Deliverable 5.1

Enhancement of Transport Simulation Frameworks with Models of Emerging Mobility Solutions

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815069.

Table of Contents

Table index

Figure Index

Summary sheet

Deliverable No.	5.1
Project Acronym	Momentum
Full Title	Enhancement of Transport Simulation Frameworks with Models of Emerging Mobility Solutions
Grant Agreement No.	815069
Responsible Author(s)	Athina Tympakianaki (Aimsun) Jordi Casas (Aimsun) Evripidis Magkos (CERTH) Josep Maria Salanova Grau (CERTH) Georgia Ayfantopoulou (CERTH) Natalia-Maria Constantinidou (CERTH) Santhanakrishnan Narayanan (TUM) Constantinos Antoniou (TUM) Ignacio Martín (Nommon) Rodric Frederix (TML) Willem Himpe (TML) Antonio D. Masegosa (UDEUSTO)
Peer Review	Maria Chatziathanasiou (CERTH), Joren Vanherck (TML), Vishal Mahajan (TUM)
Quality Assurance Committee Review	General Assembly
Date	25/07/2021
Status	Draft
Dissemination level	Public
Abstract	The general objective of this document is to design the integration framework of the demand and supply models developed in WP4 into state-of-the-art transport simulation software to enable strategic planning and assessment of emerging mobility solutions. The main outcomes of this document are the following: (i) An integrated transport simulation framework, which satisfies the requirements to model and assess the different aspects of emerging mobility solutions, (ii) Recommendations and guidelines for calibration and validation of each integrated

	model, (iii) Benchmark advancement of the enhanced transport simulation software against its current solutions with respect to the strategic development of emerging mobility systems.
Version	Issue 1 Draft 3
Work Package No.	5
Work Package Title	Decision Support Toolset
Programme	Horizon 2020
Coordinator	EMT Madrid
Website	www.h2020-momentum.eu
Starting date	01/05/2019
Number of months	36

Project partners

Organisation	Country	Abbreviation
EMPRESA MUNICIPAL DE TRANSPORTE DE MADRID SA	Spain	EMT
NOMMON SOLUTIONS AND TECHNOLOGIES SL	Spain	NOMMON
DIMOS THESSALONIKIS	Greece	THESS
ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	Greece	CERTH
STAD LEUVEN	Belgium	LEUVEN
TRANSPORT & MOBILITY LEUVEN NV	Belgium	TML
STADT REGENSBURG	Germany	REGENSBURG
TECHNISCHE UNIVERSITAET MUENCHEN	Germany	TUM
AIMSUN SLU	Spain	AIMSUN SLU
POLIS – PROMOTION OF OPERATIONAL LINKS WITH INTEGRATED SERVICES, ASOCIATION INTERNATIONALE	Belgium	POLIS
UNION INTERNATIONALE DES TRANSPORTS PUBLICS	Belgium	UITP
UNIVERSIDAD DE LA IGLESIA CATÓLICA DE DESUTO ENTIDAD RELIGIOSA	Spain	UDEUSTO

Document history

Version	Date	Organisation	Main area of changes	Comments
Issue 1 Draft 1	05/07/2021	Aimsun	Initial version	
Issue 1 Draft 2	21/07/2021	Aimsun	All sections	Authors addressed comments from the consortium internal review
Issue 1 Draft 3	25/07/2021	Aimsun	Pages 15, 25, 33	Final version

List of acronyms

BS	Bike-sharing
CO	Carbon monoxide
CO2	Carbon dioxide
COPERT	Computer Programme to estimate Emissions from Road Transport
CS	Car Sharing
DARP	Dial-A-Ride Problem
DB	Data base
DRT	Demand Responsive Transport
DTA	Dynamic Traffic Assignment
DUE	Dynamic User Equilibrium
EU	European Union
Euro 0 – Euro 6	Euro Emission Standards
FCD	Floating Car Data
ICT	Information and Communication Technology
IPF	Iterative Proportional Fitting
IPU	Iterative Proportional Update
JDFs	Junction Delay functions
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
NOx	Nitrogen Oxides
OD	Origin-Destination
PM	Particulate matter
RS	Ride Sharing

SS	Scooter Sharing
TPF	Turn Penalty Function
VDF	Volume Delay Function
VOC	Volatile Organic Compounds

Executive Summary

The overall goal of the MOMENTUM project is to develop a set of mobility data analysis and exploitation methods, enhanced transport models and advanced decision support tools, which can capture the impact of emerging transport modes (such as shared mobility services) and ICT-driven behavioural changes on urban mobility. The developed methods aim to support local authorities in the task of designing the right policy mix, to exploit the full potential of emerging mobility solutions.

The general objective of Deliverable 5.1 is to integrate the demand and supply models developed in WP4 into state-of-the-art transport simulation frameworks to enable strategic planning and assessment of emerging mobility solutions.

The specific objectives for this Deliverable are:

- Summarise the advancement and opportunities of state-of-the-art transport simulation frameworks with respect to modelling the emerging mobility solutions, due to the integration of the developed models within the MOMENTUM project.
- Design the integration framework and identify the interaction workflow between the different models and algorithms.
- Provide recommendations and guidelines for the calibration and validation of the models in the integrated framework.
- Provide a technical assessment of the capabilities of the enhanced transport simulation framework after the integration against current solutions of state-of-the-art transport simulation software with respect to modelling emerging transport systems.

The main outcomes of Deliverable 5.1 are the following:

- An integrated transport simulation framework, consisting of the models developed in W4 in previous stages of the project that combine principles of agent-based modelling approach and traditional strategic transport models. The integrated framework provides the opportunity for cities to evaluate and integrate shared mobility systems and design long-term planning strategies.
- Provides a detailed description of the interaction workflow between the different models as well as information regarding the type of input and output data for each model, programming language used, and software requirements.
- Data collection requirements, recommendation processes for calibration and validation are reported for each model, aiming to provide general guidelines for the application of the enhanced transport simulation tool.
- The enhanced capabilities of the state-of-the-art traffic simulation software Aimsun Next, after the integration of the models into the software, are identified and compared against its current solutions.

The integrated framework presented in this Deliverable will be applied to the four case study cities of the MOMENTUM project in WP6 in future stages of the project.

1. Introduction

1.1. Scope and objective

The overall goal of the MOMENTUM project is to develop a set of mobility data analysis and exploitation methods, enhanced transport models and advanced decision support tools, which can capture the impact of emerging transport modes (such as shared mobility services) and ICT-driven behavioural changes on urban mobility. The developed methods aim to support local authorities in the task of designing the right policy mix, to exploit the full potential of emerging mobility solutions.

To achieve this general goal a set of demand and supply models and algorithms for modelling different aspects of the emerging mobility concepts and solutions have been developed in and reported in Deliverable 4.1 "Transport modelling approaches for emerging mobility solutions: supply and demand models".

Following up on the description of the demand and supply models, discussed in Deliverable 4.1, the present deliverable presents the integration workflow of the new theoretical models and algorithms into state-of-the-art transport simulation frameworks, providing them with essential capabilities to reproduce multi-dimensional impacts due to the introduction of new transport mobility technology and solutions. Deliverable 5.1 further provides guidelines on data requirements, calibration, and validation techniques of the integrated models and algorithms that can be adopted for the implementation of the enhanced framework. A technical assessment of the enhanced transport simulation software capabilities in relation to modelling emerging mobility solutions is provided as well.

The proposed integrated modelling framework will be applied to the four MOMENTUM case studies (Thessaloniki, Madrid, Leuven, Regensburg) described in Deliverable 2.2 "Specification of the MOMENTUM Test Cases". The results from the deployment and evaluation of various emerging mobility scenarios, in the context of the MOMENTUM project, will be provided in subsequent deliverables.

1.2. Structure of the document

The remainder of this report consists of four sections:

Section 2 describes the integration process of the demand and supply models developed in the project into state-of-the-art transport simulation frameworks, aiming to enable the strategic planning and evaluation of emerging mobility systems. The section also highlights the enhanced capabilities of traditional strategic transport models that can be achieved by adopting principles of agent-based models within the traditional strategic approach though the integration of the developed intermediate modelling framework in the MOMENTUM project. Moreover, the integration workflow and data requirements are provided for each model.

Section 3 provides general recommendations and guidelines with respect to the calibration and validation of each model as well as the identified data requirements. The presented techniques and processes can be adopted in order to enable the integration of the models developed into state-of-the-art transport simulation frameworks.

Section 4 focuses on a technical assessment of the advancements of state-of-the-art transport simulation software after the integration with the developed modelling framework, in terms of their ability in modelling and assessing the implications of emerging transport mobility systems. The traffic simulation software Aimsun Next is used as a concrete example to demonstrate its enriched capabilities against current solutions with respect to its ability to model and assess emerging transport mobility systems and solutions.

Finally, Section 5 summarises the main conclusions and outcomes of this deliverable.

1.3. Applicable documents

[I] MOMENTUM D1.2 Data Management Plan and Open Data Policy, November 2019

[II] MOMENTUM D2.2 Specification of MOMENTUM Test Cases, February 2020

[III] MOMENTUM D3.1 Data Inventory and Quality Assessment, March 2020

[IV] MOMENTUM D3.2 MOMENTUM Data Repository, June 2020

[V] MOMENTUM D3.3 Methodologies and Algorithms for Mobility Data Analysis, December 2020

[VI] MOMENTUM D4.1 Transport Modelling Approaches for Emerging Mobility Solutions, May 2021

[VII] MOMENTUM D4.2 Open Repository of Demand and Supply Models and Algorithms for Emerging Mobility Solutions, May 2021

2. Extended transport simulation framework for emerging mobility solutions

2.1. Overview of enhanced modules

This chapter focuses on presenting the new capabilities that the developed methods bring into the existing transport simulation models as well as the integration procedure between the different modules. Existing state-of-the-art strategic transport models do not have the essential capacity to model emerging mobility solutions for strategic planning applications due to the absence of agent-based principles within the strategic four-step modelling approach. Traditional transport models (traffic assignment with modal choice) are not able to reflect the demand-supply interaction and provide realistic predictions of the network performance implications due to changes in the network supply with the introduction of new shared mobility infrastructure. On the other hand, activity-based models are able to model the demand and operational side of shared mobility systems at a disaggregate level and capture the demand-supply interactions. Nevertheless, they are computationally expensive and data-intensive. This gap between traditional strategic models and agent-based models is closed by the proposed extended four-step modelling framework, which was designed in WP4 and presented in Deliverable 4.1. The proposed framework is referred to as “Intermediate modelling approach”, as it combines principles of the disaggregate agent-based models and the strategic transport models.

The enhanced framework helps to overcome some of the limitations of traditional strategic models, mentioned above, with respect to modelling and assessing the different aspects of new mobility systems. This is achieved with the addition of modules for synthetic population generation, disaggregate mode choice and fleet management. Furthermore, modules are developed for the estimation of emissions, car ownership and induced demand. Cities have an increasing interest on such indicators in order to assess the impact of different policies towards sustainable mobility solutions. For a detailed description of the modelling schema as well as technical description of the developed methods we refer the reader to Deliverable 4.1 entitled "Transport modelling approaches for emerging mobility solutions: supply and demand models". The general modelling schema is presented on Figure 1 as a point of reference throughout this report. The framework consists of different modules, which are generally categorised into demand, supply, fleet management and sustainability (see Deliverable 4.1 for a detailed description).

The proposed framework enables efficient and consistent integration of the developed demand and supply models into state-of-the-art transport simulation frameworks in order to enhance their capabilities in performing strategic planning and evaluation of emerging shared mobility services.

Concerning the demand module, it includes the first three steps of the traditional four-step modelling approach (i.e., trip generation, distribution and mode choice). Usually, cities have a traditional strategic transport model available. However, these models are not suitable for shared mobility services, as the demand is represented by aggregate OD matrices for conventional modes, hence, an extended model that also includes demand for different shared mobility systems is required. Typically, due to insufficient or unavailability of data related to new shared mobility services, cities are unable to estimate their own mode choice model (including all conventional and shared mobility modes). Towards this aim, the integration of the synthetic population generation module with a disaggregate mode choice model provides a significant benefit to model shared mobility with an agent-based approach. As it was mentioned in Deliverable 4.1, the estimation of the disaggregate mode choice model for both conventional modes as a whole and the different shared mobility services is based on generic sociodemographic and trip characteristics of the anticipated demand for shared mobility systems. Subsequently, the model can be considered transferable to other cities, and hence, it can be integrated into transport simulation frameworks to enable strategic planning applications for shared mobility services. The main benefit of this approach is that the existing travel demand model can still be used, without the need to estimate a new mode choice model for conventional modes and shared mobility services. In particular, the estimated disaggregate demand for conventional modes as a whole can be aggregated and introduced into the existing mode choice model of a city to estimate the split between the individual conventional modes.

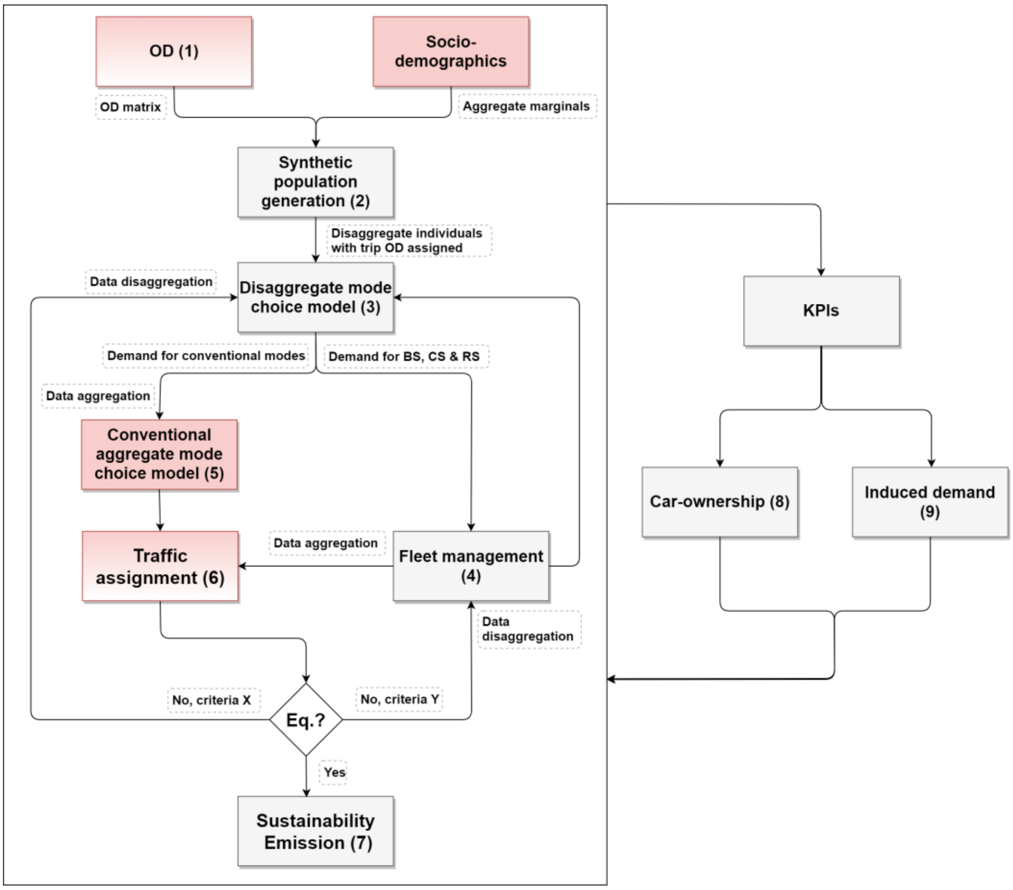

Figure 1. Intermediate modelling approach. The red colour shaded boxes indicate the unaltered existing components in the traditional four-step approach. The boxes with a red-white gradient are also already found in the traditional four-step approach, but improvements and alternative models are presented in Deliverable 4.1.

