Telefónica

Investigación y Desarrollo Chile

Urban Dynamics Estimation using Mobile Phone logs and LVA

Oscar F. Peredo, José Garcia, Ricardo Stuven and Julián M. Ortiz

10th International Geostatistics Congress, Valencia, Spain

oscar.peredoa@telefonica.com

8th September 2016

ELEGIMOS TODO_

Telefónica R+D

- R & D branch of Telefónica Group (Telecom)
- Telefónica global subscribers: 272.6 millions (June 2016)
- R & D offices in Spain, UK, Israel, Brazil and Chile
- Research topics:
  - Prediction of human behavior
  - Industrial Internet of things
  - Network infrastructure and connectivity
  - Identity and privacy

Mobile phone data

- ***Event-driven cell phone data****: triggered directly by the user or indirectly by automatic apps
  - Call Detail Records (Billing/Invoicing purposes)
  - Internet access logs (URLs, ports, bytes transferred,…)
- ***Network-driven cell phone data****: triggered indirectly by the network
  - "Local Area Code" changes
  - Handovers (antenna changes in active use)
  - On/Off device switching
  - Paging (periodic location update, 2-4 hrs)

* Extracted from (Oliver, Matic and Frias-Martinez, 2015)

Call Detail Records

- Billing data
- Voice, SMS, MMS, aggregated Internet traffic
- Large-scale longitudinal data
- All telcos store CDRs for long periods (months/ years, reasonable storage size)

Fields	Sample value
Origin number	6957491b18b029f3f95e651ba7
Destination number	bd48d4c6c086d985da6a698fa0
Date (yymmdd)	140823
Time (hhmmss)	084533
Duration (seconds or KB)	447
Origin Antenna ID	CURZFU3
Destination Antenna ID	ATLBAL1

Internet access logs

- Used to measure "QoS" (byte downstream/ upstream for each user and antenna)
- App/Service profiling (Browsers, Instant Msg,…)
- Not all telcos store this dataset for large periods (HUGE storage size)

Fields	Sample value
Origin number	6957491b18b029f3f95e651ba7
Date (yymmdd)	140823
Initial Time (hhmmss.mmm)	084531.865
Final Time (hhmmss.mmm)	084533.033
IP	216.58.208.132
Port	80
URL	http://www.google.com
Origin Antenna ID	CURZFU3

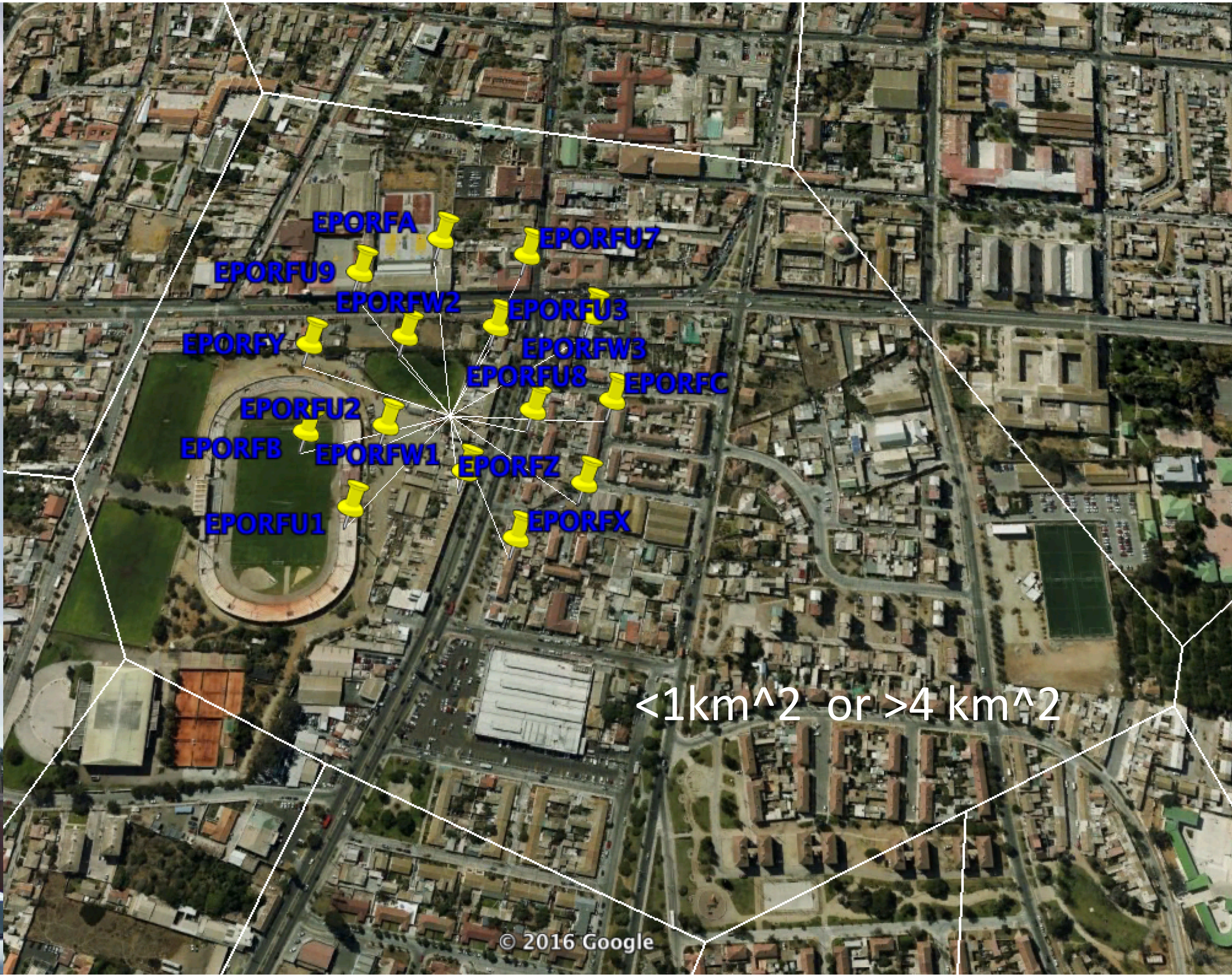
Can we estimate urban dynamics using CDRs and Internet logs?

