Ref. Ares(2025)6572244 - 12/08/2025

DATA SPACE 4 MOBILITY MODELS INTEGRATION

Project Acronym:	**23-ES-TG-DS4MOBMODELS**
Project ID:	**101176007**

Deliverable number:	**D4.1**
Deliverable name:	**Transport model design**
Work Package number:	**4. AI DATA SCIENCE AND SIMULATION TRANSPORT MODELS**
Due date:	**31/07/2025**
Delivery date:	**31/07/2025**

Lead Beneficiary:	**AIMSUN**
Main author:	**Antonio Pellicer and Jordi Casas (Aimsun)**
Contributors:	**Lidia Montero (UPC), Marta Cuatrecasas (UPC), Marc Manzanares (UPC), Allison Fernández-Lobo (UPM), Andrés Monzón (UPM)**

Co-funded by the European Union

DS4MOB MODELS

Document Information

The DS4MOBMODELS project (Grant Agreement nº: 101176007) project is co-funded by the EU. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the EU or the CINEA. Neither the EU nor the CINEA can be held responsible for them.

Revision History

Version	Date	Lead Beneficiary	Main area of changes
V 0.1	02/07/2025	AIMSUN	Initial draft
V 0.2	24/07/2025	AIMSUN	Final draft to be reviewed
V 0.3	29/07/2025	UPM/UPC	Reviewed version. Draft for final Q&A
V 1.0	31/07/2025	AIMSUN	Final

Glossary

ANG	Compressed proprietary binary file format used to store an entire Aimsun traffic model
ARIMA	Autoregressive integrated moving average family of models
ATM	Metropolitan Transport Authority (Catalunya)
AVI	Automatic vehicle identification
CNN	Convolutional neural Networks
DTA	Dynamic Traffic Assignment
ETIS	European Transport policy Information System
EU	European Union
FCD	Floating Car data
GNN	Graph Neural Networks
ICT	Information and Communication Technologies
IDM	Incident Detection Module
ITS	Intelligent Transport Systems
KNN	k-Nearest Neighbours
ML	Machine Learning
NMF	Network Mobility Forecaster
OD	Origin – Destination
RF	Random forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
RP	Revealed preference
SP	Stated preference
SQV	Scalable Quality Value
SVM	Support Vector Machines
TEN-T	Trans-European Transport Network
V2I	Vehicle-to-infrastructure communication
V2P	Vehicle-to-pedestrian
V2V	Vehicle-to-vehicle
WP	Work Package

Table of Content

Executive Summary ***6***

List of Figures ***7***

List of Tables ***8***

1. Introduction ***9***

1.1. Purpose and objectives **9**

1.2. Intended audience **9**

1.3. Project summary **9**

1.3.1. TEN-T: strategic context and Spanish corridors 10

1.4. Structure of the deliverable **11**

2. Review of Transport Models ***12***

2.1. Demand models in the literature **15**

2.2. Classification of transport simulation models **18**

2.2.1. Macroscopic Models 18

2.2.2. Mesoscopic Models 21

2.2.3. Microscopic Models 22

2.3. Data sources and fusion in transport modelling **23**

2.3.1. Types of Data Sources 25

2.3.2. Sensors and Available Variables 27

2.3.3. Surveys 29

2.3.4. ICT Data in Vehicle and Passenger Mobility 30

2.3.5. Data sources in Spain 33

2.4. Multimodal interaction and TRANS-TOOLS model in TEN-T Corridors **34**

2.4.1. TRANS-TOOL: Specific Transport Model for TEN-T 35

3. Review of Data-Driven solutions for Transport Applications ***39***

3.1. Traffic monitoring and pattern generation **39**

3.2. Incident detection **39**

3.3. Traffic predictions **40**

4. Overview of Use Cases ***41***

5. Models and Use Cases: Identifying gaps and necessary solutions ***43***

5.1. Existing transport models of the use cases **43**

5.2. Existing data-driven solutions for application **44**

5.2.1. Raw data analyzer and cleaner algorithm 45

5.2.2. Pattern generation algorithm 45

5.2.3. Network mobility forecaster 46
5.2.4. Aimsun incident detection method 46
5.2.5. Speed recommendation 46
5.3. Methodology for identifying gaps and needs 47
5.4. Use Case classification based on data-driven and simulation-based solutions 48
5.5. Preliminary identification of needs and gaps 52
6. Conclusions 54
7. References 55
Appendix 1: TEN-T network description 58
1. Context 58
1.1. TEN-T Atlantic Corridor 58
1.2. TEN-T Mediterranean Corridor 62
Appendix 2: Travel demand modelling phases 69
2.1. Demand Generation Models 69
2.2. Demand Distribution Models 69
2.3. Modal Split Models of Demand 71
2.4. Demand Assignment Models 72
2.5. Macroscopic Assignment Models in Private Transport Networks 73
2.6. Public Transport Network Assignment Models 74

Executive Summary

This deliverable presents the technical design of transport models to be applied within the DS4MOB project with the aim of supporting the implementation of the predefined use cases. It begins with a concise review of current transport modelling approaches, including simulation-based and data-driven techniques, to establish a shared understanding of the available methodologies and their practical applications.

Building on this foundation, the deliverable analyses the predefined use cases and explores how they can be addresses using the appropriate modelling approaches. This includes assessing the feasibility of applying specific models based on the nature of the use case the availability of data, and the operational context. The analysis helps identify which use cases can be deployed with exiting tools and where limitations may require further development or adaptation.

Throughout this process, several key gaps and needs have been identified – such as need to establish communication with stakeholders to know the existence of absence of calibrated transport models, the access to relevant datasets, and challenges related to cross-border and freight transport scenarios. These findings are essential for guiding the next steps in model development, stakeholder engagement, and data acquisition.

Overall, this deliverable provides a clear and structured basis for the technical implementation of the DS4MOB use cases, ensuring that the selected modelling approaches are both feasible and aligned with the use cases requirements.

List of Figures

Figure 1. TEN-T Atlantic Corridor (left) and Mediterranean Corridor (right) in Spain. 11
Figure 2. Four-Step Modelling Framework Schema 16
Figure 3. Macroscopic models: components and relations 21
Figure 4. DIKW Hierarchy. Source(Rowley, 2007) 25
Figure 5. Structure of the TT3 model system. ('TRANS-TOOLS 3', 2017) 38
Figure 6. Data-driven modules for traffic monitoring, pattern generation, predictions and incident detection developed by Aimsun. 45
Figure 7. Methodology to explore data-driven and/or simulation-based solutions for the predefined Use Cases 48
Figure 8. Classification of use cases based on data-driven and/or simulation-based approaches 49
Figure 9. TENT-T Atlantic Corridor 59
Figure 10. TEN-T Atlantic Corridor in Spain and Portugal. 60
Figure 11. TEN-T Network in Madrid Region. 61
Figure 12. TEN-T Mediterranean Corridor 63
Figure 13. TEN-T Mediterranean Corridor – Spain. 64
Figure 14. Madrid TEN-T network for Atlantic and Mediterranean Corridors. 65
Figure 15. TEN-T Network in Catalonia Region and Barcelona city. 66
Figure 16. TENT Network in Valencia Region and Valencia city. 68

List of Tables

Table 1. Classification of Measurement Data According to (Ou, 2011) 26
Table 2. Traditional Sensor Types and Obtainable Variables 27
Table 3. Types of ICT Sensors and Obtainable Variables 31
Table 4. Key multimodal urban nodes in the Atlantic and Mediterranean Corridors by Mode of transport. 35
Table 5. TRANS-TOOLS versions. 37
Table 6. Summary of predefined use cases and their description 41
Table 7. Existing transport models according to the knowledge of the DS4MOB partners. 44
Table 8. Potential Data-Driven Use Cases and minimum data requirements 50

1. Introduction

1.1. Purpose and objectives

The purpose of this deliverable is to define the technical design of transport models that will be used to address the requirements of the predefined use cases within the project. This includes identifying the most appropriate modelling approaches – whether simulation-based, data-driven or hybrid – for each use case, based on their specific characteristics, data and model availability, and operational context.

To support this goal, the deliverable is structured around the following key objectives:

- **Review existing modelling approaches.** Provide a concise overview of current transport modelling methodologies, including macroscopic, mesoscopic, and microscopic model, as well as data-driven approaches. This review sets the foundation for understanding the capabilities and limitations of each modelling type in addressing different mobility scenarios.
- **Alignment of use cases with modelling techniques**. Analyze the predefined use cases and assess their compatibility with various modelling approaches. This involves evaluating the technical feasibility of applying specific models to each use case.
- **Identification of gaps and needs**. Identify limitations in existing models or data that may hinder the implementation of certain use cases. This includes recognizing where models are missing, not calibrated, or insufficiently detailed, as well as where data availability may constrain the application of data-driven models.
- **Preliminary design of model applications**. Begin shaping the technical design of the transport models to be developed or adapted for the project. This includes outlining minimum data requirements that will be needed to apply such solutions.

1.2. Intended audience

This deliverable will be **disseminated publicly** informing all interested stakeholders about the technical design of the transport models to fulfil the Use Case requirements, as well as inform the rest of the DS4MOB project partners about the advancements in aligning modelling techniques with use case's needs.

1.3. Project summary

The DS4MOBMODELS project aims to define and create an advanced Mobility Data Space that integrates transport, traffic, and travel data across various regions, including the Spanish state, the Nouvelle-Aquitaine-Euskadi-Navarre Euroregion, and the Central Portugal region.

The primary goal of the project is to enhance the availability, quality, and interoperability of mobility data to address key national, cross-border, regional, and local use cases related to sustainable urban mobility, traffic management, and real-time operations within the Trans-European Transport Network (TEN-T).

The project will develop a data platform that will support the integration of various mobility models and data science tools to facilitate efficient, safe, and sustainable mobility management. The initiative will also establish robust data governance in line with European and national regulations, ensuring secure and interoperable access to the data for all the stakeholders in a non-discriminatory way.

The DS4MOBMODELS project aligns with the broader objectives of the European Data Strategy, the Digital Europe Programme, and national digital and mobility strategies in Spain, France and Portugal, contributing to the creation of a more integrated and resilient European mobility infrastructure.

1.3.1. TEN-T: strategic context and Spanish corridors

The **Trans-European Transport Network (TEN-T)** is a strategic infrastructure initiative by the European Union aimed at improving transport connectivity across member states by developing an integrated, efficient, and sustainable transport infrastructure. To support the development of this network, the European Commission established the **Connecting Europe Facility (CEF)**, a financial instrument designed to promote investment in key transport projects. In the transport sector, CEF plays a crucial role in funding and coordinating initiatives that enhance cross-border mobility and multimodal transport system through the seamless integration of different transport modes throughout the EU.

This project is focused on two of the nine TEN-T corridors: **Atlantic and Mediterranean, which involves Spain and the proposed pilots in cross-border regions with Portugal and France.** Atlantic and Mediterranean Corridors are focused on the transportation of people, goods or both, and have different levels of relevance within the network and specifically within the different countries involved. In Spain, these corridors cover the main urban areas and references point in the transport networks. A more detailed description of the Atlantic and Mediterranean corridors is provided in **Appendix 1**.

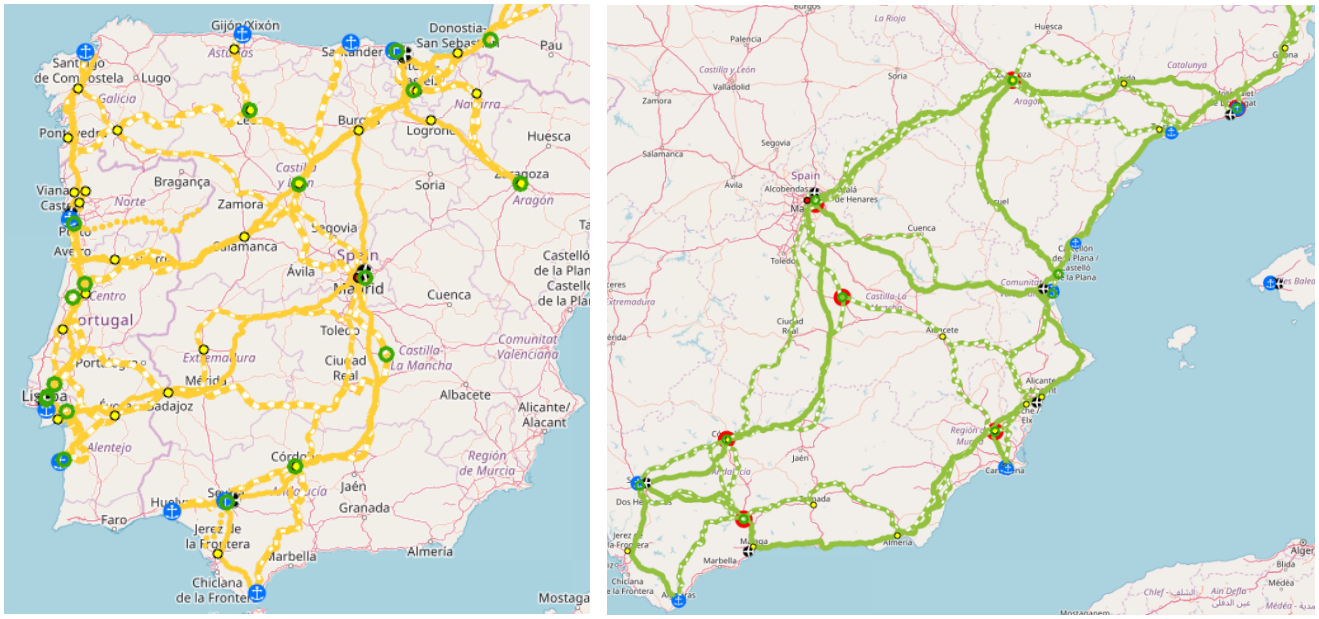


Figure 1. TEN-T Atlantic Corridor (left) and Mediterranean Corridor (right) in Spain.

1.4. Structure of the deliverable

This deliverable is structure into 6 sections with the corresponding description as follows:

- Section 2 presents an overview of transport models, their classification, and types of data sources used to support them. This section lays the foundation for understanding the data requirements essential for developing and applying transport models.
- Section 3 explores various data-driven solutions, showcasing the range of possibilities these methodologies offer for transport analysis and decision-making.
- Section 4 summarises the different Use Cases identified for the DS4MOB project, outlining their scope.
- Section 5 presents the core content of this deliverable, including the identification of existing transport models and data-driven tools, the alignment between use cases and modelling approaches, the methodology for identifying gaps and needs, and a final classification of use cases based on simulation-based and/or data-driven approaches.
- Section 6 summarises the main findings and insights derived from the deliverable.

2. Review of Transport Models

Transport Models are computational frameworks used to simulate and analyse how people and goods move through a transport system. These models are a systematic representation by mathematical expressions of the transport network dynamics within a defined area having certain socio-economic and land-use characteristics. They are powerful tools for analysing mobility patterns and assessing the impact of changes in the transport operations and infrastructure. These changes may be due to changes in the demand for transport and/or due to changes in the transport network itself.

The area covered by the model may range from local urban or metropolitan road networks to transnational or continental corridors, considering cross-border transport nodes.

In general, these models help planners forecast travel demand, evaluate infrastructure investments, and assess policy impacts, since their output can provide keys insights into the understanding of an existing or future transport problem. This can be used to identify the likely impacts that will result from a proposed project, strategy or Transport Policy. As such, Transport Models play an essential role as a decision-support tool, providing relevant and accurate information into planning and decision making (Ortúzar and Willumsen, 2011).

Traffic management models are implemented for optimizing the use of road infrastructure, reducing congestion, enhancing safety, and minimizing environmental impacts. Globally, these models are used to simulate traffic dynamics, forecast demand, and evaluate the effectiveness of different traffic control strategies for both passenger and freight flows.

Related with passengers, traffic models are used to evaluate the impacts of congestion pricing, signal timing optimization, public transport integration, and active mobility infrastructure. For freight, these models are used to simulate delivery vehicle routes, logistics chain behaviour, and access to logistics hubs. Freight-specific models must consider vehicle restrictions, delivery time windows, and both (passenger and freight focuses) the multimodal connections.

Modern traffic models increasingly interact with Intelligent Transportation Systems (ITS) platforms to enable real-time monitoring and control for decision-making operators. This includes some applications as: Adaptive traffic signal control, Vehicle-to-infrastructure communication (V2I), Automated incident detection, Smart routing and congestion management using AI and machine learning.

In Spain, macroscopic traffic flow models are primarily used, which treat traffic as a continuous flow, suitable for long-distance corridors and freight logistics planning. Mesoscopic models are also applied in metropolitan areas to capture group behaviour with some individual vehicle characteristics. In major cities such as Madrid and Barcelona, urban traffic management relies on microscopic simulation models to model individual vehicle movements, and are used for optimizing traffic signals, evaluating infrastructure changes, and designing low-emission zones.

Transport Models typically involve two main components: **supply models**, which represent the available transport infrastructure and services—such as roads, railways, ports, and terminals, and **demand models**, which estimate how many people or goods need to move between different

locations. Demand models consider factors like population, economic activity, travel behaviour, and logistics flows, while supply models assess network capacity, travel times, and modal availability. Related to their relevance in TEN-T Networks, aligning supply (infrastructure) with expected demand (mobility and freight flows), transport models help guide strategic investment decisions, and ensure efficient network performance, promoting the development of a sustainable, interoperable European transport system.

Supply Models

Transport supply models represent the active (infrastructure) and passive (vehicles) assets that define the services available in a transport system. These models define the general characteristics, physical and technical, capacity, connectivity, and performance of all elements in the network. The network includes all assets for all transport models: roads and highways, railways, airports and airspace routes, ports and inland waterways, multimodal terminals and logistics hubs, public transport systems (bus, tram, metro, subways, etc.). These models are focused on describe what is available in the system (Ortúzar and Willumsen, 2011).

To develop a transport supply model, it is essential to gather comprehensive data and information related to the following elements:

- **Definition and Zoning of the Study Area:** To define the geographical boundaries of the study area and divide it into appropriate zones. Zoning provides a spatial structure that allows for the analysis of transport flows, infrastructure distribution, and connectivity patterns. Each zone typically reflects specific characteristics, such as population density, land use, or administrative function, and serves as the basis for linking transport demand with available infrastructure. Accurate zoning is fundamental to ensure reliable modelling outcomes and meaningful interpretation of transport dynamics within the system. In the project, these *zones correspond to the Pilots and use cases: National (Spain), Regional (Andalusia, Catalonia and Valencian), Metropolitan Areas and cities (Madrid, Barcelona and Valencia), and cross-border areas (Aquitanie Region and Iberia Region).*

- **Status, Structure, and Description of the Transport Network Infrastructure:** This includes an inventory of all relevant assets, such as roads, railway lines, ports, logistic terminals or hubs, and interchanges stations or hubs, along with information on their physical condition, functional classification (local, regional, national), and strategic role in the broader network. These elements define the fixed, physical components that enable the movement of passengers and goods. Evaluating the status of infrastructure helps identify bottlenecks, capacity constraints, and areas in need of intervention.

- **Network Representation in terms of Physical and Functional Characteristics:** As a system of interconnected elements, the transport networks are typically organized as nodes (intersections, terminals, stations) and links (road segments, railway tracks, maritime corridors). Each link in the network should include a set of key attributes such as capacity, travel speed, travel time, cost, and modal classification (road, rail, maritime, etc.), along with any relevant restrictions (e.g., weight limits, access bans). This network representation allows

for the simulation of transport flows, the identification of connectivity gaps, and the assessment of network efficiency and accessibility across different modes.

- **Structure of Transport Services:** The data on the structure of transport services defines the system's operational capacity and level of service. This includes information on the maximum throughput (how many vehicles, passengers, or goods can be accommodated per unit of time) and performance indicators such as congestion levels, average speeds, and service frequency. This information is crucial for evaluating current system effectiveness and forecasting future needs.