Typically, the available OD matrices that can be used as input to the modelling framework consist of aggregate counts of the total trips performed in each potential OD pair of the area of interest. OD matrices can be derived directly from travel surveys and interviews for a sample of travellers over the area of interest, however, this is costly and difficult. Therefore, they are usually inferred from traditional traffic measurements and new data sources, such as mobile phone data or GPS data. The methods for estimating OD matrices are dependent on historical data and previously estimated OD matrices. Since the actual OD matrices and their travel patterns are unknown, "representative" OD matrices based on averages on historical days can be used as input to the framework. To improve this input, the OD matrix selection module could be used, as it uses machine learning methods to select a more representative OD matrix (from available historical matrices) according to day-based attributes (such as holidays, day of week or forecasted weather). With this approach, more representative OD matrices can be derived from groups of matrices that are clustered according to their trip volume and structure. These OD matrices provide more reliable input to the synthetic population model.

Next, the final step in the traditional four-step modelling approach is the traffic assignment. The aggregate demand for conventional modes and for new mobility services (estimated by the disaggregate and conventional aggregate mode choice models) can be fed into an existing traffic assignment model, or any of the improved static and dynamic traffic assignment (DTA) models developed within the project. This allows to forecast the impacts of the new shared mobility services on the network performance.

As mentioned above, two enhanced traffic assignment models have been developed in the project, whose integration into state-of-the-art transport simulation software has several benefits in the context of strategic planning and evaluation of emerging mobility systems. In particular, the improved static assignment model (see Deliverable 4.1 for a detailed description) allows for a more accurate modelling of the new transport systems as intrazonal trips on local roads can be modelled (by considering each link as a potential origin or destination for users), which is ignored in traditional macroscopic network models based on centroids and connectors. The second traffic assignment model that is developed is a hybrid DTA modelling framework in which the route choice set is determined by dynamic path assignment for the whole large-scale network, while vehicles are simulated using dynamic mesoscopic network loading in specific areas of interest and macroscopic network loading in the remainder network. The developed model is beneficial for the large-scale networks used in strategic transport planning where the introduction of new mobility services requires realistic knowledge of traffic conditions and travel times for the service areas. The hybrid model offers improved run times when simulation of large-scale networks is needed, however, the analysis study area that requires detailed representation considers a smaller part of the network.

With regards to the modelling and evaluation of fleet operations for emerging shared mobility services, traditional transport models are unable to do so, considering the requirement to mimic the operations (i.e., serving the requests, trip plan optimisation, fleet re-allocations, etc.) of the shared system operators. Therefore, models with capabilities to represent the operations at a disaggregate level are needed for a more realistic representation of the shared mobility. The developed fleet management modelling approach consists of planning and operational models and a shared mobility service simulation platform. These for adequate modelling and evaluation of shared mobility services towards an agent-based model. Once the modal share is estimated using the disaggregate mode choice model, the disaggregate demand for the shared mobility services is fed into the fleet management algorithms, which assign shared vehicles and simulate their operations.

Finally, emission models have been developed as part of the sustainability module to estimate emission indicators due to the inclusion of new shared mobility systems in the network. In particular, a static and a dynamic emission models have been developed offering improved capabilities compared to the models in existing traffic models. The integration of the emission models into existing transport simulation frameworks provides the capability to estimate more accurate traffic emission indicators.

The proposed framework provides other modules that allow the estimation of further KPIs, such as car ownership or induced demand that are not available in existing strategic transport models. However, these models are needed for accurate planning and evaluation of the supply for shared mobility systems. They have been designed as separate modules in the framework in order to reduce model complexity and avoid convergence issues. Nonetheless, the outputs from the application of those models can be used to provide enhanced information to the transport simulation frameworks. Specifically, the developed aggregate and disaggregate car ownership models can be integrated into strategic transport models for updating the synthetic population with more accurate information regarding the car ownership at a zone or household level.

The introduction of new transport modes and services in a network (such as shared mobility systems) is expected to change the network supply, which consequently can lead to changes in the total demand. As mentioned above, one limitation of the state-of-the-art transport models is their inability to model the behaviour of travellers with respect to seeking for alternatives in terms of departure times, shifts in activity and activity locations, etc. Activity-based models consider these interactions; however, the computational cost is higher. The induced demand module implicitly accounts for these behavioural changes by creating a feedback loop to the demand side and calibrates this feedback loop based on available elasticity estimates (e.g., from the literature). This approach adds a significant value to transport models, as it can reflect the supply-demand interaction (i.e., demand elasticity with respect to a change in supply). The output of the aggregate induced demand model may be used to update the aggregate demand input for the synthetic population generation and traffic assignment model. Furthermore, a data-driven method has been developed, which is referred to as induced demand for roundtrip station-based car sharing model. It is meant to serve special trips (e.g., trips to furniture stores) for a small special case car sharing system.

2.2. Overview of integration and interaction workflow between modules

After the development of the separate modules, the next step was to integrate the enhanced framework into existing state-of-the-art transport simulation software. This will enable the modelling of travellers´ choices and behaviour to move from their origin to the desired destination at a given time, including the planning, monitoring and management of the vehicle fleet that supplies the travel requests. The objective is to provide important KPIs both at user- and system-wide level for the evaluation of the shared mobility services as well as their impacts.

The developed modules of the proposed intermediate modelling approach (Figure 1) were presented in the deliverables of WP4. In particular, Deliverable 4.1 “Transport Modelling Approaches for Emerging Mobility Solutions: Supply and Demand Models” described in detail the theory of the developed models and algorithms. In Deliverable 4.2, entitled “Open Repository of Demand and Supply Models and Algorithms for Emerging Mobility Solutions”, the technical characteristics of the developed solutions are described, which differ mutually in terms of algorithms as well as programming languages that have been used. As mentioned in Deliverable 4.2, not all the developed modules are open source or developed with the same technologies, hence, an automatic integration of those modules is not feasible. Additionally, some of these models are directly implemented into the transport modelling software Aimsun Next, so they need to be executed independently. of the different Another reason that prevents the full integration of all the modules into a single tool is that most of the models need to be calibrated with data from each specific city whose availability and quality may differ across different use cases. Nevertheless, the output from each of the models serves as valuable input to state-of-the-art simulation software in order to enable the evaluation of the strategic planning for shared mobility services.

Therefore, the integration framework presented in this deliverable refers to the construction of common interfaces between individual components of the different modules in the modelling schema, rather a single application that could be executed automatically. The integration ensures that the input and output workflow is suitably interfaced with the corresponding modules. Therefore, the output from one module can be used as input to the next module to be executed. A module may require inputs from more than one module.

Depending on the assumptions made with respect to the penetration of trips for shared mobility services, the workflow and interaction between different demand and supply models varies. When the demand penetration of shared mobility services is low, the hypothesis is that the network travel times will not change with the introduction of the new demand for shared mobility services. Specifically, if the demand for shared trips is below a pre-specified threshold, the network travel times obtained from a traffic assignment, or any other available data source, can be considered as fixed in the implementation and evaluation of the disaggregate mode choice model as well as the fleet operational algorithms and the service simulation platform which utilise travel time information.

The integration frameworks are presented in the following sections for the low and high demand penetration for shared mobility trips, respectively. Additionally, the workflow, interaction and interface between the different models is described. In Sections 2.3–2.6, the integration and interaction workflow are presented in more detail for each individual model, following the four modular approach categories that were defined in Deliverable 4.1, namely, demand, supply, fleet management and sustainability. A description of the input and output types, and interface between the models are provided as well.

2.2.1. Low demand penetration integration

The low demand penetration scenario assumes negligible or no change in travel times due to the introduction of demand for the new mobility services. In that case, historical travel times obtained from a previous static assignment model or based on the available travel time observations can be used. These travel times are considered fixed in the disaggregate mode choice model and in all fleet management scenarios.

Figure 2 illustrates the integration framework which follows the sequence of the main modelling schema in Figure 1. The different models interact with each other through information exchange. The figure also presents the workflow, level of interaction with other models and the type of data that is being exchanged.

In the first step of the sequence, the synthetic population module is deployed in order to create individual (disaggregate) trips to be used as input to the disaggregate mode choice model. Synthetic individuals are generated based on the aggregate or disaggregate OD demand matrices and socio-demographic data that should be first collected and prepared according to the requirements of the synthetic population method that is utilised within the project [step (1) in Figure 1]. Eventually, a base synthetic population is generated, and a destination choice is assigned to each synthetic individual [step (2) in Figure 1]. If a synthetic population already exists, the integration workflow concerns the application of the destination choice model and the enhancement of the synthetic population with missing attributes through statistical matching. This process is explained in detail in the following sections.

Subsequently, the synthetic population is provided as input to the disaggregate mode choice model [step (3) in Figure 1], along with other model parameters to estimate the mode choices for shared mobility services and conventional modes. Another necessary input for the mode choice model is the trip travel times between origin and destination zones that is assigned to each individual. The travel times can be provided as skim matrices to the mode choice model after performing a traffic assignment using the existing travel demand model of the city, which represents the demand for conventional modes. As indicated in Figure 2, the execution of the traffic assignment model is an initial condition, which defines the starting point for the integration of the framework. Following, the estimated disaggregate demand for the shared systems along with data and information related to the request characteristics, service and network supply are fed into the fleet management model [step (4) in Figure 1] to optimise the trip plan solutions and simulate the operations to serve the demand for the shared services. The pre-determined KPI´s and traffic metrics obtained from the fleet management model can be further provided as updated input to the disaggregate mode choice model to repeat the estimation of the modal shifts. Finally, post processing is carried out to calculate emissions [step (5) in Figure 1] by receiving output data from the traffic assignment model.

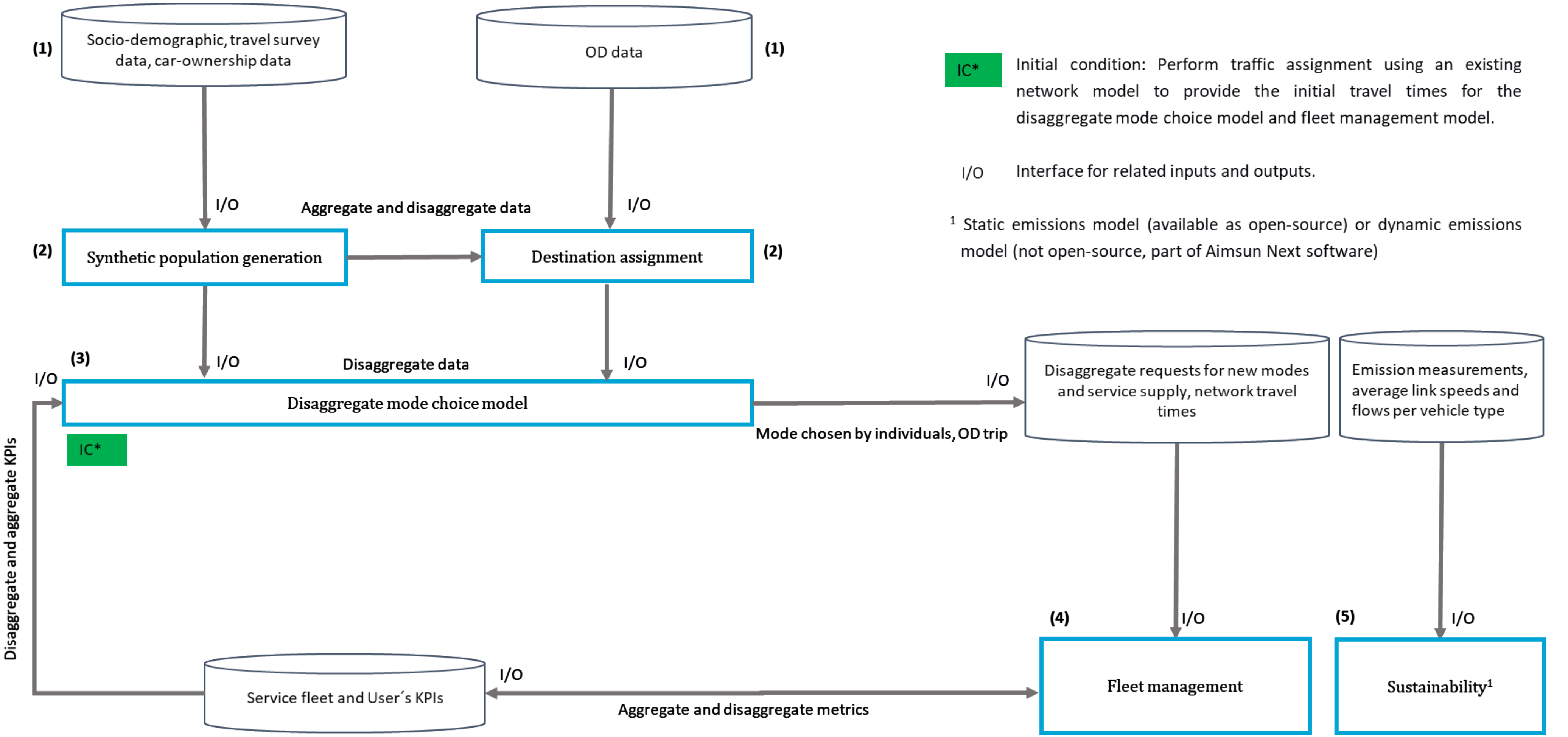

Figure 2. Integration framework for low demand penetration for shared mobility services.

The fleet management module is further extended to planning, operational and shared mobility service simulation modules. The integration and workflow between the fleet management components is presented in Figure 3 to provide a better understanding of the interactions and interface between the different models.

For a comprehensive representation of the fleet management module and integration with respect to the main framework, the traffic assignment and disaggregate mode choice modules are also presented in Figure 3. These modules provide important input for the fleet management module. Moreover, iterative interaction with these modules may be needed in the high demand scenario. The three models that constitute the fleet management module are the planning, operational and shared mobility service simulation platform Aimsun Ride. For more details about the models and algorithms we refer the reader to Deliverable 4.1. It should be noted that the simulation platform Aimsun Ride is designed as a plug-in inside the Aimsun Next traffic simulation software. Consequently, a suitable interface between Aimsun Ride, the external fleet planning and operational models is implemented to makes the interaction between the three models feasible.

The same initial condition as for the complete framework applies for the fleet management module as well. A traffic assignment needs to be performed to obtain the travel times that are required as input for the mode choice estimation and fleet management execution. As mentioned in the previous section, in the high penetration scenario travel times may need to be updated. Therefore, the traffic assignment needs to be repeated after completing each sequence and update the travel times.

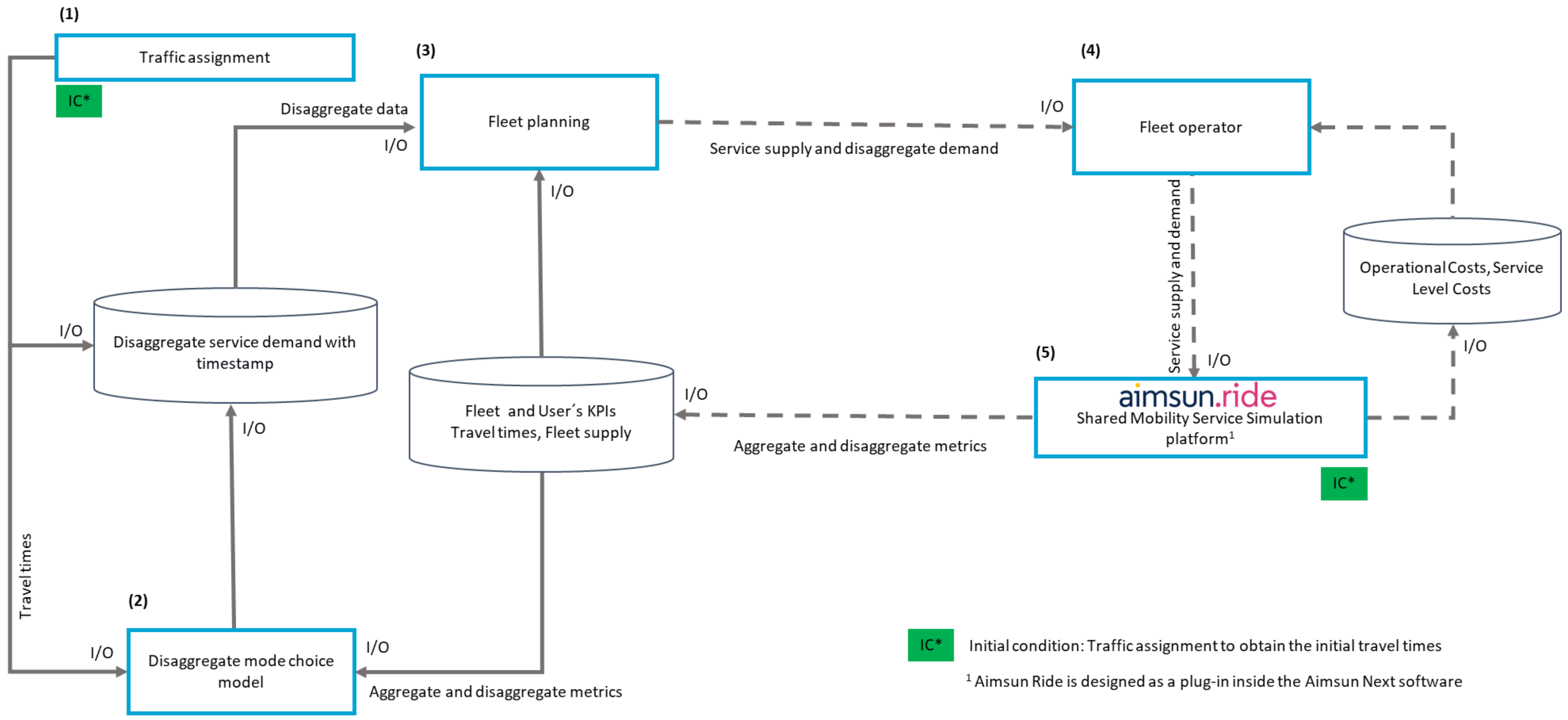

Figure 3. Integration between fleet management models.

