Spatial uncertainty propagation in ICT data analysis

Maxime Lenormand, Thomas Louail, Marc Barthélemy, J.J. Ramasco

Spatial uncertainty propagation in ICT data analysis

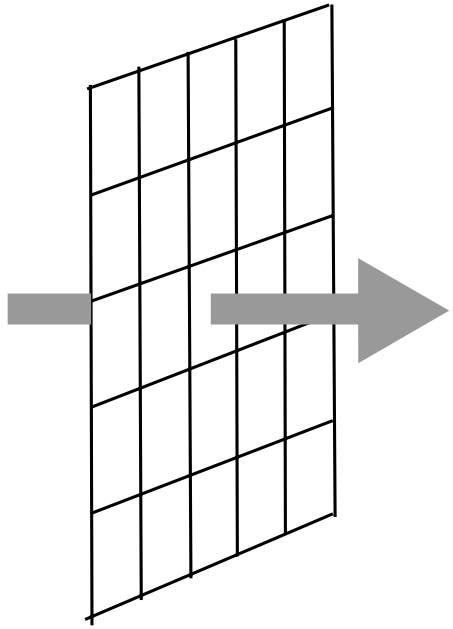
Maxime Lenormand | Irstea, France
Thomas Louail | IFISC, Spain
Marc Barthelemy | CEA, France
Ramasco JJ | IFISC, Spain

UrbanNet 2016 | Amsterdam

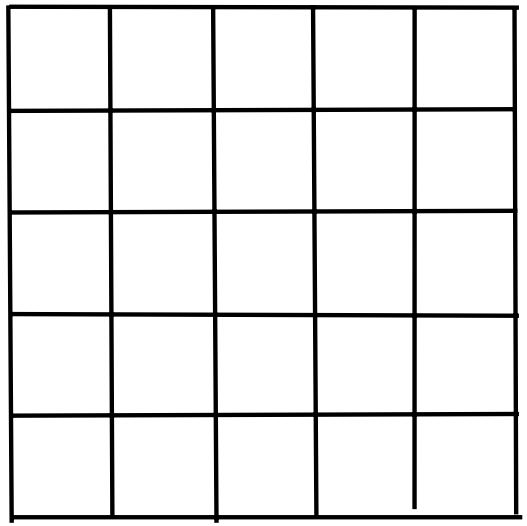
21 September 2016

Motivation

"Reality"



Data



Sampling framework

(Spatial) Uncertainty Propagation

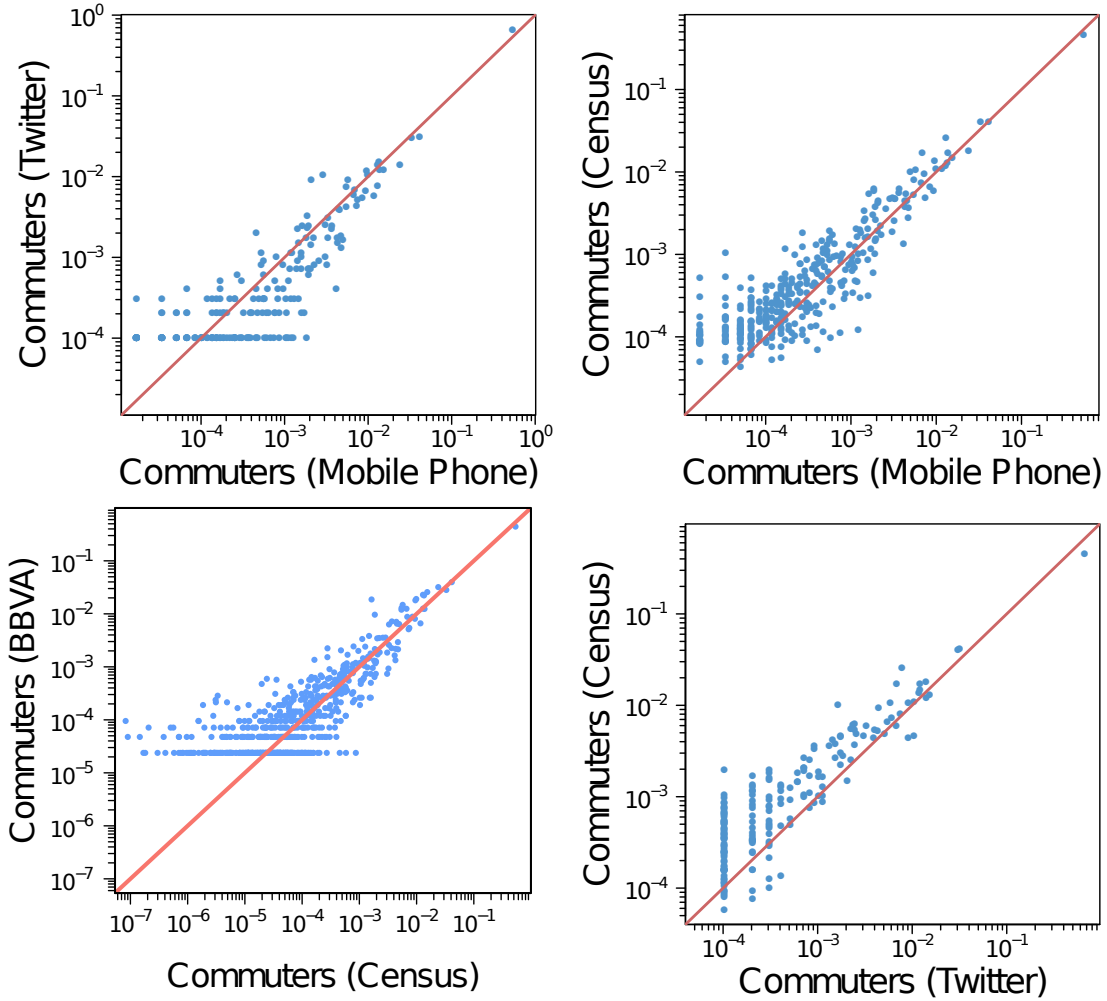
- Example: population size of Netherlands. We may estimate it as 16,800,000 milion. Perhaps in reality it is 16,967,234; hence error of -1%