- **Operational characteristics:** This describes how the transport system functions in practice, including the organization and scheduling of services, traffic control mechanisms, and the integration of different modes. This involves data on service timetables, frequencies, typical traffic patterns, and control systems such as traffic lights, signalling, and tolls, how well different transport modes are connected (through transfer points, interchange times, and transfer penalties). These elements are essential for capturing the real-world behaviour of the system.

Related to the data, to build accurate supply models' various types of datasets from different data sources are requited:

- GIS data for spatial representation
- Traffic counts and sensor data for real-time flows and capacity utilization
- Infrastructure inventories
- Service schedules
- Historical performance data

Some specific cases are:

- **Automatic Vehicle Identification and Traffic Counts:** Technologies such as RFID, ANPR (Automatic Number Plate Recognition), and inductive loops gather data on vehicle flows and can be used to infer freight and passenger movement along key corridors.
- **Toll Road Data:** It provides accurate and granular information on vehicle movements across major transport corridors. By recording entry and exit points, timestamps, and often vehicle types, this data enables the identification of real travel patterns over time and space. Its reliability and broad coverage make it especially useful for modelling traffic flows, supporting dynamic OD estimation, and complementing or even replacing traditional survey-based methods, particularly in regions where other data sources are limited or outdated.
- **Passengers, Freight and Logistics Databases:** Public transport data, customs records, freight forwarders' data, and supply chain management systems provide information on people and goods movement, essential for mapping OD matrices for both cases.

The network is typically modelled as a multilayer graph, where nodes represent key transport hubs (e.g., ports, airports, rail terminals) in specific cities and edges represent transport corridors or links between each other.

Supply models play a critical role by providing a detailed understanding of the existing infrastructure, its capacity, and operational characteristics across multimodal networks. In the transport planning for the TEN-T Atlantic and Mediterranean Corridors, which are strategic gateways for freight and passenger mobility across Southern and Western Europe, supply models enable planners to identify infrastructure gaps, assess cross-border interoperability, and prioritize investments by providing insights into capacity constraints, bottlenecks, and the potential impacts of upgrades or expansions. This information helps to enhance connectivity, efficiency, and sustainability in the European transportation systems. By simulating network performance under different scenarios, these models guide policy decisions and investment strategies that align with the TEN-T's goals of a unified, resilient, and sustainable European transport system.

As **demand models** represent the core focus of this project, they are discussed in detail in the following sections of the document.

2.1. Demand models in the literature

Transport planning draws from a wide range of scientific disciplines, from Economics and Urban Planning to Engineering; but in almost all of its aspects, it requires the use of mathematical tools. However, there is an important difference between the type of mathematical tools used in planning and those employed in more conventional fields of Engineering: in Planning, the accuracy of predictions is inherently limited due to the high degree of uncertainty involved in the social variables considered (population growth, technological changes, etc.).

Transport systems are generally very complex and involve a vast number of variables. The mathematical analysis of such models relies heavily on criteria that help distinguish relevant variables from irrelevant ones, thereby reducing the problem to a more manageable and understandable size, according to the level of accuracy deemed necessary (or possible) to achieve.

Transport Planning procedures must be applied both to the construction of a simple road intended to serve existing traffic over the next one to five years (a time frame comparable to the design and construction period), as well as to the planning of the expansion of an urban transport system over a 20-year period, or to the design of transport services in newly built cities. The basic principles are similarly applicable to all of these cases, though in practice they may be implemented in very different ways. Small-scale problems with short-term planning horizons can be analyzed with fairly high accuracy; however, large-scale problems with long-term horizons require greater technical skill and practical experience with similar issues.

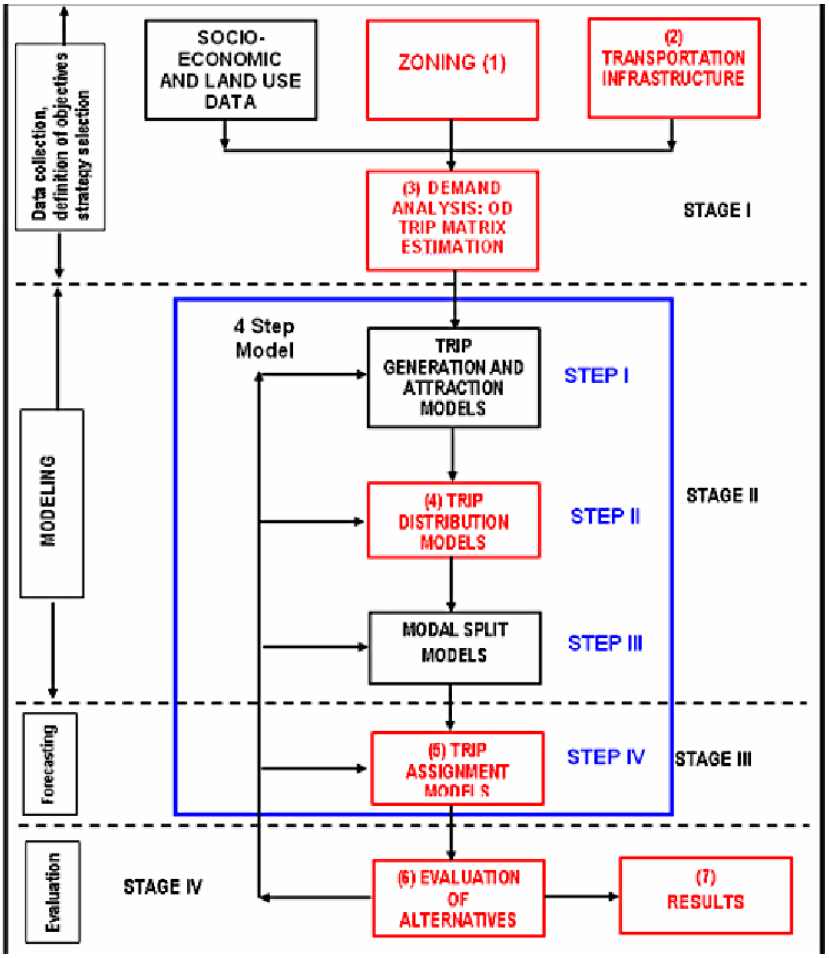


Figure 2. Four-Step Modelling Framework Schema

Over the past fifty years, a methodology for analyzing transport planning problems has been developed, based on classic transportation studies such as the Chicago Transportation Study, Penn-Jersey Study, Bay Area Transportation Study, and Northeast Corridor Study.

The key aspects of this methodology can be grouped into four phases, and it is essentially the result of the evolution of procedures initially proposed under the Federal Highways Act of 1962 in the United States. While the details of how some of the phases are carried out changed significantly in the 1970s, the basic philosophy has endured. In many cases, the trend has been toward more automated and less costly procedures—albeit at the expense of some accuracy.

The tasks included in each phase of the classical Transport Planning framework are described below (Ortúzar S. and Willumsen, 2011):

PHASE 1. Data Collection. This first stage refers to the collection of any type of data considered relevant to describing how the transportation system under study behaves and why. It includes an inventory of:

- The existing transport supply and its characteristics.
- Current mobility patterns within the study area, determined both through origin-destination surveys (including data on modal behavior) and through measurements of vehicle flow on the main roads of the transportation system.

- Planning variables such as land use, income distribution, the administrative structure of the study area, and the types of employment present.

This phase should also include the collection of historical data on the evolution of relevant variables, such as population growth, motorization rate, etc.

This phase has a dual purpose. On the one hand, it must be acknowledged that any future transport system will likely be just an extension or modification of the existing one, meaning that parts of the current transport system and mobility patterns will persist into the future. On the other hand, the second goal is to obtain numerical values for certain parameters that will appear in the mathematical models to be developed in later phases.

The aforementioned variables must generally be referenced within a common framework to facilitate their later use. That is, the geographic scope of the planning study must be divided into transport zones, in such a way that each zone possesses homogeneous characteristics with respect to the variables considered relevant in the study. This process is known as *zoning*.

PHASE 2. Modelling. The goal of this phase is to determine the relationships between the relevant variables collected in Phase 1. The initial data must be reduced to formulas and automatable procedures, some of which may remain valid for application within a future planning horizon. These relationships or models are typically classified into the following categories:

- **Trip generation models.** These models aim to relate the production and/or attraction of trips in an area to variables such as land use (industrial, residential, commercial, etc.), population density, income distribution, and employment types. They are usually multiple regression models.
- **Trip distribution models.** Given the trip productions and attractions of the various zones within the study area, the next step is to develop formulas that describe how trips from a given origin are distributed among the different destinations. That is, to reduce origin-destination mobility patterns into a few simple equations and procedures that relate zone productions and attractions with a small number of variables such as distance or the generalized cost of travel from origin to destination.
- **Mode choice models.** These are statistical models usually based on the random utility paradigm (RUM), in which the probabilities of selecting each available mode for different origin-destination pairs are determined based on service characteristics (access time, in-vehicle travel time, wait time, fares, etc.).
- **Traffic assignment models.** These are procedures based on behavioural models. Given a distribution of trips between origin-destination zone pairs (via a modal O-D matrix), the final step is to determine how these trips are distributed across the transport system along the different routes or paths available between each origin-destination pair (this also includes determining mode split). Ultimately, the aim is to relate the flow of travellers or vehicles on various routes or paths to both the characteristics of the transport supply and the O-D matrix.

This model typology is not universal. In some cases, mode choice models are treated independently from traffic assignment models. In other cases, trip distribution and assignment models have been combined into models that directly relate traveler or vehicle flows to trip generation (and to the defining characteristics of the transport system).

A more detailed description of each model is included in **Appendix 2**.

PHASE 3. Demand Forecasting. In this phase, predictions are made about future land use, population, and other factors based on the extrapolation of historical trends in those variables. This is the stage where the trip generation models developed in Phase 2 are used to estimate the future production and attraction of trips in the various zones, and the distribution models are applied to determine the future origin-destination (O-D) matrix for interzonal mobility.

PHASE 4. System Evaluation. The final phase essentially consists of assigning the future demand—estimated in Phase 3—to the various development alternatives for the transportation system.

The methodology described was developed in the context of private transportation; however, it can be directly adapted to the planning of public transport networks. In such cases, Phase 4 will include forecasting the consequences of restructuring and/or implementing new lines for existing or new transport modes, extending service for an existing mode, introducing demand-responsive public transport systems, creating special lanes for high-occupancy vehicles, and so on.

The main criticism of the described methodology stems largely from the limited availability of reliable historical data on the evolution of key variables, and from the specific ways in which the various phases have been implemented in practice in many studies.

2.2. Classification of transport simulation models

Transportation systems can be considered an archetypal example of modelling, depending on the objectives of the study, which are influenced by the geographic or spatial and/or temporal scope of the transportation system being modelled. Transportation system models can be initially classified as:

- Static models for long-term strategic planning, or macroscopic models.
- Dynamic models for operational planning or management, or mesoscopic models.
- Dynamic operational and management models, or microscopic models.

2.2.1. Macroscopic Models

Their objectives are mainly strategic in nature, and therefore long-term, and consequently static, since short-term temporal dynamics are not significant in this context. Their geographic scope encompasses regional or metropolitan transportation networks, and they consist of two fundamental components (Figure 3):

- A **supply model** that represents, with the required level of detail, the infrastructure of the transportation system, the available modes within it and their corresponding networks (road

network for vehicle circulation of different classes, public transport network and its sub-networks—bus, metro, commuter trains…; micromobility). The supply model is usually defined in terms of the graphs that represent the networks usable by each mode and, eventually, their interrelations (for example, in the case of interchange nodes where a traveller may change modes): $G^m = (V^m, A^m)$, where G^m is the graph of the network for mode m, and V^m and A^m, respectively, the sets of nodes and links of that network. The nodes, for instance in the road network, represent intersections, while the links, in the case of an urban network, represent the street segments between intersections.

- A **demand model** that represents the trips for each transportation mode between each origin and destination in the system. It is formalized in terms of mobility matrices in which each element of the matrix X^{pt}_{rs} represents the number of trips between an origin r and a destination s, for a travel purpose p (mandatory mobility, leisure, shopping, etc.) during a given time period t. In common practice, the origins and destinations correspond to Transport Analysis Zones, defined to be as homogeneous as possible in terms of the socioeconomic attributes that characterize them (Ortúzar S. and Willumsen, 2011). Formally, they are represented in the graph that models the system using fictitious nodes, called centroids, and auxiliary fictitious links that connect the centroids to the links that represent the road network, in order to define the entry and exit points of traffic flows into the road network. Since time is not explicitly used as a variable of analysis, the time periods considered are usually a full 24-hour typical day, the morning or afternoon peak hour, etc.

In macroscopic modelling, it is important to distinguish between two common practices that differ significantly in terms of modelling complexity and their respective applicability. The main difference lies in **how the demand model is constructed**—that is, how OD matrices by mode and trip purpose are generated.

One approach, simpler and less costly but with limited applicability, consists of **building the matrices from observed data**, such as mobility surveys, traffic counts, public transport ticketing data, OD matrices derived from mobile phone data, etc. One or several of these techniques may be used. The resulting matrices faithfully reproduce reality (depending on the quality and quantity of the observed data) and allow for certain operational analyses of limited complexity. In these cases, the trip generation and distribution phases (and sometimes even modal split) are not applied. The observed matrices are built directly and then passed to the assignment phase.

The **other approach**, more complex but more versatile, involves the full programming and formulation of the **four steps of the classical model**. That is, based on mathematical formulations and socioeconomic data (excluding observed mobility data), the models for trip generation, distribution, and modal split are developed. These mathematical processes then construct the mobility matrices by mode and trip purpose. These matrices will generally show a slightly higher margin of error compared to the previous method (based on observation); however, they offer the analytical flexibility needed to carry out future scenario analysis, allowing the estimation of impacts from:

- Socioeconomic developments: housing, employment, services, etc.
- Infrastructure: road or rail, whether new or modified.
- Modifications or additions to public transport services.

- Changes in mobility policies such as tolls, parking fees, etc.

Moreover, for all the aspects mentioned above, the complete four-step model allows the estimation of two highly relevant issues:

- New mobility induced by infrastructure.
- Demand shifts between different transport modes.

In general, most professional software platforms for modelling and analysing transport systems include tools for implementing the four-step model. However, in these notes we will limit ourselves to its traffic assignment model component.

Once the supply and demand models are available, the assumptions about **how they relate to each other** are formulated, for example, in terms of volume-delay functions that model the costs of using an arc of the road network for vehicular flows. These costs are expressed as the time required to traverse the arc depending on the traffic volume using it, as an implicit way of modelling traffic congestion. Similarly, costs are associated with the nodes, or intersections, based on penalties incurred when making a turn, for instance.

The **relationship between the demand and supply models**, in terms of road network usage conditions defined by the volume-delay functions, is established through a **traffic assignment procedure**. This procedure assigns the demand to the supply under those usage conditions, assuming a modelling hypothesis. This hypothesis concerns the criteria used by travelers to choose the routes for their trips between origins and destinations along the paths connecting them in the road network.

These models will be addressed in detail in the corresponding section. To conclude this summary: the most widely used modelling hypothesis assumes that travelers seek to minimize their individual travel times, choosing the routes they perceive as the shortest under the prevailing conditions of the road network. This hypothesis is formulated in terms of what is known as Wardrop's First Principle "*Travel times on all used routes are equal and less than those that would be experienced by a single vehicle on any unused route.*" (Wardrop, 1952)

Professional software platforms used in transportation planning are generally based on Wardrop's principle. The differences and variations lie in the volume-delay functions and their parameters, which must be calibrated. These functions represent the cost of using an arc in the road network, in terms of travel time, as in the example of equation (1), or as generalized cost functions that may include additional elements such as urban tolls.

As an example, without delving into the specific details of each professional software platform for transport systems, we mention the typical example of the BPR volume-delay functions (from the Bureau of Public Roads, part of the U.S. Department of Commerce, Urban Planning Division), which were proposed in 1964 in its "*Traffic Assignment Manual*" (Roads, 1964) and are still widely used today. These functions take the form:

$$t_a(x_a) = t_a^0 \left[1 + \alpha_a \left(\frac{x_a}{c_a}\right)^{\beta_a}\right] \quad (1)$$

Where t_a is the travel time to traverse arc aaa when the traffic volume on it is x_a. The function depends

a

on the parameters t^0, the free-flow travel time on arc, the effective capacity of arc a, c_a, which are physical attributes of the arc defined during the supply model construction phase, and also on the calibration parameters α_a and β_a, which determine how the traffic flow behaves on the arc. In practice, the arcs of a road network are grouped into hierarchical categories, and it is assumed that all arcs belonging to the same category follow the same type of function. Although all professional tools provide default values, the calibration process—essential for later validation—requires checking that those parameter values are accurate.

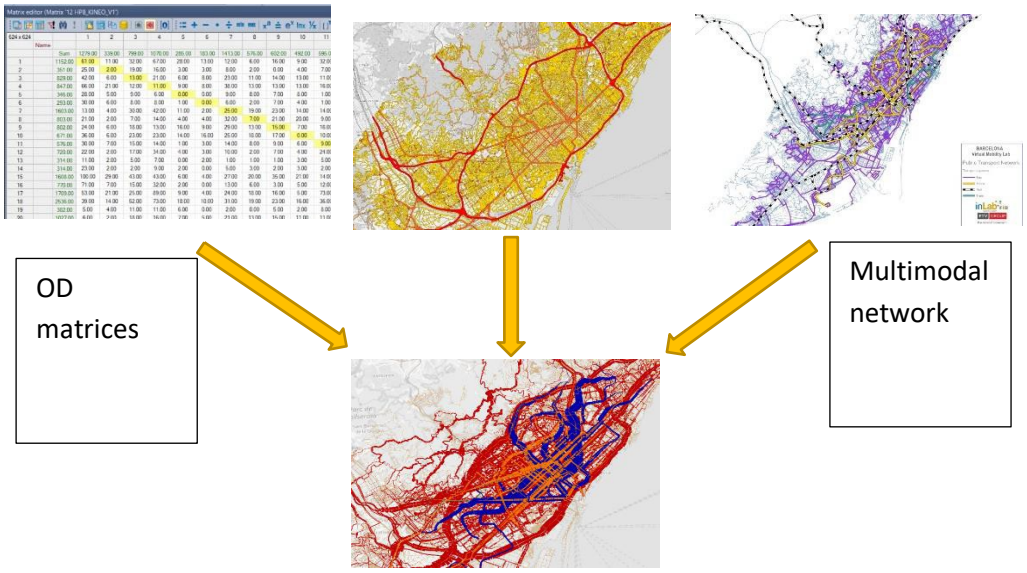


Figure 3. Macroscopic models: components and relations

2.2.2. Mesoscopic Models

These models emerged from the need to extend the capabilities of planning models to incorporate the temporal dimension into transport planning analysis, while still allowing the application of modeling-based analytical techniques to large-scale networks (such as those of metropolitan areas). They make it possible to analyze the dynamic diversions of traffic resulting from the generation and propagation of congestion, while maintaining affordable computational performance. Thanks to advancements in computing technologies, both in hardware and software, these models can be used to support traffic management systems. Like macroscopic models, mesoscopic models have two fundamental components:

- **The supply model**. Similar to that in macroscopic models in terms of the topological structure of the road network, it requires additional, more detailed information. For example, this includes data on traffic control plans for nodes corresponding to signalized intersections. Because dynamic aspects are taken into account, detailed control plans are required, including signal phases, splits, and offsets. Additional attributes are also necessary to describe the infrastructure, as the dynamics of traffic flows along links can be modelled in various ways, depending on the model's approach—for example, using fundamental diagrams of traffic for each type of link or using ad hoc parameters, whose values must be properly calibrated in each case.

- **The demand model**. As in macroscopic models, this is represented using an OD travel matrix. The key difference here is that, to capture the temporal dynamics of traffic phenomena, the OD matrix must be dynamic and capable of reflecting temporal demand variations. That is, in the case of private vehicle transport, it must take the form $X_{rs}(t)$, representing the number of trips from origin r to destination s, during time period t.