2.2.2. High demand penetration integration

As discussed in Deliverable 4.1, depending on the traffic assignment results, it might be required to iterate the modelling sequence from mode choice or fleet management.

In the high demand penetration scenario, the network traffic conditions are expected to change due to the increased demand for the new services, which implies a decrease in the aggregate demand for conventional modes. The travel times need to be updated in the disaggregate mode choice model before proceeding with a new estimation of the modal splits. Moreover, the changed travel times need to be provided in the fleet management model as they constitute a core input in the optimisation of the fleet operations. The new travel times are obtained by performing a new traffic assignment after updating demand for conventional modes as a whole in the mode choice model of the existing travel demand model. Subsequently, the traffic measurements indicators obtained from the new traffic assignment should be also given as new input to the sustainability modules to re-estimate the emissions.

The integration workflow is modified accordingly from Figure 2 and illustrated in Figure 4. The main difference between the low and high demand penetration scenario workflow is the interface that is added between the traffic assignment model, the fleet management module as well as the mode choice model in order to introduce the altered demand and supply interactions mentioned above. The workflow between the fleet management models remains the same as for the low demand penetration case (Figure 3).

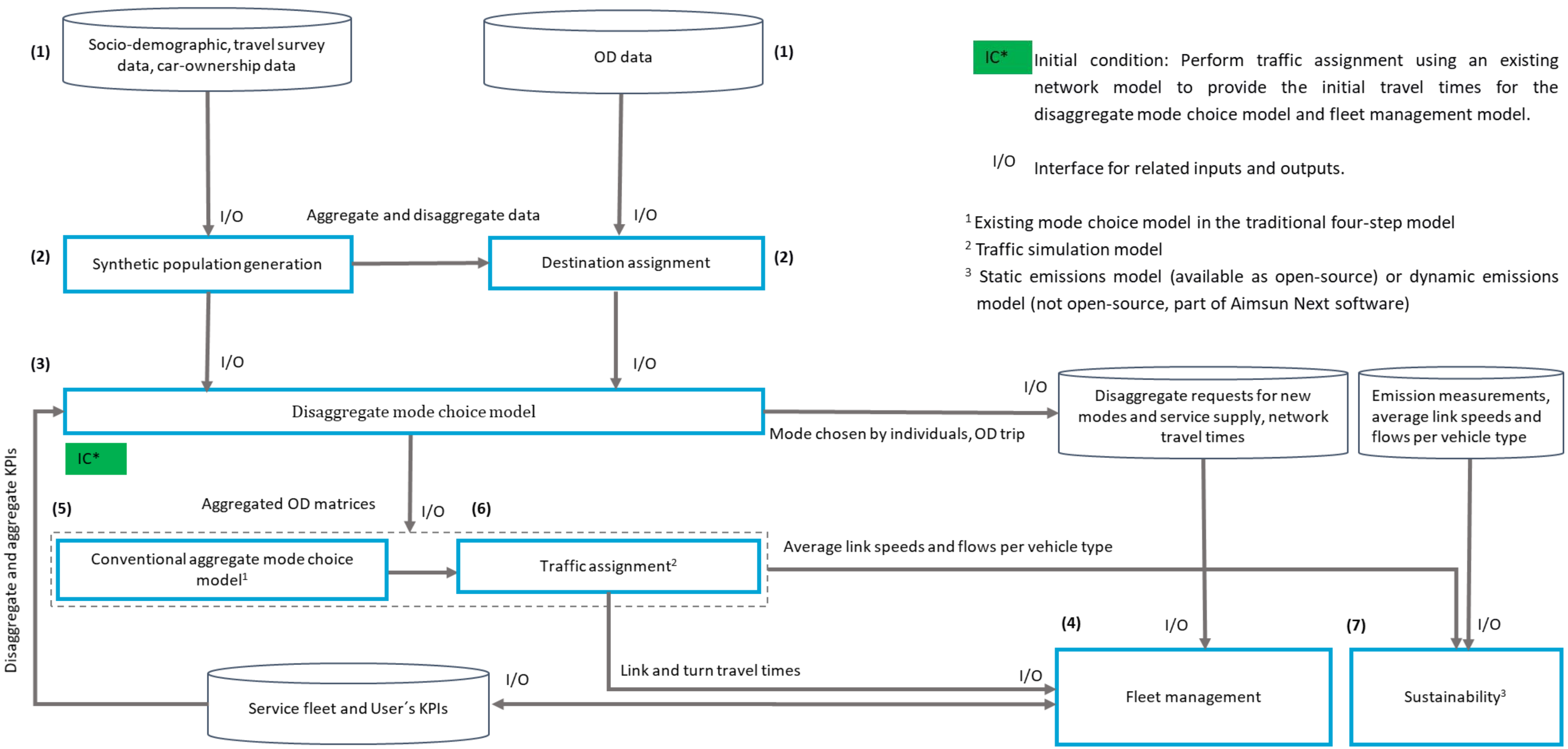

Figure 4. Integration framework for high demand penetration for shared mobility services.

2.2.3. Supplementary modules integration

Figure 2 and Figure 4 presented the integration into the transport simulation framework for the modules that belong to the main modelling schema for the low and high demand scenarios, respectively. Nevertheless, the integration is also extended to the modules for the estimation of car ownership, induced demand as well as models for the prediction of demand for new mobility services and classification of OD matrices. These models are separate from the main integration framework; however, they can be applied as post-processing steps. Following, the interface diagrams for these models are presented and described. The entire modelling sequence could be rerun with respect to the estimations from these models.

2.2.3.1 Integration of OD matrix selection and data driven share mobility estimation models

The OD matrix selection module consists of a model to classify and select a general mobility OD matrices accordingly to the type of day to be studied

Figure 5). The shared mobility demand prediction model provides a model for estimating the demand for new mobility services (Figure 6). Hence, OD matrices for conventional modes are obtained from subtracting the demand of shared mobility to the general mobility matrices. The output from these models can be integrated with the synthetic population module by providing the input OD matrices, obtained from data-driven approaches as an alternative to the traditional four-step modelling approach. Depending on the matrices to be predicted, they may include mode-specific information (shared mobility matrices, general OD matrices derived from surveys) and therefore could be used directly in fleet management module or may require disaggregation (synthetic population generation) and application of the mode choice model. For this reason, the demand prediction model (Figure 6) is not integrated into the main frameworks (Figure 2 and Figure 4), as feedback between fleet management/traffic assignment and shared mobility demand estimation is not possible due to the data-driven nature. However, these matrices can be used as input to the fleet management and traffic assignment modules. Since the current version of the data-driven shared mobility estimation module does not take into account travel time, this could be used only in the case of low penetration scenarios.

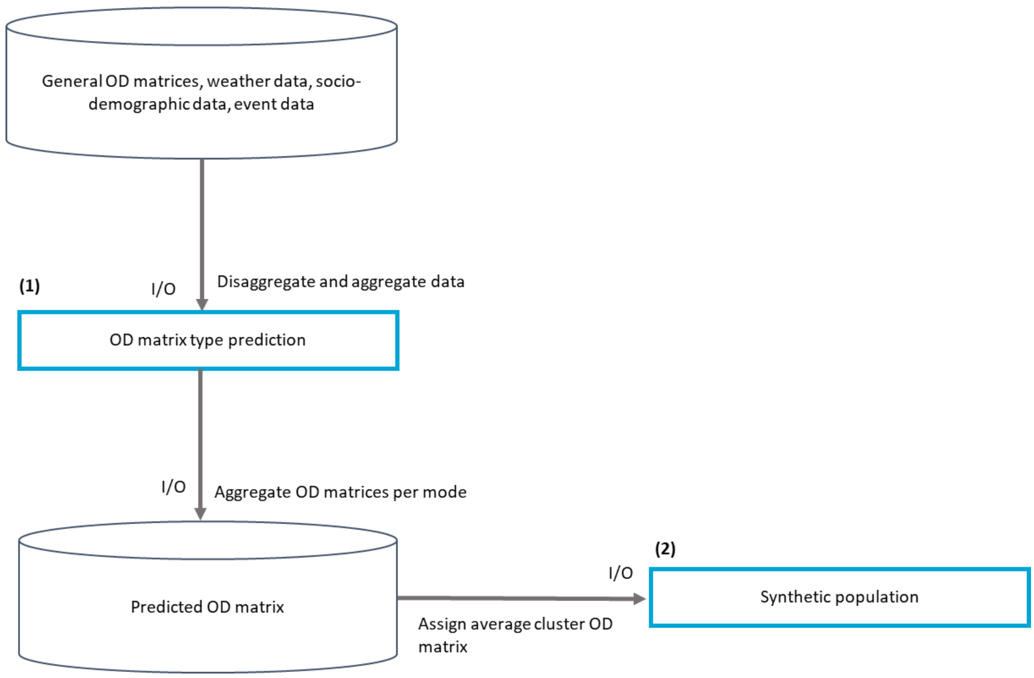

Figure 5. Integration workflow for OD matrix prediction.

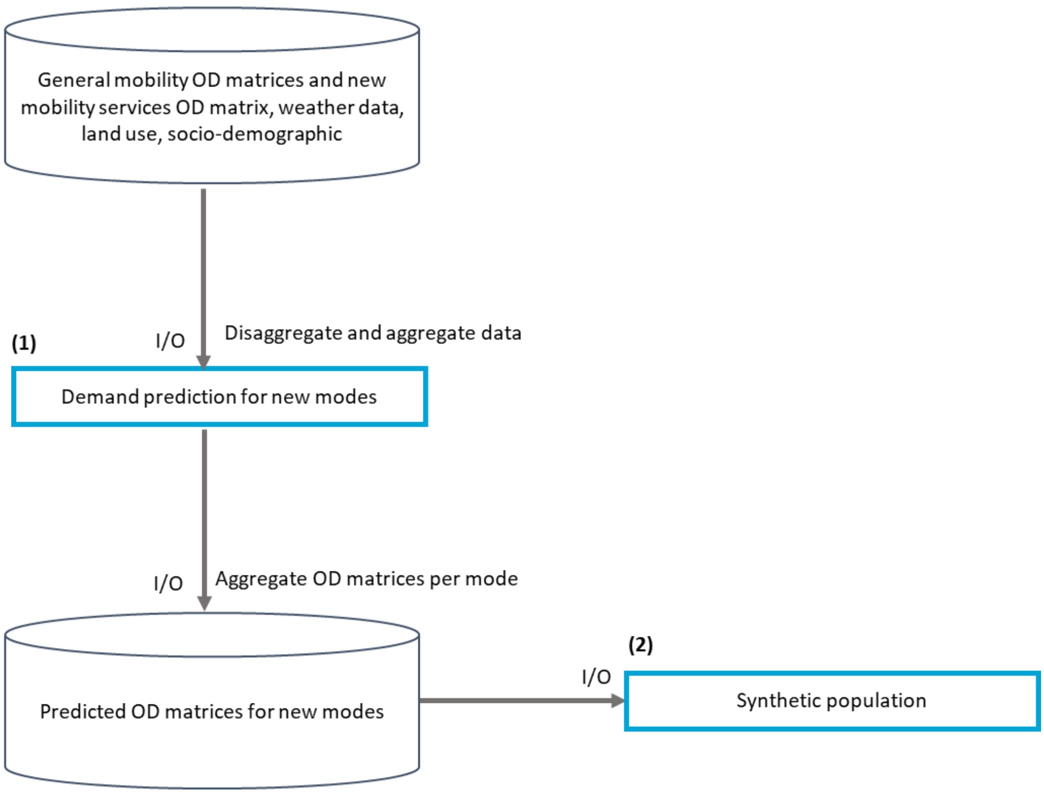

Figure 6. Integration workflow demand prediction for shared mobility services.

2.2.3.2 Integration of induced demand models

The induced demand models are also separated as aggregate (Figure 7) and disaggregate models (Figure 8), depending on the type of data they utilise. The output from the induced demand model that estimates demand elasticity (aggregate model), can be integrated with the synthetic population generation model to update the demand input. The disaggregate induced demand model was built as an external model in order to estimate the demand for special trips as well as the profile of such users, as those are attributes that are not covered in the traditional four-step models. The output from this model can be integrated with the fleet management module to simulate the demand for roundtrip station-based car sharing systems. Similar to the demand prediction model, a feedback between induced demand and OD generation step cannot be implemented.

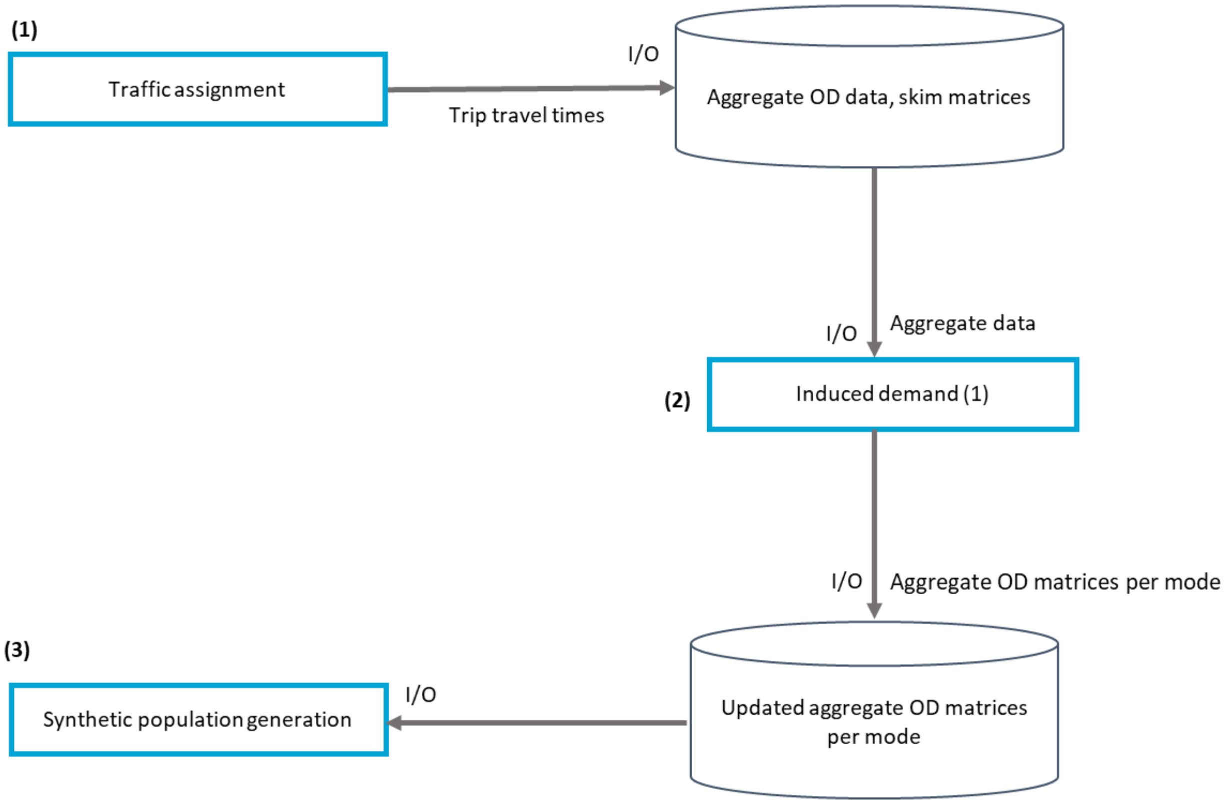

Figure 7. Integration workflow for aggregate induced demand model.

