Sorbonne Université
LIP6 Seminar
Paris, France
September 24, 2018

National Research Council of Italy

Two ways you did not know mobile networks could be useful

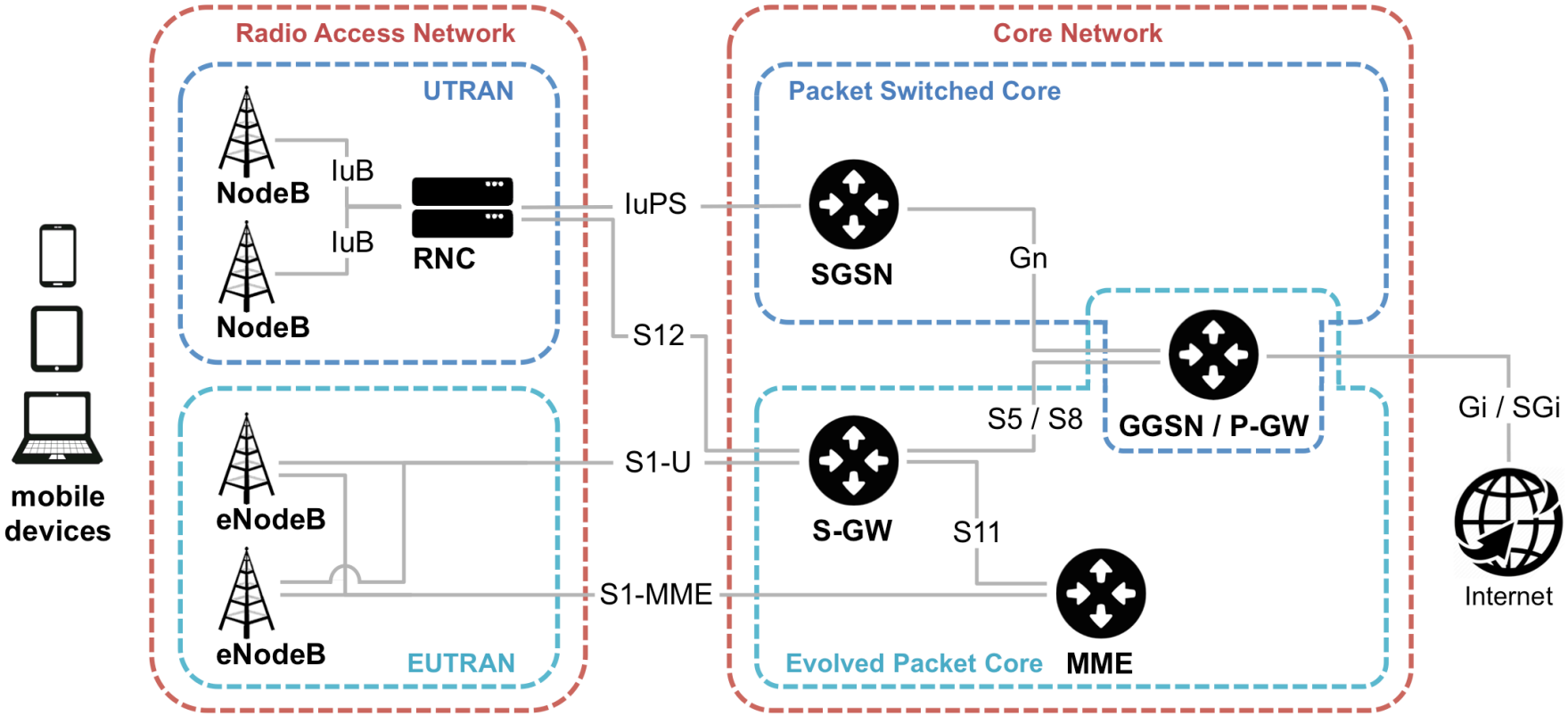
Marco Fiore

http://perso.citi.insa-lyon.fr/mfiore/
marco.fiore@ieiit.cnr.it
@marc0_fi0re

Institute of Electronics Computer and Telecommunication Engineering

What's a mobile network?

"A telecommunication system where the last link is wireless"



What's a mobile network for?

- **Evolution through generations**

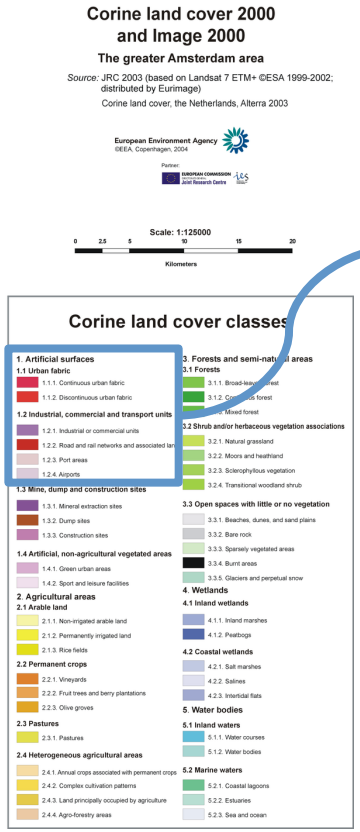
1G	1980	***TACS***		Analog voice
2G	1991	***GSM***	9.6 kbps	Digital voice, text messages, minimum data
		GPRS	20-50 kbps	
		EDGE	100-130 kbps	
3G	2001	***UMTS***	384-400 kbps	Web browsing, social media, navigation, audio streaming
		HSPA	2-3 Mbps	
		HSPA+	5-8 Mbps	
4G	2010	***LTE***	10-20 Mbps	Video streaming, machine-type communication
		LTE-A	20-40 Mbps	
5G	2020			Internet of things, automated vehicles, enhanced mobile broadband, augmented reality

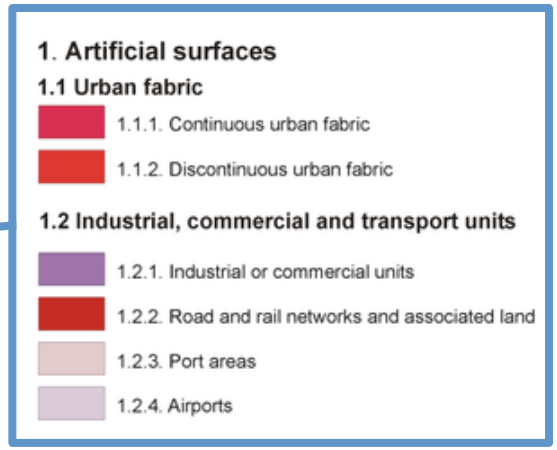
- **Can we go *beyond communication-based services*?**
  - A pervasive individual-level ***remote sensing platform***

I.1
Land use
Context and mapmaking

Land use

"The total of arrangements, activities, and inputs that people undertake in a certain land cover type"