Yes!

- Population mobility (Song, 2010)
- Transportation flows (Chiang, 2011)
- Infer socio-economic indicators (Soto, 2011)
- Exposure to air pollution (Liu, 2013)
- Privacy issues (de Montoye, 2013)
- User demographics (Dong, 2014)
- Predict crime (Bogomolov, 2014)
- Credit scores (San Pedro, 2015)
- Public health (Oliver, 2015)
- …

However,
we have a large spatial resolution…

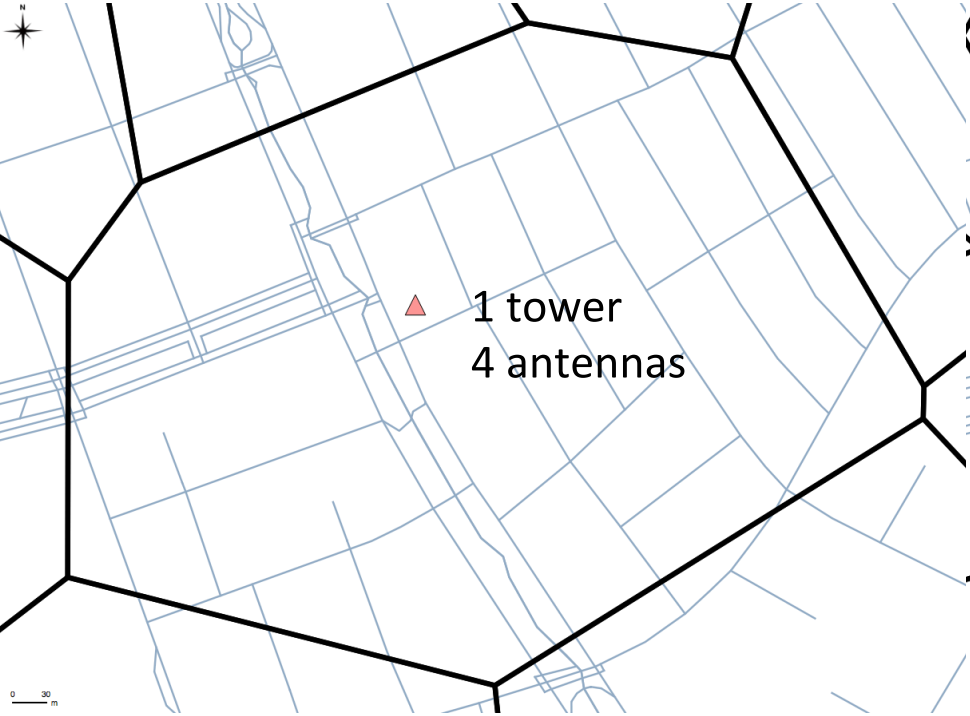
Tower-based location of subscribers

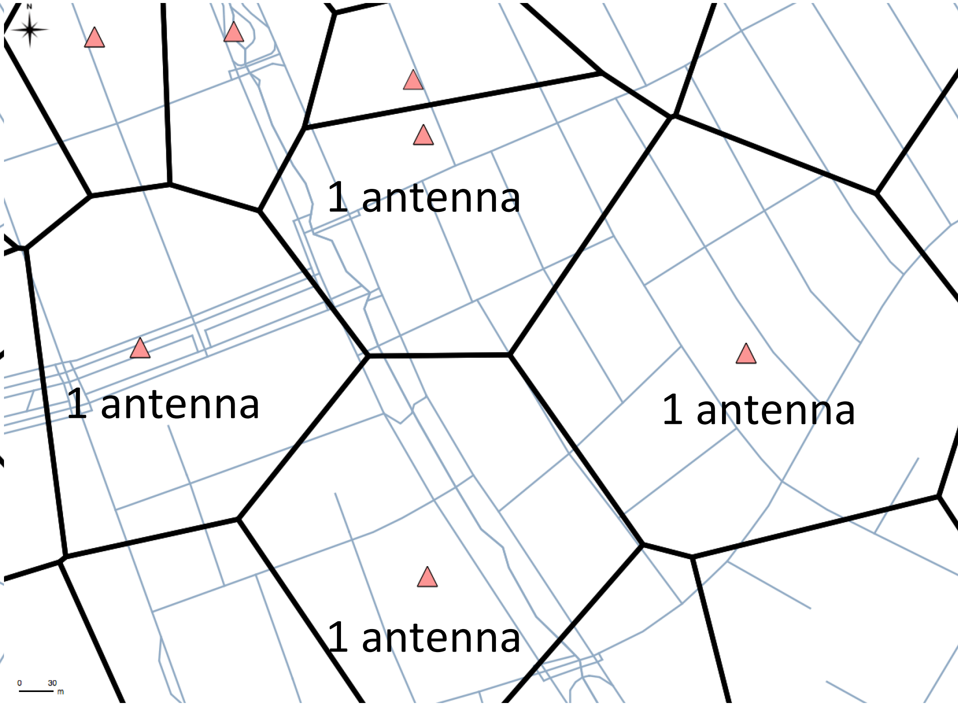


How can we improve the resolution of the estimation?

From tower to antenna (Voronoi sub-sampling)…

(Graells-Garrido & Peredo, 2016)





but the spatial resolution is still large…
(also there are several single-antenna towers)

Proposal: Urban-scale interpolation of mobile phone data

Not a new idea, but with a different source of information…

Modeling locally varying anisotropy of CO_2 emissions in the United States

Jeff B. Boisvert · Clayton V. Deutsch

AN EXPONENTIAL EXTERNAL DRIFT MODELING FOR URBAN SCALE INTERPOLATION,

Maxime Beauchamp, *INERIS, France*,
Laure Malherbe, *INERIS, France*,
Chantal de Fouquet, *Mines ParisTech, France*,
Laurent Létinois, *INERIS, France*

Case study: Santiago, Chile

Telefónica in Chile

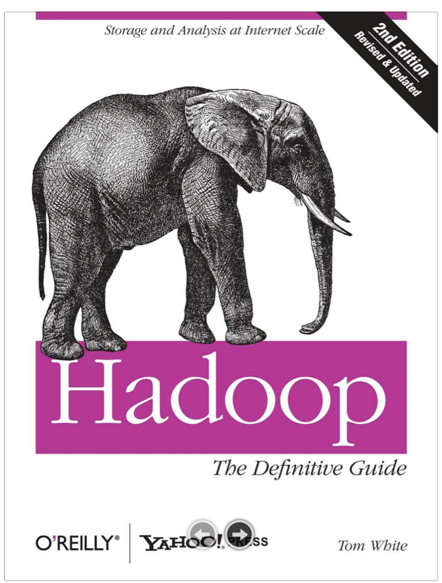
- Population of 18 million
- 35% Telefónica market share
- 40,000 cell phone antennas (tech 2G, 3G, 4G)
- One day in Chile(estimated values):
  - 120 million of CDRs (voice and data)
    - 25 GB raw
  - 6 billion of Internet logs
    - 2000 GB raw
- One year:
  - 9.1 TB of CDR
  - 730 TB of Internet logs