- Error is usually not know because reality is not know

(Spatial) Uncertainty Propagation

- Example: population size of Netherlands. We may estimate it as 16,800,000 milion. Perhaps in reality it is 16,967,234; hence error of -1%
- Error is usually not know because reality is not know
- Crosschecking information

Crosschecking mobility information



Lenormand *et al.* (2014) Cross-checking different sources of mobility information. *PlosOne*, 9(8):e105407.

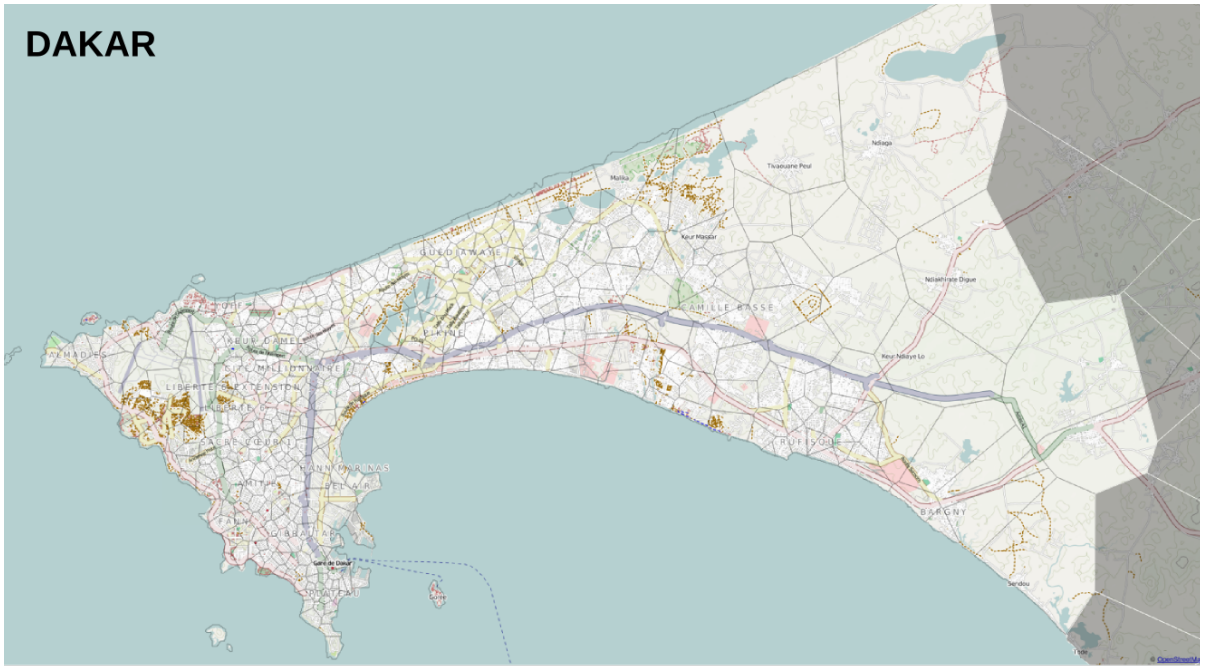
(Spatial) Uncertainty Propagation

- Example: population size of Netherlands. We may estimate it as 16,800,000 milion. Perhaps in reality it is 16,967,234; hence error of -1%
- Error is usually not know because reality is not know
- Crosschecking information
- **Uncertainty propagation analysis**

  ↳ ***Lenormand et al.*** (2016) Is spatial information in ICT data reliable? *arXiv preprint arxiv:1609.03375*.