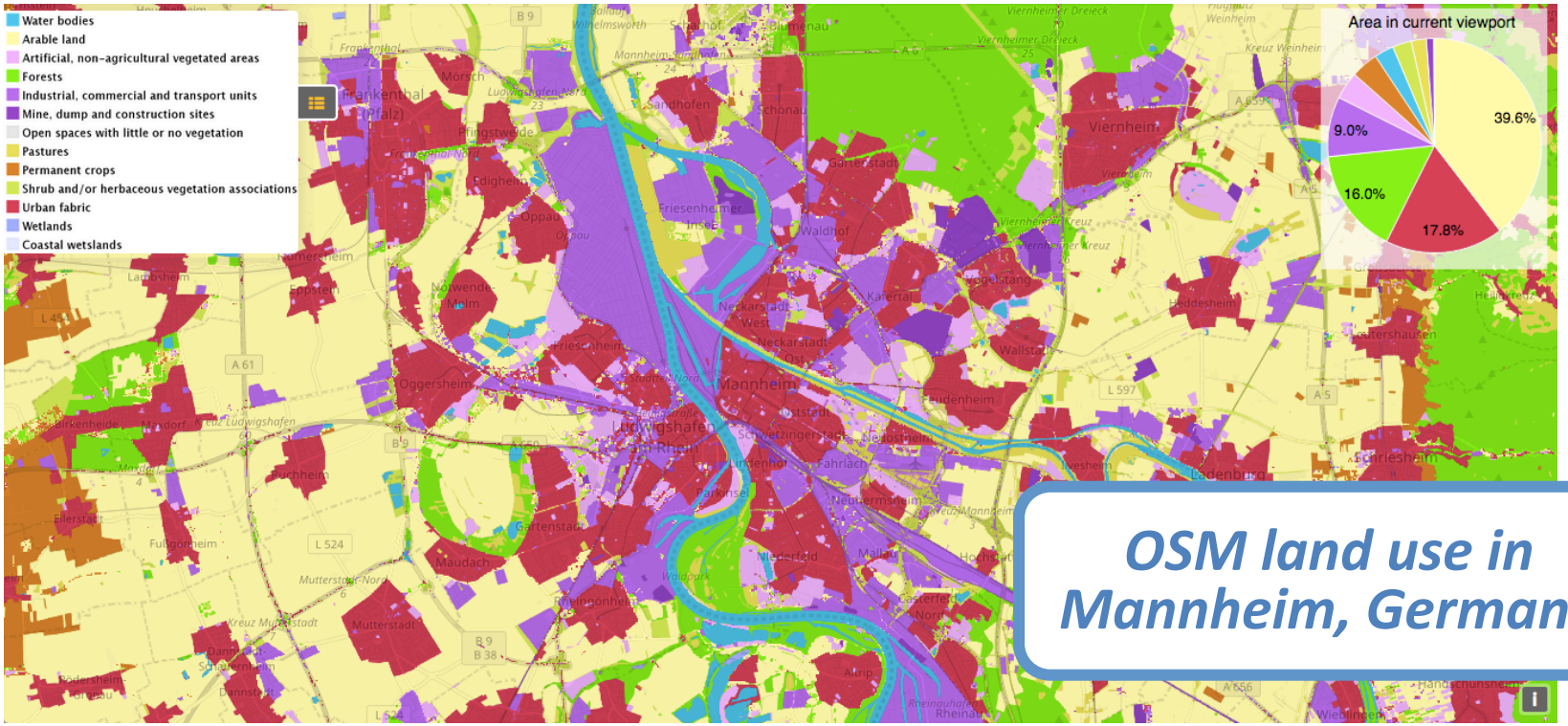




Corine land use database in Amsterdam, The Netherlands

Urban land use

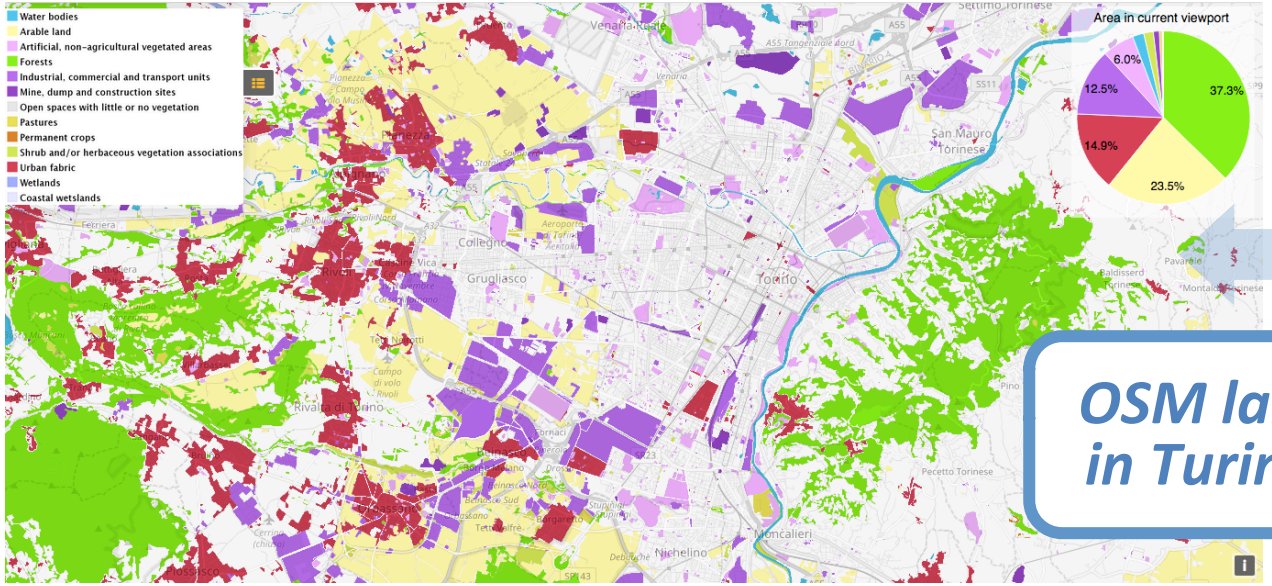
- **City land use has extensive applications**
  - Urban planning, ***zoning***, metropolitan ***transport system*** planning, demographics, social ***segregation***, etc.



OSM land use in Mannheim, Germany

Land use mapmaking

- **Traditional approaches**
  - ***Census*** data, surveys, ***satellite imagery*** processing
  - An active research field in ***geoinformatics***
- **Current techniques have significant drawbacks**
  - Time-consuming, *easily* ***outdated***, expensive, incomplete



OSM land use in Turin, Italy

I.2

Using mobile network traffic to detect land use

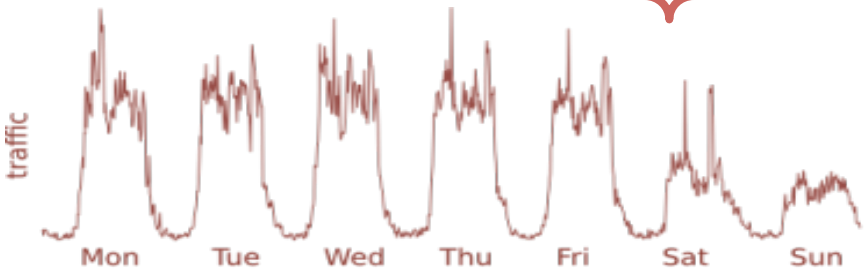
A simple hierarchical classification approach

One-slide methodology

- **Intuition** – different land uses entail ***diverse traffic dynamics***

1. Mobile traffic **signature** at one location

- metric — ***aggregate*** *mobile traffic volume*
- temporal support — *one* ***week*** [1,2]
- (filtering) — *none*
- normalization — ***standard score*** ***median*** *hourly volume*



2. Pairwise **signature distance** measure

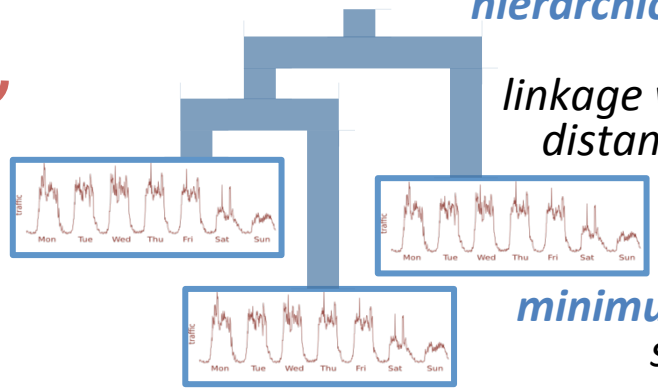
$$p_{ab} = \frac{\sum_{t\in\mathcal{T}} (\mathbf{s_a}(t) - \mu_{\hat{a}}) \cdot (\mathbf{s_b}(t) - \mu_{\hat{b}})}{\sqrt{\sum_{t\in\mathcal{T}} (\mathbf{s_a}(t) - \mu_{\hat{a}})^2} \cdot \sqrt{\sum_{t\in\mathcal{T}} (\mathbf{s_b}(t) - \mu_{\hat{b}})^2}}$$

correlation-based *distance*

3. Signature **clustering algorithm**

agglomerative ***hierarchical*** *clustering*

linkage with average distance – ***UPGMA***

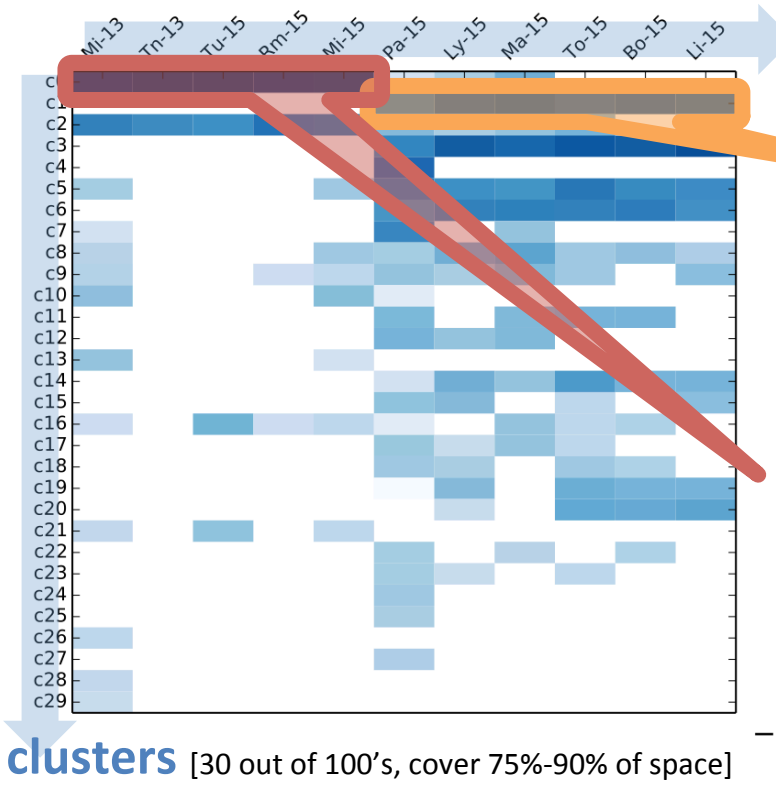


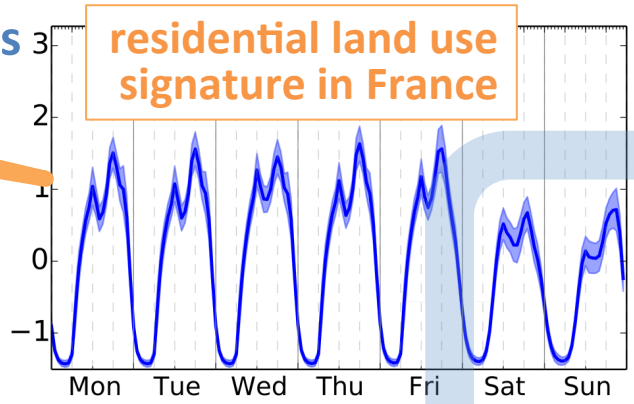
minimum skewness *stopping rule*

[1] R. Keralapura et al., ACM MobiCom 2010; [2] M.Z. Shafiq et al., ACM SIGMETRICS 2011