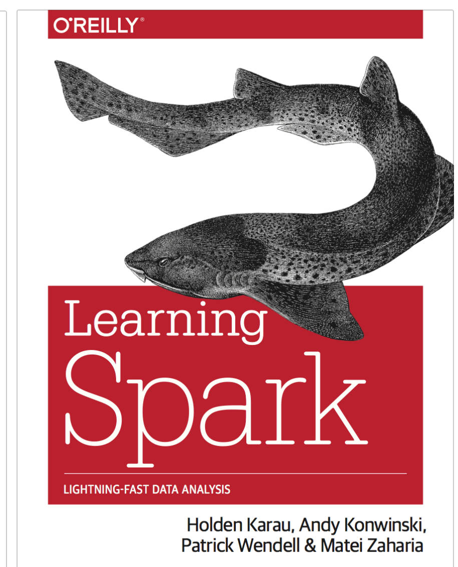
Steps in the project

0. ***Infrastructure***
1. Build mobility traces
2. Identify travelers
3. Aggregate travelers traces into antennas for each time frame
4. Extract street map
5. Build LVA field from streets
6. Run LVA estimation for each time frame

Step 0: Infrastructure

- Extraction-Transformation-Load from corporate databases: Java, Shell, Networking
- Cloud Computing provider: Microsoft Azure
- Storage: Hadoop (HDFS)
- Data-processing tools: Pig, Tez, Hive, Spark
- Parallel computing tools: MPI, OpenMP, CUDA





Steps in the project

0. Infrastructure
1. ***Build mobility traces***
2. Identify travelers
3. Aggregate travelers traces into antennas for each time frame
4. Extract street map
5. Build LVA field from streets
6. Run LVA estimation for each time frame

Step 1: Build mobility traces

From mobile phone events:

```
UID_123,TIME_001,ANTENNA_A
UID_235,TIME_002,ANTENNA_A
UID_123,TIME_003,ANTENNA_B
UID_654,TIME_005,ANTENNA_D
...
```

Large-scale complex SQL query
Join + Group_by

to mobility traces for each subscriptor:

```
{UID_123,((TIME_001,ANTENNA_A),(TIME_003,ANTENNA_B))}
{UID_235,((TIME_002,ANTENNA_A))}
{UID_654,((TIME_005,ANTENNA_D))}
...
```

Steps in the project

0. Infrastructure
1. Build mobility traces
2. ***Identify travelers***
3. Aggregate travelers traces into antennas for each time frame
4. Extract street map
5. Build LVA field from streets
6. Run LVA estimation for each time frame

