



Gruppo FS

The Mobility Leader

6th International Symposium on High-Speed Rail Socioeconomic Impacts

The Catchment Area of High-Speed Rail Stations in Italy: An Analysis Based on Mobile Phone Big Data

Antonella Falanga, Ilaria Henke, Ennio Cascetta, Mario Tartaglia, Lorenzo Vannacci, Chiara Bonvicini, Luigi Galieni

6th International Symposium on High-Speed Rail Socioeconomic Impacts Kick-Off Meeting

Contents

Index

Introduction and Literature background

Context and Objectives

Research Design: Methodology and Results

Final Summary

Page 2

Page 3

Page 6

Page 10

Page 25



Gruppo FS

The Mobility Leader



Gruppo FS

The Mobility Leader

Introduction

Introduction

Limits of Traditional Demand Estimation

From Episodic Sampling to Multi-Source Big Data

- Episodic Sampling (the old approach)
- Episodic surveys (e.g. every 5–10 years): OD surveys, traffic counts, passenger interviews

Result: a "static snapshot" with several limitations:

- Does not capture seasonal variations or daily dynamics
- Limited use in rapidly changing contexts (e.g. post-pandemic, sharing mobility, smart city)
- Lag between data collection and availability: by the time data is ready, the context has already changed



Introduction

Limits of Traditional Demand Estimation

The Big Data "Revolution" (the new paradigm)

- **Continuous data collection** via sensors, electronic ticketing, mobile data, connected fleets, apps
- **Real-time monitoring** of transport system performance and mobility behavior

Advantages

- Enables **detailed analysis of a complex phenomenon** like road mobility, which lacks consolidated and reliable official sources for its various components
- Enables **adaptive planning**: supply is adjusted based on observed real demand
- Supports **ex post evaluations** and continuous tuning of transport models and policies
- **Enables early warnings** on congestion, demand drops, strikes, missed shifts, etc.



How can we increase its value?

Big Data applications in Transport Planning

- **Public transport OD estimation**

Smart card (AFC) and GPS vehicle data combined to estimate OD matrices for bus and metro systems. Applied to calibrate gravitational models, detect underserved areas and optimize network design

- **Gravitational demand model calibration**

MND-derived demand data used to calibrate spatial interaction models linking transport demand to population, economic intensity and accessibility. Enables scenario analysis for estimating potential demand for a new HSR station as a function of catchment population and competing alternatives

- **HSR station catchment area**

Fastweb MND used to identify origin municipalities of HSR users. Home location data and Google Maps API travel times allow construction of an empirical catchment area at municipal level, replacing costly OD surveys





Gruppo FS

The Mobility Leader

Context and Objectives



Gruppo FS

Research Centre

Context and Objectives

High-Speed Rail in Italy

Currently, Italy's HSR network extends over 1,100 km; the Milan–Bologna route, part of the TEN-T Mediterranean Corridor, is one of Europe's most important rail axes

Along the corridor, stations of different types are located:

- **in city centers** (e.g. Turin, Milan, Bologna, Florence, Rome)
- **within the urban area** (Turin Porta Susa, Milan Rogoredo, Florence Belfiore)
- **on the outskirts of major metropolitan areas**, mainly to capture road traffic from wide metropolitan sectors (e.g. Naples Afragola)
- **along the line near medium-sized cities** to intercept mainly road-based users, such as Reggio Emilia Mediopadana and the proposed Tuscia station (non-metropolitan stations)



Context and Objectives



Objective of the study

To assess the catchment areas of **high-speed rail stations in Italy**, comparing different station typologies (e.g. stations in city centers, within the urban area, on the outskirts of major metropolitan areas) in order to answer the following question

- ***Do all HSR stations in Italy have the same catchment area, or do stations in medium-sized cities and large metropolitan areas have different catchment areas?***



To use big data to reconstruct the current catchment areas of high-speed rail stations in Italy

It matters to know the real catchment area of an HSR station **to decide where to build new stations and how to design them**



Gruppo FS

The Mobility Leader

Research Design: Methodology and Results

Research Design: Methodology and Results

Mobile Phone Data as a Source for Mobility Analysis

What are Mobile Network Data?