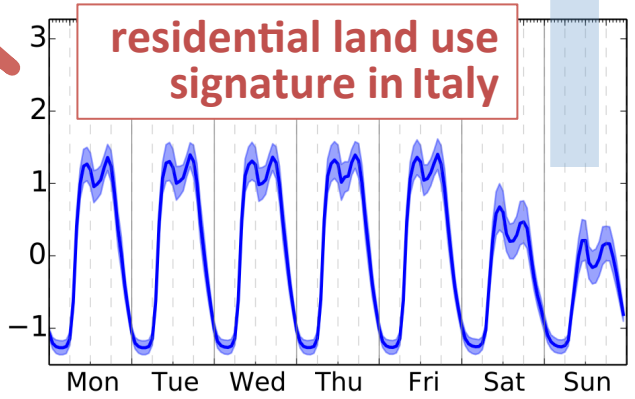
Case study

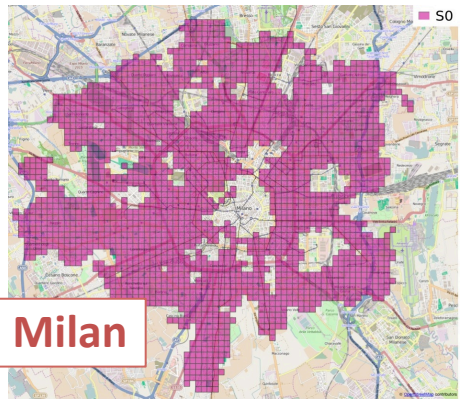
- **Real-world mobile network traffic datasets**
  - ***Orange*** 2014-15 [6 main cities in France, 4 months, antenna cells]
  - ***TIM*** BDC 2013-15 [4 main cities in Italy, 2 months, grid]



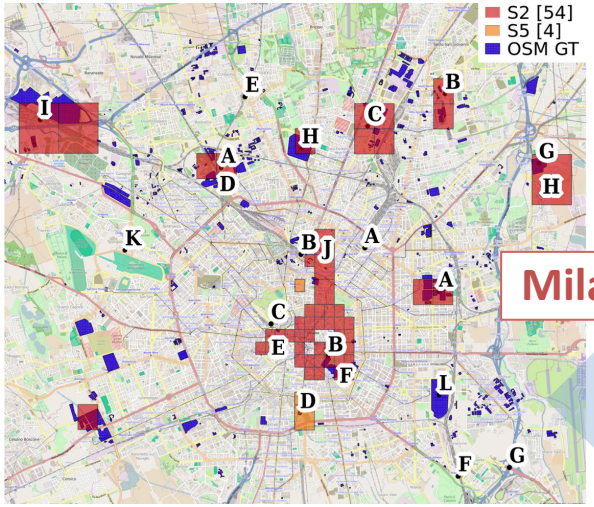


different baseline *mobile traffic demand across (culturally similar) countries*



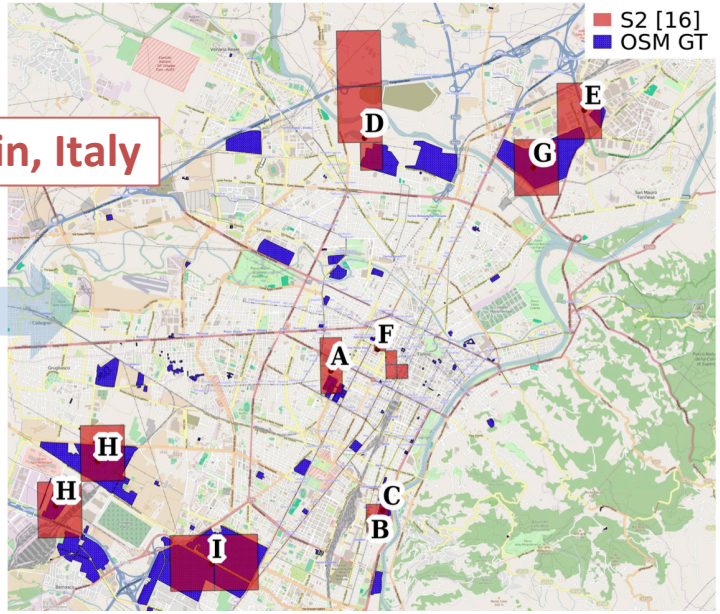


Case study

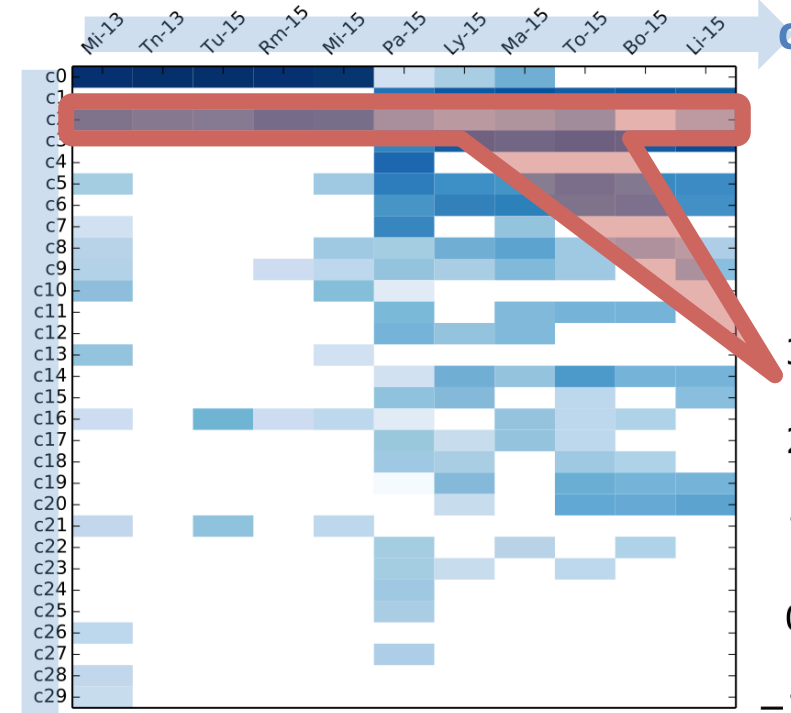


Milan, Italy

Turin, Italy

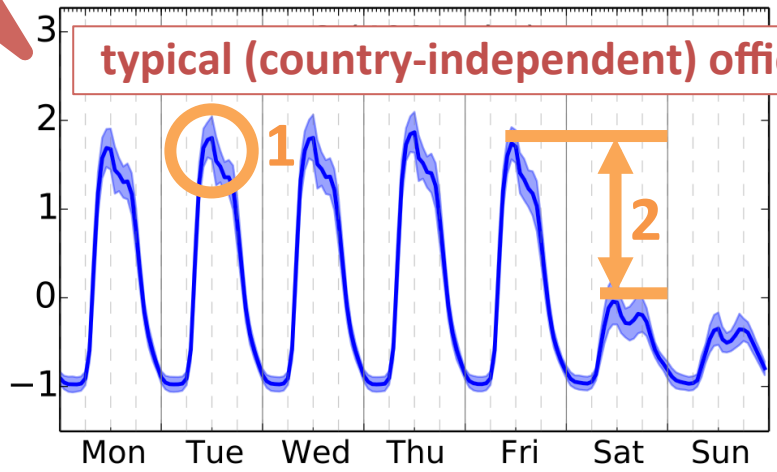


cities



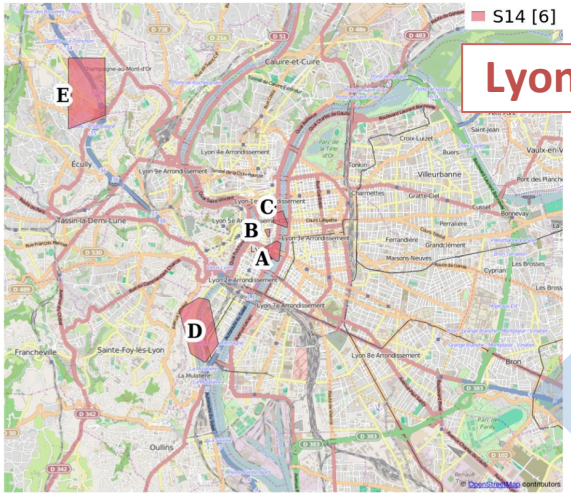
profiles [30 out of 100's, cover 75%-90% of space]

