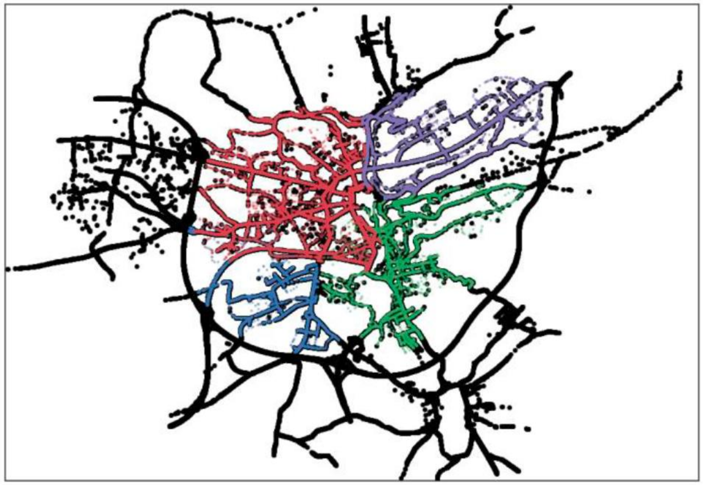




Simulation Results (2)







Simulation Results (3)

Pearson correlation coefficient: MFD velocity estimation vs. true simulated velocity

Partition	Pearson: $v^{true} \sim v^{proxy}$
City Center (red)	0.95
Hollerich (blue)	0.81
Bonnevoie-Cents (green)	0.9
Kirchberg plateau (violet)	0.85

→ Works well but simulation environment is ideal case, e.g. no stationary users, no noise in data

→ **Need a real data study!**

Real Data Study (1)

SnT
securityandtrust.lu

Goal:

Verify if 'MFDs' from real mobile network data are expressive and correlate with the traffic states

Solution:

Compare real mobile network data to floating-car data (as ground truth)

Publication:

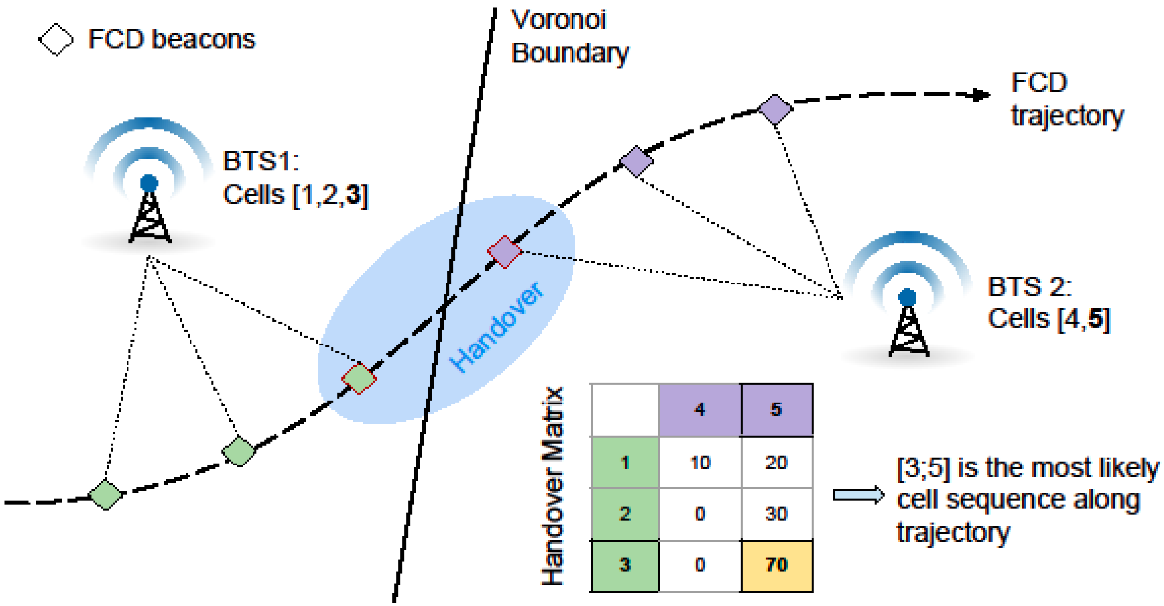
''Estimating Urban Road Traffic States Using Mobile Network Signaling Data'', T Derrmann, R Frank, F Viti, T Engel, submitted to IEEE ITSC 2017

Real Data Study (2)

Data sets (1 week of data, September 2016)

- POST Luxembourg: Aggregated Handovers and Call Counts
- Floating-Car Data of Luxembourg City

