Integrating Mobile Network Data with Official Mobility Statistics: The DREAM_PACE and MIND Data Fusion Framework for the Bologna Metropolitan Area

Daniele Di Antonio[1], Francesco Ciaffi[1], Simone Porcacchia[1], Daniel Nori[1], Giorgio Fiorillo[2], Eugenio Margelli[2], Marco Amadori[2]

[1] GO-Mobility S.r.l., Roma, Italy
[2] SRM Reti e Mobilità S.r.l., Bologna, Italy

This is the Accepted Manuscript of an article published in Statistical Journal of the IAOS (Sage). The final published version is available at https://doi.org/10.1177/18747655261495230.

Citation: Di Antonio D, Ciaffi F, Porcacchia S, Nori D, Fiorillo G, Margelli E, Amadori M. Integrating mobile network data with official mobility statistics: The DREAM_PACE and MIND data fusion framework for the Bologna metropolitan area. *Statistical Journal of the IAOS*. First published online 9 October 2026. DOI: 10.1177/18747655261495230.

This version incorporates one correction made at proof stage: the basemap attribution in the Figure 1 caption.

The DREAM_PACE project (CE0100046) was implemented through financial assistance from ERDF funds of the Interreg CENTRAL EUROPE Programme. The MIND project (CUP F55G24000020004) is co-funded by the European Union under the Emilia-Romagna ERDF Regional Programme 2021–2027. This publication reflects only the authors' views; the programme authorities are not liable for any use that may be made of the information contained therein.

Integrating Mobile Network Data with Official Mobility Statistics: The DREAM_PACE and MIND Data Fusion Framework for the Bologna Metropolitan Area

Keywords: *mobile network data; floating car data; data fusion; official statistics; demand-responsive transport; low-demand areas; SUMP*

Abstract

This paper presents a replicable framework for integrating Mobile Network Data (MND) with Floating Car Data (FCD), an ad hoc mobility survey and official administrative registers to support mobility statistics and transport planning. Developed within the DREAM_PACE and MIND projects and applied to the Metropolitan City of Bologna (Italy), the framework uses a structured data-fusion pipeline to produce origin–destination matrices, distinguish residents from non-resident users and estimate the intra-zonal trip component through calibrated gravity models. Indicators of potential trip production and attraction are used to classify the analytical zones by mobility demand intensity, identifying low-demand areas through a data-driven operationalisation of the criteria set by Italian national guidance (DM 157/2018) and the Transport Regulation Authority Resolution No. 48/2017. Fused estimates are used to define service levels for demand-responsive transport (DRT) consistent with the metropolitan Sustainable Urban Mobility Plan (SUMP). The integration of multiple data sources reveals approximately 3.7 million weekday trips in November 2024 – approximately 1.17 million more than previous survey-based estimates – largely because it explicitly accounts for non-resident and crossing trips. We discuss implementation choices, quality safeguards and limitations, and provide a modular workflow that

National Statistical Offices and local authorities can adapt to produce timely, policy-relevant mobility indicators.

1. Introduction

Official statistics on mobility and transport demand are increasingly expected to capture spatial and temporal dynamics with a level of detail that traditional survey-based systems alone cannot provide. National and local authorities need up-to-date information to design and monitor Sustainable Urban Mobility Plans (SUMPs), to assess the performance of public transport services and to plan targeted interventions in peripheral and low-demand areas. At the same time, digital traces generated by private actors – such as mobile network operators and mobility service providers – have become widely available, opening new opportunities for enhancing the evidence base of public decision-making [1,2].

Mobile Network Data (MND) offer a particularly valuable perspective because they provide large-scale, continuous coverage of movements for broad segments of the population [3,4]. However, MND are not collected for statistical purposes and present well-known challenges: market-coverage bias, the need to infer mobility patterns from raw network activity and the inherent inability to detect intra-zonal movements occurring within a single spatial unit of the operator's grid [5]. To be used in support of official statistics, MND-based indicators must therefore be embedded in a transparent framework that clarifies assumptions, quantifies uncertainty and exploits complementarities with traditional sources [6,7].

The work described in this paper was developed within two complementary project frameworks. The DREAM_PACE project (Demand-REsponsive trAnsport integrating regional Mobility networks for passengers in Central Europe), an Interreg Central Europe initiative, provides the

methodological and analytical core: the data-fusion pipeline, the classification of low-demand areas and the definition of DRT service levels [8]. The MIND project (Mobility INsight Data), funded under the Emilia-Romagna ERDF Regional Programme 2021–2027, complements DREAM_PACE by developing the data-management infrastructure and the monitoring platform that enable the integration, update and dissemination of mobility indicators for planning and contract management [9]. Together, these projects explore how MND, Floating Car Data (FCD) and an ad hoc mobility survey can be combined to produce robust and policy-relevant mobility indicators for the Metropolitan City of Bologna.

The combined effort focuses in particular on the identification of low-demand areas and on the definition of appropriate service levels for demand-responsive transport (DRT), in line with the national regulatory framework [10,11] and the objectives of the Bologna SUMP [12]. The paper proposes a practical, quality-assured framework that can be replicated in other territories and adopted by National Statistical Offices and local authorities.

The paper is structured as follows. Section 2 introduces the study area and the main data sources used in the analysis. Section 3 summarises the regulatory and policy context. Section 4 describes the methodological framework, including the data-fusion pipeline, the estimation of the intra-zonal trip component, the construction of mobility demand indicators and the identification of DRT service levels. Section 5 presents the main results, while Section 6 discusses validation and quality assessment. Section 7 reflects on the implications for official statistics and for local decision-makers, with specific attention to operational replicability and the potential for continuous mobility monitoring, and Section 8 concludes with key messages and directions for further work.

2. Data and Sources

2.1 Study area

The study area is the Metropolitan City of Bologna, a large urban and peri-urban territory in central-northern Italy that covers approximately 3,700 km² and includes the municipality of Bologna – the regional capital with approximately 400,000 inhabitants – together with a wide set of surrounding municipalities. The territory extends along an altitudinal and functional gradient from the Po Valley plain in the north to the Tuscan-Emilian Apennines in the south, passing through a central hilly belt. Three broad morphological belts can be distinguished: the plain (approximately 40% of the territory), hosting the main production centres and population concentrations; the hill area south of the Via Emilia, with mixed residential-agricultural land use; and the Apennine area (approximately 35% of the territory), characterised by low population density and settlements distributed along valley floors.

This territorial heterogeneity is reflected in travel patterns and in the structure of the public transport supply: from high-frequency urban services in the core to sparse and often inadequate connections in the peripheral and mountainous areas. The settlement system is polycentric, with strong functional interdependencies along the east-west Via Emilia corridor and a marked urban polarisation effect generated by the municipality of Bologna.

The Metropolitan City has a consolidated tradition of strategic mobility planning. The local SUMP, first adopted in November 2018 and approved by the Metropolitan Council in November 2019, defines objectives and targets in terms of accessibility, environmental sustainability and quality of public transport services [12]. The SUMP is currently undergoing an update phase launched in 2026 to align with the new public transport tendering cycle. Within this framework, the Metropolitan City and the local public transport authority are required to identify areas with weak

transport demand and to design appropriate solutions, including DRT services, in accordance with national guidelines. The study area and the analytical zoning are shown in Figure 1.

2.2 Spatial reference systems and terminology

The sources combined in this study are collected, delivered and processed on different spatial units, and the generic term "zone" would therefore be ambiguous. Four of these units form a single chain of successive aggregations, each level being obtained by grouping whole units of the level below: census cells are aggregated into analytical zones, analytical zones into MND zones, and MND zones into operator zones. Two further systems describe the administrative and survey geography: the municipalities of the metropolitan area and the macro-zones of the household mobility survey. The cell catchment areas of the mobile network operator constitute the underlying physical geography of the MND and do not coincide with any of the systems above. The seven systems are defined in Table 1 and are referred to throughout the paper by the names given there. Unit counts are reported both for the Metropolitan City of Bologna and for the full extraction perimeter, which additionally covers an external ring representing the territory outside the metropolitan boundary.

Table 1. Spatial reference systems used in the study.

System	Units within the Metropolitan City	Total units, incl. external ring	Definition	Role in the analysis
MNO cell catchment areas	n.a.	n.a.	Probabilistic service areas of the operator's radio cells. Radii of 400–800 m in built-up areas and of the order of kilometres in low-density areas; not uniform, and including enclaves and exclaves. Attribution of a device to an area is probabilistic, not deterministic	Underlying physical geography of the MND. Not used as an analytical unit
Census cells	10,683	n.a.	ISTAT census enumeration areas	Building blocks of the analytical zones; carriers of the socio-demographic attributes
Analytical zones	1,050	1,113	Aggregation of census cells, co-produced with the metropolitan transport authority and calibrated for relative	Finest analytical resolution. Level at which the gravity models are applied and at which

System	Units within the Metropolitan City	Total units, incl. external ring	Definition	Role in the analysis
			internal homogeneity of socio-economic characteristics	the demand indicators are computed
MND zones	144	207	Aggregation of analytical zones. The transport zoning system set out in the extraction specification submitted to the operator	System on which the MND analysis is conducted and on which MND-based results are reported
Operator zones	133	192	Further aggregation of MND zones, imposed by the operator for those zones where the coverage verification was not passed	System on which the MND were physically delivered, and the level at which trip detection operates: movements that begin and end within the same operator zone are excluded by construction
Municipalities	55	n.a.	Municipalities (comuni) of the Metropolitan City of Bologna	Level at which the gravity models are calibrated and at which the DM 157/2018 classification operates
Survey macro-zones	16	17	Territorial aggregations adopted by the household mobility survey. They follow municipal boundaries except in three cases: Bologna is split into a central and a peripheral macro-zone, and the municipalities of Mordano and Valsamoggia are each divided between two macro-zones. A seventeenth aggregation represents the territory outside the metropolitan area and serves only to identify trips to and from the exterior; it is not an analysis unit	Level at which the survey is designed to be representative, and at which trip rates and the distributions by purpose and mode are applied

Note to Table 1. Analytical zones and MND zones lie entirely on one side of the metropolitan boundary. One operator zone straddles it and is counted with the external ring, so that the 192 operator zones comprise 133 lying entirely within the Metropolitan City and 59 outside or astride its boundary. The relationship between the MND zones and the operator zones is directional and is easily read the wrong way round. The 207 MND zones are the zoning system originally requested in the extraction specification. The operator then verified radio coverage zone by zone and, where the verification was not passed, required further aggregation; the resulting 192 operator zones are the system on which the data were physically delivered. The zones that had been aggregated for coverage reasons were subsequently re-split through an FCD-based data-fusion procedure, restoring the 207 MND zones on which the analysis operates. The sequence is therefore 207 → 192 → 207: the finer system is the original request, degraded by the operator under coverage constraints and reconstructed by the authors, and not a refinement of a coarser delivery.

A number of technical terms recur throughout the paper with meanings that depend on these spatial systems. They are defined in Table 2 and used consistently thereafter.

Table 2. Definitions of the technical terms used in the paper.

Term	Definition
Zone / area	A zone is a unit of one of the zoning systems of Table 1 and is always named by the system it belongs to. An area is an aggregation of contiguous zones identified by the analysis and

Term	Definition
	having no independent status in any zoning system; the 23 low-demand areas of Sections 4.4 and 5 are of this kind
Macro-components	The seven-way partition of total metropolitan demand used throughout the paper as the basis for comparison across sources: trips internal to the municipality of Bologna; intra-municipal trips in the other municipalities; exchanges between Bologna and the rest of the metropolitan area; inter-municipal trips not involving Bologna; exchanges between the rest of the metropolitan area and territories outside it; exchanges between Bologna and territories outside the metropolitan area; and crossing trips
Self-contained trips	Trips whose origin and destination fall within the same municipality, comprising the first two macro-components above. Self-contained trips are defined at the level of the municipalities and are not to be confused with intra-zonal trips, which are defined at the level of the MND zones (Section 4.2)
Exchange components	The macro-components in which exactly one endpoint falls within the reference area. The term is used with respect to two reference areas, the municipality of Bologna and the Metropolitan City, which are always named explicitly, since the same trip may be an exchange with respect to one and an internal movement with respect to the other
Crossing trips	Trips that traverse the metropolitan territory with both origin and destination outside it. They lie outside the sampling frame of a household survey by construction, and are observable here because the MND user-type classification identifies them explicitly
Mobility demand intensity	The demand density of an analytical zone, measured as the number of trips generated and attracted on a typical weekday per km². It is the variable on which the classification into mobility demand categories is based, and it is the input to the definition of the DRT service levels (Table 5)
Link-level indicators	Indicators referred to the individual links of the road network: travel time, average and free-flow speed, and derived congestion measures, computed from map-matched FCD trajectories, as distinct from zone-based indicators and from indicators referred to origin–destination pairs