typical (country-independent) office-area signature

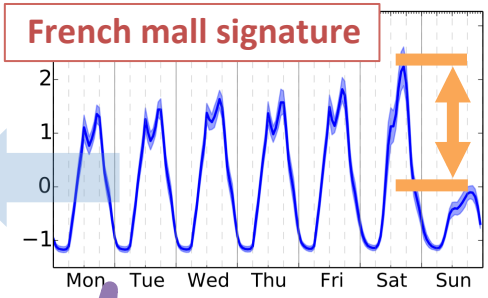


1. *morning peak*
2. *high workday to weekend ratio*

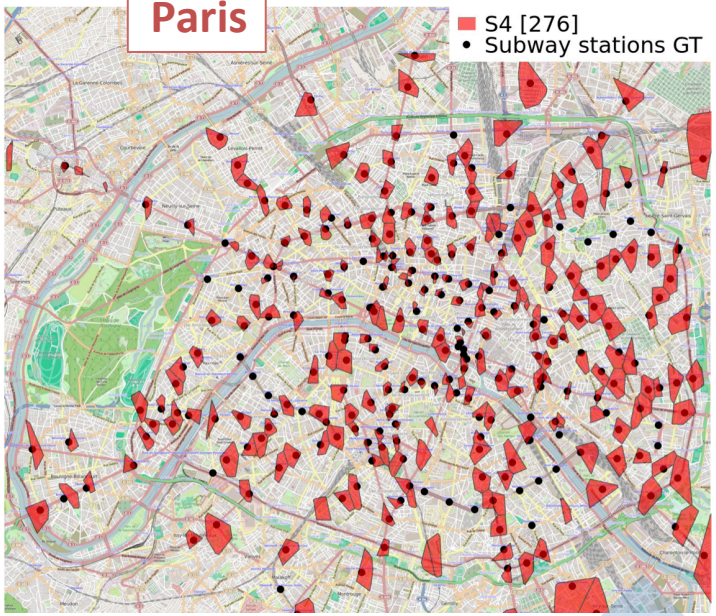
Case study



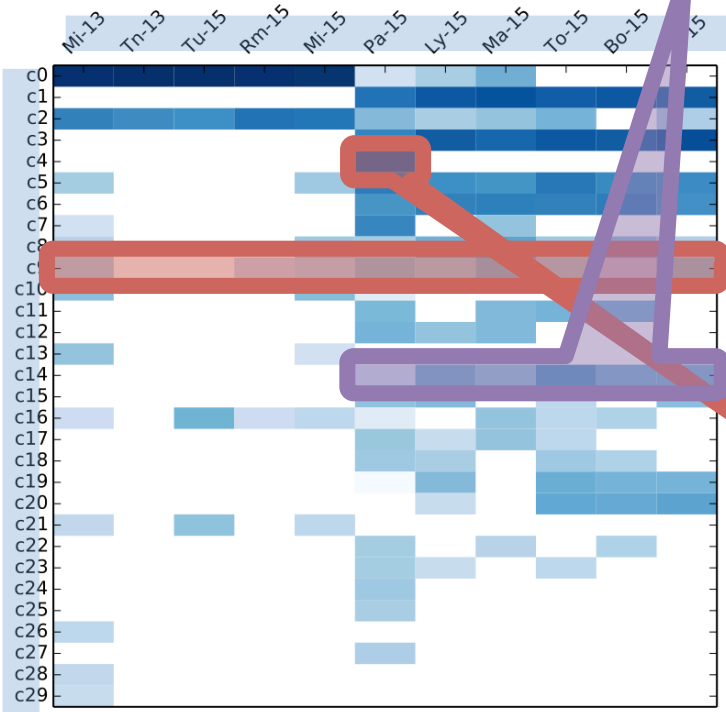
French mall signature



Paris

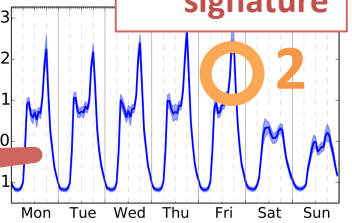


cities

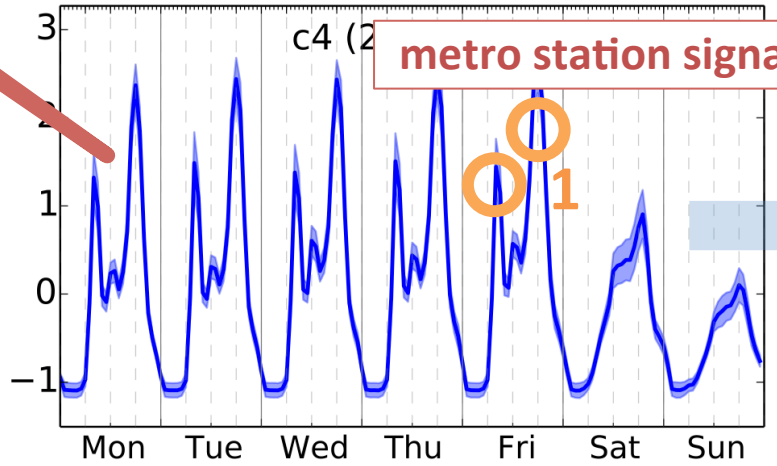


profiles [30 out of 100's, cover 75%-90% of space]

train station signature



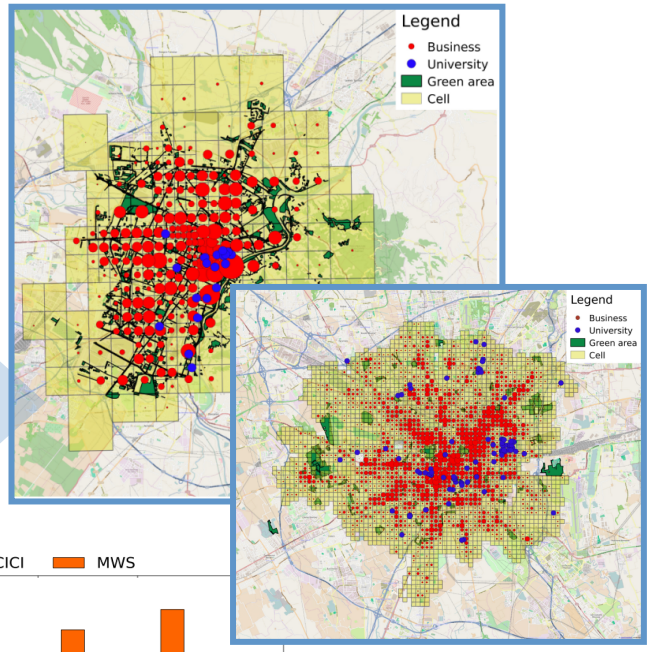
metro station signature



1. *AM-PM commuting behavior*
2. *mostly PM commuting*

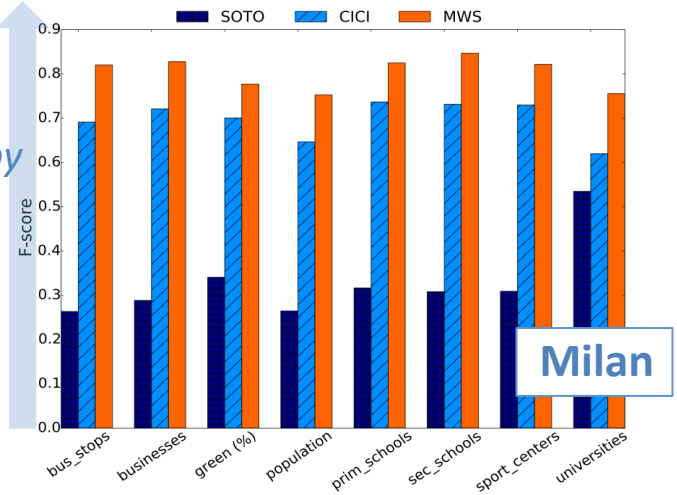
Validation

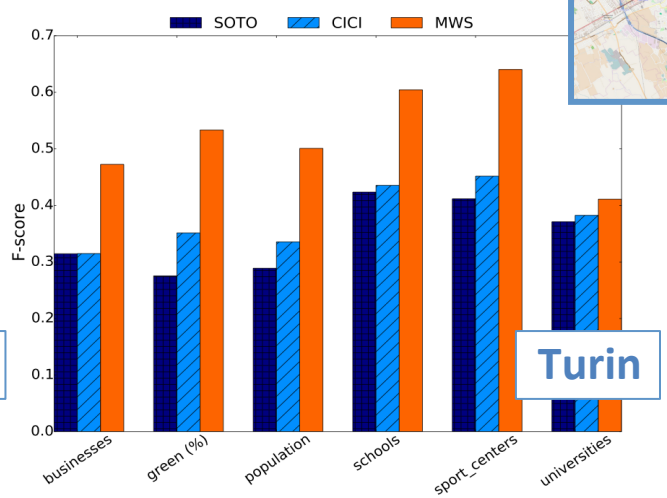
- **Ground truth land use**
  - Provided by the ***municipalities of Milan and Turin, Italy***



- **Comparative evaluation** [3,4,5]

f-score of coverage and entropy





  - A relevant ***complement*** to traditional land use mapmaking

[3] V. Soto et al., ACM HotPlanet 2011; [4] B. Cici et al., ACM MobiHoc 2015
[5] S. Grauwin et al., Geotechnologies and the Environment 2015

I.3

An alternative approach

Exploratory Factor Analysis

Methodology

- **Exploratory Factor Analysis (EFA)**

inter-related by

Kx1 vector of ***common factors***, i.e., a small number (K<<N) of complex relationships among variables

Nx1 vector of ***variables***, i.e., phenomena of interest

$$\mathbf{X} = \mathbf{\Lambda F} + \mathbf{U}$$

Nx1 vector of unknown ***unique factors*** that are specific to a single variable

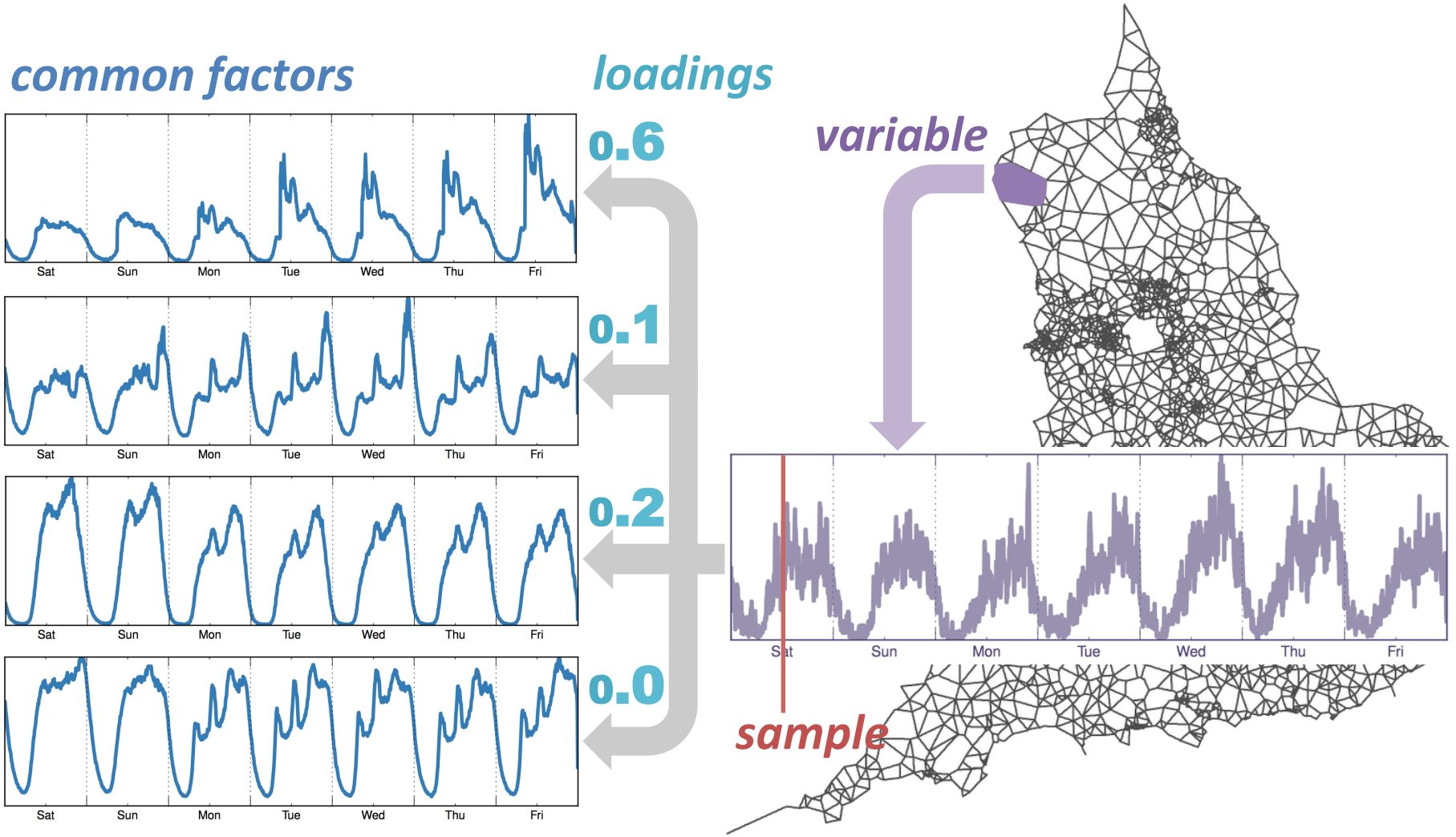
NxK matrix of unknown ***loadings*** that describe how important is each factor to every variable

