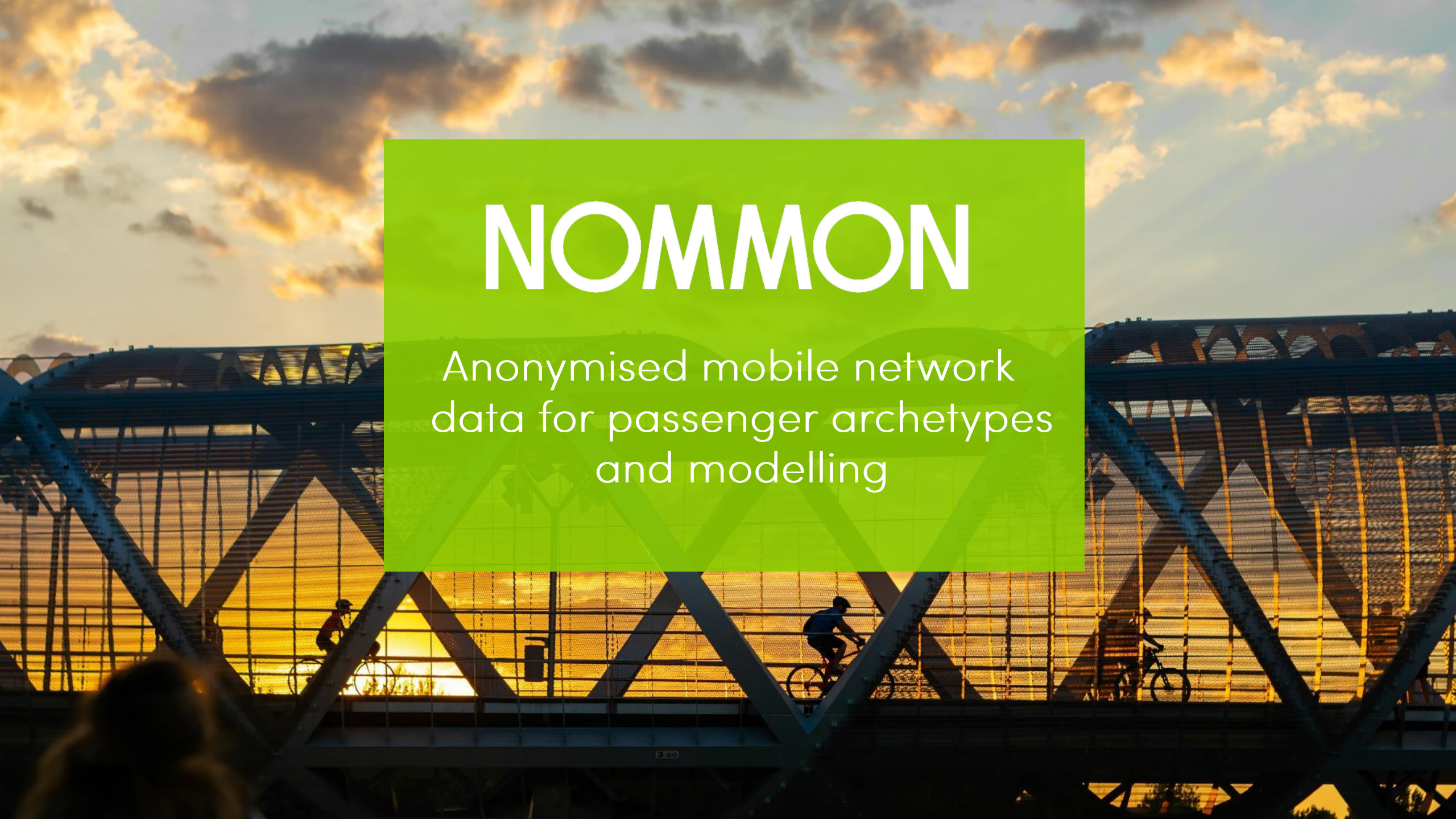

NOMMON
Anonymised mobile network data for passenger archetypes and modelling

1 CONTEXT: CHALLENGES AND OPPORTUNITIES

2 NEW DATA SOURCES

3 MOBILE NETWORK DATA

4 CASE STUDIES IN AVIATION

1. Context: Challenges and Opportunities

Context: evolution of mobility

Advances in technology are generating radical changes in mobility patterns.

Demand
- Remote work
- E-commerce

Supply
- Shared mobility
- Demand-responsive transportation (DRT)
- Mobility as a service (MaaS)
- Connected and autonomous vehicles (CAVs)
- Urban air mobility (UAM/IAM)

- Flattening of the demand curve
- Less forced mobility
- More recreational travel
- Change in distances traveled

- Less dependence on private vehicles
- More flexibility
- More multimodal

Other external factors that have also affected mobility patterns:
- COVID
- Demand management policies
- Environmental awareness

Demand data: areas of application

Why do we want this data?

Urban mobility plans

Regional transport plans

Concession evaluation

Traffic studies

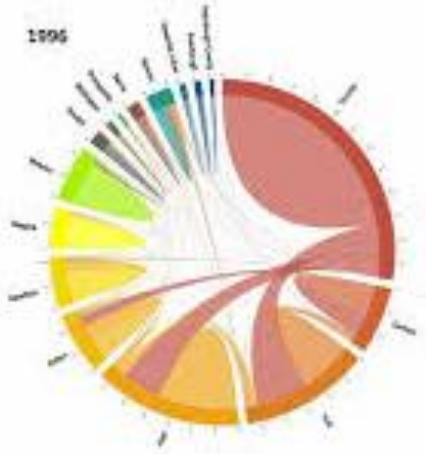
Competitor analysis

Accessibility and catchment areas

Mobility on specific days

What information are we interested in?

- Number of trips per person
- Generation/attraction of trips by area
- Distribution of trips (OD matrices)
- Mode of transport
- Trip route
- Trip purpose (work, leisure, medical appointment, etc.)
- Time of trip
- Socio-demographic profile (age, gender, place of residence, etc.)
- Trip times/speed
- Vehicle capacity
- Access/egress for multimodal trips
- ...

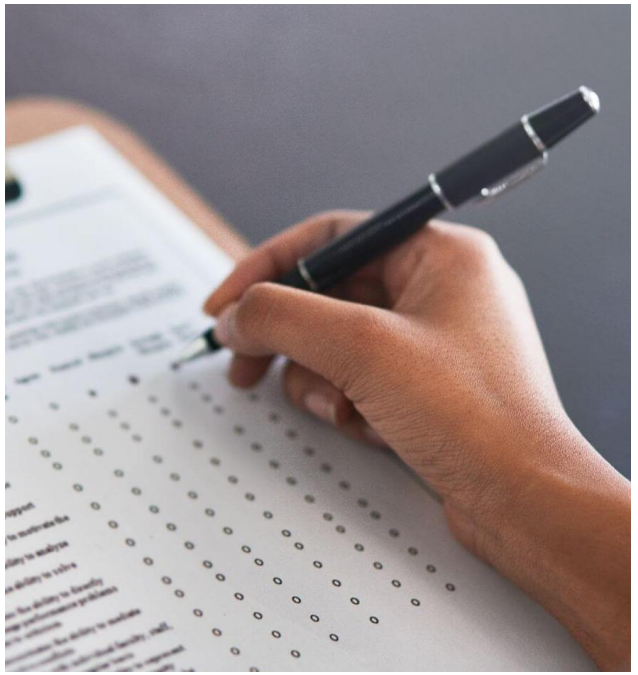


Transport surveys

- **Transport surveys** are the traditional method of gathering information on transport demand. They have several **advantages** but also **disadvantages**:

☑ Detailed information, sociodemographic characteristics, stated preferences

☒ Long execution times, costs, need for updating, response bias, sample size, need for planning (unexpected events)



Types of surveys

- Household surveys
  - Detailed information: mobility + socio-demographic profile
  - €1–4 million // duration: 1–3 years
- Intercept surveys
  - Specific points in the study area: airports
  - Information on non-residents + supplementary household survey
  - €5,000/point // timeframe: weeks
- Other types
  - Capacity: volume/type of vehicles (coils, tyres, manual, etc.)
  - Travel times: floating vehicle (e.g. place GPS in vehicle)

Context: technological developments

- "New" data, tools, and technologies:
  - Proliferation of mobile devices
  - Connected vehicles
  - New and cheaper sensors
  - Lower data processing costs
  - Big Data technologies
  - Cloud computing
  - Open data
  - Artificial intelligence
  - Image recognition

New data sources and mobility: trends

BEFORE		NOW
Complementary information	→	**Baseline** information
Specific analyses	→	**Continuous** monitoring
Independent analysis of different sources	→	Data **fusion**
Input for **traditional models**	→	**New models**

Demand information

While information on high-quality transport supply is increasingly available, **travel demand data** is, in most cases, largely **unknown**.

This leads to many planning and management **decisions** being made on the basis of **outdated/incomplete information**, at a high cost to society.

How can we leverage these new technologies and data sources to address the challenges in urban mobility planning, management, and operation?

2. New data sources

Transport demand data sources

Mobile devices

Fixed sensors

Transport demand data sources

Sensors
Bluetooth-WiFi

Cameras – OCR
(Optical Character
Recognition)

Coils

Mobile devices

Fixed sensors

Transport demand data sources

Speeds and trip times
Local OD matrices

Vehicle flow
Local OD matrices

Vehicle flow
Classified traffic counts

Mobile devices

Sensors
Bluetooth-WiFi

Cameras – OCR
(Optical Character
Recognition)

Coils

Fixed sensors

Transport demand data sources

Driving navigators

Smart transport card

Mobile applications

Mobile network

Mobile devices

Fixed sensors

Transport demand data sources

Driving navigators 

Road maps
Speeds and trip times

Smart transport card 

Demand for public transport

Mobile applications 

Door-to-door trips
High spatial resolution

Mobile network 

Door-to-door trips
Representativeness and sample size

Mobile devices

Fixed sensors

Transport demand data sources

Different data sources provide relevant information on different aspects of mobility...

...but none of them provides a complete picture.

So... What demand information do we use?

Optimal combination of data sources

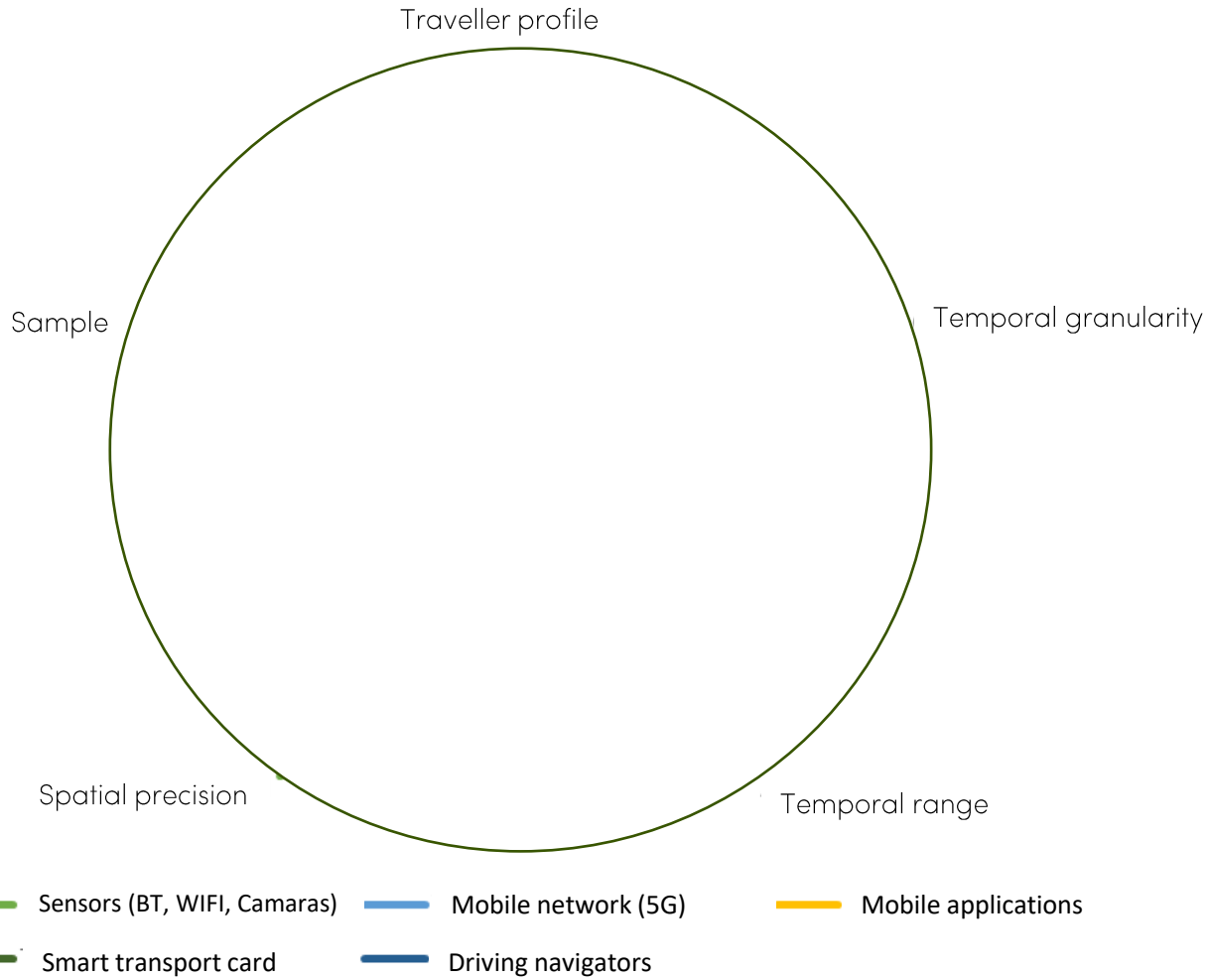
- How can we obtain this information?: optimal combination of data sources according to case study

- Data storage and processing
- Debugging and troubleshooting
- Transport engineering knowledge
- Methodology and algorithms for data analysis and fusion
- Statistical analysis and sample elevation
- Adaptation of the analysis to specific problems

Metropolitan mobility | Interurban mobility | Accessibility studies | Concession evaluation | ... Others

Comparison of data sources

What does each data source provide me with?



3. Mobile Network data

Mobile phone data: what is it?

- By mobile phone data, we refer to data collected by mobile operators (Orange, Vodafone, etc.) for billing and network management purposes.

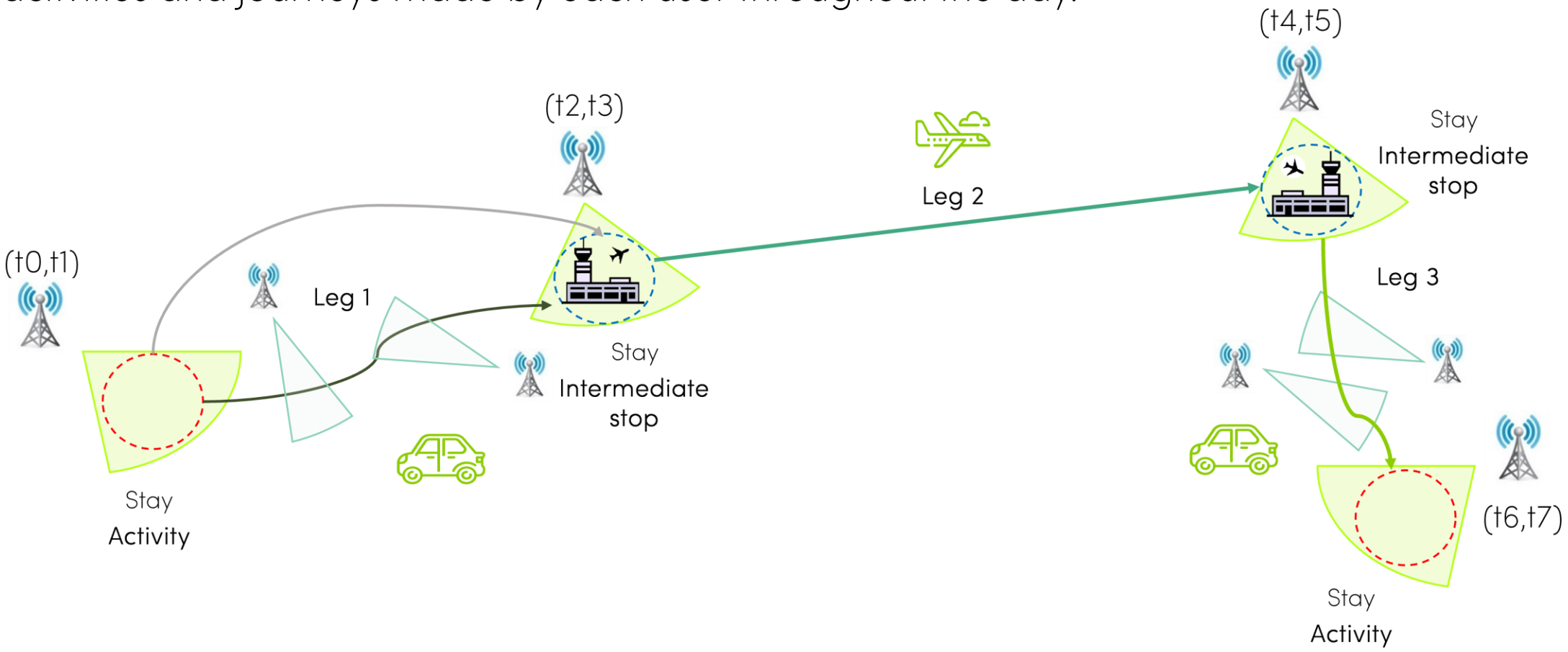
Mobile phone events
Activity logs between device and antena
(Call Detail Records (CDRs)/probes)

Mobile telephone network
location and characteristics of the antennas

Sociodemographic information
Age, gender, nationality...

Mobile phone data: what is it?

When properly processed, telephone data can be used to reconstruct the sequence of activities and journeys made by each user throughout the day.