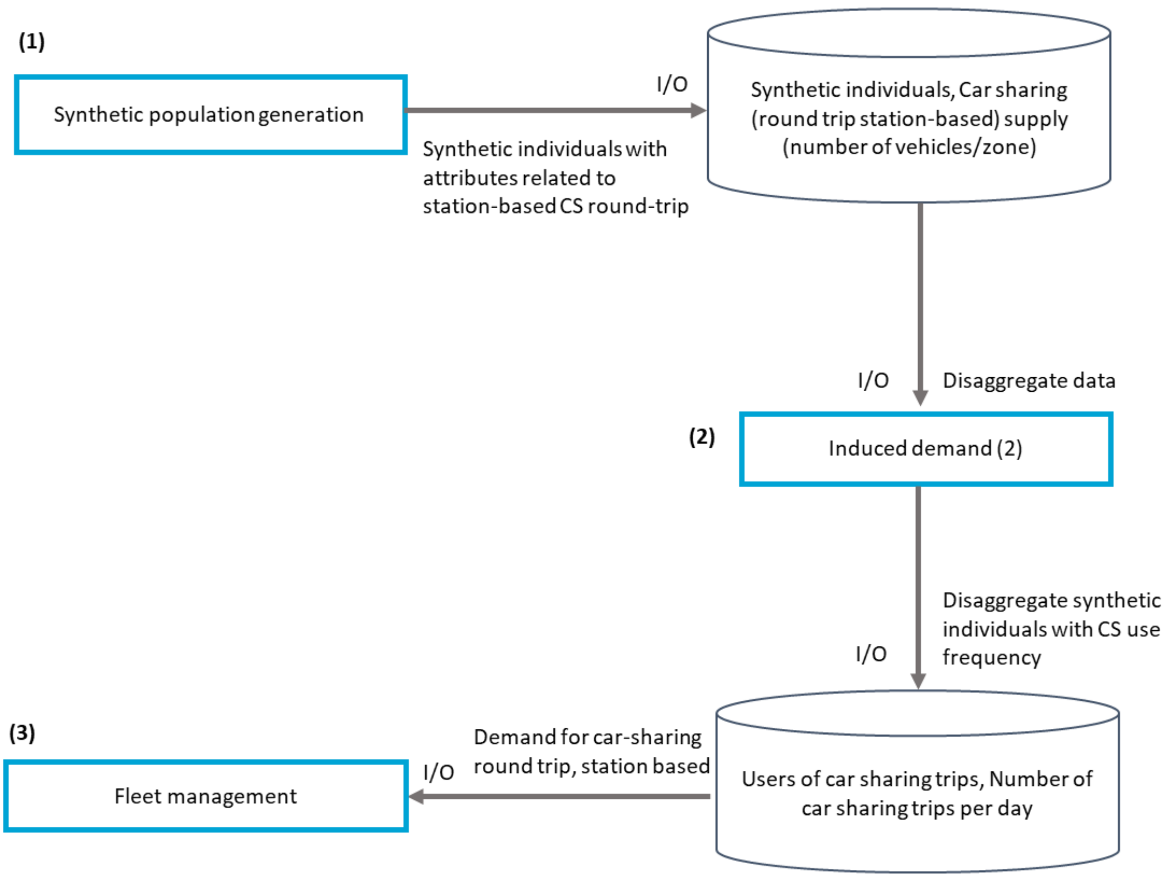

Figure 8. Integration workflow for disaggregate induced demand model.

2.2.3.3 Integration of car-ownership models

The two developed car-ownership models are categorised as aggregate (Figure 9) and disaggregate (Figure 10), depending on the associate type of inputs and outputs. The outputs from those models can be integrated into the synthetic population model to enhance the information related to car ownership of the synthetic individuals.

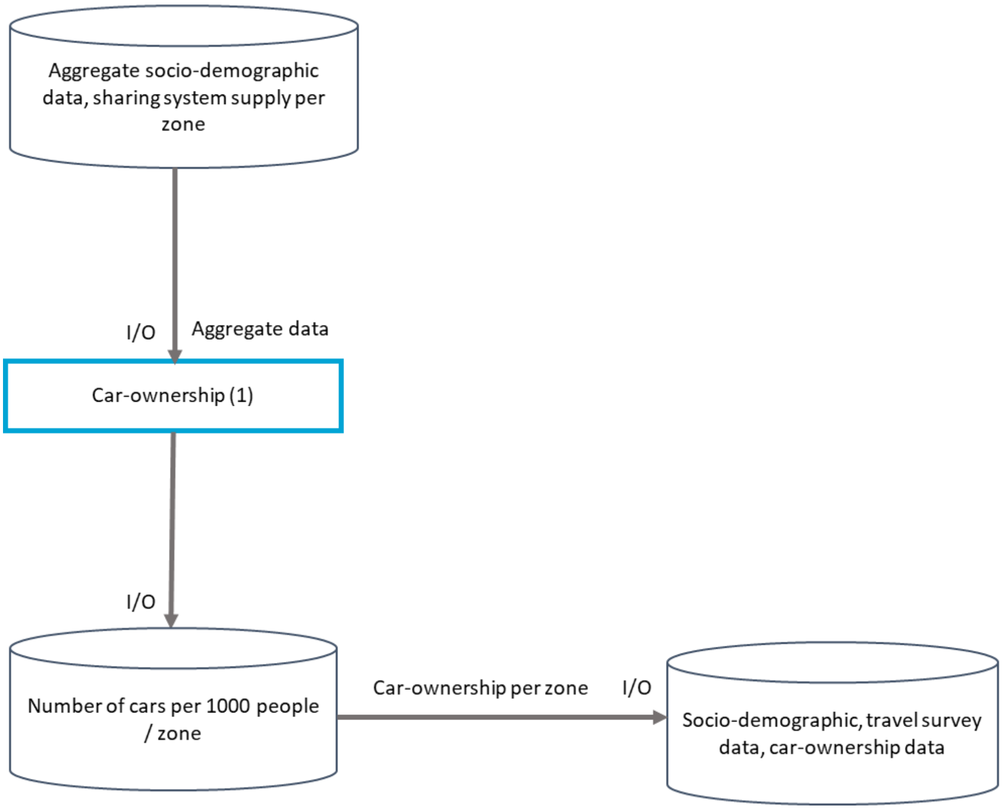

Figure 9. Integration workflow for aggregate car ownership model.

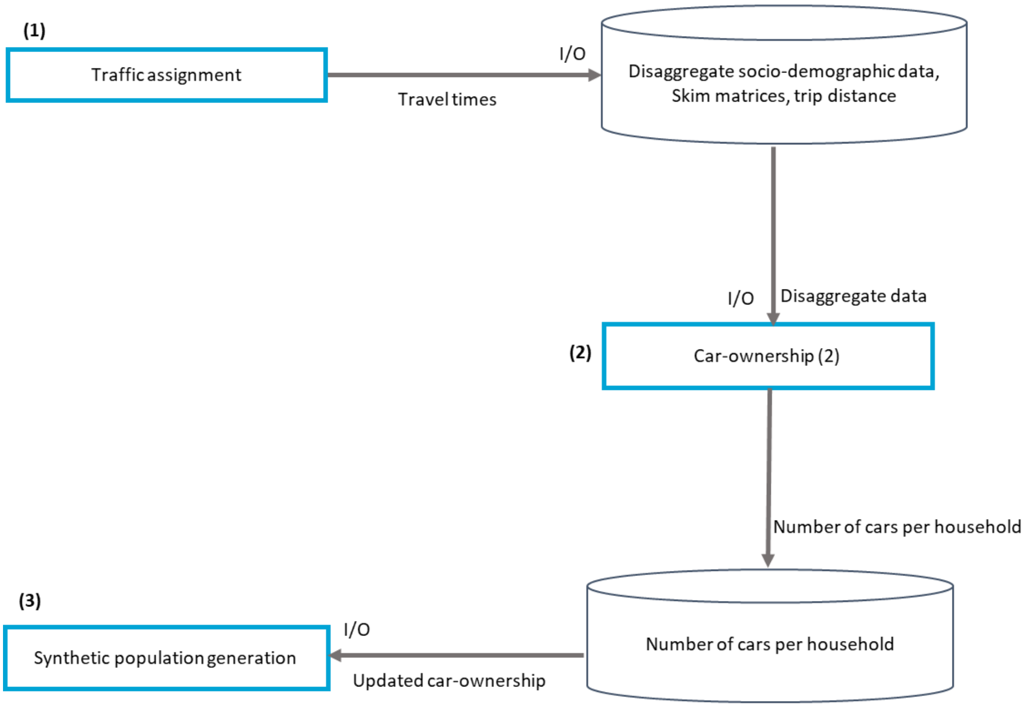

Figure 10. Integration workflow for disaggregate car ownership model.

2.3. Demand modules integration

2.3.1. Induced demand modules

2.3.1.1 Demand elasticity integration

The induced demand module, based on elasticities, is relevant in scenarios where travel times (or general costs) alter considerably with respect to a base scenario. It uses the base OD matrices and general costs/skim matrices from the traffic assignment step to calculate updated OD matrices. These updated OD matrices can be used as input for the other modules, such as the traffic assignment step.

Table 1 presents the data requirements for the implementation of the Induced demand module.

Table 1. Interface and data requirements for the induced demand module.

Module	Input		Output		Programming language	Software requirements
	Type	Format	Type	Format		
Induced demand	OD matrices (aggregate)	CSV	Updated OD matrices	CSV	Matlab	Matlab 2007b
	Skim matrices/aggregated cost	CSV				
	Parameters of calibrated model	CSV				

2.3.1.2 Demand for roundtrip station-based car sharing

A multi-method data-driven model had been developed to model the demand of roundtrip station-based car sharing in Regensburg. For more details about the model, the reader is referred to Deliverable 4.1. The demand, given the special business model of the system, has to be estimated additionally using a model external to the proposed intermediate modelling approach, as mentioned in Deliverable 4.1. Since the demand corresponding to the system is an addition to the trips in the existing OD matrices, the approach has been included as part of the induced demand module. Besides the demand per se, it is also relevant to profile the users of such a system, so that the demand can be linked to individuals.

In the integration framework presented in Section 2, this model could be placed above the fleet management step. The data-driven demand will be directly fed into the fleet simulator to simulate the trips and outputs KPIs. Given the data-driven nature, it is not possible to establish feedback loops, e.g., with the traffic assignment step, to update the demand for the car sharing system. Nevertheless, such a loop is not required, since the car sharing system in Regensburg is small and its penetration does not affect the existing travel times significantly.

Table 2 presents the data requirements for the model input as well as the module outputs.

Table 2. Interface and data requirements for the demand for roundtrip station-based car sharing module.

<table>
<tr><th rowspan="2">Module</th><th colspan="2">Input</th><th colspan="2">Output</th><th rowspan="2">Programming language</th><th rowspan="2">Software requirements</th></tr>
<tr><th>Type</th><th>Format</th><th>Type</th><th>Format</th></tr>
<tr><td rowspan="4">Demand for roundtrip station-based car sharing (linear regression, Dirichlet regression and statistical sampling)</td><td>Car sharing supply (no. of vehicles for each analysis zone)</td><td>CSV</td><td rowspan="4">Trip requests with OD</td><td rowspan="4">CSV</td><td rowspan="5">R</td><td rowspan="5">R 4.0.3</td></tr>
<tr><td>Analysis day and month</td><td>CSV</td></tr>
<tr><td>OD zones with distance between them</td><td>CSV</td></tr>
<tr><td>Travel distance distribution for the car sharing system</td><td>CSV</td></tr>
<tr><td>User profiling (multinomial logit model)</td><td>Synthetic population</td><td>CSV</td><td>Users of car sharing</td><td>CSV</td></tr>
</table>

2.3.2. OD matrix selection module integration

This module automates the selection of a representative OD matrix of the past for a near time horizon that consists of the classification of a day to a type of matrix. Types of matrices have been derived by clustering OD matrices according to their travel patterns and computing the average matrix of each cluster. The selection and classification features include weather data, calendar days, festivities and some special events, such as major football matches.

This module is proposed as an alternative to the direct estimation of OD matrices using observed data sources, such as survey or mobile phone data, in such a way that the OD matrix of a given day can be approximated beforehand. Therefore, this module can be used as a provider of OD matrices to the entire system by means of providing OD matrices into the synthetic population module (Section 2.3.3) and also can be used to provide inputs for other modules, such as the data-driven shared mobility demand predictor (Section 2.3.4.2).

As described in Deliverable 4.1, this module is used to select representative OD matrices in the short term. The module is split into three submodules, two of which feature alternative clustering methods of OD matrices to obtain the library mentioned above, and another one to perform the training and classification of clustered data to support the matrix selection process. Table 3 illustrates the software requirements and interfaces of each of the submodules in the module.

Table 3. Interface and data requirements for OD matrix selection module.

Module	Input: Type	Input: Format	Output: Type	Output: Format	Programming language	Software requirements
Base OD matrix Clustering	OD matrix distance file	CSV	OD matrix clustering model	Binary	Python	Python (3.7.6) Scipy (1.4.1) Numpy (1.18.1) Pandas (1.0.1) Scikit-learn (0.23.2) Matplotlib (3.1.3)
		CSV	Clustering of OD matrices according to travel patterns similarity: trip volume and matrix structure	CSV		
Graph-based OD matrix clustering	OD matrix distance file	CSV	OD matrix clustering model	Binary	R	R (4.0.3) R libraries: Lubridate (1.7.9.2) Tidyverse (1.3.0) Pals (1.7) Ape (5.4-1) Python libraries: Pecanpy (1.0.1)
		CSV	Clustering of OD matrices according to travel patterns similarity: trip volume and matrix structure	CSV		
OD matrix classification	Clustered OD matrices	CSV	Trained OD matrix classification model	Binary	Python	Python (3.7.6) Scipy (1.4.1) Numpy (1.18.1) Pandas (1.0.1) Scikit-learn (0.23.2) Matplotlib (3.1.3)
	Input feature data	CSV	Classified daily OD matrices			

Specifically, the module provides the methods to train two clustering schemes over a set of OD matrix distances as well as to train a cluster classification model for matrix selection from input features and the pre-trained classification model.

This module defines two clustering submodules, one based on hierarchical clustering and another one based on graph-embeddings. On the one hand, hierarchical clustering has been proposed to aggregate the data by grouping days and groups of days following their pairwise similarities. On the other hand, a clustering method based on graph-embeddings has been proposed to provide an enhanced clustering method that converts similarity relations into a graph and obtains an embedding of that graph, that is, a representation of the nodes in the form of vectors of real values that keeps the properties of the graph. Then, the clusters are obtained using the Euclidean distance and hierarchical clustering over those real vector representations of OD matrices relations.

The OD matrix clustering submodule takes as input a CSV file containing the distance metric for every pair of matrices according to the inverse of the similarity metric defined in deliverable D3.3 "Methodologies and algorithms for mobility data analysis" and generates a binary file containing the clustering model and a CSV file containing the results of the clustering process, which consist of a list of OD matrices (identified by day) and their corresponding cluster ID.

Alternatively, the graph-based clustering submodule receives as input a CSV file containing pairwise OD matrix distances and produces the binary clustering module and a CSV file containing the assignment of each OD matrix to a given cluster.