interpreted through

- **EFA solution**
  - By analyzing variable observations from a set of samples, ***EFA identifies common/unique factors, and loadings*** [6]

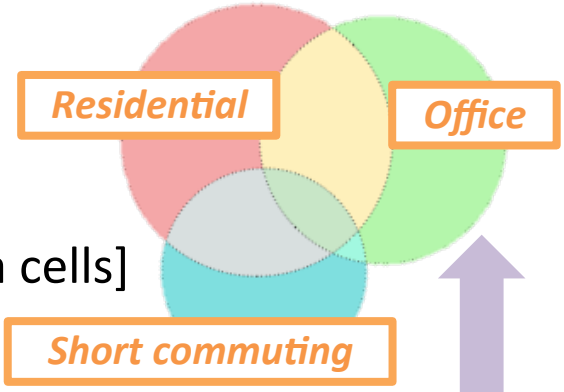
[6] S.A. Mulaik, Foundations of Factor Analysis, CRC Press, 2009

Methodology

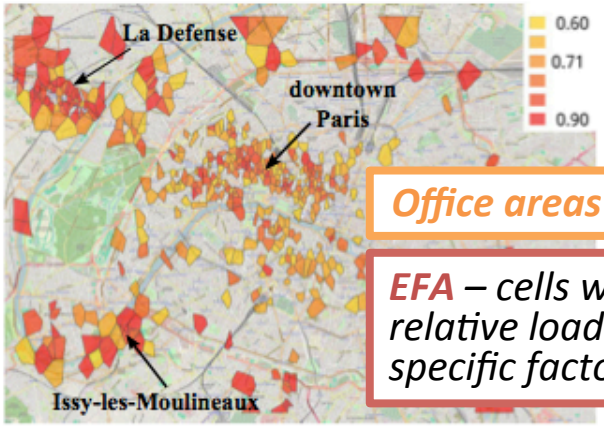


Case study

mixed land use detection

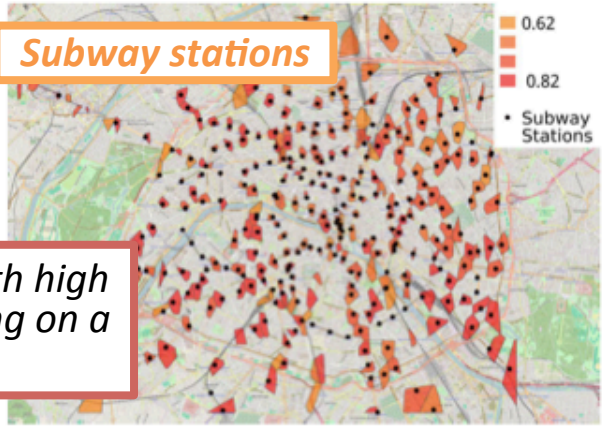


- **Orange** 2014-15 [Paris, France, 4 months, antenna cells]
  - *14 EFA factors* are identified



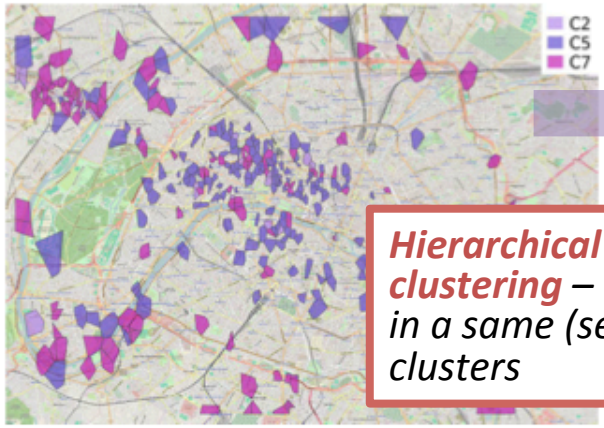
Office areas

EFA – cells with high relative loading on a specific factor



Subway stations

Railway stations



Hierarchical clustering – cells in a same (set of) clusters

- 14 factors versus *hundreds of clusters*
  - multiple signature clusters just capture *different intensities of a same land use*
  - many clusters are *unique factors*
  - traffic demands are in fact a *mixture* of actual common factors

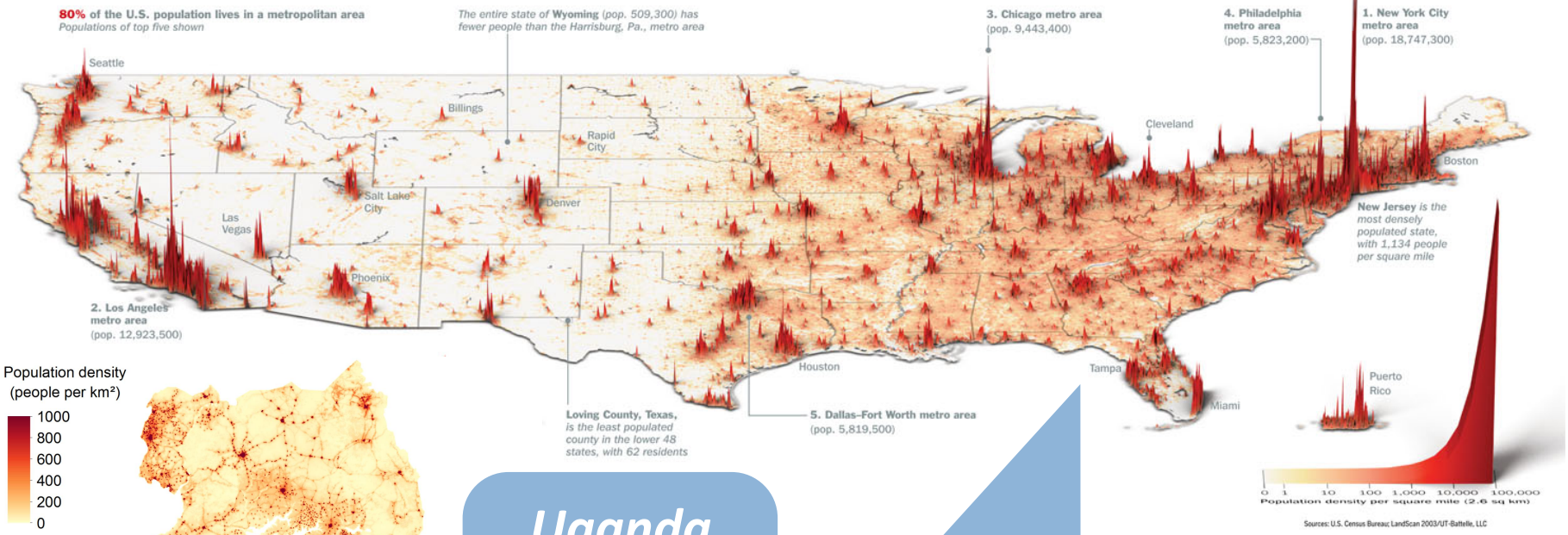
II.1

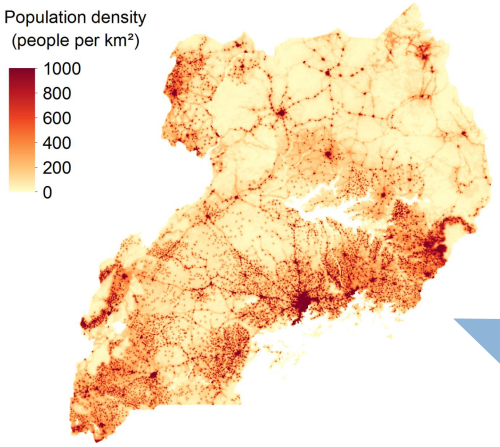
Population density

Context and dynamic estimation

Population density

> *A measurement of population per unit area or unit volume, frequently applied to humans*





Uganda population density

[WorldPop]

USA population density

[Time magazine, Oct 30, 2006]