Mobile phone data: advantages

- Like other sources of digital data (e.g., mobile applications, driving navigators, smart transport cards), mobile phone data enables:
  - Passive collection of information
  - Up-to-date information
  - Longitudinal data
  - Reduction in timeframes and costs

- However, telephone data has other characteristics that make it particularly valuable for studying mobility:
  - Large, well-distributed samples
  - Include trips by all modes
  - Observation of door-to-door trips

Mobile phone data: opportunities

- Higher quality OD matrices thanks to larger sample size
- Regular/continuous updating of information
- Analysis and modelling of more types of days and periods of the year and ability to observe demand variability and evolution
- Monitoring of pilot tests and natural experiments

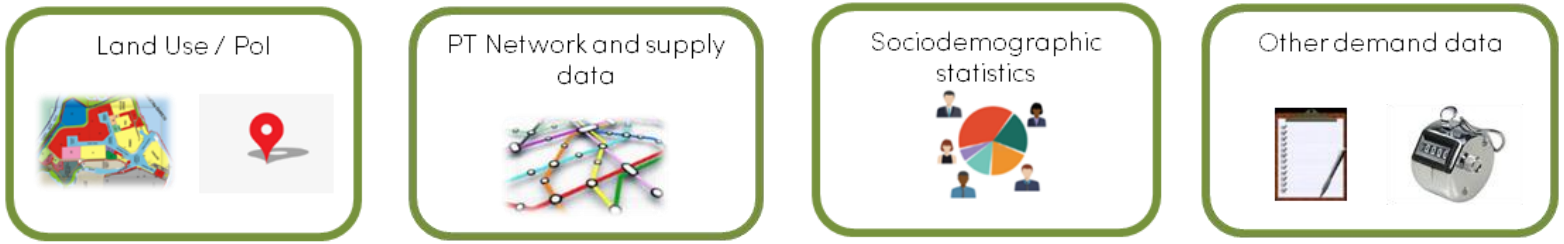
Mobile phone data: limitations

- Spatial resolution:
  - Inability to detect short trips
  - Uncertainty about the exact location of activities
  - Uncertainty about the exact start and end times of journeys
- They do not capture important information:
  - User profile (partially)
  - Trip purpose (partially)
  - Group travel/number of users in the same vehicle
  - Engine type and driving license
  - Subjective information
- Access, processing and data volume

Other external data

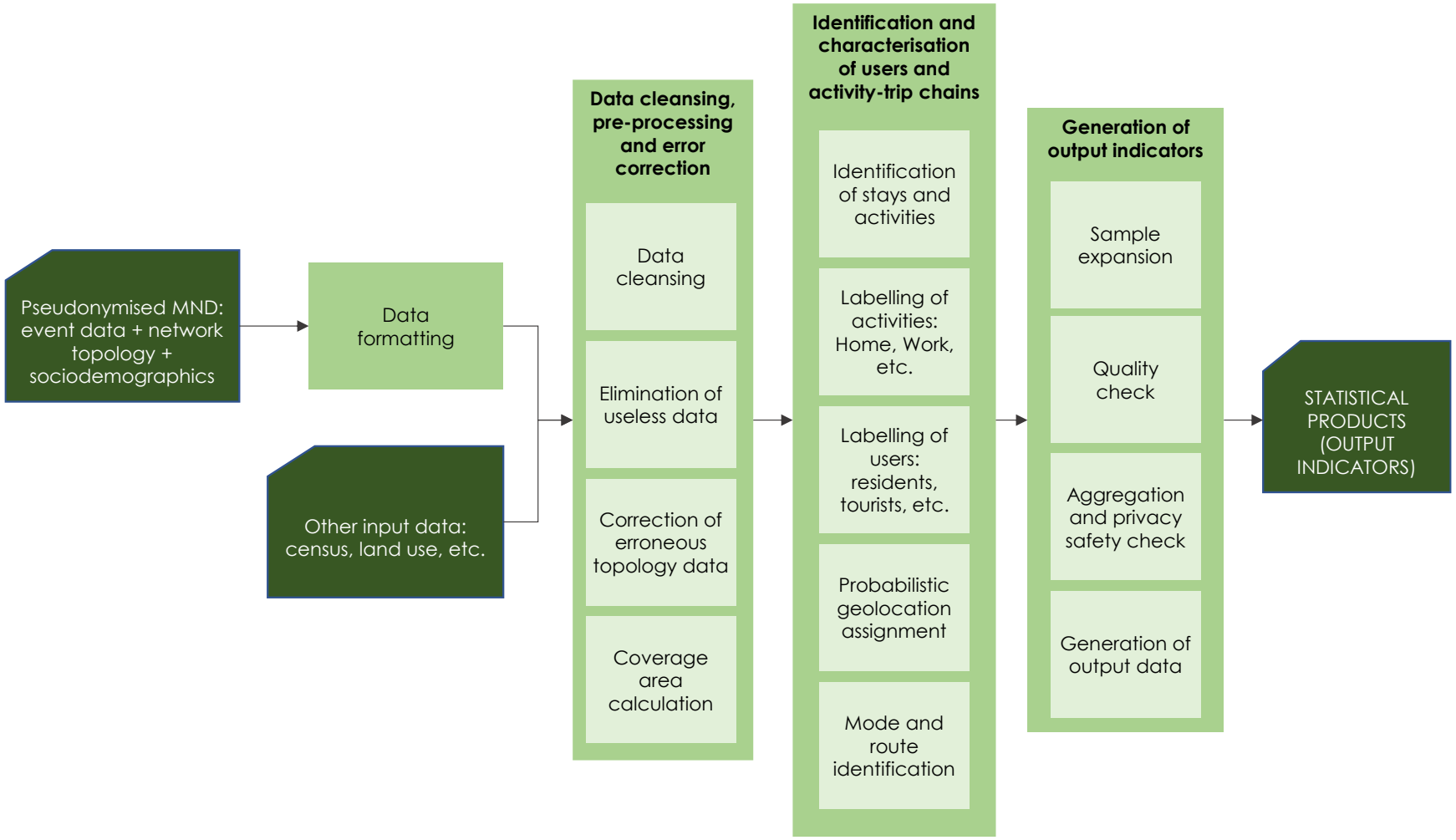
Main external data sources for extracting mobility information from mobile phone data :

- Land use and points of interest
- Transport network and services offered
- Sociodemographic statistics (e.g., census)
- Transport demand data: ticketing, traffic counts, etc.



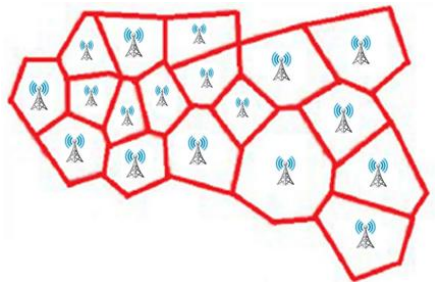
Process of generating mobility information

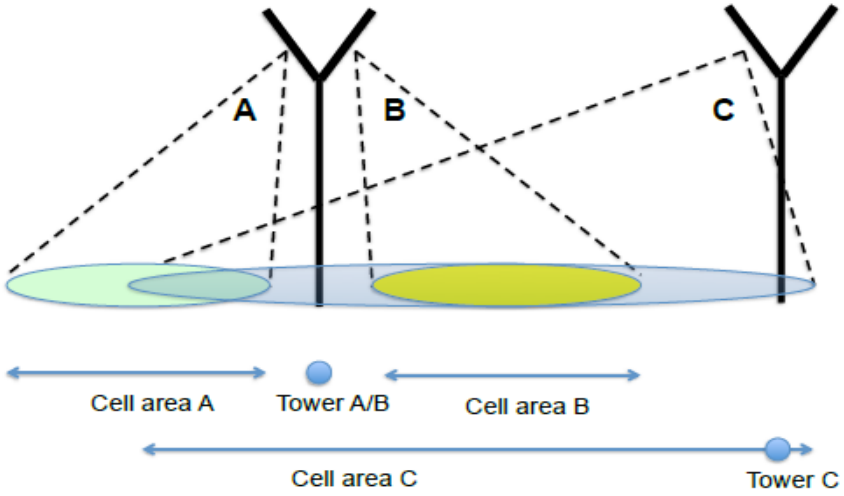
- There is no standard methodology: the importance of validation



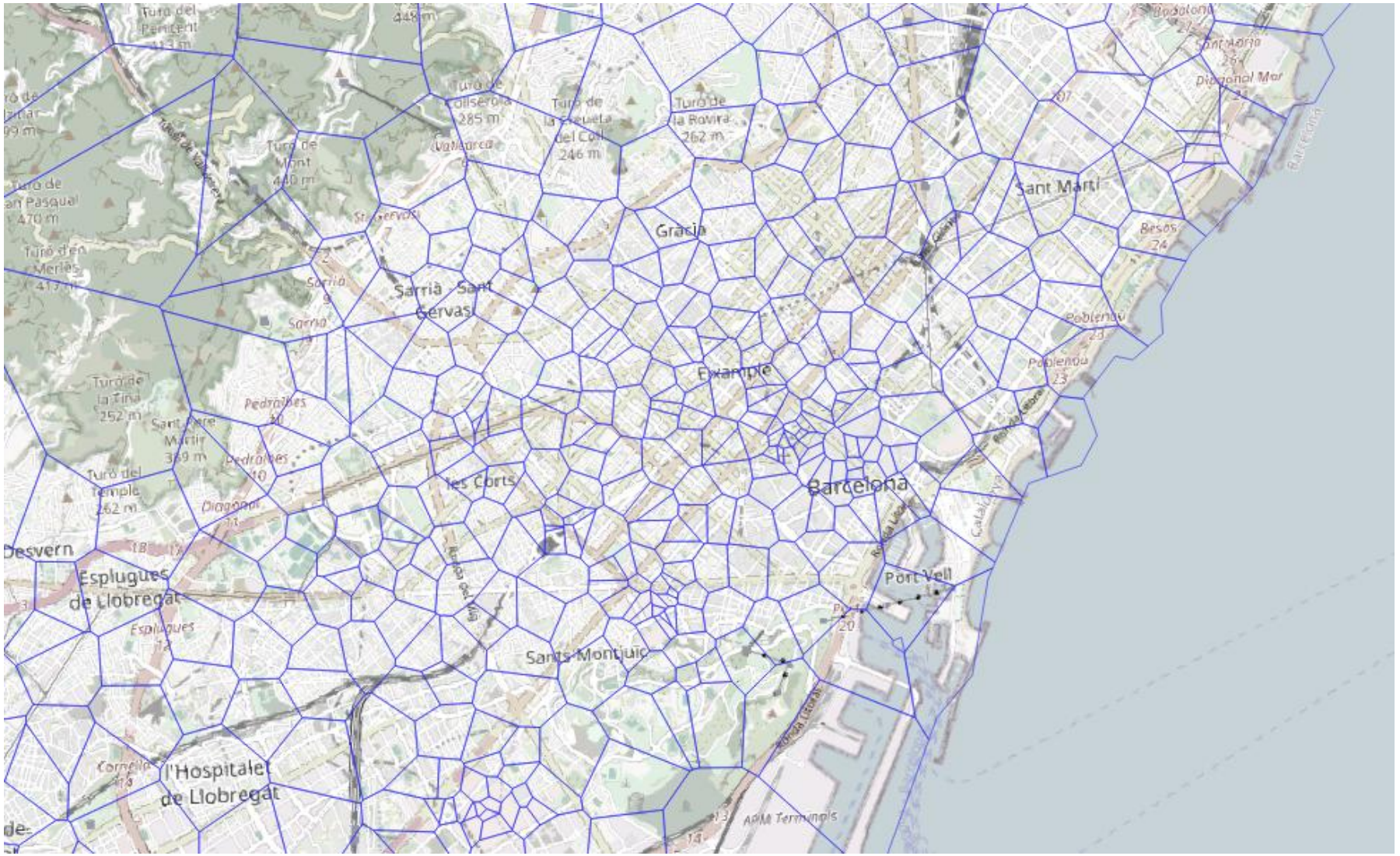
Standardisation: mobile telephone network

- Variable density of mobile phone antennas (from 50-100 metres between antennas in urban areas to kilometres in rural areas)
- Estimation of coverage areas
- Voronoi tessellation/Thyssen polygons



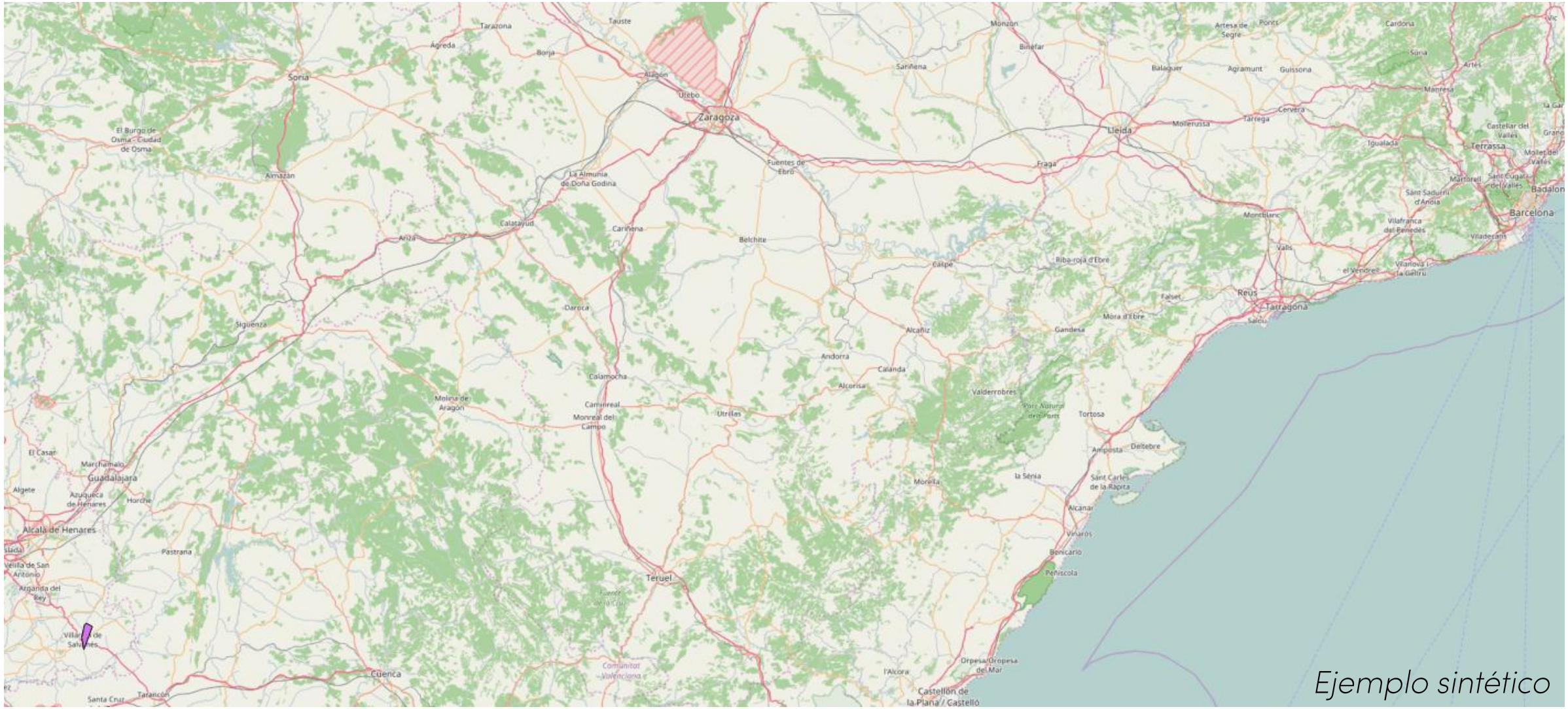


Standardisation: spatial precision

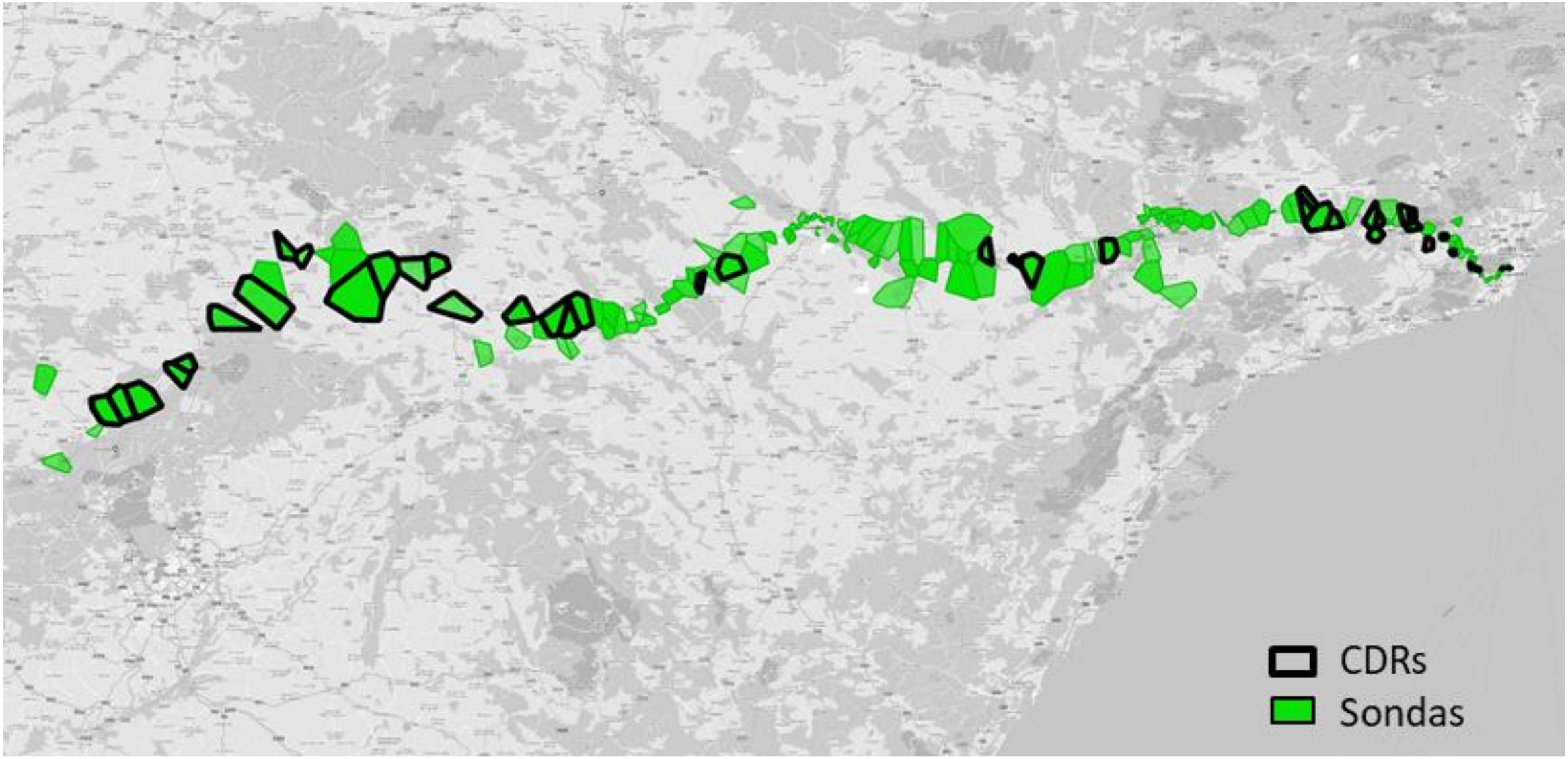


NOMMON

Standardisation: mobile phone events



Standardisation: CDRs vs probes



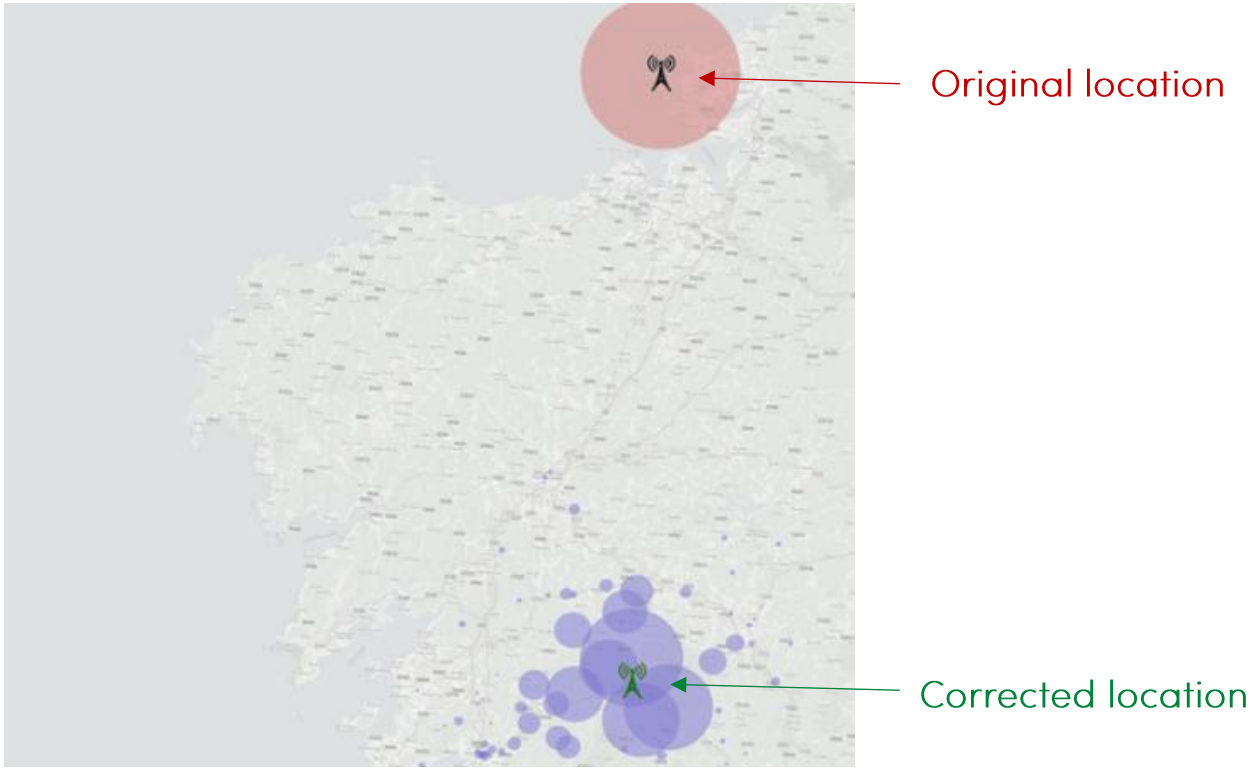
Cleaning, pre-processing, and error correction

⚠️ Very important!