The relationship between demand and supply, which determines how traffic flows dynamically propagate through the road network, is defined through an assignment model—in this case, a **dynamic assignment model**. However, not every dynamic traffic assignment (DTA) model corresponds to a dynamic user equilibrium (DUE) model. The Dynamic Traffic Assignment (DTA) problem can be considered an extension of the static assignment problem that lies at the core of the macroscopic models previously described. The fundamental difference is that the dynamic version must be capable of determining how traffic flows on links and along paths evolve over time, as a consequence of a time-dependent demand described by an OD matrix that reflects this temporal variation. As precisely described in (Mahmassani, 2001) "*dynamic traffic assignment formulations must be able to describe how traffic flow patterns evolve through the road network in time and space, and consequently provide estimates of travel times on links and paths, and their expected short-term evolution.*". The monograph by (Chiu *et al.*, 2011) prepared for the Transportation Research Board, provides a clear and practical overview of what this formulation entails and its practical implications.

2.2.3. Microscopic Models

Microscopic models represent the most detailed way to describe traffic flow dynamics by modelling the individual behaviour of vehicles, considered as particles in a fluid. The aggregation of individual dynamics reproduces the overall behaviour of the traffic flow as if it were a fluid.

These models are suitable for analysing, in great detail, the dynamic impacts of operational strategies such as ramp metering, linear speed control on highways, traffic signal coordination, real-time traffic control, or dynamic management strategies typical of Intelligent Transportation Systems (ITS). Accordingly, they can be applied in various scenarios depending on the study's objectives. Two primary types include:

- Scenarios defined by sets of intersections of different geometries, possibly including roundabouts, such as those corresponding to a section of an urban corridor, where traffic control, including signal coordination or adaptive control systems, are among the study's objectives; and highway segments with entrance ramps (which may be the subject of ramp metering studies), exit ramps, toll stations, etc.
- Scenarios involving urban networks with a mesh structure where, in addition to issues related to traffic control systems, alternative routes that vehicles may take between various origin-destination pairs in the urban network must be considered.

As in the previous cases, they consist of:

- **The supply model**, which in this case requires a detailed description of the geometry of the road network, since its characteristics influence the driving dynamics of individual vehicles. It also

requires detailed information on the control plans of signalized intersections.

- **The demand model.**
  - Defined, as is usual in traffic engineering, in terms of entry flows at each of the model's access points and the turning percentages at each of the exit or transfer points.
  - Defined by OD matrices which, as a dynamic model, will take the form $X_{rs}(t)$, representing the number of trips from origin r to destination s, during period t.

The simulation of flows generated by demand in the network, based on the usage conditions, is not carried out in this case by an assignment procedure as in macro and meso models, but rather through **behavioral models** that attempt to **emulate driver behavior**. These are essentially:

- The car-following model, which simulates the behavior of a pair of vehicles: the leader (in front) and the follower (behind).
- The lane-changing model.
- The gap-acceptance model.
- The route choice model from origin to destination for each individual vehicle.

As in other cases, professional software platforms for traffic system modeling and analysis using microscopic simulation provide default values for all parameters. However, this does not exempt users from verifying whether these are appropriate for the specific project under study, or whether the appropriate parameter values need to be estimated, which will be one of the objectives of the calibration process for that microscopic model.

2.3. Data sources and fusion in transport modelling

Transport systems **require accurate and precise data** for monitoring and managing routine operations that maximize the safety and efficiency of the transport network. The proliferation of connected vehicles capable of providing data on speeds, congestion, weather, and road conditions presents both an opportunity and a challenge for transport authorities and consultants to support Intelligent Transport Systems (ITS). The opportunity arises from the **diversity of available data**, while the challenges stem from the need to collect, store, process, and analyze the large volumes of data generated. These sources increase travelers' need for reliable traffic information.

The **challenge** lies in developing **data fusion methods** capable of handling these massive amounts of data (Kashinath *et al.*, 2021). Sensors located at specific points in the transport network are usually not effective at providing the data required by modern transport management systems. For example, OD (origin-destination) matrices cannot be obtained from such data. GPS data from fleet traces and mobile phone tracking complement these conventional data sources.

If data from different sources are complementary and independent, **their fusion can benefit transport** network managers. **Integrating multiple data sources** is necessary to develop the different analysis and increases the robustness, accuracy, and granularity of the results. Heterogenous data sources enhance the consistency, generalizability, and overall reliability of transport models. However, achieving such integration presents challenges related to data harmonization, privacy concerns, and representativeness, all of which must be addressed before the data can be used in final applications.

Efforts to improve data **interoperability, standardization, and privacy safeguards** are ongoing, both within individual countries and across borders. The advances in big data analytics, machine learning, and data fusion techniques are facilitating this work, enabling the generation of interoperable and consistency data spaces that provide more comprehensive and actionable mobility information.

ICT (Information and Communication Technologies) sources extend spatial and temporal coverage, such as weather data derived from automatic windshield wiper activation systems. The fusion of data from multiple sources offers a solution to improve transport system data. The possibilities for applying data fusion techniques are not limited to ICT data; they also apply to the analysis of traditional traffic flow data by time of day, day of the week, month or season, or recurring events. **Historical (offline) data combined with real-time (online) data** can be used to predict trends or changes in mobility patterns. Likewise, incident detection can benefit from complementing induction sensor data with new sources from connected vehicles and mobile phones. Travel time estimation can also benefit from data fusion processes.

An important element of traffic planning and management is the origin-destination matrices (OD matrices). **Estimating OD matrices** from traffic counts has traditionally been used to update an outdated matrix obtained from household surveys. New data sources are currently being used to obtain OD matrices with greater spatial and temporal detail (at an hourly level).

Data fusion relies on **fusion algorithms** that provide innovative solutions in the form of fusion applications for Intelligent Transport Systems (ITS). The techniques used so far include Bayesian inference, Dempster-Shafer evidence theory (Faouzi, Klein and Mouzon, 2009), Kalman filters, and neural networks. The first two techniques are parametric types. The architecture of data fusion systems is of great importance, but this guide will focus more on fusion methods than on architectures, for which systems engineers must ensure their capacity to achieve the desired objectives.

Access to the vast amounts of data generated by ICT applications, especially regarding the mobility of people and goods, has led some to believe that data and its proper processing alone are sufficient for decision-making.

The Knowledge Hierarchy, also known as the DIKW Hierarchy (Figure 4) is a set of models representing the seemingly structural relationships between Data, Information, Knowledge, and in some cases Wisdom. The sequence of the hierarchy, from the most basic to the most complex, is:

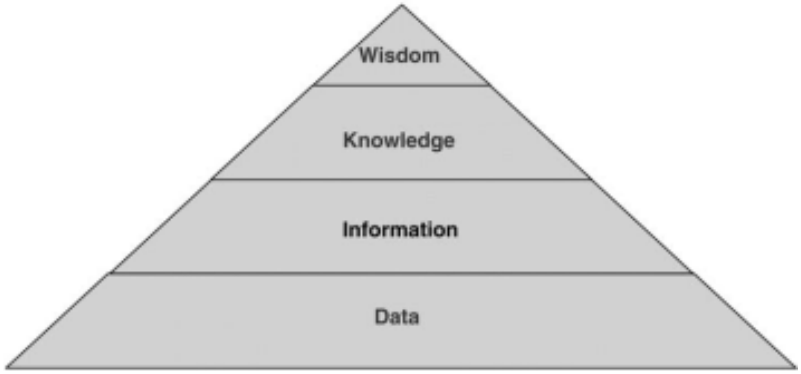


Figure 4. DIKW Hierarchy. Source(Rowley, 2007)

Ou (Ou, Lint and Hoogendoorn, 2010; Ou, 2011) indicates that the availability of traffic data from multiple sources does not increase the consistency, coherence, and explainability of traffic conditions; this data must be **transferred in a consistent and complete manner** to achieve an accurate estimation of traffic state, and this is where data fusion methods must be applied.

In Ou's work, the focus is on estimating traffic state through speeds, flows, and densities, which is of interest for the state of private vehicles on highways. The study concentrates on the highway network of the Netherlands. An interesting observation is that coverage is not uniform, and missing or poor-quality data are detected. The data from these different sensors (cameras, induction loops, or in-vehicle GPS/GSM devices) typically have **different formats, semantics, resolution, and temporal and spatial accuracy**, and they also differ in availability and reliability depending on location, time, and circumstances.

From a methodological point of view, **integrating such heterogeneous data into complete and consistent data warehouses is a complex and challenging task**. As will be seen later, particularly the semantic differences regarding space and time among these data impose strong constraints on the applicability of data fusion techniques.

2.3.1. Types of Data Sources

The structure of this document proceeds by describing each of the cited sources. Subsequently, a general section on mechanisms for data processing, filtering, and labelling is included. Following that, guidelines for processing each data source are provided. The next section will address data fusion, but it is appropriate here to refer to potentially plausible scenarios that combine conventional technologies (magnetic induction loops, magnetometers, radar, etc.) with new ICT applications (mobile phones, GPS device tracking, etc.). The primary characteristic of these scenarios is the coexistence of multiple traffic data collection technologies representing heterogeneous data sources, differing in the type of data collected, how they are collected, temporal collection patterns, units of measurement, and so forth.

To estimate traffic variables, traffic measurements are required. These **measurements can be classified into Eulerian and Lagrangian measures**. Eulerian measurements can further be classified

into local, instantaneous, and spatio-temporal measurements. Lagrangian measurements are usually represented by the vehicle's trajectory.

Local measurements are those provided by sensors measuring traffic within a spatio-temporal region whose spatial interval becomes extremely small. Here, it is important to differentiate between average temporal speed and spatial speed.

Instantaneous measurements (*Table 1*) measure traffic conditions within a spatio-temporal region whose time intervals become extremely small. They provide traffic conditions on a road section of a certain length *X* at an instantaneous time *t*. These can be thought of as a snapshot captured by a camera at time t, "freezing" the traffic state at that moment. Instantaneous density is defined as the number of vehicles divided by the length of the analyzed section.

Lagrangian measurements are those obtained by tracking an individual fluid parcel (for example, an individual vehicle) moving through space and time. These measurements can trace all positions of a single parcel over time. Vehicle trajectories are typically Lagrangian measurements. This type of measurement can be obtained through GPS technology or any tracking device. Unlike Eulerian measurements, Lagrangian measurements reflect how a vehicle experiences traffic. However, trajectories cannot be obtained for all vehicles, and sometimes only travel times are directly available. The recurring question is: how do these measurements contribute to the estimation of spatio-temporal variables?

Table 1. Classification of Measurement Data According to (Ou, 2011)

Sensor Measurement Types	Subtype	Examples
Eulerian Measurements	• Spatio-temporal measurements • Local measurements • Instantaneous measurements	• Radar measurements • Induction detector measurements • Traffic images
Lagrangian Measurements	• Complete • Incomplete	• Vehicles trajectories • Travel times obtained by cameras

Another way to categorize data is into those describing vehicle movements versus those describing individual movements.

The predominant sensors for obtaining traffic flow data are **magnetic induction loops**, which measure traffic characteristics at a specific location. These are intrusive sensors since they are installed in the pavement. Other point sensors include magnetometers, infrared sensors, and ultrasonic sensors. None of these can capture the spatial behavior of traffic flow. Installation and maintenance costs are not negligible, especially for induction loops.

RSU sensors (roadside units) mounted on poles or signs can provide multiple types of data, improve spatial coverage across one or more lanes, and incorporate various short-range detection technologies, such as those associated with V2V (vehicle-to-vehicle), V2I (vehicle-to-infrastructure),

and V2P (vehicle-to-pedestrian) in connected and cooperative vehicles. They also include internal vehicle sensors controlling subsystems like braking, speed, acceleration, lane departure, windshield wiper activation, etc. Vehicles with mobile sensors can transmit position via GPS signals or congestion/incident messages initiated by the travelers themselves. Some mobile sources are leveraged by technology companies such as INRIX, HERE, TomTom, or Waze.

Data from severe braking or activation of **incident warning signals** help identify congestion areas and incidents. Loss of traction or windshield wiper activation can trigger weather or road surface condition warnings on variable message signs.

Improvements and expansion of wireless networks and mobile device locations enable the development of location services that use mobile device users in vehicles as **traffic probes**. Automatic vehicle identification (AVI) systems, including toll tags and Bluetooth devices detected by Bluetooth antennas, allow for device re-identification via MAC addresses for estimating travel times, spatial speeds, and origin-destination analysis (OD pairs).

Other traditional sources still play a role today, such as **radio transmissions** among commercial fleets or highway maintenance teams and emergency phones on highways.

Each data source has a specific accuracy, detection area, and refresh rate. **No single data source can provide a complete description of the detailed and spatial characteristics of traffic by itself**. In principle, every data point should be accompanied by its timestamp and location (in a coordinate system that can be translated into conventional systems; this is part of the data alignment process).

2.3.2. Sensors and Available Variables

Sensors have been classified in various ways, one of which distinguishes between active and passive sensors. Active sensors transmit and receive energy; the received energy is a fraction of the transmitted energy that returns to the sensor. Passive sensors only receive energy. The received energy is a combination of energy emitted by motor vehicles, bicycles, pedestrians, and other objects of interest.

Table 2 shows the most common types of sensors and the variables available, along with a description of their most notable characteristics.

Table 2. Traditional Sensor Types and Obtainable Variables

Sensor / Method	Description	Variables
Induction Loops (active)	Detects vehicles crossing the detector by the variation in inductance of the loop caused by the metal of the vehicle: provides presence, counting, and occupancy. If double loops are used, they can estimate spot speeds. Some can classify vehicles based on the number of axles. Very reliable under varied weather conditions; however, they require constant maintenance for calibration, which necessitates lane closures.	Vehicle counting, presence, and occupancy. Also spot speed, classification into light and heavy vehicles.

Sensor / Method	Description	Variables
	They cannot be used on bridges and viaducts. They reduce pavement lifespan. They are affected by heavy vehicles. Generally, they operate at the lane level.	Eulerian measurements.
Magnetometers (passive)	They detect disturbances in the Earth's magnetic field caused by the metal of vehicles. They detect presence and movement. Less susceptible to damage from heavy traffic. Intrusive, with lower maintenance and installation requirements. They reduce pavement lifespan. Their detection area is limited, and several are usually installed to cover multi-lane roads.	Vehicle counting, presence, and occupancy by lane. Eulerian measurements.
Magnetic Detectors (passive)	They detect disturbances in the Earth's magnetic field caused by the metal of vehicles. They detect movement and the number of passengers. They require lower installation and maintenance costs than loops and can be installed on bridges. Limited detection range.	Counts and number of passengers by lane. Does not detect stopped vehicles. Eulerian measurements.
Camera (passive)	Detects reflected light and provides high-resolution images. Can capture vehicles in different lanes, can classify them basically (light, heavy), presence, counting, occupancy, and spot speed. Cameras require image processing through vision algorithms. They have issues with optimal placement and visibility due to weather conditions. They show errors due to reflections on wet pavement or shadows and are costly. Require good lighting for nighttime operation. Once installed, they do not require much maintenance.	Flow, occupancy, speed, density, and queue length. Limited vehicle classification.
Presence radar (active)	Detects the fraction of transmitted energy returned to the radar antenna. Provides counts, speed, occupancy, and detects the direction of movement across multiple lanes. Limited vehicle classification (based on length). Not affected by nighttime or bad weather conditions. Requires redundancies to capture 100% of vehicles at a stop (occlusion problem).	Vehicle counts, speed, occupancy, direction. Classification by length. Eulerian measurements
Doppler sensor (active)	Detects the fraction of transmitted energy returned to the sensor antenna. Not affected by nighttime or bad weather conditions.	Vehicle counts, spot speed, presence, occupancy. Eulerian measurements
Infrared video (pasivo)	Detects heat emission. Fine spatial and spectral resolution. Does not require external lighting. Similar data collection potential as visible spectrum devices, but less affected by weather conditions. Operates well at night. They are costly.	Vehicle counts, speed, occupancy, and presence.

Sensor / Method	Description	Variables
		Eulerian measurements
Lidar (laser radar) (active)	Detects the fraction of transmitted energy returned to the sensor's photodetector. Detailed images and good spectral resolution. Operates well at night.	Speed, vehicle count, occupancy, multilane presence in small areas
Infrared (passive)	Detects infrared radiation. Can measure counts, presence, spot speeds, vehicle length/volume, and lane occupancy. Affected by heavy rain. Traffic interruption required for installation and maintenance.	Speed, vehicle count, occupancy, presence by lane Eulerian measurements
Ultrasonic (active)	Detects the fraction of transmitted energy returned to the sensor by vehicles or the pavement. Detects vehicle presence, counts, and occupancy. Not costly to install (on a pole). Reliability affected by temperature variations and air turbulence. High occupancy requires sensor tuning. One sensor per lane.	Presence, vehicle count, occupancy by lane Eulerian measurements
Acoustic (passive)	Detects vehicle and environmental noise. Multilane. Operates well at night. Easy to install on a pole. Loses reliability in congestion and rain.	Presence, vehicle count, speed, occupancy Eulerian measurements

2.3.3. Surveys

Surveys are traditional sources of transport and mobility data. The most common surveys in transport models are: household surveys, intercept surveys and stated-preferences surveys.

Household surveys provide a comprehensive picture of the trips made by residents in a study area, including walking and cycling, over a period of about weeks or a few months. They are static snapshots that constitute an average image of the mobility profile of residents in a study area. The results of these surveys can be segmented by key variables such as household type, person type, trip purpose, mode, and time period. However, building travel matrices directly from household surveys is generally not practical due to cost reasons: the method is inherently expensive per recorded person-trip.

The main application of this type of dataset is therefore in segmenting demand data collected from other sources and for use in building car ownership models and calibrating demand models: travel willingness, trip production by purpose, trip distribution by purpose (accuracy in defining origins, destinations, and intermediate trip points directly affects the calibration of trip distribution models). Household surveys can be employed for activity analysis in the latest generation demand models.

Intercept surveys are used to quantify trips made by non-residents into the study area and therefore complement the data collected through household travel surveys.

There is no probabilistic sampling design supporting these surveys, as the goal is to obtain the largest convenient sample possible through face-to-face interaction. Therefore, there is no predetermined number of surveys; it will be sufficient to ensure the adequacy of the data collected in the two surveys that make up the Intercept Survey.

According to (Ortúzar S. and Willumsen, 2011), until the mid-1980s, it was almost axiomatic that transport demand modelling was based on information about observed choices and decisions, that is, on **revealed preference (RP) data**. Within this approach, project evaluation requires that the intervention policy be expressed in terms of changes in attributes considered influential on current behaviors. However, it has practical limitations basically associated with sample costs and the difficulty of distinguishing effects of attributes that are not easy to observe, for example, those related to concepts such as quality or convenience (latent variables).

Another traditional practical problem has been how to handle "new alternatives," where it is necessary to predict the behavior regarding a facility that is not currently available and may be radically different from existing ones.

Stated preference (SP) techniques, originating from the field of market research, emerged in the late 1970s as a tool to conduct experiments with choices related to transport systems, and thus to solve some of the problems described above.

Stated preference techniques base demand estimation on an analysis of responses to hypothetical choices; these can naturally include a wider range of attributes and conditions than those of the real system. However, these techniques were discredited at their inception because responses resulted from excessive enthusiasm toward hypothetical actions without proper contextualization of the intervention, though these issues have now been largely overcome.

Data from SP (Stated Preference) surveys are valuable for determining the subjective evaluation of the attributes that contribute to the discrete choice considered.

