D4.1: Interface to modal choice model

Deliverable ID:	**D4.1: Interface to modal choice model**
Dissemination Level:	**PU**
Project Acronym:	**Modus**
Grant:	**891166**
Call:	**H2020-SESAR-2019-2 SESAR-ER4-10-2019**
Topic:	**ATM Role in Intermodal Transport**
Consortium Coordinator:	**BHL**
Edition date:	**27 July 2021**
Edition:	**00.01.00**
Template Edition:	**02.00.02**

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Annika Paul / BHL	Project Coordinator	23/07/2021
Damir Valput / INX	Project Team	23/07/2021
Pablo Hernández / INX	Project Team	23/07/2021
Andrew Cook /UoW	Project Team	23/07/2021
Pierre Arich / ENAC	Project Team	23/07/2021
Hamid Kadour /ECTL	Project Team	23/07/2021

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Annika Paul / BHL	Project Coordinator	26/07/2021
Damir Valput / INX	Project Team	26/07/2021

Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Annika Paul / BHL	Project Coordinator	26/07/2021

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
00.01.00	28/07/2021	Submitted	Modus consortium	Submitted for review by the SJU

Copyright Statement

Modus

MODELLING AND ASSESSING THE ROLE OF AIR TRANSPORT IN AN INTEGRATED, INTERMODAL TRANSPORT SYSTEM

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 891166 under European Union's Horizon 2020 research and innovation programme.

Abstract

This deliverable is the first deliverable of WP4 of the Modus project which aims to develop highly detailed low-level results on the present and future of the mobility of passengers in Europe based on flight and passenger metrics. The purpose of this document is to describe the methodology designed and developed to translate the output results of the modal choice model into individual passenger itineraries that are going to be used by the mobility models. Additionally, it outline so-far identified data requirements and processing needs to create valid input for the rest of the models developed in Modus: flight-centred airside model RNEST, passenger-centric airside model Mercury, and the landside model (i.e. door-to-door model).

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Objectives of Modus

In the context of increasing environmental awareness, regulatory measures, capacity shortages across different modes, or the need for a more seamless and hassle-free passenger journey, the future evolution of European travellers' demand for mobility is still unknown, as well as its potential impacts on the European transport system. The optimisation and alignment of multimodal transport is therefore of utmost importance for the overall performance of the (future) European transport system, especially in regard to providing a seamless and hassle-free journey for passengers as well as mitigating (air) capacity constraints. In line with this, the high-level objective of Modus is to analyse how the performance of the overall European transport system can be optimised by considering the entire door-to-door journey holistically and considering air transport within an integrated, multimodal approach. This is pursued by:

- Identifying and assessing (future) drivers for passenger demand and supply of mobility, and how these affect passenger mode choice,
- Applying and further advancing existing models to determine the demand allocation across different transport modes, especially air and rail, and the effects on the overall capacity of these modes, and
- Developing and assessing performance and connectivity indicators which facilitate the identification of gaps and barriers in meeting high-level European (air) transport goals, and solutions to gaps can be addressed.

Modus wants to understand in a better way how air transport management (ATM) and air transport can better contribute to improve passengers' multimodal journeys and how this translates into an enhanced performance of the overall transport system. A multimodal journey from door to door comprises different steps. The focus of Modus within this door-to-door travel chain is on multimodal transport that includes as a main segment either rail or air transport in Europe. Other transport modes such as public transport are considered as access and egress modes (feeder traffic) to either the airport or the rail station.

1.2 Objectives of this deliverable

Modus WP4 is devoted to producing low-level, highly detailed results on the mobility of passengers in Europe considering flight and passenger metrics. Modus deliverable D4.1 has the objective to describe the methodology designed to develop the interface to translate the output of the modal choice model into individual passenger itineraries. The aim is twofold:

- Individual flight data need to be assigned to the schedules generated as an outcome of the modal choice model. Then, the supply and demand flow scenarios of the modal choice model are to be translated into actual passenger itineraries.
- The itineraries need to be standardised to different formats, to be used by the rest of the models used in Modus (Mercury - airside and landside model, as well as RNEST) tailoring the data to their input needs. Moreover, other data needs are discussed for the successful upgrade of these models.

1.3 Deliverable structure and content

D4.1 consists of the following sections:

- Introduction to the Modus project, WP4 and D4.1 context;
- Modal choice model - which provides the description of the fundamental idea and functioning of the modal choice model;
- Mobility modelling (Mercury and RNEST) - providing the overview of the three major mobility models used and developed further in Modus, including both airside and landside component of Mercury;
- Mapping: modal choice model to mobility models - where the block that provides the generation of passenger itineraries using the data generated by the modal choice model is described;
- Data requirements - a section which touches on the particularities of the data needs for running the models, with a particular focus on the landside model data needs which required definition of the passenger profiles that are to be simulated, landside processes (e.g. boarding) and parametrisation of the functions that model those processes.
- Next steps - a short outline of the next steps in WP4.

2 Modal choice model

In a situation of increasing environmental awareness, regulatory measures, capacity shortages across different modes, and the need for a more seamless passenger journey, the optimisation and alignment of multimodal transport in Europe are of the utmost importance for the overall performance of the (future) transport system. Within this context, the objective of the work is to identify the main drivers that influence passenger demand for mobility. We will pay particular attention to the characteristics of air and rail transport, to determine how their respective characteristics impact the passenger's choice for each mode. On the routes where both air and rail are available, we will address what motivates passenger choice for one or the other mode.

2.1 Purpose of the modal choice model and planned output

Our objective is to **estimate the demand function for air and rail services on European city pairs** where both are available. The data are collected on rail (MERITS) and air (OAG) databases for 2016, September 2017 and 2019. These data correspond to the supply side of the market and provide information on capacities supplied. The information on demand will come from the FRACS Air Transport Database for air and from average load factor of rail operators.