Fastweb's phone data fall under the category of Mobile Network Data (MND), a type of Big Data generated by the ordinary functioning of mobile telephone networks. They are not collected for statistical purposes, but are a technical by-product of communication services (calls, messages, data traffic)

How they are generated:

- Every SIM card continuously interacts with mobile network antennas (passively or actively)
- Each interaction produces a signal containing: anonymized user ID, timestamp, antenna/cell location

Fastweb's Coverage4G

- coverage close to 99% of the population
- over 200,000 mobile cells across the national territory
- ~23 million "human" SIMs → tens of billions of daily records

Sample expansion:

Fastweb applies machine learning models, local market share weights and non-sampling bias corrections to expand the SIM sample to all mobile phone users aged 16–99

Research Design: Methodology and Results

Mobile Phone Data as a Source for Mobility Analysis

Data are **anonymized, aggregated and processed in compliance with GDPR**, making individual identification impossible. The analysis concerns collective mobility behavior, not individuals as single units

Masking (Privacy by Design):

- Process that automatically aggregates records with fewer than **15 users**
- Hierarchical mechanism: less critical variables are masked first, preserving readability of the most important dimensions
- Validation of presence in the station against user's mobility trajectory and crossed cells
- Determination of the mode of transport used to arrive at or depart from the station
- Raw data is then processed using standardized algorithms, generating various outputs that are useful for mobility analysis

How can we leverage these data to estimate the catchment area of a railway station?

Research Design: Methodology and Results

Mobile Phone Data as a Source for Mobility Analysis

The methodology is built around Reggio Emilia Mediopadana — a non-metropolitan HSR station for which both traditional OD survey data and Mobile Network Data are available — allowing cross-validation of the two approaches

Step 1 — Traditional data (OD surveys) provide the reference benchmark: municipality of residence, access mode, travel purpose and destination. An **empirical catchment area** is derived from the spatial distribution of users and their access times to the station

Step 2 — MND station visitor data, disaggregated by the visitor's municipality of origin, nationality, and the type of transport used to arrive at and depart from the station - distinguishing between arrivals/departures by high-speed train, other trains or other modes of transport

Step 3 — The two catchment areas are compared in terms of geographic extent, user profiles and demand decay with distance — assessing the **reliability of Big Data** as a standalone tool where traditional surveys are unavailable

Research Design: Methodology and Results

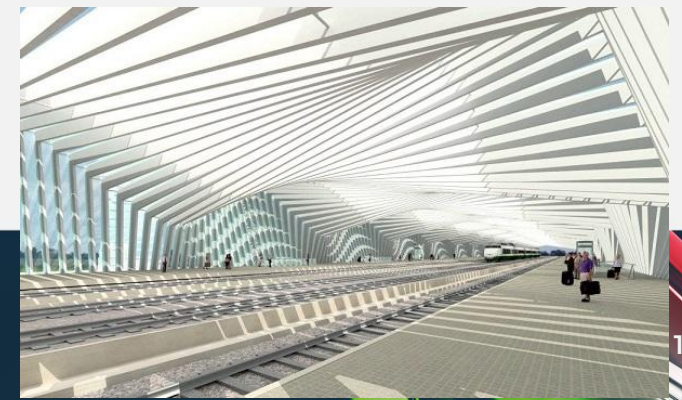
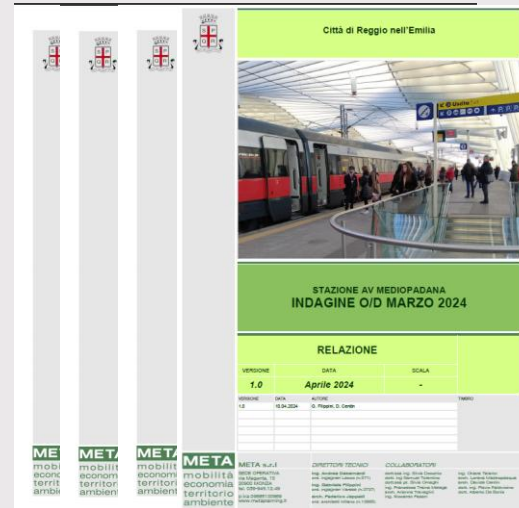
The Mediopadana HRS station

Designed by **Santiago Calatrava** and **opened in 2013** on the Milan–Bologna HSR line, **Reggio Emilia Mediopadana is located approximately 4 km from the city center**, outside the urban area, and conceived to intercept road-based traffic from a wide Po Valley catchment

Mediopadana is a **dedicated HSR-only station**, serving exclusively Frecciarossa and Italo services with no conventional rail traffic

Key context for catchment analysis, with impacts — metropolitan integration, tourism growth, urban regeneration — shaped by city size, station location and distance from the main metropolis

Annual surveys were conducted from 2016 to 2024 to monitor traffic flow and station supply