Urban population density

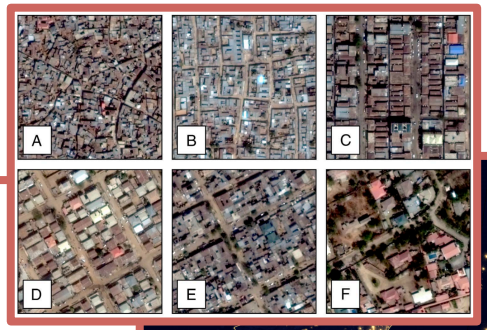
- **Urban population density has extensive applications**
  - Urban ***planning***, transportations, ***economics***, health, innovation, psychology, ***geography***, sustainability

Greater NYC tract level

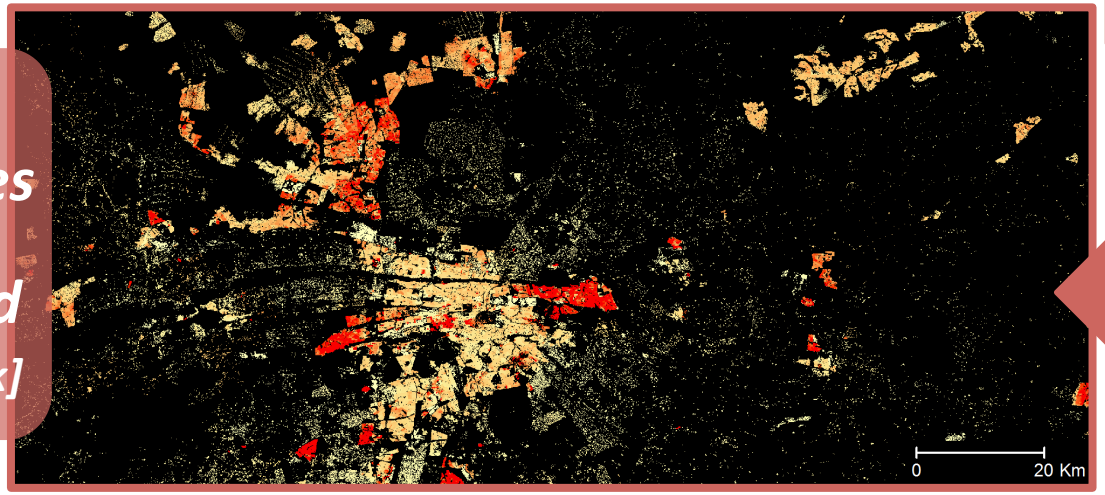
[Social Explorer]

Population density estimation

- **Traditional and advanced approaches**
  - ***National censuses***, population ***registers***, local surveys
    - often outdated, unreliable, unavailable
  - An active research field in ***geoinformatics***
    - recent breakthrough from ***neural networks*** applied to high-definition satellite imagery



23 countries fully mapped [Facebook]



  - Limited to ***static populations*** (dwelling units)

II.2

Using mobile network traffic to estimate population density

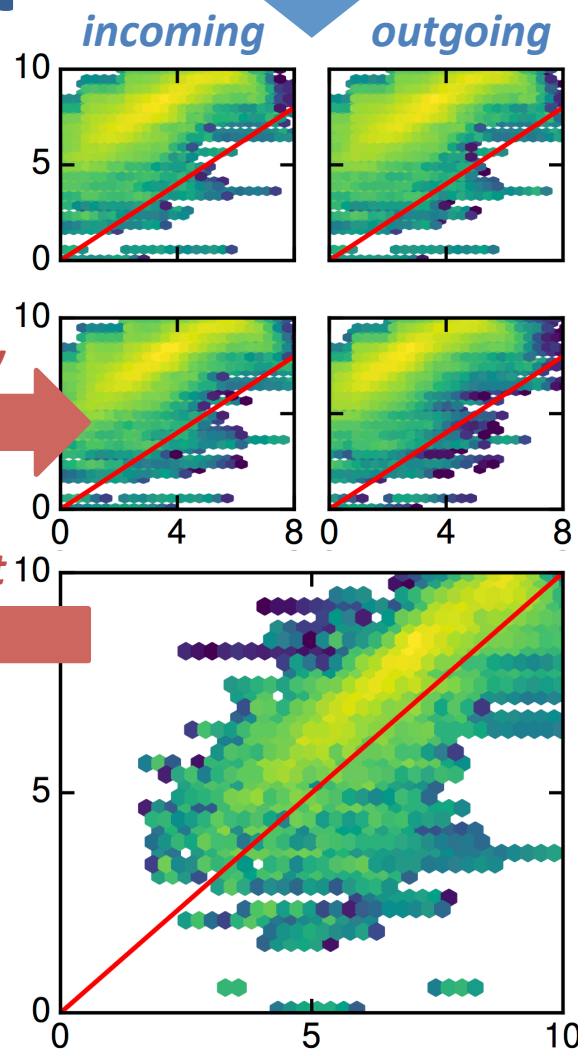
A regression model for static populations

Legacy model

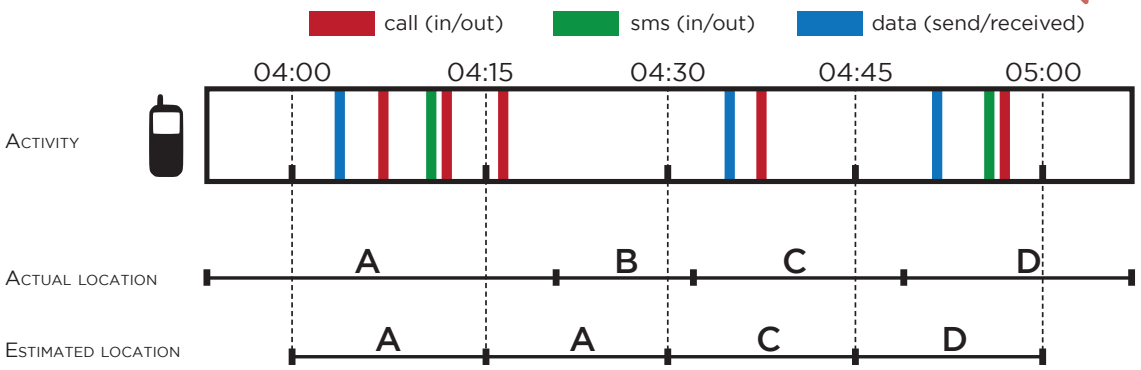
- **A power-law relationship**
  - Mobile network traffic volume and census population density [7,8]

$$\rho_i = \alpha \sigma_i^{\beta}$$

3 case studies: Milan, Rome, Turin



- **Subscriber presence metadata**
  - Rough proxy for *users' location*

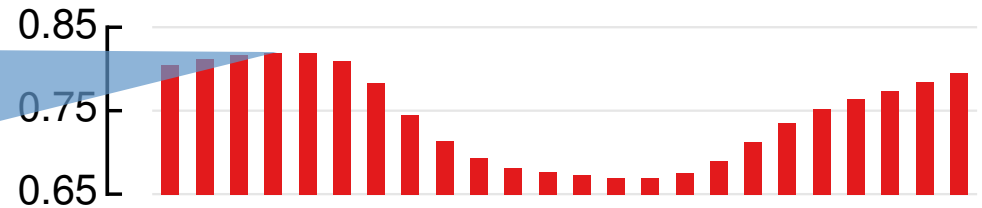


[7] Deville et al., PNAS, 2014; [8] Douglass et al., EPJ Data Science, 2015

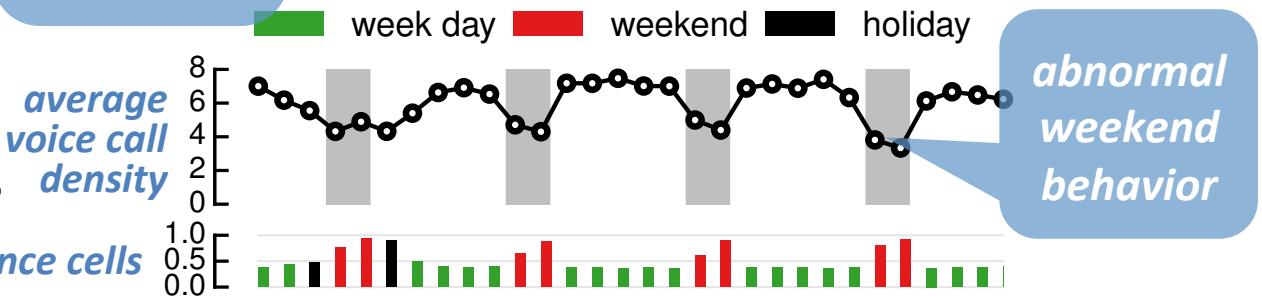
De-noising the correlation

- **Time filtering**
  - 4 to 5 am

maximum correlation overnight [0.8-0.9]