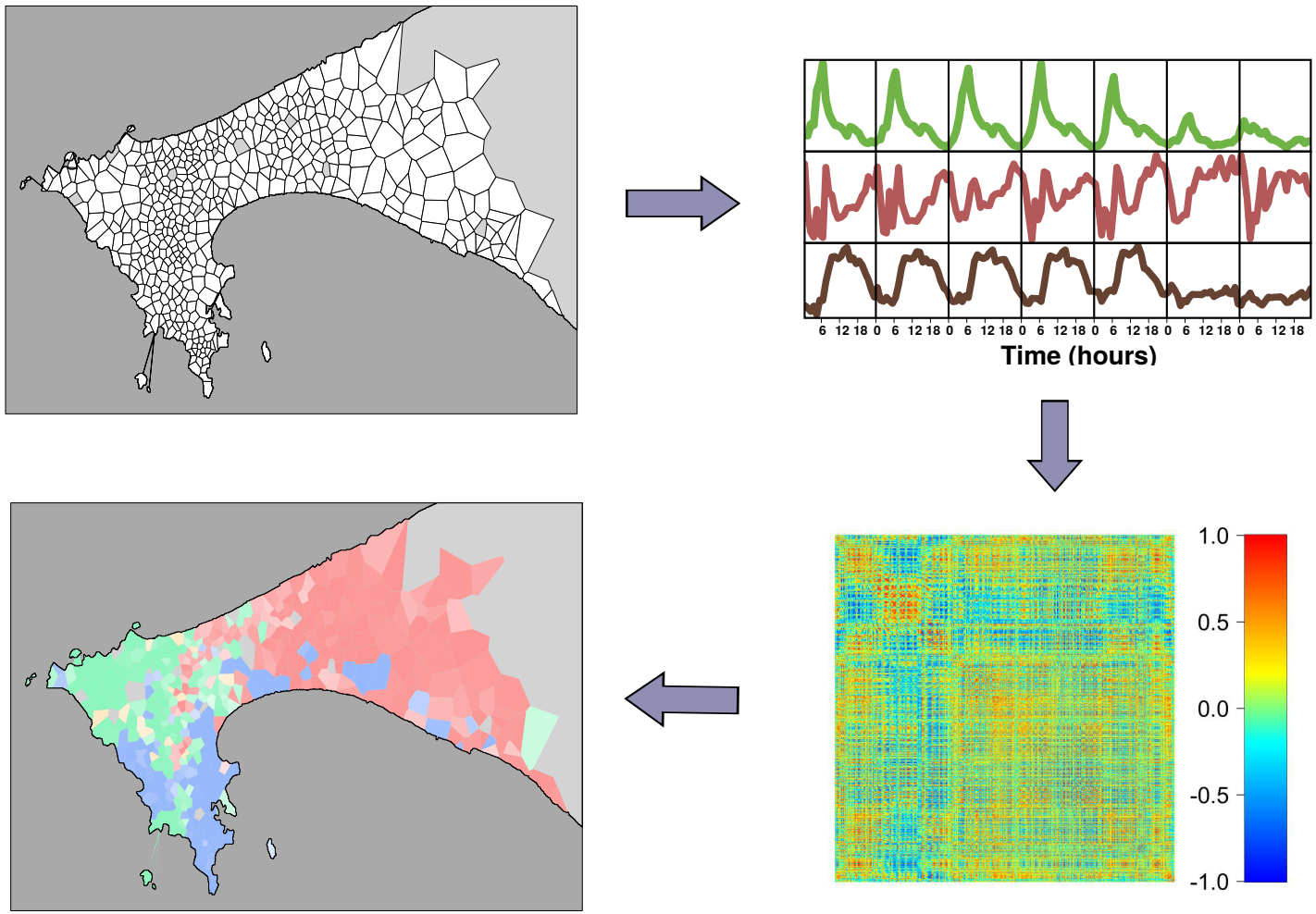
ICT Data uncertainty analysis

300,000 mobile phone users' trajectories **x 25** two-week periods



Inferring **land use** and identifying **home-work locations** from mobile phone activity