FCD ⇔ Mobile network matching: most likely handover sequence

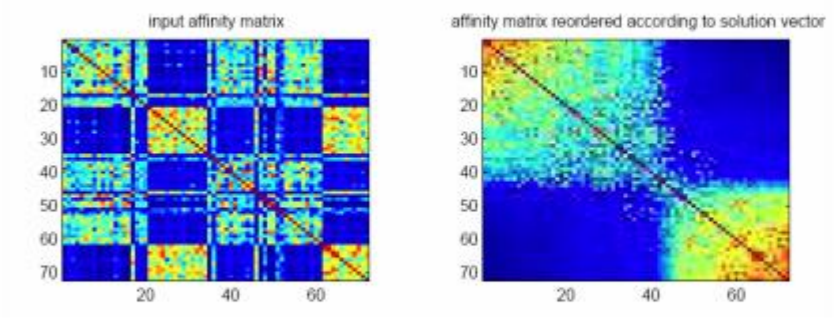


Real Data Study (3)

Clustering of mobile network:

Spectral clustering of the matrix of handovers

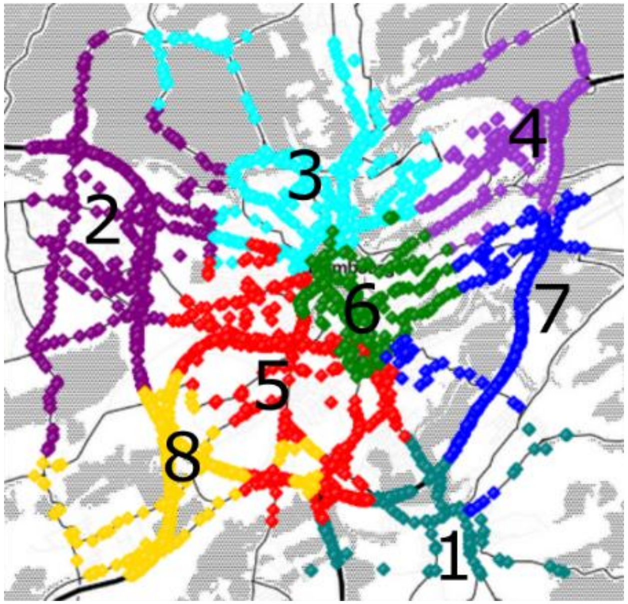
→ Creates cohesive, densely connected clusters of antennas

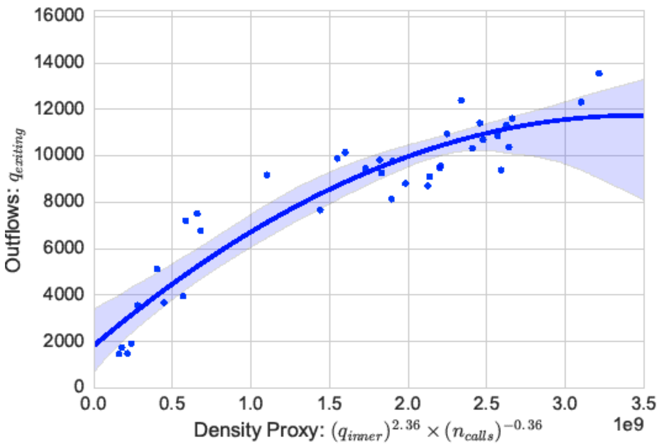


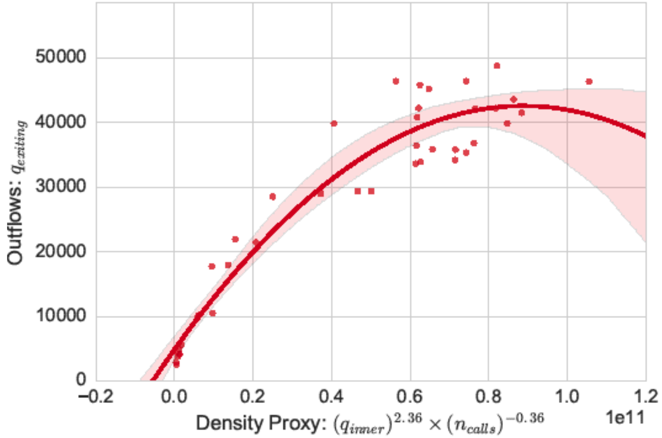
In real data, **we also have stationary users** (unlike sim. study)

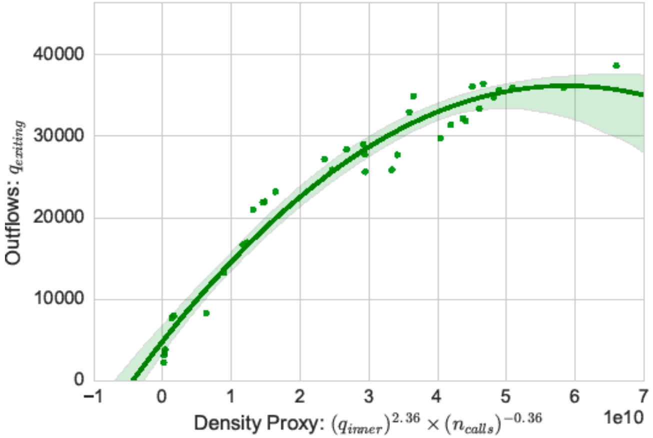
→ Approximation of vehicular density such that the outflow-to-density ratio estimates the road velocities best

