

Exploring the use of mobile phones during domestic tourism trips

Maarten Vanhoof
Orange Labs/Newcastle University

M.vanhoof1@newcastle.ac.uk
@Metti Hoof
MaartenVanhoof.com



Together with

Liane Hendrickx (KU Leuven, BE)

Gert Verstraeten (KU Leuven, BE)

Aare Puusaar (Newcastle University, UK)

Thomas Plötz (Newcastle University (UK)

Maxim Janzen (ETH Zürich)

Zbigniew Smoreda (Orange Labs, FR)

Mobile Phone Data (Call Detail Records -CDR)

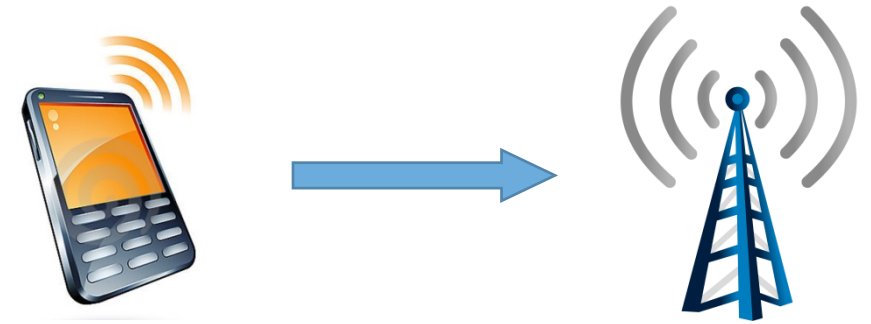
Orange dataset

25 million unique utilisateurs

13 mai – 14 october 2007 (153 days)

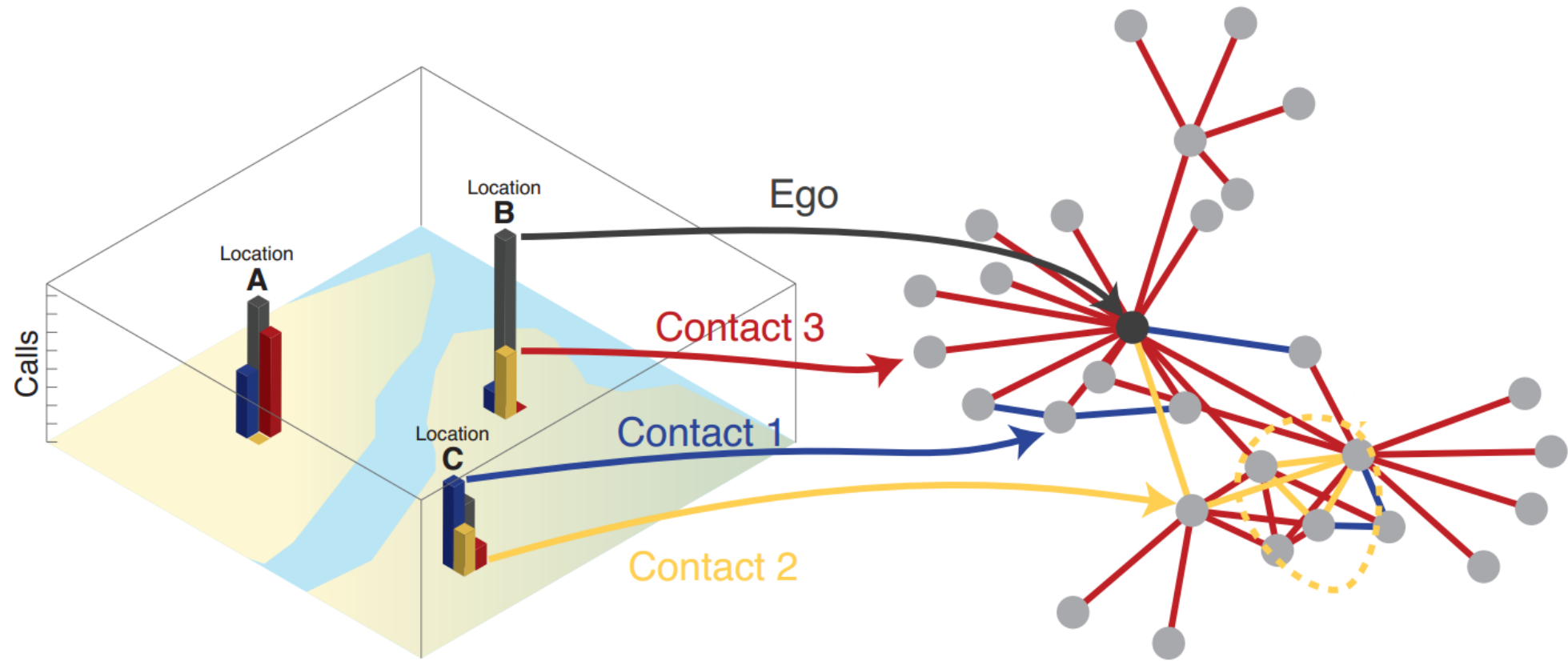
Metadata

- Caller (phone)
- Called phone
- Timestamp
- Type of event
- Duration of call/Length of text
- Location of celltower



| timestamp | caller | callee | event | duration | area id | tower id |
|---------------------|----------|-----------|-------|----------|---------|----------|
| 2007/10/01 23:45:00 | HJ123423 | R482G9342 | VO | 3656s | 1548 | 53571 |
| 2007/13/01 12:10:04 | TR234S3 | 43FG3423 | SI | 125c | 32768 | 53571 |
| ... | ... | ... | ... | ... | ... | ... |

Mobile Phone Data (Call Detail Records)



Context

- Mobile phone data has successfully been used for
 - Population density estimation (Deville et al. 2014)
 - Delineation of territory (Sobolevsky et al. 2013)
 - Movement tracking (Chen et al. 2014)
 - Transportation models (Liu et al. 2014)
 - ...
- In mobility/tourism research
 - Deriving seasonality of foreign tourists' space consumption (Ahas et al. 2007)
 - Segmenting unique and repeat visitors (Kuusik 2011)
 - Evaluation of travel distances for event visitors (Nilbe et al. 2014)
 - Detecting tourism destinations (Raun 2015)
 - Extracting magnitudes of long-distance travel (Janzen et al. 2016)

Context

Advantages:

- Huge samples (millions of users)
- Passive location detection (so no respondent fatigue)
- Long and continuous observation periods (typically several months)

Opportunities:

- To investigate specific types of activities compared to baseline behaviour
- E.g. tourism travel vs. daily routine mobility

Context

Disadvantages:

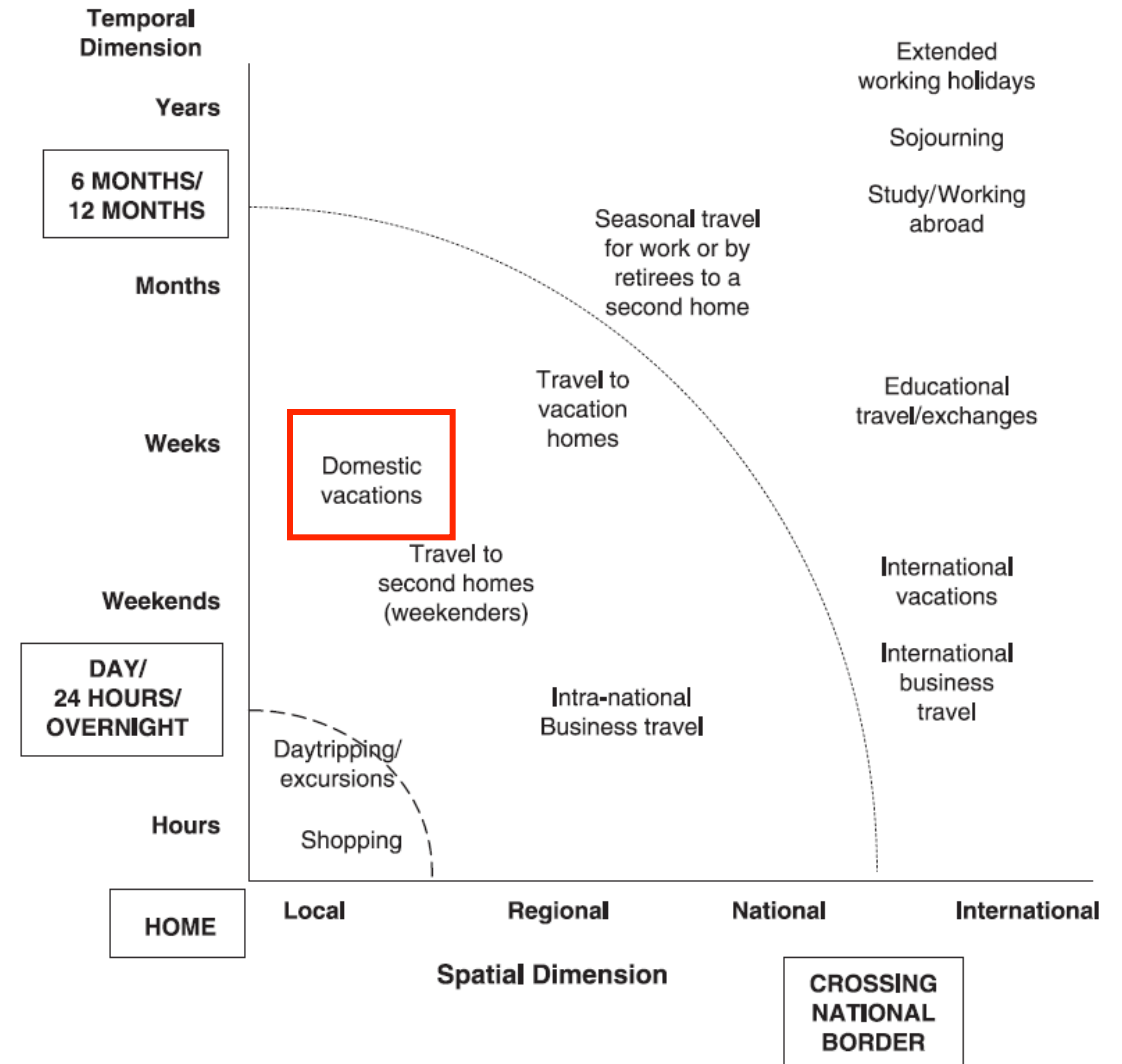
- Temporal and spatial resolution of observation is dependent on user's phone activity
- Purely technical annotation: no information on activity, context,...
- Impossible (yet?) to have proper validation data on all aspects of mobility

Challenges

- Type of trip detection
- Purpose imputation
- Social context derivation
- ...