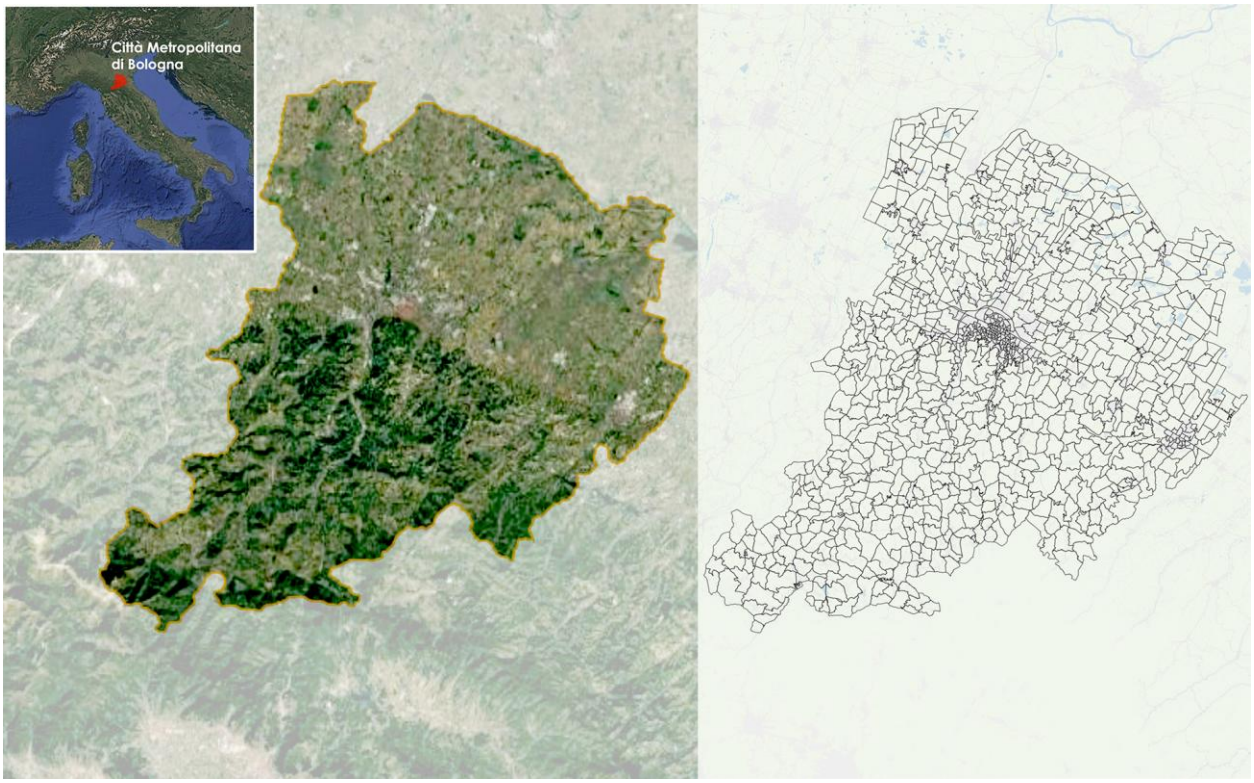


Figure 1. Study area: the Metropolitan City of Bologna (55 municipalities) and the analytical zoning, with 1,050 internal analytical zones and 1,113 including the external ring. Basemaps: Google Maps satellite imagery, left panel and inset (Imagery ©2026 Landsat/Copernicus, Data SIO, NOAA, U.S. Navy, NGA, GEBCO; Map data ©2026 Google); OpenStreetMap, right panel (© OpenStreetMap contributors).

2.3 Data overview

The analysis relies on four main types of data: Mobile Network Data, Floating Car Data, an ad hoc mobility survey and official administrative data. Table 3 summarises the four sources and their role in the workflow.

Mobile Network Data (MND). MND were provided in the form of aggregated, anonymised origin–destination matrices describing the number of trips between operator zones within the study area, between the study area and an external ring, and crossing trips through the territory. The data were acquired from a major national mobile network operator with a market share covering approximately 30% of the population. The dataset distinguishes different user groups through a residency and user-type classification (residents, commuters, visitors and crossing users) and

provides temporal detail by hour of day, day type (weekday, Saturday, Sunday) and month, covering May, July and November 2024, enabling the analysis of seasonal variations and the identification of event-specific patterns. A key characteristic of MND is that trip detection relies on algorithms that identify changes in the user's location between the operator's cell catchment areas (Table 1). Movements that begin and end within the same operator zone are therefore excluded by construction, as they do not satisfy the stationarity thresholds applied in the extraction process; the configuration of those thresholds is described in Section 2.4 and the recovery of this component in Section 4.2.

Floating Car Data (FCD). FCD were used as a complementary source for characterising private vehicle movements. The dataset consists of anonymised GPS traces from a sample of connected vehicles (approximately 4% penetration rate) circulating on the road network of the metropolitan area during October 2023, the most recent period for which the source was available at the time of the analysis. Raw GPS data were processed into trajectories and trips, from which indicators were derived at three levels of aggregation: on the links of the road network; on zones, as generated and attracted vehicle flows, aggregated to the spatial system each use requires; and on origin–destination relations. The link-level indicators of Table 2 are travel time, average and free-flow speed, and derived congestion measures. These indicators complement the person-based picture obtained from MND and survey data.

In the data fusion, FCD play three roles. First, at a highly aggregate level, provincial totals and macro-components, they serve as an independent benchmark for overall trip volumes. Second, the origin–destination indicators provide an independent measurement of the intensity of mutual exchange between adjacent operator zones, relative to the generative and attractive potential of each; this is what allows outliers in the MND matrices to be identified and corrected on the operator

zones as delivered, before the zonal re-split. The outliers concerned, commonly referred to as "bounce" or "ping-pong" effects, are the spurious movements generated when a stationary device is alternately attributed to neighbouring cells; they inflate the exchange between adjacent operator zones and are otherwise indistinguishable from genuine short-range mobility, particularly in dense urban areas. Third, FCD provide the basis of the re-split procedure that restores the 207 MND zones from the 192 operator zones delivered, as described in Section 2.4.

Mobility survey. An ad hoc household mobility survey was first conducted in 2016, as the basis of the metropolitan SUMP, and updated in 2023. It provides detailed information on trip purposes, modes of transport and the socio-demographic characteristics of respondents. For the purpose of this work, the travel behaviour recorded in the 2016 wave, trip rates and the distributions by purpose and mode, is adopted throughout, so as to remain consistent with the planning instruments built on it; origins and destinations, and the volume of non-resident users, are instead re-estimated from the mobile network data, independently of any change in travel behaviour. The 2023 wave allows that change to be assessed, but such a comparison is outside the scope of this paper. The 2016 wave targeted the resident population aged 14 to 85 and collected information on daily trip chains, including origin, destination, mode, purpose and time of day; its design ensures statistical representativeness at the level of the 16 survey macro-zones, with a sample of 5,500 interviews. Trip purposes were grouped into five categories: work, study, shopping, social activities and other. From a territorial perspective, the survey provides the municipality-level totals of self-contained flows (intra-municipal trips) that MND cannot observe, on which the gravity models are calibrated (Section 4.2; Appendix A).

Official administrative data. Survey data were expanded to the resident population using auxiliary information from official statistics, in particular the Metropolitan Statistical Atlas

("Atlante Metropolitano") and demographic small-area statistics at census-cell level produced by the Italian National Statistical Institute (ISTAT). For each of the 1,050 analytical zones within the Metropolitan City, the resident population was disaggregated by age group, gender and employment status, creating a multifactorial description of the potential demand for mobility. Census cells were used as the building blocks for constructing the analytical zoning that balances the need for spatial detail with the requirement to preserve statistical consistency.

The analytical zones were obtained by aggregating the 10,683 ISTAT census cells according to five criteria, applied in the following order of precedence: (i) *administrative containment* — each analytical zone lies within a single municipality, so that the system nests exactly into the 55 municipalities; (ii) *density gradients* — zone boundaries were placed along local minima of residential and productive settlement density, so that each zone is internally homogeneous in settlement type; (iii) *consistency with the transport network and physical barriers* — boundaries follow major infrastructure and natural barriers rather than cutting across them; (iv) *purpose-driven resolution* — because the study focuses on low-demand areas, these were zoned at proportionally finer resolution than the major urban centres, so that spatial detail is concentrated where the analysis requires it; (v) *statistical viability* — the total number of zones was kept low enough for the cell counts entering the data-fusion process to remain reliable. The zoning was co-produced with SRM Reti e Mobilità, the transport authority of the Metropolitan City, so that local knowledge of settlement patterns and service catchments could inform boundary placement rather than being applied mechanically.

The national register of vehicles maintained by the Automobile Club of Italy (ACI) was used to perform a first expansion step of the FCD, by matching the sample with flow statistics from the regional traffic counting system.

2.4 MND spatial units and preprocessing

Spatial units and the aggregation chain. The native spatial units of the mobile network data are the operator's cell catchment areas, whose properties are set out in Table 1. Those properties matter here for a single reason: irregular extent varying with the degree of urbanisation, discontinuous shape with enclaves and exclaves, and probabilistic rather than deterministic attribution of a device to a unit together mean that the native geography does not constitute a partition of space in the sense that a transport zoning system does. It cannot therefore serve as the spatial reference of an analysis whose results must be attributable to identifiable and stable portions of territory. This is why the analysis is not conducted on the units on which the phenomenon is physically observed, and why the extraction was specified on a zoning system defined for the purpose.

The history of that system is summarised in the note to Table 1; what it involves at each step is the following. The operator verified radio coverage for each of the 207 requested MND zones and, where the verification was not passed, required the zone to be merged with one or more adjacent zones; the 192 operator zones on which the data were delivered are the result. The aggregation is limited in extent but unevenly distributed. One hundred and eighty of the 192 operator zones coincide with a single MND zone; the remaining twelve comprise 27 MND zones between them, as eleven pairs and one grouping of five. Seven of the twelve lie entirely within the metropolitan area and comprise 17 of its 144 MND zones, and four of those seven, including the grouping of five, fall in Apennine municipalities, where cell coverage is sparsest. The merging preserves the boundaries of the MND zones: no MND zone is divided between two operator zones. It does not preserve administrative boundaries: whereas the MND zones nest exactly within the 55 municipalities, four of the operator zones lying within the metropolitan area extend across more than one. The aggregation imposed at this step is thus the only point in the chain at which the

correspondence between the analytical and the administrative geography is broken, and restoring the MND zones restores it.

Three operations are then performed, in a fixed order. First, the delivered flows are examined for outliers at the level at which they are delivered: FCD-based indicators of mutual exchange between adjacent operator zones, related to the generative and attractive potential of each, identify anomalous over-representation of short-distance flows: the residual "ping-pong" effect, in which a stationary device alternately attributed to neighbouring cells produces spurious movements. These are corrected. Second, the operator zones that had been merged are re-split into the MND zones they comprise: the flow delivered for an aggregated operator zone is apportioned among its constituent MND zones in proportion to the generation and attraction weights obtained for each of them from FCD, relative to the operator zone total. Third, the analysis proceeds on the 207 MND zones. The order is not incidental: the artefact arises in trip detection, which operates on the operator zones, so correcting it at that level prevents it from being carried into the re-split, and it is neither addressed nor re-introduced further downstream.

The second of these operations is ours and not the operator's, and we treat it as such. It recovers a component that would otherwise be lost, namely the flows between distinct MND zones belonging to the same operator zone, which the delivered matrices do not distinguish. It recovers it, however, by estimation rather than by observation, and it introduces uncertainty in the correspondence between the geography on which the data were delivered and the geography on which they are analysed. Both FCD-based operations, moreover, rest on data covering October 2023, the most recent period for which floating car data were available: the parameters they yield were estimated and validated on that period and applied to the observation months of the mobile network data.

We regard the re-splitting and the temporal transfer of these parameters as sources of uncertainty in their own right, and return to both in Section 7.1.

Stationarity thresholds. The mobile network data used in this study were delivered as pre-processed origin–destination matrices, and two layers of that pre-processing must be distinguished. The first is algorithmic: signal cleaning, the handling of coverage interruptions and the attribution of devices to radio cells are performed by the operator through proprietary procedures to which we have no access and which we are therefore unable to report. The second is parametric: the stationarity thresholds that separate movement from stationary presence, and which consequently determine what is registered as a trip, were not delegated to the operator but configured jointly with it. This subsection reports the second layer.

The thresholds are not a single global value. They are defined at the level of the operator zones, the level at which trip detection operates (Table 1), and were differentiated as a function of two characteristics of each zone. The first is spatial extent: the larger the zone, the longer the threshold must be, since a device may legitimately require more time to traverse it and an insufficient threshold would fragment a single movement into several. The second is settlement type: the denser the zone, the shorter the threshold must be, since a long threshold suppresses genuine short-range movements, which in dense areas account for a substantial share of mobility. The two requirements act in opposite directions, and the configuration consists in resolving them zone by zone. How a zone-level threshold is subsequently applied to the underlying cell catchment areas belongs to the operator's algorithmic layer and is not known to us.

The configuration followed the same sequence as the zoning chain described in Section 2.2. Thresholds were first set on the 207 MND zones of the extraction specification. Where the coverage verification required further aggregation, the threshold of the resulting operator zone was

re-adapted to the extent and settlement type of the aggregated unit; where an operator zone coincides with a single MND zone, the initial configuration was retained.

The retained configuration was determined iteratively rather than analytically. Successive threshold configurations were generated with the operator and the resulting matrices compared against a set of external benchmarks: aggregate trip generation rates and their relative ordering across population strata, derived from the mobility survey; the age structure of the observed mobile population, compared with official demographic statistics; gross trip totals at aggregate level; and origin–destination matrices previously produced and validated for territorial contexts comparable in geographical setting and in the size of the zoning units. The configuration retained is the one that minimised divergence from this set as a whole. Comparison at this stage is made on aggregate quantities; the by-stratum comparison is used as a relative check on internal coherence, that the mobility rates of the different strata stand in the expected relation to one another, and not as a match on levels.

The epistemic status of this procedure is worth stating explicitly, since it conditions how the resulting estimates should be read. There is no single correct stationarity threshold: there is a combination of thresholds that minimises divergence from external evidence, and identifying it requires iteration on outputs. Because the algorithmic layer is closed to the data user, iteration on outputs is the only form of control available at this level of data access. It is nonetheless a real form of control, and it is what allows the analysis to rest on a numerical base whose behaviour has been verified against independent sources.