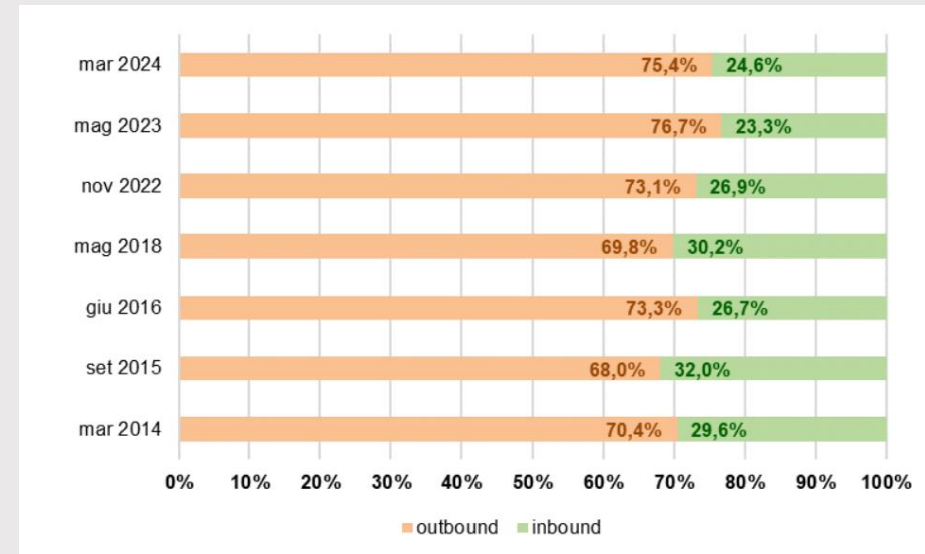
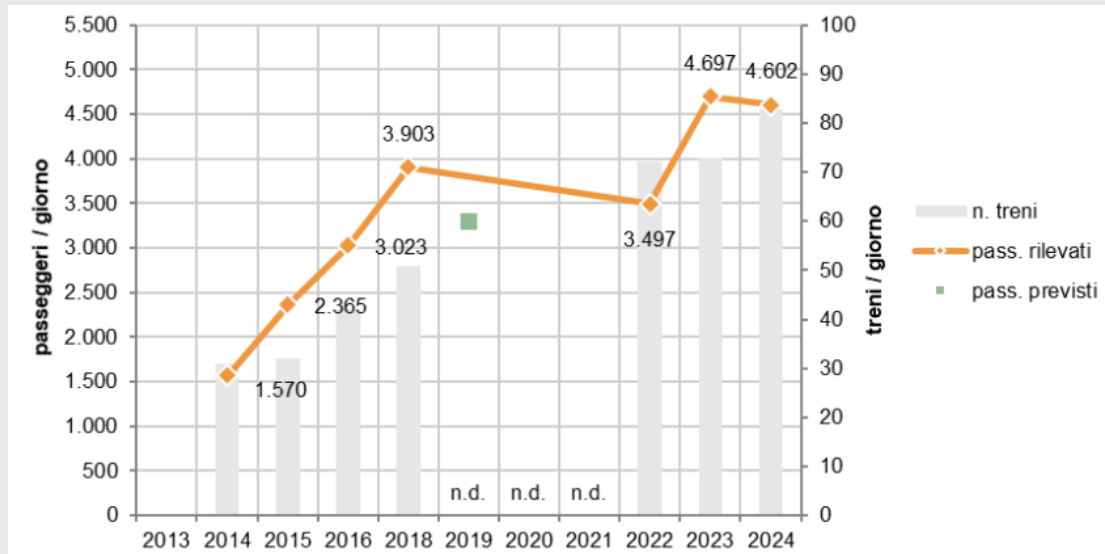
Research Design: Methodology and Results

The demand of the Medio Padana HSR station

In 2024 approximately **2 million passengers/year**

Between May 2018 and March 2024, passengers at Mediopadana station increased by **approximately 20%**

The station mainly serves residents of its area of influence (66%) rather than those from outside