Finally, the classification submodule receives the relation of OD matrices and their assigned clusters in a CSV file along with another CSV file containing machine learning features per OD matrix and produces a binary classification model trained using the input data and a file that assigns its class to each given data instance.

2.3.3. Synthetic population module integration

This synthetic population module is the primary component in the workflow as it generates the synthetic individuals, which represent the disaggregated demand to be provided to the mode choice module for the estimation of the modal split for conventional and shared mobility trips. For the scope of the MOMENTUM project, the objective is to synthesise a simplified representation of the actual population, based on sociodemographic and other relevant information (i.e., household and individual attributes). The selection of specific household and individual attributes, and categories to be used in the synthetic population generation depends on the data availability in the census and samples from travel surveys. For more details with respect to the methodology adopted for the population synthesis as well as the selection of the key attributes, the reader is referred to Section 4.3 in Deliverable 4.1.

The main inputs related to this synthetic population module are aggregate sociodemographic data representing the entire population (census) as well as samples of disaggregate data from travel surveys with their corresponding geographic zoning system. Given those input data, a basic synthetic population is generated. Subsequently, the destination choice sub-module is used in order to assign a destination zone (or specific location) to each synthetic individual. This process requires as input available OD matrices, aggregate (e.g., obtained from a travel demand model) or disaggregate (e.g., estimated from mobile phones). Finally, in cases where some important variables are not available in both the census and survey data, which is a requirement for deploying the selected synthetic population method, a statistical matching model is applied to assign the missing attributes based on mutual variables in the synthetic population and travel survey data.

The final output from the synthetic population module (obtained after the synthetic population generation, addition of a destination and any missing attributes to the synthetic individuals) mainly serves as input to the disaggregate mode choice model. However, it may also be used as input to the disaggregate car ownership model to estimate the car ownership levels at the level of households, as basis for the car ownership of the synthetic individuals if a re-run of the main modelling framework is required. Moreover, the synthetic population module may receive as input during a re-run of the main modelling scheme the output from the aggregate car ownership module, which is used to estimate the number of cars per capita per zone. A detailed description of the car-ownership module integration workflow is provided in Section 2.5.1.

Table 4 summarises the input and output type and format as well as the interface requirements for this module.

Table 4. Interface and data requirements for the synthetic population module.

Module	Input		Output		Programming language	Software requirements
	Type	Format	Type	Format		
Synthetic population generation	Sociodemographic data	CSV	Synthetic individuals/households Household weights	CSV	Python	Python 2.7.9
	Samples of households and individuals	CSV				
	Geographic areas	CSV/SHP				
	Geographic mapping (between different zone levels)	CSV/SHP				
	Marginals for household and individual attributes	CSV				
Destination choice	Synthetic population	CSV	Synthetic population with trip distance	CSV	Python	Python 3.7.1
	OD matrices (aggregate or disaggregate)	CSV				
Statistical matching	Synthetic population	CSV	Enhanced synthetic population with additional attributes	CSV	Python	Python 3.7.1
	Travel survey data	CSV				

2.3.4. Mode choice modules integration

2.3.4.1 Disaggregate mode choice model

Table 5 summarises the input and output type and format as well as the interface requirements for the disaggregate mode choice model. The inputs are related to the demand, network supply as well as information related to the supply of the shared service systems.

Table 5. Interface and data requirements for the disaggregate mode choice model.

Module	Input		Output		Programming language	Software requirements
	Type	**Format**	**Type**	**Format**		
Disaggregate mode choice model	Synthetic population	CSV	Synthetic individuals with mode assigned (choices: conventional systems as a whole, bike sharing, car sharing and ridesharing)	CSV	R	R 4.0.3
	Travel times/skim matrix	CSV				
	Bike-sharing supply	CSV				
	Sharing vehicle availability (from fleet management)	CSV				

2.3.4.2 Data-driven shared mobility demand estimation

The data-driven shared mobility demand estimation module performs an estimation of the volume of shared mobility trips to be expected in a given service for a set of OD pairs and dates. This model uses as input features the figures of different segmentations of the general OD matrix, weather data from the nearest weather station and land use values of each zone. For further detail, the reader can refer to Section 4.4.2 of Deliverable 4.1.

This module requires general OD matrices and shared mobility data. Since OD matrices are typically not available in real-time, the resulting selected matrices from the OD matrix selection module (Section 2.3.2) would be used as input. Similarly, the trip estimation obtained by this model can be used as an alternative module for the disaggregate mode choice model (Section 2.3.4.1) or directly as an input to the fleet management module (Section 2.4).

This module can be used to perform an estimation of shared mobility demand data from a given service using a pre-trained model. Furthermore, the module provides the methods to perform the training of a model for a given SM service operating in one or more cities taking as input a historical register of features associated to the output (training data).

Table 6 summarises the input and output interfaces of the submodule of the data-driven shared mobility demand estimation model and their different software requirements. The implemented module has been divided into four submodules that run individual processes forming a single flow where each submodule provides the inputs for the next one.

The first submodule (pre-processing) is in charge of the individual pre-processing of each data source to adapt and convert each input data source into a CSV table that contains information aggregated by origin location, destination location, date and time (each data source up to the relevant level). The data fusion step is in charge of combining each of the pre-processed data sources into a single file that is ordered by origin location, destination location, date and time of each observation.

After pre-processing and data-fusion, the data augmentation submodule applies to the entire dataset time aggregations into days by summing trips and averaging other values. This module also augments the sample of shared mobility trips to increase shared mobility data variability for model training (See deliverable D4.1 for details) and segments the dataset into four according to inter-quantile ranges of the OD pair distance.

Finally, the OD pair trip prediction submodule performs both training and prediction. When training, it takes as input a large amount of pre-processed data that contains the observed trips in shared mobility per OD pair and generates a binary file that contains the trained prediction model and a file containing trip predictions per origin location, destination location and date. When the model is already trained, the OD pair trip prediction submodule receives the trained model and a set of prediction data that does not contain shared mobility trips (prediction variable) and returns a CSV file containing trip predictions per origin location, destination location and date.

Table 6. Interface and data requirements for the Data-driven shared mobility demand model.

Module	Input		Output		Programming language	Software requirements
	Type	Format	Type	Format		
Pre-processing	General OD matrices	CSV	Processed General OD matrices	CSV	Python	Python (3.7.6) Numpy (1.18.1) Pandas (1.0.1) Scikit-learn (0.23.2)
	Shared mobility OD matrices	CSV	Processed Shared mobility OD matrices			
	Daily Weather data	JSON	Processed Daily Weather data			
	Land use data	SHP	Processed Land use data			
Data Fusion	Processed General OD matrices	CSV	Model input data (hourly)	CSV		
	Processed Shared mobility OD matrices					
	Processed Daily Weather data					
	Processed Land use data					
Data augmentation	Model input data (hourly)	CSV	Model input data (daily and segmented by distance)	CSV		
OD pair trip prediction	Model input data (daily and	CSV	Trained classification model	Binary		

	segmented by distance)		Estimated trips per OD pair	CSV		

2.4. Integration of fleet management modules and simulation of shared mobility services

The fleet management module workflow integrates algorithmic procedures and simulation for shared mobility services to address both the strategic (planning) and operational aspects of designing accurate and robust mobility services.

Sections 2.4.1 and 2.4.2 discuss the fleet management planning and operational processes, respectively, while Section 2.4.3 one presents the operational scale of the integration framework.

2.4.1. Fleet planning models

The goal of the fleet management planning module is to decide the parameters that will affect and determine the performance of the service in the long term. The problems that the module can handle involve Demand Responsive Transit (DRT), Ride Sharing (RS), Bike Sharing (BS), Scooter Sharing (SS), Car Sharing (CS). Those problems can be generalised into two wider categories given the operational model they follow.

Firstly, DRT and RS belong to the category of the shared trip (--), so the outputs needed for planning such services should embody the number of vehicles to serve the estimated demand, the actual location of stops (in case of fixed station DRT), the capacity of each vehicle and the suitable service area. To do so, the model should process demand density input to get a more realistic view of the spatial distribution of demand, a set of stop candidates (or service spots), and the distribution related to each one. Additionally, the distance and time matrix are useful for service level and network performance measures. The module also contains a pre-processing step that decides stop candidates' location and initial fleet size and capacity. During the simulation experiments, those parameters were refined to choose the optimal subset.

The second category contains the BS, SS, and CS as they constitute resource sharing services. In that case, the core strategic parameters should depict the number and the location of docks, the capacity of each one, and the number of resources, along with the rebalancing strategy and the area suitability analysis. In fact, to retrieve those results, a dataset of trips across that area is needed, preferably for micro-mobility means. Moreover, the service its available fleet (if there is already present) and the distance matrix among possible docks should be supplied as input.

Table 7 summarises the input and output type and format as well as the interface requirements for this module.

Table 7. Interface and data requirements for the fleet planning models.

Module	Input		Output		Programming language	Software requirements
	Type	**Format**	**Type**	**Format**		
Fleet management planning: DRT	Demand Density	CSV, JSON, XLSX	Number of Vehicles	JSON	Python 3.8	Python 3.8 NumPy (1.19.2) Pandas (1.1.3) Geopandas (0.9.0) Scikitlearn (0.24.2) OR-tools (8.2.8710)
	Stations Position	CSV, JSON, XLSX	Final Location of stops	JSON		
	Demand Distributio n for each Stop	CSV, JSON, XLSX	Capacity	JSON		
	Vehicle Position	CSV, JSON, XLSX	Schedule	JSON		
			Operational Program	JSON		
Fleet management planning: RS	Demand Trips	CSV, JSON, XLSX	Valuable Areas	JSON		
	Vehicles Available	CSV, JSON, XLSX	Matching Method	JSON		
	Distance-Time Matrix	CSV, JSON, XLSX	Fleet Size	CSV		
Fleet management planning: BS, CS, MM	Demand Trips	CSV, JSON, XLSX	Valuable Areas	JSON		
	Vehicles Available	CSV, JSON, XLSX	Rebalancing Strategy	JSON		
	Distance-Time Matrix	CSV, JSON, XLSX	Fleet size	JSON		
			Stations	JSON		

2.4.2. Fleet operational models

The same categorisation of services made for fleet planning applies in the operational part of the fleet management module. In fact, the operational part of the module is used into the simulation experiments, as the service manager (see Section 2.4.2). 2.4.3The two basic mathematical problems to be addressed are the Dial-a-Ride Problem (DARP) and the rebalancing/relocation problem. The DARP is mainly used for the DRT and RS services while the rebalancing models should handle the resource sharing services like BS, SS, CS. The main difference, between this two, is that the DARP aims to match users in a same route. This tries to find the optimal combination of vehicles and passengers given their constraints. On the other hand, rebalancing focusses on relocating the available resources to maximise the vehicles availability and the expected profit. Thus, the DRT receives as inputs a set of trip requests with origin destination and time windows as well as the capacity and the location of the vehicles. The output of the algorithm returns an optimal sequence of actions to serve those requests. The rebalancing algorithm differs significantly as it needs to comply with the constrains posed by the set of docks with surplus (pick-up unit) and a set of shortage (deliver units) to decide the pick-up and delivery is the strategy. Hence, it also uses mathematical models similar to DARP, but there is no specific request pair with known pick-up and delivery. The outcome is both the sequence and the number of units that should be picked up or delivered in any station.

Table 8. Interface and data requirements for the fleet operational models.

<table>
<tr><th>Module</th><th colspan="2">Input</th><th colspan="2">Output</th><th>Programming language</th><th>Software requirements</th></tr>
<tr><td></td><td>Type</td><td>Format</td><td>Type</td><td>Format</td><td rowspan="7">Python 3.8</td><td rowspan="7">Python 3.8

Numpy (1.19.2)

Pandas (1.1.3)

Geopandas (0.9.0)
Pulp (2.4)

OR-tools (8.2.8710)</td></tr>
<tr><td rowspan="3">Fleet management operational: DRT, RS</td><td>Trip Requests</td><td>JSON</td><td rowspan="3">Sequence of actions</td><td rowspan="3">JSON</td></tr>
<tr><td>Distance/Time matrix</td><td>JSON</td></tr>
<tr><td>Fleet size and Capacity</td><td>JSON</td></tr>
<tr><td rowspan="3">Fleet management operational: BS, CS</td><td>Docks demand or surplus</td><td>JSON</td><td rowspan="3">Sequence of actions</td><td rowspan="3">JSON</td></tr>
<tr><td>Vehicles Available</td><td>JSON</td></tr>
<tr><td>Distance-Time Matrix</td><td>JSON</td></tr>
</table>

2.4.3. Integration with Aimsun Ride simulator for shared mobility services

The Aimsun Ride simulation platform for shared mobility services is an advanced tool that enables the deployment and assessment of various scenarios related to new shared mobility applications. The main input files for the shared-mobility services simulation platform Aimsun Ride (Aimsun, 2020) are related to the definition of the service area, the requests and operator's information. The files should be provided in JSON format. In particular, the requests file has a collection of fields related to the requested trips to be executed during the simulation (i.e., origin, destination, departure time, etc.). Furthermore, each request can include several user-defined fields that are, in turn, provided to the operator as additional attributes (e.g., accessibility constraints for specific individuals, walking time constraints, etc.). The input related to the operator includes information about the specific service that is being evaluated (e.g., car sharing, DRT, etc.) as well as the fleet specifications. These specifications can be, among others, the vehicle type (e.g., bus, taxi, bike, etc.), origin position of the vehicles, number of seats in each vehicle, location of stations (for station-based services), etc. Furthermore, Aimsun Ride uses link travel times that reflect the prevailing traffic conditions generated by the total demand in the network (demand for conventional modes). Subsequently, Aimsun Ride computes the path costs that are further provided as travel information to the operator. The interface is flexible in utilising travel times obtained from different sources: 1) free-flow travel times, 2) external travel times (historical travel times), 3) simulated travel times using the Aimsun Next traffic simulation software. The low demand penetration scenario assumes no change in travel times due to the introduction of demand for the new mobility services. Hence, historical travel times can be used as input and considered fixed in all fleet management scenarios. In this case no simulation of the network demand is required; rather external travel times can be imported which may be obtained from a previous static assignment performed in any traffic simulation model or from available travel time observations.

In the case that the traffic simulator Aimsun Next is used, a network model needs to be available to interact with the Aimsun Ride platform. Depending on the assumptions related to the change in travel times due to the demand shared-mobility services, the network model can be modelled at different traffic flow resolutions, namely, microscopic, mesoscopic, macroscopic or hybrid. For further details about the traffic assignment module, we refer the reader to Section 0. The frequency of interaction between the simulator and the fleet operator varies depending on the assumptions concerning the travel times. The simulated travel times can be considered fixed throughout the fleet operations or may be updated iteratively by performing several traffic assignments. If the travel times are to be considered fixed, the traffic assignment is performed only once (low demand penetration scenario). However, if the travel times are expected to change with an increase in the demand for new services (high demand penetration scenario), the frequency of the traffic assignment runs will be equal to the number of fleet management scenarios to be evaluated after a percentage change in the demand for new mobility services. The simulator will then pass the updated travel information obtained from the simulator to the operator. Several iterations between the simulator and operator might be needed in order to converge to an optimal solution.

After an execution, Aimsun Ride produces a file with all the events executed and KPIs that can be used to assess the service quality. Optionally, a recording file is also produced to visualise the simulation. The events file is a binary file that can be exported to JSON format, which will contain a list of events with information for the different event types that have been executed (either for a request or a vehicle in the fleet). This information is related to the state of a request or a vehicle executing it, the position and used capacity of vehicles and the path chosen. For more details on the Aimsun Ride simulation platform we refer the reader to Deliverable 4.1.

Table 9 summarises the minimum information required as input to the Aimsun Ride model as well as the main outputs and interface requirements for this model.

Table 9. Interface and data requirements for the Aimsun Ride simulator for shared mobility services.