User-type classification and the attribution of residence. The delivered matrices classify users as residents, commuters, visitors or users in transit. This classification rests on a residence attribution rule which is known to us and which we report here, because the population it defines

does not coincide with the registered population. A device is attributed to a municipality when it is present there overnight, in the interval from 20:00 to 08:00, on more than half of the nights in the twelve months preceding the analysis. Attribution is therefore made to an individual municipality and not to the metropolitan area as a whole, and what it identifies is habitual overnight presence rather than administrative registration. Section 6 quantifies the difference between the two populations.

Disaggregation and masking. A request for mobile network data is specified jointly along three dimensions: spatial detail, temporal detail, and the profiling variables by which the counts are broken down. The three interact multiplicatively. The provider releases output under a confidentiality regime, and the finer the disaggregation requested along any dimension, the smaller the counts falling in each cell of the resulting table and the larger the share of them that is withheld. Detail along one dimension is therefore obtained at the cost of completeness along the others, and a maximal request does not yield a maximal amount of information.

Before the extraction is executed, the operator reports which entries a given specification would cause to be withheld and in what proportion. We used this facility to design the extraction rather than to audit it afterwards, revising the specification so as to keep the withheld share within limits we judged tolerable for the analytical purpose. The levels of detail reported in Table 1 and in Section 2.3 are the outcome of that exercise.

Entries withheld at the release stage were reconstructed downstream under apportionment assumptions. We do not report the procedure in detail, but we state its status, which is what matters for the interpretation of the results: it is a reconstruction resting on assumptions, not an observation, and its influence on the estimates is proportional to how much of the output it has to cover. This is why containing the extent of masking when the extraction is specified matters more than the

sophistication of the reconstruction applied afterwards. The residual uncertainty is discussed in Section 7.1.

2.5 Data quality notes

Each data source has strengths and limitations that must be considered when they are combined. For MND, a first concern is the representativeness of the operator's customer base relative to the entire population. Market-share information and benchmarking against survey-based indicators were therefore used to derive expansion factors and to identify segments where coverage may be weaker [4,5]. A second concern is the spatial resolution: the MND used in this study were delivered as pre-processed OD matrices on the 192 operator zones, a system coarser than the 207 MND zones on which the extraction was specified and than the 1,050 analytical zones used for demand analysis (Section 2.4). Two distinct constraints act on what is delivered, and they are easily conflated. The first is radio coverage, which determines the zoning system on which the data can be released at all. The second is the confidentiality regime governing the release [13], which acts not on the zoning but on the contents of the table: the finer the disaggregation requested, spatially, temporally and by profiling variable jointly, the smaller the counts falling in each cell and the larger the share of them that is withheld. The reconstruction of the MND zones, the reconstruction of withheld entries and the transcoding between the MND zones and the analytical zones each introduce uncertainty in the attribution of flows to territory. A further concern arises from the probabilistic relationship between the operator's cell catchment areas and any zoning system defined over them, which can lead to misattribution of trips near cell boundaries.

For FCD, the main limitations are the penetration rate of monitored vehicles, the uneven spatial distribution of traces, a potential bias towards specific fleet or user types and the restriction to

private vehicle trips. To mitigate these issues, we applied a traffic-count-based expansion, consistency checks, outlier removal and temporal aggregation over periods long enough to stabilise the indicators.

Survey and administrative data, despite their statistical design, are affected by sampling error, non-response and possible under-reporting of specific trip types. The survey sample is stratified on a set of population profiling variables, and is designed to be representative at the level of the 16 survey macro-zones for estimates produced by stratum, such as the trip generation rates used here. Aggregate quantities that do not require stratification can be estimated below that level.

To reduce the impact of these limitations, the project adopted a cross-source perspective from the outset. Quality checks were carried out both within each source and across sources, with particular attention to the consistency of total volumes, modal split and boundary flows between MND zones and between the study area and the external ring [6]. Event days, such as major cultural or sporting events (e.g. football matches at the city stadium), were identified and treated separately to avoid distorting typical weekday indicators.

Table 3. Data sources and their role in the data-fusion workflow.

Data source	Description and coverage	Role in the data-fusion pipeline	Key limitations
Mobile Network Data (MND)	Aggregated, anonymised OD matrices from a major national operator (~30% market share). User-type classification (residents, commuters, visitors, transit). Hourly detail; May, July and November 2024; three day types. Spatial resolution: 207 MND zones (144 internal), delivered by the operator on 192 operator zones and restored through an FCD-based re-split (Section 2.4).	Primary source for total trip volumes for all user types. Inter-zonal flows. Seasonal and day-type variation.	Market-coverage bias. Coarser spatial resolution (207 MND zones) than the analytical zoning (1,050 internal zones). Intra-zonal trips excluded by construction. Preprocessing performed by the operator; parameter configuration and iterative calibration are documented in Section 2.4.
Floating Car Data (FCD)	Anonymised GPS traces from connected vehicles (~4% penetration rate). Processed into trajectories, trips and link-level indicators (Table 2). Metropolitan road network coverage. Observation period: October 2023.	Independent benchmark, at a highly aggregate level (provincial totals and macro-components), for the magnitude of MND-derived trip volumes and for the aggregate share of car traffic, after expansion via the ACI vehicle register. Identification of ping-pong	Low and uneven penetration rate. Potential fleet/user-type bias. Restricted to private vehicle trips. Requires traffic-count-based expansion.

Data source	Description and coverage	Role in the data-fusion pipeline	Key limitations
		outliers in the exchange between adjacent operator zones, before the re-split. Basis of the re-split restoring the 207 MND zones from the 192 operator zones delivered (Section 2.4).	
Mobility survey	Ad hoc household survey, conducted in 2016 and updated in 2023 (Section 2.3). ~5,500 interviews. Resident population aged 14–85. Trip chains with origin, destination, mode, purpose, time of day. Designed to be representative at the level of the 16 survey macro-zones. Generation indices actualised to 2024 using updated ISTAT population data.	Derivation of trip-generation indices by user profile (gender, age group, employment status). Modal split and trip-purpose distributions. Gravity model calibration at the level of the 55 municipalities (Section 4.2). Baseline for validation (Section 6).	Sampling variability at resolutions finer than the 16 survey macro-zones, the level at which the survey is designed to be representative (Section 2.5). Non-response and under-reporting of short trips. Covers registered residents only.
Official administrative data	ISTAT census and demographic statistics (census-section level; 10,683 sections). Metropolitan employment register ("Atlante Metropolitano"). ACI national vehicle register. Regional permanent traffic counting system.	Population disaggregation by analytical zone (age, gender, employment) for survey expansion. Census sections as building blocks for the analytical zoning. Vehicle register for FCD expansion. Traffic counts for flow validation.	Census data updated infrequently (2011, 2021). Employment register may lag. Traffic counts limited to equipped sections.

Note to Table 3. All data refer to the Metropolitan City of Bologna. Observation periods: MND, May, July and November 2024; FCD, October 2023; survey waves, 2016 and 2023. The observation periods of the sources do not overlap; Sections 2.3 and 2.4 describe how the workflow accounts for this.

3. Regulatory and Policy Context

The identification of low-demand areas and the design of DRT services in the Bologna metropolitan area are guided by the Italian national regulatory framework for local public transport. Resolution No. 48/2017 of the Italian Transport Regulation Authority (Autorità di Regolazione dei Trasporti, ART) provides principles and criteria for the regulation of scheduled and demand-responsive services [10]. These include requirements for defining public-service areas, delineating mobility catchment areas and structuring service lots for competitive tendering. In particular, Measure 1 of the Resolution requires an assessment of actual and potential mobility demand, Measure 2 concerns the identification of low-demand areas and Measure 3 addresses the definition of the most suitable modes and types of transport to meet low demand, taking into

account economic, environmental and contextual variables as well as the principles of modal integration and sustainability.

Decree No. 157/2018 of the Ministry of Infrastructure and Transport (DM 157/2018, art. 4) further specifies the conditions under which areas may be classified low demand areas, by defining a set of primary and secondary indicators based on territorial, demographic and socio-economic characteristics [11]. Primary indicators include: the generative potential of daily trips (municipalities generating fewer than 3,000 systematic trips per day according to ISTAT commuting statistics, with 2019 as the reference year are classified as low demand); the demographic composition of the population (municipalities where more than 10% of the population is in the education-age group of 5–24 years or more than 10% consists of residents over 70 are flagged as having fragile demand characteristics); and the degree of peripherality according to the National Strategy for Internal Areas, with municipalities classified as intermediate, peripheral or ultra-peripheral qualifying as indicators of low demand [14,15]. Secondary indicators include altitude range (above 600 metres, indicating significant orographic complexity) and territorial dispersion (settlement fragmentation with critical average distances between municipal centres and hamlets).

At the local level, the Bologna SUMP translates these national principles into strategic objectives and operational targets [12]. The Plan emphasises the need to guarantee a minimum level of public transport service to all residents, with particular attention to peripheral and rural areas, while promoting a shift towards more sustainable modes. Within this framework, the local public transport authority is required to periodically review the classification of the analytical zones, to assess whether existing fixed-route services are adequate and to propose the introduction or adjustment of DRT solutions where appropriate.

The DREAM_PACE project was conceived as a tool to support this policy process. By integrating MND, FCD and survey-based information, the project provides a coherent and up-to-date picture of mobility demand and accessibility across the metropolitan territory. One of the distinctive contributions of this work is the operationalisation of the regulatory concepts defined by DM 157/2018 and ART Resolution No. 48/2017 through a data-driven approach: rather than relying solely on the indicators prescribed by the decree – which are largely based on static socio-demographic data (ISTAT commuting statistics; census population; peripherality classification) – the framework enriches the identification of low-demand areas with observed mobility data derived from the multi-source fusion process, thereby providing a more granular and demand-sensitive classification than would be possible using the regulatory indicators alone [8,9].

4. Methods

4.1 Workflow architecture

The methodological framework is organised as a modular workflow comprising four main stages that link raw input data to policy-oriented outputs (Figure 2).

Stage 1: Spatial and temporal harmonisation. The first stage focuses on harmonising the different sources in terms of spatial and temporal units. A common analytical zoning of 1,113 analytical zones, 1,050 of them within the Metropolitan City, was constructed from ISTAT census cells (construction criteria in Section 2.3), calibrated to ensure relative homogeneity of socio-economic characteristics within each analytical zone. MND OD matrices, which use the coarser resolution of the 207 MND zones, are mapped onto the analytical zoning through a transcoding matrix. FCD trajectories are matched to the analytical zoning and to the road network, yielding link-level indicators. Survey data are expanded to the target population using socio-demographic

weights (gender, age group, employment status and survey macro-zone of residence) and reallocated to the analytical zones. For each analytical zone, socio-economic attributes – population by age group, gender and employment status – are computed from ISTAT data and the metropolitan employment register, and further cross-tabulated to create user profiles that serve as the basis for subsequent expansion.

Stage 2: Construction of integrated OD matrices. In the second stage, two parallel processes produce complementary OD matrices. On one hand, trips leaving and returning home are estimated for residents only, by applying survey-derived mobility indices (average number of daily trips by user profile, disaggregated by gender, age group and employment status) to the updated population of each analytical zone; these trips are further disaggregated by purpose and mode, based on the distributions observed in the survey at the level of the 16 survey macro-zones. On the other hand, total movements of all users – residents and non-residents alike, including city users, commuters and transit traffic – are calculated from MND, enriched and corrected through FCD. The FCD correction specifically targets outliers in short-distance flows between adjacent operator zones: by comparing the ratio of mutual exchange between adjacent operator zones against their respective generative and attractive potential as measured by FCD, cases of anomalous over-representation are identified and removed. This cross-referencing eliminates spurious trips that do not correspond to actual movements, improving the quality and consistency of the information flow, particularly on short inter-municipal routes and in high-density urban areas.

Stage 3: Estimation of the intra-zonal component. The third stage addresses a critical gap in MND-derived matrices: movements that begin and end within the same MND zone are excluded by construction from the MND extraction process. A dedicated estimation procedure reconstructs

this component. It is calibrated at the level of the municipalities, applied at the level of the analytical zones and finally expressed at the level of the MND zones; it is detailed in Section 4.2.

Stage 4: Transformation into policy indicators. In the final stage, the integrated OD matrices and estimated intra-zonal trips are combined to produce the full set of policy-oriented indicators. This includes normalisation by surface area, computation of generative and attractive potential, classification of the analytical zones into mobility demand categories and mapping to DRT service levels. These indicators directly support the identification of low-demand areas in accordance with the national regulatory framework and the objectives of the SUMP.

Figure 2 depicts the pipeline as three blocks, with each stage stating the spatial system on which it operates. The first block carries the survey-based chain: the mobility survey, expanded through census and employment registers, yields user profiles and generation and attraction indices on the analytical zoning, and from these a resident OD estimate. The second block carries the big-data sources: stationarity thresholds are defined on the operator zones by zone extent and settlement density; the MND matrices are extracted along the zonal chain, with 207 MND zones requested, 192 operator zones delivered and 207 restored through the FCD-based re-split; and the FCD, map-matched at link level, support the ping-pong correction of adjacent-zone exchange on the operator zones as delivered. A quality-check stage closes this block: mobility KPIs are checked at aggregate and macro-component level, and the feedback loop labelled iterative re-configuration returns to the stationarity thresholds, the iterative calibration of the preprocessing parameters described in Section 2.4. The third block integrates: the gravity models, calibrated on the 55 municipalities, provide the intra-zonal shares of the movements MND cannot observe; the sources are transcoded and fused on the analytical zoning; generated and attracted volumes are corrected and scaled against the benchmarked volumes and the resident OD estimate; and the classification of Section

4.3, followed by the filter on the planned SUMP public transport network, leads to the low-demand areas and the DRT service levels of Section 4.4.