Research question(s)

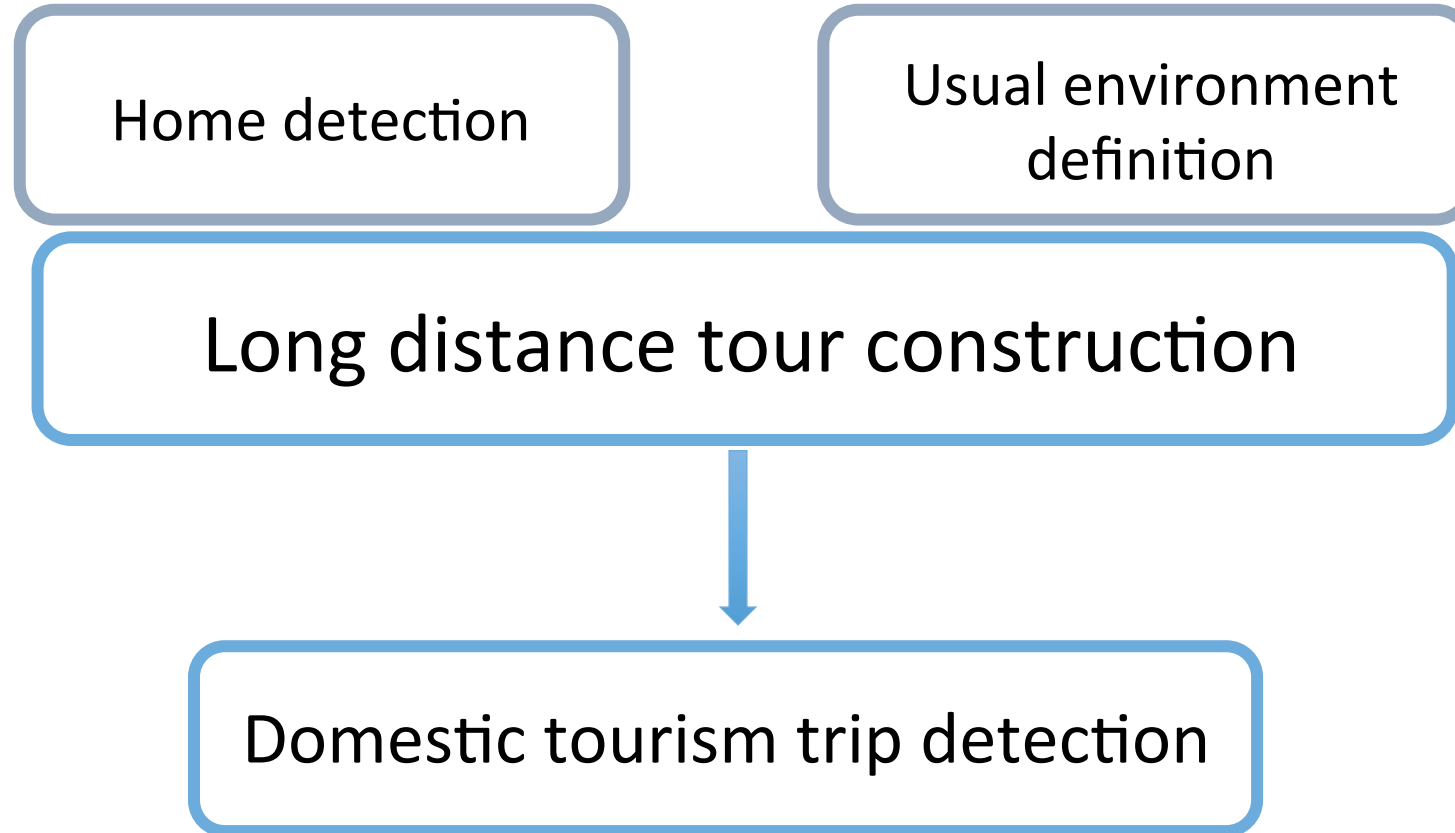
Is it possible to distinguish just one 'type of mobility':
domestic tourism trips
from CDR data?



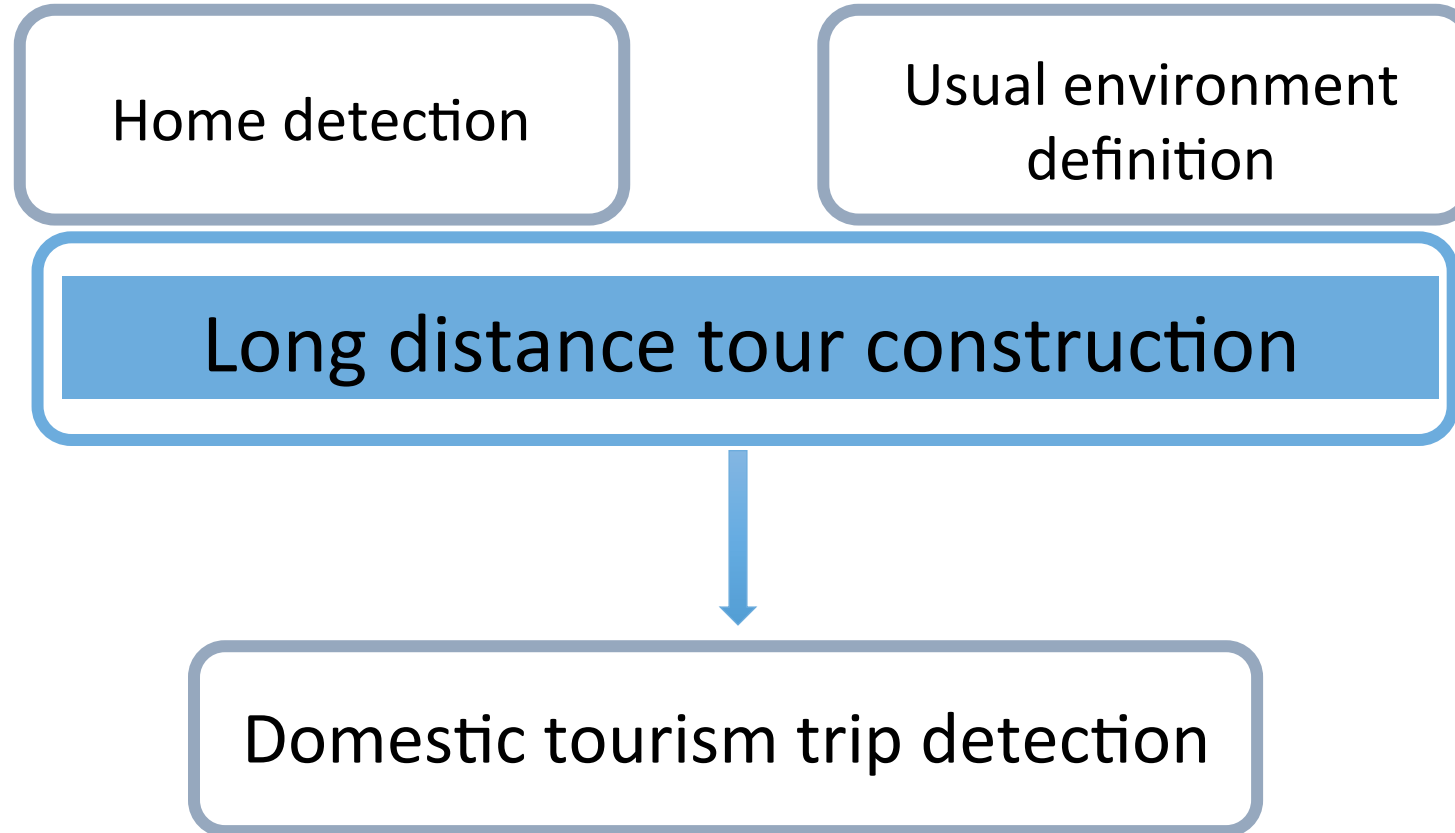
Research question(s)

1. Is it possible to distinguish just one 'type of mobility': **domestic tourism trips** from CDR data?
2. Given that no validation dataset is available; how plausible are temporal and spatial patterns of the detected domestic tourism trips?
3. Can we link types of mobility with activation of social networks?

Methodology

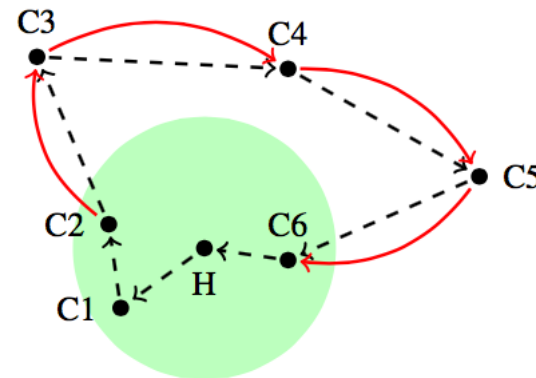


Methodology






Extracting long-distance trips from CDR data

(a) Perfect tour reconstruction

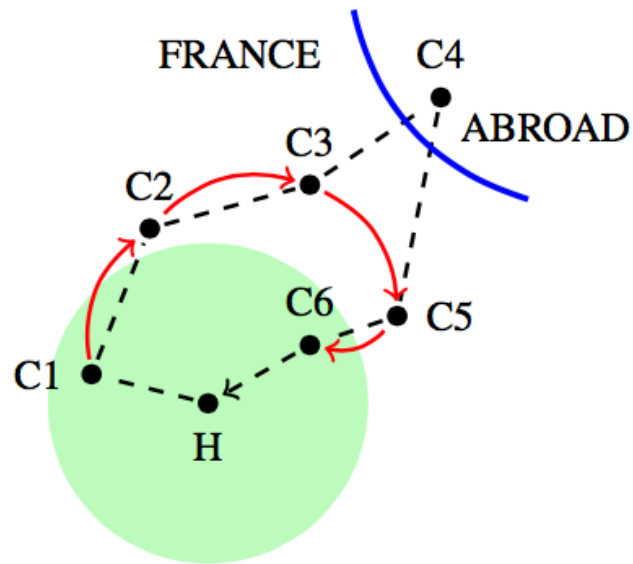


Legend

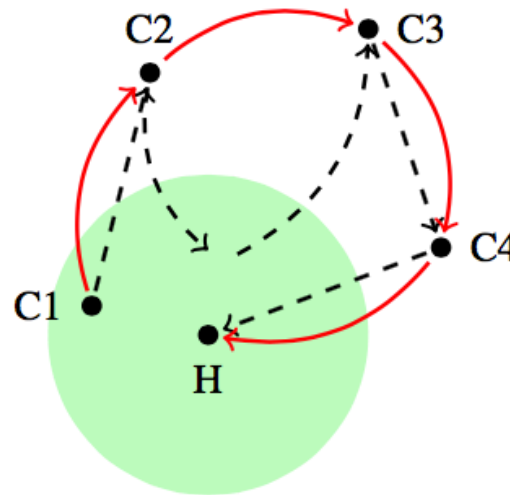
H - Home anchor, C1...C6 - CDR positions,  - User environment,  - Reconstructed tour,  - Real world tour