Some examples:

- Signal jumps
- Incorrect times
- Duplicate/erroneous events
- Record types
- Device types
- Incorrect antenna locations
- Valid sample selection

Example of a poorly positioned antenna

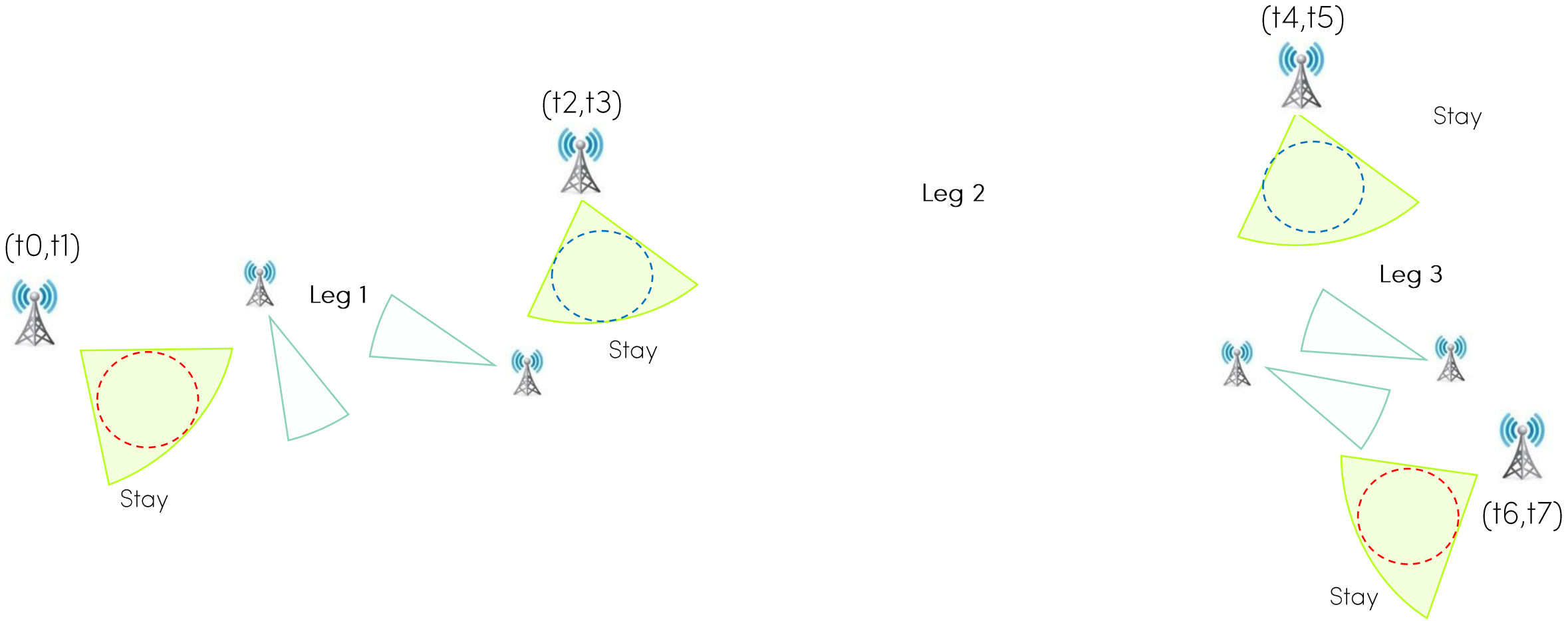


Reconstruction of activity and travel diaries

1. Detection of stays and movements: clustering of records
2. Identification and labelling of activities: longitudinal analysis
3. Location assignment: fusion with land use and points of interest
4. Identification of trips and characterisation of stages: map matching
5. Distinction between personal and professional mobility: longitudinal analysis
6. Classification/profiling of users

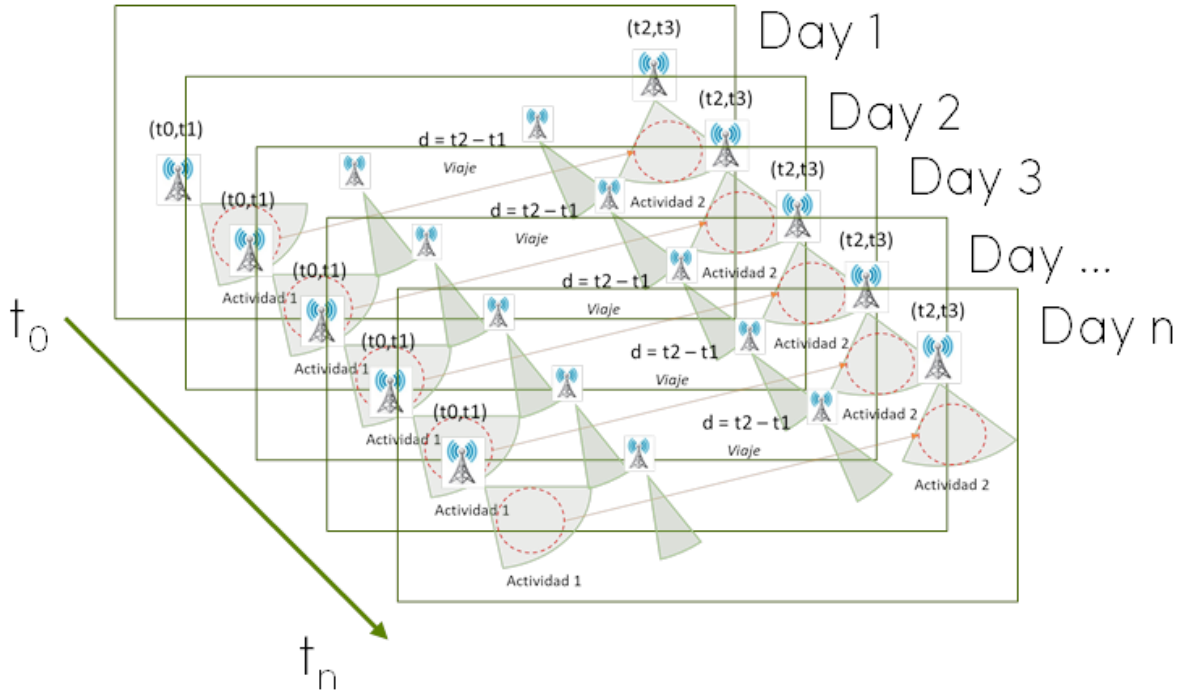
Reconstruction of activity and travel diaries

1. Detection of stays and movements: clustering of records



Reconstruction of activity and travel diagrams

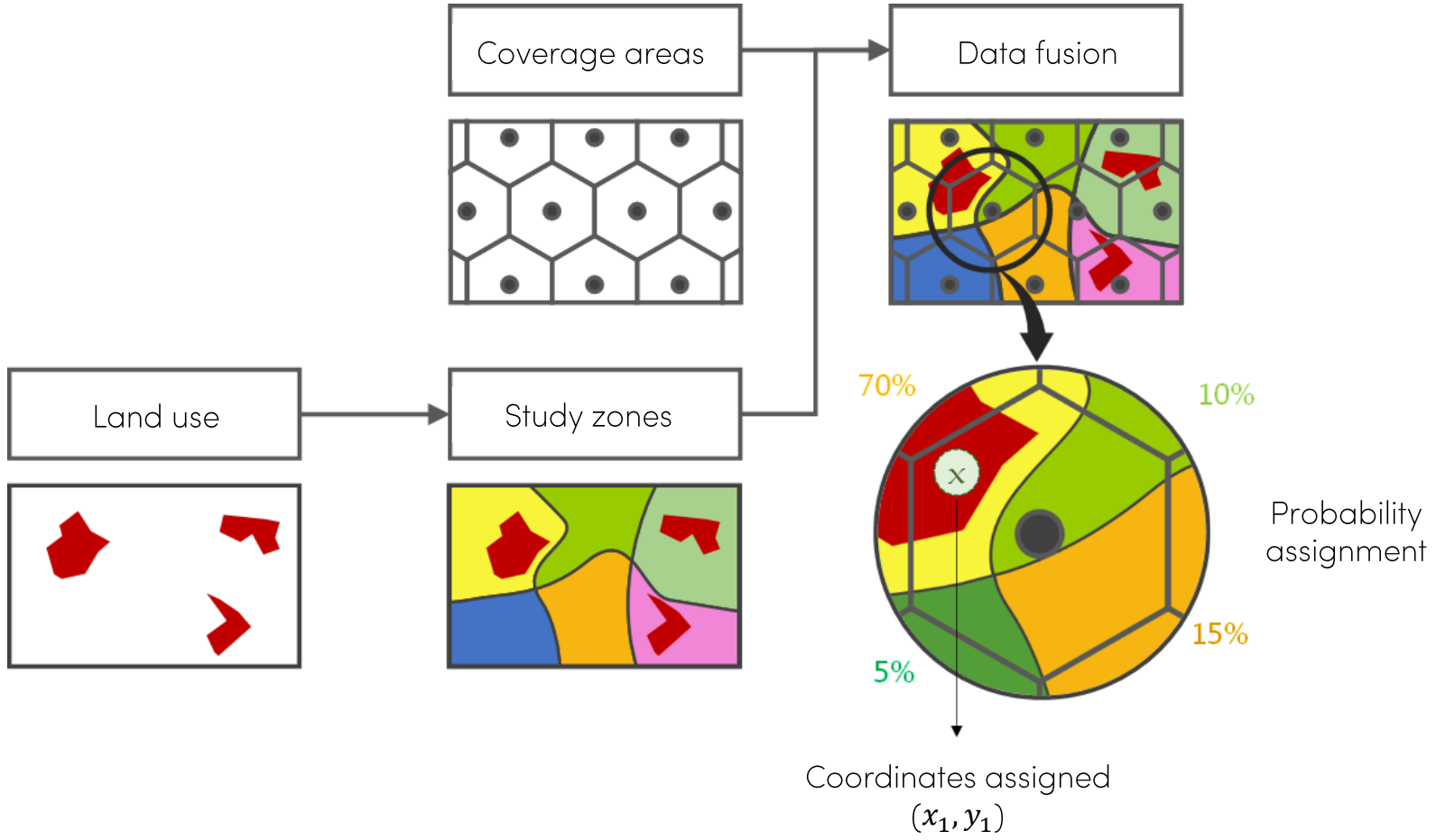
2. Identification and labelling of activities: longitudinal analysis



- Place of residence
- Place of work/study
- Second residence
- Other activities

Reconstruction of activity and travel diaries

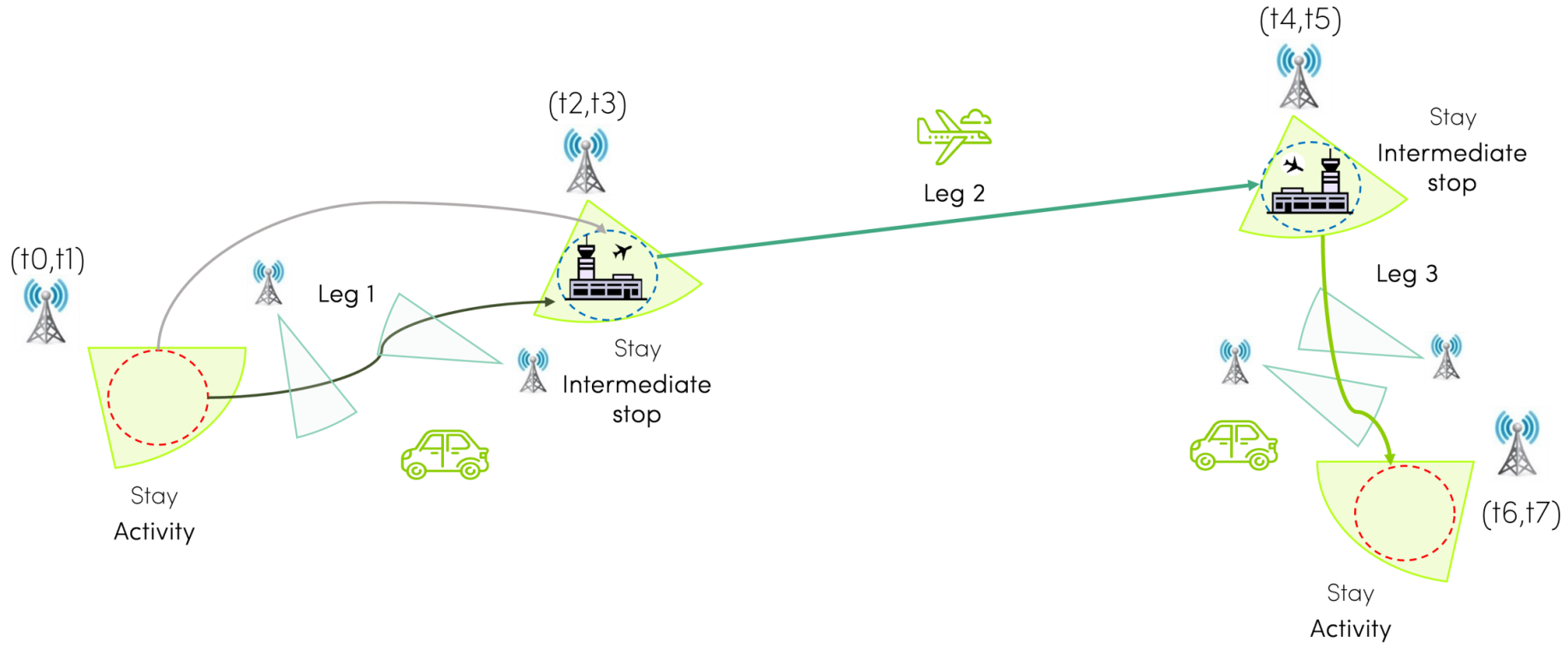
3. Activity location assignment: fusion with land uses and points of interest



Reconstruction of activity and travel diaries

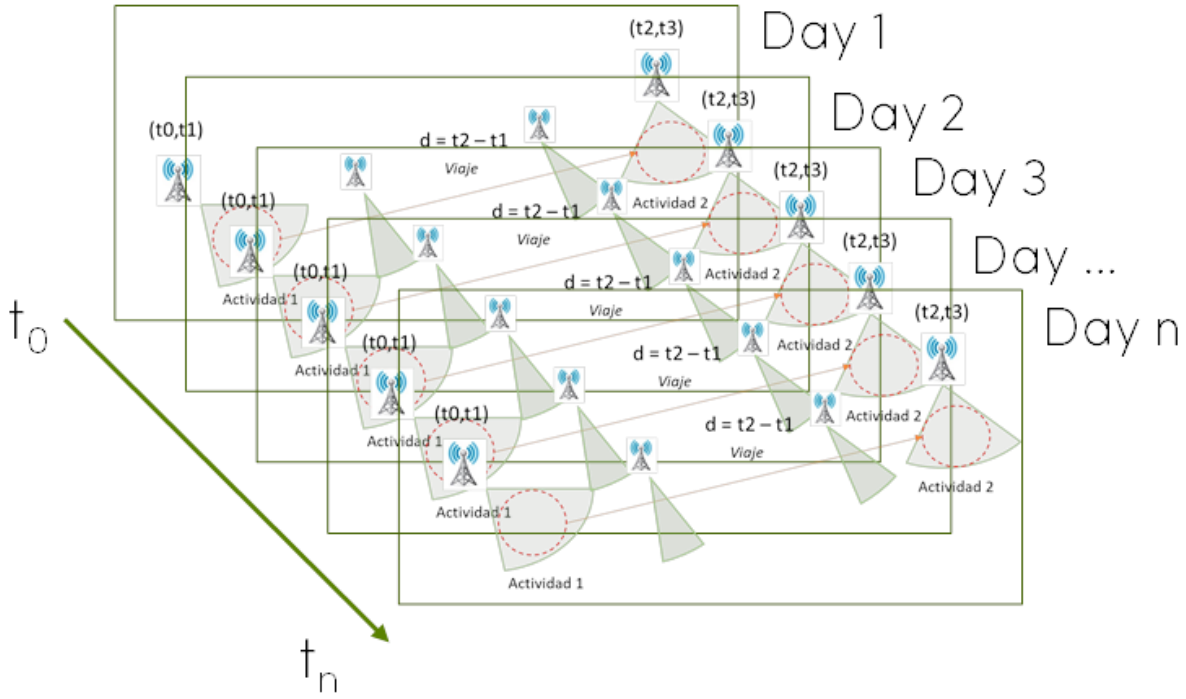
4. Identification of journeys and characterisation of stages: map matching

   - Difficulty in distinguishing between modes that generate similar records: short journeys in urban environments, private vehicle vs bus, etc.



Reconstruction of activity and travel diaries

5. Distinction between personal and professional mobility: longitudinal analysis

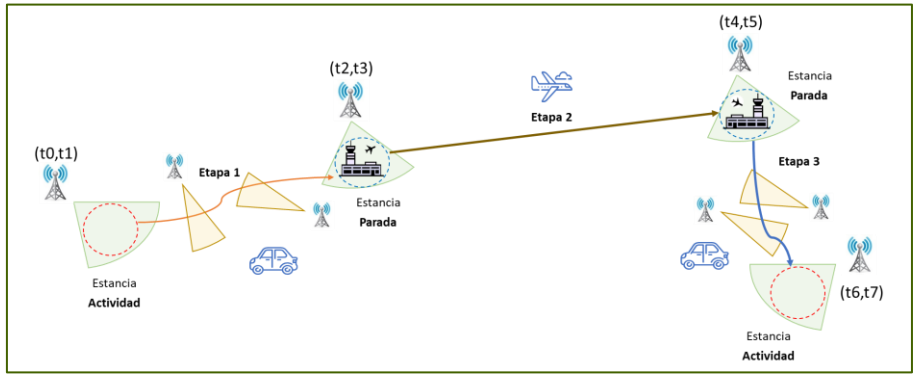


- Trip frequency
- Identification of professional drivers: HGV, LGV, taxi
- ...

NOMMON

Reconstruction of activity and travel diagrams

6. User classification/profiling



Data from MNO's CRM

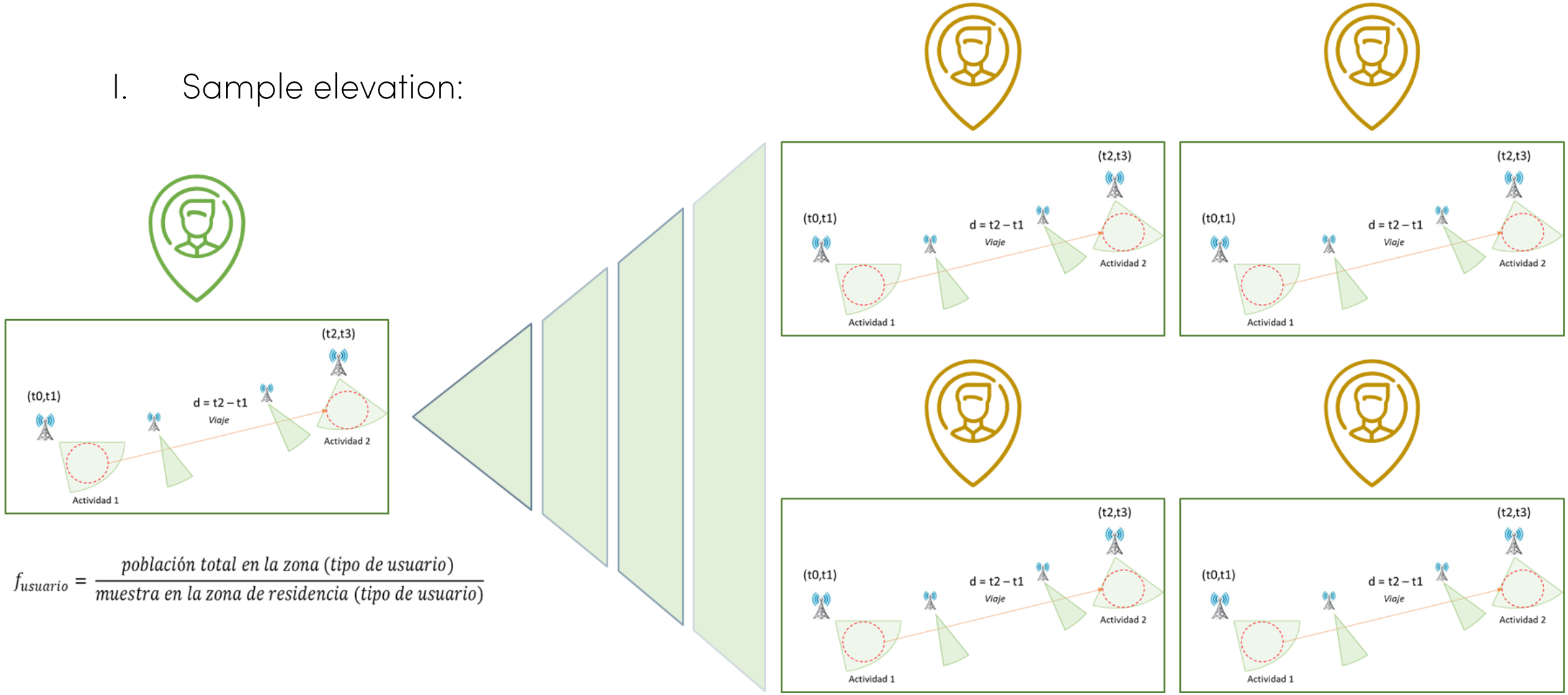
Web navigation

Official statistics

- Age
- Gender
- Nationality
- Other: based on contract data, browsing profiles or statistical cross-referencing based on place of residence (e.g. income level)

Sample expansion and generation of indicators

I. Sample elevation:



$$f_{usuario} = \frac{población\ total\ en\ la\ zona\ (tipo\ de\ usuario)}{muestra\ en\ la\ zona\ de\ residencia\ (tipo\ de\ usuario)}$$

NOMMON

Sample expansion and generation of indicators

II. Generation of final indicators :

- Post-processing and aggregation, with statistical disclosure control (SDC) measures that eliminate the risk of re-identification.