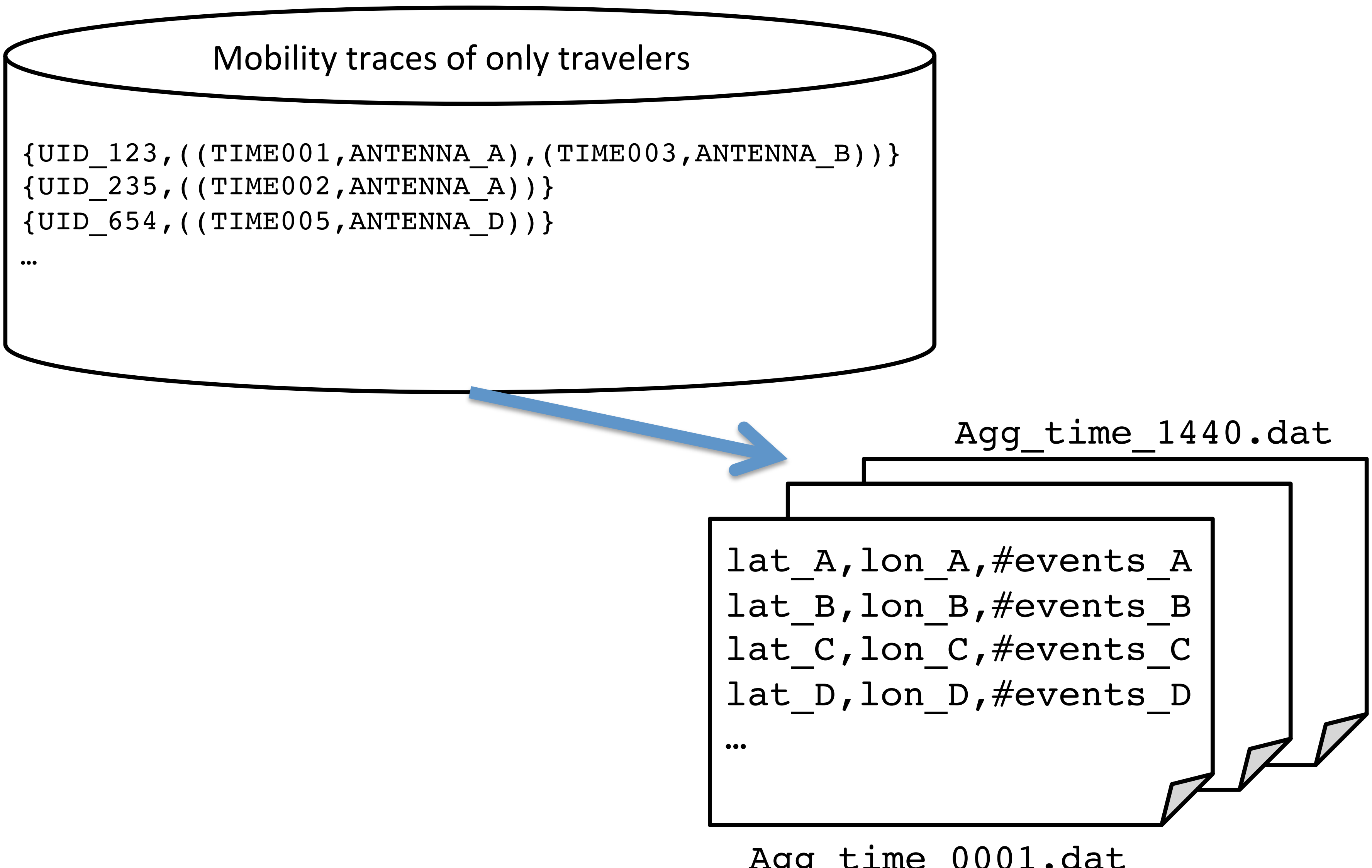
Step 2: Identify traveling subscribers

- Several authors: (Isaacman, 2012) (Phithakkitnukoon, 2012), (Kung, 2014), …
- Home/Work travelers
  - Home: Most frequent antenna during nighttime hours across several weeks
  - Work: Most frequent antenna during business hours across several weeks
  - If distance(home,work)> threshold then OK
  - Validation using most recent Census

Steps in the project

0. Infrastructure
1. Build mobility traces
2. Identify travelers
3. ***Aggregate travelers traces into antennas for each time frame***
4. Extract street map
5. Build LVA field from streets
6. Run LVA estimation for each time frame