The final objective of our analysis is to estimate air and rail respective market shares on the city pairs where they compete. The model will give insights on the substitution path between these transport modes, depending on several of their characteristics. We will provide some elasticities of demand of the drivers included into the demand function. These elasticities will give some measure of demand sensitivity to any change in these drivers. Then, the estimated model will allow assessing the main determinants in the choice of transport modes.

We use the traditional **econometrics methods on panel data** to estimate the demand function. To **validate** our methodology, we will perform the tests that are required when using panel data, in particular the non-stationarity test. We will include in the demand function the drivers of demand identified, including supply attributes and proxy for passenger characteristics. These proxies are the usual socio-economic variables that are used in transportation literature and have been shown to impact demand.

3 Mobility modelling

The following section will present the mobility models to be used in the Modus project.

- Passenger air mobility modelling: Mercury
- Research Network Strategic Tool (R-NEST)
- Landside model: modelling of passenger door-to-door itineraries

These mobility models will be fed by data of individual passenger itineraries obtained from the modal choice interface. It should be noted that the development of the models will be carried out in parallel and independently of each other, with no communication of results between them. The development of the three models is intended to provide a 360-degree low-level, scalable and integrated model of air and ground passenger transport in Europe, along with passenger mobility metrics that enable quantitative insights into the performance of the European transport network.

3.1 Passenger air mobility modelling: Mercury

Mercury allows for fast time simulations under various scenarios, tracking each passenger almost individually throughout one day of operation, the results of which can be used to assess the performance of the system — i.e., how the overall state of network changed after introducing a certain mechanism into the system. The current version of Mercury, developed during the ER3 Domino project and coordinated by the University of Westminster, relies on a very extensive dataset merged from various data sources. Among others, these datasets include information on airport curfew times, European airspaces and airports, estimates of compensation costs under different circumstances, passenger itineraries and flight schedules, and Base of Aircraft Data (BADA), which provides information on the performance of various aircraft models. Ad hoc distributions (for delays, taxi times etc) and causal links (e.g. if the aircraft is delayed by more than 5 minutes, speed up) are used to run the simulations.

The simulator has evolved a lot through the years, and has now the following features which are relevant for Modus:

- Individual passenger tracking through a day of operation, including missed connection, compensation, duty of care etc.
- Explicit models of aircraft, allowing computations of reactionary delay.
- Explicit regulations at arrival airports in case of capacity limitations. Arrival managers manage explicit arrival queues.
- Complex decisions from the airlines. Airlines can rebook passengers, change aircraft speed, swap flights, etc. based on a highly detailed cost model taking into account compensation, fuel, ATC charges, curfew fees etc.
- Several exploratory SESAR mechanisms which can be turned on or off. Included are dynamic cost indexing, improved UDPP processes, and extended arrival managers.
- A detailed trajectory description, allowing to compute fuel consumption and CO_2 emissions.

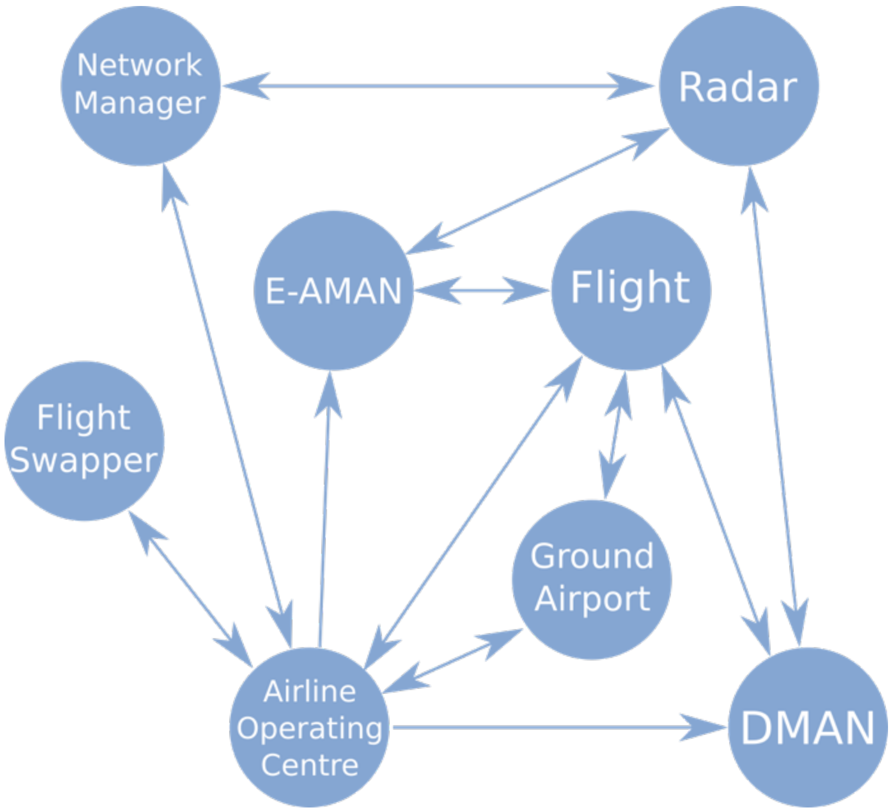


Figure 1 Agents implemented in the current version of Mercury-ABM

Moreover, an extensive set of metrics can be computed by Mercury on the simulation results, including flight departure delay (broken down by reason), passenger arrival delay (and missed connections, rebookings etc.), cost of operation (broken down by type, e.g. maintenance) etc., through which we can assess the impact of simulated mechanisms on different stakeholders (e.g., passengers, airlines, airports). Mercury's scalability and modularity also allows us to incorporate novel modules and mechanisms into the system as needed, such as it will be done with the **incorporation of the door-to-gate simulator** in Modus (more on which you can find in Section 5), as well as to define and calculate any new metrics as needed.