OD Matrix

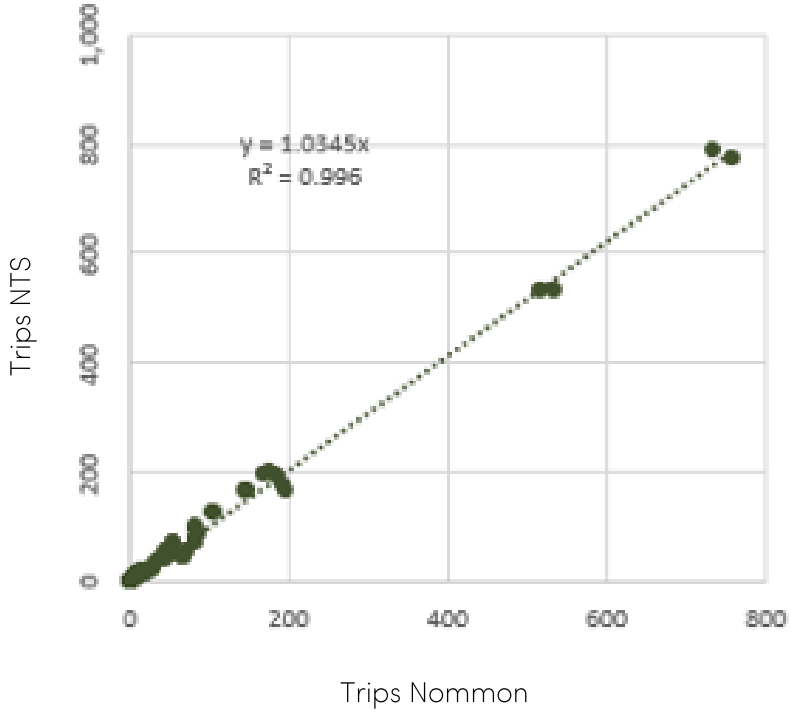
Origen	Destino	Periodo	Residenci	Género	Edad	Próposito_origen	Próposito_destino	Rango de	Viajes
20	06	Tarde	Locales	Hombre	+65	Frecuente	Frecuente	0-1000	7.48
10	01	Mañana	No_locales	Mujer	10-24	Frecuente	Frecuente	+10000	37.23
02	04	Mediodía	Locales	Mujer	10-24	No frecuente	Trabajo/Estudio	5000-10000	3.01
01	17	Tarde	Locales	Hombre	+65	No frecuente	Frecuente	0-1000	56.30
20	07	Tarde	No_locales	Hombre	45-65	No frecuente	No frecuente	5000-10000	67.37
04	15	Mediodía	Extranjeros	Hombre	+65	NA	NA	0-1000	60.42
23	12	Mañana	Locales	Mujer	+65	Frecuente	Casa	+10000	35.18
16	13	Mañana	Locales	Mujer	+65	Casa	Casa	0-1000	58.05
06	19	Resto	No_locales	Hombre	45-65	No frecuente	Casa	5000-10000	82.68

Importance of validating results

- There is no standardised/agreed methodology.
- Different methodologies and algorithms can lead to very different results.
- Importance of validating the analytical solution and comparing it with other sources.
- Key metrics:
  - Trip generation rates
  - Temporal distribution of trips
  - Distribution of trip distances
  - Correlation between OD matrices
  - Modal split
  - ...

Validation of results

Comparison with UK National Travel Survey



4. Case studies in aviation

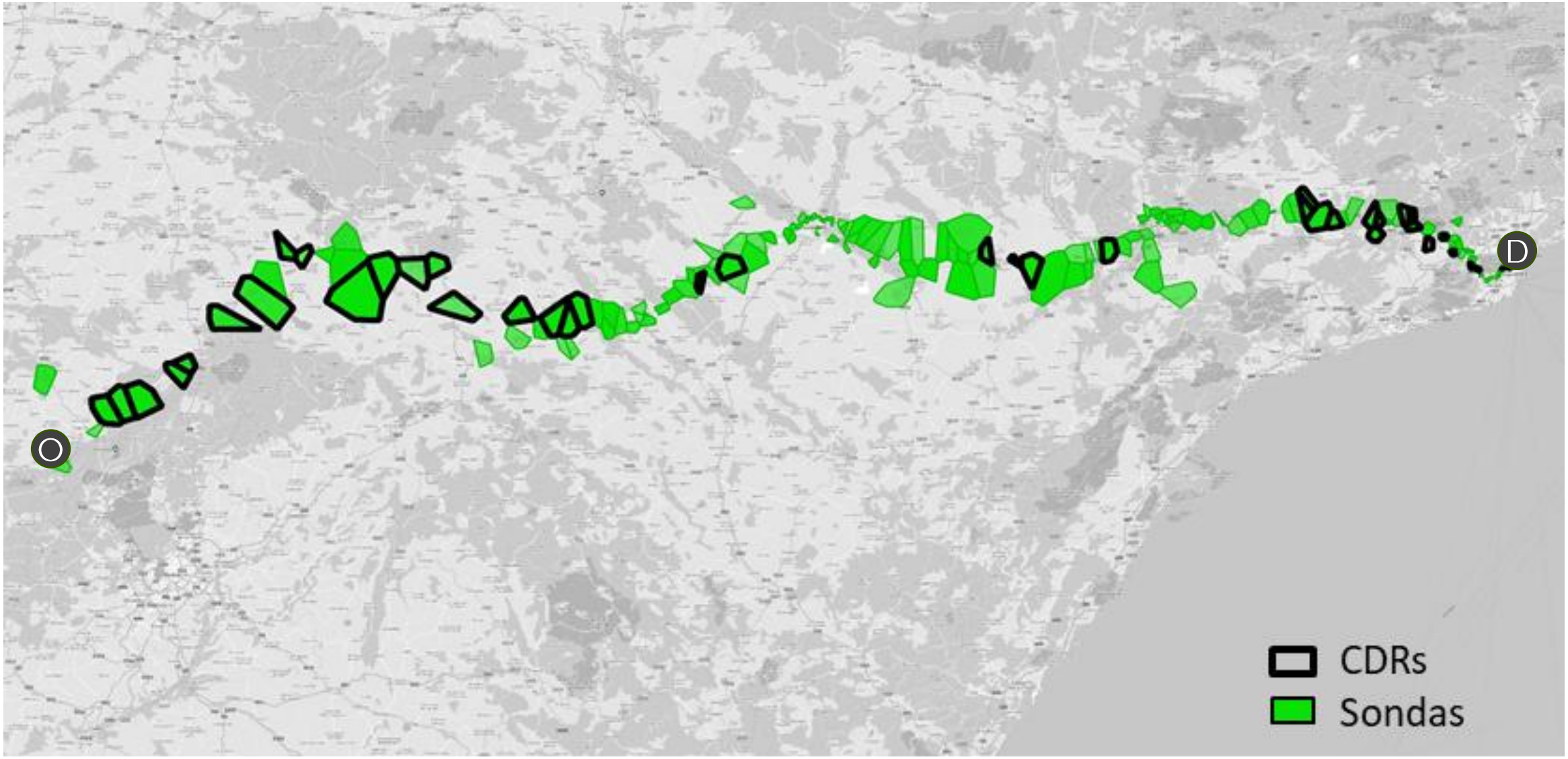
1 MULTIMODALITY: Mode detection and passenger archetypes

2 AIRPORT PLANNING AND MANAGEMENT: Catchment areas and traveller profiling

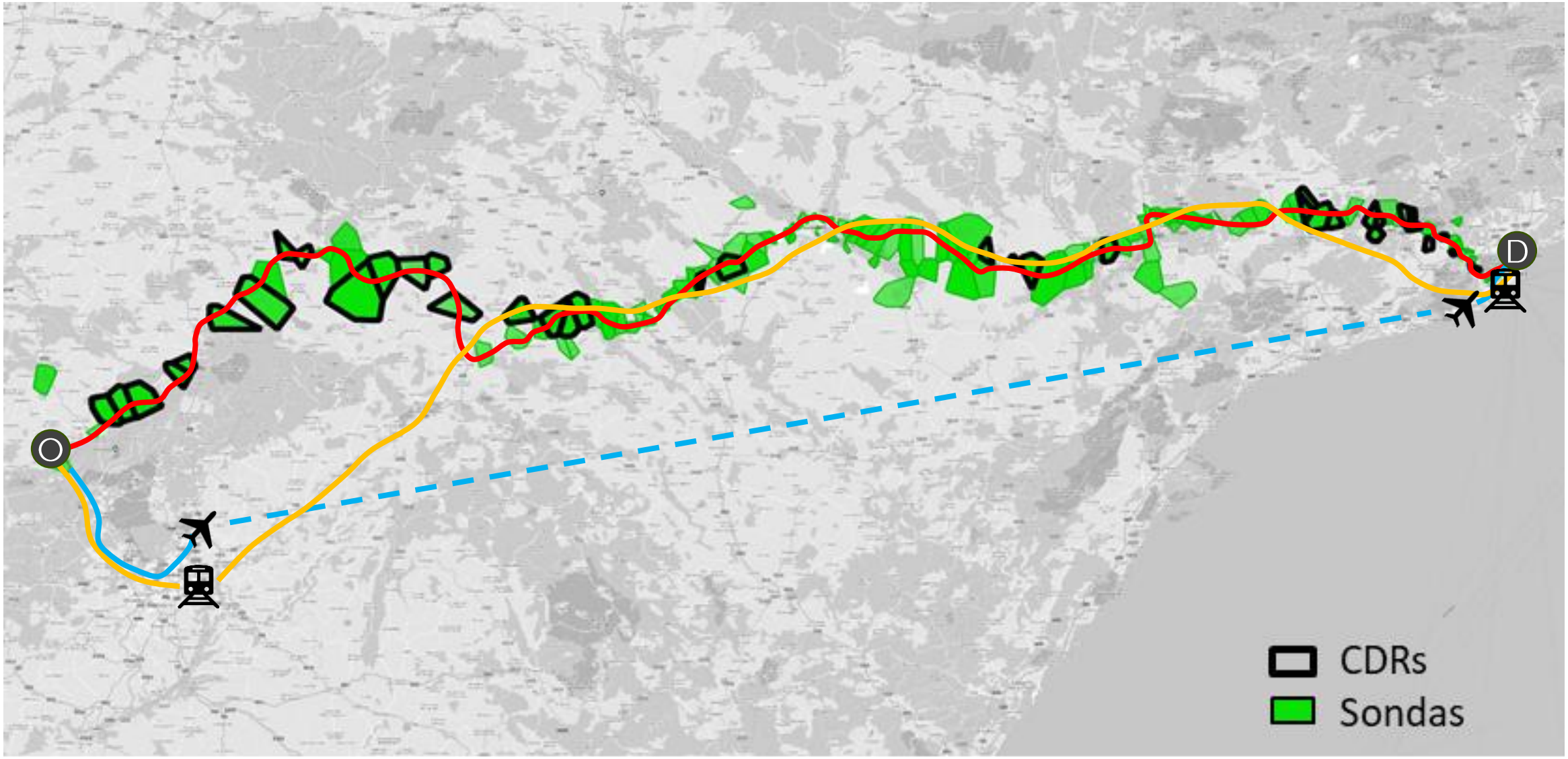
3 AIRLINES: New routes and demand prediction

4 CONNECTIVITY: IAM services

Multimodality: Mode detection



Multimodality: Mode detection



Multimodality: Passenger archetypes

- Objectives:
  1. **Exploit mobile network data to extract passenger archetypes based on the long-distance travel mobility patterns** (at passenger level) observed in the data.
     - Apply **unsupervised ML** to identify clusters of passengers according to their behaviour: number of trips, duration, type of destination region, distance, time of year, etc.
  2. Determine links that need to be served by air/rail exclusively or simultaneously (e.g. due to limiting capacity in either of the networks);

     Minimize the overall passenger journey times,

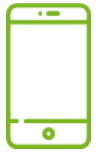
     Optimizing connection times; Minimize impacts of everyday stochastic disruptions.
- Input data
  - Supply and policies
  - Demand for each passenger archetype -> MND

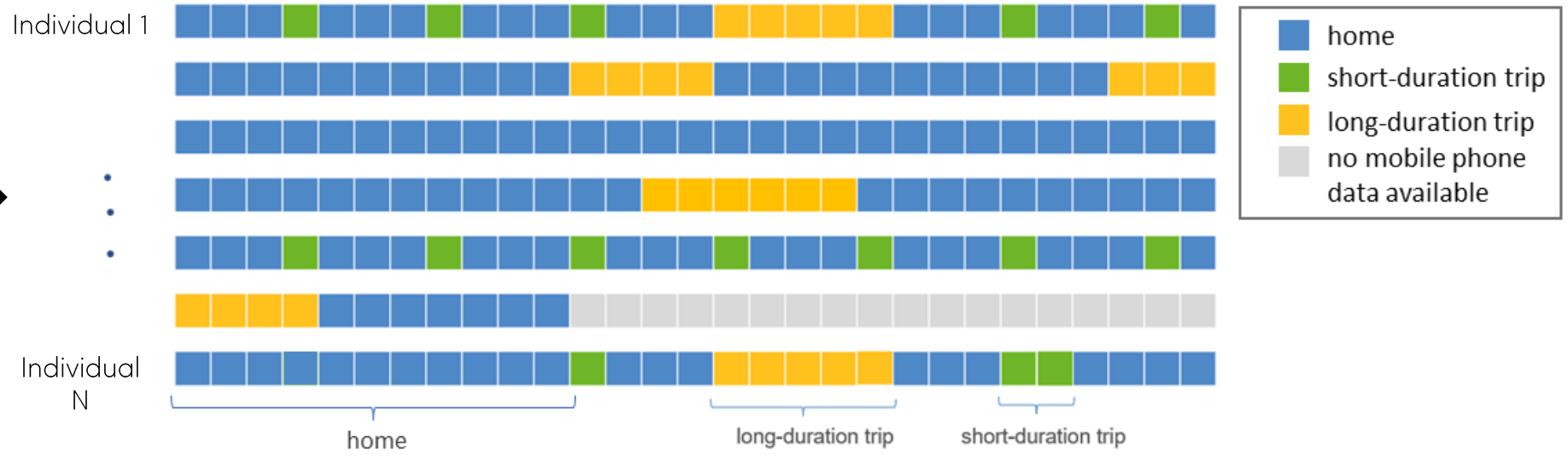
Multimodality: Passenger archetypes

Co-funded by the European Union

A longitudinal analysis to observe where the individuals spends the night throughout the year

Mobile Network data



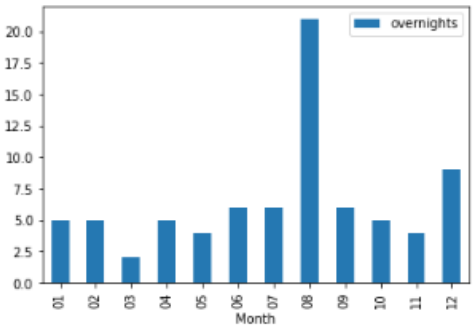
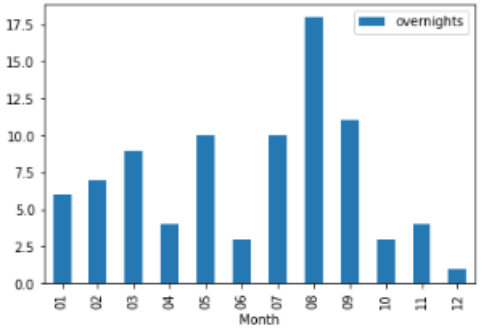
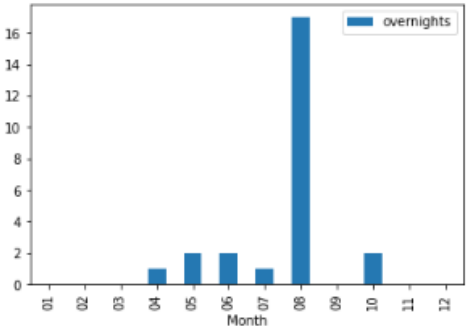
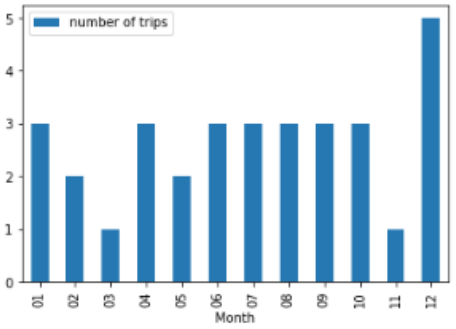
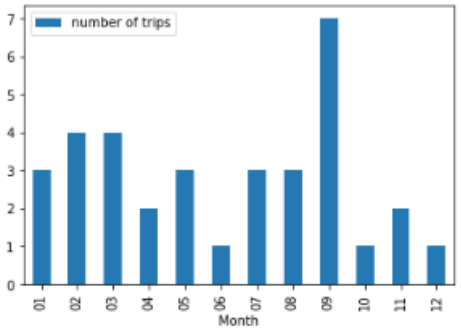
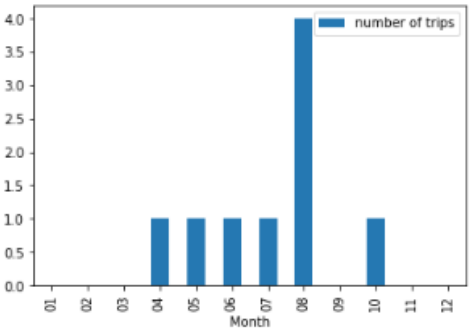
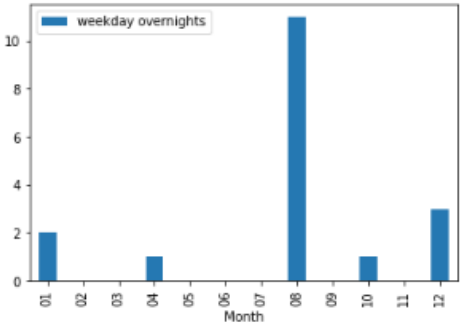
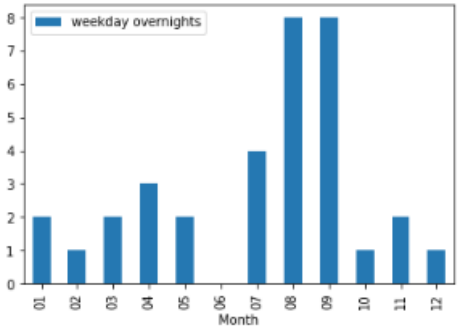
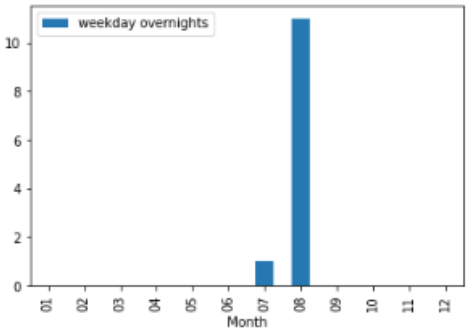