2.3.4. ICT Data in Vehicle and Passenger Mobility

The topics covered are not intended to be exhaustive; this section is subject to change in the short and medium term:

- Bluetooth/Wi-Fi
- Mobile telephony
- GPS devices
- Data from autonomous and connected vehicle equipment.
- Social media (GPS location of the person, enriched with contextual information from the message shared on social platforms)
- Bank card transactions (GPS location of the point of sale terminal where the transaction occurs, enriched with contextual consumption information)

- Data from shared mobility and MaaS applications (GPS location of vehicles + booking records, origin and destination of users' trips) – this can be a specific case or adapted from public transport data
- Especially in public transport, where the mass collection of data from electronic fare media (contactless cards) enables a rethinking of demand estimation models, and access to real-time vehicle tracking data via Operational Support Systems opens new perspectives on how the system operates.
- Trip matrices: Mobile telephony, Location-Based Services, specific applications used during research campaigns (apps on smartphones)
- Travel times: (Google, TomTom, Inrix…)

Sensors and available variables:

Table 3. Types of ICT Sensors and Obtainable Variables

Sensor/Method	Description	Variables
GNSS	Satellite Navigation Systems that provide geospatial positioning with global coverage, either autonomously or with augmentation systems.	Coordinates Lagrangian measurements
GPS traces	Vehicle location via satellites. Available for fleets and in navigation systems. Vehicle tracking reports (i.e., GPS traces in commonly used terminology) have in recent years become a promising data source for estimating traffic conditions in cities (Li *et al.*, 2017). Companies such as INRIX, HERE, TOM-TOM, among others, have incorporated the exploitation of GPS trace data from data providers into their business model.	Coordinates, counts, instantaneous speeds, direction(heading) Lagrangian measurements
AVI Bluetooth	Vehicle identification using Bluetooth antennas. BMS (Blue Media Scanner) antennas scan the Media Access Control identifier (MAC-ID) of detectable Bluetooth (BT) devices within their communication range. As more modern portable electronic devices—such as mobile phones, in-car navigation systems, headsets, etc.—are equipped with BT, usage is increasing with the improved availability of these devices (Chung, 2001; Bhaskar and Chung, 2013)	Travel times between two Bluetooth antenna positions Lagrangian measurements
Smartphones	Geographical location based on mobile network cell towers. Mobile telephony data	Counts Lagrangian measurements

Sensor/Method	Description	Variables
	are those recorded by mobile network operators (MNOs) as a result of the interaction between mobile devices and the operator's antenna network. The primary purpose of this source is the management of the mobile telephone network.	
Autonomous and Connected Vehicles	Autonomous and connected vehicles represent a challenge, both for their widespread deployment in the market and for the implications related to their capacity to be a massive source of data production and consumption. The use of connected vehicle data (Otonomo, 2022) considered in a very optimistic light, covers areas related to Artificial Intelligence (AI) and Machine Learning (ML). To date, most players have overlooked opportunities to monetize data from these vehicles, a significant oversight considering how companies in other industries are aggressively generating value from data (McKinsey, 2021).	Coordinates, counts, instantaneous speeds, direction (heading) Lagrangian measurements
Shared Mobility Applications	Operational data from shared mobility services are those collected by the operators of these services as a result of monitoring the vehicle fleet and its use by customers. Since these are digital services, where reservation and vehicle access are done through mobile applications, all customer interactions with the service are recorded. The primary purpose of this data source is therefore the management of the mobility service itself.	Lagrangian measurements
Intelligent Public Transport Payment Systems	Data from intelligent public transport payment systems are those recorded by public transport operators as a result of the validation of an electronic transport ticket by service users. The primary purpose of this source is the collection of service fares by the operator through intelligent payment systems.	Lagrangian measurements

2.3.5. Data sources in Spain

Particularly in Spain, some available data sources related with the **supply** in the transport network are:

- **Mobility Databases:** The **Spanish Statistical Office (INE - Instituto Nacional de Estadística)** (*INE. Instituto Nacional de Estadística*, 2025), publishes detailed freight transport statistics, including goods movement by road, rail, maritime, and air transport. Other data sources are the **Observatory of Transport and Logistics (OTLE)** (Ministerio de Transportes y Movilidad Sostenible, 2025b) and the **Metropolitan Mobility Observatory (MMO)** (TRANSyT UPM, 2025) which annually publish a huge number of indicators related to the freight and passenger mobility in the country.
- **Traffic Sensors:** The **Spanish Ministry of Transport and Sustainable Mobility** operates a network of traffic counting stations throughout the national road network, providing continuous data on vehicle flows, types, and speeds (*Portada | Ministerio de Transportes y Movilidad Sostenible*, 2025).
- **European and Regional Data Portals:** Spain participates in EU-wide data initiatives such as TENtec (European Commission, 2013; European Environment Agency, 2025), which provide harmonized infrastructure and traffic data, useful for cross-border and multimodal travels analysis.

Related with the **demand:**

- **Household Mobility Surveys:** In some cities is conducted a survey that collects comprehensive data on daily travel patterns of the citizens, including trip purposes, modes, distances, and times. These surveys are not very frequently and just developed in main cities. This information is recompiled by the MMO.
- **Mobile Phone Data and Big Data Sources:** The **Ministry of Transport and Sustainable Mobility** has developed a mobility repository based on mobile phone data. The goal of this repository is to provide a robust, detailed, reliable, and up-to-date source of data that can be used to transport planning, policy development, and crisis response by delivering accurate metrics on trip frequencies, distances, and movement trends between Spanish provinces and some regions in France and Portugal. The Ministry has enabled 3 main modules for consulting specific analysis that are daily updated and carried out automatically.
    - **Basic Studies,** for daily mobility analysis of the Spanish residents across the country.
    - **Complete Studies,** for a detailed mobility analysis with more indicators, and statistical values for specific days.
    - **Routes in highways,** for analysis routes followed in road travel between each origin-destination.

  Additionally, there is available the option to download the data in different formats for carrying out customized studies.

  All mobility data collected is made publicly available via the HERMES Platforma (Ministerio de Transportes y Movilidad Sostenible, 2025), an open data portal which different resources available for mobility analysis.

- **Rail and Public Transport Operators Data:** Transport authorities, consortiums, urban and interurban bus companies, RENFE (Spain's national railway company), regional and local transport authorities provide ridership and ticketing data, enabling detailed analysis of passenger flows across rail and urban public transport networks.

Spain is increasingly integrating ITS into its traffic management framework and working on Projects like **DATAHUB**, and **DGT 3.0** aim to create a real-time interoperability data space and transport hub application. Recently, he Ministry of Transport and Sustainable Mobility has published ***pySpainMobility***, a Python library aimed to facilitate the access and management of Spain mobility data (Beneduce *et al.*, 2025; Ministerio del Interior del Gobierno de España, 2025; *Portada | Ministerio de Transportes y Movilidad Sostenible*, 2025).

2.4. Multimodal interaction and TRANS-TOOLS model in TEN-T Corridors

Modelling the TEN-T network involves capturing its complex structure, traffic flows, and interactions across the different transport modes considered in the network. Consequently, the multimodal interactions are crucial in the operation, management and control.

Multimodal nodes in the TENT-T corridors act as pivotal modal exchange points (ports handle maritime traffic, rail-road terminals bridge long-haul trains with trucks, and airports facilitate passenger interchanges) forming a backbone for seamless integrated operations across the complete network in Europe in this case. This enables efficient and sustainable multimodal logistics and passengers' mobility (Trasporti e Territorio, 2022).

The transport networks in the Atlantic and Mediterranean Corridors involved in these nodes are (European Commission, 2024):

- **Rail:** High-speed passenger lines (Madrid-Lisbon, Madrid-Valencia, Barcelona-Valencia, Basque Y, Madrid–Barcelona, Barcelona–Perpignan) and Rail Freight Corridors (RFC) for the Atlantic the link is Lisbon – Madrid – Bilbao – France, and for Mediterranean Madrid – Valencia – Barcelona – France.
- **Road:** Highway axes are heavily used by both freight transport and coach services. They play a key role in passenger mobility while also connecting ports with inland logistics hubs. Running parallel to rail infrastructure, roads concentrate high passenger mobility volume, provide essential flexibility, and last-mile accessibility. The most important highways in these corridors are A-2 – Madrid–Barcelona, A-3 – Madrid–Valencia, A-1 / A-8 – Madrid–Bilbao (through Burgos/Irún), A-5 / A-6 / A-62 – Madrid–Lisboa (through Extremadura).
- **Maritime Inland waterways:** Spain and Portugal rely heavily on maritime transport for international freight flows, with major ports such as Valencia, Barcelona, Bilbao, and Lisbon that serve as key logistics hubs. These ports are linked to inland areas via both rail and road corridors, ensuring efficient multimodal connectivity. Related to inland waterways, the

relevance is very low integrated just for rivers like the Douro and Tagus in Portugal with a limited use.

Table 4 summarises the key multimodal nodes for passenger mobility and logistic distributions by the main modes of transport for the Atlantic and Mediterranean Corridors which are relevant for the project.

Table 4. Key multimodal urban nodes in the Atlantic and Mediterranean Corridors by Mode of transport.

Mode of Transport	Atlantic Corridor	Mediterranean Corridor
Railroad terminals	Madrid, Vitoria–Júndiz, Bilbao (ES) Lisbon (PT)	Barcelona, Madrid, Valencia (ES)
Airports	Madrid–Barajas, Bilbao (ES) Lisbon–Humberto Delgado, (PT)	Barcelona–El Prat, Madrid, Valencia (ES)
Public Transport in Urban Nodes	Madrid, Bilbao, San Sebastian (ES) Lisbon, Coimbra (PT)	Barcelona, Madrid, Valencia (ES)
Seaports	Port of Lisbon / Setúbal (PT)	Port of Valencia, Port of Barcelona, (ES)
Inland Waterways	River Douro (PT/ES) River Tagus (PT)	

The freight stations in Madrid (Dry Port of Madrid – Coslada and Madrid – Abroñigal) integrate with ports via standard-gauge freight corridors, reinforcing modal shift to rail (Ministerio de Transportes y Movilidad Sostenible, 2025b). This highlights the crucial role of this urban node in linked both corridors and the different modes of transport involved.

Additionally, It is relevant to mention that through the POLIS Urban Nodes Taskforce, as a result of efforts developed since 2025 and concatenated different milestones along the period, it is expected that , each urban node in the TEN-T Network by 2030 are requited the development of multimodal passenger hubs and by 2040 have access to at least one multimodal freight terminal (POLIS Network, 2025).

2.4.1. TRANS-TOOL: Specific Transport Model for TEN-T

TRANS-TOOL (Tools for Transport Forecasting and Scenario Testing) Project was a key initiative funded by the European Union to develop a pan-European transport model to serve as a support tool in the planning, analysis and evaluation of transport policies at the European level.

The first version of this project was developed between 2004 and 2006, with the objective of developing the first version of a European network model integrating passenger and freight transport. The idea was developing a comprehensive European transport network model that included both passenger and freight transport, along with intermodal connections, addressing the limitations of earlier models used across Europe. Building upon the strengths of existing modelling frameworks while introducing key enhancements that would serve as the foundation for an integrated tool to support transport policy-making at the EU level.

The result was the model registered under the same name *TRANS-TOOLS,* which currently is a European-wide transport model designed to simulate large-scale freight and passenger transport flows. Its main purpose is supporting the European Commission in the evaluation of infrastructure, transport policies and regulatory measures, as well as in the development and analysis of the Trans-European Transport Network (TEN-T). Since this is directly related to the transport models used in Europe, in particular to evaluate policies such as the TEN-T network, sustainability strategies, logistics, and demand projections.

The model combines different sub-models and databases to provide an integrated analysis of the European transport system. It focuses on both the supply side (available infrastructure and services) and the demand side (mobility of people and goods) (*Tools for Transport Forecasting and Scenario Testing - TRIMIS projects*, 2004).

TRANS-TOOLS was developed as the largest and most comprehensive European transport model in terms of countries covered, population covered, and geographical scale. It covers all modes (cars, trucks, trains, canal ships, sea ships and air transport) as well as both freight and passenger transport. It considers THE European Transport policy Information System (ETIS) as the data basis for assessing transport performances. The model is compound by 4 main modules:

- **Freight Demand Module:** It is composed of three integrated sub-modules. The Trade Module uses the ETIS origin-destination (O/D) matrix to forecast freight flows, including information on regions of origin, transhipment points, destination, transport mode, commodity group, and volume in tonnes. The Modal Split Module, based on the NEAC model, refines these forecasts by adjusting the stable modal split from the Trade Model, producing a detailed O/D matrix that includes mode-specific distributions. Finally, the Logistics Module, adapted from SLAM (a component of the SCENES model), assesses how changes in logistics and transport systems—such as the number and location of warehouses—affect the spatial patterns of freight movement across Europe.
- **Passenger Demand Module:** It simulates passenger transport at the European level using a combination of existing models, including SCENES, Vaclav, and ASTRA. Its primary output consists of unimodal origin-destination matrices at the NUTS-3 regional level, disaggregated by transport mode, trip purpose, number of passengers, and number of vehicles. These matrices serve as key inputs for the network assignment stage of the overall modelling framework.
- **Economic Module:** It forecasts the monetary impacts of different policy scenarios, providing relative economic changes by sector compared to a baseline scenario. These changes are passed to the NEAC model to assess further implications. Additional outputs of this module include key policy evaluation indicators such as regional GDP impacts and equivalent variation, disaggregated by region, year, and scenario, enabling detailed socio-economic analysis across Europe.
- **Network Assignment Module:** It is responsible for generating the final transport flow outputs of the model. It operates at a unimodal level, applying a stochastic assignment method based on probit models. This approach uses a multinomial normal distribution to calculate the probability of selecting specific links or routes, resulting in a realistic and probabilistic allocation of transport demand across the network infrastructure.

The TRANS-TOOLS model was developed using a methodology that emphasized modularity, scalability, and user-friendliness, enabling its continuous improvement and evolution in subsequent projects. From the outset, special attention was given to ensuring that the model remained accessible and adaptable to new modeling approaches, provided that the required data file formats were maintained. This flexible design has allowed the development of several updated versions over the years through various projects and has supported its application across a wide range of case studies—ensuring its ongoing relevance, reliability, and practical value (*Tools for transport forecasting and scenario testing | TRANS-TOOLS | CORDIS*, 2004).

The *Table 5* summarizes the subsequent versions and developments of the model through different projects since its first release.

Table 5. TRANS-TOOLS versions.

Years	Project Name	Objective	Source
2007 - 2009	WORLDNET - European Transport Network Model Refinement Regarding Freight and Intermodal Transport to and from the Rest of the World	Improve the forecasting capabilities and scenario developments of the TRANS-TOOLS (TT2) model in order to refine the European transport network model regarding freight and intermodal transport to and from the rest of the world. The aim was to facilitate the use and uptake of the model within other countries and regions. And attain a more precise representation of the freight flows between European countries and the rest of the world.	(*European Transport Network Model Refinement Regarding Freight and Intermodal Transport to and from the Rest of the World - TRIMIS projects*, 2007)
2011 - 2017	TRANSTOOLS 3 (TT3) - Research and Development of the European Transport Network Model TRANS-TOOLS Version 3	Upgrade and further develop the existing TT2 by improving the methodological basis, data foundation, identifying and solving known deficiencies and make the software faster, efficient and user centric. TT3 provide an improvement tool for supporting the evaluation and development of better transport policies.	(*Research and development of the European Transport Network Model – Transtools Version 3 | TRANSTOOLS 3 | Projekt | Results | FP7*, 2011)

Currently, TT3 remains the most integrated, multi-modal transport model designed for EU-wide policy analysis, emphasizing modularity and data transparency. The last version is TT3, although some organization uses customized one. TT3 uses standard software from Esri, Microsoft and Rapidis. This version consists of the same three main sub-models developed since the first one: Passenger demand, freight demand, and assignment. Related to passenger model includes the short and long distance for trips shorter or longer than 100km. Figure 5. Structure of the TT3 model system.is the structure of the model system ('TRANS-TOOLS 3', 2017).

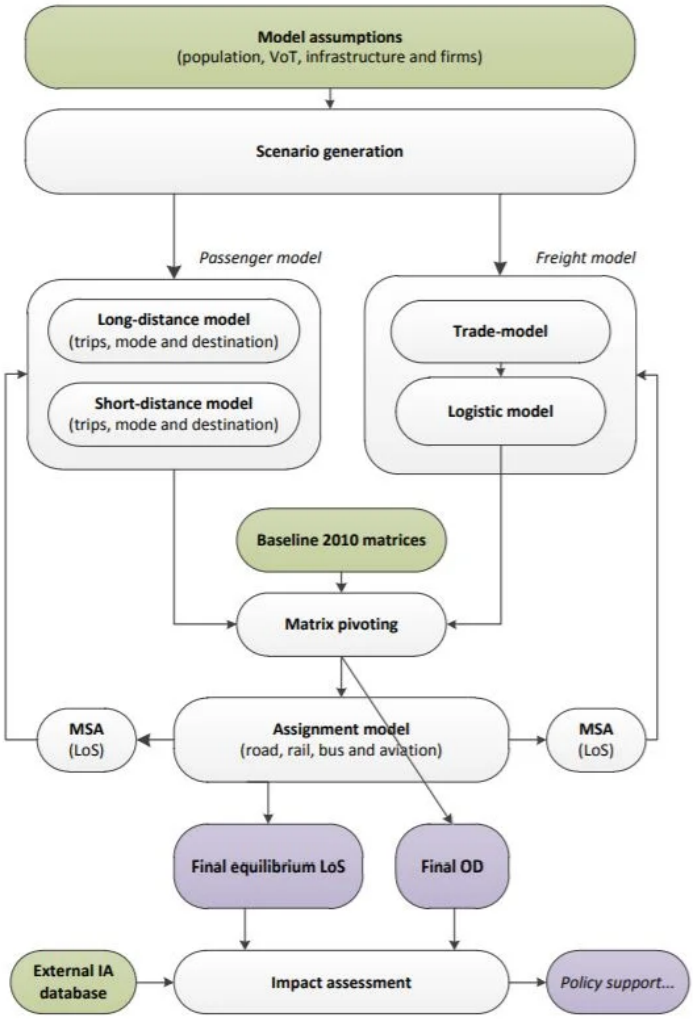


Figure 5. Structure of the TT3 model system. ('TRANS-TOOLS 3', 2017)

The relevance of the **TT3 model in the TENT-T network** context lies in that it represents a model system which has been primarily aimed at investigating large-scale in Europe. The sub models cover all relevant trip purposes across 42 countries in the continent as well as connections to the other parts of the world. The models deal with choice of mode, destination choice and frequency choice, all information relevant for understanding and defining the transportation network from passenger mobility and logistic. Moreover, even at the aggregated European level it is primarily aimed at the benchmarking of different specific objectives, so it is a useful tool for the development of the project that involves transport area at the European context, and a reference point to new developments (*TRIMIS: TRANS-TOOLS Projects*, 2004; De Jong *et al.*, 2017).

3. Review of Data-Driven solutions for Transport Applications

The increasing availability of real-time and historical transportation data has ushered in a new era of data-driven tools that complement and enhance traditional simulation approaches. Leveraging techniques from machine learning, statistics analysis and big data processing, these tools can uncover hidden patterns, predict demand, optimize operations and support adaptive, evidence-based planning and operations. In the context of transport planning and smart mobility, data-driven methodologies are particularly impactful in three core areas: traffic monitoring, incident detection and traffic predictions.

3.1. Traffic monitoring and pattern generation

Traffic monitoring involves continuous collection, processing and analysis of traffic-related data from various sources such as loop detectors, cameras, GPS devices, mobiles phones, and connected vehicles. Data-driven approaches enhance traditional monitoring by enabling real-time visualization of traffic conditions, identifying congestion patterns, and detecting anomalies in flow and speed.

Traffic pattern generation refers to the extraction of recurring traffic behaviours and trends from large datasets, which can be used to inform traffic management and predictive modelling. The procedure effectively performs clustering by grouping days based on selected features – such as the day of the week – that, according to the engineer's experience, minimize both the number of resulting patterns and the intra-group traffic flow variance. Each group's representative pattern, or exemplar, is typically defined as the centroid, calculated using the Euclidean mean of all days in the group. This approach can be generalized to other methods of exemplar selection (Torrent, 2020).