Finally, being a pre-tactical/tactical model, **Mercury requires as input a set of flight schedules linked to passenger itineraries**. The consortium has experience in generating these realistic inputs for the model from strategic schedules and passenger flows as shown in the ER1 project, **Vista**, where 2035 and 2050 scenarios were successfully analysed. For the development of interface to modal choice model and transformation of the data to input suitable for Mercury we rely on this experience as well as already developed piece of software from Vista, which is called *Schedule Mapper*, on which we build in Modus. More details on how this data preparation activity is performed and on Schedule Mapper are provided in Section 4.

3.2 Flight-centric network model: RNEST

R-NEST is a model-based simulation tool, sharing the same base as the EUROCONTROL NEST tool. R-NEST is dedicated to research activities for pre-evaluating advanced ATM concepts. The tool is a standalone desktop application combining dynamic ATFCM simulation capabilities with powerful airspace design and capacity planning analysis functionalities.

R-NEST offers an intuitive, planner-orientated interface with a low barrier to entry for new users. It is a powerful scenario-based modelling engine, capable of running a broad range of complex, operationally relevant analyses and optimization functionalities.

R-NEST can process and consolidate large quantities of data spanning multiple years, but allows to drill down into details.

Simulation Capabilities

Delays

R-NEST **dynamically** simulates network operations and allows detection and observation of delays that characterize the degradation of the network performance.

R-NEST is able to capture:

- **ATFCM** delays, thanks to an emulation of the **CASA** (Computer Assisted Slot Allocation) algorithm used by the Network Manager to respond to network constraints and **an integrated STAM (Short Term ATFCM Measures) model** developed following the guidance of the SESAR concept.
- **Non-ATFCM** delays, mostly generated by internal ATM network disturbances. These delays can result from various causes e.g. handling, passengers or baggage problems. (source: EUROCONTROL/CODA)
- **Reactionary** delays, incurred by delays affecting previous flights using the same aircraft. It is through reactionary delays that problems at one airport propagate through the network.

4D traffic distribution

R-NEST can calculate 4D flight trajectories for a given route network, taking into account aircraft performance data, route restrictions and flight level constraints, SID & STAR and military area opening times. The traffic can be distributed via shortest, cheapest (minimizing route charges) or optimum (minimizing overloads) routes.

Future traffic samples

R-NEST can generate future traffic samples using traffic growth forecasts provided by STATFOR. Airport capacities and curfews can be used to constrain traffic growth.

The EUROCONTROL Network Research unit has developed a tool (FIPS – Flight Increase Process) which allows future traffic samples to be created that completely respect the temporal distribution of the baseline sample (i.e. the same peaks are observed in the demand distribution at each airport) but take into account the planned airport hourly capacities.

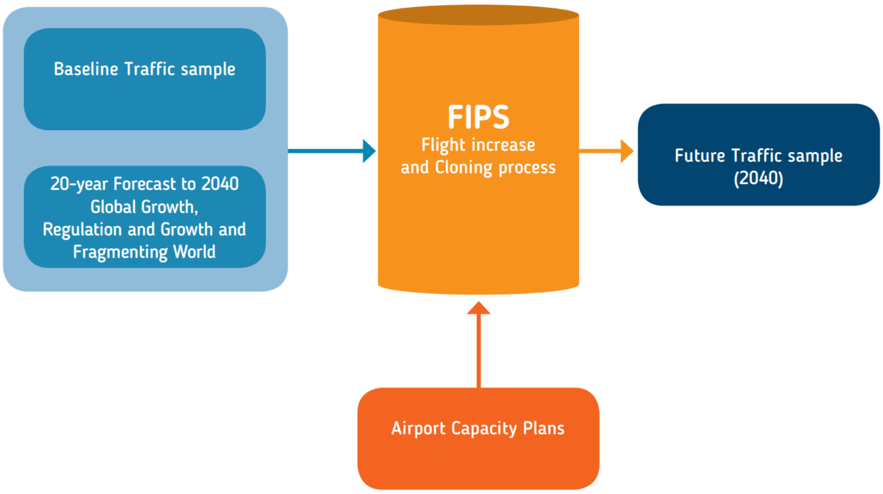


Figure 2 Traffic Increase Process

Future traffic samples are constructed directly from the baseline traffic sample, which in our case will be selected days from summer 2019. Growth figures are then applied to the baseline traffic sample, from the 2040 forecast prepared for *Challenges of Growth.* One out of the four scenarios, *Regulation and Growth* (most likely), will be used to build the future traffic picture.

Another major component of the modelling environment is the airport capacities.

From Local to Network Performance indicators

R-NEST is able to measure at the network level the improvements generated by the local implementation of a new ATM concept.

R-NEST offers global indicators evaluation including route length extension for flight efficiency, fuel consumption, ATFCM/Non-ATFCM delays, CO_2 and NO_x emissions.

R-NEST can process large quantities of data spanning over multiple years. A strategic view allows users to detect trends or carry out detailed analysis.

3.3 Landside model: modelling of passenger door-to-door itineraries

In this section, we briefly present the structure of the landside (door-to-gate) model, followed by a description of its data and parametrisation requirements. The model will be an update of the version produced in the project DATASET2050 [1, 2]

3.3.1 Structure of the landside model

This section presents the **scope** of the **landside** or **door-to-gate model** (passengers included, journey steps assessed, and geographic scope). The details of the **data needs of the input** fed to the model are described in Section 5 of Data requirements.

Within the Flightpath 2050 (European Commission, 2011a) "Meeting societal & market needs" the second specified goal is for "*90% of travellers within Europe are able to complete their journey, door-*

to-door within 4 hours. Passengers and freight are able to transfer seamlessly between transport modes to reach the final destination smoothly, predictably and on-time."

Based on this statement, the model scope will be the following:

3.3.1.1 Passengers

The model includes any type of **air traveller within Europe,** covering all journeys for which air transport has any contribution in the door-to-door segment, even when the air segment is shorter in time or distance than ground/sea transport segment.