Real Data Results (1)









Real Data Results (2)

Cluster	Area	$\mathbf{r}(\tilde{v})$	$\mathbf{r}(p')$
1	Hesperange, Ring: A3 LU $\leftrightarrow$ FR	0.33	0.49
2	Ring: A6 LU $\leftrightarrow$ BE	0.60	0.45
3	Limpertsberg, uptown	0.10	-0.04
4	Kirchberg	-0.08	-0.22
5	Cessange, Gasperich	0.51	0.58
6	Train station, downtown	0.52	0.29
7	Ring: A1 LU $\leftrightarrow$ GER	0.65	0.54
8	Ring: A4 LU $\leftrightarrow$ Esch/Alz.	0.59	0.7

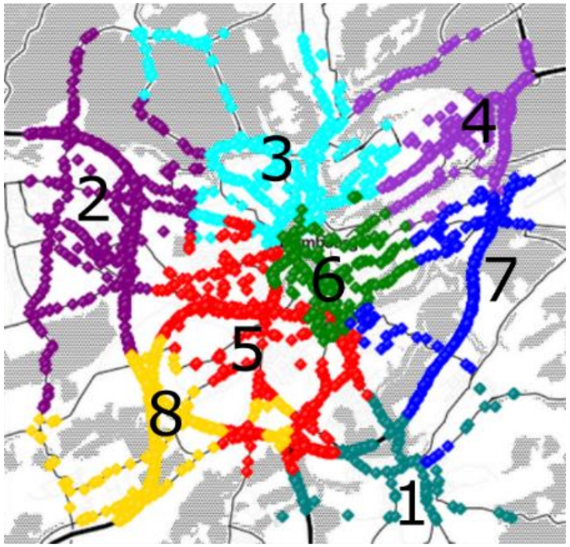


Table I: Pearson correlation coefficients by cluster: Traffic state vs. velocity proxy and derivative of the profile function

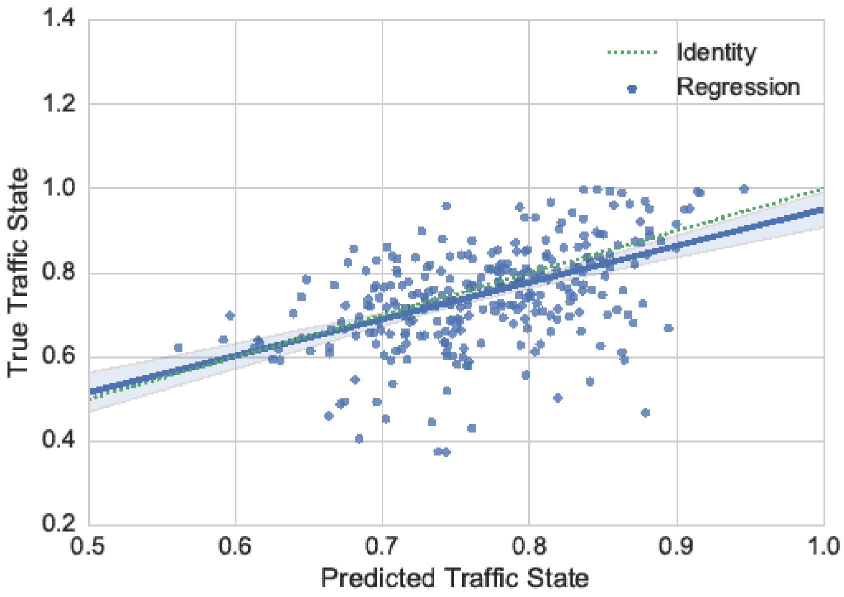
Mobile network MFDs are good for some clusters, bad for others
→ Density proxy model might be too simple or use too little features
→ Highest correlations on highways, urban needs more data

Real Data Results (3)

Linear regression model:

Predicting traffic state using:

- MFD function of current cluster, learnt over the 3 training days
- mobile network data in cluster at prediction time interval



→ Mean Absolute Percentage Error: 11.9%

→ simple model achieves decent results

Conclusion

- Results show that mobile network data can - in some environments - produce MFD-like profiles
- These profiles approximate the underlying road network behavior
- We **can** use mobile network data to predict traffic states
- **But:** it remains to be seen whether we can we also estimate the underlying road MFDs directly
  → More data needed

Future work

- **Better models of density estimation** & traffic state regression
- Direct comparison of emerging MFDs:
  mobile network vs. road network
- Try other clustering methods
- Evaluate the impact of adding additional, aggregate mobile network metrics

Thank you! Enjoy the cocktails and dinner!

For questions: Find me at my poster ☺