3.2. Incident detection

Incident detection focuses on the identification of unexpected events on the transport network, such as accidents, road obstructions or vehicle breakdowns. Data-driven techniques use real-time data streams – such as vehicle trajectories, speed variations etc. – to automatically detect anomalies that may indicate an incident.

Machine learning and statistical models can be trained to recognize patterns associated with disruptions, which enables a faster and more accurate detector. Recent advances in unsupervised machine learning have introduced a wide variety of promising methods for traffic anomaly and incident detection. A noteworthy contribution in this area is the study by (Papadatou, Fafoutellis and Vlahogianni, 2025), which introduces a method for real-time detection of non-recurrent congestion patterns by combining deep learning models with clustering techniques. This approach extracts typical traffic patterns from raw data and identifies deviations. Complementing this, (Gao, Ozbay and Hu, 2025) propose an enhanced model for urban traffic disruptions, adding an adaptive mechanism to refine anomaly scores over time.

3.3. Traffic predictions

Traffic prediction involves forecasting future traffic conditions based on current and historical data. This includes short-term predictions (e.g. the next 15-60 minutes) for operational management, as well as long-term forecasts for planning purposes. Forecasting traffic supply contributes significantly to anticipate and prevent traffic congestion. Predictive models use statistics methods, time-series analysis, and increasingly, deep learning techniques to anticipate traffic volumes, travel times, and congestion levels. Two main categories of data-driven approaches are: (1) Statistical (logistic regressions, autoregressive integrated moving average family of models ARIMA, Nonparametric regression, Bayesian, Hidden Markov, Kalman filter) (Smith, Williams and Keith Oswald, 2002; Guo, Huang and Williams, 2014; Qi and Ishak, 2014; Wang, Deng and Guo, 2014) and Machine Learning methods (Support Vector Machines SVM, k-Nearest Neighbours KNN, Random forest RF) (Vlahogianni, Karlaftis and Golias, 2014; Lana, del Ser and Vlahogianni, 2018) and more specifically deep learning (Recurrent Neural Networks RNN, Convolutional neural Networks CNN, Graph Neural Networks GNN, etc.) (Ma *et al.*, 2017; Yu *et al.*, 2017).

4. Overview of Use Cases

Table 6 provides an overview of the predefined use cases as initially described in Deliverable D2.2. For each use case, a brief description is included along with its corresponding typology, which refers to whether the use case is primarily planning-oriented or operational in nature. This description is essential for guiding the methodological development and ensuring that the selected data-driven and/or simulation-based approaches are well aligned with the specific characteristics and objectives of each use case.

At this stage, the information remains provisional and subject to refinement, as the project progresses and further input is gathered – particularly through stakeholders' engagement and technical validation – this section will be updated to reflect a more detailed and finalized definition of each use case. the updated version will be included in the next iteration of this deliverable.

Table 6. Summary of predefined use cases and their description

Sector	#	Use case	Description	Typology
Public Transport	1	Optimization of bus routes and frequencies	Analysis and improvement of bus routes to optimize public transport efficiency, adjusting frequencies according to passenger flows in national, regional, and urban networks	Planning
	2	Bus fleet route analysis and optimization	Analysis of bus routes to optimize fleet management, identifying efficiency opportunities in national, regional, and urban transportation networks	Operation
	3	Intermodal node demand analysis	Analyze passenger demand at interconnection points between train, tram, and bus services to optimize transportation network efficiency and user experience across national, regional, and urban environments	Planning
	4	Real-time PT user information	Provide instant, accurate travel information for passengers across multiple transport modes (airplane, train, tram, bus) at national, regional, and urban levels.	Operation
	5	Cross-border passenger transit analysis	Analyze passenger movements and transit patterns across borders using bus, car, and motorcycle transportation	Operation
	6	Urban transport peak demand analysis	Analyze passenger demand peaks in urban transportation systems, focusing on train, tram, and bus services across regional, metropolitan, and urban environments	Planning
Freight Transport	7	Port-based freight flow analysis	Analyze freight transportation flows originating from ports, focusing on train and truck logistics at the national level	Operation
	8	Airport-origin freight flow analysis	Analyze goods transportation flows originating from airports, focusing on air and truck logistics at the national level	Operation
	9	Logistics platform location in corridors.	Identify optimal locations for logistics platforms along transportation corridors, focusing on train and truck logistics at the national level	Planning

Sector	#	Use case	Description	Typology
	10	Cross-border freight transit analysis	Analyze goods transportation movements across borders, focusing on train and truck logistics in cross-border regions	Operation
Traffic	11	Main road segment capacity analysis	Comprehensive analysis of traffic capacity and flow dynamics on primary road segments, examining interactions between cars, buses, trucks, and motorcycles at the national infrastructure level	Planning
	12	Predictive bottleneck resolution	Identify and address traffic bottlenecks using predictive information for cars, trucks, and motorcycles at the national level	Operation
	13	Highway and main road congestion monitoring	Monitor traffic congestion on highways and primary roads for cars, buses, trucks, and motorcycles at the national level	Operation
	14	Special operations congestion prediction	Predict traffic congestion during special events or operations for cars and trucks at the regional level	Planning
	15	Real-time network traffic mapping	Create a dynamic, real-time traffic map of the regional transportation network	Operation
Sustainability	16	Transport emissions evaluation	Assess emissions generated by cars, buses, trucks, and motorcycles at the national level	Planning
	17	Informed shared mobility	Analyze and optimize car-sharing practices and mobility patterns at the national level.	Operation
	18	Freight emissions analysis	Evaluate emissions generated by air, train, and truck transportation, with specific focus on truck logistics at the national level	Planning
	19	Electric vehicle charging infrastructure	Analyze and optimize electric vehicle charging network for cars at national level.	Operation
	20	Urban mobility tracking by mode and distance	Monitor urban mobility patterns for tram, bus, car, motorcycle, bicycle, and pedestrian transportation in metropolitan and urban areas	Operation
	21	Low emission zone monitoring	Track and analyze transportation in Low Emission Zones, including tram, bus, car, motorcycle, bicycle, pedestrian, and truck movements in metropolitan and urban areas	Operation
	22	Active mobility tracking	Monitor and analyze bicycle and pedestrian mobility patterns in metropolitan and urban environments	Operation

5. Models and Use Cases: Identifying gaps and necessary solutions

In this section, we present a structured approach to aligning use cases with appropriate modelling techniques, specifically on the data-driven and simulation-based models presented in the previous sections. On one side, we have a set of use cases that reflect real-world requirements and operational challenges for stakeholders. On the other, we have a suite of models – each with distinct capabilities and limitations – that must be assigned to these use cases in a way that maximizes relevance and effectiveness.

By assigning models to use cases, we aim to uncover where current capabilities fall short, and what additional resources or developments are needed. This process not only helps in identifying gaps and unmet needs but also provides a clear path towards resolving them through targeted solutions.

To support this, we introduce in this section a methodology for gap and need identification, which includes classifying use cases based on the type of model they utilize. This classification enables a clearer understanding of model applicability and limitations. A summary of the key gaps and needs, along with insights into how these can be addressed is also presented in this section.

5.1. Existing transport models of the use cases

At this stage of the project, no formal feedback has been received from stakeholders regarding the availability and status of existing transport models relevant to the identified use cases. This represents a critical gap in the information needed to ensure accurate model-to-use-case alignment.

Moreover, the project partners do not currently own or have direct access to transport models or mobility data from the cities associated with the use cases. As a result, there is a clear need to engage with local stakeholders and data owners to request access to these resources. This step is essential to validate the applicability of existing models, assess their calibration status, and determine whether they can be effectively used and integrated into the project.

In the absence of this information, we will, for the time being, present a set of simulation-based transport models that the project partners have previously used in other contexts or projects. These models are known to exist and have been applied in similar domains. However, their status – such as whether they have been updated, calibrated, or adapted to new data – is unknown.

As the project progresses, a dedicated effort will be made to:

- Request detailed information from stakeholders regarding existing models and data sources.
- Assess the relevance of these models for the current use cases.
- Identify any gaps that may require the development of new models or the adaptation of existing ones.

Based on the knowledge and expertise of the DS4MOB project partners, as well as insights gained from their involvement in previous projects, several transport models have been identified in *Table 7*, which are well-suited to the location of each pilot. However, as partners do not currently have direct access

to these models, it is not possible to determine their present status. Further engagement with local stakeholders and data providers will be necessary to assess their applicability.

Table 7. Existing transport models according to the knowledge of the DS4MOB partners.

Pilot	Location	(Potential) Existing model	Software	Resolution	Owner
Spanish National Pilot	Spain	National transport model[1]	Visum	Macroscopic	Spanish Government
Regional Pilot: Andalucia	Andalucian region	-	-	-	-
Regional pilot: Catalonia	Catalonian region	Public (and somehow private) transport model for Catalonia	Visum	Macroscopic	Metropolitan Transport Authority (ATM)
Regional Pilot: Comunitat Valenciana	C. Valenciana region	-	-	-	-
Local pilot: metropolitan area of Madrid	Madrid	Model of Madrid city (up to the M30 ring road)	Aimsun Next	Macroscopic	Madrid City Council
Local pilot: metropolitan area of Valencia	Valencia	Model of Valencia city (not the metropolitan area)	Aimsun Next	Microscopic	Valencia City Council
Local pilot: metropolitan area of Barcelona	Barcelona	Model of the Eixample of Barcelona (not the metropolitan area)	Aimsun Next	-	-

[1] Source: https://cdn.mitma.gob.es/portal-web-drupal/hermes/documento_diseno_general_del_mnt.pdf

5.2. Existing data-driven solutions for application

In scenarios where simulation-based transport models are not available, data-driven approaches offer viable alternatives. This section presents several existing solutions developed by Aimsun, which leverage data analytics and machine learning techniques to support transport planning and operations without relying on traditional simulation frameworks. Figure 6 shows the main data-driven modules that will be described in the subsequent sections.

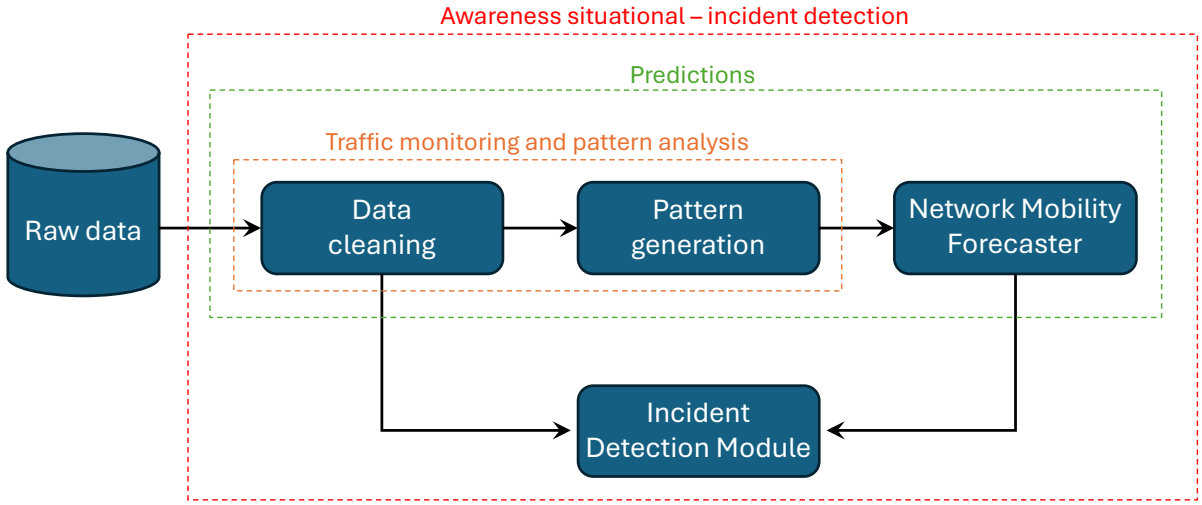


Figure 6. Data-driven modules for traffic monitoring, pattern generation, predictions and incident detection developed by Aimsun.

5.2.1. Raw data analyzer and cleaner algorithm

The purpose of the raw data analyzer algorithm is to perform an initial scan of time-series transport data, such as traffic flow, travel demand, or speed. The first step provides a comprehensive overview of the historical raw data by identifying and flagging values such as “NaN”, "zero", "outlier", "negative" entries. The algorithm also assesses the reliability of each data point and visualizes the results through heatmaps and statistics summaries for all variables in the dataset. These heatmaps typically display average traffic metrics (e.g., flows), helping to highlight patterns and anomalies.

In addition, the analyzer estimates the reliability of each traffic detector by evaluating key aspects such as start and end times, data duration, days of week covered, and overall data completeness.

Once the raw data has been assessed, the next step involves applying a data cleaning algorithm. This process filters out unrealistic or erroneous values and imputes missing data where necessary. The choice of filters and imputation techniques depends on the quality and characteristics of the input data. Users can activate or deactivate specific filters and imputation methods as needed, allowing for a flexible and adaptive approach.

5.2.2. Pattern generation algorithm

Once the raw data is cleaned and pre-processed, the next step involves applying the pattern generation algorithm. This algorithm is designed to extract daily traffic patterns from the dataset using unsupervised Machine Learning (ML) techniques, specifically clustering methods.

The goal is to identify typical daily profiles – such as flow patterns – by grouping similar days together. The clustering-based process consists of three main stages:

- Extracting daily patterns using clustering.
- Extracting explanatory variables.

- Training the pattern matcher to predict a pattern that should be assigned to each day.

The implementation of the pattern generator algorithm uses agglomerative hierarchical clustering or k-means. These two methods are chosen because the user can specify the number of clusters, and they scale well with the number of data samples. Once the clustering is performed, it is needed to calculate the pattern that represents each cluster. The latter process consists of two independent algorithms in which forecasts are merged to give a single forecast. Both forecasts are then combined considering the amount and quality of intraday data

5.2.3. Network mobility forecaster

After identifying the representative traffic patterns in the cleaned data, the next step is to make predictions to estimate future traffic conditions. These predictions are then used as input for incident detection. The Network Mobility Forecaster (NMF) algorithm provides short- and mid-term forecasts—ranging from the next 15 to 60 minutes or even the next few hours—by leveraging historical, real-time, or near-real-time data.

As a first step for the proper operation of the NMF algorithm, the input data and the corresponding detectors must be mapped onto the network graph at regular time intervals (e.g., every 5 minutes). Flow and speed measurements, along with their corresponding timestamps and detector IDs, are then matched with the road sections in the Athens use case. Next, a supervised Machine Learning (ML) model is trained and tested using cross-validation; in this case, four folds were used. Metrics such as the GEH Statistic, Root Mean Squared Error (RMSE), and Scalable Quality Value (SQV) are employed to evaluate the quality of the predictions.

5.2.4. Aimsun incident detection method

Aimsun's Incident Detection Module (IDM) is an anomaly detection algorithm specialized in early detection of incidents based on historical or real-time traffic data. Using the clean data and the NMF predictions of the previous steps, the algorithm applies different filters to detect anomalies as potential traffic incidents and also can combine timeseries from various variables and sensors (e.g., flow and speed). This latter capability makes the algorithm adaptable to real-world conditions where the temporal and spatial resolution of the sensors is never the same.

The algorithm starts by analysing the temporal dependencies of each timeseries, e.g. flow of each loop detector, and estimates the level of outlierness of every datapoint (as defined above, the recommended timeseries length is, at least, 30 days). Then, the level of outlierness of each datapoint is modified considering numerous factors, but mainly the distance between the measured outlier and the expected datapoint for every 5-minute time interval. Finally, the IDM applies a dynamic and personalized thresholding to flag datapoints as candidate incidents or not.

5.2.5. Speed recommendation

The speed recommendation module is designed to enhance driving efficiency, safety and sustainability by dynamically suggesting optimal speeds based on real-time traffic conditions and specific operational

goals. Depending on the context in which the module is deployed, it adapts its recommendations to meet different objectives.

For example, in scenarios involving traffic congestion, the module analyses speed data across road sections and recommends speeds that minimize abrupt speed changes. By smoothing out speed transitions – especially between congested and free-flowing segments – it helps reduce stop-and-go driving, improve traffic flow, and lower emissions.

Another application of the Speed Recommendation module involves the use of pollution data to support environmentally sustainable traffic management. In this scenario, the module identifies road sections where pollution levels exceed predefined thresholds. When the pollution protocol is activated for any of these sections, the system recommends a constant speed across the affected areas to reduce emissions and improve air quality.

5.3. Methodology for identifying gaps and needs

To ensure that each use case is addressed with the most appropriate and feasible approach, we adopt a methodology that **evaluates both data-driven and simulation-based solutions** to accomplish the requirements of use cases. This process is designed to identify any gaps in data availability, modelling capabilities, or methodological requirements that may hinder the successful implementation of a solution.

The methodology showed in Figure 7 begins with an exploration of existing data-driven solutions that could be applied to the predefined use cases. In this phase, we first identify whether there are **existing methods, algorithms or models** that can address the specific objectives of the use case. Once potential solutions are identified, the minimum data requirements to apply these specific solutions are identified. Data needs to be acquired from partnerships with stakeholders or from open data sources if available. This is crucial because the lack of data can lead to the infeasibility of the use case. It is also important to identify if the data-driven solution depends on the existence of a calibrated transport model. If such model is required, we need to identify if the model is available and whether its calibration is adequate for the intended analysis.

If all necessary data is available, no additional modelling is required, and the scope of the solution fits the objectives of the use case, the data-driven method can be directly applied to the use case. However, if the data is insufficient, the model is unavailable or uncalibrated or the solution does not align well with the use cases objectives, alternative approaches need to be proposed.

In cases where data-driven methods are not viable, we shift our focus to simulation-based solutions. This involves determining whether a transport simulation model exists that is relevant to the use case. If a model is available, we will assess its calibration status and its ability to simulate the required scenarios. If the model is not calibrated, we might evaluate the feasibility of performing a calibration using available data.

If neither calibration nor model development is feasible within the project constraints, we may consider redefining or adapting the use case to better align with the available tools and data. This iterative and adaptive approach ensures that each use case is addressed in the most efficient and realistic manner.