- Where they travel
- How often they travel
- How long they travel for

NOMMON

Multimodality: Passenger archetypes

sesar JOINT UNDERTAKING
Co-funded by the European Union

	Agent 1	Agent 2	Agent 3
Number of overnights			
Number of trips			
Number of overnights in weekdays			

Multimodality: Passenger archetypes

1. **Load overnights information:** 259.000 agents (~5% total sample)

2. **Calculate long-distance trip diary for each agent**:
   1. Characteristics: start date, end date, duration, destination, distance, weekday
   2. Parameters:
      1. Long-distance trips: > 100 km
      2. Holidays: summer (July and Augst), holy week, Christmas (from 22 December to 6 January)

3. **Calculate features**

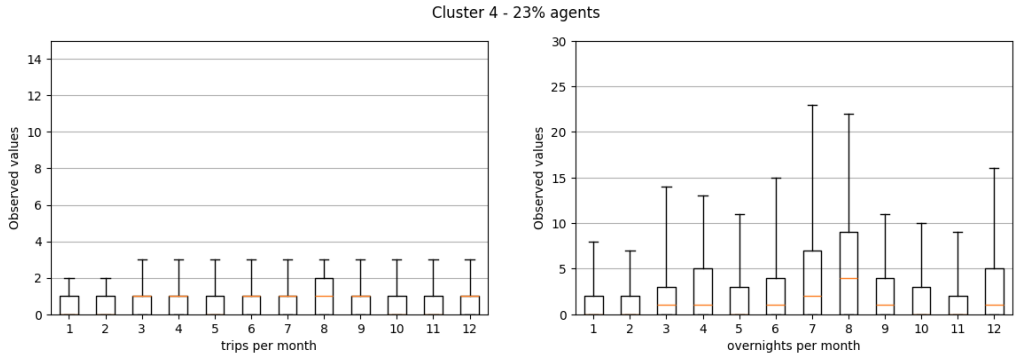
4. **Preprocessing:**
   1. Correlation analysis: no feature removal
   2. Scaling: Min-Max scaler
   3. PCA: amount of variance > 0.98

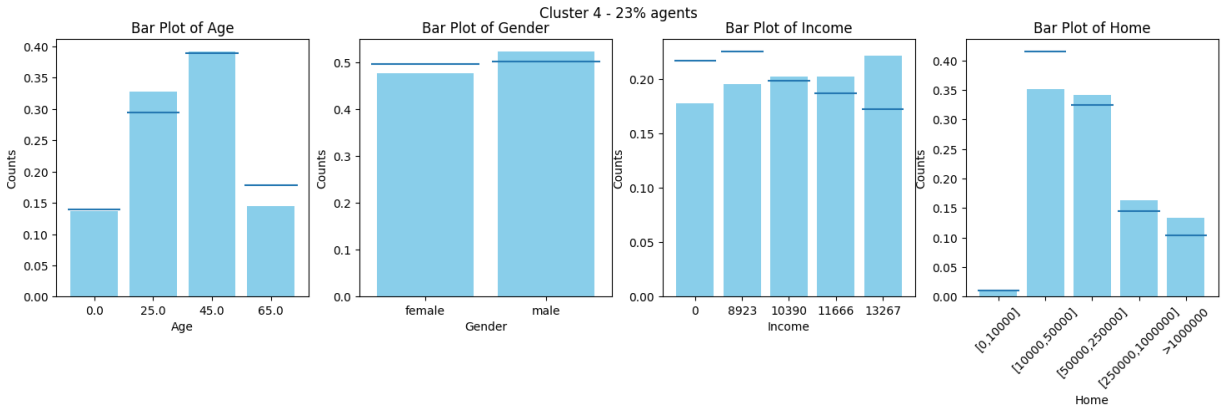
5. **Clustering:**
   1. Filter non-traveler group
   2. KMeans clustering: number of cluster analysis → optimal is 5 or 6

6. **Profile analysis of each cluster:**
   1. Age, gender, income, home population
   2. Travel characteristics: business, summer travelers, etc.

Multimodality: Passenger archetypes

ID = 4





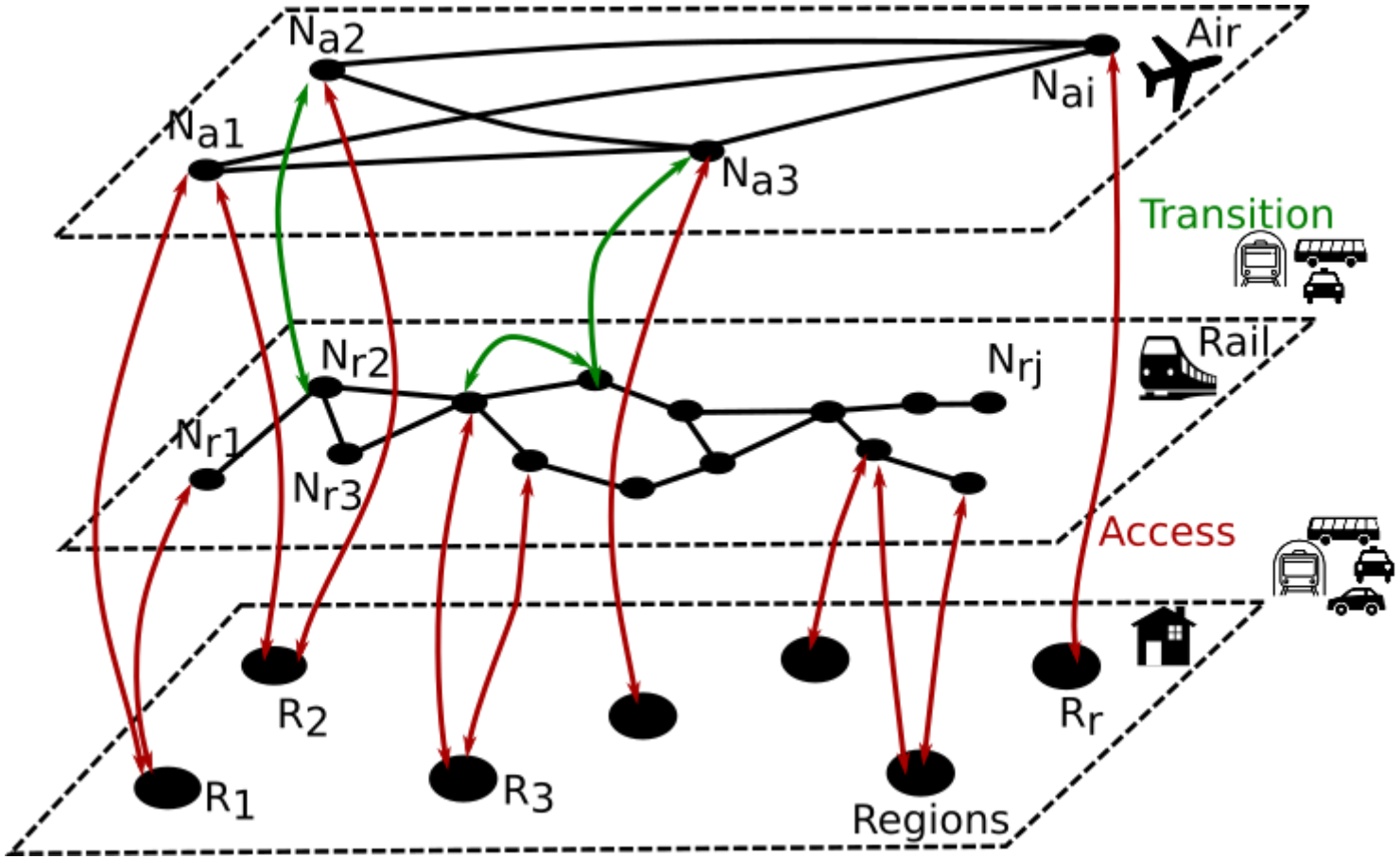
Passenger archetype	Long-distance actives
	Young, high-income individual who frequently travels to various destinations, covering different distances

feature	metric	Cluster 4
number_of_trips	mean	9.57
	std	8.27
number_of_overnights	mean	39.41
	std	34.65
average_distance	mean	422842
	std	421834
average_duration	mean	4.51
	std	3.70
number_of_trips_weekday_perc	mean	0.08
	std	0.13
number_of_overnights_abroad_perc	mean	0.09
	std	0.15
variation_distance	mean	0.48
	std	0.47
variation_overnight_month	mean	1.48
	std	0.44
cluster_size	%	23

Passenger archetype	N = 6
Non-traveler	Cluster -1
Sporadic long-haul traveler	Cluster 0
Occasional weekday traveler	Cluster 1
Domestic summer traveler	Cluster 2
Sporadic **international** traveler	Cluster 3
Long-distance actives	Cluster 4
International urbanite	Cluster 5

NOMMON

Multimodality: MultiModX

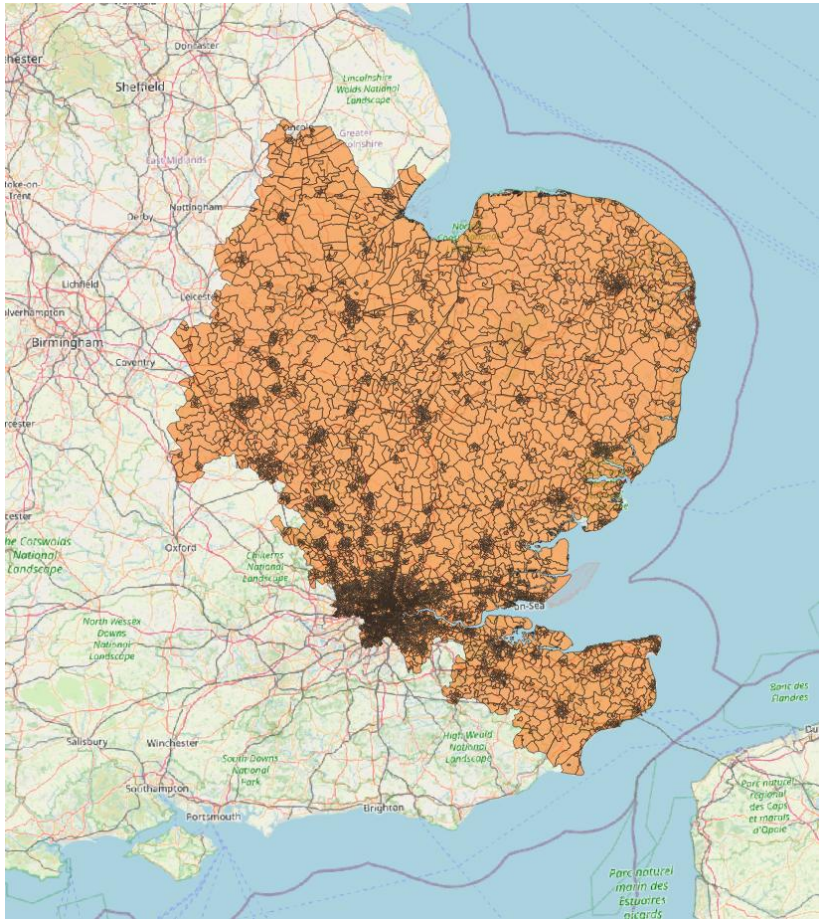


Multimodality: MultiModX

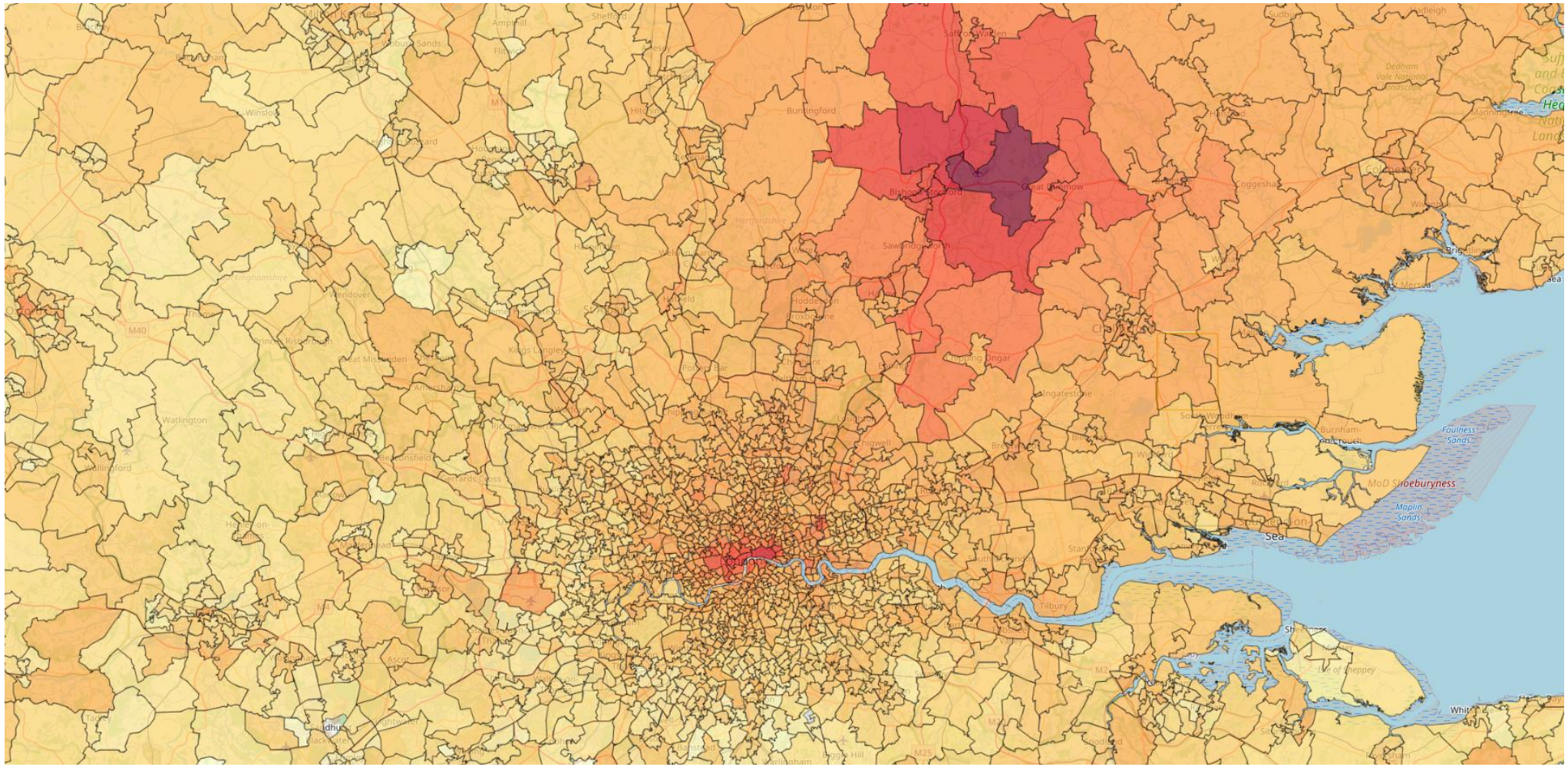
- Using the MND with the passenger archetypes:
  - Path assignment -> Logit model
  - Schedule coordination -> Optimization of wait times, overall travel times…
- Coordinated schedules for multimodal trip selection

Airport planning & management: Catchment areas

- MAG project:
  - Analyze catchment area of flights leaving any of their airports
  - Analyze destinations of competing airports with origins in the catchment area of their airports
  - Adjust with ticketing data for MAG airports
- Advantages of MND:
  - Real origin/destination of trips
  - Help with demand of new routes

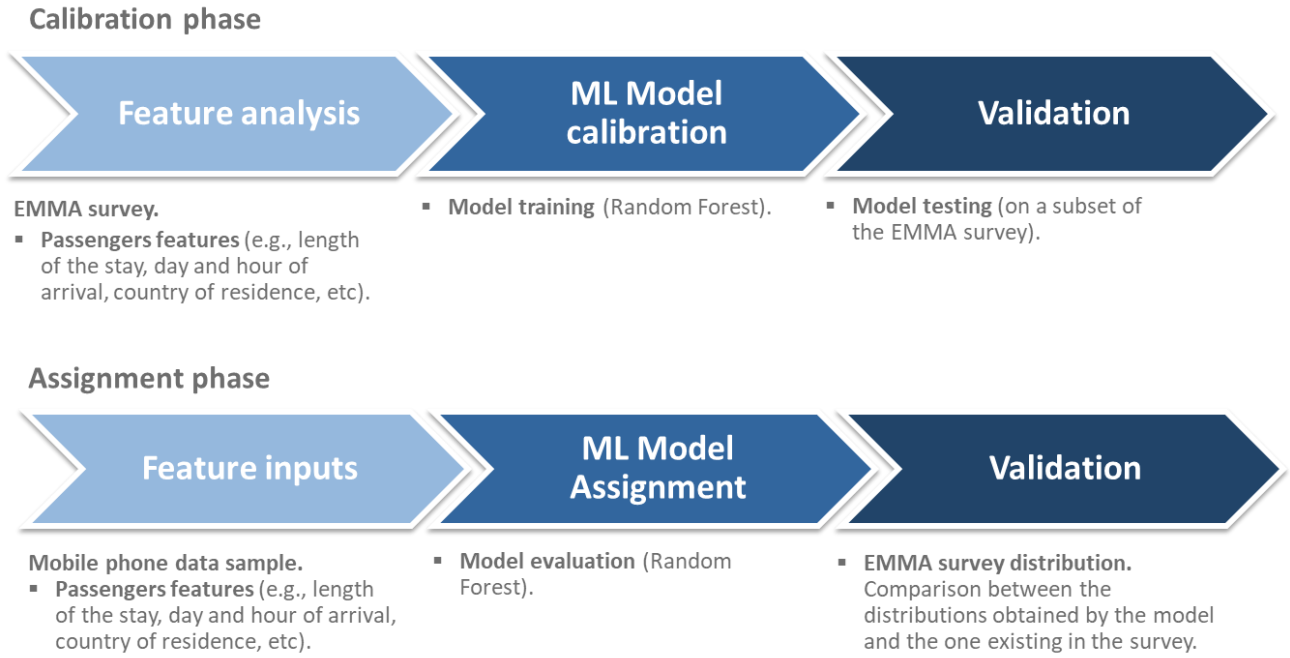


Airport planning & management: Catchment areas



Airport planning & management: Traveller profiling

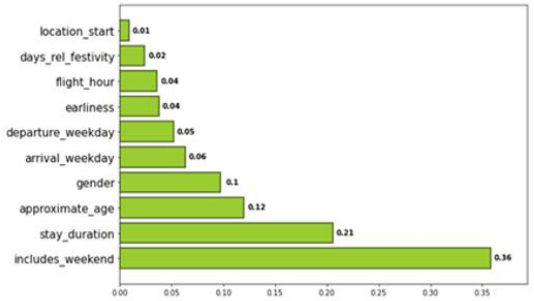
- TRANSIT project:
  - Enrich trip diaries obtained from mobile network data for air travel with information on trip purpose and airport access mode by using AI and passenger survey data.