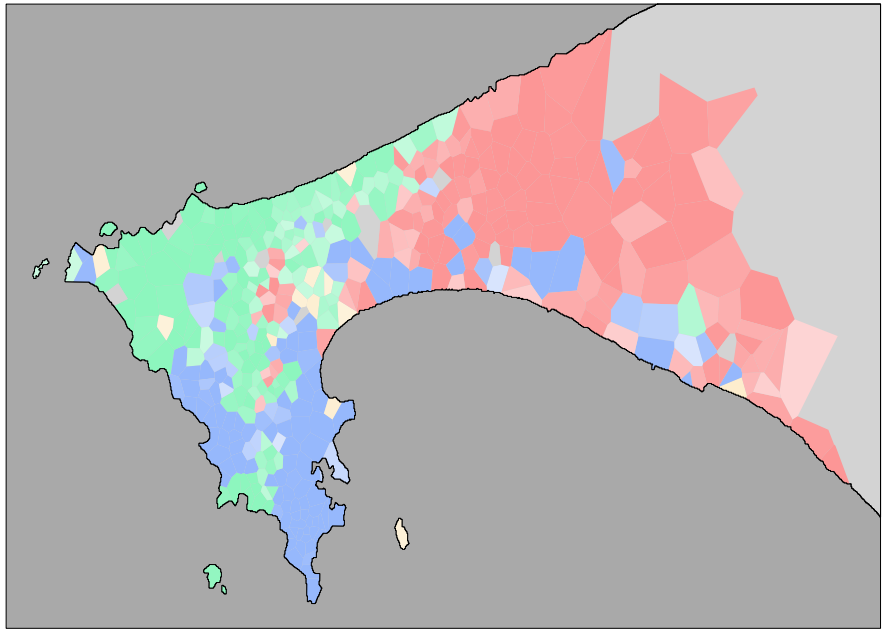
Land use detection

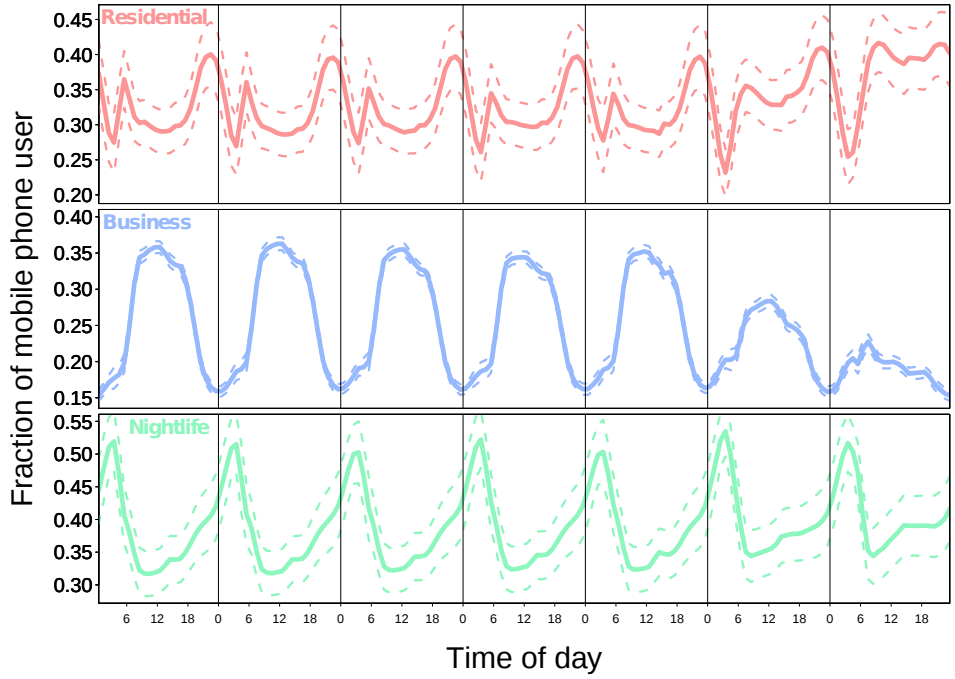


Lenormand *et al.* (2015) Comparing and modeling land use organization in cities. Royal Society Open Science 2, 15052015.