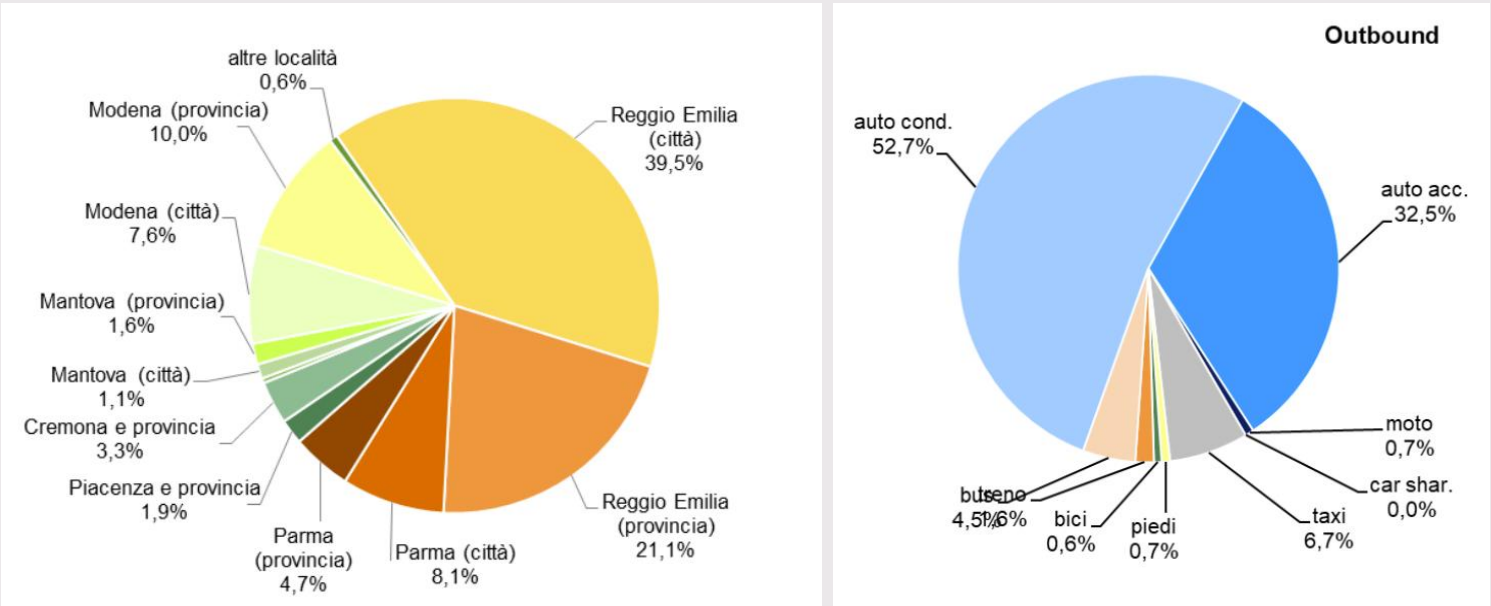
- **outbound trips:** trips made by residents of the Mediopadana area who go outside
- **inbound trips:** trips by residents outside the basin that go to the Mediopadana area

Source: OD surveys 2024, Meta srl, 2025

Research Design: Methodology and Results

The demand of the Medio Padana HSR station

The Mediopadana HSR station mainly serves the **city of Reggio Emilia and its hinterland**
Passengers by location of residence: **over 60% of users reside in the province of Reggio Emilia, 18% in Modena and the Province, 13% in Parma and the Province, 3% in Cremona, 2% in Piacenza, 3% in Mantua and 1% in other locations**



Self-mainway of access (85%) , Minority public transport (6.1%)

Source: OD surveys 2024, Meta srl, 2025

Research Design: Methodology and Results

Data source

Based on home location data, we distinguished **outbound and inbound trips***

Travel times to the station were computed via **Google Maps API**, considering car travel with the fastest route on a typical weekday, during both peak and off-peak hours