Passenger archetypes covered in the landside model

In the landside model, the door-to-door journey components, as described above, may vary according to passenger requirements, preferences and behaviour along the journey. Therefore, different passenger archetypes are considered, which are elaborated in more detail in Modus Deliverable D3.2. The seven archetypes defined within Modus - (1) Business Flyer, (2) Digital Gen Z Flyer, (3) Environment-minded Flyer, (4) Premium Flyer, (5) Cultural Jetsetter, (6) Holidayer, and (7) Golden Senior Traveller - are assigned distinct characteristics in regard to their travel behaviour, which are translated into respective parameters in the landside model.

For example, each archetype exhibits a different value of (saving travel) time, i.e. the willingness to pay to reduce travel time differs by travel purpose or overall travel journey time. Passengers travelling for business purposes are therefore assumed to exhibit a higher willingness to pay to reduce their overall time spent travelling. For the landside model, this implies that passengers with a business background are eager to reduce the airport access and egress time (door-to-kerb) as much as possible, and therefore choose according transport modes, and hence reduce their respective time for this travel segment or component. Another characteristic of each passenger archetype is the assumed frequency of travel, for example. This parameter translates into a high or low familiarity with airport processes, for example, and hence the time required to accommodate with relevant procedures and thus resulting times spent in processes.

Along these descriptions, each passenger archetype exhibits distinct characteristics which can be translated into according parameters for the different components in the landside model.

3.3.1.2 Journey

The landside model aims at uncovering insights into the current situation of the holistic, European mobility. For these reasons, this model adds to the following legs to the air transport system (i.e. air legs) modelled in Mercury, that way **completing the journey**:

- Door-to-Kerb, multi-modal, public/private transport;
- Kerb-to-Gate, includes airport processes, check-in, baggage drop-off, security, immigration and boarding;
- Gate-to-Gate, from boarding to alighting (with connections), including off-block, taxiing-out, take-off, route, landing, taxiing-in and in-block;
- Gate-to-Kerb, from alighting to deboarding, luggage reclaim, immigration and customs;
- Kerb-to-Door, multi-modal, public/private transport.

3.3.1.3 Geographical coverage

Passenger journeys in the model include those within 32 European countries: current EU member states (EU-27 plus Iceland, Liechtenstein, Norway and Switzerland (EFTA). Each mobility process of departing and arriving to and from any pair of EU locations is also included. Trips from a non-EU origin are not considered at this moment, due to the lack of information of the door-gate process. Non-EU destinations are also excluded, given the lack of information in their gate-door processes.

According to the prioritisation of the city archetypes for which the synthetic data is to be produced, we aim at producing various simulations corresponding to various city archetypes, described in Section 4 of this deliverable, and scenarios that are currently being defined in Modus.

4 Mapping: modal choice model to mobility models

In this section we present the objective behind the interface developed between the modal choice model and the mobility models (Mercury and R-NEST). The mapping interface seeks to convert the **high-level flows** of the econometric modal choice model into **future supply and demand scenarios** that can be translated into individual **passenger itineraries** which will then be fed into passenger mobility model Mercury, which required this kind of input to be able to run the simulations, and used as an additional input for R-NEST. Although simple in concept, the conversion of the high-level flows of the economic model into individual itineraries is a very demanding task, not only from a computational point of view, but also because of the large number of combinatorial possibilities and many constraints. The project builds on work done in previous projects, in particular Vista [4] where this kind of module has been built, in order to streamline and reduce the total effort required.

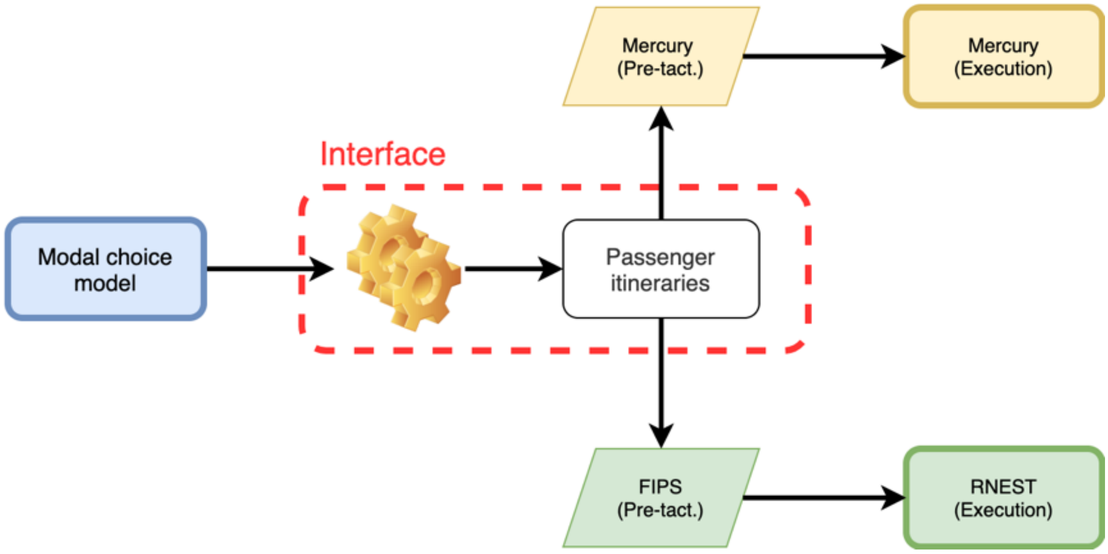


Figure 3 High-level mapping interface

The first task of the modal choice model interface is to convert the high-level flows identified in the modal choice model into **individual schedules**. The main complexity faced in scheduling is due to the large number of possibilities and the multiple constraints. These include among other hard constraints such as crew, aircraft and airport slots. The scheduling process begins with the **data collection** on airports, past schedules and high-level strategic flows. Then the **average travelling times** between OD (Origin-Destination) pairs is computed. In addition, the probable **departure times** are also calculated and the decision tree for the **turnaround times** is developed. These calculations rely on the concept of a "**pattern**", i.e. the successive steps an aircraft takes over a given period of time, as well as how the aircraft of an airline together provide the desired capacity. Patterns can be of two types: either a ‘loop’, i.e. the first and the last airport are the same, or an open pattern. As a simplification, it is established that each airline uses only back and forth for their aircraft as this is the actual majority usage as well as because it considerably improves computing times. Continuing with the scheduling process, for each airline the network is trimmed by removing excess aircraft and/or the network is grown by adding aircraft to meet demand. To decide whether or not an airline should remove current patterns or add some a **'master' cost function** is used for every pattern. This function takes into account the difference between demand and supply on each leg, the price of the tickets, the costs and the number of patterns. Finally, with all this, the new schedules are computed. Once the schedules are available, they are used in conjunction with the modal choice model information on passenger flows to generate individual **passenger itineraries**. This is done in a three stage process:

- **Options for passenger flows:** in this stage the possible options available for the passengers in each flow is computed. In the case where a passenger is making one or several connections, the minimum connecting time required to change between flights is considered when making the computation of possible options.
- **Passenger assignment optimisation:** Once the options are computed the passengers need to be assigned one of the options. One of the main things to keep in mind at this stage is to ensure that the flight's capacities are respected. The objective being to maximise the number of passengers that are assigned to an option while maintaining the number of passengers on the flights lower than the capacity of the flight and not assigning more passengers from an itinerary flow than the volume of that given flow.
- **Passenger 'fillers':** After the optimisation stage, it is possible that some flights have a load factor unrealistically low. This could be due to a possible mismatch between passenger flows, schedules and flight capacity constraints. Therefore, a new target load factor occupancy on flights is set. This may mean that extra itineraries are needed. These will be generated as "fillers" of single leg passengers on those flights. The number of passengers added as fillers is defined by the load factor of the aircraft, ensuring realistic load factors for the flights.

Translating supply and demand data from the modal choice model into passenger-specific itineraries is not an easy or straightforward task because there is usually not enough granularity in the data to achieve this. This is why we must rely on certain assumptions or generalisations when developing mobility models. For this reason, the Modus project has worked on the definition of different **supply and demand scenarios**, as well as a series of archetypes. During the modal choice model development a total of four different demand and supply scenarios have been defined as well as seven **traveller archetypes** (for more details refer to deliverable 3.2).

Similar to the traveller archetypes, and to facilitate the analysis during the development of the interface for the modal choice model, airport and railway connection archetypes have been developed. The aim of this is to create and define city archetypes based on combinations of these two. This **archetype centred approach** was used as it provides the possibility of obtaining meaningful results identifying non-universal patterns that apply to well-defined subsets of scenarios, providing meaningful differentiation and thus avoiding the problem of overgeneralization. For the definition of the **airport archetypes**, we opted for inheriting the framework from the DATASET2050 project [3]. Four airport archetypes were defined, based on current (2015) ACI EUROPE passenger and connectivity data, and recent data from other sources (for more details see Appendix 6.1). The initial airport archetypes classification is:

1. **Main Hub**: This group covers the key hub airports, of which four have been included: London Heathrow, Paris Charles de Gaulle, Frankfurt Main and Amsterdam Schiphol.
2. **Secondary Hub**: This group captures the secondary level of hub airports: Madrid-Barajas, Munich, Rome Fiumicino, London Gatwick, Barcelona El Prat, Paris Orly, Copenhagen Kastrup, Zürich, Dublin, Brussels National, Düsseldorf, Lisbon Portela and Helsinki Vantaa.

3. **Large/Medium**: The third group covers the next tier of busy airports: Oslo Gardermoen, Palma de Mallorca, Manchester, Stockholm Arlanda, Vienna International, London Stansted, Berlin Tegel, Milan Malpensa, Athens International, Geneva International, Hamburg, Málaga, London Luton, Nice Côte d'Azur, Prague Václav Havel, Warsaw Frederic Chopin, Edinburgh, Alicante, Gran Canaria, Stuttgart, Milan Orio al Serio, Cologne Bonn, Budapest Ferihegy, Birmingham, Milan Linate, Venice Marco Polo and Berlin Schönefeld.
4. **National/Regional**: The final archetype covers the remaining 156 EU-28 and EFTA airports ranked in the top 200, which have not already been included in archetypes (1) to (3).

The rail connection archetype have been defined internally in the project and mainly depending on the availability and access from the airports to a high-speed railway (HSR) line. A total of 4 different archetypes have been defined:

- HSR directly connected to the **airport**
- HSR connected to the **city only**
- HSR connected to the **region only** and/or good mainline rail
- **No HSR** in country and not good mainline rail in region

Table 1 Definition of city archetypes

Airport archetype	Railway connections archetype			
	HSR directly connected to the **airport**	HSR connected to the **city only**	HSR connected to the **region only** and/or good mainline rail	No HSR in country and not good mainline rail in region
Main Hub	arch-1	arch-2	arch-3	arch-4
Secondary Hub	arch-5	arch-6	arch-7	arch-8
Large / Medium	arch-9	arch-10	arch-11	arch-12
National / regional	arch-13	arch-14	arch-15	arch-16

Building joint, ***city* archetypes**, instead of focusing on specific airports and/or railway stations, allows us to consider more holistically, movements between 'Paris' and 'London' and the future of such flows, rather than being tied to specific constraints at particular airports, for example. In the model itself, as developed in WP4, in line with the various scenarios under which the model is run, different levels of 'promotion' will be applied to the city archetypes, to reflect improved connectivity in the future, for example by promoting X% of arch-2 city archetypes to arch-1 archetypes (the best connected). This is further elaborated in the parallel Modus Deliverable 3.2 (Demand and supply scenarios and performance indicators). A first evaluation of possible airport/city archetypes for the study has been made, mainly taking into account the availability of data and their variety in order to keep the range of study as large as possible. This first list can be found in Table 1 Please note that this is still an initial assessment and the final selection used in the project may vary.