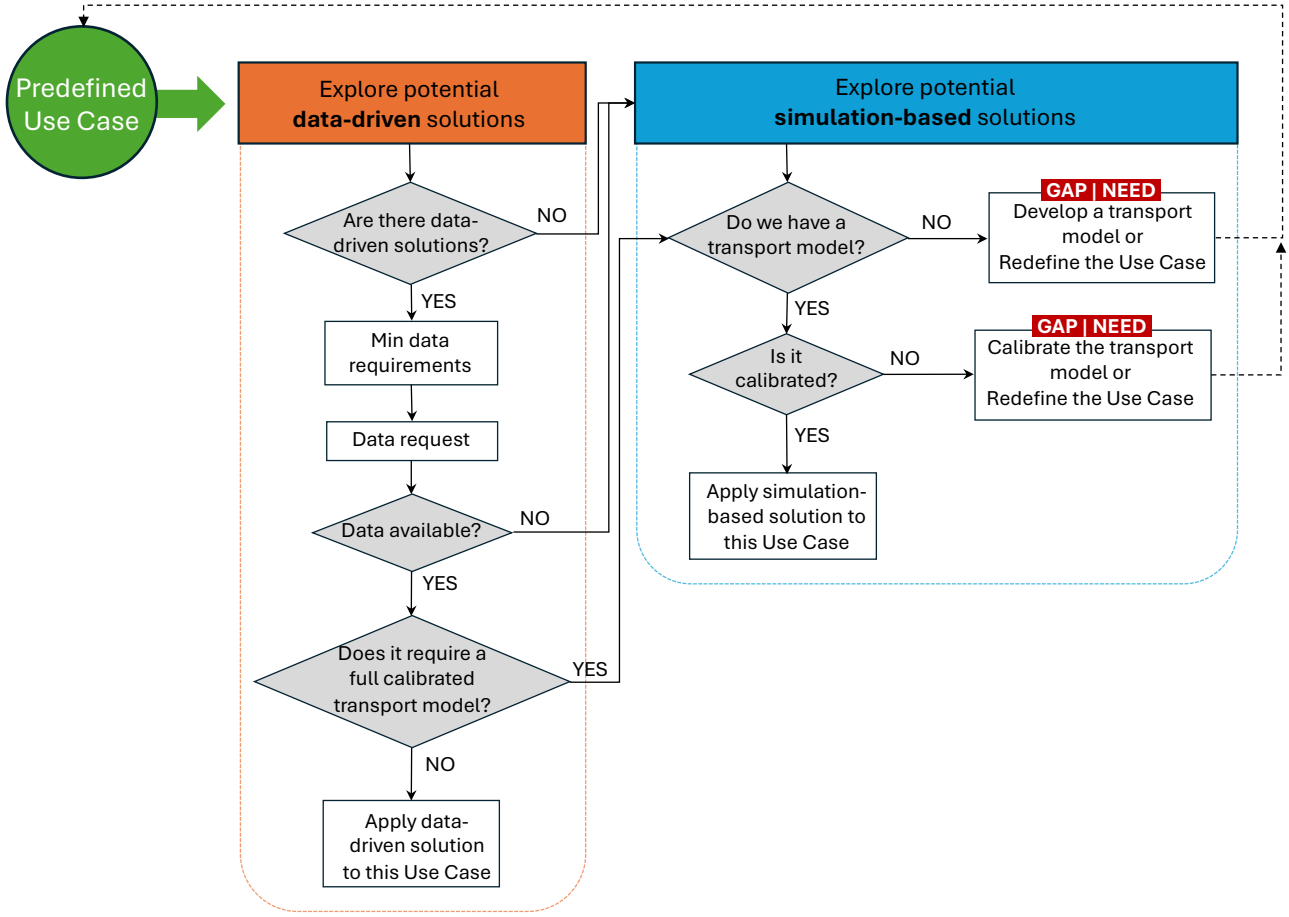


Figure 7. Methodology to explore data-driven and/or simulation-based solutions for the predefined Use Cases

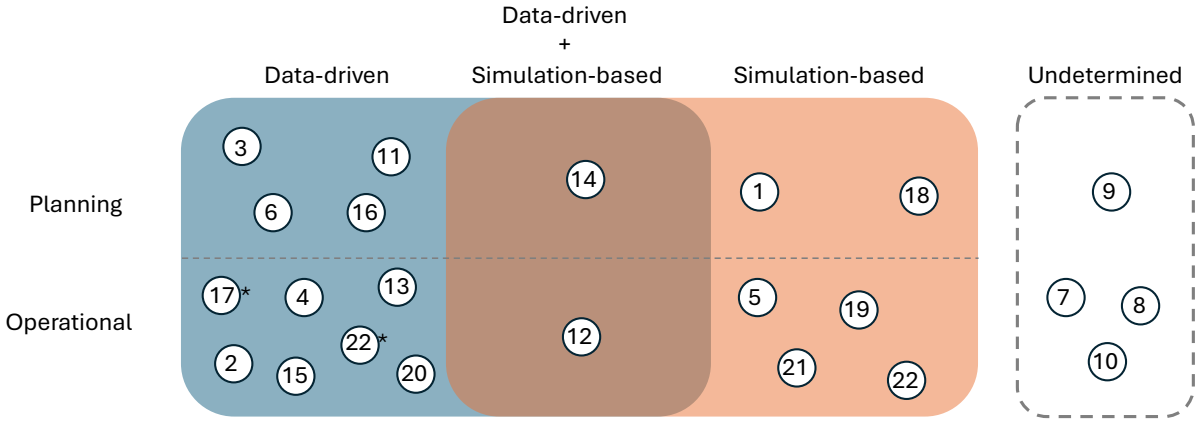
5.4. Use Case classification based on data-driven and simulation-based solutions

The application of the methodology described in the previous section has enabled a **preliminary classification** of predefined use cases **according to the most suitable type of solution**: data-driven, simulation-based, or a hybrid approach combining both. This classification aims to guide the selection of appropriate techniques for each use case. In this version 1 of the deliverable, the classification has been carried out independently of data availability. That is, the feasibility of implementing data-driven approaches has not yet been validated against actual data sources. The same happens with simulation-based models. These aspects will be addressed in coordination with WP2 ("Identification, definition and validation of mobility and logistics use cases") and WP3 ("Data interoperability and fairness"), and the outcomes will be reflected in Version 2 of this deliverable.

The classification also distinguishes between **planning and operational use cases**, following the framework established in Deliverable D2.2. Planning use cases typically involve long-term strategic decision, whereas operational use cases are more concerned with short-term or real-time decision-making.

Figure 8 illustrates the **current classification of use cases** along these two axes: the type of approach and the temporal focus. A small number of use cases remain unclassified due to insufficient detail at this point in the project. These will be revisited and refined in the second version of this deliverable, once additional information becomes available.

It is important to emphasize that this classification is provisional and subject to change. As the project progresses and more data becomes accessible, the suitability of each approach will be reassessed to ensure alignment with both technical feasibility and project goals.



Public Transport:
1. Optimization of bus routes and frequencies
2. Bus fleet route analysis and optimization
3. Intermodal node demand analysis
4. Real-time PT user information
5. Cross-border passenger transit analysis
6. Urban transport peak demand analysis

Freight Transport:
7. Port-based freight flow analysis
8. Airport-origin freight flow analysis
9. Logistics platform location in corridors.
10. Cross-border freight transit analysis

Traffic:
11. Main road segment capacity analysis
12. Predictive bottleneck resolution
13. Highway and main road congestion monitoring
14. Special operations congestion prediction
15. Real-time network traffic mapping

Sustainability:
16. Transport emissions evaluation
17. Informed shared mobility
18. Freight emissions analysis
19. Electric vehicle charging infrastructure
20. Urban mobility tracking by mode and distance
21. Low emission zone monitoring
22. Active mobility tracking

NOTES:
17* Apply the Data-driven tool only for the analysis of mobility patterns. No optimization involved.
22* Only bicycle mobility patterns

Figure 8. Classification of use cases based on data-driven and/or simulation-based approaches

Following the initial classification of use cases, those identified as suitable for data-driven solutions have been further examined to define their potential scope. *Table 8* presents the **predefined use cases alongside the envisioned data-driven approaches** that could be applied to each. These approaches are aligned with the existing capabilities developed by Aimsun, presented in Section 5.2, and serve as a foundation for further discussion and integration with potential solutions that might be proposed by other project partners. This collaborative process aims to broaden the scope and applicability of each use case.

For use cases that involve hybrid approaches, the table also specifies whether a calibrated transport model is required. In such cases, the solution combines both data-driven and simulation-based methodologies, leveraging the strengths of each to address complex mobility challenges.

Additionally, a **high-level summary of the minimum data requirements** has been included for each use case. These requirements represent the essential datasets needed to implement the proposed data-driven solution. If these data are not available, the use case cannot be developed under the defined scope and may need to be reconsidered or reclassified.

In contrast, certain use cases have not been associated with a specific data-driven solution. These are expected to rely primarily on simulation-based approaches due to the nature of the problem. This distinction is particularly important for WP2 and WP3, as it provides a clear indication of where data collection efforts should be prioritized and where simulation capabilities will play a more central role.

This information will be instrumental in upcoming workshops with stakeholders, where WP2 and WP3 will work collaboratively to refine the final definition and implementation strategy for each use case.

Table 8. Potential Data-Driven Use Cases and minimum data requirements

Area	Use Case	Description	Potential Use Case scope with data-driven methods	Does it require a calibrated transport model?	Minimum data requirements
Public Transport	1	Optimization of bus routes and frequencies	-	-	-
	2	Bus fleet route analysis and optimization	Prediction of travel times for buses between bus stops	No	Network geometry (ANG + network of FCD provider) + historical FCD of buses with speeds and travel times + traffic counts + locations of bus stops
	3	Intermodal node demand analysis	Prediction of incoming and outgoing passengers per intermodal hub	No	Network geometry + historical number of passengers accessing and leaving each node + location of nodes
	4	Real-time PT user information	Prediction of arrival times of buses per station	No	Network geometry (ANG + network of FCD provider) + real-time FCD of buses with speeds and travel times + traffic counts + locations of bus stops
	5	Cross-border passenger transit analysis	-	-	-
	6	Urban transport peak demand analysis	Generate demand patterns based on historical data	No	Network geometry (ANG + network of FCD provider if FCD is considered) + traffic counts and/or historical FCD

Area	Use Case	Description	Potential Use Case scope with data-driven methods	Does it require a calibrated transport model?	Minimum data requirements
Freight Transport	7	Port-based freight flow analysis	-	-	-
	8	Airport-origin freight flow analysis	-	-	-
	9	Logistics platform location in corridors.	-	-	-
	10	Cross-border freight transit analysis	-	-	-
Traffic	11	Main road segment capacity analysis	Congestion prediction, incident detection and speed recommendation based on traffic conditions	No	Network geometry with capacity of sections (ANG + network of FCD provider if FCD is considered) + historical FCD + traffic counts + list of incidents for the same period of the counts and the FCD + ranges of speed
	12	Predictive bottleneck resolution	Traffic prediction, incident detection and response plans evaluation	Yes	Network geometry (ANG + network of FCD provider if FCD is considered) + historical FCD + traffic counts + list of incidents for the same period of the counts and the FCD + response plans
	13	Highway and main road congestion monitoring	Traffic prediction, detection of incidents, speed recommendations	No	Network geometry (ANG + network of FCD provider if FCD is considered) + historical FCD + traffic counts + list of incidents for the same period of the counts/FCD
	14	Special operations congestion prediction	Traffic prediction for nonrecurrent conditions caused by expected events	Yes, if response plans are assessed	Network geometry (ANG + network of FCD provider if FCD is considered) + historical FCD + traffic counts + response plans
	15	Real-time network traffic mapping	Traffic monitoring and predicting	No	Network geometry (ANG + network of FCD provider if FCD is considered) + real-time data of FCD + traffic loop detectors
Sustainability	16	Transport emissions evaluation	Traffic prediction and estimations of pollutants caused by traffic	No	Network geometry (ANG + network of FCD provider if FCD is considered) + FCD + traffic loop detectors (flow and speed) + distribution of transport fleet (e.g., number of heavy vehicles, light vehicles, etc)

Area	Use Case	Description	Potential Use Case scope with data-driven methods	Does it require a calibrated transport model?	Minimum data requirements
	17	Informed shared mobility	Generate demand patterns based on historical data	No	Network geometry (ANG + network of FCD provider if FCD is considered) + traffic counts and/or historical FCD
	18	Freight emissions analysis	-	-	-
	19	Electric vehicle charging infrastructure	-	-	-
	20	Urban mobility tracking by mode and distance	Generate demand patterns based on historical data	No	Network geometry (ANG + network of FCD provider if FCD is considered) + traffic counts and/or historical FCD
	21	Low emission zone monitoring	-	-	-
	22	Active mobility tracking	Generate demand patterns based on historical data	No	Network geometry (ANG + network of FCD provider if FCD is considered) + traffic counts and/or historical FCD

5.5. Preliminary identification of needs and gaps

Based on the provisional classification of use cases presented in the previous section, several gaps and needs have been identified that require further attention as the project progresses. These insights are critical for refining the scope of the use cases and ensuring their feasibility from both a methodological and data availability perspective.

Gaps identified include the following:

It remains unclear how **cross-border use cases** will be addressed, particularly in terms of data integration, governance, and interoperability across different jurisdictions. These aspects pose significant challenges, especially when aligning data standards, access protocols, and institutional responsibilities across national boundaries. On the technical side, it is currently assumed that a cross-border transport simulation model does not exist. As a result, these use cases are expected to rely primarily on data-driven approaches. However, if a transport model were to be developed specifically for this purpose, it would require substantial time and resources which may not be feasible within the current project scope. This uncertainty highlights the importance of further investigation and coordination with relevant stakeholders to determine the most viable path forward for cross-border scenarios.

We have also identified **several challenges** in assigning appropriate methodological approaches to use cases within the **freight transport sector**. One of the main limitations is that existing transport

simulation models typically do not account from train-based logistics chains, meaning that only road freight – primarily truck movements – can currently be simulated. This significantly restricts the scope of the simulation-based solutions for multimodal freight scenarios. Furthermore, there is a lack of clarity regarding the **availability and accessibility of data** related to freight flows originating from key infrastructure nodes such as airports and ports. Without reliable data sources or established mechanisms for accessing this information, it becomes difficult to define and implement robust data-driven solutions for freight-related use cases.

Needs identified include the following:

It is crucial to explore **data availability and the existence of transport models** in close collaboration with stakeholders, as these elements will directly influence the feasibility and design of the proposed use cases. Establishing the necessary agreements to access and use stakeholder data within the project is essential. In particular, the absence of key datasets may lead to a change in the methodological approach or even a redefinition of certain use cases.

Equally important is understanding the characteristics of existing transport models -whether they are calibrated, their geographical scope, the level of detail, and the specific study areas they cover. This information must be systematically gathered and reflected in the corresponding use cases to ensure methodological consistency and technical viability. These aspects will be a central focus in the upcoming **stakeholder workshops** led by WP2, where alignment between data, models and use case requirements will be further refined.

6. Conclusions

This deliverable has outlined the technical design framework for transport models intended to support the implementation of predefined use cases within the DS4MOB project. By briefly reviewing existing modelling approaches – both simulation-based and data-driven- it has established a foundational understanding of the methodologies available and their respective strengths and limitations. By exploring the various data sources and fusion techniques, the deliverable also highlights the potential for integrating heterogeneous datasets to enhance mode accuracy.

Through the analysis of current use cases, the deliverable has sought to align each case with the most appropriate modelling technique – either simulation-based or data-driven approach. This alignment exercise has revealed both opportunities and constraints, highlighting where existing models and data sources are sufficient, and where gaps remain that may hinder implementation.

Additionally, the deliverable has identified several critical needs, including the availability of calibrated transport models for the use cases, access to relevant datasets to apply data-driven solutions, and the feasibility of applying certain approaches in cross-border and freight-related scenarios. These insights will be complemented in the upcoming tasks within the project, specifically in the stakeholder workshop that will take place in the next months. A second version of this deliverable will be updated.

Overall, this deliverable serves as a strategic starting point for the technical realization of the DS4MOB use cases. It provides a clear direction for refining model designs and addressing limitations.

7. References

Autoridad Portuaria de Barcelona (APB) (2025) *Port de Barcelona | Inicio*. Available at: https://www.portdebarcelona.cat/es (Accessed: 28 July 2025).

Beneduce, C. *et al.* (2025) 'pySpainMobility: a Python Package to Access and Manage Spanish Open Mobility Data'. arXiv. Available at: https://doi.org/10.48550/ARXIV.2506.13385.

Bhaskar, A. and Chung, E. (2013) 'Fundamental understanding on the use of Bluetooth scanner as a complementary transport data', *Transportation Research Part C: Emerging Technologies*, 37, pp. 42–72. Available at: https://doi.org/10.1016/J.TRC.2013.09.013.

Chiu, Y. *et al.* (2011) *Dynamic Traffic Assignment: A Primer*, *Transportation Network Modeling Committee*. Available at: https://doi.org/10.1016/j.trd.2016.06.003.

Chung, Y. (2001) 'Travel time estimation algorithm using GPS probe and loop detector data fusion', in *Transportation Research Board 80th Annual Meeting [CD-ROM], Washington, DC*. Available at: https://www.academia.edu/3865872/Travel_time_estimation_algorithm_using_GPS_probe_and_loop_detector_data_fusion.

De Jong, G. *et al.* (2017) 'Modelling production-consumption flows of goods in Europe: the trade model within Transtools3', *Journal of Shipping and Trade*, 2(1), p. 5. Available at: https://doi.org/10.1186/s41072-017-0023-9.

Empresa Municipal de Transportes de Valencia S.A.U (2025) *EMT Valencia*, *EMT Valencia*. Available at: https://www.emtvalencia.es/ciudadano/index.php (Accessed: 28 July 2025).

European Commission (2013) *TENtec Information System and TEN-T map library*. Available at: https://transport.ec.europa.eu/transport-themes/infrastructure-and-investment/trans-european-transport-network-ten-t/tentec-information-system-and-ten-t-map-library_en (Accessed: 28 July 2025).

European Commission (2024) *TEN-T Newsletters - Towards the Mediterranean European Transport Corridor*. Available at: https://ec.europa.eu/newsroom/mytent_t/items/844637/en (Accessed: 28 July 2025).

European Commission (2025) *TEN-T newsletters*, *European Commission: Mobility and Transport*. Available at: https://transport.ec.europa.eu/transport-themes/infrastructure-and-investment/trans-european-transport-network-ten-t/ten-t-newsletters_en (Accessed: 28 July 2025).

European Environment Agency (2025) *European Environment Agency: Data and maps*, *European Environment Agency*. Available at: https://www.eea.europa.eu/data-and-maps (Accessed: 28 July 2025).

European Transport Network Model Refinement Regarding Freight and Intermodal Transport to and from the Rest of the World - TRIMIS projects (2007). Available at: http://trimis.ec.europa.eu/project/european-transport-network-model-refinement-regarding-freight-and-intermodal-transport-and (Accessed: 28 July 2025).

Faouzi, N.-E.E., Klein, L.A. and Mouzon, O.D. (2009) 'Improving Travel Time Estimates from Inductive Loop and Toll Collection Data with Dempster–Shafer Data Fusion':, *https://doi.org/10.3141/2129-09*, (2129), pp. 73–80. Available at: https://doi.org/10.3141/2129-09.

'Four Stages Transportation / Land Use Model | The Geography of Transport Systems' (2018), 19 January. Available at: https://transportgeography.org/contents/methods/spatial-interactions-gravity-model/transportation-land-use-four-stages-model/ (Accessed: 28 July 2025).

Furness, K.P. (1965) 'Time function iteration', *Traffic Engineering and Control*, 7, pp. 458–460.

Gao, J., Ozbay, K. and Hu, Y. (2025) 'Real-time anomaly detection of short-term traffic disruptions in urban areas through adaptive isolation forest', *Journal of Intelligent Transportation Systems*, 29(3), pp. 269–286. Available at: https://doi.org/10.1080/15472450.2024.2312809.

Guo, J., Huang, W. and Williams, B.M. (2014) 'Adaptive Kalman filter approach for stochastic shortterm traffic flow rate prediction and uncertainty quantification', *Transportation Research Part C: Emerging Technologies*, 43, pp. 50–64.

INE. Instituto Nacional de Estadística (2025) *INE*. Available at: https://www.ine.es/ (Accessed: 28 July 2025).

Jaspers Joint Assistance (2014) 'The Use of Transport Models in Transport Planning and Project Appraisal The Use of Transport Models in Project Appraisal'.

Kashinath, S.A. *et al.* (2021) 'Review of data fusion methods for real-time and multi-sensor traffic flow analysis', *IEEE Access*, 9, pp. 51258–51276. Available at: https://doi.org/10.1109/ACCESS.2021.3069770.

Lana, I., del Ser, J. and Vlahogianni, E.I. (2018) 'Road Traffic Forecasting: Recent Advances and New Challenges', *IEEE Intelligent Transportation Systems Magazine*, 10(2), pp. 93–109.

Li, W. *et al.* (2017) 'Citywide Estimation of Traffic Dynamics via Sparse GPS Traces', *IEEE Intelligent Transportation Systems Magazine*, 9(3), pp. 100–113. Available at: https://doi.org/10.1109/MITS.2017.2709804.

Ma, X. *et al.* (2017) 'Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction', *Sensors*, 17(4).

Mahmassani, H.S. (2001) 'Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Applications', *Networks and Spatial Economics 2001 1:3*, 1(3), pp. 267–292. Available at: https://doi.org/10.1023/A:1012831808926.

McKinsey (2021) 'Unlocking full value from data from connected cars |'. Available at: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/unlocking-the-full-life-cycle-value-from-connected-car-data#.

Ministerio de Transportes y Movilidad Sostenible (2025a) *Plataforma HERMES*. Available at: https://www.transportes.gob.es/ministerio/proyectos-singulares/plataforma-hermes (Accessed: 28 July 2025).