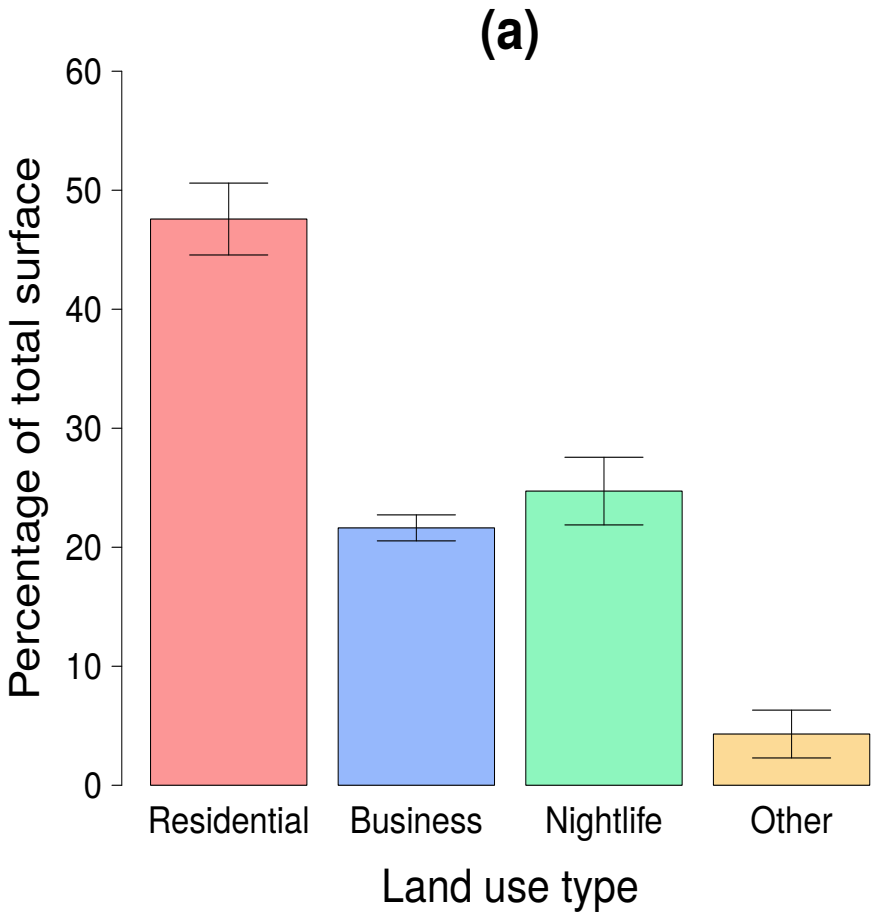
Land use detection

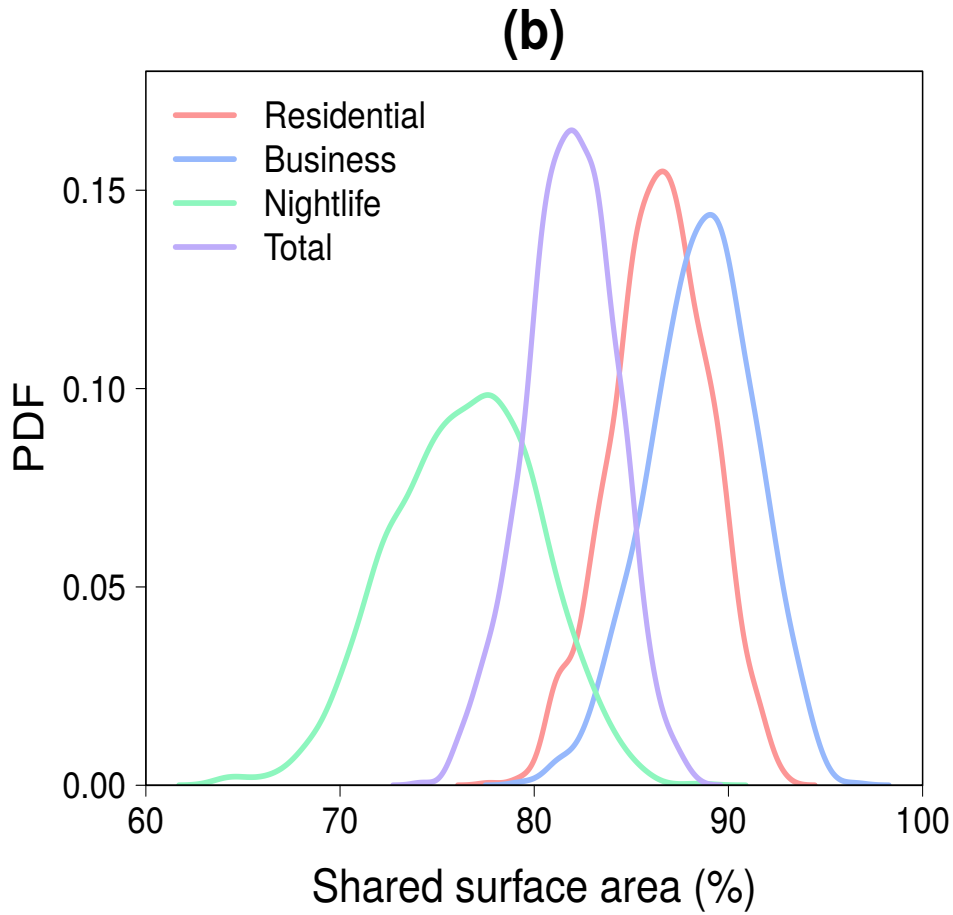
Extraction of 50 independent samples based on 150,000 users activity during on week