Airport planning & management: Traveller profiling

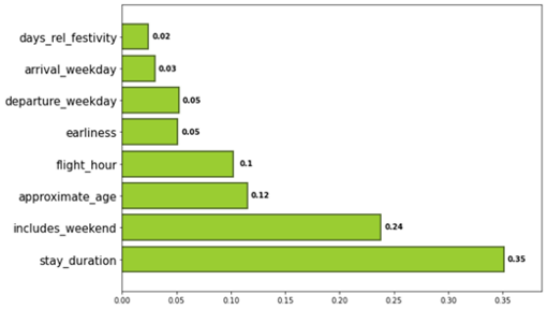
Trip purpose: Madrid

		F1-Score All EMMA Variables	F1-Score EMMA variables in Mobile Network Data	Training Sample Size
Validation	Mean CV Score	0.80	0.78	-
Test	Leisure	0.88	0.88	6,231
	Business	0.70	0.67	2,627
	Average	0.79	0.77	8,858
	Weighted Average	0.83	0.82	8,858



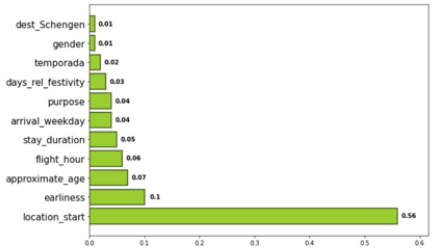
Trip purpose: Palma

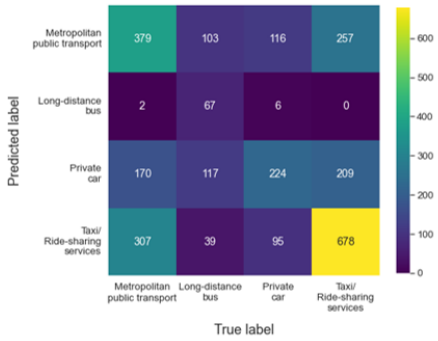
		F1-Score All EMMA Variables	F1-Score EMMA variables in Mobile Network Data	Training Sample Size
Validation	Mean CV Score	0.75	0.75	-
Test	Leisure	0.96	0.97	3,904
	Business	0.49	0.55	205
	Average	0.73	0.79	4,109
	Weighted Average	0.94	0.95	4,109



Access/Egress: Madrid

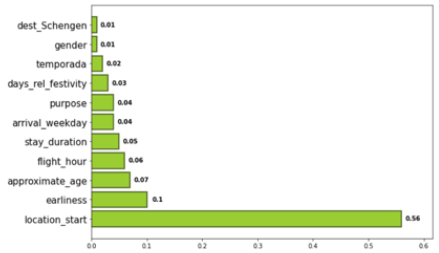
		F1-Score All EMMA Variables	F1-Score EMMA variables in Mobile Phone Data	Training Sample Size
Validation	Mean CV Score	0.47	0.44	-
Test	Metropolitan public transport	0.40	0.44	2,509
	Long-distance bus	0.35	0.33	247
	Private car	0.54	0.39	2,206
	Taxi/Ride-sharing	0.63	0.60	3,342
	Average	0.48	0.44	8,304
	Weighted Average	0.53	0.49	8,304

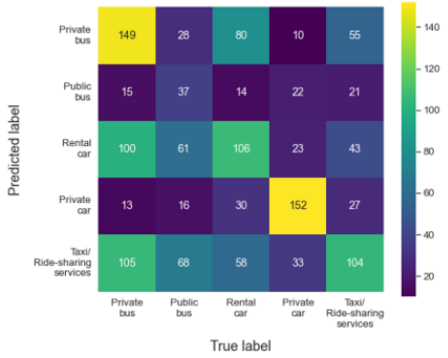




Access/Egress: Palma

		F1-Score All EMMA Variables	F1-Score EMMA variables in Mobile Phone Data	Training Sample Size
Validation	Mean CV Score	0.49	0.42	-
Test	Private bus	0.56	0.48	983
	Public bus	0.20	0.26	344
	Rented car	0.42	0.38	956
	Private car	0.78	0.62	774
	Taxi/Ride-sharing	0.41	0.31	1,052
	Average	0.48	0.41	4,109
	Weighted Average	0.50	0.42	4,109



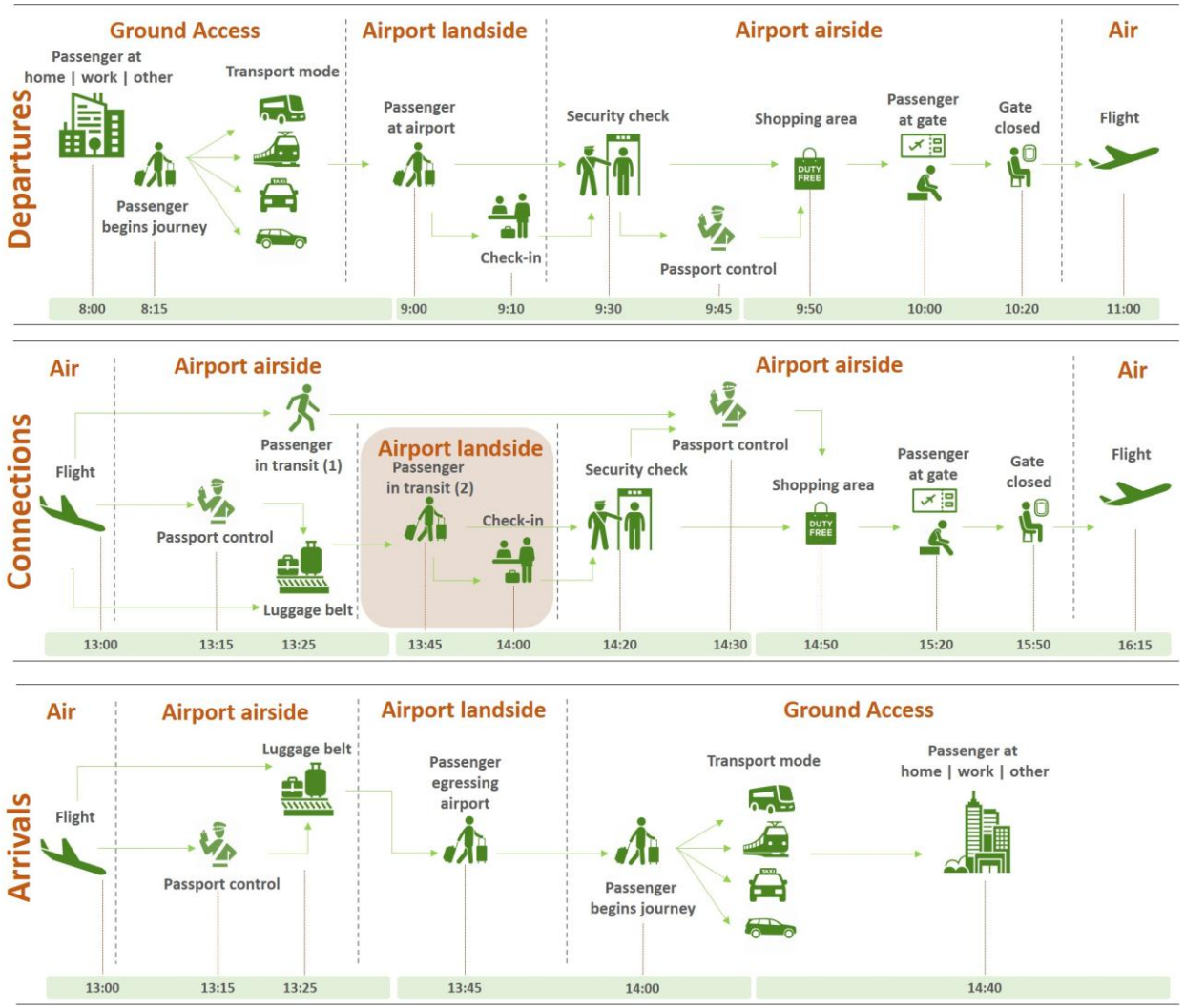


NOMMON

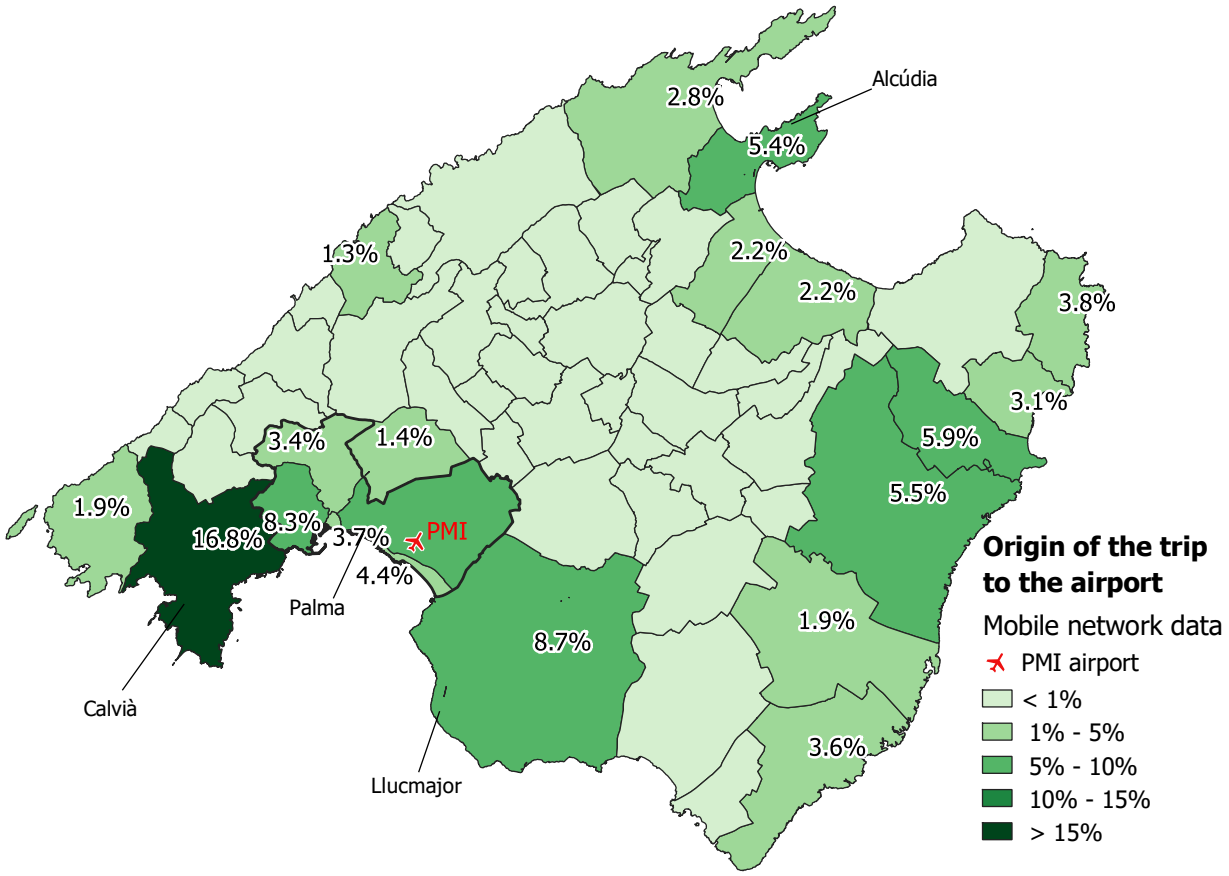
Airport planning & management: Traveller profiling

IMHOTEP project (Palma de Mallorca airport):

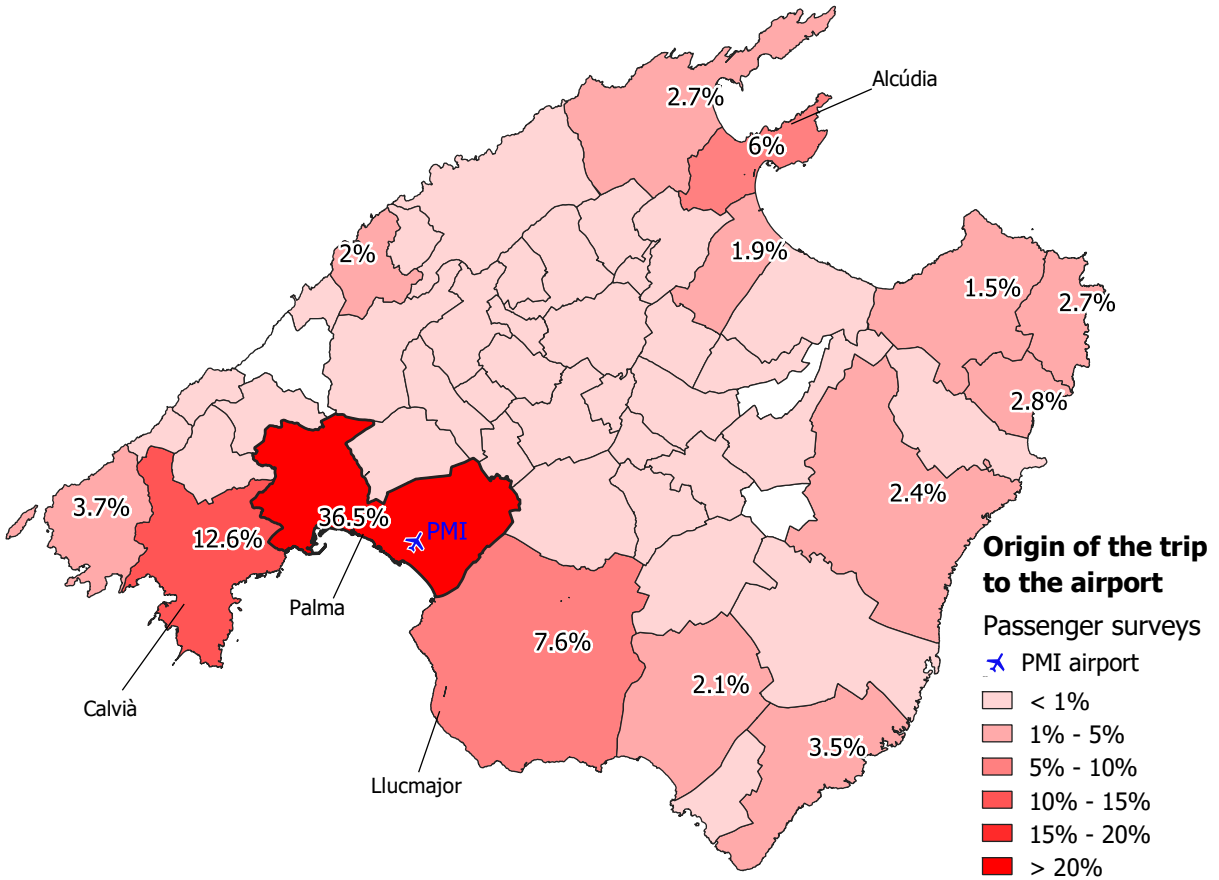
- The ability to characterize door-to-door air travel behavior is essential to design a system able to satisfy passenger preferences, needs and constraints -> passenger-centric
- Fusion of MND with other data sources:
  - Flight schedules
  - Flight ticketing data
  - Passenger surveys
  - PT surveys
  - Boarding card reader (BCR) data
  - Geolocated mobile apps



Airport planning & management: Traveller profiling



Mobile Network Data



Passenger surveys

Airport planning & management: Traveller profiling

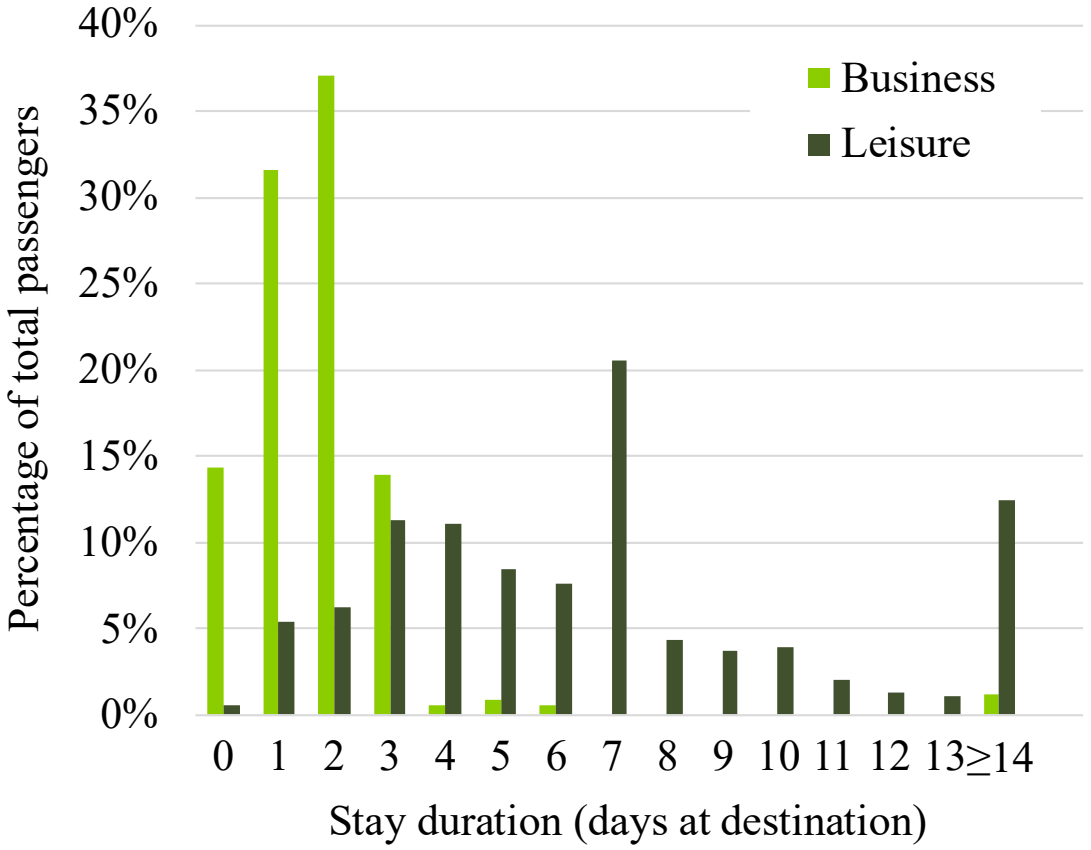
Calibration phase

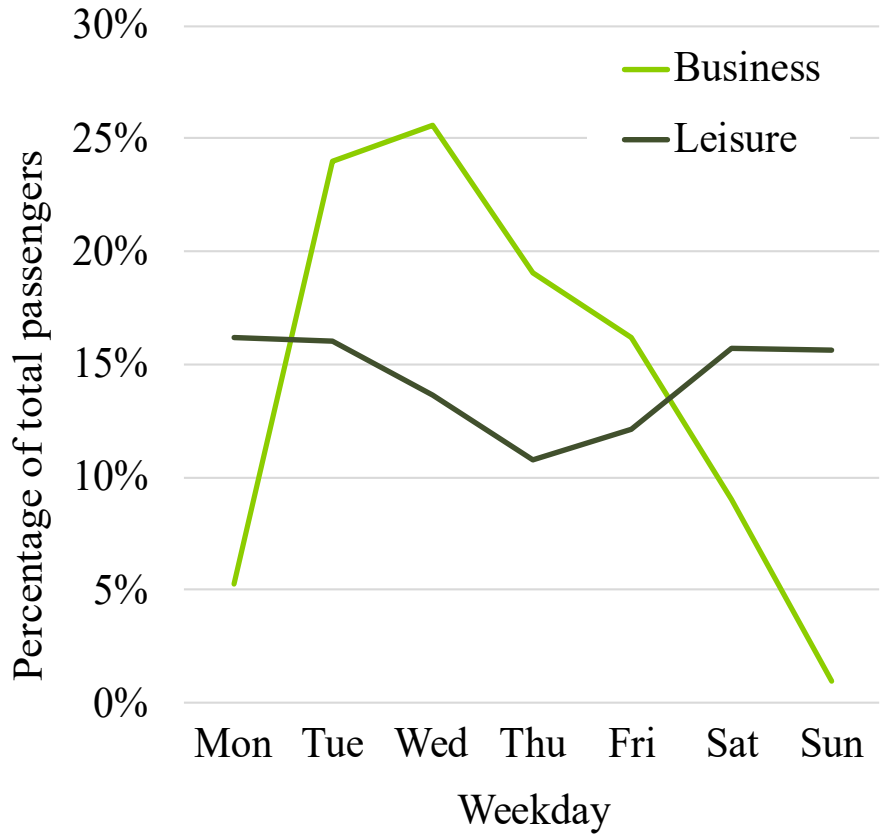
Inputs	ML Model Calibration	Validation
	▪ **Select a ML algorithm** and its parameters to minimize classification error	▪ **Test a model** on a subset of the survey sample

Assignment phase

Inputs	ML Model Assignment	Validation
▪ **Mobile network data** ▪ Passenger features (length of stay, destination, etc.) + target variable	▪ **Apply the ML classification algorithm**	▪ **Compare the results** with other data sources