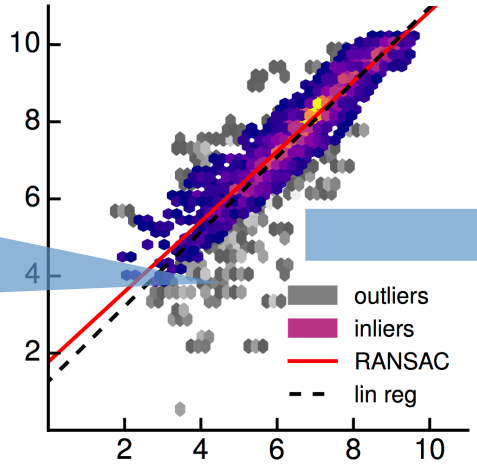
- **Day filtering**
  - Working days

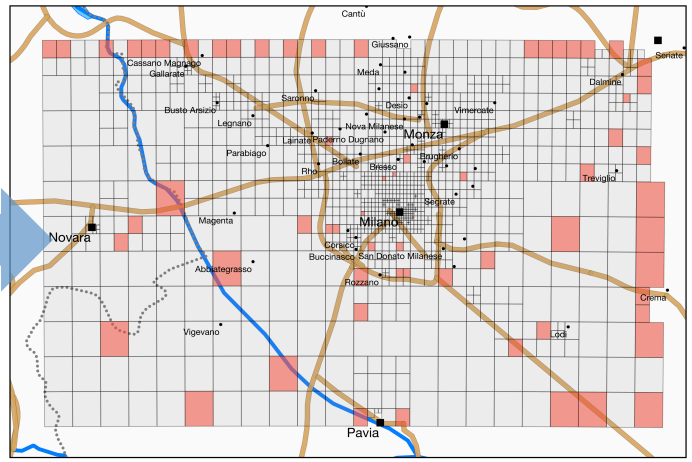


abnormal weekend behavior

- **Cell filtering**
  - Border effect

non-correlated points belong to specific geographical cells





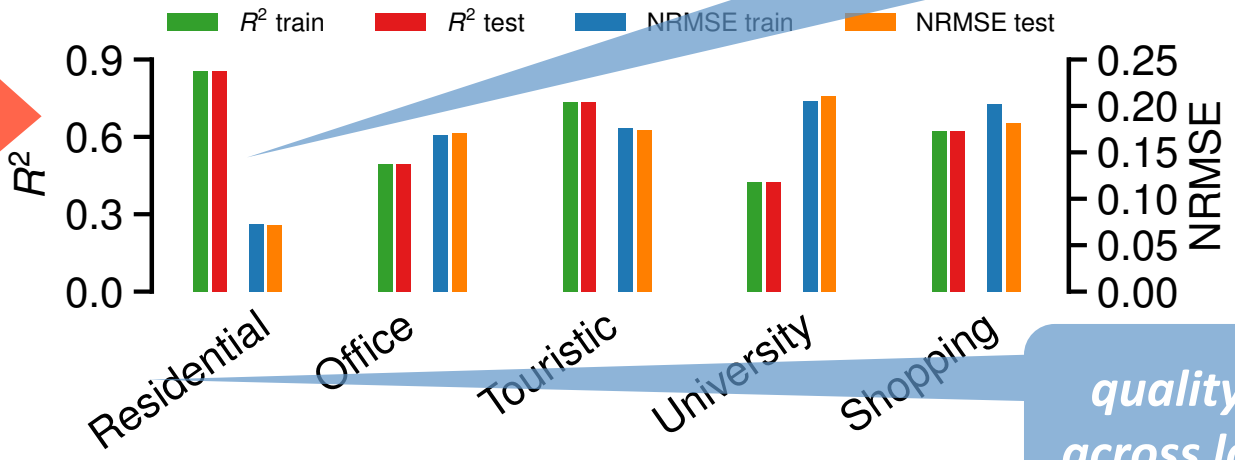
A glance at results

- **Overall estimation quality**

good result in residential areas (accounting for 70-80% of total)

R^2 = 0.8, 8.7% error

21% gain over state of art [8]



quality varies across land uses

- **Model reuse across cities is possible**

R^2 = 0.64–0.84, ~10% error

test

training

	R^2			$NRMSE$		
	Milan	Rome	Turin	Milan	Rome	Turin
MILAN	0.80	0.67	0.84	0.087	0.102	0.088
ROME	0.82	0.73	0.78	0.083	0.093	0.103
TURIN	0.77	0.64	0.84	0.094	0.108	0.087

[8] Douglass et al., EPJ Data Science, 2015

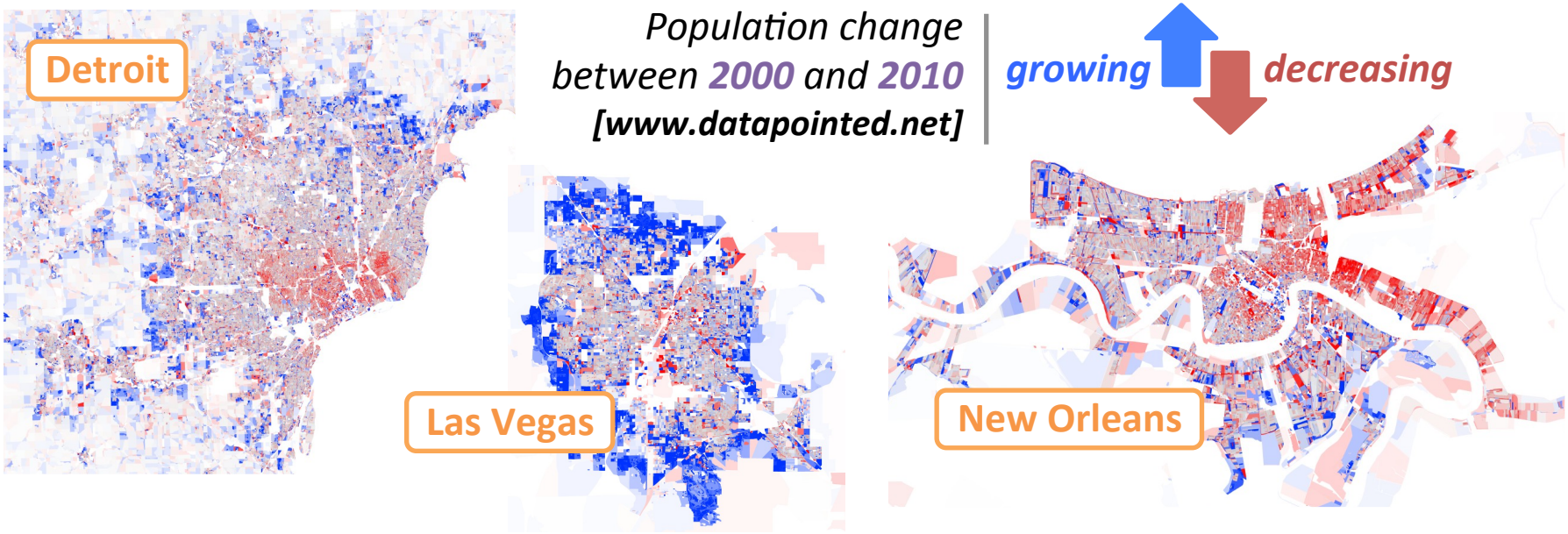
II.3

Towards dynamic urban population densities

A multivariate model

Dynamic population density

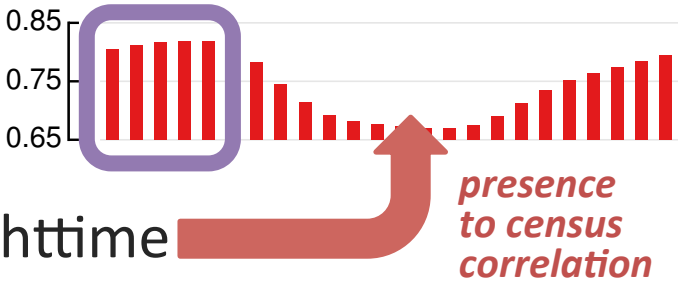
- **Population density is a time-varying phenomenon**
  - Current estimations capture ***long-timescale*** dynamics



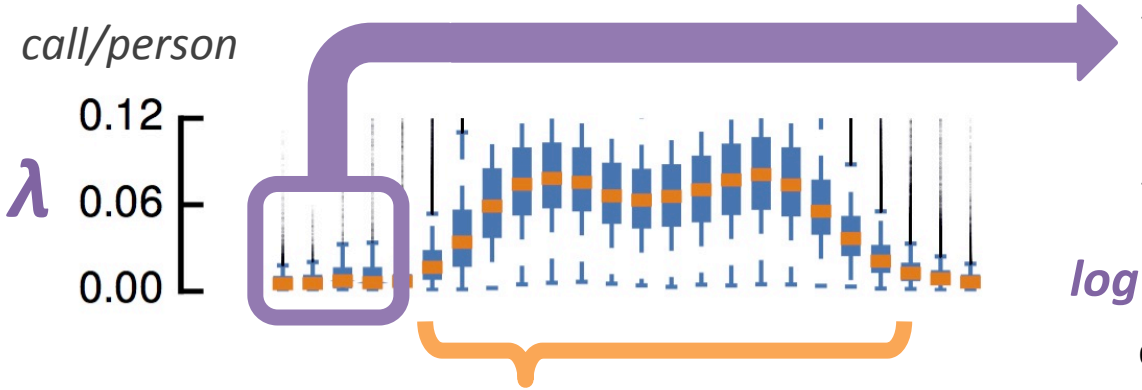
- **What about short-timescale fluctuations?**
  - People distributions in urban areas vary ***within minutes***
  - Mobile network metadata has ***suitable granularity***

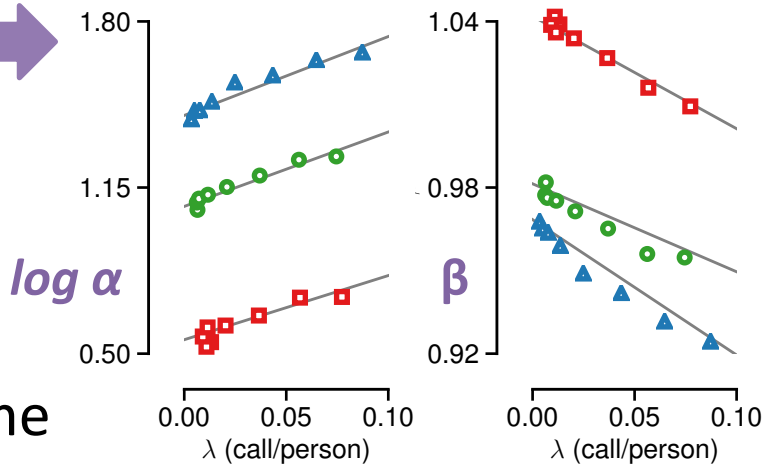
Estimating dynamic populations

- **Major problem: no ground truth**
  - ***Cannot train*** a regression model
  - ***Cannot trust*** a model trained on nighttime