Extracting long-distance trips from CDR data

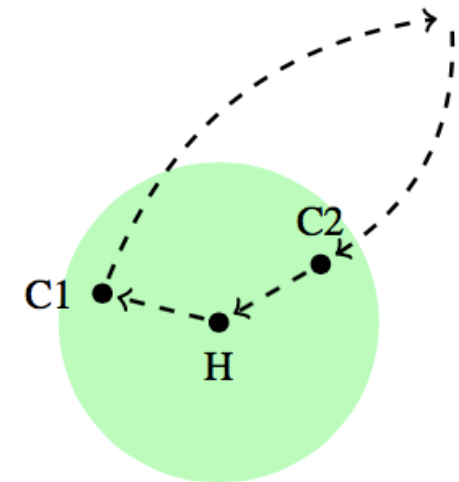
(a) Missed tour abroad





(b) Two tours merged



(c) Tour completely missed



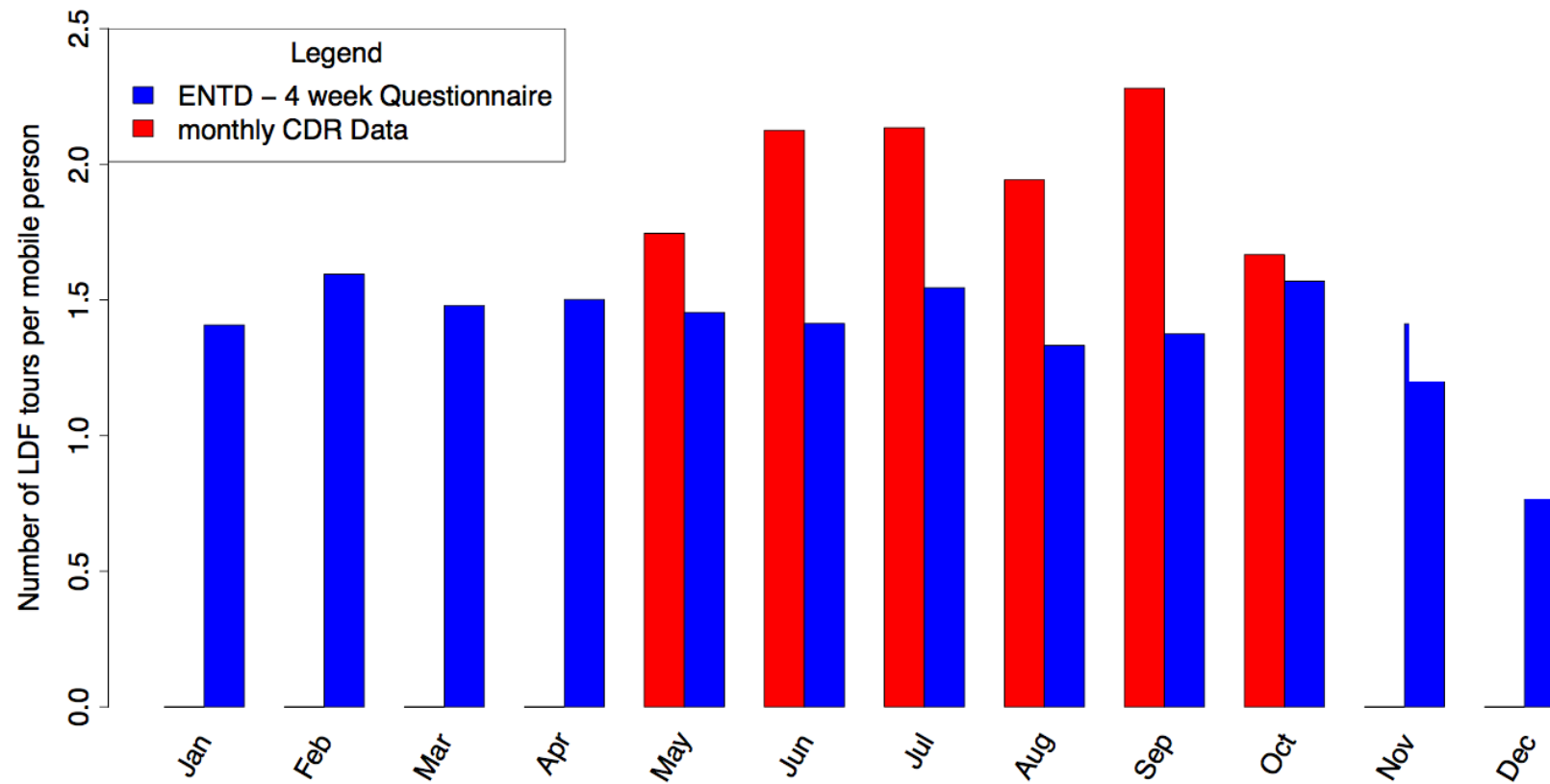
Legend

H - Home anchor, C1...C6 - CDR positions,  - User environment,  - Reconstructed tour, - > - Real world tour

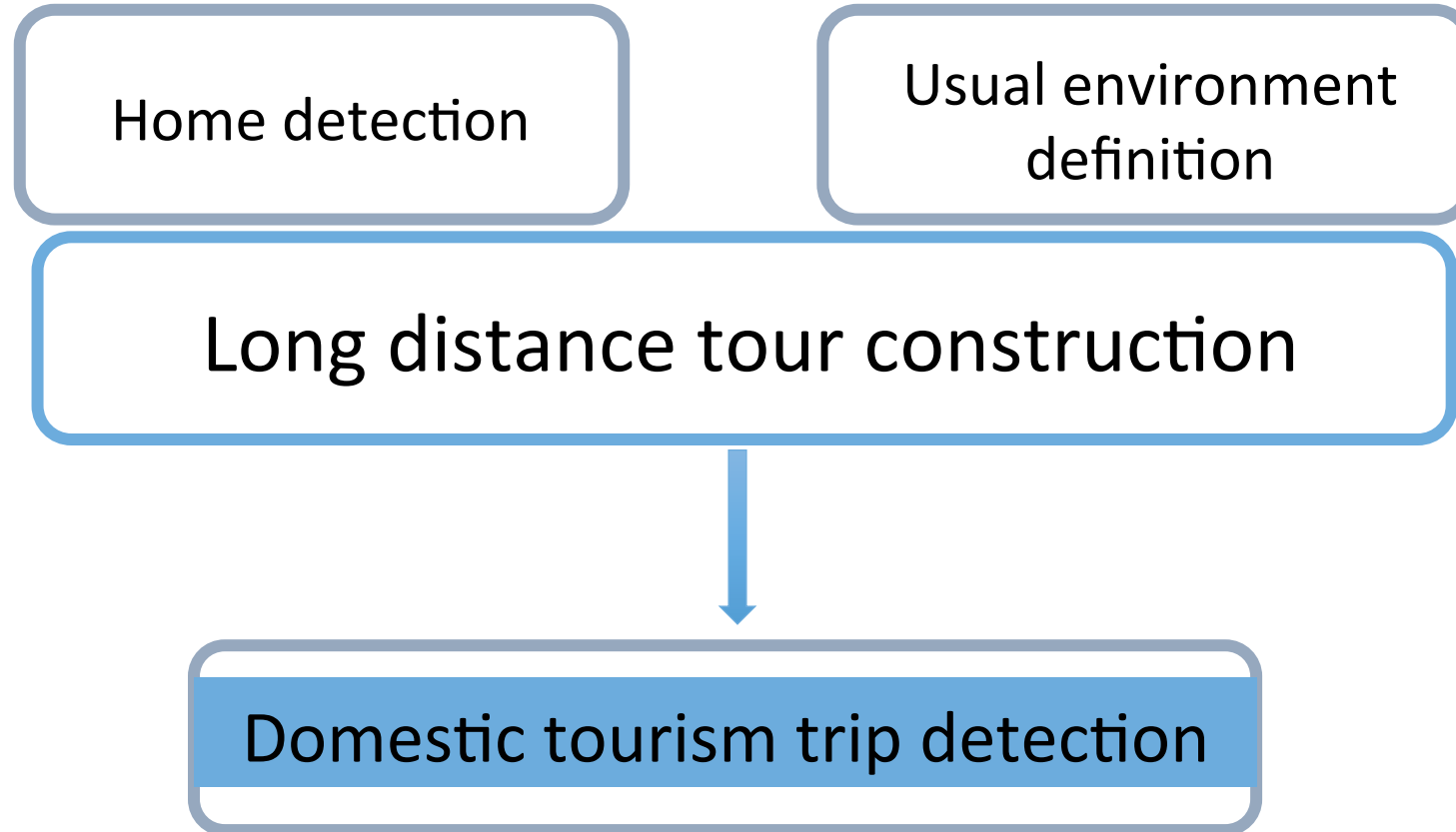
Extracting long-distance trips from CDR data

| | CDR Data | | ENTD 2008 | | ENTD 2008 | |
|-------------------------------|----------------|-------|-----------|-------|-----------|-------|
| Reporting/Tracked Interval | 30 days (June) | | 28 days | | 91 days | |
| Surveyed Persons | 1'388'941 | | 18'632 | | 18'632 | |
| LDF Mobile Persons | 814'381 | 58.6% | 4'796 | 25.7% | 8'743 | 46.9% |
| Selected for further analysis | 79'874 | | 4'796 | | 8'743 | |