<table>
<tr><th rowspan="2">Module</th><th colspan="2">Input</th><th colspan="2">Output</th><th rowspan="2">Programming language</th><th rowspan="2">Software requirement</th></tr>
<tr><th>Type</th><th>Format</th><th>Type</th><th>Format</th></tr>
<tr><td rowspan="6">Aimsun Ride</td><td>Road network</td><td rowspan="2">JSON/SHP /Map</td><td rowspan="6">KPIs related to service quality:
• Travel times (users and fleet)
• Travelled distance
• Number of served/unserved requests
• Waiting times
• Capacity at stations (for station-based shared services) and vehicles
• Event log file
• Recorded file for visualisation</td><td rowspan="6">JSON/ Exporting capability of the UI</td><td rowspan="6">Python/C++</td><td rowspan="6">Aimsun Ride</td></tr>
<tr><td>Service area</td></tr>
<tr><td>Disaggregate demand requests</td><td rowspan="2">JSON</td></tr>
<tr><td>Vehicle fleet locations and specifications</td></tr>
<tr><td>Link travel times</td><td>External or Aimsun Ride[1]</td></tr>
<tr><td>Station locations</td><td>JSON/SHP</td></tr>
</table>

[1] Simulated travel times can be obtained in case a transport network model for the area of interest is available and calibrated, so as to replicate the prevailing traffic phenomena and travel demand patterns.

2.5. Supply modules integration

2.5.1. Car ownership module integration

Two different car ownership models had been developed, an aggregate and a disaggregate model, technical details of which can be obtained in Deliverable 4.1. The aggregate model is based on the aggregate socio-demographic characteristics of city districts/traffic zones (e.g., total population) and the supply of shared mobility services. Once the main modelling framework shown in Section 2.2 is run, this model will be used to (re-)estimate the number of cars per capita per zone. Then, depending upon the extent of change in car ownership levels, the main modelling framework could be re-run. During such a re-run, the output of this model will be used as input for the synthetic population generation.

The position of the disaggregate model with respect to the main modelling framework is similar to that of the aggregate model, i.e., this model will be used to estimate the car ownership levels (but at the level of households) after the completion of the sequence of the main modelling framework. This will be used as basis for the car ownership of the synthetic individuals during the re-run (if required) of the main modelling framework. However, while the aggregate model just needs census data and the shared mobility supply data from the fleet management module, the disaggregate model requires the outputs from (i) synthetic population generation and (ii) traffic assignment, besides the supply data from the fleet management module.

Table 10. Interface and data requirements for the car ownership model.

Module	Input		Output		Programming language	Software requirements
	Type	Format	Type	Format		
Aggregate car ownership	Aggregate sociodemographic data per zone (Census)	CSV	Number of cars per 1000 individuals per zone	CSV	R	R 4.0.3
	Sharing system supply per zone	CSV				
Disaggregate car ownership	Synthetic population	CSV	Synthetic individuals and households Household weights	CSV	R	R 4.0.3
	Travel times/skim matrices	CSV				

2.5.2. Traffic assignment modules integration

In relation to the modelling framework, travel times serve as one of the initial inputs of many of the modules. Typically, travel times are provided by running analytical or simulation-based traffic assignment models towards an equilibrium solution, as real travel time data are usually unavailable or have very limited coverage. Depending on the assumptions regarding the impact of demand changes for shared mobility on the network traffic conditions, several iterations of the modelling framework (Figure 4) may be performed (high penetration demand case). Hence, the traffic assignment model will interact iteratively with other modules to provide the updated travel times. Travel times (represented as skim matrices or at a link level) are required for the disaggregate mode choice model as well as the fleet management module. Depending on the scope of the analysis in terms of how accurate the traffic dynamics in the network should be replicated, static or dynamic traffic assignment models can be utilised.

2.5.2.1 Integration of assignment methods for urban environments

The traffic assignment module for urban environments is a static assignment algorithm in which traffic originating from or destined towards a zone is distributed over all links within the zone, instead of only over the connectors. We refer to Deliverable 4.1 for more details. The required input is similar to traditional assignment methods: an

OD matrix that describes the traffic demand from zone to zone; and a description of the road network including the supply parameters that determine the link performance (e.g., number of lanes, capacity, free speed).

Table 11. Interface and data requirements for traffic assignment model for urban environments.

Module	Input		Output		Programming language	Software requirements
	Type	Format	Type	Format		
Traffic assignment for urban environments	OD matrices (aggregate)	CSV	Network performance macroscopic variables per link/routes/system	CSV	Matlab	Matlab 2020
	Road network	SHP				

2.5.2.2 Dynamic traffic assignment model

The dynamic traffic assignment (DTA) module implements a hybrid macroscopic-mesoscopic network loading developed in the Aimsun Next traffic simulation software (Aimsun, 2020). The hybrid DTA framework offers the flexibility to model the new mobility services at a strategic level, while it also enables more detailed evaluation of their impacts on specific study areas that can be modelled at a disaggregate level. The model is presented in detail in Deliverable 4.1.

The minimum required inputs to support a DTA model are the common inputs that define a simulation scenario configuration:

- **Supply (network) data:** a road network representation is required to build a network model (i.e., sections, nodes, turns, centroids, number of lanes, capacity, maximum speed).
- **Demand data:** aggregated demand matrices, namely time-dependent OD matrices, are required to represent the number of trips between each origin and destination zones in space and time.
- **Traffic control data:** depending on the application and study network, traffic control plans and signalised intersections can be also specified.
- **Public transport data:** input related to public transport (lines, stops, time schedules, etc.) may be also provided.

The hybrid DTA model provides as output various network performance variables, such as link travel times, speeds, etc. Specifically, the output travel times from the hybrid DTA model can be further received by the fleet operational algorithms to optimise the travel plans of the requests for shared mobility services. Furthermore, the disaggregate mode choice model uses as input the simulated travel times, in a skim matrix format, to estimate the new modal splits. Finally, the traffic assignment model interacts with the dynamic emissions model as it provides the simulated speeds and flows per vehicle type to estimate the network-wide emissions.

Table 12 summarises the input and output type and format as well as the interface requirements for this module.

Table 12. Interface and data requirements for the hybrid dynamic traffic assignment model.

Module	Input		Output		Programming language	Software requirements
	Type	Format	Type	Format		
Hybrid Traffic Assignment	Scenario configuration Road network Aggregate demand matrices Traffic control plans Public transport	Aimsun Next data model	Network performance macroscopic variables per link/routes/ system	DB/CSV	C++	Aimsun Next

2.6. Sustainability modules integration

2.6.1. Emission module: Static emission model integration

The static emission module uses the output of the traffic assignment model and converts the modelled speeds and traffic volumes on each link of the network into emissions. The pollutants include CO, CO_2, NO_x, PM and VOC. The emission factors used to calculate the emissions depend on the vehicle fleet. As the fleet composition differs between EU members states, the emission module uses different factors for each country. Moreover, the vehicle fleet constantly changes, so average emission factors are quickly outdated as new, cleaner, vehicles replace older vehicles. In the calculation, the TML fleet-model[1] is linked to COPERT 5[1], so as to construct a data set of fleet-average emission factors per pollutant, per EU member state per year from 2016 until 2050.

[1] https://www.tmleuven.be/en/navigation/Fleet-Model

[1] COPERT is the EU standard vehicle emissions calculator. It uses vehicle population, mileage, speed and other data such as ambient temperature and calculates emissions and energy consumption for a specific country or region.

Table 13. Interface and data requirements for the static emission model.

Module	Input		Output		Programming language	Software requirements
	Type	Format	Type	Format		
Static emission model	Link characteristics (flow, speed, length)	CSV	Emissions per link	CSV	Matlab	Matlab 2007b
	Configuration (year + country)					

2.6.2. Emission module: Dynamic emission model integration

The data-driven dynamic emission module estimates emissions curves, which depend on dynamic speed profiles and different pollutant types. In addition, the estimated emissions are distinguished per vehicle type (car, bus, taxi, truck, etc.), engine combustion type (diesel and petrol) and vehicle Euro vehicle class (Euro 0 – Euro 6). Therefore, the model requires as input measured traffic emissions per vehicle type, observed average speeds and flows per vehicle type and road link.

Currently, the estimated pollutants include CO_2 and NO_x. Subsequently, the estimated emissions curves can be introduced into the traffic simulation software Aimsun Next to predict the network-wide traffic emissions. The vehicle fleet needs to be specified, including the vehicle types, engine fuel types as well as vehicle Euro class. The emission model can be integrated with traffic assignment models, static or dynamic, that will provide the simulated traffic indicators (speed and flows) to predict the traffic emissions. Table 14 summarises the input and output type and format as well as the interface requirements for this module.

Table 14. Interface and data requirements for the dynamic emission model.

Module	Input		Output		Programming language	Software requirements
	Type	Format	Type	Format		
Dynamic emission model	Vehicle types	CSV	Emissions per link	DB/Exporting capability of the UI	Python, C++	Aimsun Next
	Emissions measurements (e.g., CO_x,NO_x, PM, HC) per vehicle type					
	Average speed per vehicle type/link					

	Flow per vehicle type/link					

3. Recommendations for calibration and validation

3.1. Data requirements, calibration, and validation of the integrated models

The objective of this chapter is to provide general recommendations and guidelines regarding the type of data required for building, calibrating, and validating the integrated models for the application of the enhanced transport simulation software. The calibration and validation guidelines provided in this deliverable are focused on the models and algorithms that were developed or enhanced within the MOMENTUM project. Other external models that are not part of the modelling schema, but whose output may be required as input in the integration framework, are assumed to be already calibrated and validated.

For the models in Sections 3.1.1.2, 3.1.1.5 and 3.1.3.1, the estimation step includes the process of determining the model form (e.g., linear regression or multinomial logit), the model specification (i.e., the independent variables) and the coefficients based on the input data. The calibration step, recommendations for which are presented in the subsequent sections, refers to the process of refining the estimated coefficients to make the model more suitable for a given case study, without changing the model form and specification.

3.1.1. Demand modules

3.1.1.1 Induced demand module – Elasticity

In theory, the described variable demand models can be calibrated based on observations on demand and mode choice changes with respect to changes in an attribute x. Such data might be obtained through observations, stated- or revealed preference surveys. For the linear models, a simple linear regression can suffice, while for the logit model, state-of-the-art software packages, e.g., Biogeme (Bierlaire, 2020), can be used.

However, in practice, it is not always feasible nor desirable to obtain enough data for a data-driven calibration. Especially when only ballpark estimates for the variable demand are required, a less resource-consuming estimation method is preferable. We propose to calibrate the model based on demand elasticities that can be found in literature.

The calibration is derived from the definition of the found elasticity. Assume that variable x is changed, which might influence the utilities U^j of several alternatives j. The x-elasticity of the alternative's demand D^j is commonly defined as:

$$\eta_j = \frac{\partial D^j / D^j}{\partial x / x} = \frac{\partial D^j}{\partial x} \frac{x}{D^j}.$$

It quantifies how a relative change in x (e.g., monetary cost or travel time of i) leads to a change in demand D^j. Using $D^j = P^j D$, where P^j is the share of the total demand choosing for alternative j, the elasticity can be expressed as:

$$\eta_j = \frac{x}{D^j}\frac{\partial(P^j D)}{\partial x} = \frac{x}{D^j}\left[D\frac{\partial P^j}{\partial x} + P^j\frac{\partial D}{\partial x}\right].$$

The changed demand D^j can thus stem from either the demand for other alternatives D^k, or from the part of the population that decides (not) to make the travel OD at the time period under study. Applying this equation to a given mode choice model (see Deliverable 4.1 for more details), the parameters of the induced demand module can be estimated.

Care must be taken when using elasticities from literature, as elasticities typically depend on the present values of demands D^j, x and modal split.

3.1.1.2 Induced demand module - Demand for roundtrip station-based car sharing

This model has been specifically built for the Regensburg car sharing system. As mentioned in Deliverable 4.1, this model consists of a multi-method framework, i.e., a linear regression for estimating the total demand per day, a Dirichlet regression for distributing the total demand to the individual stations and a multinomial logit model for user profiling.

Though it is not possible to transfer the Dirichlet regression model to other cities (since the dependent variables correspond to the eight stations in Regensburg and cannot be generalised), the linear regression model and the multinomial logit model can be transferred to other cities (with a similar car sharing service scheme). It is possible that the coefficients for other cities differ from those of Regensburg. If suitable data is available, the coefficients can be calibrated in such cases. The possibility of adapting the linear regression model under different conditions is summarised in Table 15 and the multinomial model in Figure 11.

Table 15. Suggested approach for calibrating the demand model of roundtrip station-based car sharing system.

Model	Type of available data	Suggested approach
Total demand per day (linear regression model)	Operator data (historical data on trip bookings)	Coefficients can be calibrated using conventional ordinary least squares method and then the model can be used.
	No data	Calibration of the coefficients for a particular case-study cannot be performed and only direct application of the model with the existing coefficients is possible.

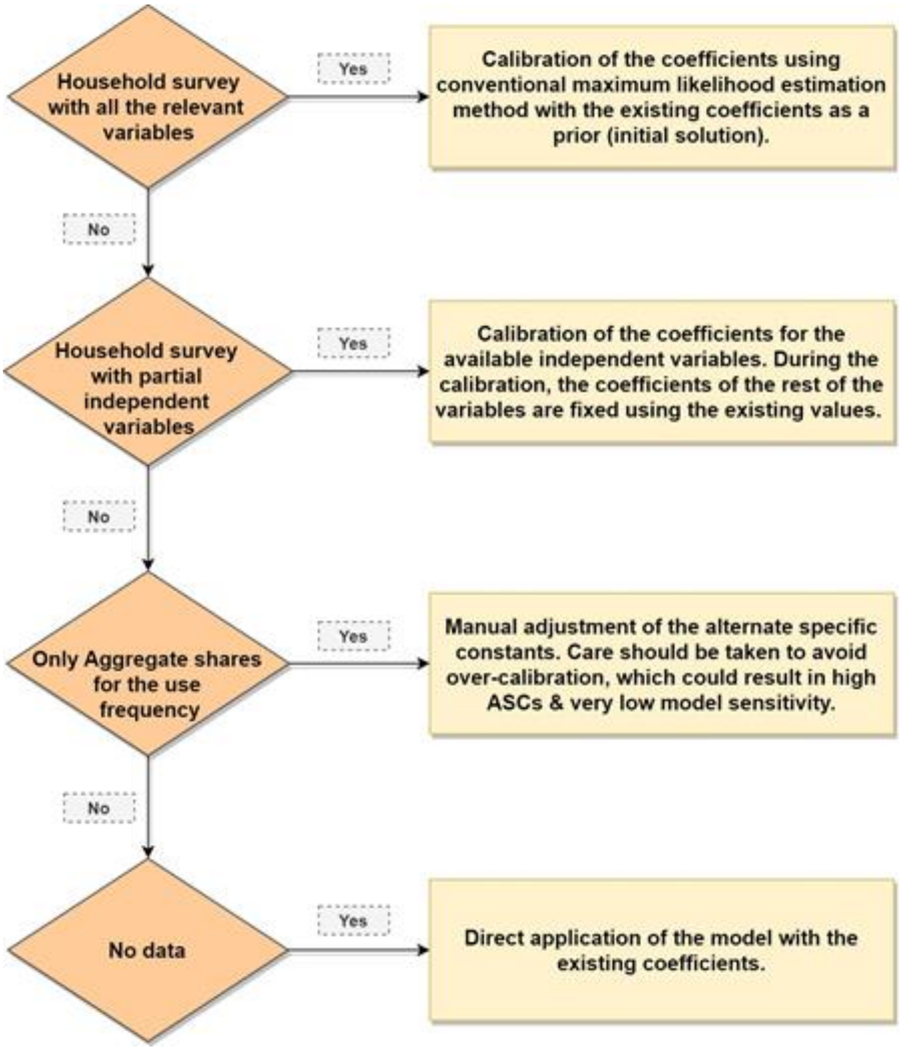

Figure 11. Suggested approach for calibration of the user-profiling (multinomial logit) model of roundtrip station-based car sharing system.

Besides the aforementioned calibration process, sensitivity analysis could be carried out to analyse the effect of the independent variables, and if necessary, the existing coefficients could be manually adjusted. Particularly, this step is important in case of direct application of the model with existing coefficients. When a change in the input variable results in an unexpected large change in the output variable, the coefficient value of the corresponding variable could be reduced in small steps until the required effect on the output variable is achieved. Similarly, if the effect on the output variable is smaller than expected, then coefficient value could be increased in smaller steps.