Table 2 Initial Airport/City archetype selection

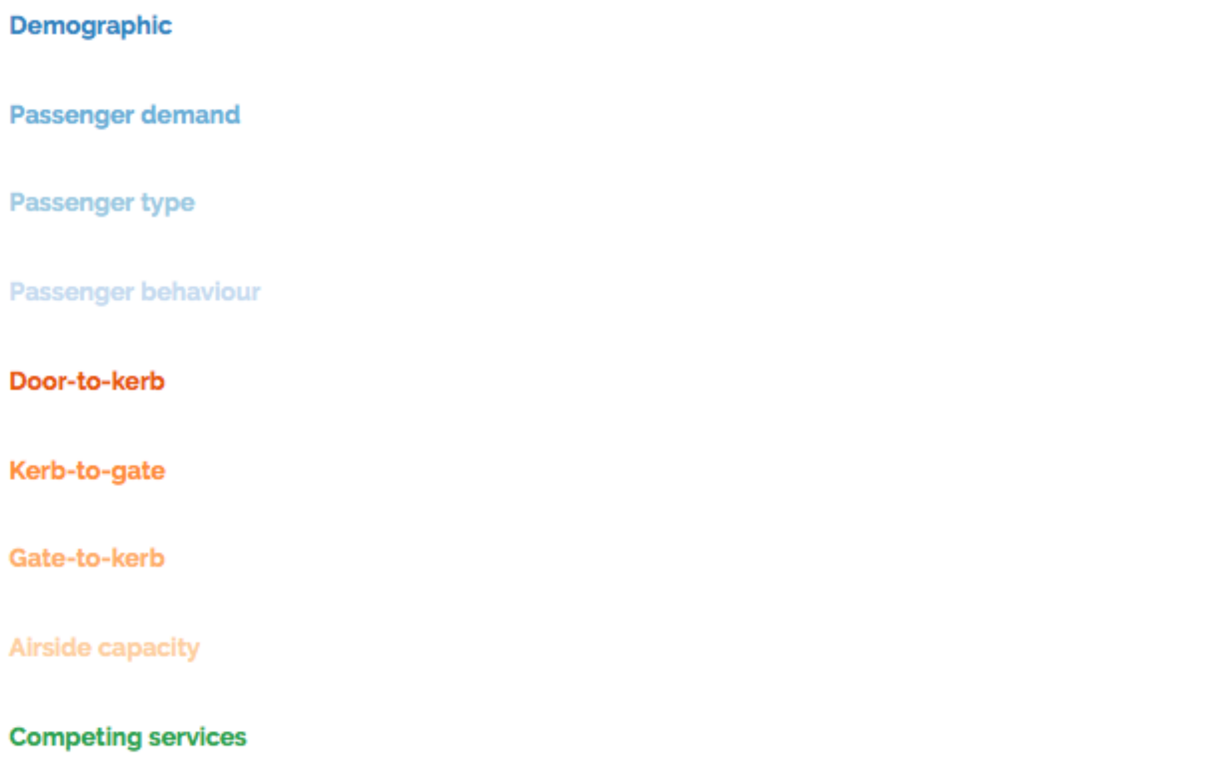
Initial City Archetype classification	Airport	Airport archetype	Railway connection	Railway connection archetype
arch-1	Paris Charles de Gaulle	Main Hub	Good interregional, direct HSR	HSR directly connected to the **airport**
arch-2	London Heathrow	Main Hub	Good interregional (via city centre), no direct HSR	HSR connected to the **city only**
arch-3	Madrid Barajas	Secondary Hub	Good interregional (via city centre), no direct HSR	HSR connected to the **city only**
arch-3	Brussels National	Secondary Hub	Good interregional, no direct HSR	HSR connected to the **city only**
arch-3	Munich	Secondary Hub	Good interregional, no direct HSR?	HSR connected to the **city only?**
arch-4	Stockholm Arlanda	Large/Medium	Good interregional, no direct HSR	HSR connected to the city only
arch-4	London Stansted	Large/Medium	Good interregional (via city centre), no direct HSR	HSR connected to the **city only**
arch-5	Charleroi	National/Regional	Near - good interregional, not HSR (20 mins to Charleroi Sud)	HSR connected to the **region only** and/or good mainline rail

5 Data requirements: landside mobility model

5.1 Data requirements and parametrisation

The datasets used by the current version of the landside model are classified in the following nine groups, covering all the mobility elements required for the model: Demographic, Passenger demand, Passenger Type, Passenger behaviour, Door-to-Kerb, Kerb-to-Gate, Gate-to-Gate, Gate-to-Kerb, Airside capacity and Competing Services.

Each of those groups has a different temporal coverage as it can be seen below. For instance, demographic data is quite abundant but specific Gate-to-Kerb information is scarce; especially forecast data. A picture of the coverage of the datasets can be consulted at http://visual.innaxis.org/mobilityDataSETs/.



