ThesisDay@SystemX 2018

Mobile Phone Data Analysis for the Estimation of Daily Urban Dynamics

Danya BACHIR

Mounim El Yacoubi[1] (directeur), Vincent Gauthier[1](co-directeur), Mostepha Khouadjia[2] (encadrant)

[1] Samovar – Télécom SudParis, [2]IRT SystemX

1. CONTEXT

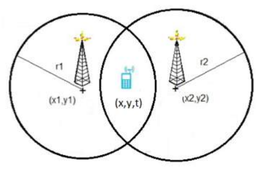
- **Growing urban population** and flows.
- **Ubiquitous MP (Mobile Phone) data available** (big data).
- Provide urban **mobility knowledge**.
- Cities need **smarter transport**.

How can we use MP data to estimate daily urban mobility ?

2. OBJECTIVES & CHALLENGES

Objectives:
- Population time variant **densities & count**. ✓
- Population **flows by transport modes**.

Challenges:
- Recent topic, few research.
- MP Data is sparse, noisy.
- Partial population knowledge.
- Unsupervised models.
- Few Validation Data.

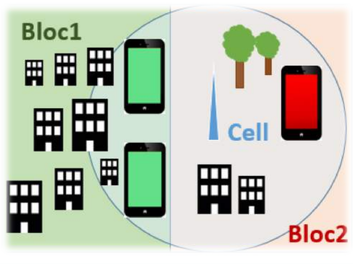


ex: phone covered by 2 overlapping cells → handover, ping pong…

3. RESEARCH METHOD

1. **MP Presence count**
2. **Spatial Mapping** →

$$\sigma_j = \sum w_i . \sigma_i$$ j :bloc i:cell



3. **Land Use Filtering** for noise reduction:

→ Clustering on blocs MP activity.

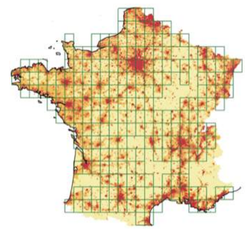
4. **Static Calibration** with census data (constant in time):
   $\rho = \alpha.\sigma^{\beta}$; ρ: pop density ; σ: MP presence.

5. **Dynamic rescaling** with user day activity.
6. **Validation** on Stadium events.

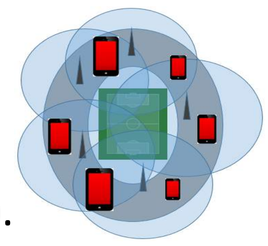
4. CASE STUDY & APPLICATIONS

Case Study:
- Ile-de-France 2 months data

(in partnership with Bouygues Telecom).

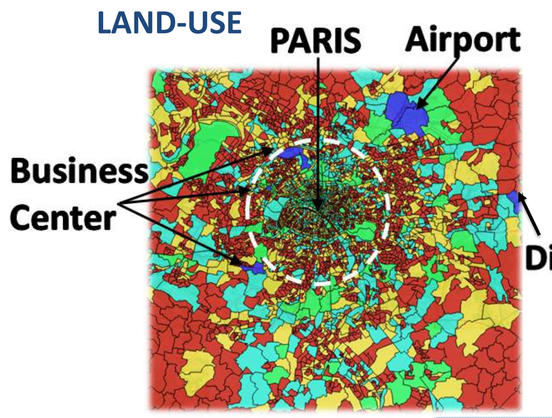


Use cases: Daily Population
- Densities at **census bloc scale** (IRIS).
- Counts for **indoor spaces** (stadiums, shopping centers, museums, train stations, tunnels, closed transport hubs).



5. RESULTS

LAND-USE



SPATIAL MAPPING PERFORMANCE

	Area		Hybrid	
Granularity	$Corr_{dens}$	$Corr_{vol}$	$Corr_{dens}$	$Corr_{vol}$
Bloc	0.58	0.21	0.65	0.40
1st neighbors	0.73	0.44	0.78	0.49
City	0.91	0.90	0.93	0.91

Cluster Size(%)	$Corr_{AVG}$	$Corr_{MEDIAN}$
ALL 100	0.60	0.59
0 62	0.79	0.81
1 0.1	0.93	0.94
2 11	0.75	0.76
3 25	0.82	0.86
4 2	0.71	0.71
0 + 3 87	0.74	0.77

EVALUATION METRICS
- Pearson Corr.
- R^2
- NRMSE

MODEL PARAMETERS

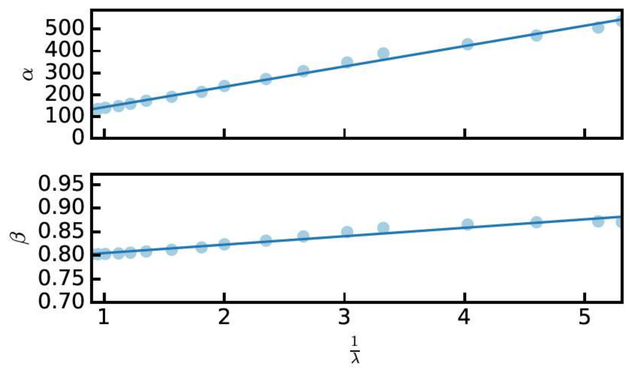
$$\alpha(t) = \frac{1}{f}\frac{1}{p^{event}(t)}$$

$$\hat{\rho}(t) = \alpha(\lambda,t).\sigma(t)^{\beta(\lambda,t)}$$

$$\alpha(\lambda,t) = \left(\hat{a_\alpha}\frac{1}{\lambda(t)} + \hat{b_\alpha}\right)$$

$$\beta(\lambda,t) = \left(\hat{a_\beta}\frac{1}{\lambda(t)} + \hat{b_\beta}\right)$$

Dynamic coeff. estimated with static coeff. corrected by mean user activity



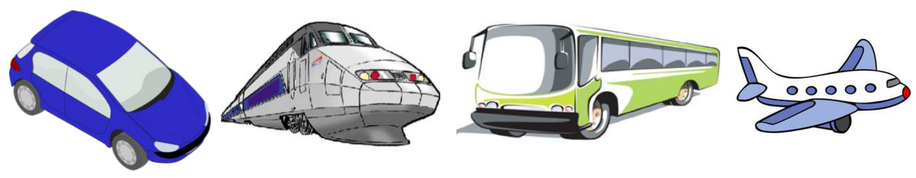
VALIDATION

	MAX	MED
ALL	0.143	0.148
NATIONAL	**0.126**	0.143
INTERNATIONAL	0.157	0.142

NRMSE for estimated population in Paris area Stadiums during 10 sport events.

6. NEXT WORK: TRANSPORT MODE

- Semi-Supervised Multimodal Cell Scoring.
  - Baseline model with spatial features.
  - Spatio-temporal feature based model.
- Unsupervised Trajectory mode Inference.



REFERENCES

- Deville et al, Dynamic population mapping using mobile phone data, 2014
- Khodabandelou et al, Population estimation from mobile network traffic metadata, WoWMoM 2016

Scientific domain: Data science and Interaction
Program: Smart Territories
Project: Modeling Mobility Solutions (MSM)

Doctoral school: Sciences et Technologies de l'Information et de la Communication (STIC)
Institution: Université Paris-Saclay

Contacts:
danya.bachir@irt-systemx.fr
mostepha.khouadjia@irt-systemx.fr