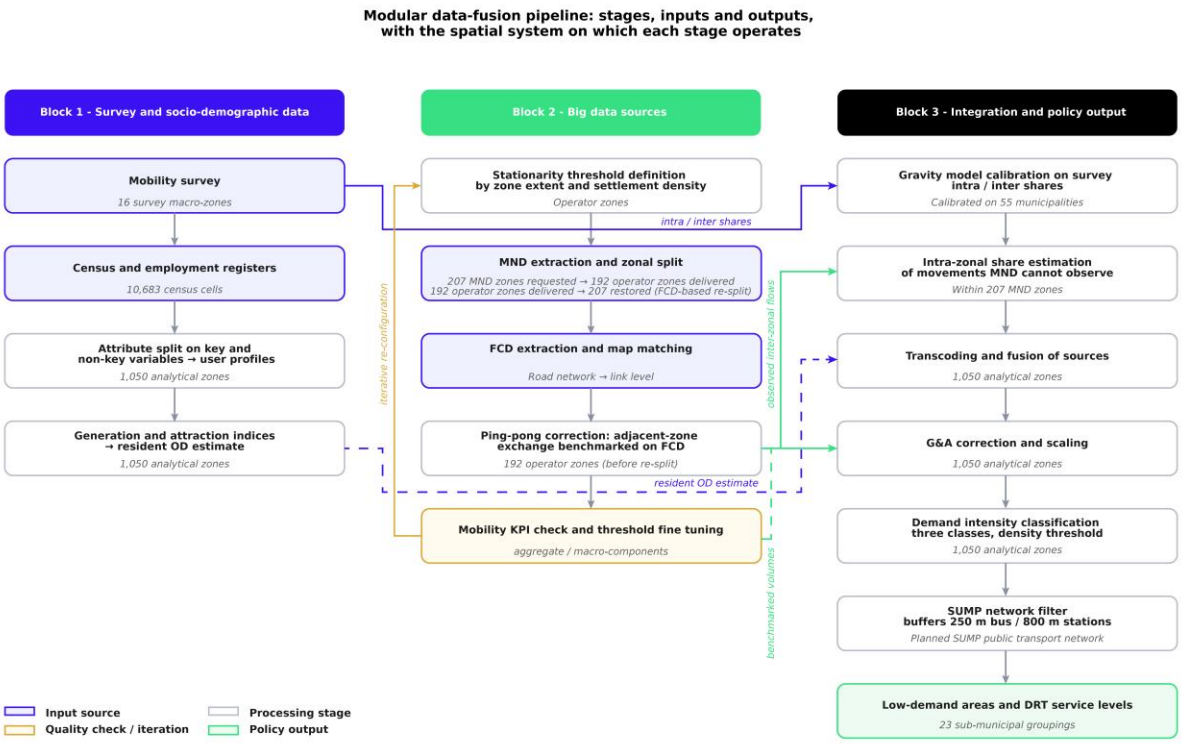


Figure 2. The modular data-fusion pipeline: stages, inputs and outputs, with the spatial system on which each stage operates.

4.2 Estimation of the intra-zonal component of the MND matrices

Trips that begin and end within the same MND zone play a non-negligible role in overall mobility, especially in the larger MND zones, which combine residential and service functions. MND-based OD matrices do not contain these trips: the trip-detection algorithms applied during data extraction only register movements between distinct spatial units, and any displacement that begins and ends within the same operator zone does not satisfy the stationarity thresholds and is therefore excluded by construction [5]. If left unaddressed, this exclusion would compromise the robustness and representativeness of the demand estimates, particularly in the peripheral MND zones, where a

substantial share of trips may be short and internal but still relevant for assessing accessibility to basic services.

To reconstruct the intra-zonal component, we developed a four-step estimation procedure.

The procedure operates across three spatial levels, and the distinction between them is essential to what follows. The gravity models are **calibrated** at the level of the municipalities, where the mobility survey provides the observed share of *intra-municipal* trips. They are then **applied** at the level of the analytical zones, which distribute that share within each municipality. The resulting flows are finally **expressed** at the level of the MND zones, at which the *intra-zonal* component is defined. Intra-municipal and intra-zonal trips are therefore not equivalent: the former is the quantity the model reproduces in calibration, the latter the quantity it is used to estimate. A trip internal to a municipality but running between two different MND zones is observed in the MND matrices; an intra-zonal trip is not.

A further distinction concerns the level at which the exclusion originates. Trip detection operates on the operator zones, so the component missing from the delivered matrices is that internal to an operator zone. For the operator zones that coincide with a single MND zone the two components are the same; for those resulting from further aggregation, the missing component also comprises the flows between distinct MND zones belonging to the same operator zone. These are recovered by the FCD-based re-split described in Section 2.4, so that the residual component to be estimated by the gravity models is, in every case, that internal to a single MND zone.

Step 1: Zoning transcoding. A correspondence matrix was constructed between the 207 MND zones and the 1,113 analytical zones (144 and 1,050 respectively within the Metropolitan City), together with their correspondence to the 55 municipalities. This step ensures spatial compatibility

between sources operating at different resolutions and allows information to be transferred across the three levels.

Step 2: Gravity model calibration. Based on data from the mobility survey, two separate gravity distribution models were calibrated to estimate the share of self-contained (intra-municipal) trips for each municipality. The distinction between two models – one for the municipality of Bologna and one for the remaining municipalities of the metropolitan area – reflects the markedly different intra-municipal mobility patterns observed in the two contexts. In the dense urban core, a larger share of trips is internal owing to the concentration of services, employment and educational facilities. In the more dispersed peripheral municipalities, by contrast, the share of internal trips is driven primarily by local errands and school-related travel. The models were calibrated by least squares at the level of the municipalities. For the fifty-four municipalities other than Bologna, calibration reproduces for each municipality the ratio of intra-municipal to total generated trips derived from the survey. For Bologna, whose territory is split by the survey into a central and a peripheral macro-zone, calibration reproduces the four observed relations between the two, the two internal ones and the two exchanges, which together constitute the intra-municipal component of the municipality. Calibration uses the survey at the level of the municipalities. The quantity involved is a single aggregate share, the ratio of intra-municipal to total generated trips, which does not require stratification and therefore rests on large counts; its estimation at municipal level is admissible for the purpose of parameter estimation.

Step 3: Application to fine-grained zoning. The calibrated gravity models were then applied to the 1,050 analytical zones of the metropolitan area to reconstruct internal OD relationships at the finest available spatial resolution. This step distributes the estimated intra-municipal trip share across the analytical zones of each municipality, taking into account the socio-economic

characteristics of each analytical zone (population, employment, land use) as attraction and generation proxies.

Step 4: Expression of the intra-zonal component at the level of the MND zones. The zone-to-zone flows estimated in Step 3 were finally aggregated to the MND zones, yielding for each MND zone the share of trips that both begin and end within it, relative to the total it generates and attracts. This is the intra-zonal component absent from the MND matrices, and expressing it at this level ensures consistency between the gravity-model estimates, the survey data and the MND-derived inter-zonal matrices.

The estimated intra-zonal trips are added to the inter-zonal flows to obtain total generated and attracted volumes for each MND zone. FCD data and local knowledge were used to validate the plausibility of the resulting values, in particular by comparing the estimated internal trip rates with vehicle-flow patterns observed in MND zones where FCD coverage is sufficient.

4.3 Indicators and classification

For each analytical zone in the study area, we compute a set of indicators derived from the integrated mobility picture. These indicators operationalise the concepts of demand intensity and territorial vulnerability defined in the national regulatory framework, while enriching them with observed mobility data produced by the data-fusion process.

The framework computes two groups of indicators, defined at different spatial levels. The first captures the generative and attractive potential of each analytical zone, measured as the total number of trips originating in or destined for it on a typical weekday and normalised by surface area (trips per km²), so as to provide a measure of demand density comparable across analytical zones of different size. The second, derived directly from the MND and therefore defined at the

level of the MND zones, comprises the peak-to-off-peak ratio, the composition of users by type (residents and non-residents) and the temporal stability of demand across the three observed months (May, July and November 2024).

The identification of low-demand areas rests on the first group alone, which is defined at the finest available level and is directly comparable across analytical zones. The second group supports the wider reading of the demand matrix: the characterisation of temporal regimes and of the resident and non-resident composition of demand. Through the data fusion these indicators are likewise expressed at the level of the analytical zones: the temporal and user-type structure observed at the level of the MND zones is carried to each analytical zone by the composition of its resident population, while the non-resident component retains the value of the MND zone it belongs to.

The density values of the first group are the basis of the classification by mobility demand intensity: the analytical zones are grouped into three classes of low, intermediate and high intensity, delimited by two thresholds on demand density, 2 and 80 trips per km² per day, which capture the generative and attractive potential of each analytical zone. A separate pair of time-normalised thresholds, 0.5 and 5 trips per hour per km², is used to assess the temporal concentration of demand, and enters the subsequent characterisation of the low-demand areas (Section 4.4) rather than the delimitation of the three classes. The indicators and the thresholds applied to them are summarised in Table 4. The thresholds were not derived from the national regulatory framework, whose treatment of demand indicators is less developed than this analysis required, but from the operational purpose of the study: the sizing of public transport services and their financial sustainability given the demand that can realistically be captured. Below the lower threshold, no scheduled service can be operated at a cost per passenger compatible with ordinary funding arrangements, which is what qualifies these areas for demand-responsive solutions; above the

upper threshold, demand density is sufficient to justify a high-frequency corridor; the intermediate class comprises analytical zones for which neither conclusion holds and where the service decision depends on factors beyond demand density. The thresholds were implemented as adjustable parameters in an interactive dashboard, and the final values were selected jointly with SRM Reti e Mobilità by observing how the classification responded across the parameter range and assessing the resulting patterns against detailed local knowledge. This is an expert-assisted rather than a purely algorithmic procedure. The demand-intensity class is not merely descriptive: it enters the subsequent stage of the analysis as an input, since the assignment of a demand-responsive service type to each low-demand area depends on it alongside the temporal and spatial regularity of demand (Section 4.4).

The classification feeds into the identification of low-demand areas. Here, the framework operates on two complementary levels. The first level applies the indicators prescribed by the DM 157/2018 [11] – generative potential (ISTAT commuting data), demographic composition and degree of peripherality – to identify municipalities that satisfy the regulatory definition of low demand. The second level enriches this statutory classification with the observed mobility demand indicators produced by the data-fusion process. By computing the actual generative and attractive potential of each analytical zone from the integrated MND–FCD–survey matrices, the framework provides a considerably more granular picture: it can detect low-demand conditions at sub-municipal scale, identify pockets of weak demand in municipalities that would not be flagged by the regulatory indicators alone (for example, dispersed rural areas within otherwise non-peripheral municipalities) and, conversely, recognise analytical zones that satisfy the regulatory criteria but exhibit higher-than-expected mobility owing to specific local attractors. This data-driven

operationalisation bridges the gap between the static, census-based indicators of the ministerial decree (DM) and the dynamic, multi-source evidence produced by the fusion pipeline.

4.4 Mapping to DRT service levels

The final stage of the method links mobility demand indicators to DRT service design through a structured four-step procedure [8,11].

Step 1: Mapping of generative and attractive potential. Each analytical zone is characterised in terms of trip density (movements generated and attracted per km²) and assigned to one of the three mobility demand intensity classes defined in Section 4.3.

Step 2: Load estimation along main corridors and application of the network filter. For analytical zones along the main transport corridors, the expected volumes on board (passenger-kilometres) are estimated and compared with the available capacity of planned services, to assess the feasibility of conventional modes (rail, high-frequency bus). Candidate analytical zones are then re-analysed in light of the public transport network planned in the SUMP, to assess whether existing or planned fixed-route services already provide adequate coverage [12]. Coverage is assessed by applying catchment buffers to the planned network: 250 m around bus services and 800 m around rail stations, the accessibility distances adopted by the SUMP monitoring framework itself, drawn from the national indicator set for SUMP monitoring [16]. Using the plan's own accessibility standard keeps the filter consistent with the network it assesses. Analytical zones falling within the buffers, and those with low generative potential, are excluded from the demand-responsive analysis: the 23 low-demand areas are the residual of this filter and of the subsequent spatial grouping, rather than the direct output of the density classification.

The analytical zones that pass the network filter constitute the set of candidate analytical zones. Contiguous candidates were then grouped into 23 low-demand areas on the basis of spatial adjacency, so as to define territorial units of a size and shape compatible with the operation of a demand-responsive service. It is these areas, rather than the individual analytical zones, that are characterised in the step below.

Step 3: Spatial and temporal characterisation of mobility demand. For each remaining low-demand area, the temporal profile and spatial profile of demand are assessed along two dimensions:

- *Temporal dispersion:* the weight of peak-hour trips (school and work peaks) on total daily demand. High temporal concentration indicates that demand is clustered in predictable time windows; low concentration indicates a more dispersed pattern throughout the day.
- *Spatial dispersion*: the weight of the most-generated and most-attracted analytical zones on overall demand for each area. High spatial concentration means that trips are directed predominantly towards a few dominant destinations (e.g., the nearest urban centre); low concentration indicates a dispersed pattern across multiple origins and destinations.

Step 4: Assignment of DRT service profiles. The service types considered are adapted from the classification of flexible transport services set out by Engels, Ambrosino and Boero [17], which orders services by the number of elements fixed in advance: route, stopping points and passage times. Their five scenarios are collapsed into four here: the two intermediate cases, in which predefined stops are served on request within a corridor and within an area respectively, are treated as a single type. Four types are considered: base itinerary with fixed stops, suited to linear areas polarised on a single axis and usable as an extension of a fixed line where demand is predictable; base itinerary with deviations and fixed stops, suited to settlements with nearby hamlets and to

demand that is partly polarised but with local secondary requests; free routing between predefined points, suited to dispersed settlement patterns with no preferential axes and demand that is diffuse and variable over time; and free routing between any set of points, suited to fragmented contexts with irregular but significant demand, and viable only where the road network and the dispatching technology allow it.