- **A multivariate relationship**
  - α and β can be written as functions of the ***activity level λ***





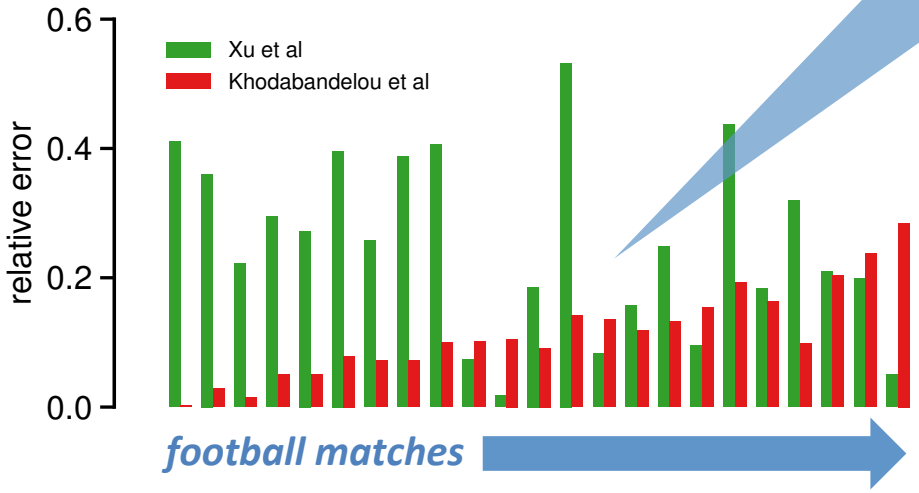
  - λ is a ***measurable*** function of time

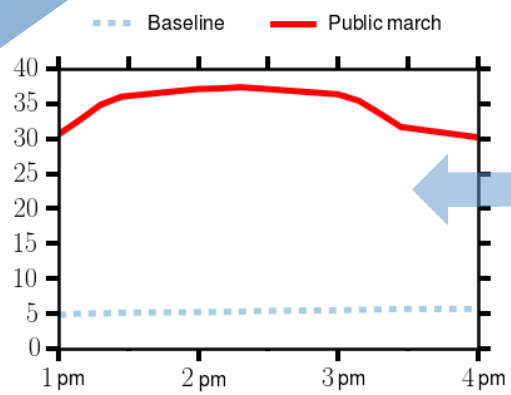
$$\hat{\rho}_i(t) = e^{(\hat{a}_\alpha \lambda_i(t) + \hat{b}_\alpha)} \cdot \sigma_i(t)^{(\hat{a}_\beta \lambda_i(t) + \hat{b}_\beta)}$$

Another glance at results

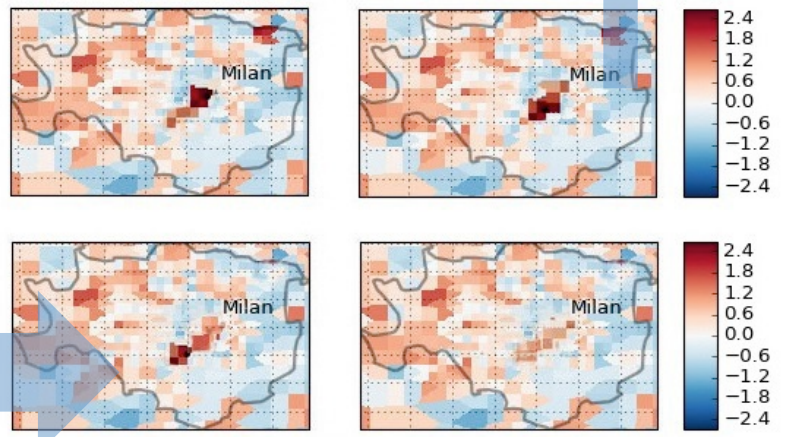
- **Validation**
  - Sports events ***attendance***
  - ***Ground truth*** by organizers

10% error versus 25% of state of art [9]





public march participation estimated at 37,000 (peak)



- **Model exploitation**
  - Morning/afternoon ***commuting***
  - Emergence of ***social events***

[9] Xu et al., ACM UbiComp, 2015

III

Outlook

And perspectives

Outlook

- **Summary**
  - Mobile network data analysis can complement existing land use mapmaking, especially for ***up-to-date mixed land use***
  - Mobile network metadata analysis complements static and enables ***dynamic population density*** estimation

Only two examples

- **Takeaway message**

> *"There is more to mobile networks than "plain" communication-based services"*

  - Mobile network ***unique features***
    - *(i)* ***pervasiveness***, *(il)* very low (additional) ***costs***, *(iii)* active/passive ***individual*** monitoring, *(iv)* decent level of ***spatiotemporal detail***

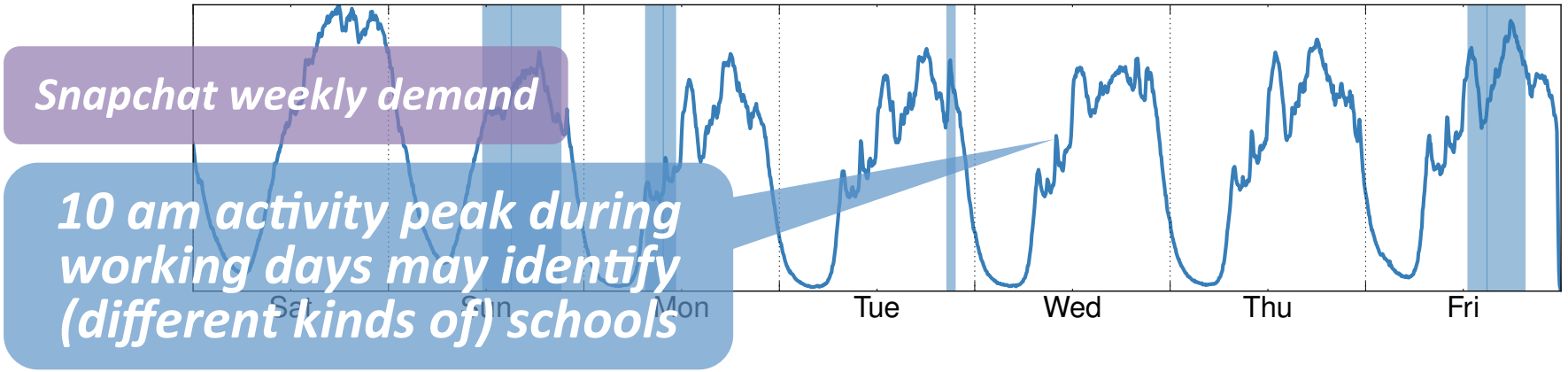
Outlook

- **What is happening now**
  - A ***growing multidisciplinary*** research effort [10,11]
    - also fueled by ***open data challenges*** (e.g., D4D [12] and BDC [13])
  - Operators start understanding this ***added value***
    - increased ***CAPEX*** on monitoring/sensing facilities
    - development of dedicated solutions (e.g., Telefónica ***4th platform*** [14])
    - provisioning of data-driven services (e.g., Orange ***Flux Vision*** [15])
  - Unison with ***pure networking*** goals
    - consistency with a ***cognitive network management*** vision [16]

[10] D. Naboulsi et al., IEEE Communications Surveys and Tutorials, 2016
[11] V. Blondel et al., EPJ Data Science, 2015; [12] V. Blondel et al., arXiv:1210.0137 [cs.CY]
[13] Telecom Italia Big Data Challenge, http:/www.telecomitalia.com/bigdatachallenge
[14] Telefonica Smart Steps, http://dynamicinsights.telefonica.com/smart-steps/
[15] Orange Flux Vision, http://www.orange- business.com/fr/produits/flux-vision
[16] 5GPPP, https://5g-ppp.eu/cognative-network-management-for-5g/

Perspectives

- **What will (possibly) happen next**
  - There is much ***unexploited (meta)data*** in 3G/4G networks
    - e.g., rich ***per-mobile service*** and ***per-user*** information



  - An opportunity for 5G and beyond-5G architectures to be ***"general-purpose systems"*** rather than just "networks"
    - fine-grained ***localization*** (e.g., via mmWave), high-frequency ***tracking*** (e.g., via edge passive probes), ***near-real-time*** provisioning

?

Thanks!

http://perso.citi.insa-lyon.fr/mfiore/

marco.fiore@ieiit.cnr.it

@marc0_fi0re

References

- **Survey**
  - D. Naboulsi, M. Fiore, R. Stanica, S. Ribot, *"Large-scale Mobile Traffic Analysis: a Survey,"* ***IEEE Communications Surveys and Tutorials***, 18(1), 2016
- **Land use mapmaking**

  [http://mobile-traffic-analysis.project.citi-lab.fr/]
  - A. Furno, R. Stanica, M. Fiore, *"Comparative Evaluation of Urban Fabric Detection Techniques Based on Mobile Traffic Data,"* ***ACM/IEEE ASONAM***, Paris, France, 2015
  - A. Furno, M. Fiore, R. Stanica, *"Joint Spatial and Temporal Classification of Mobile Traffic Demands"*, ***IEEE INFOCOM***, Atlanta, GA, USA, 2017
  - A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, Z. Smoreda, *"A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas"*, ***IEEE Transactions on Mobile Computing***, 16(10), 2017
- **Population density estimation**

  [https://doi.org/10.5281/zenodo.1012194]
  - G. Khodabandelou, V. Gauthier, M. Fiore, M. El-Yacoubi, *"Estimation of Static and Dynamic Urban Populations with Mobile Network Metadata,"* ***IEEE Transactions on Mobile Computing***, pre-print