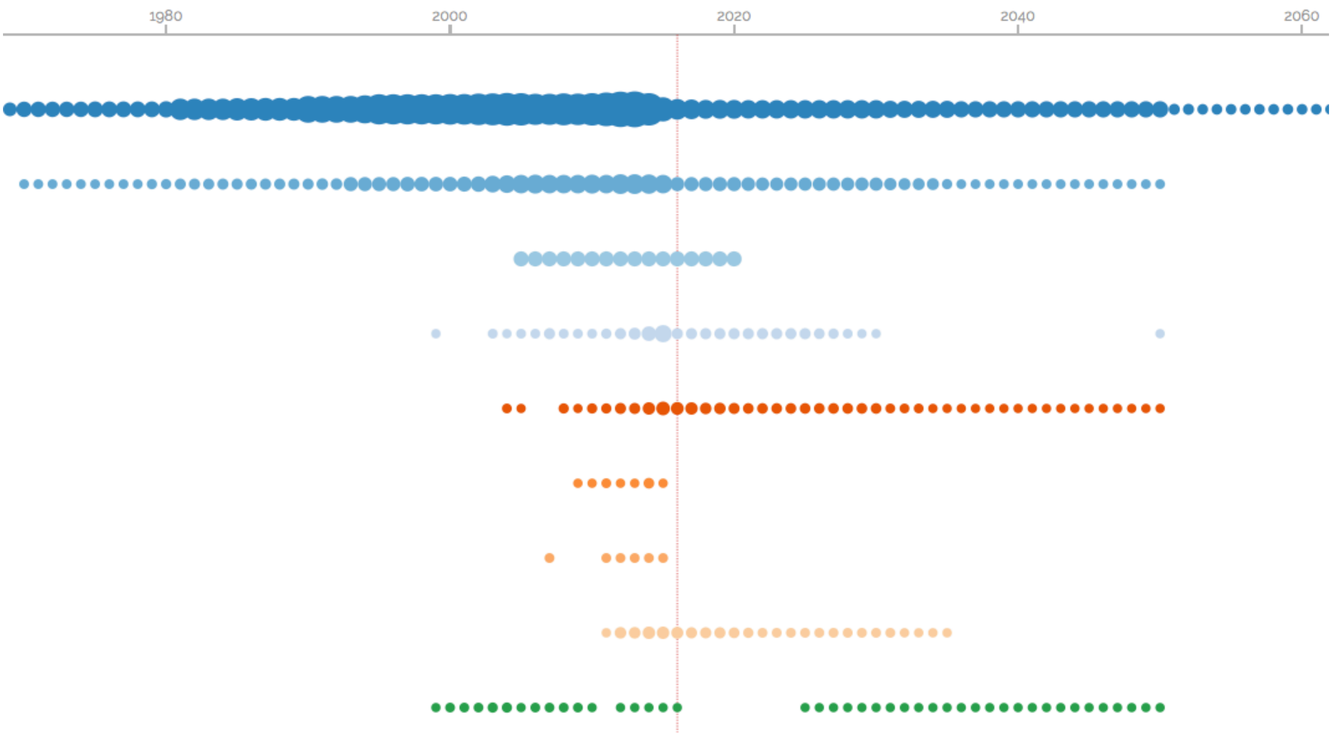


Figure 4 Temporal coverage of the different groups of data sources used in the landside model

Some of these datasets, forecasts and analysis only include information about particular EU regions/airports. Additionally, some of the datasets overlap; showing the same conceptual transport process but from different stakeholders. Also, at times information differs in terms of quantitative figures, depending on the source.

5.2 Stochastic approach adopted in the landside model

This section presents a general model of mobility and the stochastic approach.

If a data-acquisition system had kept track of all air passenger movements from door-to-door since the beginning of the commercial air transportation system, we would have a longitudinal study from their departure door until the utmost final destination for each passenger's journey in the history of air transportation. Each passenger corresponds to a sample from a **random variable** called, "air transportation door-to-door journey length".

Although the concept has been simplified, it is true that very little is known about its distribution probability function. For instance, we do not know details on the distribution shape or average. We do not know if the average has been growing with time (i.e. passengers, on average, spend more time to reach their destinations) or if certain processes (e.g. Kerb-to-Gate) have become a significant burden over the last years.

Instead of considering the "Air Transportation Door-to-Door journey length for each passenger" as a random variable, it would be better understood as a series of time-dependent random variables, or **stochastic process**.

The description of the air transport door-to-door journey length is what it is mathematically called a **non-stationary stochastic process**. If X(t) is a random variable that depends on time, or simply a stochastic process, it would be stationary if all X(t) have the same distribution for every t. To illustrate, rolling the same dice would be a stationary stochastic process, as every time we roll the dice we have the same probability of obtaining the values in range; independently from the actual result of rolling the dice as it is different every time. The important property is that the probability remains constant over time.

This does not happen in the air transportation system, except if the time period under analysis satisfies the following conditions:

- The time period is small enough to negate the effect of long term trends on demand behaviour
- There is no disruptive technology on the supply profile
- There is no disruptive exogenous factor, fuel prices, weather, etc.

The "Air Transportation Door-to-Door journey length for each passenger" can be considered as a **cyclostationary process**. Which is a weaker version of the non-stationary process in which the distribution of the X(t) does vary over time, but only through a cyclic or seasonal manner i.e. there is a T such that X(t) follows the same distribution as X(t+T) for all t. For instance the cycles can be weekly, monthly or even seasonal. A cyclostationary process can be interpreted as a piece-wise stationary process.

Assuming a time frame such that the process is cyclostationary the challenge in describing this door-to-door process resides in finding **the probability distribution function for each cycle**. A *priori* in every cycle it would be desirable that the "Air Transportation Door-to-Door journey length for each passenger" probability density function has the following properties:

1. Only positive values, no reverse time-travel allowed.
2. Long tailed, a small amount of passengers with longer travelling times (i.e. remote EU areas).
3. Possible local maxima around the geometric centres representing each EU country regional passengers.

5.3 Parametrisation using conditional probabilities

Each passenger contributes to the previous distribution by adding one sample point, and the set of all passengers is known in statistics as the (complete) sample space. In addition to this, each passenger has a number of properties or parameters associated, or simply called **factors**. For instance each passenger can be classified according to geographical data, origin-destination or temporal data, as well as other data such as aircraft type, airline operator, or even the archetype of passenger.

Each factor divides the sample space into several subspaces. Only when the factors are not mutually exclusive do the subspaces not intersect, when the set of factors is exhaustive, every element of the sample space belongs to at least one factor. Factors are usually ordered in layers, and each layer should contain a set of mutually exclusive and exhaustive factors. Simple examples of these layers are yes/no factors, e.g. connecting passengers and not connecting passengers.