Ministerio de Transportes y Movilidad Sostenible (2025b) *Portal del Observatorio del Transporte y la Logística en España | OTLE*. Available at: https://otle.transportes.gob.es/ (Accessed: 28 July 2025).

Ministerio del Interior del Gobierno de España (2025) *Dirección General de Tráfico*. Available at: https://www.dgt.es/inicio/ (Accessed: 28 July 2025).

Mouter, N. (2021) 'Standard transport appraisal methods', in *Advances in Transport Policy and Planning*. Elsevier, pp. 1–7. Available at: https://doi.org/10.1016/bs.atpp.2021.02.001.

Ortúzar, J. de D.O. and Willumsen, L.G. (2011) *Modelling transport*. Wiley-Blackwell.

Ortúzar S., J. de Dios. and Willumsen, L.G. (2011) *Modelling transport*. Wiley-Blackwell.

Otonomo (2022) '4 Ways AI and Machine Learning Use Connected Vehicle Data in Transportation'. Available at: https://otonomo.io/blog/ai-in-transportation/.

Ou, Q. (2011) 'Fusing Heterogeneous Traffic Data: Parsimonious Approaches using Data-Data Consistency'. TRAIL Research School. Available at: https://repository.tudelft.nl/islandora/object/uuid%3A7118e92f-bf63-47ac-8c50-01a35647068a (Accessed: 2 October 2021).

Ou, Q., Lint, H. van and Hoogendoorn, S.P. (2010) 'Fusing Heterogeneous and Unreliable Data from Traffic Sensors', *Studies in Computational Intelligence*, 281, pp. 511–545. Available at: https://doi.org/10.1007/978-3-642-11688-9_19.

Papadatou, K.N., Fafoutellis, P. and Vlahogianni, E.I. (2025) 'Revisiting the Non-recurrent Traffic Incident Identification Problem for Real-Time Applications Using Theory-Aware Unsupervised Learning', *Data Sci. Transp*, 7(12). Available at: https://doi.org/10.1007/s42421-025-00126-4.

POLIS Network (2025) 'Urban Nodes', *POLIS Cities and regions for Transport Innovation*. Available at: https://www.polisnetwork.eu/what-we-do/working-groups/urban-nodes/ (Accessed: 28 July 2025).

Portada | Ministerio de Transportes y Movilidad Sostenible (2025). Available at: https://www.transportes.gob.es/ (Accessed: 28 July 2025).

Qi, Y. and Ishak, S. (2014) 'A Hidden Markov Model for short term prediction of traffic conditions on freeways', *Transportation Research Part C: Emerging Technologies*, 43, pp. 95–111.

Research and development of the European Transport Network Model – Transtools Version 3 | TRANSTOOLS 3 | Projekt | Results | FP7 (2011) *CORDIS | European Commission*. Available at: https://cordis.europa.eu/project/id/266182/results (Accessed: 28 July 2025).

Roads, B. of P. (1964) *Traffic Assignment Manual for Application with a Large, High Speed Computer*. Edited by U.S.D. of Commerce. Available at: https://books.google.es/books?id=AvNUR_O_JEcC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.

Smith, B.L., Williams, B.M. and Keith Oswald, R. (2002) 'Comparison of parametric and nonparametric models for traffic flow forecasting', *Transportation Research Part C: Emerging Technologies*, 10(4), pp. 303–321.

Tools for transport forecasting and scenario testing | TRANS-TOOLS | CORDIS (2004) *CORDIS | European Commission*. Available at: https://cordis.europa.eu/project/id/502644/es (Accessed: 28 July 2025).

Tools for Transport Forecasting and Scenario Testing - TRIMIS projects (2004). Available at: http://trimis.ec.europa.eu/project/tools-transport-forecasting-and-scenario-testing (Accessed: 28 July 2025).

Torrent, F. (2020) 'How to extract patterns from traffic data for better insights into mobility', *Aimsun blogs*. Available at: https://www.aimsun.com/innovation-blog/how-to-extract-patterns-from-traffic-data-for-better-insights-into-mobility/ (Accessed: 23 July 2025).

'TRANS-TOOLS 3' (2017) *Rapidis*. Available at: https://www.rapidis.com/transtools-3-transport-model-for-europe/?lang=en (Accessed: 28 July 2025).

TRANSyT UPM (2025) 'Observatorio de la Movilidad Metropolitana – OMM'. Available at: https://observatoriomovilidad.es/ (Accessed: 28 July 2025).

Trasporti e Territorio (2022) 'Development of multimodal digital mobility services', *TRT Trasporti e Territorio*. Available at: https://www.trt.it/en/PROGETTI/development-of-multimodal-digital-mobility-services/ (Accessed: 28 July 2025).

Travel Demand Modelling | Australian Transport Assessment and Planning (no date). Available at: https://www.atap.gov.au/tools-techniques/travel-demand-modelling/index (Accessed: 28 July 2025).

TRIMIS: TRANS-TOOLS Projects (2004). Available at: http://trimis.ec.europa.eu/search (Accessed: 28 July 2025).

Vlahogianni, E.I., Karlaftis, M.G. and Golias, J.C. (2014) 'Short-term traffic forecasting: Where we are and where we're going', *Transportation Research Part C: Emerging Technologies*, 43, pp. 3–19.

Wang, J., Deng, W. and Guo, Y. (2014) 'New Bayesian combination method for short-term traffic flow forecasting', *Transportation Research Part C: Emerging Technologies*, 43, pp. 79–94.

Wardrop, J.G. (1952) 'ROAD PAPER. SOME THEORETICAL ASPECTS OF ROAD TRAFFIC RESEARCH.', in *Proceedings of the Institution of Civil Engineers*. Thomas Telford - ICE Virtual Library, pp. 325–362. Available at: https://doi.org/10.1680/IPEDS.1952.11259.

Yu, H. *et al.* (2017) 'Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks', *Sensors*, 17(7), p. 1501.

Appendix 1: TEN-T network description

1. Context

The **Trans-European Transport Network (TEN-T)** is a strategic infrastructure initiative by the European Union aimed at improving transport connectivity across member states by developing an integrated, efficient, and sustainable transport infrastructure. To support the development of this network, the European Commission established the **Connecting Europe Facility (CEF)**, a financial instrument designed to promote investment in key transport projects. In the transport sector, CEF plays a crucial role in funding and coordinating initiatives that enhance cross-border mobility and multimodal transport system through the seamless integration of different transport modes throughout the EU.

This project is focused on two of the nine TEN-T corridors: **Atlantic and Mediterranean, which involves Spain and the proposed pilots in cross-border regions with Portugal and France.** Atlantic and Mediterranean Corridors are focused on the transportation of people, goods or both, and have different levels of relevance within the network and specifically within the different countries involved. In Spain, these corridors cover the main urban areas and references point in the transport networks.

1.1. TEN-T Atlantic Corridor

The **Atlantic Corridor** connects the Iberian Peninsula with the northwestern regions of Europe. The complete network is highlighted in Figure 9. The major nodes and urban centres that includes are in Spain, Portugal, and France. This corridor though the railway component features (new high-speed rail links and parallel conventional lines), provide cross-border continuity between Lisbon, Madrid, Porto, Vigo, Paris, Strasbourg, Mannheim and Le Havre (Jaspers Joint Assistance, 2014; Mouter, 2021).



Figure 9. TENT-T Atlantic Corridor.

In Spain, this corridor includes main urban points around the country, such as, Madrid, Bilbao, Santiago de Compostela, San Sebastian, Santander, Gijon, Zaragoza, Badajoz, Cordoba and Sevilla. These nodes play a key role in the multimodal transport of people and goods in different modes (rail, road, ports, and airports) connecting major ports, urban centres, industrial zones, and logistics hubs. Related to Portugal, the main core in general is the Alentejo region which is the cross-border with Spain, and there are principal cities as Lisbon and Coimbra. Figure 10 is focused on these countries. The sections or principal connections in this network are (European Commission, 2025):

- Algeciras – Bobadilla – Madrid
- Sines / Lisboa – Madrid – Valladolid
- Lisboa – Aveiro – Leixões/Porto
- Aveiro – Valladolid – Vitoria – Bergara – Bilbao/Bordeaux – Paris – Le Havre/Metz – Mannheim/Strasbourg



Figure 10. TEN-T Atlantic Corridor in Spain and Portugal.

Specifically focusing on the Spanish nodes involved in this corridor, one main urban node two cross-border regions are part of the project: **Madrid, the Nouvelle-Aquitaine – Euskadi – Navarre Euroregion (France–Spain) and the Central Portugal Region (Portugal–Spain).** A brief description of each is provided below.

- **Madrid** is a strategic core urban node at the heart of Spain's transport system and a key junction for both the Atlantic and Mediterranean Corridors. As the capital city, it hosts major logistics hubs, intermodal terminals, and the largest airport in the Iberian Peninsula. This is one of the most important air hubs in Europe, offering extensive international connections and serving as a key gateway between Europe and the rest of the world. The central location of the city supports multimodal connections across Spain and to neighbouring countries. Madrid has two high-speed train stations (Atocha and Chamartín), 13 urban transport interchanges and 18 intermodal areas. All together supporting the multimodality within the city and the metropolitan area that connect the region with the TEN-T network. Figure 11 shows the TEN-T network in the region and the two main modes of transport present. Figure 11 shows the TEN-T network in the region.

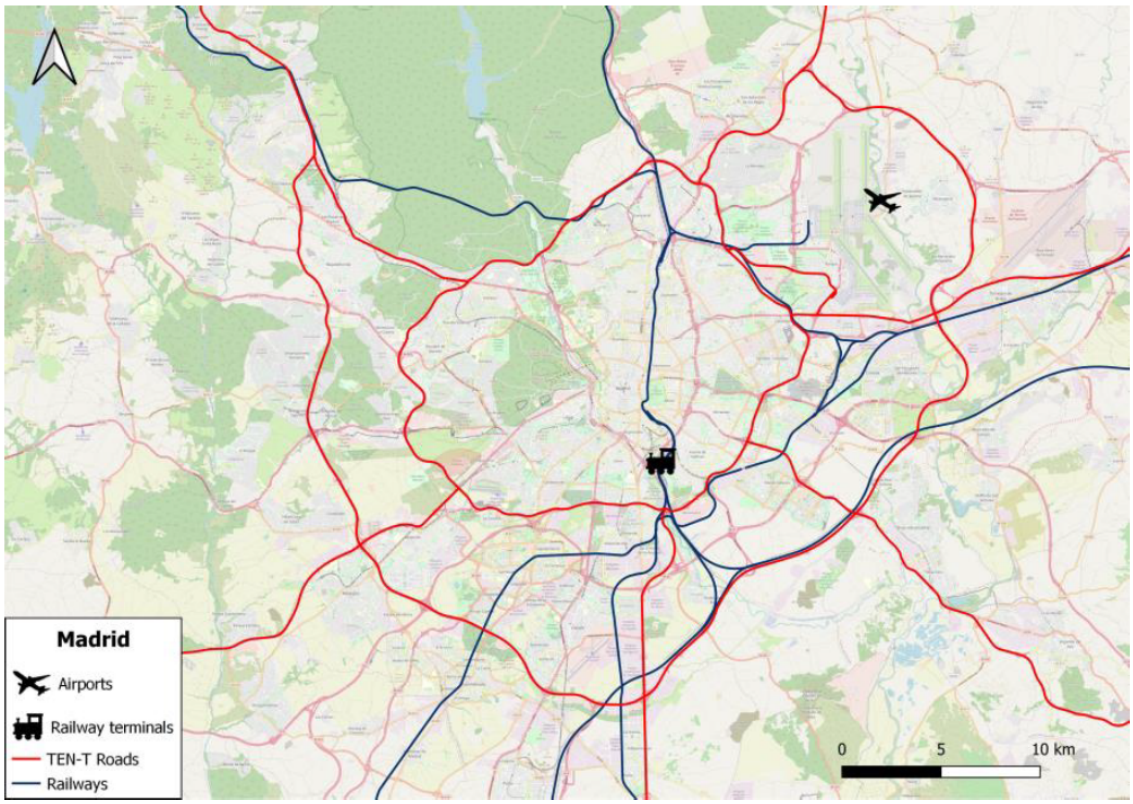


Figure 11. TEN-T Network in Madrid Region.

- **Nouvelle-Aquitaine – Euskadi – Navarre EuroRegion (France – Spain)** represents a dynamic transboundary area fostering strong economic, cultural, social, and infrastructural cooperation between southwestern France and northern Spain. It encompasses key urban centres such as Bordeaux, Bayonne, San Sebastián (Donostia), and Pamplona, combining a rich industrial base with significant passenger mobility and logistical assets. Related to rail services, High-speed rail (TGV) connects Bordeaux and Paris, with regional rail linking key cities. Freight corridors link the Atlantic ports of Bordeaux and Bayonne to inland logistics hubs. The region benefits from a dense highway network, including motorways that facilitating efficient freight and passenger road transport. Major Atlantic ports such as Bordeaux and Bayonne offer container terminals and logistics services, connecting regional economies with global maritime routes. Airports like Biarritz and Bilbao support passenger and cargo flights, strengthening regional accessibility.

- **Iberia Central West Region (Central Portugal Region and Central West Spanish Region) (Portugal – Spain)** is a strategic cross-border area for the cooperation between both countries. This transnational region covers the Portuguese districts of Coimbra, Castelo Branco, Leiria, most of Aveiro, Viseu, and Guarda, the northern part of Lisbon, and part of Santarém, as well as the Spanish Extremadura community. The region acts as a critical land bridge connecting Portugal's central regions to Spain and onward to European core markets. Its transport services support TEN-T's objectives by fostering cross-border interoperability, reducing territorial fragmentation, and promoting balanced regional development. This transnational region is served by a diverse mix of transport services. There is conventional rail lines connect cities like Coimbra, Leiria, and Castelo Branco with Extremadura. Some key highways in each

country and their counterparts in the other provide vital road freight and passenger links across the border. Although predominantly inland, proximity to Atlantic ports like Lisbon and Sines enhances export-import flows through multimodal logistics chains. Lastly, there is a relevant airport in Lisbon.

The Atlantic Corridor has strong multimodal dimensions, utilising rail, road, inland waterway and maritime routes. The Corridor entails a railway network of more than 13,500 km, over 5,700 km of roads and 775 km of inland waterways in four river basins. It interconnects Multimodal Freight Terminals (MFTs) located in 25 ports (of which 5 maritime/ inland and 5 inland waterway ports) with those 27 rail-road terminals along the Corridor. Key projects for the Corridor in this area include the Y Basque high-speed railway project and its connection to France with the Grand Project du Sud Ouest (GPSO) and the new high-speed rail connection between Madrid and Lisbon ('Four Stages Transportation / Land Use Model | The Geography of Transport Systems', 2018; *Travel Demand Modelling | Australian Transport Assessment and Planning*, no date; European Commission, 2025).

The key multimodal nodes in this corridor considered in the project, in the above descripted areas are:

- **Railroad terminals:** Madrid, Vitoria and Bilbao (ES).
- **Ports:** Lisbon/Setúbal (PT), Leixões–Porto (PT).
- **Airports:** Madrid (1), Lisbon (1).

1.2. TEN-T Mediterranean Corridor

The Mediterranean corridor spans from the western Mediterranean coast to the Hungarian-Ukrainian border. It runs broadly east-west axes, since it connects Europe from West to East as shows Figure 12. It is a major freight axis, especially for goods entering Europe via Mediterranean ports. Spanning from Spain to Hungary, passing through France, Italy, Slovenia and Croatia, six countries are involved with the addition of Ukraine, that represented the inclusion of 267 km of railway and 258 km of road network (European Commission, 2024).

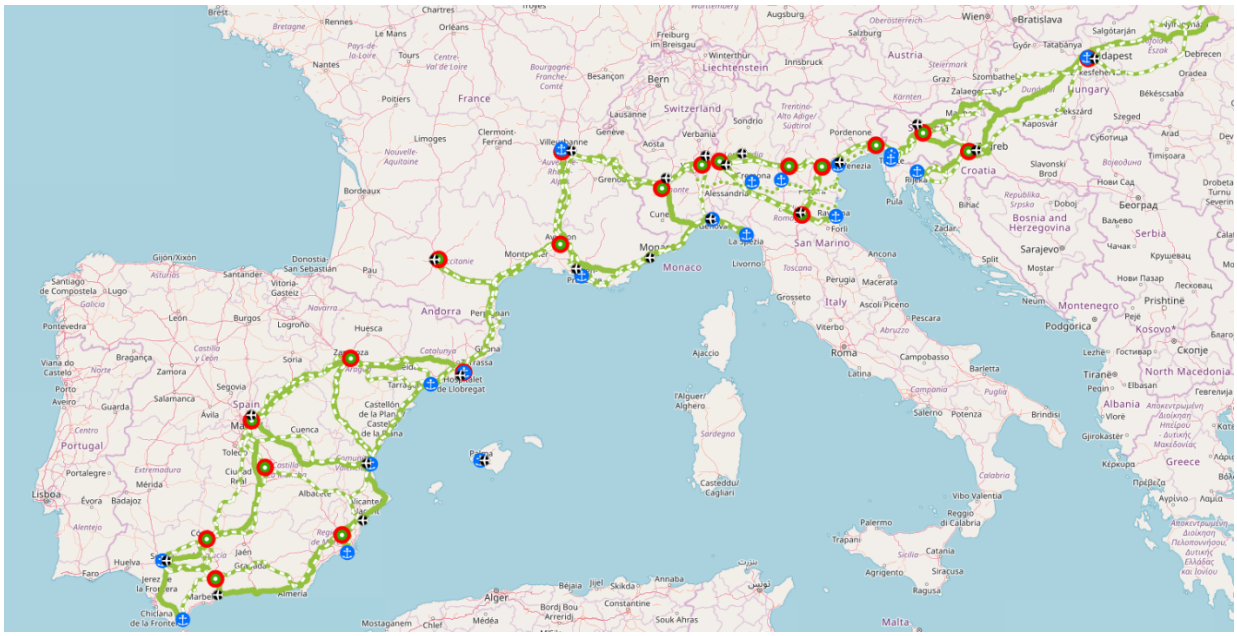


Figure 12. TEN-T Mediterranean Corridor.

In Spain, as can observed in Figure 13, the main urban nodes are in Barcelona, Zaragoza, Tarragona, Madrid, Alicante, Valencia, Málaga, Almería, and Algeciras. Similar to the Atlantic, for this corridor the links can be described as:

- Algeciras – Bobadilla –Madrid – Zaragoza – Tarragona
- Sevilla – Bobadilla – Murcia
- Cartagena – Murcia – Valencia – Tarragona
- Tarragona – Barcelona – Perpignan – Marseille/Lyon – Torino – Novara – Milano – Verona – Padova – Venezia – Ravenna/Trieste/Koper - Ljubljana – Budapest
- Ljubljana/Rijeka – Zagreb – Budapest – UA border

The Mediterranean Corridor encompasses over 10,000 km of rail connections (core network) and more than 7,000 km of road networks. It features 21 airports, 22 ports, and 26 Rail-Road Terminals (RRT) (Jaspers Joint Assistance, 2014; 'Four Stages Transportation / Land Use Model | The Geography of Transport Systems', 2018; *Travel Demand Modelling | Australian Transport Assessment and Planning*, no date; European Commission, 2024)



Figure 13. TEN-T Mediterranean Corridor – Spain.

Madrid, Barcelona y Valencia are the urban areas involved in the pilots proposed in this project.

- **Madrid** plays a crucial role in the project development since is the linked point between the two corridors as show Figure 14.