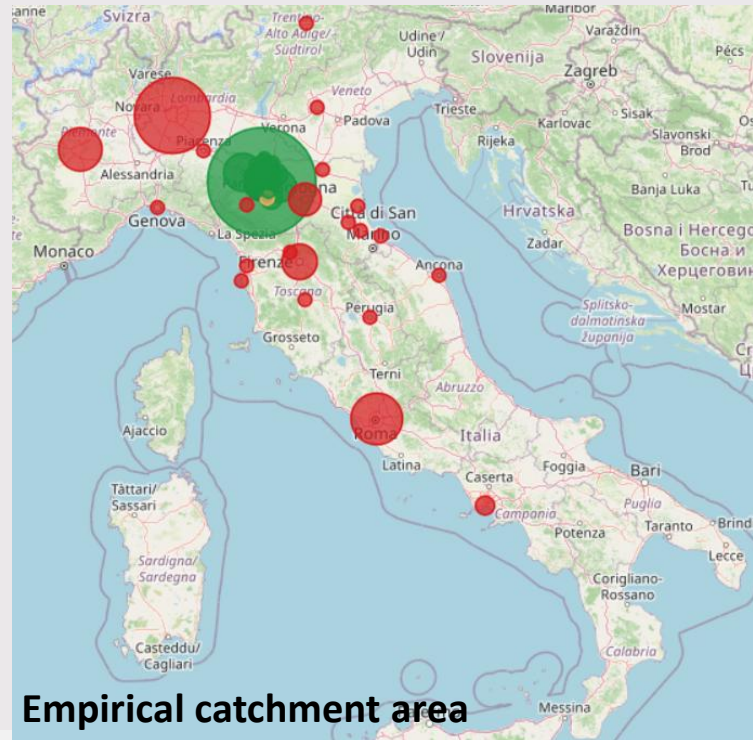
```
Comuni con tempo trovato: 251
Premi Invio per chiudere...

(base) C:\Users\39340\Desktop\ANTONELLA\Lavori 2026\Alto-Padana stazione\6th
=====
Calcolo tempi percorrenza -> 44.6918,10.6336
=====

Lettura file Excel: ReggioEmiliaOttobre2025.xlsx
Comuni da interrogare: 251

Inizio chiamate API (puo' richiedere 5-10 minuti)...

[1/251] Alba Adriatica (Abruzzo)
punta: 221 min | morbida: 232 min
[2/251] Atessa (Abruzzo)
punta: 300 min | morbida: 313 min
[3/251] Chieti (Abruzzo)
punta: 267 min | morbida: 279 min
[4/251] L'aquila (Abruzzo)
punta: 281 min | morbida: 292 min
[5/251] Pescara (Abruzzo)
punta: 266 min | morbida: 274 min
[6/251] Vasto (Abruzzo)
punta: 304 min | morbida: 315 min
[7/251] Matera (Basilicata)
punta: 464 min | morbida: 476 min
[8/251] Castrovillari (Calabria)
punta: 502 min | morbida: 506 min
[9/251] Fuscaldo (Calabria)
punta: 568 min | morbida: 562 min
[10/251] Lamezia Terme (Calabria)
punta: 597 min | morbida: 592 min
[11/251] Ariano Irpino (Campania)
punta: 422 min | morbida: 433 min
[12/251] Arzano (Campania)
punta: 370 min | morbida: 382 min
[13/251] Boscoreale (Campania)
punta: 380 min | morbida: 395 min
[14/251] Caivano (Campania)
```



- ***outbound trips** are journeys made by people residing in the region where the HSR station is located (e.g. for Mediopadana AV: users residing in Reggio Emilia)
- **inbound trips** are journeys made by people residing outside the catchment area who travel to the area where the HSR station is located

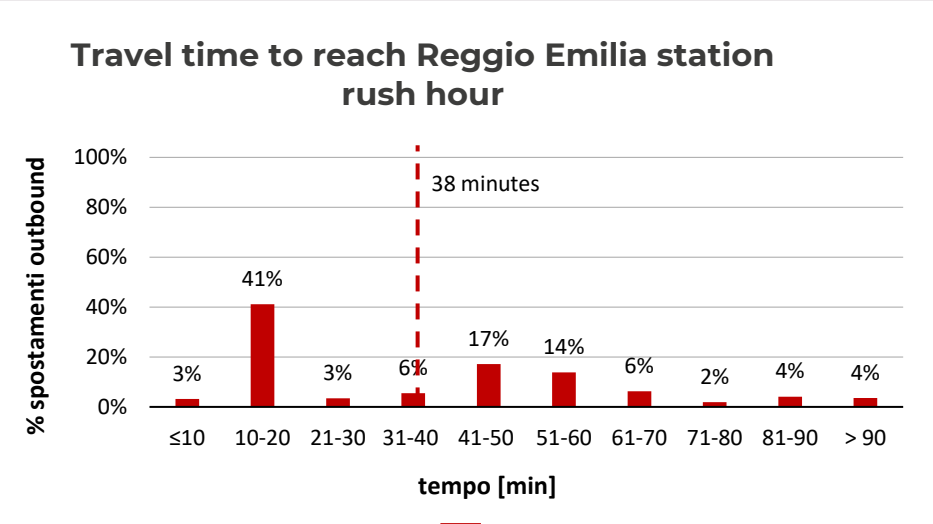
Research Design: Methodology and Results

Data source

Based on the survey results, it was possible to estimate the distribution of access times to Reggio Emilia HSR station