To avoid overfitting when calibrating the coefficients with new household survey data, it is advisable to split the available dataset into a training and validation sets: the former is used during the calibration process, while the latter (usually smaller) dataset is used only for validation. A better approach than this is to implement 5-fold cross validation (James et al., 2013).

To support easy change of existing coefficients, a separate coefficient file has been included for the respective models in the corresponding directory in the MOMENTUM GitHub repository. Therefore, the user does not need any coding knowledge in R (the language in which the models are scripted) to change the coefficient values.

3.1.1.3 OD matrix selection module

The OD matrix selection module is intended for the estimation of OD matrices in a short-term future (some weeks) from recent data. This module is composed of two machine learning models:

- **Clustering:** the clustering model aggregates OD matrices according to their trip volume and geographical properties. Each daily matrix will be assigned to a cluster with similar OD matrices and each of the clusters will have associated an OD matrix that will consist on the average of all the OD matrices assigned to that cluster.
- **Classification**: the training of a Decision Tree classification model takes as input all the data instances from the clustering using as class the assigned cluster and produces a model that, given a set of day-related features, predicts the cluster to which each matrix should be assigned. Once this assignment is performed, the predicted OD matrix will be the associated average OD matrix of each cluster.

While document Deliverable 4.1 “Transport Modelling Approaches for Emerging Mobility Solutions: Supply and Demand models” discusses the training and evaluation of the model for the Madrid case study, it is possible to train this model for other cities. For this training, a set of OD matrices for a zoning scheme of the area of study should be used. Concisely, training data is composed of OD matrices from days in the past whilst the target of the module is to select the most appropriate representative OD matrix (cluster) for specific days in the near future. As a result, the OD matrices selected for model training should comprise the entire historical data for several months in the recent past.

The reason for using historical data is the trade-off between large datasets and focusing on recent data: the availability of large amounts of data is useful to perform more robust estimation of each cluster OD matrix as well as achieving better machine learning inference capabilities in the supervised setting (classification). Nonetheless, the use of too old data may negatively impact model performance, since mobility patterns may have changed and thus, they can add biases in the system.

OD matrix availability is a potential bottleneck for the correct development of this project. OD matrices might be available for every day of study, but their cost can limit the total amount of OD matrices to be used in training. In these cases, using relevant months in mobility analytics like February, October, April or November is recommended in combination with a subset of months from summer (July or August).

Regarding model re-calibration, it is worth noting that new OD matrices can be generated with daily frequency in spite of some estimation delays. This implies that the model could be re-trained as frequently as daily to include such new data. However, it is expected that the individual contribution of a single OD matrix will not significantly change the model and, therefore, weekly or monthly re-trainings could be scheduled. These re-training processes should consider removing the oldest data accordingly.

Finally, hyper-parameter tuning of the model should be performed in order to adjust each model the best for the given data. In the case of clustering, the obtained clusters are highly dependent on the hyperparameters of the clustering methods used (e.g., cut-off thresholds in hierarchical clustering methods). As for both clustering methods proposed, below we give indications on the calibration of the most important hyperparameters:

- **Hierarchical clustering**:
  - *Cut-off threshold*. The calibration of the parameter that determines the number of clusters is highly data-dependent and it is not possible to establish even a range of values. In this sense, we recommend that the threshold is set according to the resulting number of clusters. This number of clusters should be in a range between 5 and 20 clusters, as a higher number may reduce the performance of machine learning based classification methods. Within that range, it is

recommended to keep a small number of clusters to avoid excessive imbalance between OD matrices in each cluster. Finally, we recommend an expert review of the resulting clusters to avoid outliers and noise in the definition of these clusters.

- **Graph-embeddings-based clustering**:
  - *Threshold μ for similarity graph*. As indicated in Deliverable 4.1, this parameter establishes when there is an arc between two OD matrices in the similarity graph. This value should be set within the range between the 65th and 75th percentile ensuring that only matrices with a high similarity are connected.
  - *Node2Vec parameters*. In the case of the Node2Vec algorithm (Grover & Leskovec, 2016) parameters, it is recommended to keep the same parameterisation, i.e.: p=1, q=20, random walks=500, random walk length=10, window size=10, feature vector dimension=128.

Hyper-parameter tuning of the Decision Tree algorithm should include, at least, the maximum depth of the tree. It is recommended that the minimum number of allowed leaf points and the information criteria (entropy or Gini index) should be adjusted as well. For that purpose, cross-validation should be used to perform an extensive grid search process over all parameters.

3.1.1.4 Synthetic population module

The data collection requirements and calibration guidelines for the synthetic population module may vary depending on the data availability for the city of interest.

The data requirements are defined by the type of method that is used for the generation of a synthetic population as well as the data availability. Synthetic population generation is based on census data and sample data from travel surveys. Within the MOMENTUM project, the open-source PopGen synthesiser (MARG, 2016) is used for population synthesis. It employs Iterative Proportional Fitting (IPF) and Iterative Proportional Updating (IPU) methods to compute the household weights (see Deliverable 4.1 for more details). This method requires that the selected variables to synthesise the population are present in both the census and sample data.

The synthetic population generation provides synthetic individuals, usually characterised by general sociodemographic attributes. At this stage, a specific trip is not assigned to the individuals. For this purpose, the destination choice model can be applied to assign a specific destination to each synthetic person, using information from available OD matrices for the period of interest. Nevertheless, interesting attributes related to the profile of the individuals that choose a shared mobility service for their trips may not have been included in the synthetic population, as this information may not be available in the census and the synthetic population yet. An example of such an interesting attribute is an individual has a public transport subscription, which is considered an important factor for an individual to choose e.g., a car sharing system. Nevertheless, these new attributes may be known for every observation in household travel surveys. In such cases, statistical matching can be applied to combine a sample of the synthetic individuals with available data from household travel surveys. The feasibility of this approach relies on the assumption that both data sets share common attributes with the same definition and categorisation (such as age class, gender, number of cars, etc.) that are considered sufficiently correlated with the new attributes of interest (available only in one data source).

Some cities may already have a synthetic population generated for the scope of other applications. In this case, it is not necessary to deploy the full synthetic population module, rather the proposed statistical matching model could be applied to enrich the synthetic population with the missing variables.

Table 16 summarises three possible application cases with respect to the population synthesis and provides a short description for the suggested approach for each case. More information related to the synthetic population methods can be found in Deliverable 4.1.

Table 16. Suggested approach for implementation process of the synthetic population module.

Application case	Suggested approach
Case 1: Synthetic population not available	Consideration of adequate variables and their categories as input to the population synthesiser. The selection should be based on both the data availability in the data sources to obtain the marginals and sample data as well as the desired characteristics of the synthetic individuals for a specific case study.
Case 2: Synthetic population available, but the individuals do not have a trip assigned.	The destination choice model can be applied to assign a specific destination to each synthetic person, using information from available OD matrices for the period of interest.
Case 3: Synthetic population available but missing important attributes	The statistical matching model can be applied to enrich the existing synthetic population with additional information. It is important to define a set of adequate matching attributes associated with the new information to be added.

3.1.1.5 Disaggregate mode choice model

The approach suggested in Section 3.1.1.2 for the multinomial model on car sharing use frequency is also applicable to the disaggregate mode choice model, since the model type is the same (i.e., multinomial logit model). It is recommended to carry out the sensitivity analysis to analyse the effect of the independent variables, and if necessary, the existing coefficients could be manually adjusted. Similar to the car sharing use frequency model presented in Section 3.1.1.2, the coefficients of this model are stored in a separate file in MOMENTUM GitHub repository such that they can easily be changed.

Besides the ones suggested in Section 3.1.1.2 (Figure 11), the following two cases might be of interest to cities:

1. Application of the model for one or two shared mobility service, instead of all the three services. Given the IIA (Independence of Irrelevant Alternatives) property of the multinomial logit models, it is possible to consider just one or two of the three shared mobility services included in the model without the need for additional calibration of the coefficients.
2. Inclusion of variables in addition to the existing ones.

Table 17 discussed a possible approach to deal with each of the two cases.

Table 17. Suggested approach for the application of the disaggregate mode choice model.

Application case	Suggested solution approach
Case 1: Application for one or two shared mobility services	A manual correction is required in the model script, i.e., remove the utility specification of the shared service which is not of interest. It is to be noted that setting the coefficient values corresponding to the non-relevant mode to zero will not result in the intended effect (i.e., the utility will be zero, which can lead to some probability being assigned to the non-relevant mode). This is a consequence of the fact that only differences in utilities between modes are of importance in multinomial logit models, not their absolute values.
Case 2: Inclusion of variables in addition to the existing ones	If the additional variables of interest are not correlated with the existing ones (i.e., one does not influence the other), it is possible to externally estimate the coefficients for them and include them in the coefficient file. Besides this, the additional variables must be manually encoded in the utility specification in the model script. It should be noted that the base choice should be kept as conventional system as a whole during the external estimation. If a different mode is chosen as a base choice, it is incorrect to integrate the additional coefficients with the existing ones.

3.1.1.6 Data-driven shared mobility demand estimation

The data-driven shared mobility demand model provides an estimation of the total trips to be observed in an OD pair for a given date as a function of general mobility represented by OD matrices derived from mobile phone data, weather records for each time period and land usage at each origin and destination locations. The proposed model trains a Random Forest model using historical data from a shared mobility service and its input features. Although document Deliverable 4.1 reports the calibration of a model for a Bike sharing and a Moto sharing service in Madrid, this model can be calibrated for other cities by executing the training subcomponent of the model using their OD matrices, weather data and land usage.

Since the model prediction relies on observations that have been taken in the same time period, the risk of training the model using outdated data is limited, as, in principle, weather events or drops in general mobility should be captured by the model. Hence, this model can be trained with as much data as available. In any case, there are three issues to take into consideration for the training of the shared mobility demand model:

- **State of the service**: even though the service is expected to follow consistent patterns through time, this assumption only holds from the moment it has been completely deployed and operates normally. For instance, data from the ramp-up phase of the service, where it still has a small number of trips that grows very fast, should be avoided as it does not reflect the real functioning of the service.
- **Punctual disruptions**: in some cases, specific problems of the system or in the city may cause punctual disruptions of the service where the observed patterns will probably be inconsistent. Unless there is an additional data source that can explain this issue, these type of data points should not be considered.
- **Data scarcity problems**: the model distributes trips among a grid zoning scheme that, in some cases, may result in very few originated and destined trips in many of the zones. When this happens, the training can be impacted since the variability of the training data would be limited due to the lack of trips (and not the

lack of patterns). The solution for this issue would be to aggregate data, so the trips observed have larger variability. At present, the data preparation phase of the module performs time aggregation into days and weekdays, but a target zoning scheme can be configured.

Regarding hyperparameter tuning, the number of trees in the forest, the maximum depth of each tree and the minimum samples in any tree node to perform a split has been adjusted following a grid search strategy using cross-validation. Any calibration attempt of the model should consider the tuning of at least the maximum tree depth and the number of trees in the forest.

3.1.2. Fleet management modules

3.1.2.1 Fleet planning module

The fleet management planning process requires some effort and parameter definition in strategic parameters that can affect the overall performance of the system. Critical parameters like the service network density, fleet size, and capacity of vehicles or docks directly affect most of KPIs like waiting time, access time (e.g., walking time), travel time and number of accepted requests. The first variable the user should define is the mean walking distance of each passenger expected to walk to access the service. The lower the goal the higher the number of stops/docks needed in the system and probably the higher the capacity and size of fleet. Moreover, the user can define the average trip duration, which affects the number of stops each vehicle can serve as well as the fleet size. Especially, the lower the desired waiting or trip duration time the higher the need for available vehicles. Those two parameters are critical to define the optimal number of system parameters during simulation experiments.

3.1.2.2 Fleet operational module

Operational algorithms are more fixed and task-specific so that they do not require a lot of effort for calibration. However, rebalancing is a process that usually occurs either after a fixed number of time steps or based on some unbalancing condition. This choice should be made by the users based on their needs, but it may also be considered as a tuning parameter. A more frequent and narrower rebalancing window leads to more expensive operational costs that may not be profitable in some cases. For instance, if a route takes place every 30 minutes it could generate costs that do not return as profits from that action. In particular, rebalancing operation has a specific cost C, while renting a bike returns a profit P. Hence, rebalancing within small time-windows may not be feasible. The time step is a critical parameter defined by the user. Similarly, the waiting time of the DRT line can make a significant impact on the overall performance of the service. Waiting too long will dissatisfy customers while serving immediately each request leads to low occupancy routes that are not profitable. This demonstrates that the fulfilment boundary can play a key role in the net performance of the system.

3.1.2.3 Simulation for shared mobility services with Aimsun Ride

As described in Section 2.4.3, besides the input files related to the demand requests and fleet information, the minimum input data to perform the simulation of shared-mobility services in Aimsun Ride is the network graph and link travel times. The travel times may be external or simulated using the Aimsun Next software. Hence, the data input and calibration requirements for the Aimsun Ride model differentiate between the following two cases:

1. Utilisation of external/historical travel times: If external link travel times are available and representative for the network and time period of interest, no simulation of the network demand is required. Hence, a network graph and link travel times would be sufficient inputs.

Travel times obtained through simulation in Aimsun Next: Requires a calibrated and validated transport network model to perform the traffic assignments and obtain the trave times. The network model can be built at the adequate traffic flow resolution for the scope of the examined application.

Table 18 summarises the required data and suggested approach for each case. More specific guidelines for the calibration and validation of a transport network model are provided in Section 3.1.3.3 and are applicable to the second case presented on Table 18.

Table 18. Suggested approach for calibrating the Aimsun Ride simulation platform.

Application case	Suggested approach
Case 1: Utilisation of free-flow or external/historical travel times for the path cost calculations	The minimum required data are link travel times and a network graph to represent the network geometry and characteristics. No calibration and validation of the network demand and supply parameters is required. The demand and supply for shared-mobility services is simulated on the provided network graph according to the travel plan solutions provided by the operator. External travel times can be obtained from an existing (calibrated and validated) travel demand model or observed measurements (if available). Different travel times may be provided per shared-service mode.
Case 2: Travel times obtained from traffic simulations inside the Aimsun Next	A network model in Aimsun Next is needed with calibrated demand and supply inputs and parameters at the desired traffic flow resolution (microscopic, mesoscopic, macroscopic, or hybrid). The Aimsun Ride platform interfaces with the Aimsun Next simulator to obtain the travel times that can be calculated by performing traffic assignments (static or dynamic) and will be further provided to the fleet operational algorithms.

3.1.3. Supply modules

3.1.3.1 Car ownership modules

Both the aggregate and disaggregate car ownership models are usable in multiple cities. The approach suggested in Table 17 is also applicable for the aggregate car ownership model, as both are linear regression models (a difference is that a census data is required here and not a sharing system operator data). Concerning the disaggregate car ownership model, it is recommended to follow the approach suggested in Figure 11.

As suggested in sections 3.1.1.2 and 3.1.1.5, a sensitivity analysis has to be carried out to analyse the effect of the independent variables, and if necessary, the existing coefficients could be manually adjusted. To avoid over-calibration, it is advisable to implement 5-fold cross validation (James et al., 2013). Similar to the car sharing demand model and the disaggregate mode choice model presented in Sections 3.1.1.2 and 3.1.1.5, the coefficients of the aggregate and disaggregate car ownership models are stored in a separate file (for each) in MOMENTUM GitHub repository such that they can easily be changed.

3.1.3.2 Traffic assignment model for urban environments

The traffic assignment module for urban environments does not differ substantially from traditional assignment models. Therefore, similar recommendations on calibration and validation that are valid for traffic assignment in general, are also applicable here.

The traffic assignment module for urban environments introduces the concept of access points (see Deliverable 4.1). These access points are hypothetical connections at the boundary of each zone (Figure 12). They are introduced to reduce the calculation time by reducing the number of destinations. These access points determine the attraction towards that zone and they are used to initialise the shortest path search upstream from that zone.