For each low-demand area, the appropriate service profile is assigned by combining the two-by-two matrix of Step 3 with the demand-intensity class of the analytical zones composing the area, the regulatory requirements on minimum service levels [10,11] and the strategic objectives of the SUMP [12]. The capacity characteristics taken into account are those typical of the vehicle fleet used for demand-responsive operation; together with the acceptable detour and a viable service frequency, they underlie the economic reading of the demand-density thresholds of Section 4.3: below given load levels, the cost per transported passenger rises beyond what ordinary funding arrangements can sustain. The output includes both a service-type recommendation and an indication of whether a minimum activation threshold is met, a threshold below which the service is not activated. Table 5 reports the service types, their assignment criteria and the capacity assumptions; these recommendations are directly usable in the planning and tendering process.

Table 4. Indicators and thresholds used to identify low-demand areas.

Indicator	Data source	Threshold / criterion	Spatial unit	Type
Level 1 — Statutory classification (DM 157/2018, art. 4)				
Generative potential of daily trips	ISTAT commuting statistics (2019)	< 3,000 systematic trips/day (work + study)	Municipality	Primary
Demographic composition (young or elderly population)	ISTAT demographic registers (resident population, 31 December 2023)	> 10% of resident population aged 5–24, or > 10% aged over 70	Municipality	Primary
Degree of peripherality	National Strategy for Internal Areas (SNAI)	Classification as intermediate, peripheral or ultra-peripheral	Municipality	Primary
Altitude range	ISTAT / regional geodata	> 600 m (significant orographic complexity)	Municipality	Secondary
Territorial dispersion	ISTAT settlement data	Critical average distance between municipal centres and hamlets	Municipality	Secondary
Level 2 — Data-driven classification (multi-source fusion)				
Generative and attractive potential (observed)	Fused MND–FCD–survey OD matrices	Two thresholds on demand density, 2 and 80	Analytical zone	Core (classification)

Indicator	Data source	Threshold / criterion	Spatial unit	Type
		trips per km² per day, delimiting the low, intermediate and high intensity classes; the low-intensity class qualifies for the DRT analysis		
Peak-to-off-peak demand ratio	Fused OD matrices (hourly structure from MND)	Reading of temporal regimes; no fixed threshold	Analytical zone	Characterisation (Section 4.3)
User-type composition	Fused OD matrices (MND residency classification)	Share of residents and non-residents; reading of demand stability	Analytical zone	Characterisation (Section 4.3)
Temporal concentration of demand	Fused OD matrices (hourly detail)	Weight of peak-hour trips (school and work peaks) on total daily demand, high vs. low; a separate pair of time-normalised thresholds, 0.5 and 5 trips per hour per km², enters this characterisation (Section 4.4, Step 3)	Low-demand area	Characterisation (Section 4.4)
Spatial dispersion of destinations	Fused OD matrices	Weight of the dominant destination(s) on total attracted demand, high vs. low concentration (Section 4.4, Step 3)	Low-demand area	Characterisation (Section 4.4)

Note to Table 4. Level 1 applies the statutory indicators prescribed by DM 157/2018 (art. 4) at municipal level; municipalities satisfying all three primary indicators plus at least one secondary indicator qualify as "priority low-demand areas". Level 2 enriches the classification using observed mobility data from the multi-source fusion pipeline at sub-municipal scale; its characterisation indicators are expressed at the level of the analytical zones through the data fusion (Section 4.3), while the Step 3 dimensions are assessed for each low-demand area (Section 4.4). The Level 2 thresholds were implemented as adjustable parameters in an interactive dashboard, and the final values were selected jointly with SRM Reti e Mobilità through an expert-assisted rather than purely algorithmic procedure (Sections 4.3 and 6).

Table 5. DRT service types, assignment criteria and capacity assumptions. Service types are assigned to the 23 low-demand areas (Section 4.4).

Service type	Route and stops	Demand pattern (spatial × temporal)	Suited context
Base itinerary, fixed stops	Fixed route, fixed stops and timetable, concentrated windows	High spatial × high temporal concentration	Linear areas polarised on a single axis; extension of a fixed line where demand is predictable
Base itinerary with deviations	Core fixed route, on-request deviations, fixed stops	High spatial × low temporal concentration	Settlements with nearby hamlets; demand partly polarised with local secondary requests
Free routing between predefined points	Flexible routing, fixed set of stops, concentrated windows	Low spatial × high temporal concentration	Dispersed settlement, no preferential axes; demand diffuse and variable over time
Free routing, any set of points	Fully flexible door-to-door	Low spatial × low temporal concentration	Fragmented contexts, irregular but significant demand; viable only where network and dispatching technology allow

Note to Table 5. Service types are adapted from the classification of flexible transport services in [17], whose five original scenarios are consolidated here into four types. Capacity assumptions refer to the vehicle fleet typical of demand-responsive operation. The assignment reflects the economic constraint underlying the demand-density thresholds: below given load levels, the cost per transported passenger rises beyond what ordinary funding arrangements can sustain, which is what the minimum activation threshold safeguards.

5. Results

The integrated framework produces a detailed picture of daily mobility in the Metropolitan City of Bologna. On a typical weekday in November 2024, the total number of trips, including the intra-zonal component, is estimated at 3,679,000. This figure is 1,170,000 higher than the previous estimate based on the earlier survey and on assumptions regarding non-resident access to the metropolitan area [12]. The difference decomposes exactly into three components: 264,000 additional trips by residents, concentrated in movements between municipalities other than Bologna, where the resident component exceeds the survey-based estimate by 77.6%; 652,000 trips by non-resident city users, previously estimated only from limited "city-gate" access data; and 253,000 crossing movements by non-residents, which were not considered at all in the earlier survey framework.

Spatial patterns confirm the strong concentration of mobility in the urban core and along the main radial and tangential corridors. The figures below refer to all users, resident and non-resident. Internal movements within Bologna alone account for 1,044,000 daily trips, followed by intra-municipal movements in other municipalities (954,000). Exchanges between Bologna and the rest of the metropolitan area amount to 461,000 weekday trips, while trips between non-Bologna municipalities contribute a further 478,000. Exchanges with territories outside the metropolitan area (487,000) and crossing movements (255,000) together account for 741,000 daily trips.

Figure 3 reports the weekday temporal profiles derived from mobile network data. Panel (a) compares the typical weekday profile in the three observed months. Seasonality changes the shape of the profile, not only its level: the daily maximum falls at 07:00 in November and moves to 17:00 in May and July. The school calendar is directly legible in the profile: the time bands tied to school mobility contract far more than the others: at 07:00 the July volume is 24.2% below November, at 13:00 it is 20.1% below, at 14:00 21.6% below. The evening commuting peak at 17:00, by contrast,

cedes only 6.6%, and at 20:00 July exceeds November by 16.1%. The share of trips in the 12:00–14:00 band falls from 18.7% in November to 16.2% in July, while the 21:00–23:00 share rises from 7.0% to 8.9%. The former is the effect of the non-school period, in which students and school staff do not express return mobility at that time. The comparison is possible only net of the component internal to the MND zones, whose identification through the gravity models is available for November alone; it refers to the weekdays of the three months, and is restricted to trips internal to the metropolitan area, since including crossing trips would attenuate the seasonal differences.

Panel (b) contrasts a typical May weekday with the day of a major concert, in hourly shares of each day's own total. On the event day overall mobility is 29.6% higher, which in absolute terms lifts the entire curve, morning hours included; once normalised, the morning difference vanishes and only the evening deviation remains. The sequence the panel isolates is inflow–event–outflow: the differential is positive at 18:00 (+0.69 pp) and 19:00 (+0.85 pp), vanishes at 20:00 and 21:00, when spectators are inside the venue and do not move, and reaches its maximum at 23:00 (+1.43 pp). The vertical band marks the 20:00–22:00 duration of the concert and coincides exactly with the interval of null differential: what the figure captures is the event, not daily variability.

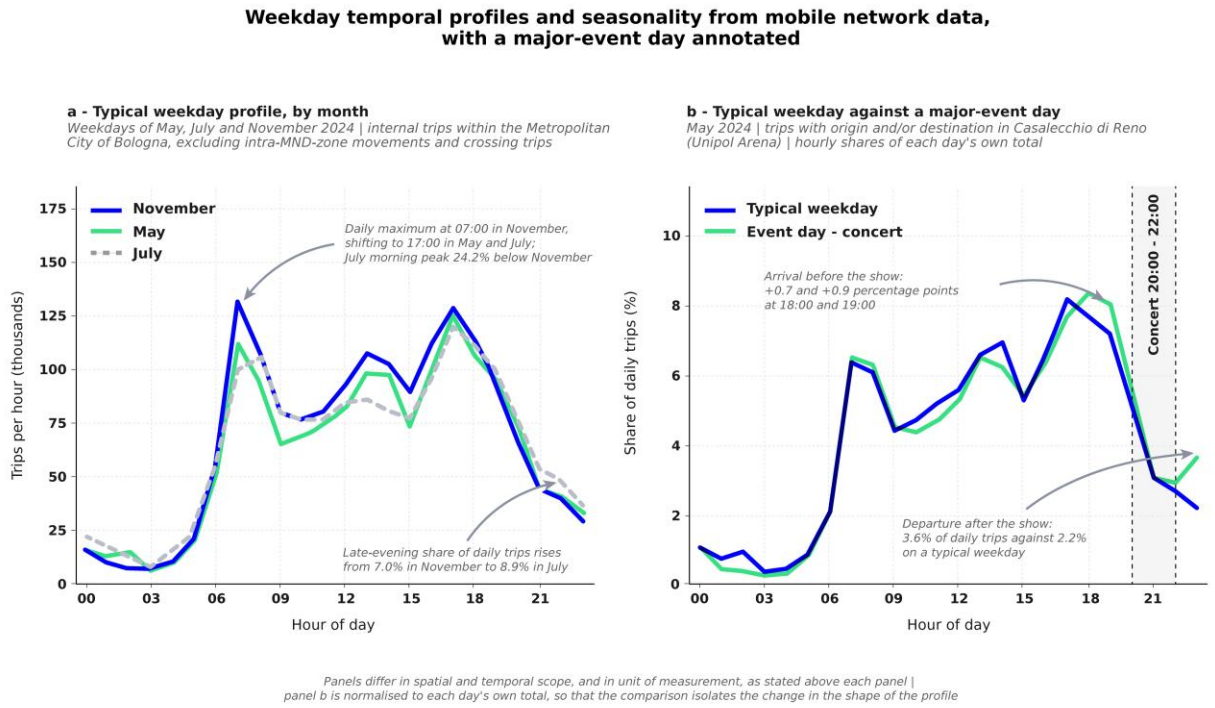


Figure 3. Weekday temporal profiles from mobile network data. Panel (a): typical weekday profile in May, July and November 2024, for trips internal to the metropolitan area, excluding intra-MND-zone movements and crossing trips. Panel (b): hourly shares of each day's own total on a typical weekday and on the day of a major concert (May 2024), restricted to trips with origin and/or destination in the municipality hosting the arena; the shaded band marks the 20:00–22:00 duration of the event.

The figures above refer to the total population observed by the mobile network. Within it, non-resident users account for 905,187 movements on a typical weekday, 24.6% of all trips, and their distribution across macro-components is far from uniform. Non-residents dominate precisely the components whose trips have at least one end outside the metropolitan city: they account for 99.4% of crossing trips (both ends outside), 77.8% of the exchange between Bologna and external territories, and 67.1% of the exchange between the other municipalities and external territories; these three components together comprise 600,172 movements, two thirds of all non-resident mobility. A household survey of resident families cannot, by design, observe most of these movements: those who make them, people passing through the area or reaching it from territories outside, largely do not belong to the population from which the sample is drawn. Of the 254,821

crossing trips estimated from mobile network data, the survey records 2,148, the small share made by metropolitan-area residents. This is not under-reporting but a property of the sampling frame. The resident component of the matrix can therefore be compared with the survey; that comparison, restricted to residents, is presented in Section 6 (Figure 4). No independent benchmark exists for the non-resident component. Figure 5 reports its breakdown by macro-component, derived from mobile network data alone and to be read as an unvalidated estimate.

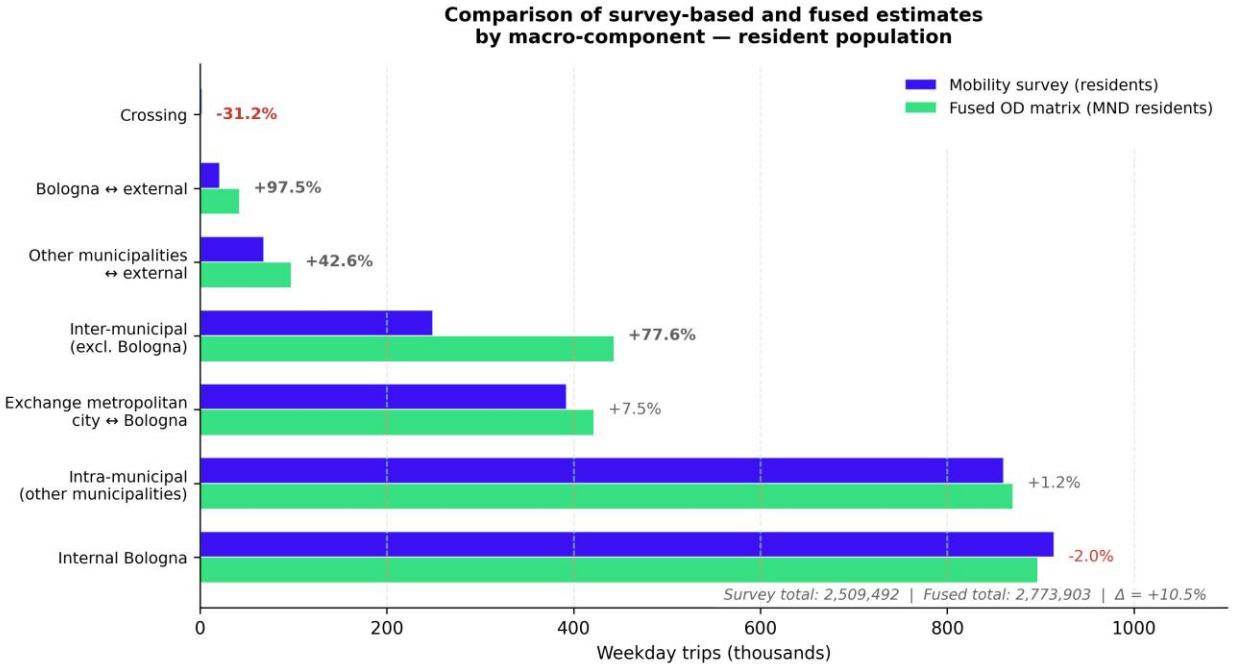


Figure 4. Survey-based and fused estimates of daily trips by macro-component: resident population, typical weekday. Macro-components as defined in Table 2 (Section 2.2); Metropolitan City of Bologna.