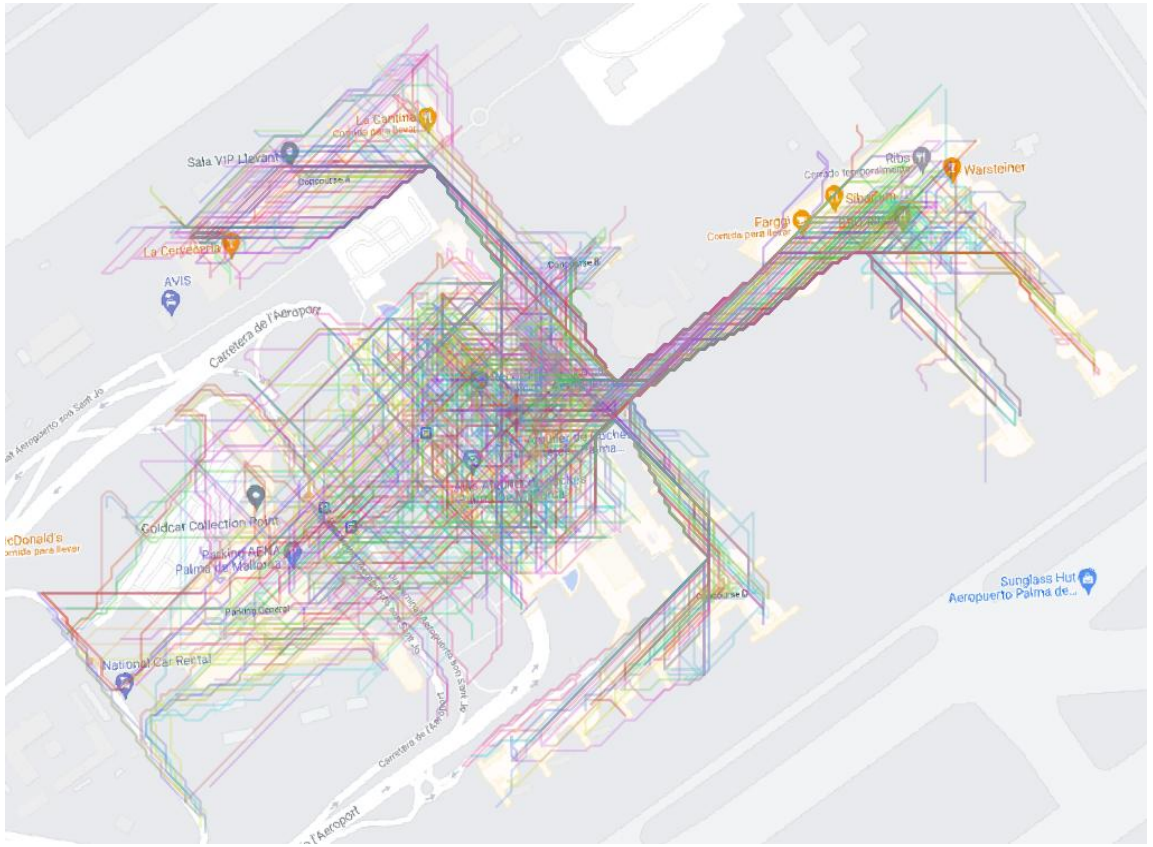
Airport planning & management: Traveller profiling





Airport planning & management: Traveller profiling

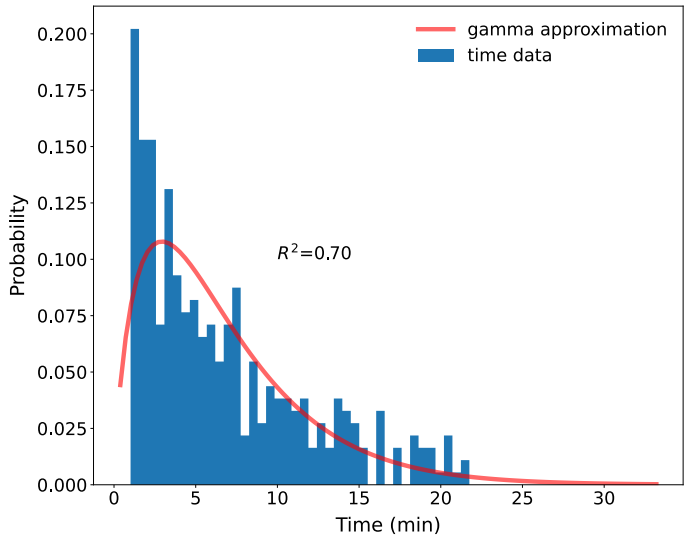
Airport planning & management: Traveller profiling



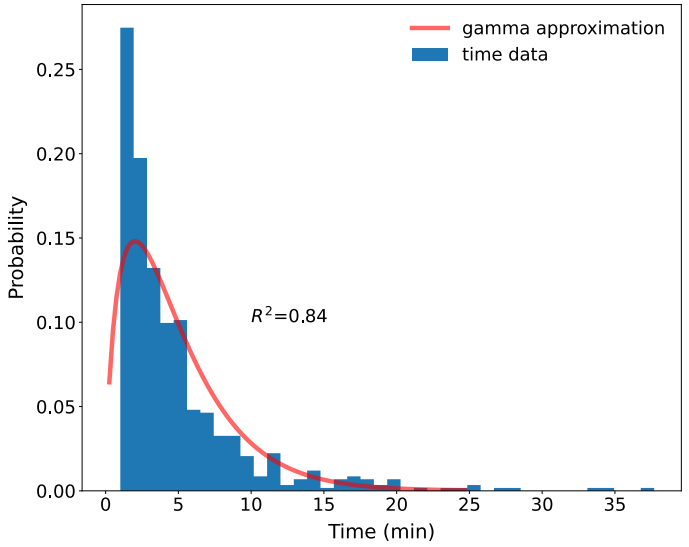
Airport planning & management: Traveller profiling

- From trajectories, passenger stay times information on the different airport areas can be derived
- BCR data allows the assignment of arrival times to security control
- Mobile apps data opens the door to a more detailed understanding of behavior within the terminal
- Obtained enhanced activity-travel diaries

Passport control area

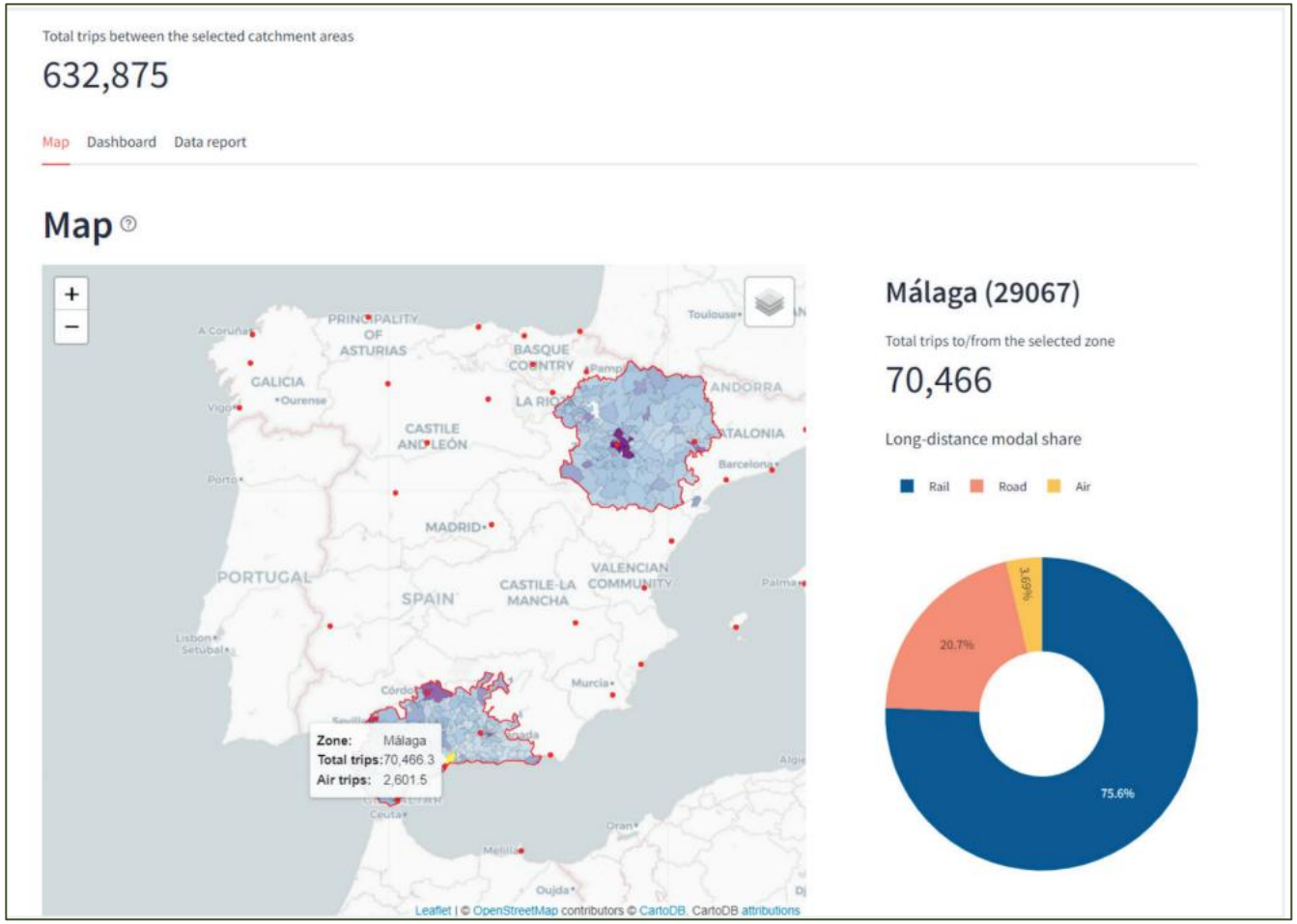


Commercial area



Airlines: demand prediction for new routes

- Mobility analysis between the Spanish cities of Zaragoza and Santiago de Compostela before 2022
- No air route existed, and the forecast of the demand captured by such a route if it were implemented
- Predictions were then compared with the actual demand for the new air route opened by Ryanair in 2023
- WisePax provides accurate demand forecasts for new air connections

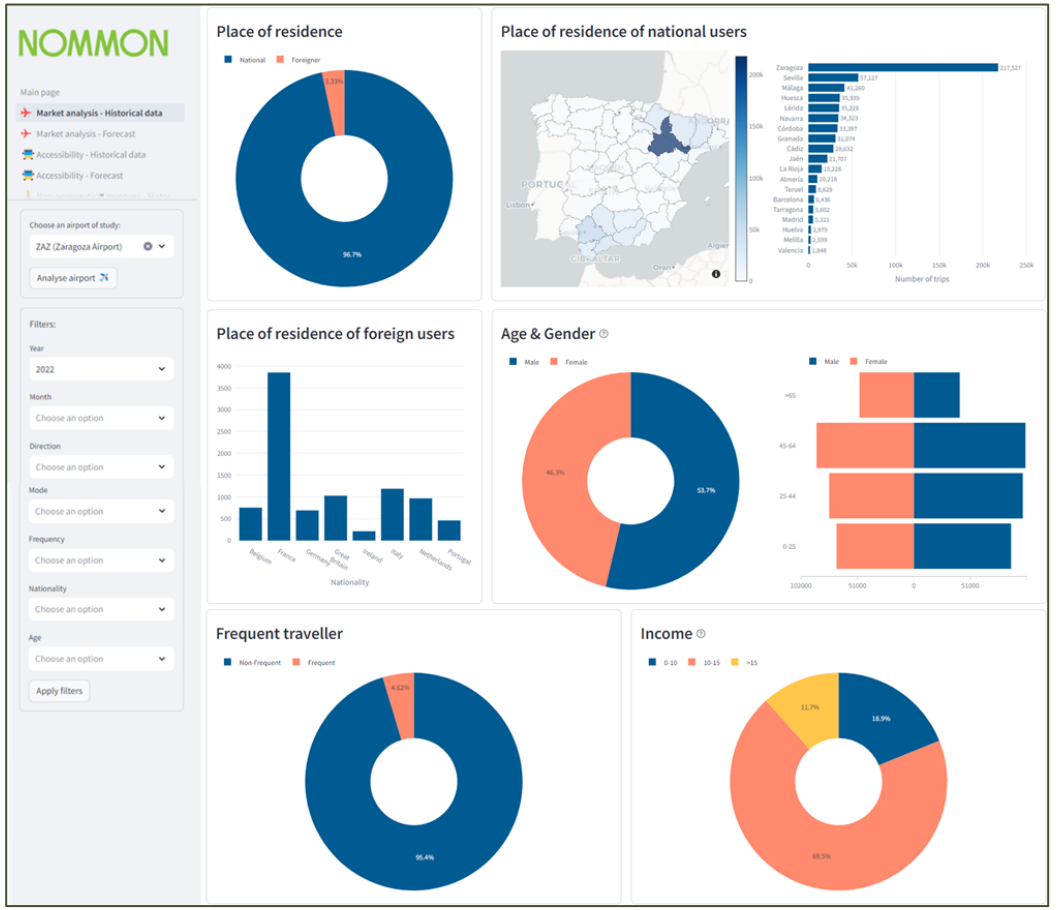


NOMMON

Airlines: demand prediction for new routes

Modal share

Passenger profile



Airlines: demand prediction for new routes

Connectivity: IAM services

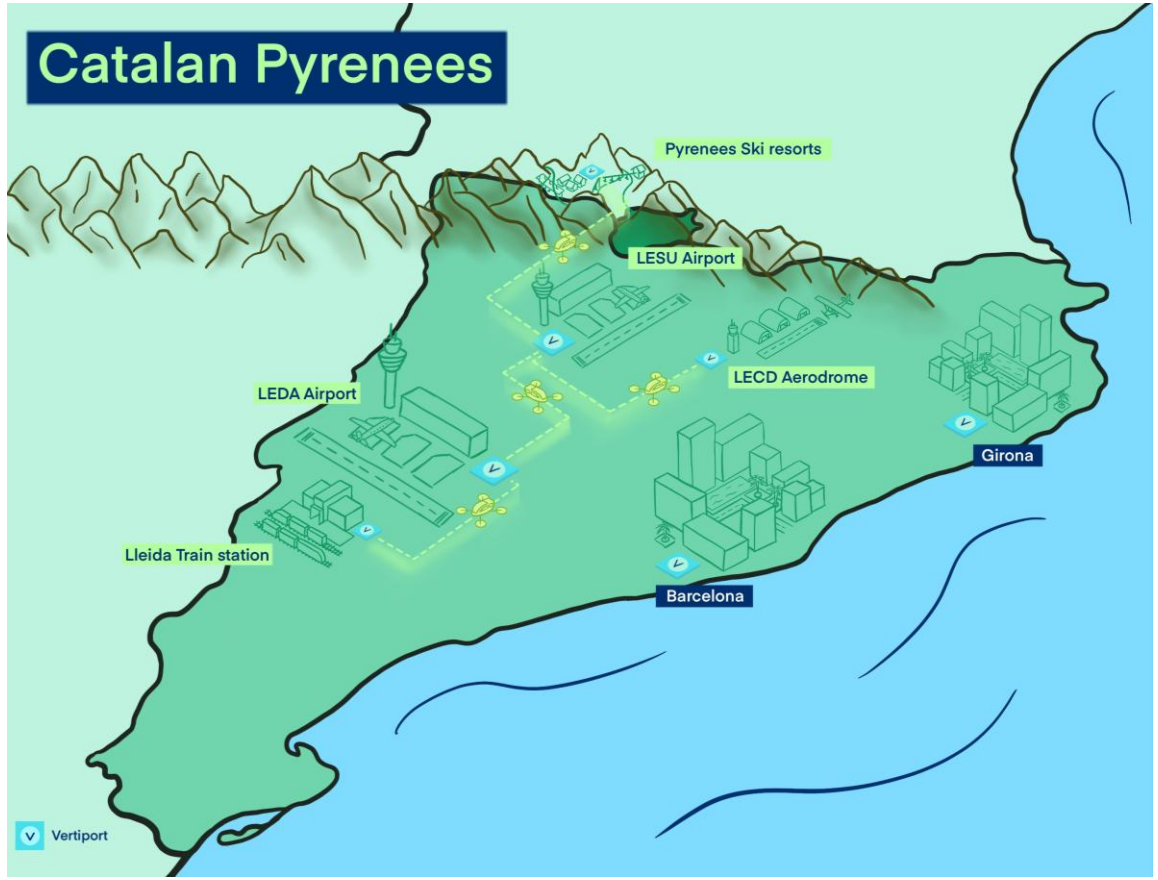
- Planning Regional-Scale Multimodal Operations for Innovative Air Mobility Services (PRIAM)
- Planning the integration of IAM infrastructure and services into regional multimodal transport networks
  - Passenger behaviour modelling:
    - IAM adoption (*Which passenger groups are more prompt to use IAM services?*)
    - Potential demand (*Which is the potential demand for IAM?*)
    - IAM induced demand and modal shift (*Which passenger demand is IAM able to capture? Which demand are these services going to induce?*)
  - IAM service characterization
    - IAM network planning and service optimization (*Which is the optimal location for placing vertiports? How IAM services should operate to optimise the performance of the service?*)

Connectivity: IAM services

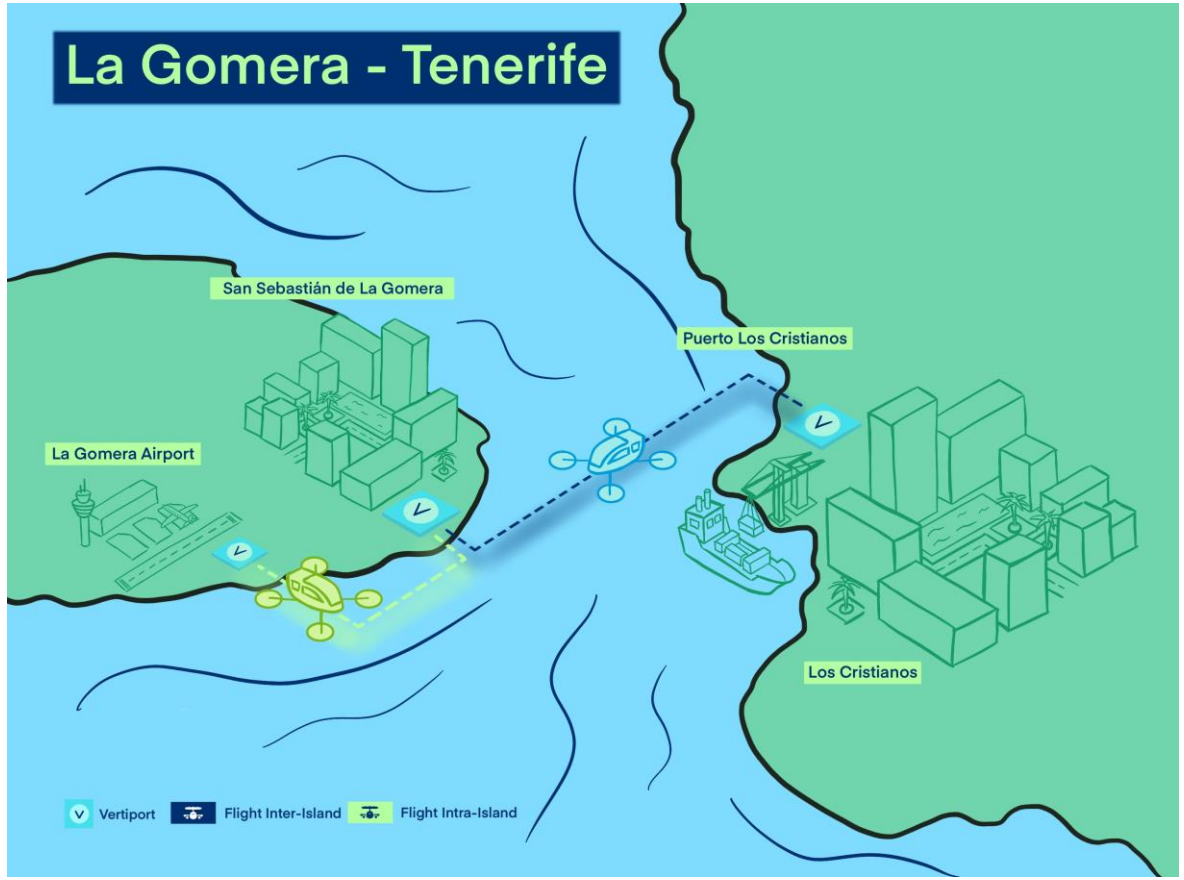
PRIAM

sesar JOINT UNDERTAKING

Mountainous regions



Inter/Intra-island connectivity



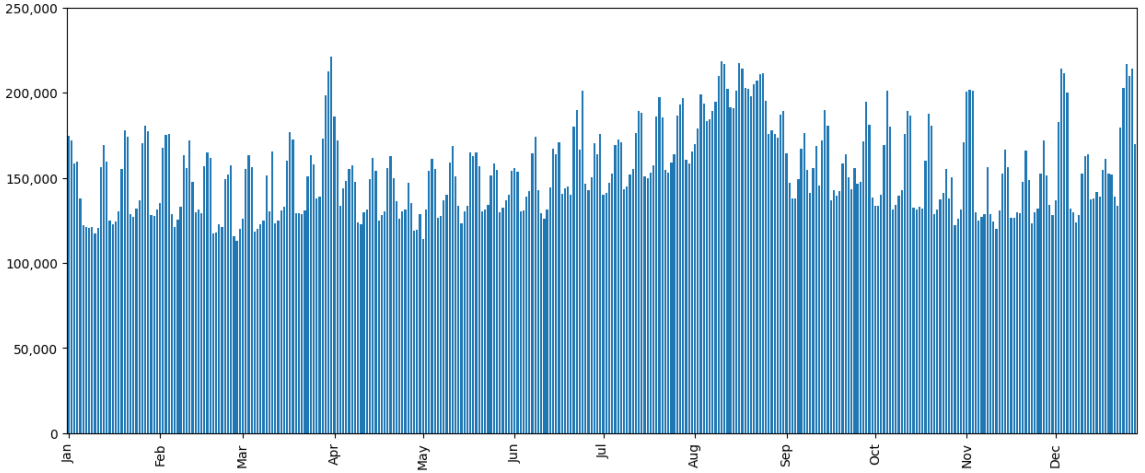
Connectivity: Passenger behaviour modelling