Extracting long-distance trips from CDR data

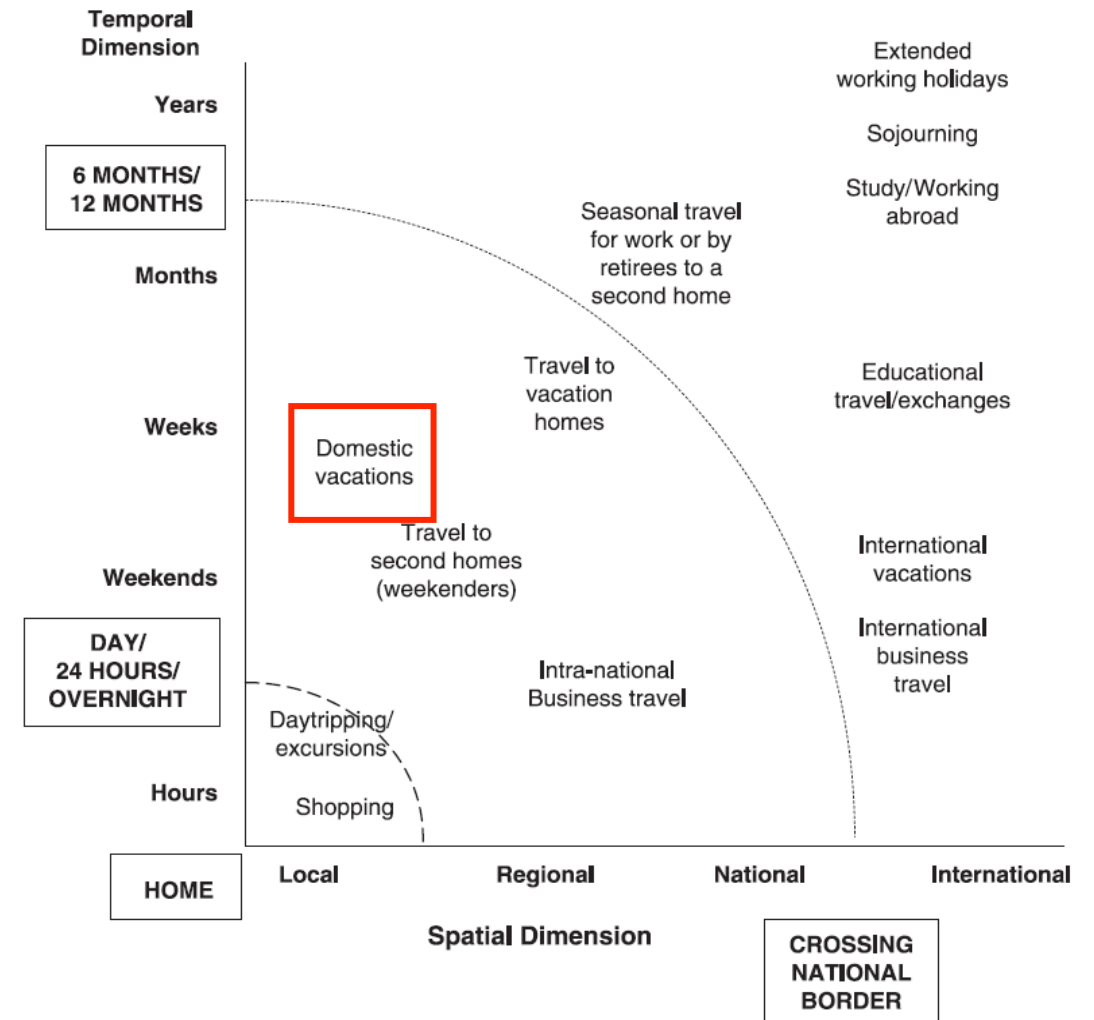


Methodology



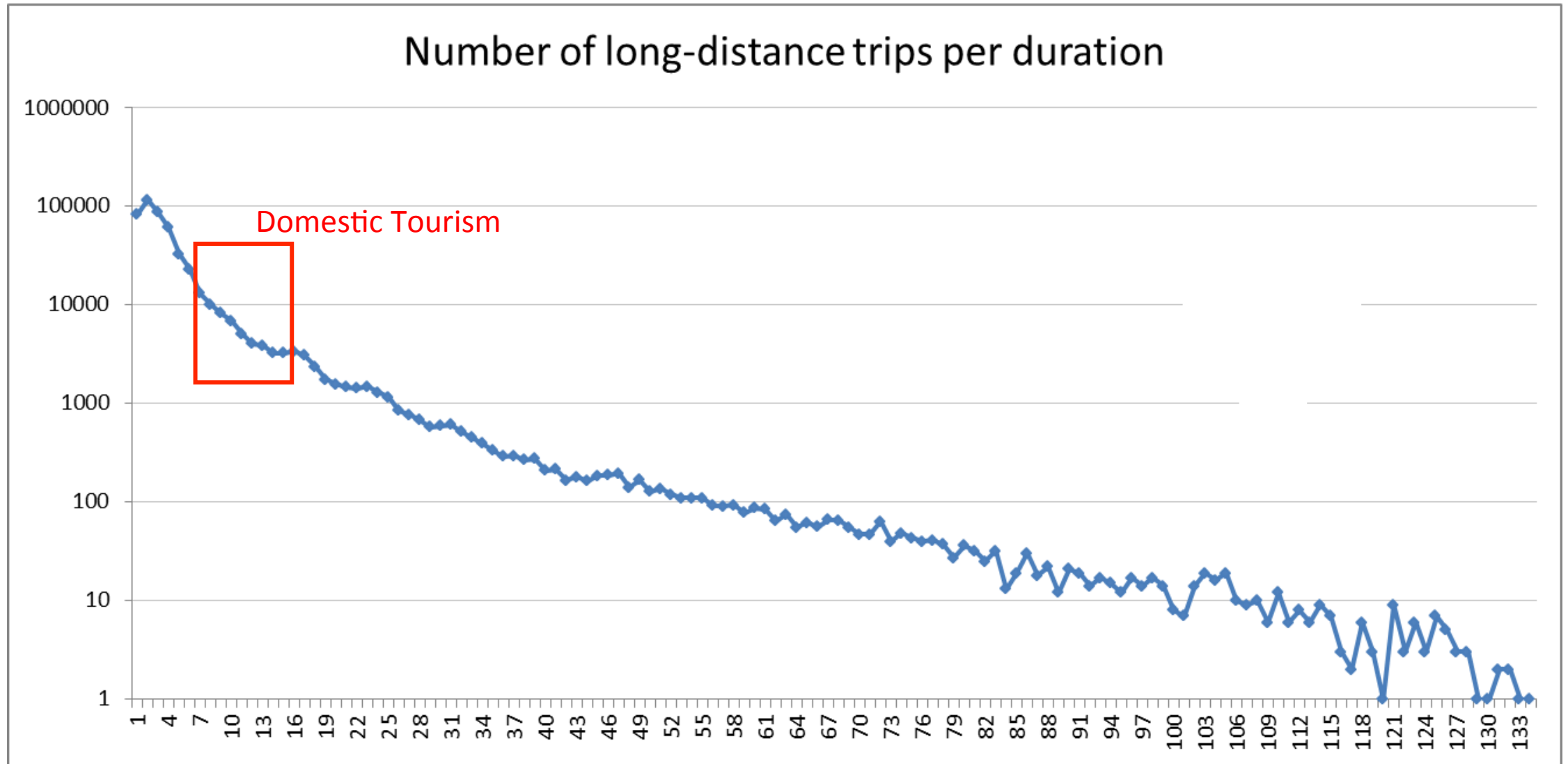
Simple heuristics for defining domestic tourism trips

1. Duration between 7 and 15 days.



Bell and Ward, 2000; Hall, 2003, 2005

Limitation of duration between 7 and 15 days

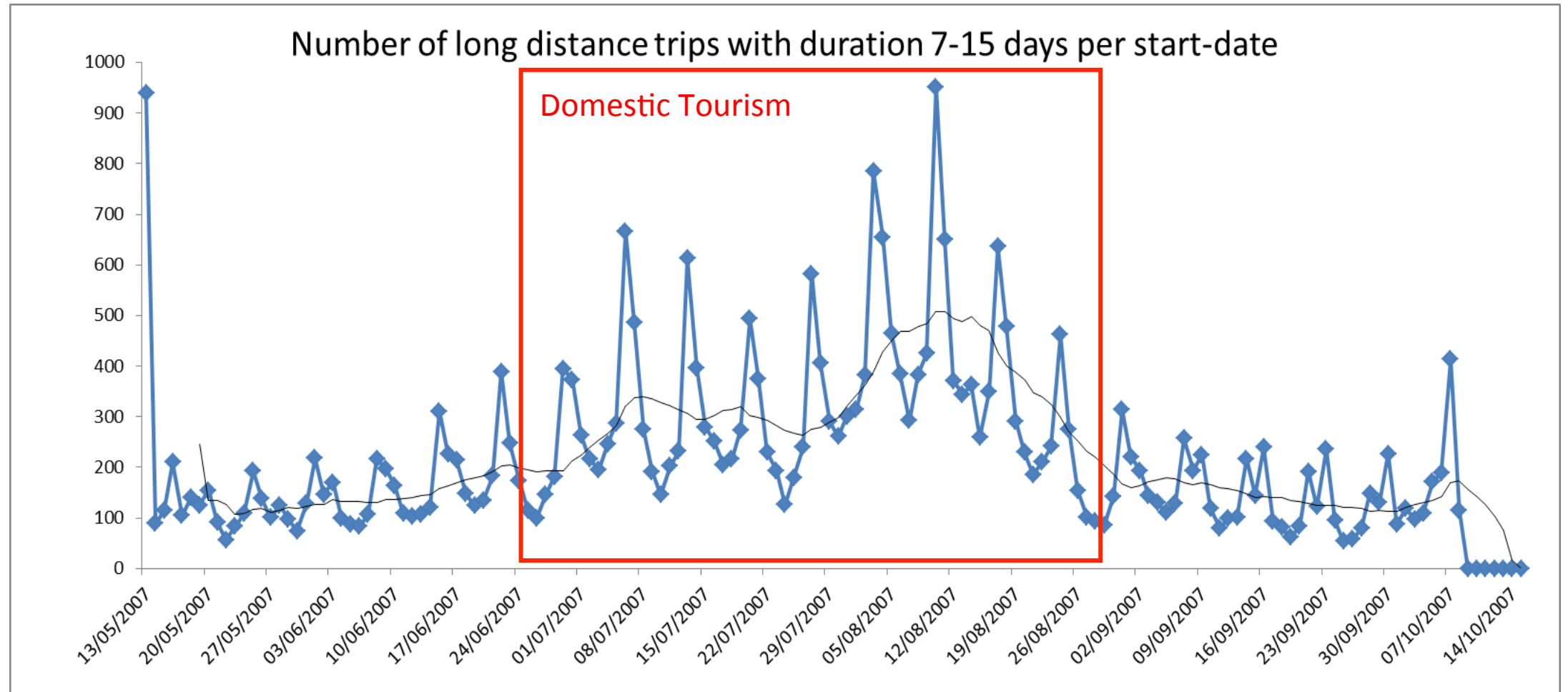


Simple heuristics for defining domestic tourism trips

1. Duration between 7 and 15 days
2. Limitation of start-date to July and August

| | Professional purpose (in %) | Personal purpose (in %) |
|-----------|-----------------------------|-------------------------|
| January | 35.1% | 64.9% |
| February | 17.6% | 82.4% |
| March | 30.8% | 69.2% |
| April | 18.1% | 81.9% |
| May | 18.7% | 81.3% |
| June | 20.7% | 79.3% |
| July | 12.5% | 87.5% |
| August | 5.6% | 94.4% |
| September | 20.7% | 79.3% |
| October | 24.4% | 75.6% |
| November | 18.5% | 81.5% |
| December | 16.5% | 83.5% |