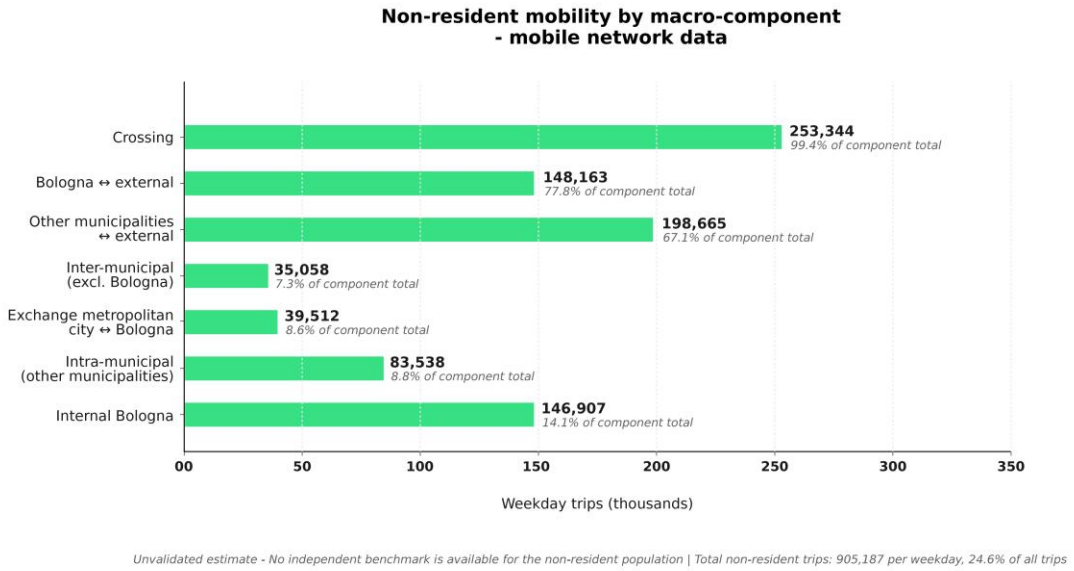


Figure 5. Non-resident mobility by macro-component, from mobile network data alone: unvalidated estimate. Typical weekday; macro-components ordered as in Figure 4.

The integration of survey information allows these movements to be characterised by purpose and mode. A large share of trips originating from or destined for the urban core is performed using public transport, while trips between peripheral areas are markedly more car-dependent. Seasonal analysis across the three observed months (May, July and November 2024) shows that total weekday volumes remain broadly stable, but the composition by user type and trip purpose shifts – with July showing lower systematic (work and study) mobility and higher leisure-related movements. Weekend mobility also reveals significant volumes: approximately 3.4 million trips on a typical Saturday and 2.9 million on a typical Sunday, confirming that non-systematic mobility is a structural component of the metropolitan system.

Building on the indicators described in Section 4.3, the framework classifies the analytical zones by demand intensity at both the regulatory and the data-driven level. The application of the DM 157/2018 (art. 4) indicator methodology identifies 11 municipalities that satisfy all three primary indicator criteria – low generative potential, high share of dependent population and high degree

of peripherality – covering an area of approximately 696 km² with roughly 32,500 residents. All 11 municipalities also exceed at least one secondary indicator threshold (altitude range or territorial dispersion), qualifying as "priority low-demand areas".

The data-driven classification based on the integrated demand estimates, however, produces a more differentiated picture. By computing the actual generative and attractive potential of each analytical zone from the fused matrices and applying the demand-density thresholds of Section 4.3 and the subsequent filtering and grouping steps of Section 4.4, the framework identifies 23 low-demand areas (covering 733 km² and approximately 27,600 inhabitants). The calibration of the thresholds reflects the capacity and cost characteristics of the different service types. This alternative classification differs from the statutory one in two important respects: it operates at sub-municipal scale, identifying pockets of weak demand that the municipal-level regulatory approach cannot detect; and it uses observed trip densities rather than census-based proxies, thereby capturing demand patterns that may diverge from what socio-demographic indicators alone would predict. The comparison between the two classifications provides the local transport authority with a richer evidence base for deciding where and how to intervene.

The two classifications diverge in both directions, and Appendix B reports the outcome for each municipality. Of the 11 municipalities identified by DM 157/2018, 8 overlap with at least one data-driven low-demand area and 3, namely Castel di Casio, Castiglione dei Pepoli and Loiano, overlap with none; conversely, 27 municipalities that the decree does not identify contain a portion of at least one such area. Seventeen municipalities are identified by neither classification. The divergence follows from the difference in units and criteria rather than from disagreement about the same object: a municipality that is low-demand as a whole may still concentrate its demand in a few analytical zones, while a municipality that is not low-demand overall may contain pockets

that are. The data-driven layer indicates where within the territory a different service model is warranted; the classification that carries regulatory force remains the statutory one.

The mapping to DRT service profiles yields concrete recommendations for each identified area. In some areas, a basic fixed-itinerary service focused on school and work peaks appears sufficient to guarantee access to essential destinations. In others – especially where demand is more evenly distributed throughout the day but remains below the threshold for fixed-route services – a flexible feeder profile connected to high-frequency corridors is suggested. In areas with low spatial and temporal concentration, a fully on-demand service is recommended. These recommendations are consistent with the principles of the SUMP and respect the constraints set by DM 157/2018 and ART Resolution No. 48/2017 [10,11,12].

6. Validation and Quality Assessment

Validation is carried out at several levels. At the most aggregate level, that of the seven macro-components defined in Table 2, total trip volumes and mobility indices derived from the integrated framework are compared, for the resident population, with those obtained from the mobility survey and from the modelling work carried out for the Bologna SUMP [12]. The results show good agreement in terms of overall trip volumes when the comparison is restricted to comparable macro-components. For mobility that is predominantly local and self-contained – trips internal to the municipality of Bologna and intra-municipal trips in other municipalities – the fused estimates and the survey-based figures are closely aligned, with differences of -2.0% and +1.2% respectively. For inter-municipal and exchange components, however, the fused estimates are systematically higher: +77.6% for inter-municipal trips outside Bologna, +97.5% and +42.6% for exchange trips with territories outside the metropolitan area, and +7.5% for exchanges between the rest of the

metropolitan area and Bologna. Figure 4 reports the comparison by macro-component for the resident population. Both terms refer to a typical weekday: the survey collects weekday trip chains, and the fused estimates are computed for the average weekday, so that the comparison is homogeneous in day type. These are precisely the trip categories that household mobility surveys are least equipped to capture. Respondents reconstruct a habitual chain of movements and tend to omit trips falling outside it: inter-municipal trips within the metropolitan area are frequently of this kind, and so, for different reasons, are trips crossing the metropolitan boundary, which are usually made for occasional purposes outside the respondent's habitual set. What the two categories share is not their length but their atypicality. Exchange trips with external territories additionally fall largely outside the survey sampling frame.

Overall, the resident component of the fused matrix yields 2.77 million weekday trips against 2.51 million from the survey (+10.5%). The two figures do not refer to the same population. The survey samples households recorded in the municipal registers, whereas the mobile network data attribute residence in a municipality to any device habitually present there overnight, irrespective of registration (Section 2.4). The population so defined additionally comprises non-registered residents, temporary workers, long-term visitors and students domiciled but not officially registered. It is a segment that household surveys and administrative registers are structurally unable to capture. Two mechanisms therefore act in the same direction and on the same components: a wider resident population, and the under-reporting of atypical trips described above. Our data do not allow their contributions to be separated, and we do not attribute the difference to either alone. What the pattern does establish is that the difference is not an artefact of the fusion: a distortion introduced by the procedure would not have spared the self-contained components, which agree with the survey to within 2%.

The distribution of trips by mode and purpose is then examined. Survey-based modal split and purpose shares are applied to the integrated OD matrices, and the resulting indicators are compared with independent information where available – for example, aggregate public transport ridership statistics and traffic counts on key corridors. While some discrepancies emerge at a detailed level, the overall patterns are consistent. In particular, private vehicle flows derived from FCD reproduce the volumes recorded by the regional permanent traffic counting system on the main corridors, confirming that the expansion and fusion procedure preserves the observed structure of vehicular mobility. The aggregate modal split and trip-purpose distributions obtained from the survey are coherent with the total volumes produced by the multi-source pipeline, with no systematic bias identified at the level of the survey macro-zones.

Cross-boundary checks are performed for flows between adjacent MND zones and for trips from the metropolitan area to external territories. FCD-based estimates of vehicle movements are particularly useful here, as they provide an independent benchmark for MND-based person flows and for survey-based expansions. Differences are analysed to identify potential biases, such as an over- or under-representation of specific user groups in the operator's customer base.

The robustness of the classification was explored by implementing the demand-density thresholds as adjustable parameters in an interactive dashboard, in which changing a threshold immediately reassigns the analytical zones between classes and updates the resulting map. Observing the classification across the parameter range shows that the identification of the low-demand areas is stable under reasonable variations of the thresholds, although individual borderline cases may change class. These borderline cases were discussed with SRM Reti e Mobilità and municipal stakeholders, so that final decisions incorporate qualitative knowledge of the territory rather than

being driven solely by numerical thresholds: an expert-assisted procedure, which we present as such.

7. Discussion

The experience of the DREAM_PACE and MIND projects illustrates both the potential and the challenges of integrating MND, FCD and traditional data sources in support of official mobility statistics [1,2,6].

On the positive side, MND provide a level of temporal continuity and spatial coverage that is difficult to achieve with surveys alone, while FCD enrich the analysis with detailed information on network performance and vehicle movements. When combined with statistically representative survey data and administrative registers, these sources enable the construction of indicators that are both spatially detailed and grounded in established statistical practice. The three-month observation period (May, July and November 2024) and the distinction between weekday, Saturday and Sunday profiles capture seasonal and day-type variations that a single-wave survey cannot reproduce.

At the same time, the use of privately collected digital traces raises methodological and institutional questions. From a methodological perspective, assumptions on market coverage, user behaviour and data pre-processing must be made explicit, validated and, where necessary, corrected through benchmarking with survey-based evidence [5,7]; Section 2.4 reports how this was done here, and Section 7.1 discusses what remains beyond correction. From an institutional perspective, cooperation between statistical offices, transport authorities and private data providers is essential to ensure sustainable access to data, respect for privacy regulations and the long-term reproducibility of indicators [13].

Before turning to the structural constraint that macro-level access imposes (Section 7.1), two limitations specific to this application deserve explicit statement.

The first concerns the calibration of the gravity models, which leaves a residual discrepancy with the survey-based shares whose structure differs across two groups of municipalities. In the municipalities of the first belt around Bologna the model systematically overstates the self-contained share. These municipalities have an attraction to and from the core comparable to that of the core itself, and the exponent estimated for the general regime is too steep to reproduce their behaviour; in principle each would have warranted separate treatment, as Bologna did. We chose not to do so, because calibrating a dedicated parameter for each would have meant estimating several parameters from a single calibration point each, a fragility we judged worse than the bias it would remove.

In the Apennine municipalities the discrepancy is of a different kind: large in absolute terms but with no systematic direction. Its main source is the resolution at which the reference data are reliable. The survey is designed to be representative at the level of the 16 survey macro-zones, whereas calibration operates at municipal level; part of the dispersion observed in the municipal comparison is therefore sampling error in the reference term rather than error in the model. A second, independent source concurs in the same municipalities: four of the seven operator-zone aggregations internal to the metropolitan city fall in the Apennines, so that the FCD-based re-split of Section 2.4, a reconstruction by assumption, operates precisely where the dispersion of the municipal comparison is largest. The two sources are distinct and concurrent, and the available data do not separate their contributions.

The distinction matters for interpretation. The systematic component falls on municipalities that are not low-demand areas and therefore outside the primary object of the study. The dispersed

component falls on municipalities that are within it, but the classification of low-demand areas does not rest on individual municipal estimates: it is derived from demand-density thresholds applied to the analytical zoning, and borderline cases were reviewed with the transport authority, as described in Section 4.3. We regard this as a genuine limitation, and a larger survey sample in the Apennine municipalities as the natural way to address it. Appendix A, Table 6 reports the calibration residuals on which this assessment rests.

The second limitation concerns validation coverage. The absence of an independent benchmark for the non-resident component is a significant limitation of the framework in its current form: one quarter of the estimated mobility rests on mobile network data alone (Section 5, Figure 5).