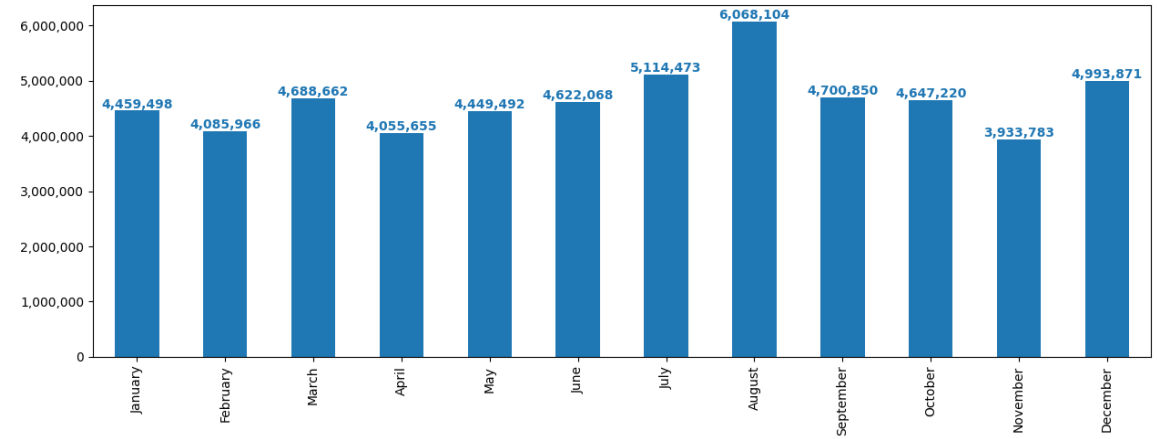
PRIAM

sesar JOINT UNDERTAKING

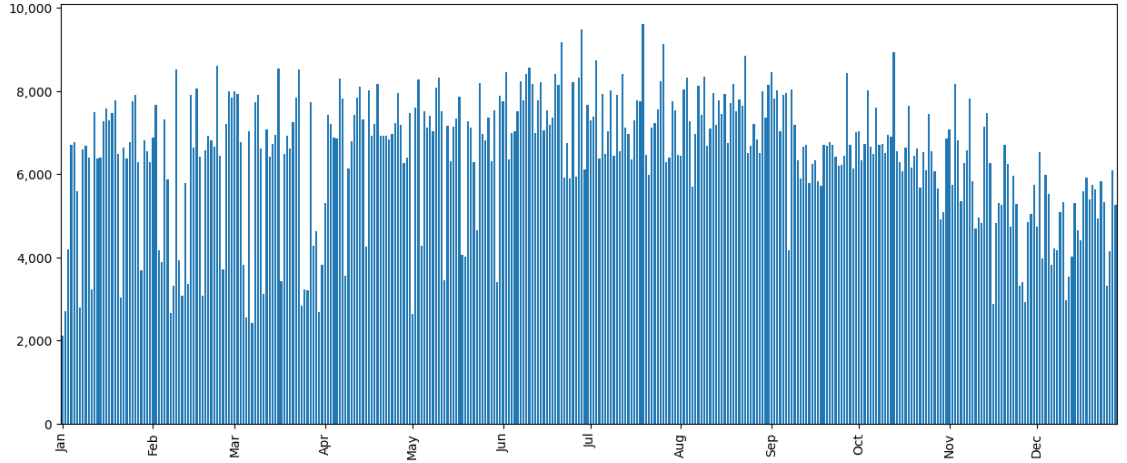
Use of Mobile Network Data to analyse the demand in the areas of study

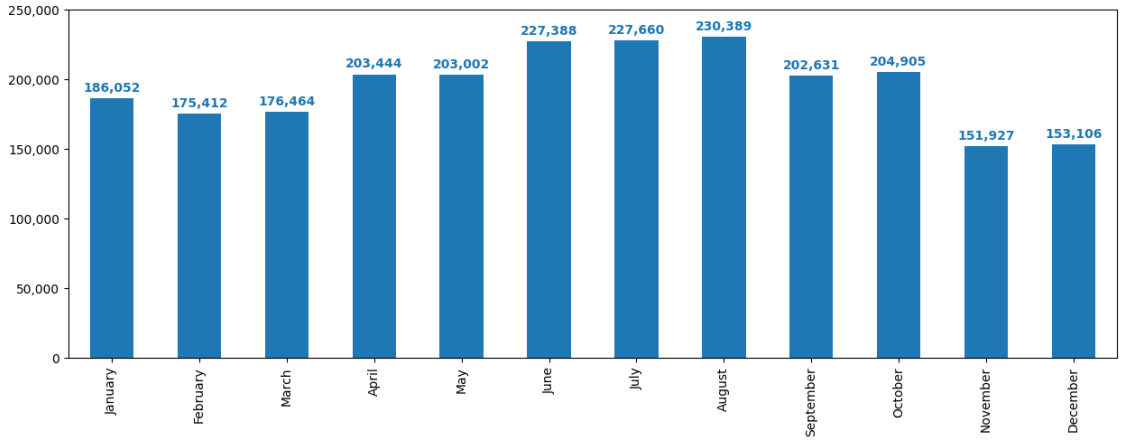
Daily and monthly trips to/from the Catalan Pyrenees in 2024





Daily and monthly trips between La Gomera and Tenerife in 2024





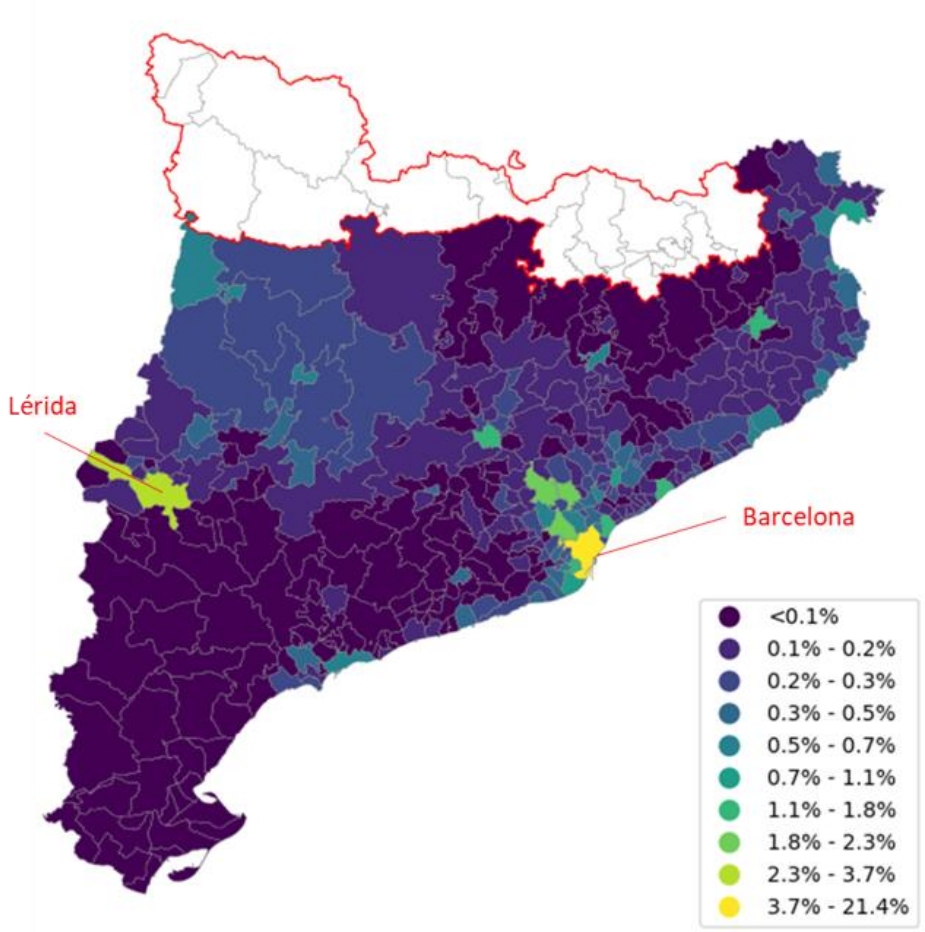
Connectivity: Passenger behaviour modelling

PRIAM

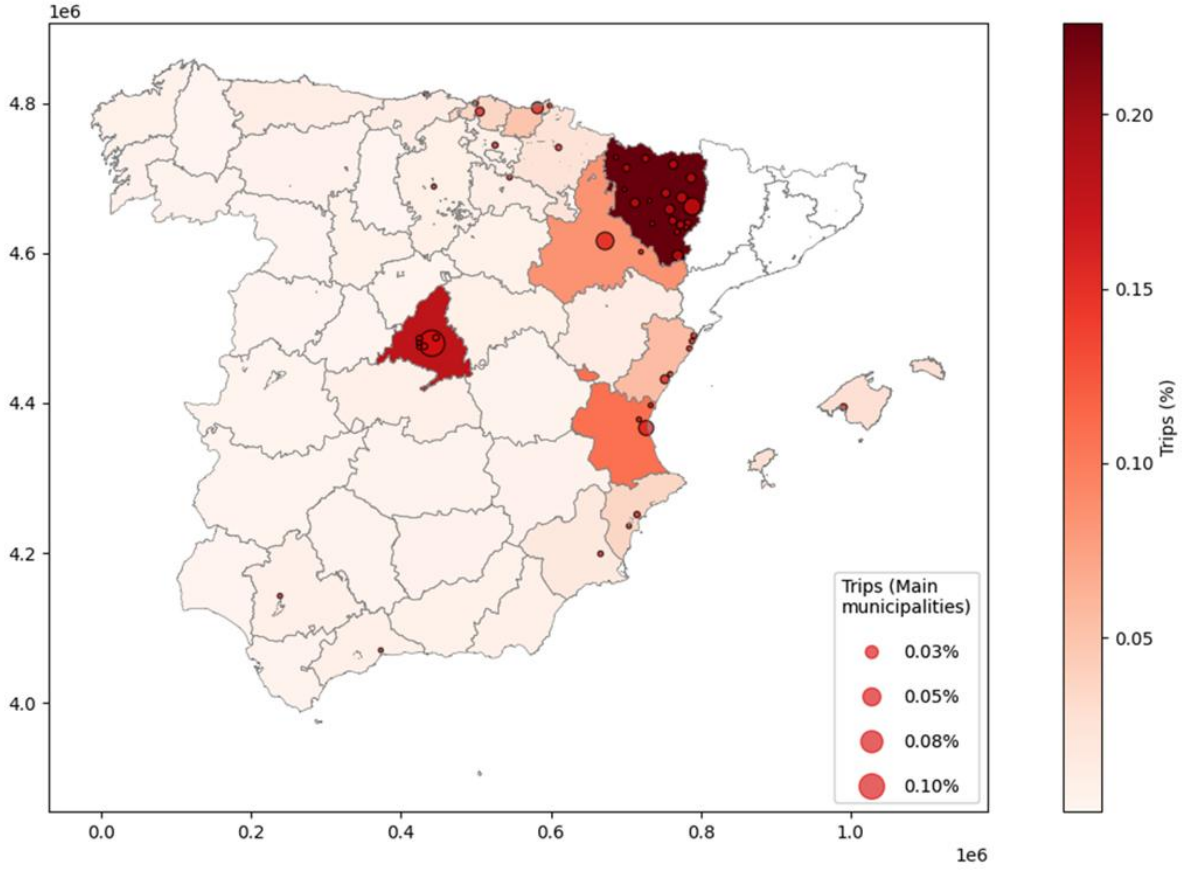
sesar JOINT UNDERTAKING

Catalan Pyrenees

Origin/Destination of the trips to/from the Catalan Pyrenees (Catalonia) – 92,5%



Origin/Destination of the trips to/from the Catalan Pyrenees (rest of Spain) – 7,5%



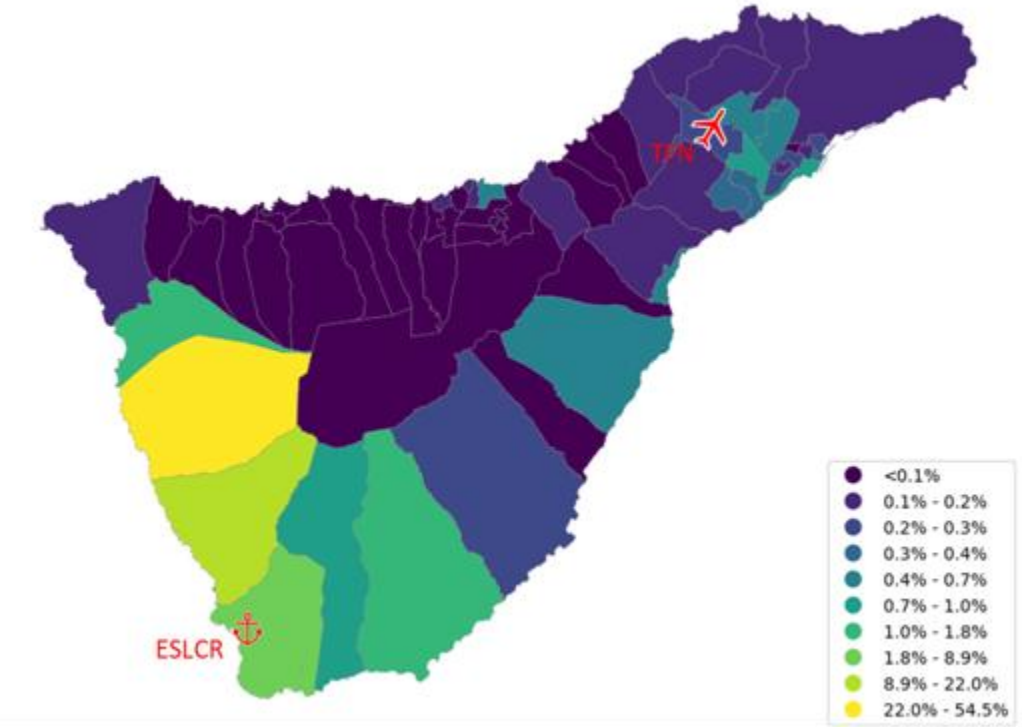
Connectivity: Passenger behaviour modelling

PRIAM

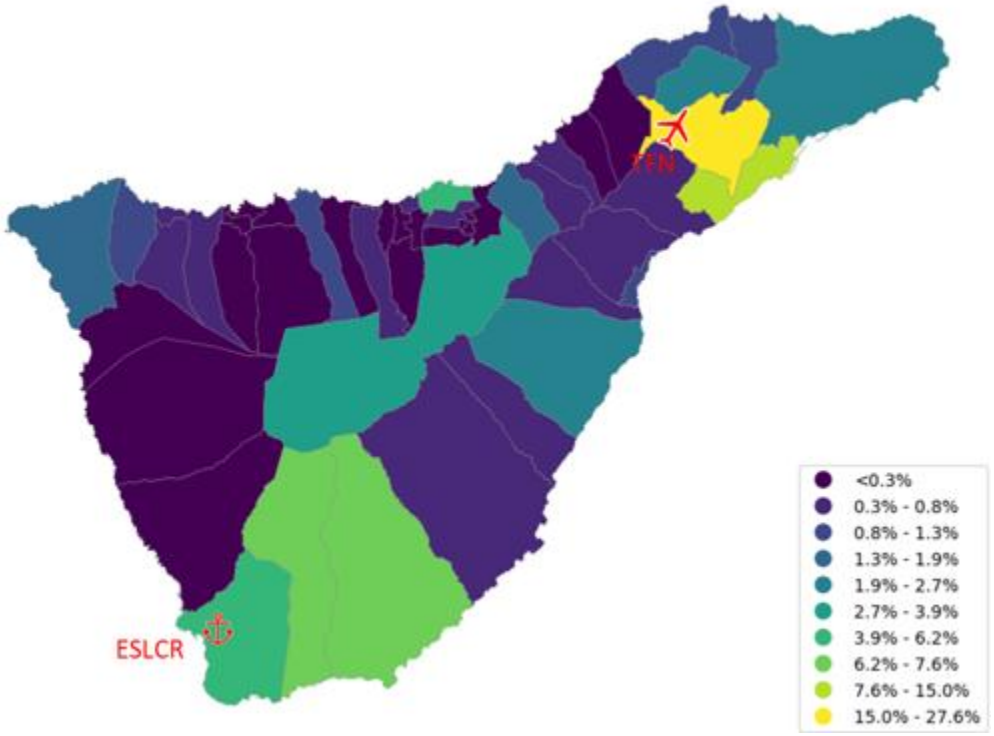
sesar JOINT UNDERTAKING

Tenerife-La Gomera

*Origin of the **ferry** trips between Tenerife and La Gomera (98,9%)*



*Origin of the **air** trips between Tenerife and La Gomera (1,1%)*



Thank you very much!

Federico Lucas

Delivery & Service Engineer

www.nommon.es

federico.lucas@nommon.es

Madrid

Pl. Carlos Trías Bertrán, 4
28020 Madrid, Spain
+34 91 072 62 61
nommon@nommon.es

London

82 William Court, Hall Road
London, NW8 9PB, UK
uk@nommon.es

Bogotá

Carrera 2B #69A – 32
Bogota D.C., Colombia
colombia@nommon.es

São Paulo

Av. Roque Petroni Junior,
1089 – Suite 713.
São Paulo – SP – Brazil
04707-900
brasil@nommon.es

Links to papers:

- WISEPAX – Santiago-Zaragoza Case Study: https://www.nommon.es/case-studies/wisepax-mobile-network-data-forecast-demand-zaragoza-santiago-air-route/
- TRANSIT – Airport accessibility surveys and mobile phone records data fusion for the analysis of air travel behaviour : https://www.sciencedirect.com/science/article/pii/S2352146523013157
- IMHOTEP – Data fusion for the analysis of air travel behavior: Application to Palma de Mallorca Airport: https://whova.com/xems/whova_backend/get_event_s3_file_api/?event_id=sesar_202212&file_url=https://d1keuthy5s86c8.cloudfront.net/static/ems/upload/files/1670000687_ebydu_SIDs_2022_paper_105_final.pdf&eventkey=4c68ac23dcb73ce4b5be87ca1b5b0ccaed6016dfcefe0c4b6af300caca061224
- MULTIMODX – Identification and Characterisation of Passenger Archetypes based on annual Longdistance Travel Patterns: https://www.sesarju.eu/sites/default/files/documents/sid/2024/papers/SIDs_2024_paper_041%20final.pdf
- MultiModX website: https://multimodx.eu/
- MultiModX linkedin: https://www.linkedin.com/company/multimodx-eu/
- PRIAM website: https://www.priam-project-sesar.eu/
- PRIAM linkedin: https://www.linkedin.com/company/priam-project-sesar/
- Nommon website: https://www.nommon.es/

Madrid
Pl. Carlos Trías Bertrán, 4
28020 Madrid, Spain
+34 91 072 62 61
nommon@nommon.es

London
82 William Court, Hall Road
London, NW8 9PB, UK
uk@nommon.es

Bogotá
Carrera 2B #69A – 32
Bogota D.C., Colombia
colombia@nommon.es

São Paulo
Av. Roque Petroni Junior,
1089 – Suite 713.
São Paulo – SP – Brazil
04707-900
brasil@nommon.es