Limitation of start-date to July and August

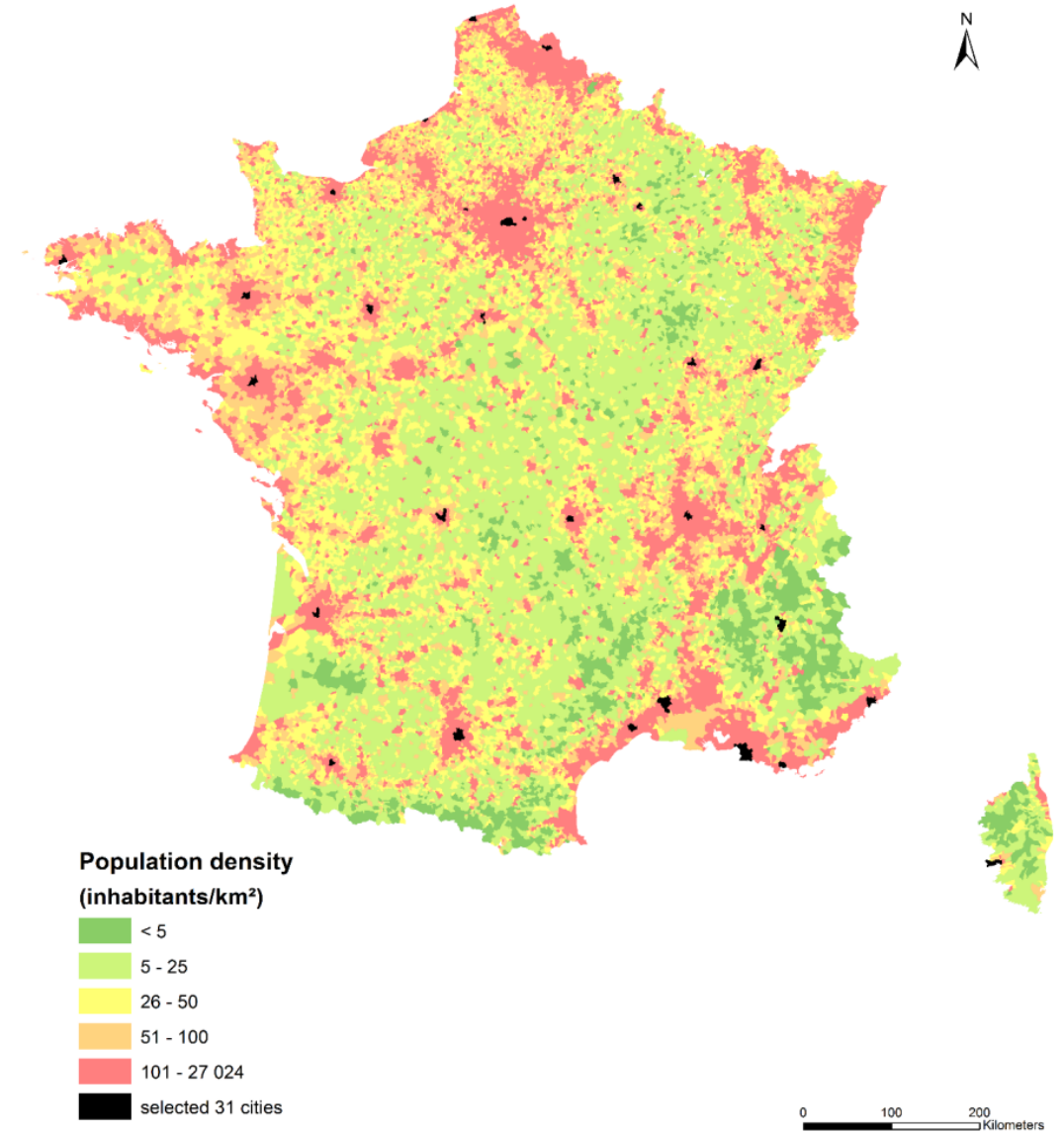


Results: Investigating domestic tourism trips

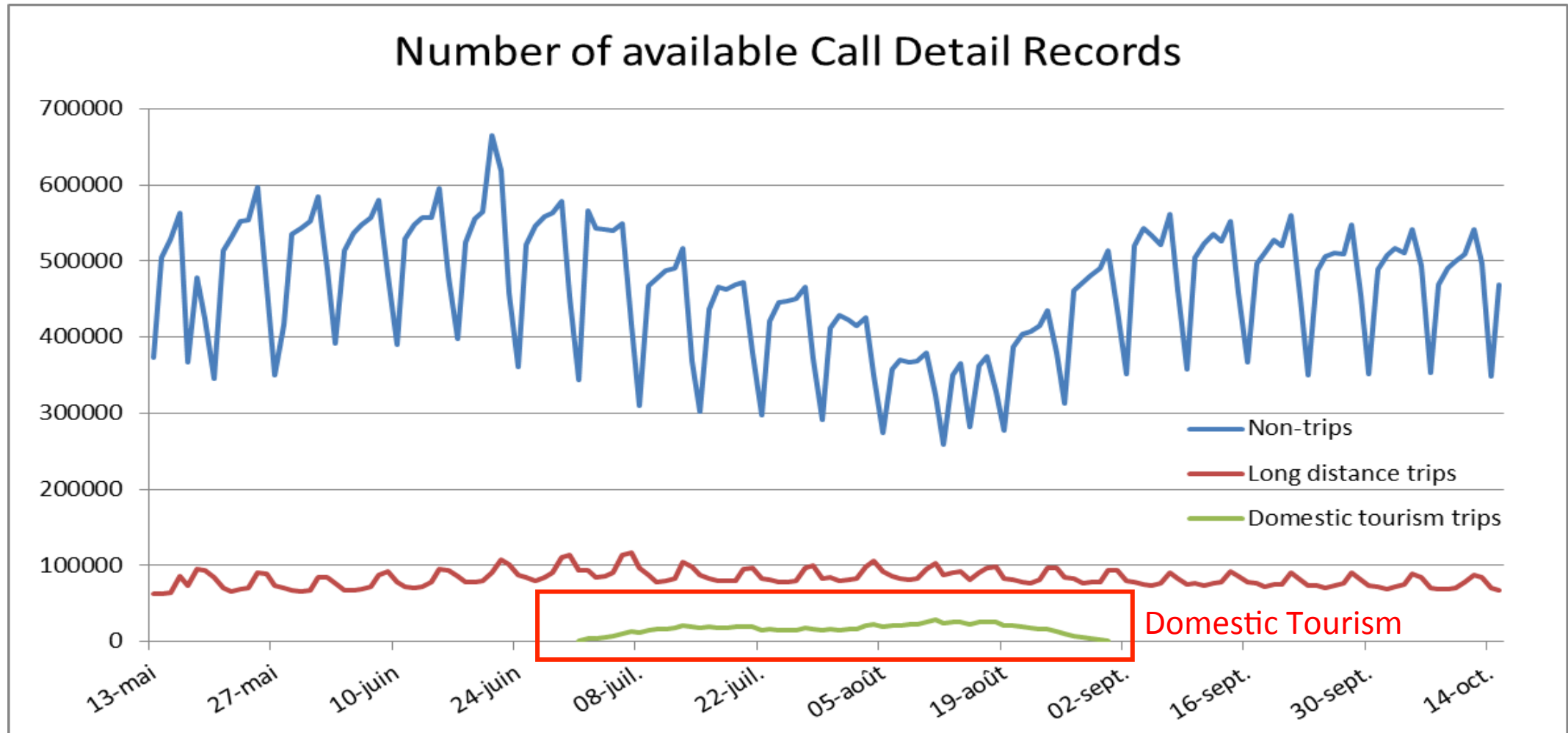
1. General properties
2. Temporal properties
3. Spatial properties
4. Social network characteristics