For National Statistical Offices, the framework proposed in this paper offers a concrete example of how MND can be integrated into the production of official-like statistics without replacing traditional sources. The approach is aligned with the ongoing efforts of the European Statistical System (ESS) to develop an open methodological standard for MNO data, as outlined in the Eurostat position paper on reusing MNO data for official statistics [6] and implemented through the Multi-MNO and MNO-MINDS research projects. In particular, the modular and multi-source design of the pipeline described here directly addresses the key requirements identified by the ESS: openness, transparency, modularity and support for the fusion of MNO data with non-MNO data. The modular structure of the workflow means that individual components – such as the gravity-based intra-zonal estimation or the FCD-based outlier correction – can be updated or replaced independently without rebuilding the entire pipeline. The MIND project further supports this perspective by developing the data-management infrastructure necessary for systematic, repeated production of mobility indicators, moving beyond one-off analytical exercises towards a sustainable monitoring capability [9]. For local authorities, the framework provides a structured

way to translate complex mobility information into operational decisions on DRT services and on the classification of the analytical zones.

7.1 Limitations arising from macro-level data access

Mobile network data can be accessed at three levels. At the *nano* level they consist of raw device–time–cell events; at the *micro* level, of summaries for individual devices; at the *macro* level, of counts aggregated above a confidentiality threshold. Privacy regulation and commercial practice mean that statistical producers and researchers ordinarily receive macro-level output, as we did. The consequence is structural rather than practical: any selection bias correlated with the mobility variable of interest cannot be corrected downstream, because the aggregation has removed the information that the correction would require.

In our experience this constraint is not prohibitive, provided three conditions are met; they are worth stating because they are the practical precondition for mobile network data to enter official statistics. The first is that the purpose of the extraction be defined before the data are requested, since it determines which parameters matter and which distortions are tolerable. The second is that the mitigation measures and the independent benchmarks used for validation be calibrated on that purpose rather than adopted generically. The third, and the one most often absent, is genuine technical dialogue with the provider: the iterative configuration of preprocessing parameters described in Section 2.4 is possible only where the provider will run alternative configurations and expose their outputs to comparison, and the same holds for the reporting, in advance of the extraction, of what a given specification would cause to be withheld.

Where these conditions hold, iteration on outputs converges on a configuration whose residual error, measured against independent sources, is small enough not to affect the conclusions the data

are used to support. The agreement between the fused matrix and the survey on the self-contained components, reported in Figure 4, is evidence of such convergence: a preprocessing configuration distorting the data systematically would not have spared those components.

What none of this delivers is a quantified uncertainty interval attributable to preprocessing. We can characterise the direction and the plausible magnitude of the effects; we cannot propagate them formally, because propagation would require knowledge of the preprocessing that macro-level release does not convey. Closing that gap, through standardised preprocessing metadata supplied alongside aggregated output, would remove the last obstacle to the routine use of these data in official statistics rather than the principal one.

Two further limitations follow from the same constraint. The first concerns reconstruction: both the re-splitting of the aggregated operator zones and the recovery of entries withheld at release restore, by assumption, information that the delivered data do not contain (Section 2.4). The re-splitting affects twelve of the 192 operator zones; in neither case is the effect on the estimates measurable. The second concerns temporal transfer: the parameters derived from floating car data were estimated for October 2023 and applied to observation months of 2024, while the trip rates derived from the mobility survey are those of the 2016 wave, carried forward with the population actualised to 2024 (Section 2.3). Both rest on an assumption of structural stability which we state rather than test.

7.2 Operational replicability and mobility monitoring

A distinctive feature of the framework is that it was designed from the outset to be not only replicable in different territories but also updatable over time. The modular architecture of the data-fusion pipeline allows the integrated outputs – both the aggregate OD matrix and the detailed

database for low-demand areas – to be regenerated whenever one or more input sources are refreshed, without the need to rebuild the entire processing chain from scratch. This property transforms the framework from a one-off analytical exercise into a continuous monitoring tool, capable of tracking the evolution of mobility behaviour and of assessing the impact of policy interventions, infrastructure changes and demographic shifts.

Three distinct update scenarios have been formalised, each involving a different subset of the processing chain [8].

The first scenario concerns the update of socio-economic data – for example, when new ISTAT population estimates or employment registers become available. In this case, the demographic base of each analytical zone is recalculated, the survey-derived mobility indices are reapplied to the new population structure, the share of intra-zonal trips is re-estimated to reflect the updated internal composition of the analytical zones, and the variation in trip volumes at the municipal level is propagated to the MND matrix. Both final outputs are then regenerated. This procedure allows the mobility picture to track demographic change (population growth, aging, suburbanisation) with minimal additional data acquisition.

The second scenario involves the **update of the mobility survey** – for instance, when a new survey campaign is conducted in preparation for a SUMP revision. Here, the mobility indices by user profile are re-estimated from the new survey, percentage changes by municipality are calculated and applied to the MND matrix, the intra-zonal estimation is recalibrated and the survey-based OD matrix is reconstructed on the updated volumes. This ensures that changes in travel behaviour captured by the new survey are reflected consistently across all outputs.

The third scenario covers the update of MND and FCD data. Since both sources are acquired from private providers and can in principle be refreshed periodically, this scenario allows the most

comprehensive update. The FCD-based outlier correction algorithm is reapplied, variation indices are calculated for each MND zone, the intra-zonal share is recalibrated on the new data, and the correction coefficients linking MND and survey matrices are recalculated. Both outputs are regenerated, providing an updated snapshot of mobility demand that is consistent with the new observational evidence.

In each scenario, only the affected stages of the pipeline need to be re-executed, while the remaining components and parameters are kept stable. This incremental approach substantially reduces the cost and time required for updates, making it feasible to maintain a periodically refreshed mobility knowledge base – an essential condition for effective transport planning and contract management.

The same modularity supports the transferability of the framework to other metropolitan areas. The core requirements are: the availability of standard official data (ISTAT census, employment registers) as the statistical backbone; access to MND from at least one mobile operator with sufficient market coverage; a sample of FCD for vehicle-flow benchmarking; and a mobility survey – even a pre-existing one – to calibrate modal split, trip-purpose distributions and intra-zonal shares. The alignment with the ART regulatory framework [10,11], which applies uniformly across Italian metropolitan areas, further facilitates adoption: the indicators and classification thresholds defined for Bologna can be recalibrated to local conditions while maintaining methodological consistency.

The MIND project contributes directly to this operational dimension by developing the data-management infrastructure – including data ingestion, processing and dissemination tools – that enables the systematic, repeated production of mobility indicators for planning and contract management [9]. The long-term vision is a mobility monitoring system that can be queried by

planners, regulators and researchers to obtain up-to-date indicators at any point in time, thereby supporting both strategic planning (SUMP preparation and revision) and operational management (DRT contract monitoring, service-level verification, assessment of the effects of network changes).

8. Conclusions

This paper has presented a replicable methodological framework for integrating Mobile Network Data, Floating Car Data, survey and register-based information to support official statistics on mobility and the planning of demand-responsive transport services. Developed within the DREAM_PACE and MIND projects and applied to the Metropolitan City of Bologna, the framework produces a coherent set of indicators describing origin–destination flows, the intensity of mobility at the level of the analytical zones and the identification of low-demand areas through a data-driven enrichment of the regulatory criteria.

The results demonstrate that the integrated approach can reveal previously unobserved aspects of mobility – most notably the magnitude of non-resident and crossing trips – while remaining consistent with established survey-based indicators at aggregate levels. The decomposition of the 1.17-million-trip discrepancy with earlier estimates into identifiable components strengthens confidence in the reliability of the fused estimates: additional resident trips, concentrated in movements between municipalities other than Bologna, trips by non-resident city users, and crossing movements. The derived classification of the analytical zones and the mapping to DRT service profiles provide directly usable inputs for planning and regulatory decisions, and the two-level comparison between the statutory DM 157/2018 classification and the data-driven classification offers a richer evidence base for local decision-makers.

Future work will focus on three directions. First, refining the treatment of uncertainty through error propagation along the fusion chain, a direction that is partly methodological and partly dependent on providers supplying standardised preprocessing metadata alongside macro-level output (Section 7.1). Second, extending the framework to incorporate additional data sources, such as smart-card transactions and public transport ticketing data, which could strengthen validation and enable real-time monitoring. Third, moving towards closer integration with the regular production processes of official statistics – a direction already initiated by the MIND project – for example through the development of standardised indicator sets and documentation that can be disseminated to a broader audience of planners, regulators and researchers.

Acknowledgements

The DREAM_PACE project (CE0100046) was implemented through financial assistance from ERDF funds of the Interreg CENTRAL EUROPE Programme. The MIND project (CUP F55G24000020004) is co-funded by the European Union under the Emilia-Romagna ERDF Regional Programme 2021–2027. This publication reflects only the authors' views; the programme authorities are not liable for any use that may be made of the information contained therein.

The mobile network data used in this study were provided by Vodafone Italia S.p.A. (now Fastweb S.p.A.) through the Vodafone Analytics service, which collected, irreversibly anonymised and aggregated the data into origin–destination matrices. The authors would like to thank Andrea Zaramella, Dario Di Sorte and Luigi Mancino (Fastweb S.p.A.) for their valuable contribution on mobility insights from mobile network operator big data.

The authors used generative AI tools in preparing this revision: to assist in drafting and editing the revised text and the response to the reviewers, to compute the summary statistics of Table 6 from

the authors' data, to prepare draft versions of Figures 2, 3 and 5 that were subsequently produced by the authors' graphics team, and to assist in compiling the reference list. All content was checked by the authors against the original sources and data, and the authors take full responsibility for it.

References

1. Ricciato F, Widhalm P, Craglia M, Pantisano F. Estimating population density distribution from network-based mobile phone data. Luxembourg: Publications Office of the European Union; 2015. Report No.: EUR 27185 EN.
2. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, et al. Dynamic population mapping using mobile phone data. Proc Natl Acad Sci USA 2014;111(45):15888-93.
3. Blondel VD, Decuyper A, Krings G. A survey of results on mobile phone datasets analysis. EPJ Data Sci 2015;4:10.
4. Iqbal MS, Choudhury CF, Wang P, González MC. Development of origin–destination matrices using mobile phone call data. Transp Res Part C Emerg Technol 2014;40:63-74.
5. Calabrese F, Diao M, Di Lorenzo G, Ferreira J, Ratti C. Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. Transp Res Part C Emerg Technol 2013;26:301-13.
6. Eurostat. Reusing mobile network operator data for official statistics: the case for a common methodological framework for the European Statistical System – 2023 edition. Luxembourg: Publications Office of the European Union; 2023. Report No.: KS-FT-23-001. Available from: https://ec.europa.eu/eurostat/web/products-statistical-reports/w/ks-ft-23-001

7. International Telecommunication Union. Handbook on the use of mobile phone big data for measuring the information society. Geneva: ITU; 2022. Available from: https://unstats.un.org/wiki/spaces/MPDMIS/overview

8. DREAM_PACE Consortium. Deliverable "D1.2.3 Co-designed solutions blueprint of integrated DRT implemented/tested through pilot activities", Annex 2 "BOLOGNA LIVING LAB - Analysis of potential demand and definition of areas with low demand at metropolitan level (in terms of geographical, socio-economic, temporal characteristics and relationships between territories), with identification of a methodology and parameters to define the different degrees of demand and possible implementation/indication of tools to replicate this analysis at different times and in the future". 2025. Available on DREAM_PACE project website (https://www.interreg-central.eu/wp-content/uploads/2025/11/D1.2.3_Co-designed-solutions-blueprint-integrated-DRT-With-Annexes.pdf, pages 99-152).

9. MIND Consortium. Mobility INsight Data: innovative project management software to improve the quality of transport planning. Bologna: MIND project (PR FESR Emilia-Romagna 2021–2027, Priority 3, Action 2.8.2); 2025. Available from: https://srmbologna.it/?page_id=9369

10. Autorità di Regolazione dei Trasporti. Delibera n. 48/2017: Atto di regolazione recante la definizione della metodologia per l'individuazione degli ambiti di servizio pubblico e delle modalità più efficienti di finanziamento. Turin: ART; 2017. Available from: https://www.autorita-trasporti.it/delibere/delibera-n-48-2017/

11. Ministero delle Infrastrutture e dei Trasporti. Decreto ministeriale n. 157 del 28 marzo 2018: Definizione dei costi standard dei servizi di trasporto pubblico locale e regionale e dei criteri

di aggiornamento e applicazione. Gazzetta Ufficiale della Repubblica Italiana, Serie Generale n. 126, 1 June 2018.

12. Città Metropolitana di Bologna. Piano Urbano della Mobilità Sostenibile (PUMS). Approved by Metropolitan Council Decision No. 54, 27 November 2019. Available from: https://pumsbologna.it/
13. European Parliament and Council. Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation). OJ L 119, 4 May 2016, p. 1-88.
14. Istituto Nazionale di Statistica (ISTAT). *La mobilità territoriale — Trasferimenti di residenza e spostamenti quotidiani* [Territorial mobility — Changes of residence and daily commuting]. Commuting indicators from the Permanent Census, reference year 2019, current municipal geography. Rome: ISTAT. Available from: https://storymaps.arcgis.com/stories/744b95b92a1248c3a18550e79c0af252; municipal dashboard: https://istat.maps.arcgis.com/apps/dashboards/8ec70613105d4846849ca4cb49d416fb.
15. DREAM_PACE Consortium. Deliverable "D1.2.3 Co-designed solutions blueprint of integrated DRT implemented/tested through pilot activities", Annex 4 "BOLOGNA LIVING LAB - Study on potential integrations between demand assessment methodologies and parameters (and related tools, if any) and citizen and metropolitan planning tools (e.g., the future SUMP)“. 2025. Available on DREAM_PACE project website (https://www.interreg-central.eu/wp-content/uploads/2025/11/D1.2.3_Co-designed-solutions-blueprint-integrated-DRT-With-Annexes.pdf, pages 200-236).