Step 3: Aggregate traveler events by antenna



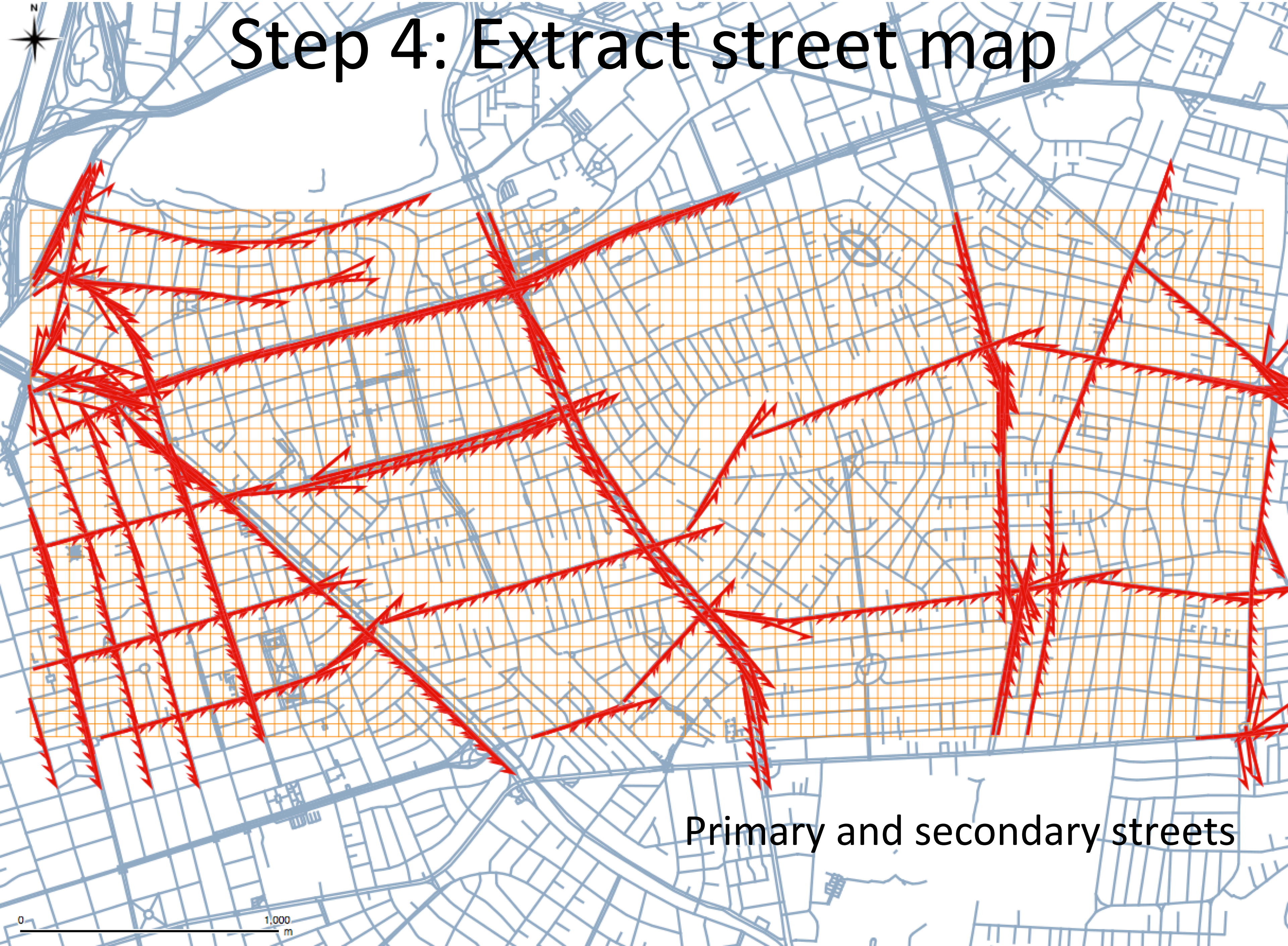
Steps in the project

0. Infrastructure
1. Build mobility traces
2. Identify travelers
3. Aggregate travelers traces into antennas for each time frame
4. ***Extract street map***
5. Build LVA field from streets
6. Run LVA estimation for each time frame