General properties

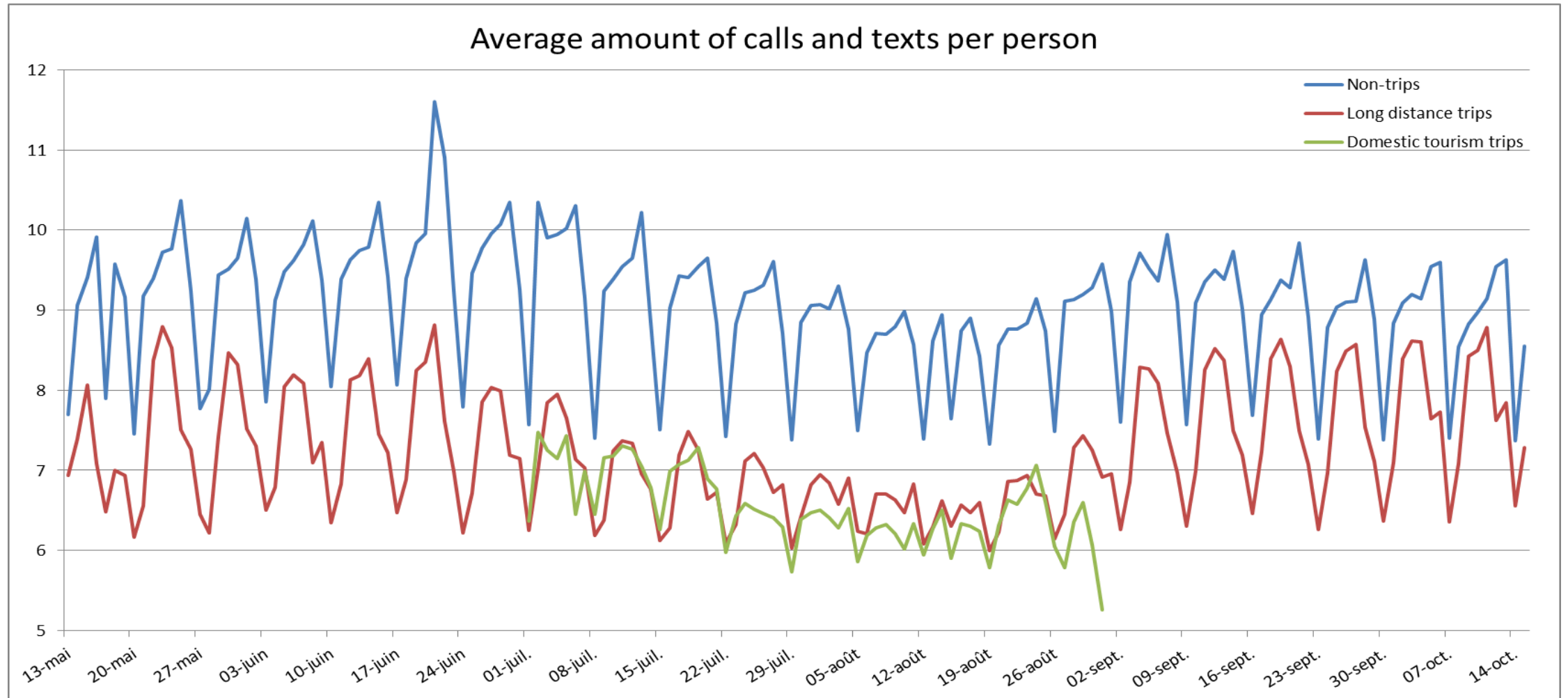
| Database | Users | Trips |
|--|--------|---------|
| Selection in 33 cities | 69 000 | / |
| Long-distance trips | 65 759 | 491 226 |
| Duration of 7-15 days | 25 488 | 34 633 |
| Starting in July or August = Domestic Tourism Trips | 16 024 | 18 380 |



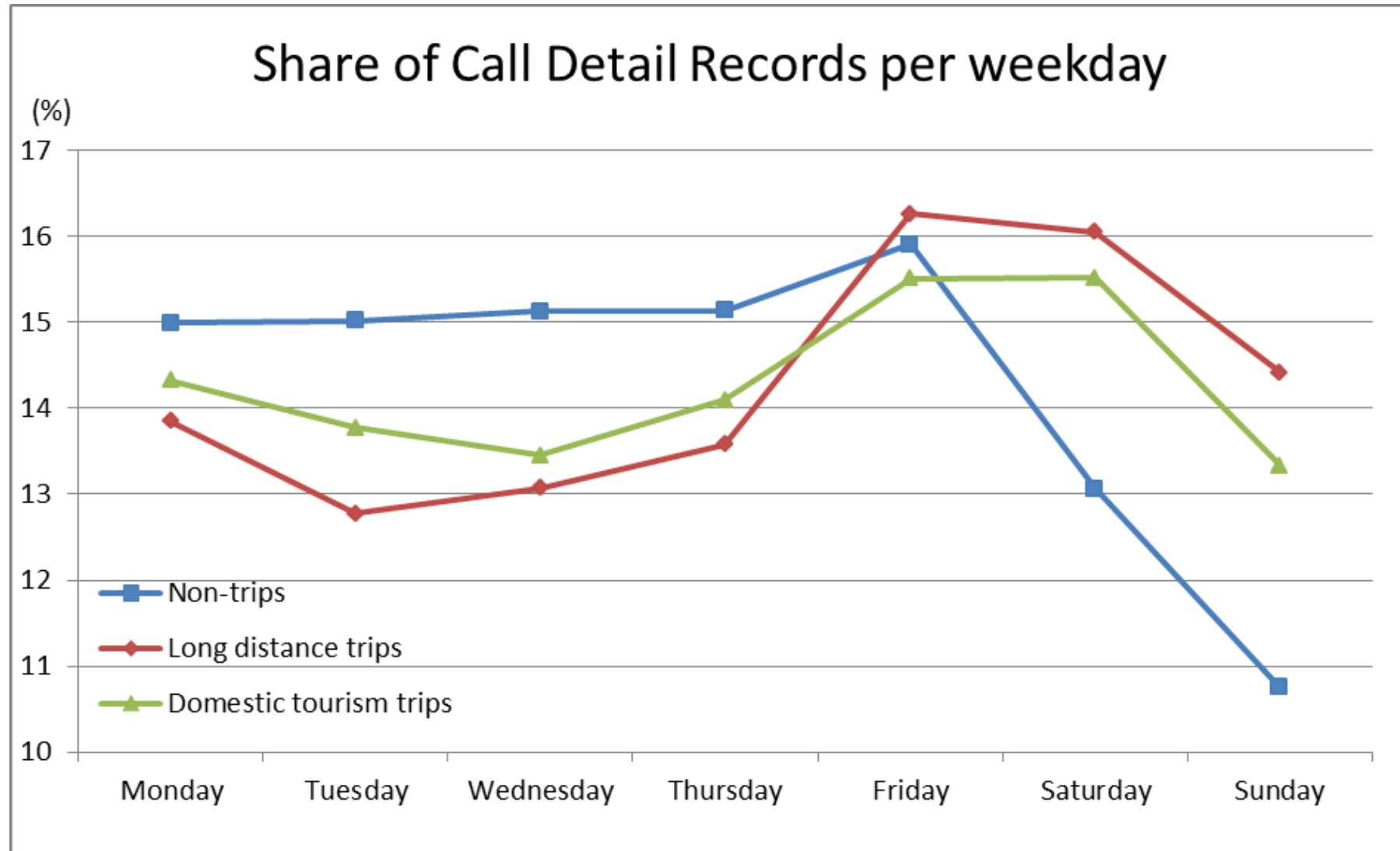
Temporal properties



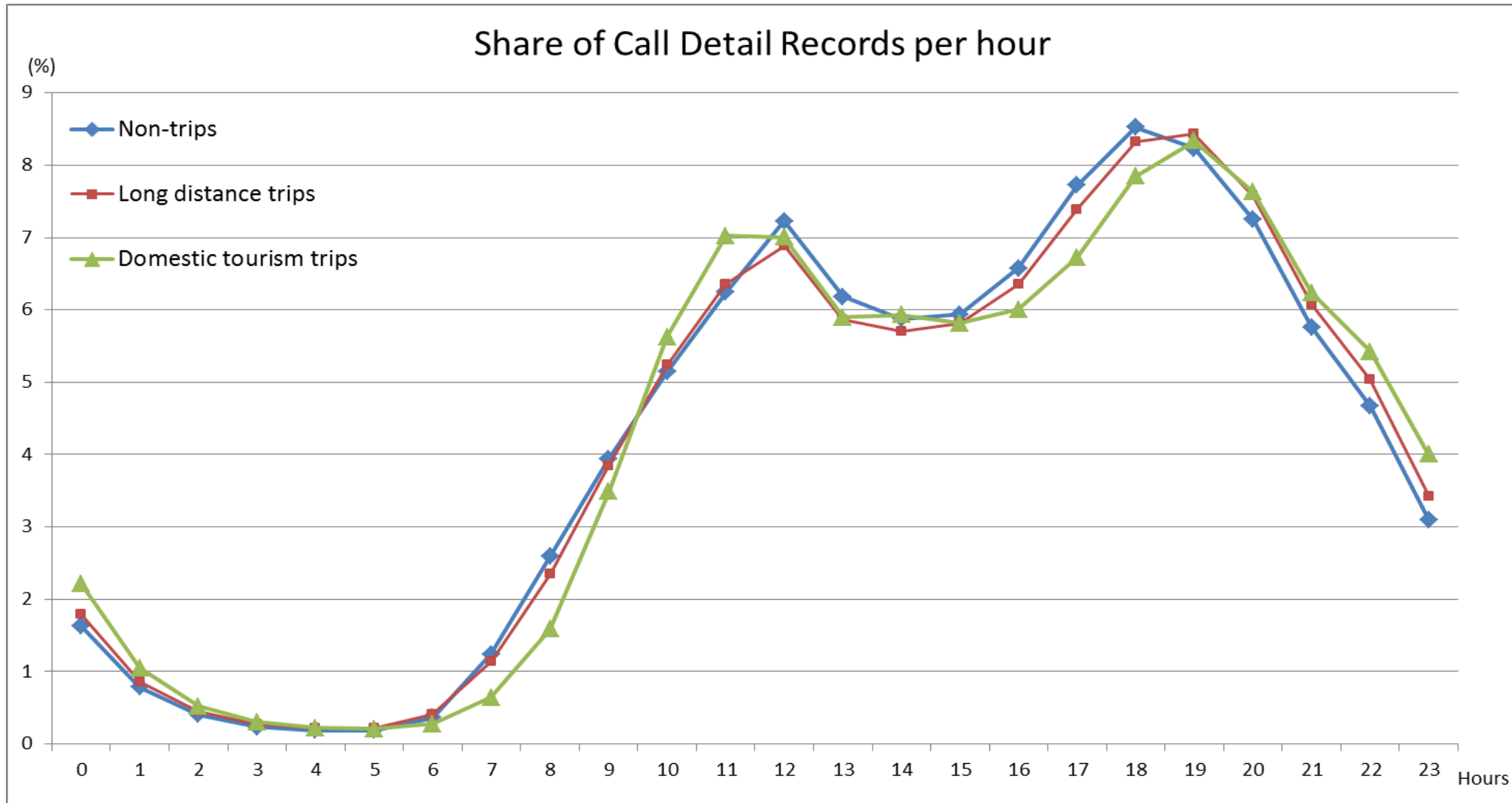
Do people call more during DDT?



Do people call at different days during DTT?