Land use detection

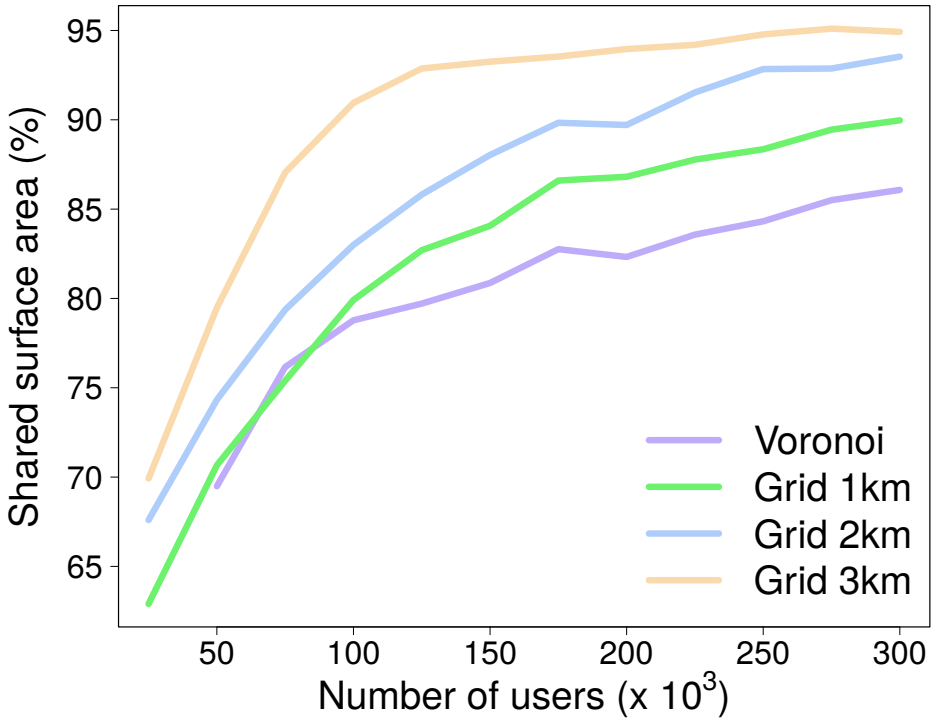




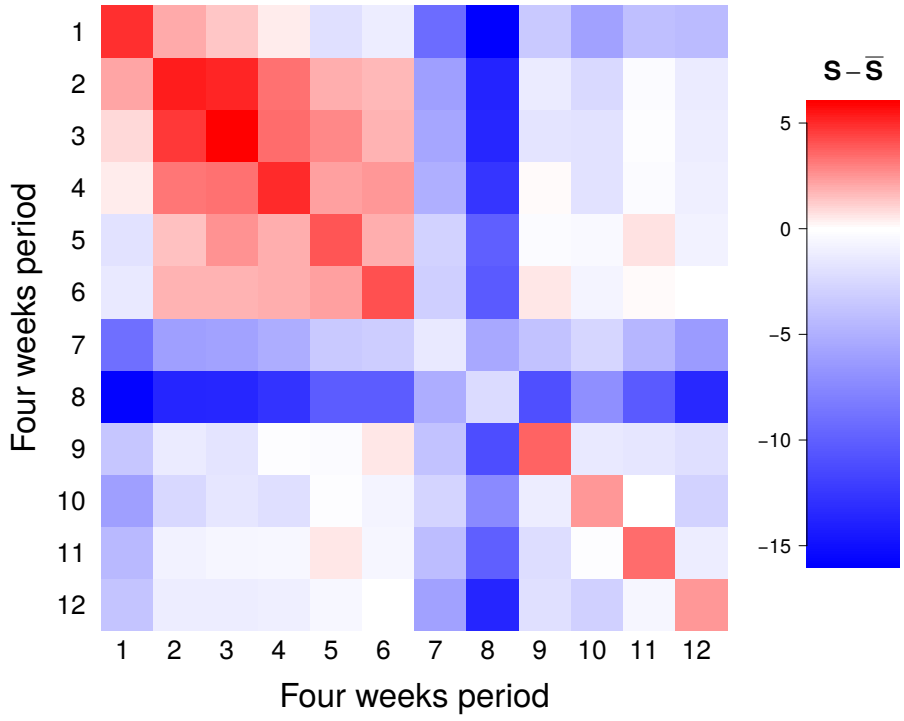
$$SSA(S_l, S'_l) = 2\frac{A_{S_l \cap S'_l}}{A_{S_l} + A_{S'_l}}$$