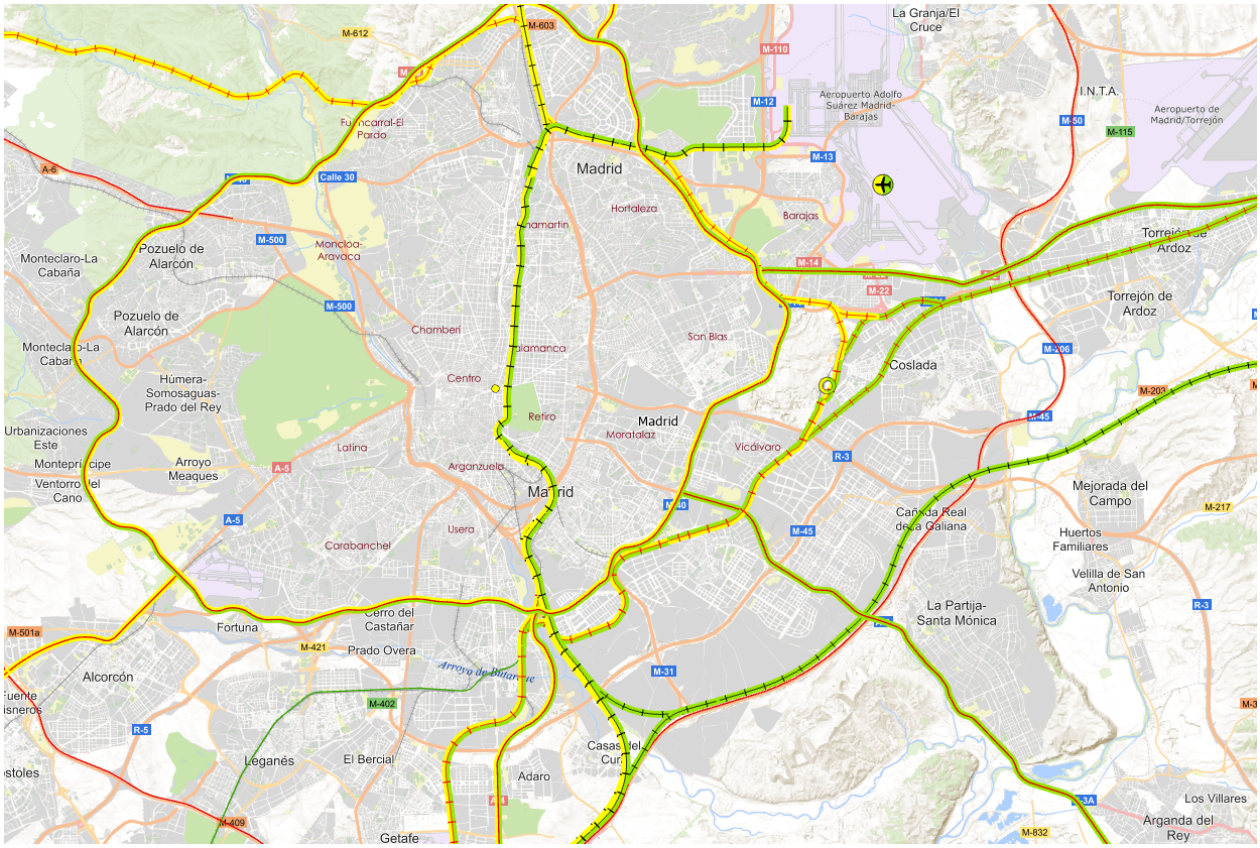


Figure 14. Madrid TEN-T network for Atlantic and Mediterranean Corridors.

- **Barcelona** is a major Mediterranean port city as well as essential node in the Mediterranean Corridor. Having the Port of Barcelona, one of the busiest ports in Europe. This port is a special case since it uniquely integrates sea, rail, road and air, functioning as a true multimodal hub. This port plays a vital role in supporting sustainable freight flows and facilitating trade between Europe, Asia, and North Africa. Barcelona is also served by the El Prat International Airport, one of the busiest airports in Southern Europe (Autoridad Portuaria de Barcelona (APB), 2025). The city has one high-speed train stations and 32 urban interchanges that connect the different transportation systems in the territory. Figure 15 shows the TEN-T network in the Catalonia Region. As can be observed, Barcelona concentrates the mentioned three different modes of transport.

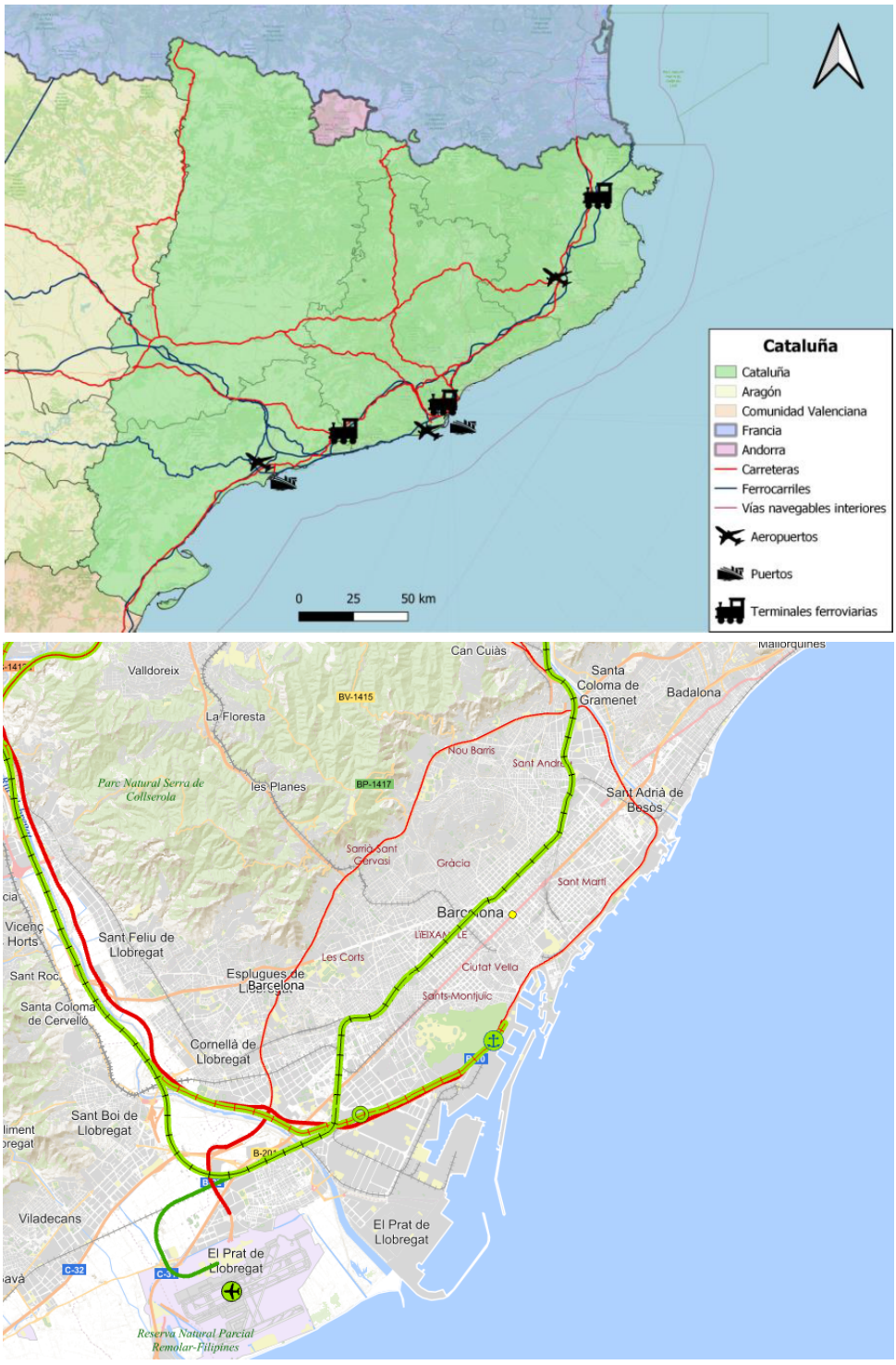


Figure 15. TEN-T Network in Catalonia Region and Barcelona city.

- **Valencia** port acts as a multimodal hub where maritime, rail, and road networks converge, facilitating the smooth transit of goods and passengers across the eastern coast, serving as a vital gateway for maritime trade between Europe, North Africa, and Asia. The city is one of Spain's primary logistical and transport hubs for import-export activities, especially for containerized goods. The Port of Valencia is Spain's second busiest container port and ranks among the top in the Mediterranean, handling millions of tons of cargo annually. Furthermore,

similar to Madrid and Barcelona, the city actively is pursuing sustainable urban mobility plans and innovations to reduce environmental impacts, aligning with the EU's Green Deal objectives (Empresa Municipal de Transportes de Valencia S.A.U, 2025)



Figure 16. TENT Network in Valencia Region and Valencia city.

Appendix 2: Travel demand modelling phases

2.1. Demand Generation Models

Trip generation models develop different estimation procedures for the production and attraction of trips in the various transport zones into which the study area is divided.

- **Trip production** is related to the socio-economic characteristics of residents, such as income level, housing structure, household size, motorization rate, residential density, accessibility, age, lifestyle, etc.
- **Trip attraction** in zones is associated with land use, types and quantities of employment, availability of school places, etc.

Trip generation models are generally classified into:

- Linear regression models, which are further divided into:
  - Zonal-level models (aggregated)
  - Household-level models (disaggregated)
- Category analysis models

Household-level trip generation models typically use variables such as income level, motorization rate, household structure (e.g., number of members working or studying), and household size. Other independent variables are more commonly used in zonal-level generation studies.

The predicted production and attraction of trips for each zone becomes an input for the trip distribution model, which then creates a mobility demand matrix between origin-destination pairs.

2.2. Demand Distribution Models

The demand distribution problem assumes knowledge of the total trips produced and attracted in each transport zone and addresses the question of how the trips produced in a given zone are distributed among the possible destinations.

The determination of the destination chosen for a trip is usually done without modelling the route used for access[1], simply associating to each destination *j* an attraction measure of the zone that reflects the opportunity of activities it offers, but incorporating the travel cost factor between zones *(i,j)*, so that the distribution attempts to satisfy the following two objectives: *minimize* the travel cost to a destination with the *highest* attraction offer.

1 Models that consider demand distribution simultaneously with the routes used for travel are called *combined distribution/assignment models*.

The functions that determine the number of trips originating from *i* and destined for *j*, based on the travel cost *(i,j)*, are called *demand distribution functions*.

The simplest demand distribution models (called *simple constraint models*) consider only the production totals at origins as constraints to satisfy and use a *multinomial logit function* to represent the demand for each destination.

Demand distribution models with *double constraints* determine the total trips d_{ij} such that they satisfy both the total number of trips produced by the origin zones and the total number of trips attracted to each destination. It is reasonable to assume that the demand between zones *(i,j)* should be proportional to the total number of trips originating in zone *i*, P_i, and the total trips attracted by destination zone *j*, A_j, but also to the generalized travel cost between both zones, t_{ij}.

The gravity model assumes that the number of trips between any two transport zones follows a law similar to the law of gravitation between two bodies. The number of origin-destination (O-D) trips is assumed to be proportional to the total trips produced in the origin zone and the total trips attracted to the destination zone (this is analogous to the masses of the two bodies), and inversely proportional to the square of the travel cost (usually travel time) between the origin and destination zones (analogous to the square of the distance between the two bodies in the classical physical model).

Entropy models are characterized by the exponential decline in the demand relationship (i,j) as travel cost increases. They assume that the distribution of travel demand, *dij*, besides satisfying the constraints on total productions and attractions in each zone, results from the aggregation of individual origin-destination selection decisions. There may be many combinations of individual decisions (states) that result in the same demand distribution pattern *dij*. The basic hypothesis is that all these states are equally probable in a given distribution, and therefore the probability associated with a distribution *dij* is proportional to the number of states that can produce it: the set of trips *(i,j)*, *dij*, most likely to occur in reality is the one with the highest number of possible states. The entropy demand distribution model seeks the most probable travel demand pattern *dij* that satisfies the total trip production and attraction constraints across zones.

Often, when a mobility matrix from a previous study is available and a short-term demand forecast is needed, along with aggregated information on expected growth in trip productions/attractions per zone, simplified forecasting methods can be used. These do not require the calibration of demand distribution functions and have been empirically proven to provide reliable short-term estimates. These methods are called growth factor methods.

Growth factor methods are classified into three major groups:

- Uniform growth factor methods.
- Single-constraint growth factor methods.
- Double-constraint growth factor methods.

Uniform growth factor methods are the simplest and, in contrast, the ones that produce the greatest inaccuracies in forecasts. If the only information available for making the forecast is the overall growth factor for the entire study area, denoted by τ, the base matrix will simply be multiplied by this coefficient, yielding the forecast mobility matrix for the target year.

Single-constraint growth factor methods apply when information is available on the expected growth in productions (or attractions) of the zones. In this case, if τ_i denotes the expected growth in productions for zone *i*, an update of the base matrix can be performed by multiplying each row (or column) *i* by its corresponding growth factor τ_i.

Double-constraint growth factor methods assume that information is available on the expected growth of both productions and attractions for each zone. This implies different growth rates for both the row and column components associated with transportation zone *i*, leading historically to the development of numerous iterative methods to obtain the future forecast matrix that satisfies both sets of growth factors (by row and by column) simultaneously. All these methods involve the computation of an intermediate set of correction coefficients applied to the rows or columns of the matrix as appropriate. The process concludes when these correction factors result in a future mobility matrix that satisfies both production and attraction growth factors.

The best-known matrix balancing algorithm (growth with double constraints) was proposed by (Furness, 1965) and is also known as the biproportional algorithm. Its basic convergence requirement is that the new row and column totals (determined by the respective growth factors) must coincide.

2.3. Modal Split Models of Demand

Modal split models are probably the most important within the realm of Transportation Planning models. This is due to the role played by the public/private transport dichotomy in maintaining acceptable congestion levels in major Western cities.

The factors that influence the choice of travel mode can be classified into several groups:

- Traveler characteristics:
  - Availability of a private vehicle.
  - Household structure.
  - Income level.
  - Population density.
- Trip characteristics:
  - Purpose of the trip (work-related trips are more likely to use public transport).
  - Time of day the trip is made.
- Transportation system characteristics:
  - Travel time (including waiting time) for each mode.
  - Monetary cost of the trip in each mode.
  - Availability and cost of parking for private transport.
  - Comfort and convenience of each mode.
  - Safety and reliability of the service.

In most cases, modal split models have been applied after trip distribution, considering one or two trip characteristics as significant variables (usually in-vehicle travel time), and requiring the calibration of an exponential-like curve that represents the share of trips by each mode, based on the difference in travel cost (mainly time).

2.4. Demand Assignment Models

Demand assignment models, also known as *transport network simulation models*, require a systematic representation of the related elements. In most studies, it becomes essential to digitize the data and process it using mathematical programming software developed *ad hoc* for the specific type of problem.

The fundamental elements involved in describing planning problems on transport networks are:

- The transport supply, represented by a *directed graph*. The basic components of a transport graph are:
  - *Nodes*: A set of points representing intersections (in private transport graphs) or stops of the transport lines included in the model (in public transport graphs).
    Nodes are classified into *centroids* and *base node*.
    Centroids model the center of gravity of each transport zone considered in the previous zoning of the geographical area defined by the mode.
  - *Arcs*: A set of segments that connect the nodes. These are usually *directional* and represent roads (in regional transport graphs) or streets (in urban transport graphs) linking intersections in private transport networks, or interstation segments in public transport networks. Arcs are classified into *connectors* and *normal arcs*; connectors are the segments that link centroids with base nodes.
  - *Penalized turns*: A set of turning movements that are either not allowed or are penalized in some way (e.g., delay or cost).

- The transport demand, which represents the *flow of trips (or vehicles)* between transport zone pairs (according to the zonification of the area) over a specific period (average day, morning peak period, etc.) and by the modes of transport included in the graph model (e.g., in public transport models: bus, metro, commuter rail, etc.). The link with the graph model is established through the nodes called *centroids*.
  This modal mobility is usually represented in the form of *origin-destination matrices (O/D matrices)*.

The arcs in the transport graphs typically have an associated "impedance", depending on the type of network. This impedance may represent distance, travel time on the arc or generalized cost of using the arc. Defining this impedance attribute allows for the determination of minimum-cost paths, based on that variable, between transport zones (or equivalently, O/D pairs).

The most frequently used impedance functions rely on travel time as the main component. This is due to its strong influence on route choice, as demonstrated by empirical studies, and because nearly all other

possible impedance measures are highly correlated with travel time (i.e., the traversal of an arc), though they are more difficult to calibrate.

Because of congestion, the travel time on roads or streets is an increasing function of traffic flow. Therefore, a behavioural function that relates traffic flow to travel time on each arc offers a more realistic modelling than assigning a constant travel time per arc.

The *travel time in the absence of traffic flow*, denoted by t_o, is known as the *free-flow travel time* of the arc. It represents the minimum travel time, as vehicles do not experience delays due to interaction with other vehicles. As traffic volume increases along an arc, the travel time also increases until reaching a certain level of service, commonly referred to as the capacity of the arc. Beyond this point, travel time increases exponentially. The functions that model this behaviour are commonly known in the literature as *flow-delay functions*.

The most widely used functions in Transportation Engineering that belong to this category are the BPR functions (Roads, 1964)**.**

The most commonly used transport network simulation models calculate the route assignment of travel demand (i.e., the O/D trip matrix) onto the transport graph; this is known as the demand assignment problem**.**

2.5. Macroscopic Assignment Models in Private Transport Networks

The equilibrium assignment problem in private transport networks (this term will be used from here onward) can be mathematically formulated as a mathematical programming problem, the exact solution of which yields the so-called equilibrium volumes, defined by (Wardrop, 1952) to satisfy the following hypotheses:

The assignment of the demand matrix to the transport graph determines, for each origin-destination pair, the use of minimum-cost paths (impedance, usually defined via flow-delay functions). That is,

- *All used paths p_{ij} between an O-D pair (i, j) (those assigned a portion of demand) have identical cost, and this is the minimum cost.*
- *All unused paths p^k_{ij} between the O-D pair (I,j) have a cost greater than the minimum.*

Solving the assignment problem exactly is often computationally expensive when dealing with networks that model real-world scenarios—both in terms of computer processing time, due to the high dimensionality of transport graphs, and because of the computational complexity inherent in solving the assignment problem itself. However, successful algorithms have been designed and implemented to solve the equilibrium assignment problem for large-scale models.

The ad hoc methods that solve the underlying mathematical program adapt non-linear optimization techniques from mathematical programming to the calculation of network flows. The applied algorithms have evolved chronologically in both efficiency and complexity:

- Frank and Wolfe.

- Conjugate gradient and PARTAN (Parallel Tangents) methods.
- Convex combination methods.
- Simplicial decomposition methods.

2.6. Public Transport Network Assignment Models

The problem of route choice and passenger assignment, where d_{ij} passengers wish to travel from zone *i* to zone *j*, in public transport networks typically involves greater algorithmic complexity than its counterpart in private transport networks.

Here are the key differences, from the assignment perspective, between both types of networks:

- Arc capacity in public transport networks is more related to the capacity of vehicles and their service frequency, rather than the physical capacity of the road segment. Some inter-station segments may be shared by several lines or modes (e.g., bus), while others are segregated (e.g., metro).
- Travel time between zones includes In-vehicle time, waiting time at stops, walking time to reach the stop, and transfer time between line sor nodes (if transfers exists).
- Monetary cost of trips is tied to the fare policy in the study area. Fare calculation may depend heavily on the chosen modal combinations (i.e., the trip strategy).

The task of finding optimal travel strategies from every origin to every destination is known as the public transport demand assignment problem. Strategically, the goal is to minimize total in-vehicle time and total waiting time. Mathematically, the problem is formulated as a linear program with nonlinear constraints.

One of the most difficult and characteristic aspects of public transport assignment is the common lines problem. This arises when some O-D pairs share route segments served by multiple transit lines. Unlike private vehicle users who choose a specific route, public transport users may wait at a stop and take the first arriving service among several. This type of route choice behaviour is more complex and requires detailed modelling. Additionally, the fare system adds further complexity to the model, especially when different routes or mode combinations imply different fare structures.