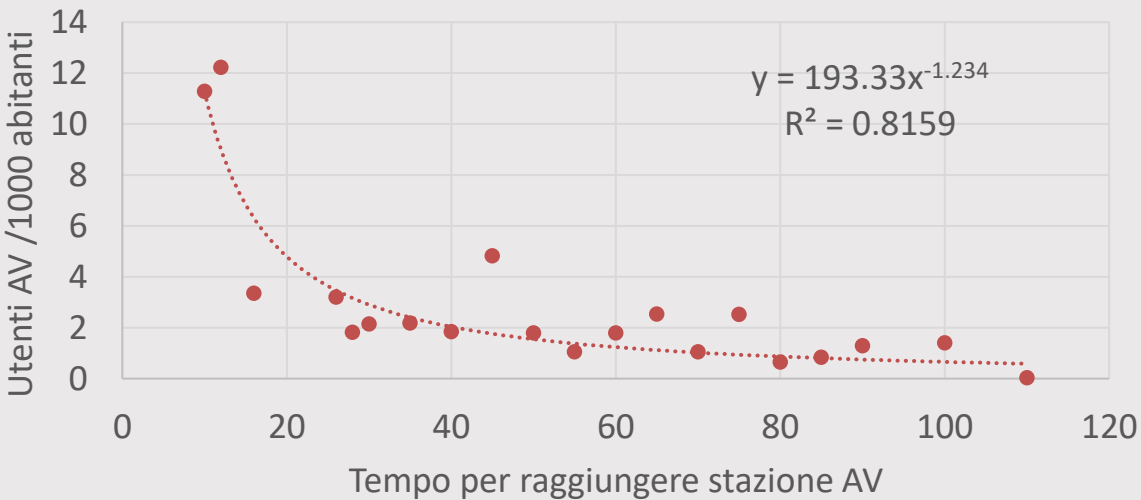
Outbound trips

travelers who boarded by municipality of origin



average time = 38 minutes
Within 60 minutes 84% of travelers

Strong correlation between access time to the HSR station and the HSR usage rate in each municipality of residence (users per 1,000 inhabitants)

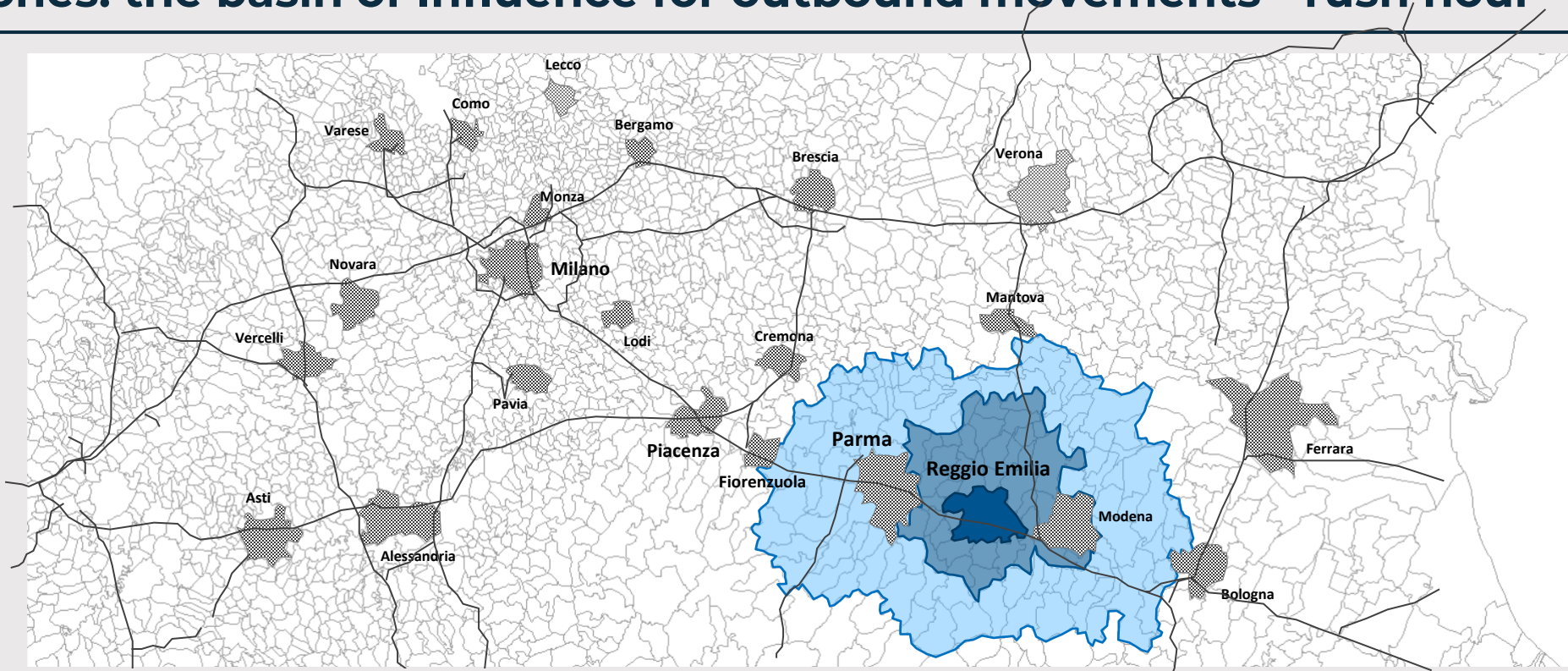


¹Travel time was estimated via Google Maps considering car travel with the fastest route on a typical weekday, departing at 8:00 am