Step 4: Extract street map

Step 4: Extract street map

Step 4: Extract street map

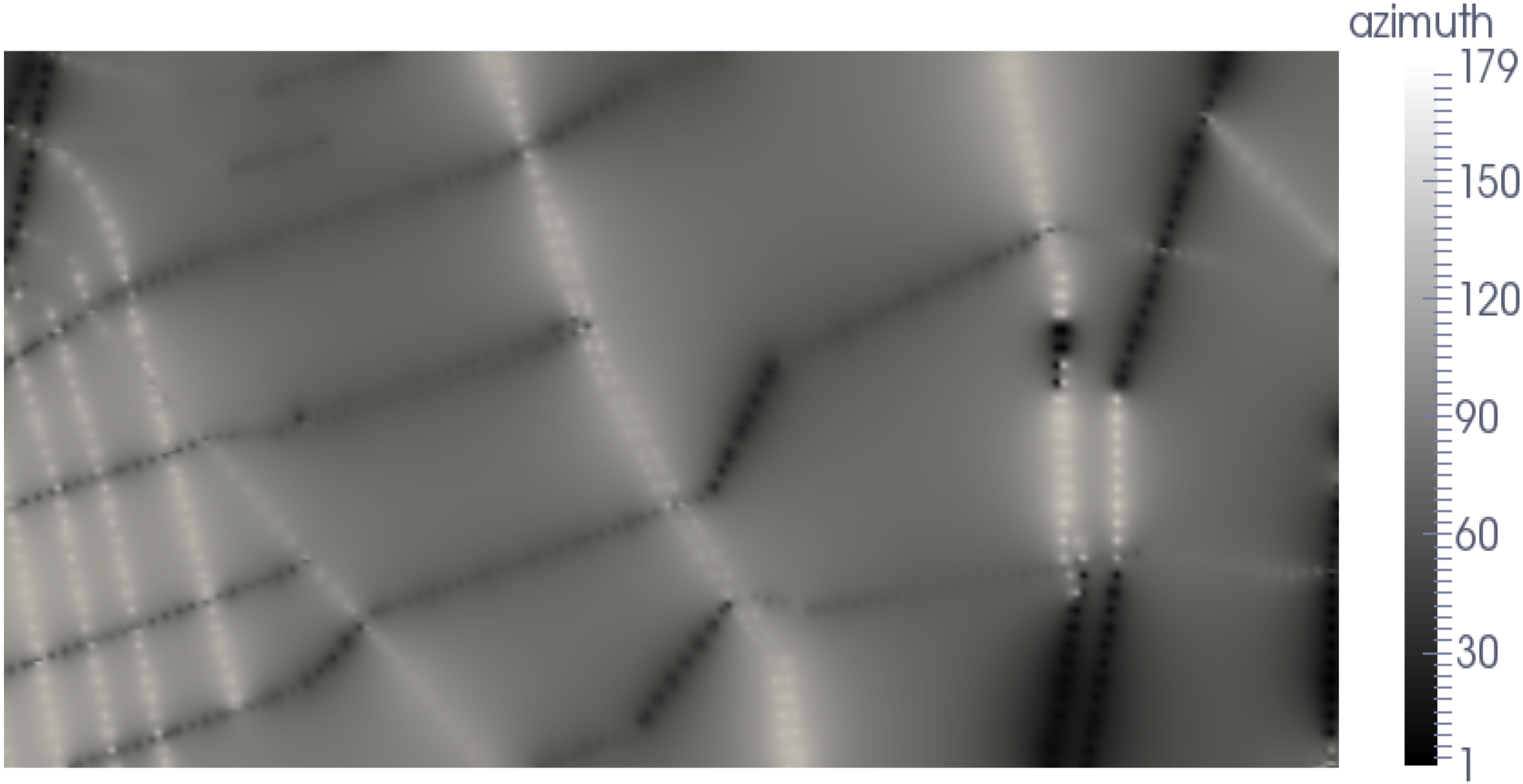


Primary and secondary streets

Steps in the project

0. Infrastructure
1. Build mobility traces
2. Identify travelers
3. Aggregate travelers traces into antennas for each time frame
4. Extract street map
5. ***Build LVA field from streets***