The model requires setting an initial cost of the access points. For each access point we initialise this value to the average distance towards each link within the zone. This is the standard option if no extra data is available for calibration. Alternative approaches based on available data are also possible. Floating car data (FCD) or traffic counts at the access points can inform which access points are used most frequently. The idea is to calibrate the costs of the access points such that the distribution matches the observed patterns more closely. For this, optimisation algorithms such as SPSA (Spall, 1987) can be used.

In a final step, the algorithm propagates the flows arriving at the access points to each of the destination links. Also, weights are used for each link to distribute traffic within the zone. Traffic from and towards this zone is distributed in proportion to these weights. As a good initial estimate for these weights, we recommend to use the link length. Additional data on parking use (if available) can be used to provide a better estimate. Possible data sources include data on available public and private parking spaces for each link, or data of the number of buildings within a street. Similar to the weights of the access points, these weights can be fine-tuned even further using optimisation algorithms such as SPSA.

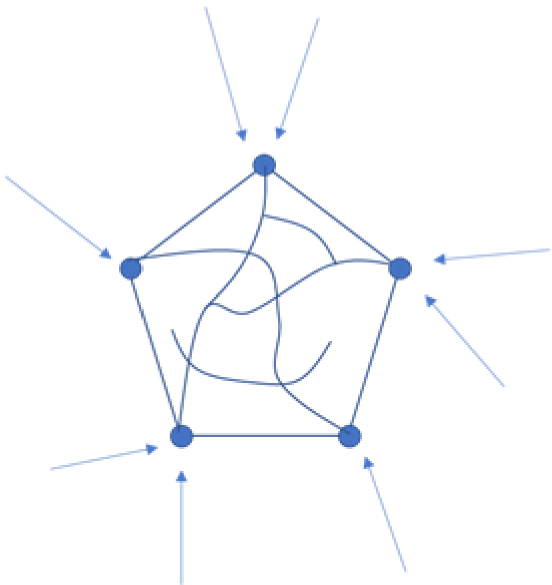

Figure 12. Access points towards a zone.

3.1.3.3 Dynamic traffic assignment model

In Deliverable 4.1 the hybrid DTA model was introduced. The model has been implemented as part of the traffic simulation software, Aimsun Next (Aimsun, 2021). In case that a more realistic representation of the traffic phenomena is desired, a hybrid macroscopic-mesoscopic model may be a suitable approach for the purpose of evaluating the new mobility services at a strategic level. The service area can be modelled with a mesoscopic traffic flow resolution in order to represent more accurately the prevailing traffic conditions due to the background traffic (demand for conventional modes), whereas the rest of the network can be modelled as a macroscopic model for higher computational efficiency. The benefits of the hybrid approach were discussed in detail in Deliverable 4.1.

This section provides general guidelines and insights on the type of data required for building, calibrating, and validating DTA models in order to capture the real traffic phenomena, focusing on presenting the requirements for building a road transport model in Aimsun Next as well as the calibration procedure for the hybrid macroscopic-mesoscopic DTA model. DTA models consist of a number of supply and demand models and parameters that need to be defined. Calibration and validation techniques for DTA models are well-established in the literature (Balakrishna, 2006).

The main data requirements, calibration and validation steps of a network model are:

Data collection and preparation: Involves the collection of data required for the development of the network graph and network model, including the definition of the study area to be modelled, collection of available traffic measurements for the desired analysis periods (e.g., day, morning-peak, etc.), available OD matrices, traffic control information, public transport data.

Network building: Involves coding the road network and verification. After the network has been created, a standard network review process should be carried out to ensure consistency and identification of any errors related to network geometry, road type configurations, traffic signals, etc.

Network model calibration: Involves adjustment of network supply and demand (origin-destination matrices, route choices) parameters to reproduce traveller behaviour and traffic conditions as close as possible to the observed ones. Specific calibration targets need to be set as well as the calibration of model parameters such that the road capacities, route choices match the observed measurements to an acceptable level.

The estimation of travel demand, usually expressed as origin-destination (OD) matrices, is an important and challenging process in the calibration of transport models to capture the underlying travel demand patterns including the route choices throughout the network. The estimation of OD matrices in Aimsun Next can be performed for static as well as dynamic (time-dependent) OD matrices, depending on the simulation type that is used (macroscopic, microscopic/mesoscopic) and availability of prior demand matrices. In particular, the demand calibration process for the hybrid model consists of the three following steps to adjust static prior matrices to better fit observed traffic conditions and split the matrix into a set of profiled matrices:

1. Macroscopic OD adjustment – This step adjusts the static (e.g., hourly) OD matrix to fit the observed data (usually vehicle counts) available for the same time-period using a macroscopic model.
2. Macroscopic OD departure adjustment – This step splits the hourly matrices into a set of finer profiled matrices (e.g., 15-minute), maintaining the total number of trips for each OD pair.
3. Dynamic OD adjustment – This step adjusts the time-varying OD matrices (e.g., with a 15-minute profile) based on time varying observed traffic data using the hybrid (macroscopic-mesoscopic) network loading.

Validation: Involves statistical comparison of the simulated output with observed data. Depending on the type of ground truth data that are available, comparisons can be performed for select route travel times, link/turn traffic counts, link speeds and travel times, etc.

Table 19 summarises the required data to be collected and prepared in order to create a network graph and build a network model when hybrid DTA is to be used.

Table 19. Input data requirements and approach for calibrating and validating the hybrid DTA model.

Model	Type of data	Approach
Hybrid DTA	**Calibration**	
	Network supply	• Creation and coding of road network elements representing the road network geometry: sections, nodes, turning movements. • Definition of geometric characteristics • Section capacities • VDF, TPF, JDF functions (for macroscopic models)
	Demand	• Aggregate time-sliced OD matrices per vehicle class or input flow patterns at entry sections to the road model and turn percentages at intersections • Specification of the DTA models which control route choice
	Individual vehicle behavioural models (for microscopic and mesoscopic models)	• Specification of car following, lane change, gap acceptance models
	Traffic control	• Specification of intersection control types • Traffic signal types and control schemes (phasing, timings) • Specification of motorway metering types and metering plans • Specification of associated detector locations
	Public transport data	• Public transport schedules • Geographic information (stops, lines, etc.)
	Validation	
	Network model performance	• Congestion propagation at bottlenecks • Volume, speed, density from detectors • Travel times along major paths

3.1.4. Sustainability modules

3.1.4.1 Emission module – Static emission model

The emission module does not require additional data on emissions for calibration. We expect that the accuracy of the module will mainly depend on the accuracy of the modelled traffic flows, and less on the accuracy of the estimated emission factors. Therefore, similar recommendations as for the traffic assignment are applicable here.

3.1.4.2 Emission module – Dynamic emission model

The proposed dynamic emissions model can be used to assess the impact of new transport modes and services with respect to the network emissions. The estimation methodology and model specifications were presented in Deliverable 4.1, where example speed-emissions curves were build based on available emissions data and dynamic speed profiles on the Madrid M30 Highway pilot site as part of the C-Roads project (https://www.c-roads.es/).

The emissions model is data-driven and consists of the estimation of a functional relationship to predict the pollution levels and provide emission indicators given available traffic emissions and speed measurements over time. As might be expected, the estimated curves presented in Deliverable 4.1 are not transferable across different cities and countries, as the main factors (road characteristics, traffic conditions, driving style, vehicle performance, etc.) affecting the level of emissions can vary significantly across different networks. Hence, the application of the dynamic emissions model to any other city requires the adjustment of the speed-emissions curves so as to fit the available sensor data on pollution indicators (CO_2, NO_x) and speeds, provided per vehicle type and engine combustion type (diesel and petrol) for the specific area of interest. We refer the reader to Deliverable 4.1 for a more detailed description of the speed-emissions curves estimation process.

Furthermore, a combination of data-driven and simulation-based approaches for estimating the emissions can be more adequate for large-scale urban road networks. For this purpose, the estimated emissions curves can be integrated in a simulation environment by adding them in the Aimsun Next traffic simulation software. The simulation-based analyses can provide emissions estimations, beyond the observed data, generated by each vehicle in all sections of the road network, based on dynamic speed profiles as well as vehicle type within the fleet. The aggregated results can provide the total network-wide emissions impacts under various scenarios with respect to the network demand and supply.

In this case a calibrated and validated transport network model as well as the percentages of vehicle types in the fleet are required for the simulation analysis. Depending on the scope of the analysis the model can be calibrated for a microscopic, mesoscopic, macroscopic, or hybrid traffic flow resolution, following the calibration and validation recommendations presented in Section 3.1.3.2 for the traffic assignment model.

4. Benchmark capabilities of the enhanced transport simulation software against current solutions

Many cities and transport authorities rely on the utilisation of traditional four-step modelling approach for strategic transport planning applications. However, the conventional strategic models are not suitable for modelling shared mobility systems for which there is a growing interest. Towards this goal, the MOMENTUM project contributes with the development of new and enhanced models in terms of the demand and supply representation and processes as well as the incorporation of various explanatory and predictive models. In Deliverable 4.1, the extended modelling framework was presented, which combines the desired principles of the disaggregate approach of the agent-based models within the traditional strategic transport modelling approach. Section 2 explained the potential of the enhanced modelling framework and presented its integration workflow into state-of-the-art transport models, enabling the incorporation of emerging mobility solutions.

The adoption of the integrated models and algorithms by state-of-the-art transport strategic transport models like Aimsun Next, would allow to plan and evaluate developments of innovative mobility services and technologies prior to their implementation. In this chapter, the technical assessment of the enhanced model capabilities to replicate phenomena required for emerging mobility solutions is presented.

4.1. Improvement of the Aimsun Next software

The Aimsun Next traffic simulation software (Aimsun, 2020) is used as an example to demonstrate the enriched capabilities of state-of-the-art transport simulation software after the integration with the supply and demand models and algorithms as developed within the MOMENTUM project.

More specifically, some of the new models are developed inside the Aimsun Next software, while other models and algorithms interoperate with the software. Table 20 summarises the improvements of Aimsun Next for the models that are developed inside the software, which are the hybrid DTA, the shared mobility service simulation platform Aimsun Ride as well as the dynamic emission model. The new capabilities of the Aimsun Next software are compared against its current solutions. On top of that, the integration with the other models developed in the MOMENTUM project, offer significant capabilities to the simulation software, even though they are not implemented directly inside the software. Their interoperability was described in Section 2. The potentials of the enhanced transport simulation software with respect to the strategic planning and evaluation of emerging mobility systems are presented on Table 21. The enhancements are compared to the current capabilities of the Aimsun Next traffic simulation software, however, the improvements are general and relevant for other state-of-the-art transport simulation software.

Table 20. Enriched capabilities of the Aimsun Next software after the integration and implementation of the models developed within the MOMENTUM project.

Model	Description	Aimsun Next software technical assessment	
	Type	Current capabilities	Enhanced capabilities
Hybrid DTA	Dynamic route choice model with mesoscopic network loading model in specific areas of interest, whereas static macroscopic network loading model in the remainder of the network.	Simulation using one traffic flow resolution (microscopic, mesoscopic, macroscopic) or Hybrid model (mesoscopic-microscopic)	Integration of mesoscopic and macroscopic models enabling the simulation of large-scale networks at macroscopic level, while the service areas for shared mobility services can be modelled with more detail at a more disaggregate level (mesoscopic).
Shared mobility service simulation platform (Aimsun Ride)	Simulation platform to assess and simulate trip plans for shared mobility systems and provide KPIs at a service and user level.	Limited to simplified fleet operational algorithms and type of shared mobility services	Integrated with new, more advanced fleet management operational algorithms for different shared mobility applications (DRT, car sharing, bike-sharing, carpooling).
		Limited to free-flow or simulated travel times from Aimsun Next	Utilisation of historical travel times (e.g., from other simulation software, or available travel time measurements).
Dynamic emission model	Calculation of network-wide emissions using estimated speed-emissions functions per pollutant and vehicle type.	Different emissions modelling models available	The new data-driven model is integrated with the Aimsun Ride platform to derive emissions for the shared mobility trips per vehicle fleet type.

Table 21. Interoperability and enhanced capability of Aimsun Next after the integration with the models developed within the MOMENTUM project.

Model	Description	State-of-the-art transport simulation software capabilities	
	Type	**Current capabilities**	**Enhanced capabilities with MOMENTUM**
Synthetic population	Generation of a synthetic population using sociodemographic and other data	Limited to the traditional four-step modelling approach for the generation and distribution of aggregate demand per mode.	Disaggregate OD demand (individual trips) as input to the simulator.
Disaggregate mode choice	Estimation of modal split between conventional modes and shared mobility services	Limited to the aggregate mode choice model of the four-step model. Simplified disaggregate mode choice model for conventional and shared mobility modes inside software.	Disaggregate demand trips with their mode choice as input to the simulator.
Data-driven shared mobility demand	Demand estimation for shared mobility	Not available or integrated in the software[1]	Possibility to update the synthetic population model and fed the synthetic individuals directly into the fleet management model
Fleet management	Strategic planning and fleet operations	Fleet operation applications with Aimsun Ride	Possibility to combine strategic level planning with operational level (dynamic interactions)
	Operational algorithms for shared mobility systems	operational algorithms to execute trip plans of requests for shared services	Advanced operational algorithms that provide optimised trip plan solutions for the execution of the demand requests
Induced demand	Estimate demand elasticity	Not available or integrated in the software[1]	Possibility to update the synthetic population due to demand elasticity.
	Demand estimation for roundtrip station-based car sharing systems (specific use-case)	Not available or integrated in the software[1]	Possibility to adopt the method if data for a new application are available.

Car ownership	Estimation of car ownership at different geographic levels (household, traffic zones)	Not available or integrated in the software[1]	Possibility to enhance the synthetic population with new or more accurate information.
Traffic assignment (static)	Improved traffic assignment model, accounting for intrazonal trips suitable for shared mobility systems.	Traffic assignment models available.	Possibility to enhance the fleet planning and operational model by providing more accurate travel times. Suitable when a DTA model is not available.
Emissions (static)	Detailed emission model considering geo-spatial dimension to produce better traffic emissions estimates	Static and dynamic emission models available	Possibility to obtain better traffic emissions estimates for macroscopic emissions estimation approach.
OD clustering and classification	Selection of representative OD matrices given specific characteristics (day, season, etc.)	Similar models available, but not integrated into the software	Depending on availability of historical OD matrices, derive input OD matrices for the synthetic population and traffic assignment models.

[1] It should be noted that the description "Not available or integrated in the software" refers to Aimsun Next. Typically, these models are not available in state-of-the-art transport simulation frameworks.

5. Conclusions

In this deliverable, we have presented the integration of the models and algorithms developed in MOMENTUM project into a state-of-the-art transport simulation framework, which can properly capture and simulate the impact of emerging mobility concepts and solutions.

The main outcomes of this deliverable could be summarised as follows:

1. An integrated transport simulation framework, consisting of the models developed in MOMENTUM , which satisfies the various developments required to model and assess the different aspects of emerging mobility solutions.
2. A detailed description of the integration design and process as well as the data interaction between the various models that are developed within the MOMENTUM project.
3. Useful recommendations and guidelines of the different models integrated into state-of-the-art transport simulation software that can be adopted for the application of the enhanced framework for modelling emerging mobility solutions.
4. A reference example of the enhanced capabilities of the integrated framework against the current solutions of state-of-the-art transport simulation software is provided using the Aimsun Next traffic simulation software.

The application of the enhanced transport simulation framework to the four case study cities in the MOMENTUM project as well as the evaluation of the results, will be presented in WP6.

References

Aimsun (2020). *Aimsun Next 20 User´s Manual.* Aimsun Next Version 20.0.2, Barcelona, Spain.

Balakrishna, R. (2006). *Off-line calibration of dynamic traffic assignment models* (Doctoral dissertation, Massachusetts Institute of Technology).

Bierlaire, M. (2020). Biogeme.

Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Resampling methods. In *An introduction to statistical learning* (pp. 175-201). Springer, New York, NY.

MARG (2016). *PopGen: Synthetic Population Generator*. Mobility Analytics Research Group. http://www.mobilityanalytics.org/popgen.html.

Spall, J. C. (1987, June). A stochastic approximation technique for generating maximum likelihood parameter estimates. In *1987 American control conference* (pp. 1161-1167). IEEE.