Land use detection

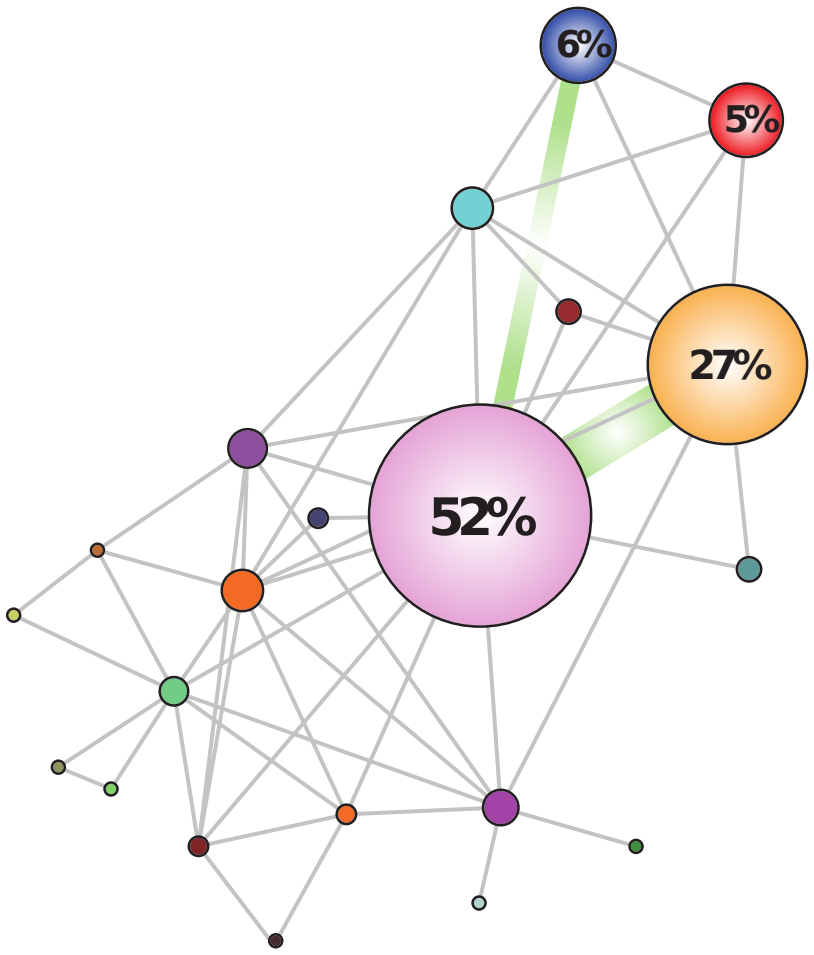
Size & Scale



Time



Most frequented locations



Home

Most frequented location between 7pm and 7am

Work

Most frequented location between 8am and 5pm on weekdays

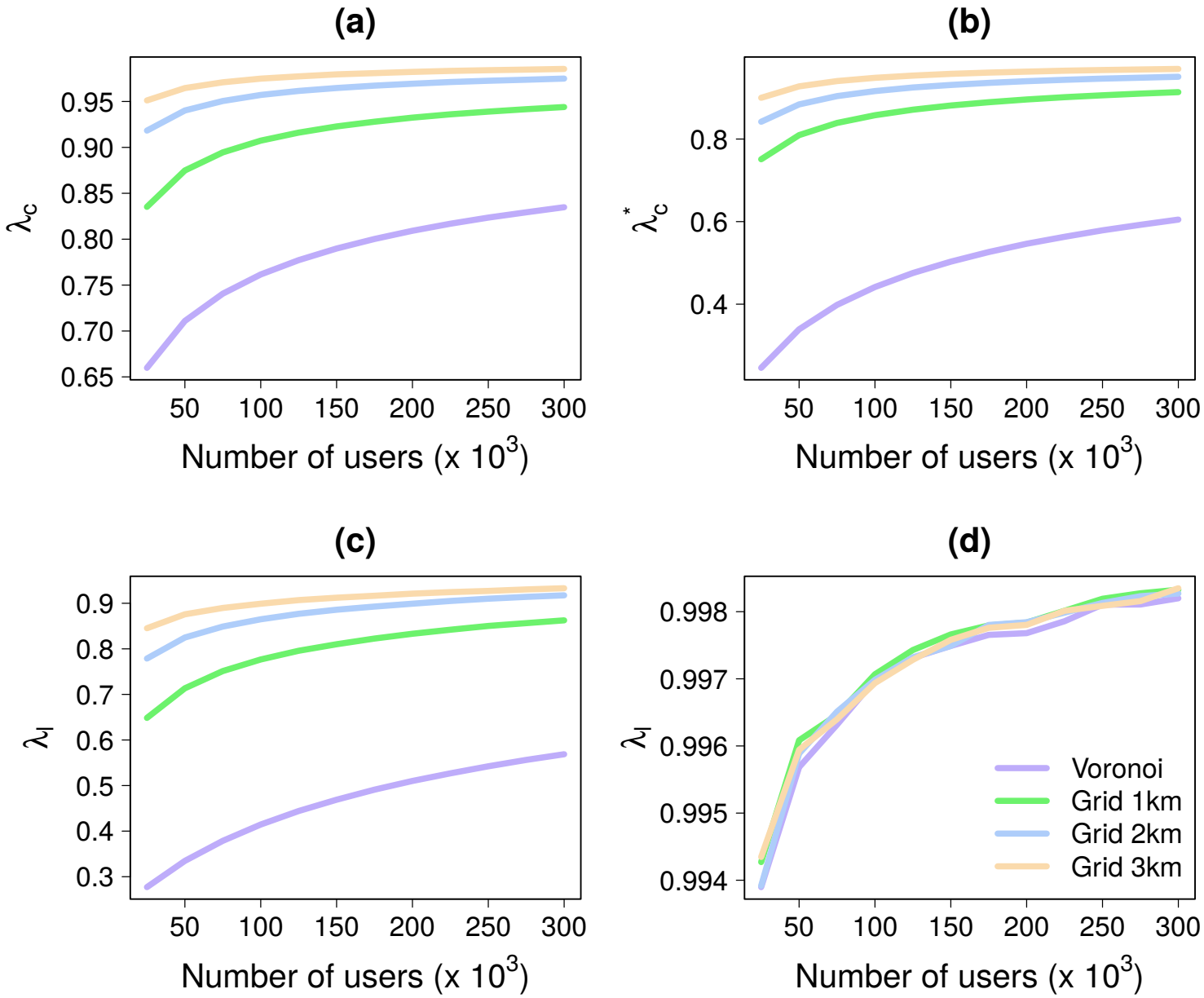
↓

Origin-Destination Matrix

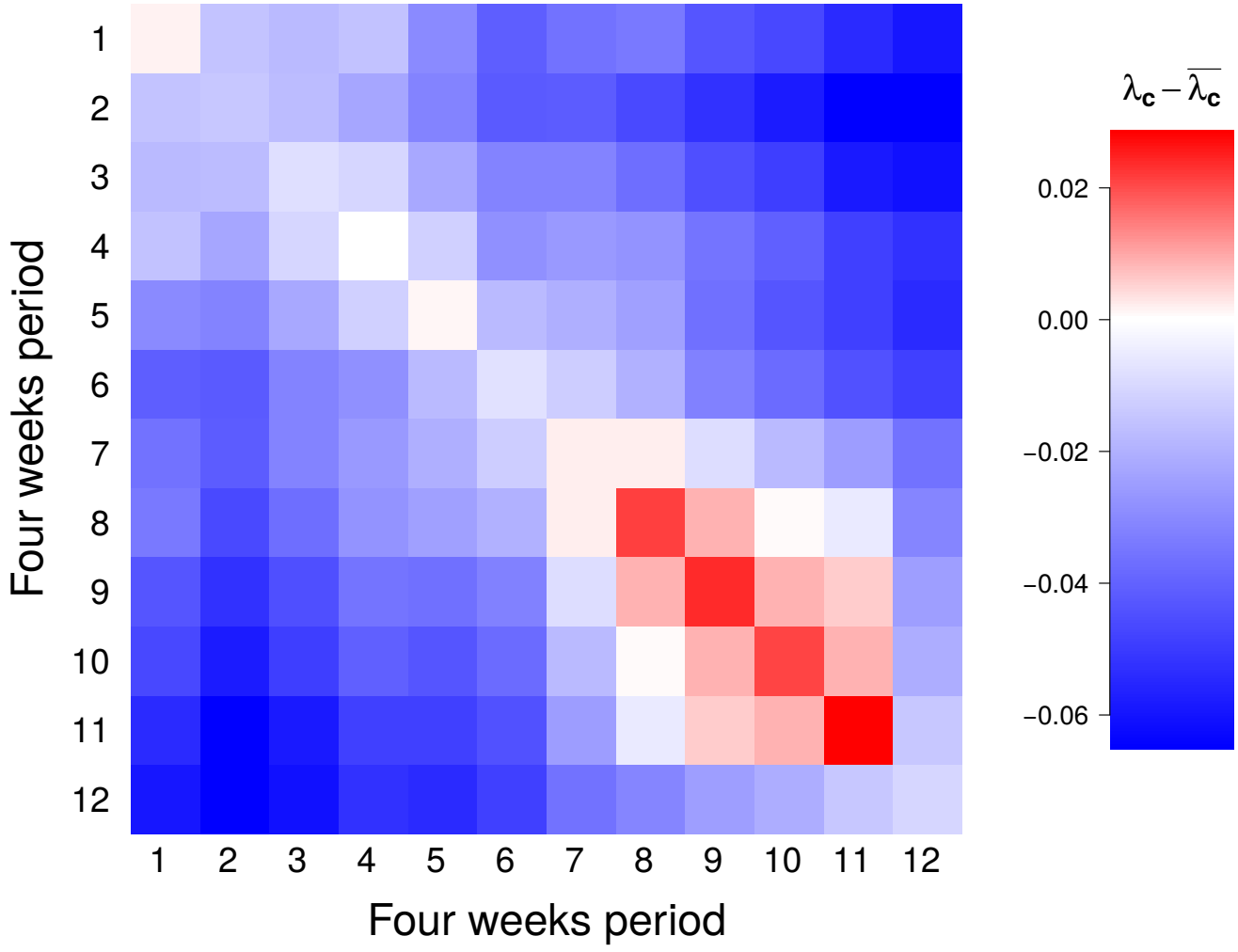
Tij: *number of individuals living in cell* ***i*** *and working in cell* ***j***