6. Run LVA estimation for each time frame

Step 5: Build LVA field from streets



Steps in the project

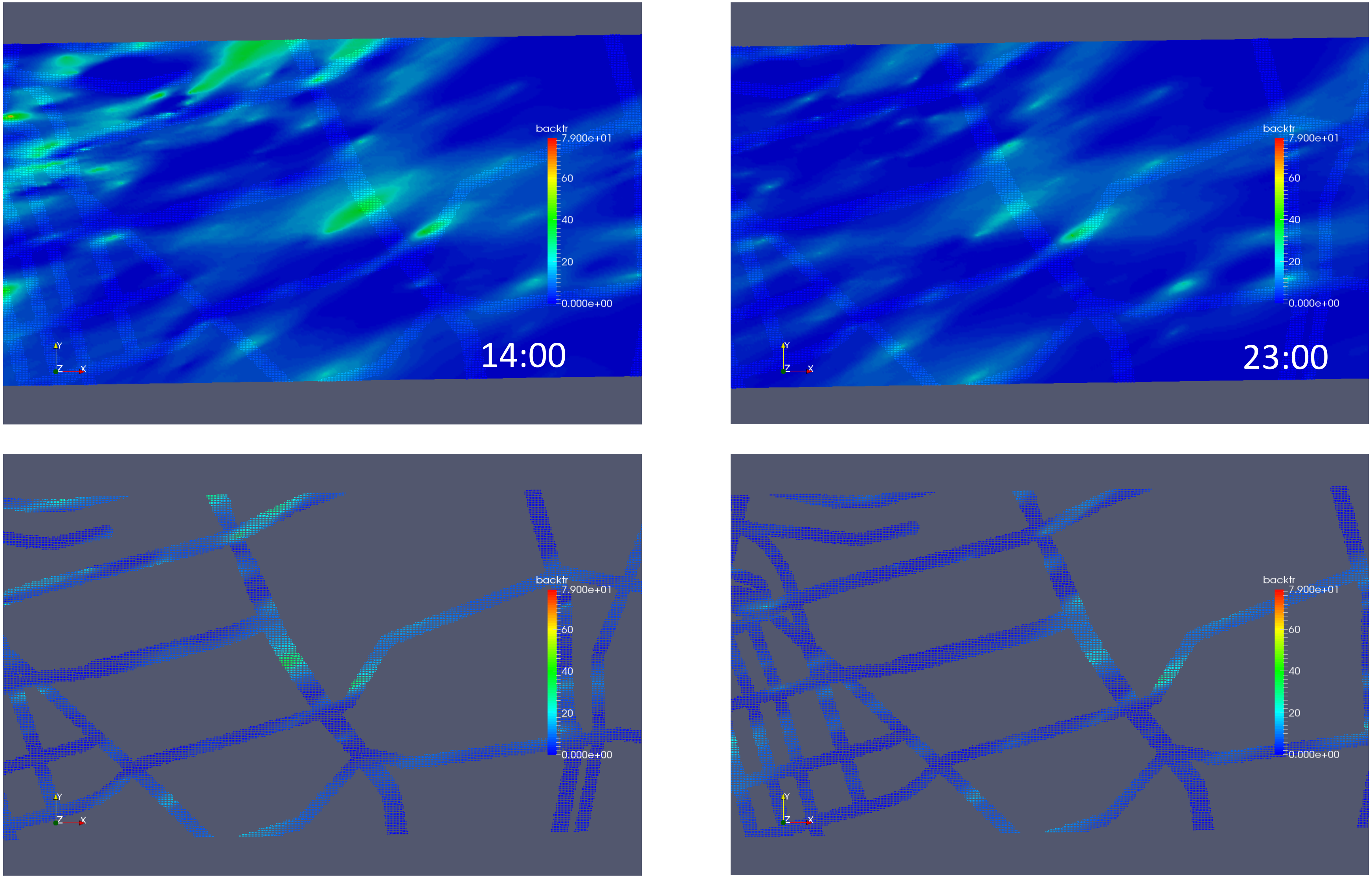
0. Infrastructure
1. Build mobility traces
2. Identify travelers
3. Aggregate travelers traces into antennas for each time frame
4. Extract street map
5. Build LVA field from streets
6. ***Run LVA estimation for each time frame***