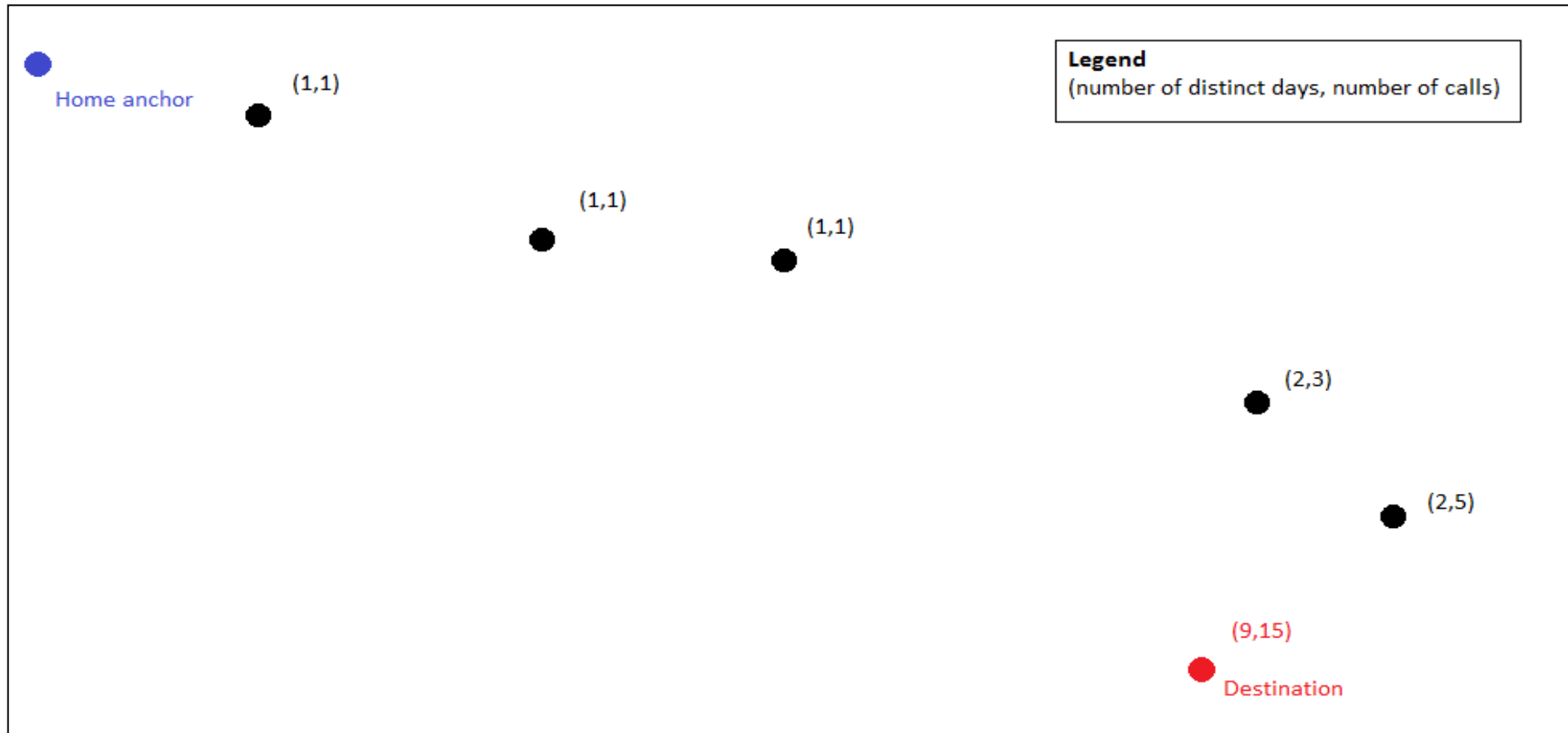
Do people call at different hours during DDT?



Spatial properties

1. Destination detection by DCR algorithm
2. Expert evaluation
3. Spatial patterns of domestic tourism in France

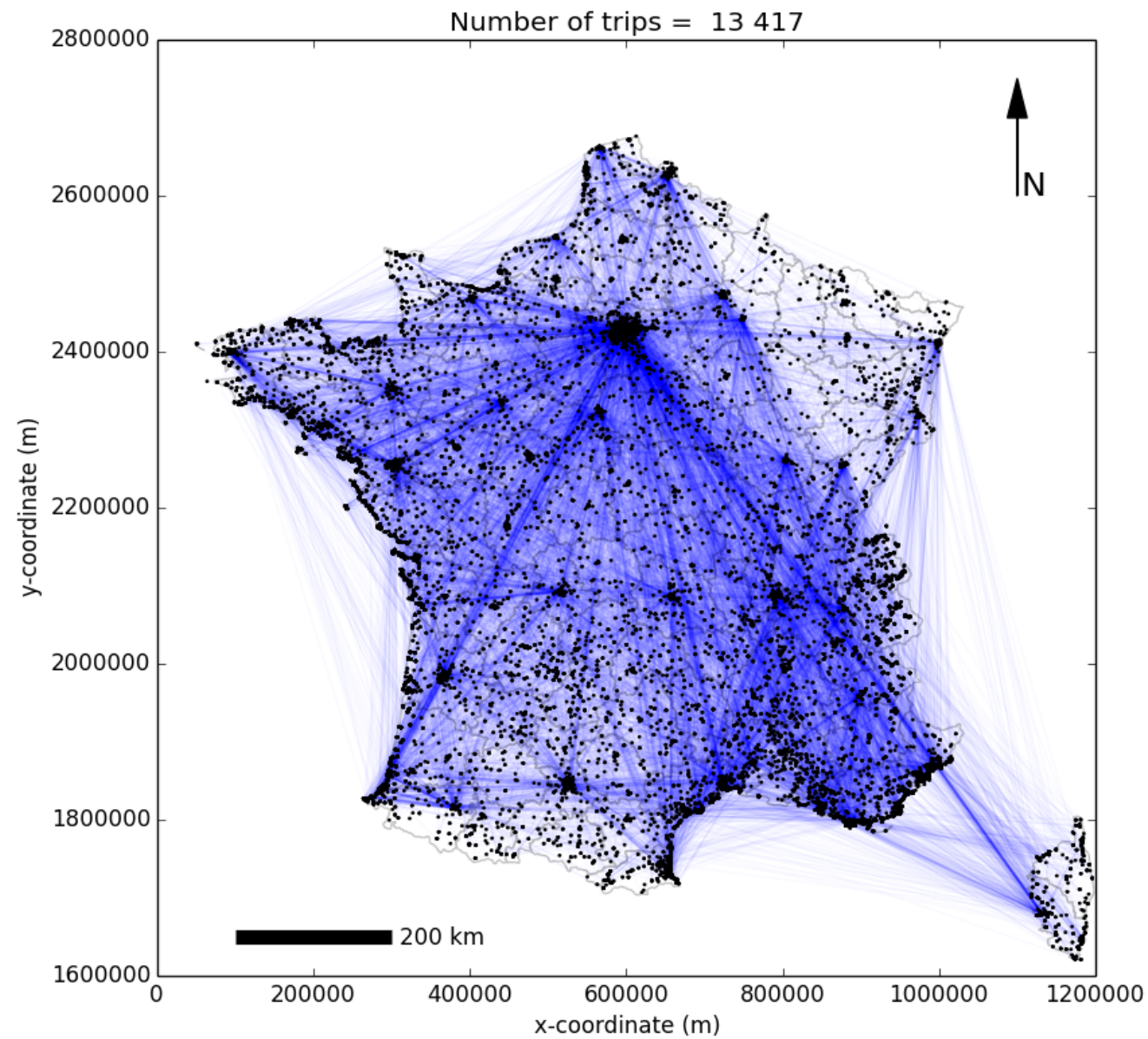
Destination detection of domestic tourism trips