16. Città metropolitana di Bologna. PUMS — Monitoraggio biennale 2021–2022 [SUMP — Biennial monitoring report 2021–2022]. Annex 1: Glossario degli indicatori. Bologna: Città metropolitana di Bologna; 2024. Available from: https://pumsbologna.it/Engine/RAServeFile.php/f/monitoraggio_2021_2022/Allegato%201%20-%20Glossario.pdf

17. Engels D, Ambrosino G, Boero M. *Scenari realizzativi e tipologie di servizio* [Implementation scenarios and service types]. Chapter 3. In: Ambrosino G, Romanazzo M, editors. *I servizi flessibili di trasporto per una mobilità sostenibile* [Flexible transport services for sustainable mobility]. Rome: ENEA; 2002. ISBN 88-8286-014-0. Available from: https://www.enea.it/it/seguici/pubblicazioni/pdf-volumi/2002/v2002-servizi_flessibili.pdf

Appendix A. Gravity models for self-contained trips

A.1 Purpose. The gravity models supply one quantity: the share of movements that begin and end within the same spatial unit, as against those crossing its boundary. Movements internal to an MND zone are absent from the operator's output by construction, and this share is what allows that component to be recovered. The models apportion rather than generate volumes, and the mobile network data remain the quantitative anchor throughout.

A.2 Functional form and parameters. Two regimes are distinguished. For relations within the municipality of Bologna,

$N_{ij} = (POP_i \times ADD_j) / d_{ij}^4$

and for all other relations, including exchanges between Bologna and the remaining municipalities,

$N_{ij} = (POP_i \times ADD_j) / d_{ij}^{5.5}$

where POP_i is the resident population of the origin analytical zone, ADD_j the number of employees in the destination analytical zone, and d_{ij} the distance between them. The apportionment coefficient is $p_{ij} = N_{ij} / \Sigma_j N_{ij}$.

A.3 Calibration. The two exponents were estimated by minimising the sum of squared deviations across the 55 municipalities of the Metropolitan City between the intra-municipal component implied by the mobility survey and the intra-municipal component implied by the models, both applied to the same survey totals. These totals are expressed in home-based outbound movements, one record for each departure from the residence, rather than in individual trips. For Bologna the calibration uses the four relations between the two survey macro-zones into which the municipality is divided. The models are estimated on municipalities, applied on the analytical zoning, and their output is expressed on the MND zoning: using the correspondence between analytical zones and MND zones (Appendix B, Table 7), the resulting shares are aggregated to obtain, for each MND zone, the intra-zonal share and hence the multiplicative coefficients applied to the observed matrix.

The estimated exponents are steeper than those commonly reported. These models are required to reproduce one quantity, the intra- and inter-municipal shares recorded in the survey, and conventional values did not reproduce them, while the estimated values do. Since the models apportion trips between two categories rather than represent the distribution of travel over distance in any behavioural sense, fit to the observed shares is the criterion we apply.

A.4 Calibration and validation. Table 6 reports, for each of the 55 municipalities, the intra-municipal share implied by the models against the share recorded in the survey. This comparison is the calibration, and we report its residuals as such. Validation is a separate question, and it is answered downstream: whether the fused matrix reproduces the survey on the components the

models feed. It does, within 2 per cent for movements internal to Bologna and within 2 per cent for intra-municipal movements in the remaining municipalities (Figure 4).

Table 6. Calibration residuals of the gravity models, by municipality.

Municipality	Size quartile	Survey intra-municipal share (%)	Model intra-municipal share (%)	Residual (pp)
Alto Reno Terme	Q2	60.9	87.2	+26.3
Anzola dell'Emilia	Q3	48.8	70.4	+21.6
Argelato	Q3	59.5	30.4	−29.1
Baricella	Q2	59.1	78.2	+19.2
Bentivoglio	Q2	59.1	52.9	−6.2
Bologna	Q4	91.5	99.5	+8.0
Borgo Tossignano	Q1	44.6	81.1	+36.5
Budrio	Q4	62.7	79.0	+16.3
Calderara di Reno	Q3	49.8	90.3	+40.4
Camugnano	Q1	62.2	16.2	−46.1
Casalecchio di Reno	Q4	55.0	98.8	+43.8
Casalfiumanese	Q1	44.3	31.5	−12.9
Castel d'Aiano	Q1	61.7	46.0	−15.8
Castel del Rio	Q1	44.5	92.4	+47.8
Castel di Casio	Q1	60.3	32.9	−27.4
Castel Guelfo di Bologna	Q1	54.2	55.3	+1.1
Castel Maggiore	Q4	48.3	49.9	+1.7
Castel San Pietro Terme	Q4	55.9	72.7	+16.9
Castello d'Argile	Q2	58.2	62.4	+4.2
Castenaso	Q3	51.1	54.0	+3.0
Castiglione dei Pepoli	Q2	60.9	41.3	−19.6
Crevalcore	Q3	65.1	93.3	+28.2
Dozza	Q2	54.8	71.5	+16.7
Fontanelice	Q1	43.0	14.6	−28.3
Gaggio Montano	Q2	60.0	49.2	−10.8
Galliera	Q2	58.8	56.2	−2.6
Granarolo dell'Emilia	Q3	47.4	59.1	+11.7
Grizzana Morandi	Q1	60.3	45.3	−15.0
Imola	Q4	86.5	93.1	+6.7
Lizzano in Belvedere	Q1	60.7	77.3	+16.6
Loiano	Q1	45.8	48.7	+2.9
Malalbergo	Q3	59.2	50.7	−8.5
Marzabotto	Q2	60.2	54.0	−6.1
Medicina	Q4	55.0	88.6	+33.6
Minerbio	Q3	59.1	54.5	−4.6
Molinella	Q4	63.4	95.6	+32.2
Monghidoro	Q1	45.9	75.3	+29.4
Monte San Pietro	Q3	65.2	49.9	−15.3
Monterenzio	Q2	44.8	94.8	+50.0
Monzuno	Q2	59.6	40.5	−19.1
Mordano	Q2	55.1	92.9	+37.8
Ozzano dell'Emilia	Q3	53.6	22.7	−31.0
Pianoro	Q4	55.2	81.3	+26.1
Pieve di Cento	Q2	60.0	76.0	+16.0
Sala Bolognese	Q3	65.1	15.5	−49.7
San Benedetto Val di Sambro	Q1	61.3	66.4	+5.1
San Giorgio di Piano	Q3	58.7	12.0	−46.7
San Giovanni in Persiceto	Q4	65.4	81.8	+16.3
San Lazzaro di Savena	Q4	54.4	78.6	+24.1
San Pietro in Casale	Q3	58.8	88.8	+30.0
Sant'Agata Bolognese	Q2	65.1	79.1	+13.9
Sasso Marconi	Q4	55.2	54.1	−1.1
Valsamoggia	Q4	64.8	71.1	+6.2
Vergato	Q3	61.3	39.6	−21.6
Zola Predosa	Q4	54.4	86.3	+31.9

Quartile	n	Mean residual (pp)	Mean absolute residual (pp)	SD (pp)	Min	Max
Q4 (largest)	14	+18.8	18.9	13.6	−1.1	+43.8
Q3	14	−5.1	24.4	28.9	−49.7	+40.4
Q2	14	+8.5	17.8	20.9	−19.6	+50.0
Q1 (smallest)	13	−0.5	21.9	27.5	−46.1	+47.8
All	55	+5.5	20.7	—	−49.7	+50.0

A.5 Structure of the residuals. Across the 55 municipalities the mean residual is +5.5 percentage points and the mean absolute residual 20.7 points. The residuals are not homogeneous, and their structure differs between two groups of municipalities, discussed qualitatively in Section 7.

In the municipalities of the first belt around Bologna the models systematically overstate the self-contained share: the mean residual over the ten municipalities bordering the core is +20.3 points, positive in nine of the ten, with the largest values in Casalecchio di Reno (+43.8), Calderara di Reno (+40.4) and Zola Predosa (+31.9). Over the quartile of the largest municipalities the mean residual is +18.8 points with a standard deviation of 13.6, a systematic and unidirectional bias rather than dispersion.

In the Apennine municipalities the pattern is the opposite in kind: over the eleven municipalities identified by DM 157/2018 the mean residual is −3.0 points while the mean absolute residual is 21.5, and over the quartile of the smallest municipalities the corresponding figures are −0.5 and 21.9. Individual values are large in both directions, while the mean is close to zero: Camugnano −46.1, Castel del Rio +47.8, Fontanelice −28.3. The discrepancy has no systematic direction, which is what sampling error in the reference term produces.

The dispersion is not monotone in municipal size: the mean absolute residual is 21.9 points in the smallest quartile and 24.4 in the third, and the correlation between the absolute residual and the logarithm of municipal size is −0.16. Sampling error in the reference term is therefore one component of the dispersion, not the whole of it.

A.6 Relation to the discrepancy discussed in Section 6. The divergences between the fused matrix and the survey are not distributed evenly across macro-components: they fall on the inter-municipal and external-exchange components, which are the categories a household survey of resident families is least equipped to record, while the two self-contained components are reproduced within 2 per cent. This concentration is **compatible with** the mechanisms set out in Section 6; it does not discriminate between them. What the pattern does support is the narrower inference stated there: the difference is not an artefact of the fusion procedure, since a procedure distorting the data systematically would not have spared the self-contained components.

Note to Table 6. Size quartiles are defined on the survey-based totals, in decreasing order: Q4 comprises the 14 largest municipalities, Q3 and Q2 the following 14 each, Q1 the 13 smallest.

Appendix B. Municipalities of the Metropolitan City of Bologna

The study covers the 55 municipalities of the Metropolitan City of Bologna. Table 7 lists them and reports, for each, the number of analytical zones and MND zones it contains and its status under the two classifications of low demand described in Section 4.3: the regulatory classification, which applies the indicators of DM 157/2018 to entire municipalities, and the data-driven classification, which applies the demand-density thresholds to the analytical zoning and identifies 23 low-demand areas.

Both zoning systems nest exactly within municipal boundaries: no analytical zone and no MND zone extends across more than one municipality, so that the counts in Table 7 sum to the 1,050 analytical zones and the 144 MND zones internal to the metropolitan city. The operator zoning on which the data were delivered does not have this property, and is therefore not reported by municipality; the aggregation chain and its restoration are described in Section 2.4.

Low-demand areas, by contrast, are groupings of portions of municipal territory and do not follow municipal boundaries. Fifteen of the 23 areas extend across more than one municipality, and a single municipality may contain portions of up to four different areas. The area identifiers listed for each municipality therefore do not sum to 23 across the table.

Municipality names follow the current ISTAT denominations. Two of the 55 are the result of mergers: Valsamoggia, established on 1 January 2014 from the former municipalities of Bazzano, Castello di Serravalle, Crespellano, Monteveglio and Savigno; and Alto Reno Terme, established on 1 January 2016 from Granaglione and Porretta Terme. The metropolitan city comprised 60 municipalities until 2013 and 55 from 2016 onwards. Both zoning systems and all sources used in the classification are defined on the current municipal boundaries.

Table 7. Municipalities of the Metropolitan City of Bologna: zone counts and low-demand status.

Municipality	Analytical zones	MND zones	DM 157/2018	Data-driven areas
Alto Reno Terme	16	2	No	1
Anzola dell'Emilia	14	3	No	23
Argelato	12	2	No	15
Baricella	10	1	No	21
Bentivoglio	13	3	No	—
Bologna	163	16	No	—
Borgo Tossignano	7	2	No	—
Budrio	22	2	No	17, 19
Calderara di Reno	11	2	No	—
Camugnano	15	2	Yes	3
Casalecchio di Reno	16	2	No	—
Casalfiumanese	19	3	No	5, 9
Castel d'Aiano	11	2	Yes	2
Castel del Rio	7	2	Yes	5
Castel di Casio	13	2	Yes	—
Castel Guelfo di Bologna	8	1	No	—
Castel Maggiore	12	4	No	—
Castel San Pietro Terme	29	3	No	9, 10, 13
Castello d'Argile	7	2	No	—
Castenaso	13	2	No	—
Castiglione dei Pepoli	15	3	Yes	—
Crevalcore	24	3	No	16
Dozza	8	2	No	—
Fontanelice	7	2	No	5
Gaggio Montano	16	3	Yes	2
Galliera	12	2	No	—
Granarolo dell'Emilia	9	2	No	—
Grizzana Morandi	18	3	Yes	3
Imola	70	4	No	9, 11, 12, 14
Lizzano in Belvedere	13	2	Yes	1
Loiano	10	2	Yes	—
Malalbergo	11	2	No	21
Marzabotto	15	3	No	6
Medicina	30	2	No	12, 19
Minerbio	12	2	No	17
Molinella	23	1	No	17, 18, 19

Municipality	Analytical zones	MND zones	DM 157/2018	Data-driven areas
Monghidoro	12	2	Yes	4
Monte San Pietro	24	3	No	7
Monterenzio	20	3	No	5, 10
Monzuno	18	3	No	4
Mordano	8	1	No	—
Ozzano dell'Emilia	16	3	No	10
Pianoro	28	4	No	8
Pieve di Cento	6	1	No	—
Sala Bolognese	9	2	No	15
San Benedetto Val di Sambro	19	2	Yes	4
San Giorgio di Piano	8	2	No	—
San Giovanni in Persiceto	23	3	No	15, 20
San Lazzaro di Savena	23	2	No	—
San Pietro in Casale	14	2	No	22
Sant'Agata Bolognese	6	1	No	—
Sasso Marconi	23	4	No	6
Valsamoggia	46	5	No	6, 7
Vergato	16	3	No	2, 6
Zola Predosa	20	2	No	23
Total	1,050	144	11 Yes	23 areas

Note to Table 7. Analytical zones and MND zones nest exactly within municipal boundaries; counts sum to 1,050 and 144 respectively. Low-demand area identifiers refer to the 23 areas of the data-driven classification (Section 4.3); an em dash indicates that no data-driven area falls within the municipality.