Song et al. (2010) Limits of predictability in human mobility. *Science* 327, 1018-1021.

Most frequented locations

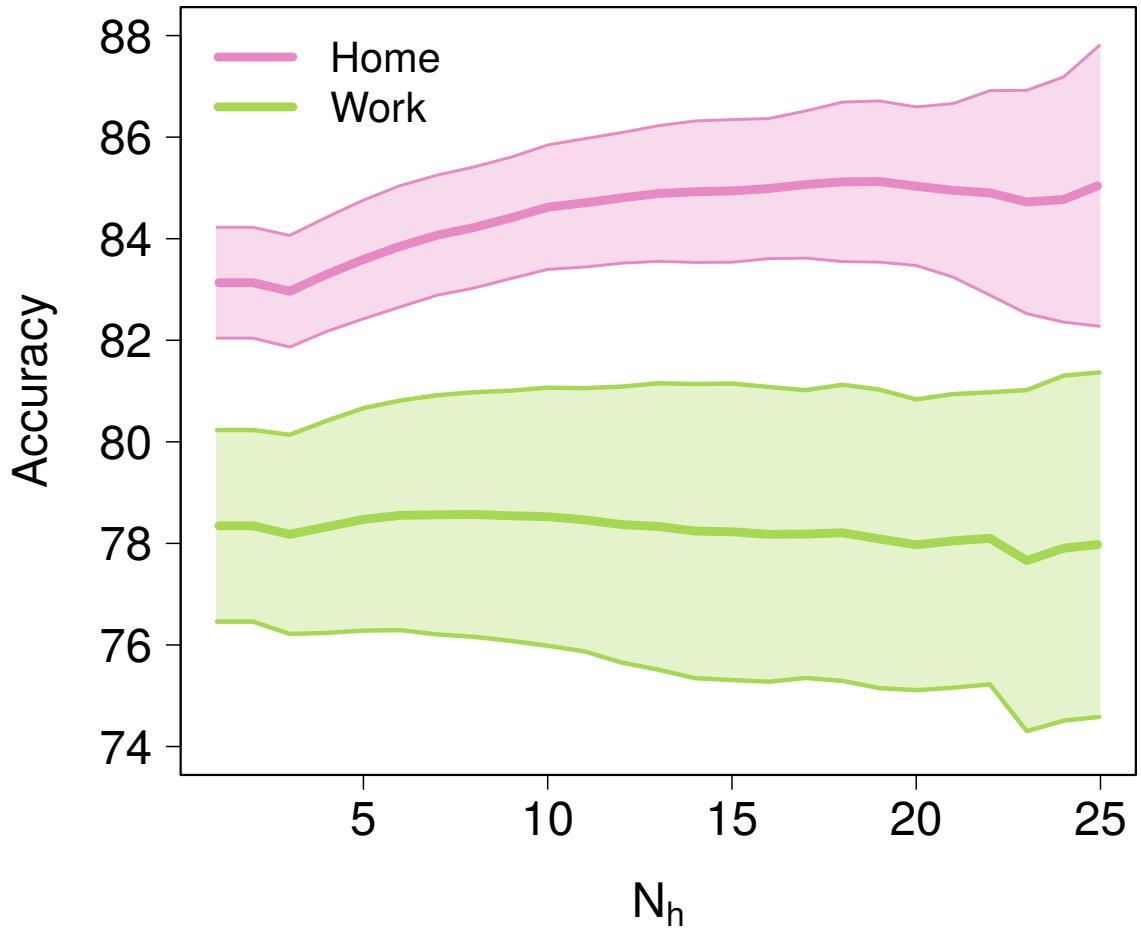


Most frequented locations



Most frequented locations

Robustness of home-work locations identified for each user



Take home messages...

- Good agreement between land uses identified from 100,000 users activity signals, with an average of 75% of shared surface area.
- Uncertainty on the journey-to-work commuting network is highly dependent of the spatial resolution.
- More studies in this spirit need to be done to assess the biases and uncertainty associated with this new data sources.

Lenormand et al. (2016) Is spatial information in ICT data reliable? *arXiv preprint arxiv:1609.03375*.

Self-promotion

ictdataccuracy.github.io