Source: Authors' elaborations based on OD survey data 2024, Meta srl, 2025

Research Design: Methodology and Results

Isochrones: the basin of Influence for outbound movements - rush hour



- Time ≤ 40 min from Reggio Emilia HSR station
- $41 \leq \text{Time} \leq 60$ min from Reggio Emilia HSR station
- Highway network

OUTBOUND TRIPS		
HSR staion	population served Isochronous 40	population served Isochronous 60
REGGIO EMILIA	about 850,000	Over 1 Mln

Research Design: Methodology and Results

Choice of the model

Specification of a **gravity model** to estimate for each Municipality i the number of users of the Reggio Emilia high-speed station

Y_i = number of HSR travelers observed to/from the Municipality i towards Reggio Emilia station

The probability of observing y_i HSR users in Municipality i is described by the function:

$$P(Y_i = y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

μ_i represents the expected average number of HSR users of Municipality; i depends on the characteristics of Municipality i and its distance from the Reggio Emilia station

Parameter estimation was performed using the Poisson Pseudo Maximum Likelihood (PPML) approach, which allows a robust estimate of the expected mean of the counts even in the presence of overdispersion and many zeros

Research Design: Methodology and Results

Specification of the model

$$\log(\mu_i) = \beta_0 + \beta_1 \log(Pop_i) + \beta_2 \log\left(\frac{Employees_i}{Pop_i}\right) + \beta_3 T_{Reggio,i} + \beta_4 C_i$$

Where:

Pop_i = population of the municipality i

$Employees_i$ = number of employees in the municipality i

 $\left(\frac{Employees_i}{Pop_i}\right)$ represents the **economic intensity** of the municipality

$T_{Reggio,i}$ = access time to the Reggio Emilia Medio Padana HSR station

$C_i = \log\left(\frac{T_{RE,i}}{\min_{j \in A} T_{j,i}}\right)$  It measures Reggio Emilia's **competitive advantage over other HSR stations in the area** (in this case we considered Rogoredo and Bologna)

Where:

$T_{RE,i}$ = time from municipality i to Reggio Emilia

A = set of alternative stations

$T_{j,i}$ = time from municipality i to alternative station j = best available alternative

$C_i > 0 \rightarrow$ Reggio Emilia is **worse** than the best alternative

$C_i = 0 \rightarrow$ Reggio Emilia is **equivalent**

$C_i < 0 \rightarrow$ Reggio Emilia is **faster** (competitive advantage)

Research Design: Methodology and Results

Results of the model

Parameter	Coefficient	Robust Std. Err.	z	p-value
Intercepts	-10.19	2.35	-4.33	0.00
log(Pop)	1.41	0.12	11.94	0.00
log(Employees/Pop)	1.07	0.48	2.23	0.03
Time _{Reggio}	-0.04	0.02	-1.43	0.15
C	-2.56	0.71	-3.59	0.00

Population: HSR demand grows with the demographic size of the municipality

Employees/Population: For the same population, municipalities with greater economic intensity generate more HSR demand

Time to Reggio: As station access time increases, HSR demand decreases significantly