Step 6: Calculate estimations using Jeff's LVA Fortran/C++ code

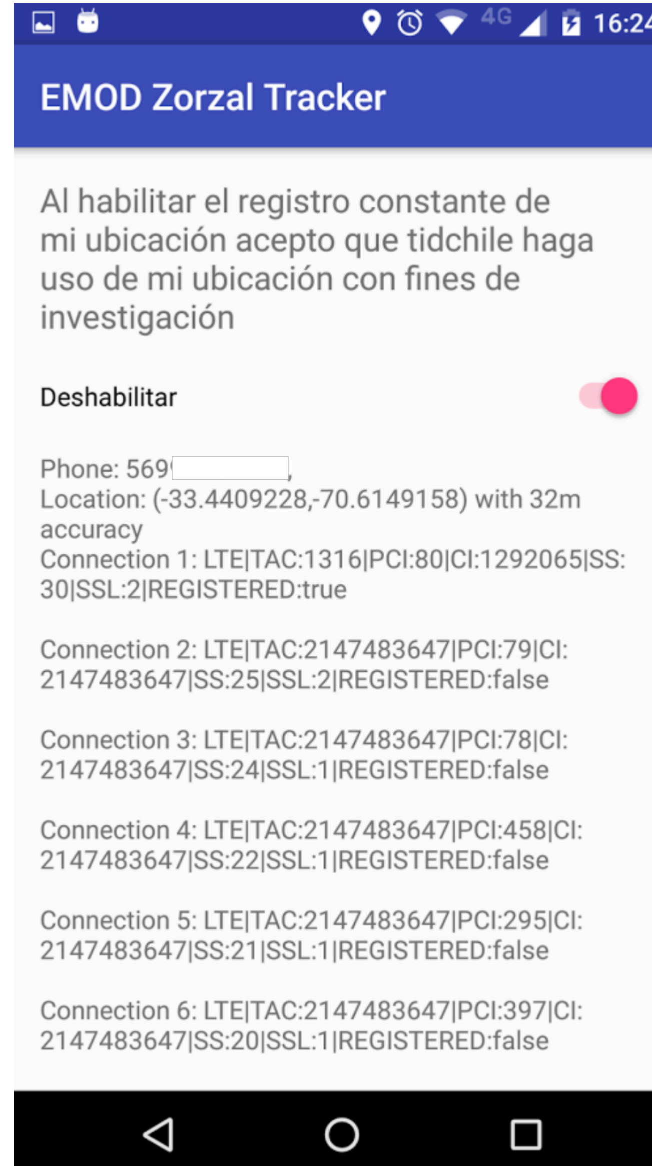
- Some code modifications:
  - OpenMP/CUDA execution of Dijkstra
  - OpenMP operations for L-ISOMAP

- Parallel/Distributed execution of LVA estimations for 1440 time frames

Step 6: Calculate estimations using Jeff's LVA Fortran/C++ code



Validation: GPS measurements



Conclusions

- On-going telecom project, with strong roots in geostatistics
- Can we do this in real-time? (<10 secs)
- Cell phone antennas are "sensors" of human behaviour
- The battle for your data: ISP (telcos) vs OTT (apps)

Thank you!

Contact: oscar.peredoa@telefonica.com