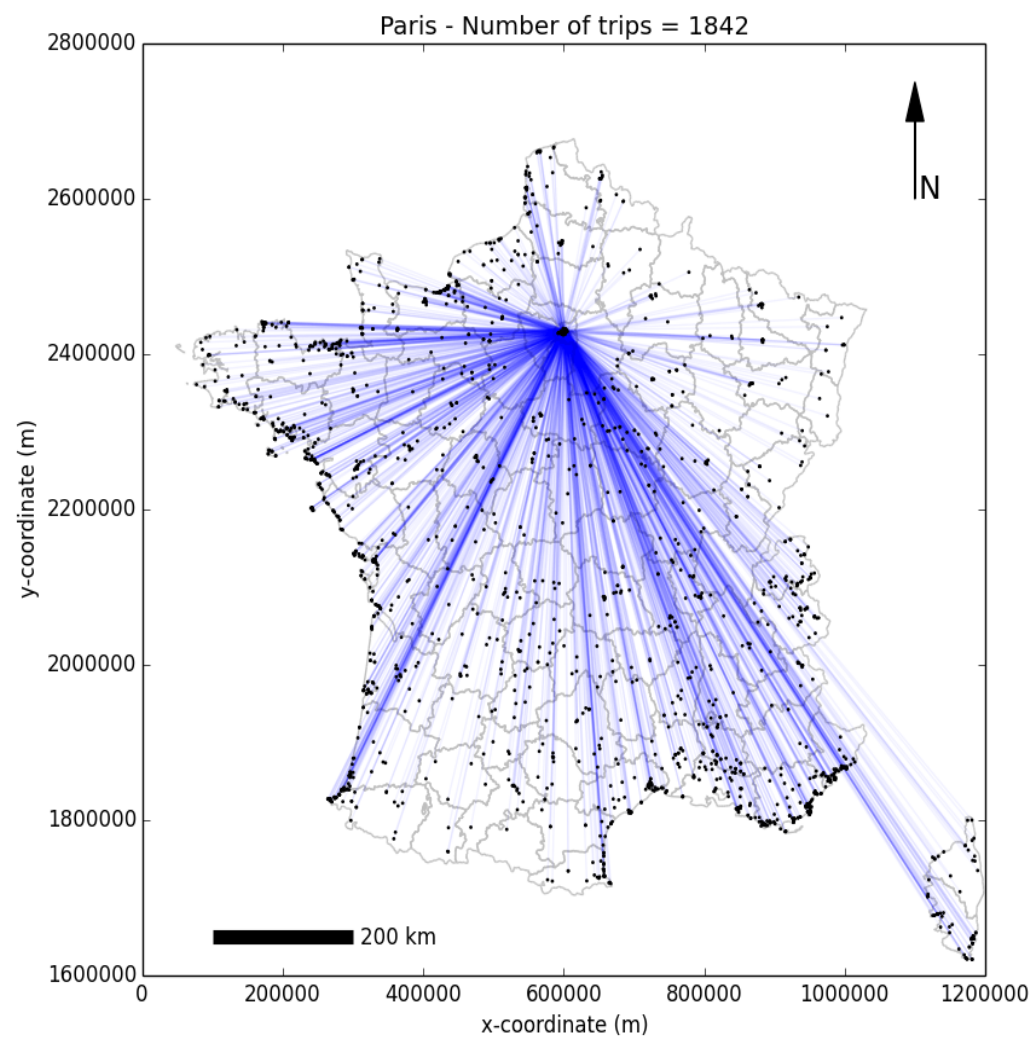
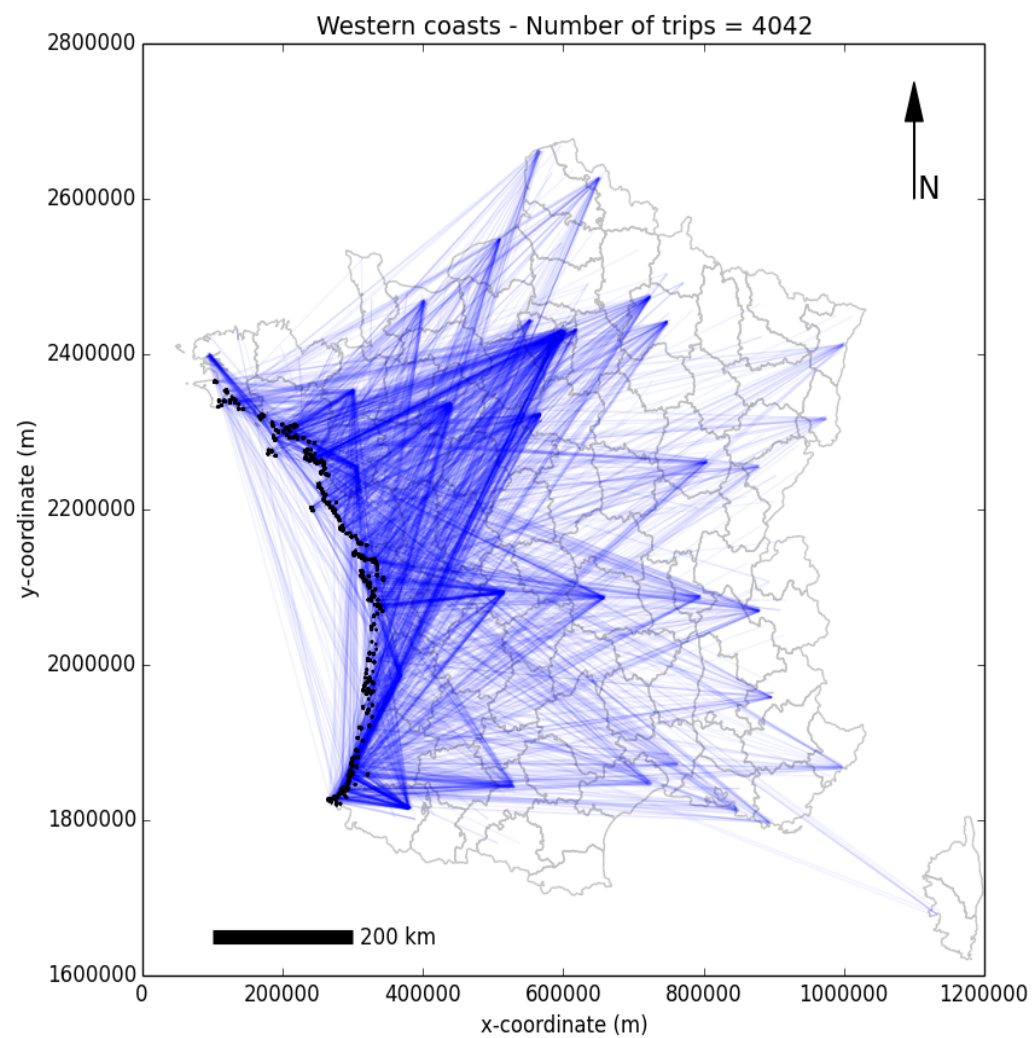


Expert validation

| | | IF DESTINATIONS ARE FOUND | | | IF NO DESTINATIONS ARE FOUND | |
|---------------------|-----------------------|---------------------------|---------|-----------|------------------------------|---------|
| distinct days ratio | number of calls ratio | % correct | % wrong | % dubious | % correct | % wrong |
| 0.3 | 0.1 | 89.0% | 4.9% | 6.1% | 94.4% | 5.6% |
| | 0.2 | 86.2% | 6.2% | 7.7% | 82.9% | 17.1% |
| 0.4 | 0.1 | 95.7% | 1.4% | 2.9% | 64.5% | 35.5% |
| | 0.2 | 86.6% | 10.4% | 3.0% | 63.6% | 36.4% |
| 0.5 | 0.1 | 88.0% | 10.0% | 2.0% | 66.0% | 34.0% |

Spatial patterns





Social network characteristics (one user)

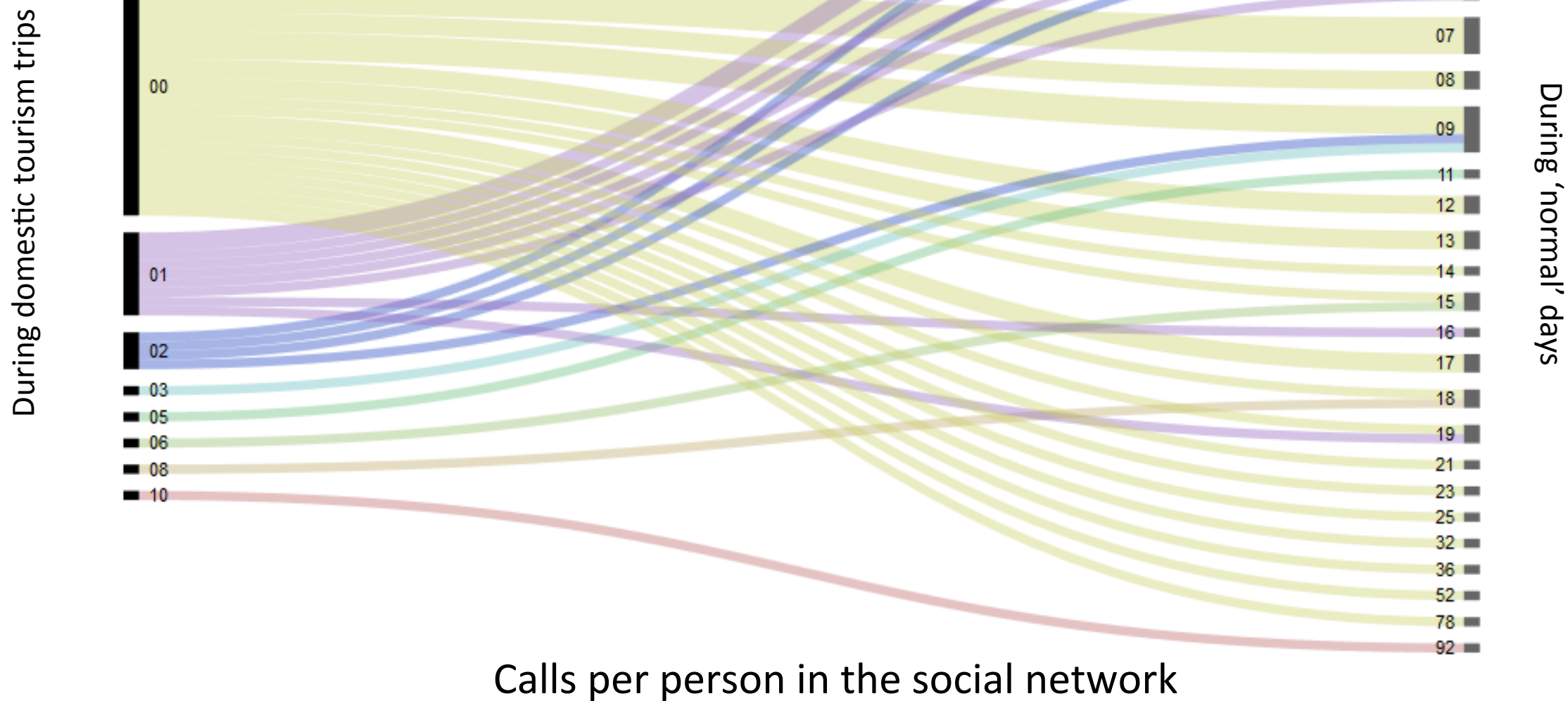
Describing social network activation during different phases of mobility

- By means of standard network measures
- By investigating the changing social patterns

Standard network measures

| | Trip id | Days | Activities | Degree | Entropy |
|-------------------------------|---------|------|------------|--------|---------|
| Non-trips | / | 128 | 882 | 145 | 1.77 |
| Long-distance trips | 1 | 1 | 5 | 3 | 0.46 |
| | 2 | 4 | 22 | 10 | 0.87 |
| | 3 | 4 | 13 | 6 | 0.73 |
| | 4 | 3 | 13 | 7 | 0.77 |
| | 5 | 4 | 7 | 4 | 0.55 |
| Domestic tourism trips | 1 | 10 | 50 | 19 | 1.11 |

Changing calling patterns



Conclusion

- Extracting long distance trips from CDR data seems plausible, even shows underestimation by surveys
- Simple heuristics can be applied to long distance trips to define 'domestic tourism trips'
- The detected domestic tourism trips are not exhaustive but:
 - Temporal properties are plausible (higher amount of DDT's during summer) and call behaviour differs from baseline behaviour (less activities on cellphone, different weekly and hourly patterns of calls)
 - Spatial patterns of DDT can be discovered by means of a simple destination detection algorithm parameterized by expert validation. Spatial patterns seem plausible with coast and mountainous areas as main destination, and cities a clear second.
 - There is potential to study changing social behaviour during different phases of mobility. This was showcased by means of analysis for one person, but needs more work to scale up to larger populations.

Thank you,

The end.

M.vanhoof1@newcastle.ac.uk

@Metti Hoof

MaartenVanhoof.com