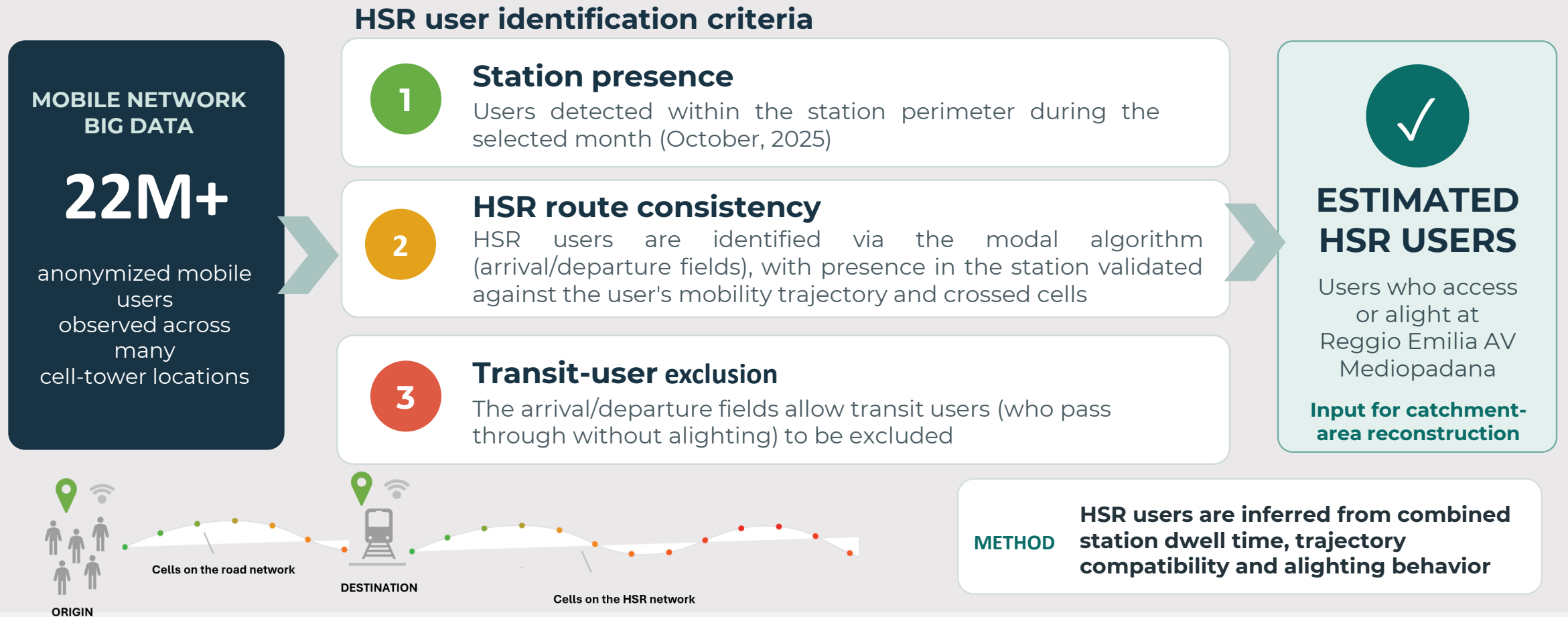
Competition with other high-speed stations: As the relative access time to Reggio increases compared to the best alternative, demand decreases, highlighting a territorial competition effect between stations

Model indicators

Number of observations	262
Total observed	4341
Total estimated	4341
Log-likelihood	-1390,65
Pseudo R2 (McFadden approx.)	0,89

Research Design: Methodology and Results

Analysis From Mobile Network Data to HSR Demand



Research Design: Methodology and Results

Analysis From Mobile Network Data to HSR Demand

Medio Padana HSR Station

Indicator	Traditional data (OD survey)	Mobile Network Data
HSR Users	111,714 survey respondents	113,932 monthly unique visitors (Oct 2025)
Residents of Emilia Romagna	75.4%	77.8%

Key finding: The two sources yield consistent geographic catchment patterns (Δ = +2.4 pp), validating Mobile Network Data as a reliable standalone tool for catchment area estimation where traditional surveys are unavailable



Gruppo FS

The Mobility Leader

Final Summary

Key Findings

- **A wide, multi-provincial catchment area:** Reggio Emilia Mediopadana draws users well beyond its own province — 60% from Reggio Emilia, 18% from Modena, 13% from Parma, and the remainder from Cremona, Piacenza, Mantua and other areas
- **Access time is the main driver:** average access time to the station is 38 minutes, and 84% of travelers reach it within 60 minutes, mostly by car
- **The gravity model confirms the key drivers of demand:** HSR demand grows with population and economic intensity, and falls with access time and with competition from alternative HSR stations (PPML estimation, Pseudo $R^2 \approx 0.89$)
- **Reggio Emilia as a validated case study:** built on the annual OD surveys conducted from 2016 to 2024, this validated case study provides a solid, replicable foundation for the catchment-area analysis

Contribution and Next Steps

- **Methodological contribution:** integrating a detailed OD survey with a gravity model estimated via Poisson Pseudo Maximum Likelihood provides a robust, well-identified estimate of a non-metropolitan HSR station's real catchment area
- **Policy relevance:** understanding the real catchment area of a non-metropolitan HSR station supports better decisions on where to build new stations and how to design their access systems
- **Next step — a Big Data-based validation:**
 - 1) A procedure based on Mobile Network Data will be tested as a near real-time complement to traditional surveys, validated on Reggio Emilia against the OD survey results already obtained
 - 2) By complementing survey-based evidence, Mobile Network Data enable a more realistic and near real-time understanding of where people are and how they move, overcoming many of the limitations of self-reported data



Gruppo FS

The Mobility Leader

Thank you

#FSResearchCentre

<https://www.fsitaliane.it/content/fsitaliane/en/fs-research-centre.html>



Gruppo FS

Research Centre