If we now restrict the random variable "journey length" to certain subset of the passengers, say group B, then the previous probability function P becomes parametrized by B. Formally we are interested in the conditional probability of A being true, P(A) assuming (or restricted to) B, That is usually represented as P(A|B) or simply $P_B(A)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

The concept can be generalized to continuous distributions as well. The formulation is a bit more tedious as it involves the geometry of the subset B as well, but the concept remains the same. For instance when conditioning over a simple X=c set the formula for the probability distribution function becomes:

$$f_{Y|c}(y) = \frac{f_{X,Y}(c,y)}{f_X(c)}$$

In general for any subset B of passengers, the conditional probability distribution function can be represented as $f_B(x)=f(x\text{^}B)/f(B)$. For instance if B represents connecting passengers, that is passengers that do not have the direct flight connection between origin and destination airports, then one would expect the probability distribution function of this passengers $f_B(x)$ to move further right as journey lengths increase notoriously for this passengers.

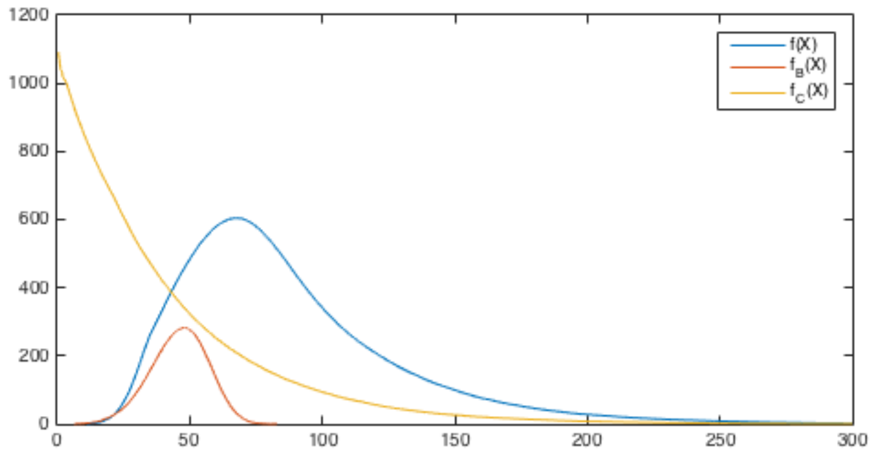


Figure 5 Components: the full distribution f(X) in blue is the aggregation of the distributions fB(x) and fC(X), red and yellow.

However, in reality we don't know the full "journey length" distribution f(x), but rather conditional distributions on a subset B of passengers, namely $f_B(x)$. If we knew $f_B(x)$ for subsets B's such that they cover almost all passengers, the previous formula could also be used forward $f(x\text{^}B)=f_B(x)*f(B)$ to generate the whole "journey length" distribution.

This formula helps to understand how a particular subset of passengers contribute to the overall "journey length" distribution. If the journey length $f_B(x)$ is large enough, even for the smallest set of passengers f(B) the impact could be significant as the total impact is the multiplication of both quantities $f(x^B)=f_B(x)*f(B)$.

Understanding how the different factors $f_B(x)$ affect the total distribution would help to understand where the bottlenecks of the system may be. Bottlenecks can be defined as elements of the system that are hindering the goal of four hours door-to-door and overall performance.

5.4 Door-to-door journey components

Current door-to-door journeys are a succession of (finite) consecutive events. The journey can be divided into phases, with each phase containing its own duration and most importantly do not overlap with other phases. To illustrate, in the DATASET mobility door-to-door model five phases are considered: door-to-kerb, kerb-to-gate, gate-to-gate, gate-to-kerb and kerb-to-door. If X_i represents the random variable "length of the i-th phase" the total journey length would be just the addition

$$X = \sum X_i$$

When looked at individually each random function X_i has an associated probability distribution function $f_{Xi}(x_i)$, however the phases are usually not independent and one must consider the joint probability distribution, $f_{X1,...Xn}(x_1,..., x_n)$ which is the probability of the first phase X_1 being exactly x_1 AND the second phase X_2 being x_2 and so on, simultaneously.

Given the joint probability distribution, $f_{X1,...Xn}(x_1,..., x_n)$ the probability density function of the total journey length X (sum of all X_i) is given by the formula:

$$f_X(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x-\sum_{i\neq 1} x_i} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$$

Only when $X_1...X_n$ are independent random variables, the joint distribution becomes the product $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\ldots f_{X_n}(x_n)$ and the integral can be simplified using convolution. In addition if the variables are identically distributed all functions f_X are in fact the same and the previous integral simplifies greatly. However the different phases in a passenger journey may not be considered in general as independent, as most processes are usually related to each other e.g. passenger missing a flight connection due to traffic on door-to-kerb segment. Neither are identically distributed. It is necessary to reconstruct the whole joint probability distribution, e.g. consider the whole journey length for each passenger, it is not enough to consider each phase separately.

The following five phases have been identified as processes executed as part of a door-to-door journey:

1. Door-to-kerb, surface/underground - The portion of the trip in which the passenger moves from the door of the building of her/his origin (home, office, hotel or any other building unrelated to the trip) to the airport's kerb. The airport kerb is understood as the last point of the mean of transport chosen to get to the airport: e.g.: metro station exit, car parking space, airport taxi stop.

2. Kerb-to-gate, airport side - It begins when the passenger reaches the kerb of the airport (concluding their previous door-to-kerb process) and ends when the passenger crosses the boarding gate door. It involver processes such as obtaining a boarding pass, luggage drop-off, security checks, immigration and customs, and everything else up to boarding.
3. Gate-to-gate, air side → modelled in Mercury
4. Gate-to-kerb, airport side - The phase in which passengers move from the last airport gate to the kerb of that same airport. In case of connections, those will be included in the gate-to-gate process, restricting gate-to-kerb for the very final destination airport.
5. Kerb-to-door, surface/underground - The process in the journey in which the passenger moves from the last airport's kerb to the "door" of the building of her/his origin (home, office, hotel or any other building unrelated to the trip).

The parametrisation of the stochastic processes modelling these processes will be done relying on the synthetic data developed for the selected city archetypes, as described in the following section.

5.5 Landside model parametrisation using synthetic data on catchment areas

Synthetic datasets representing hypothetical routes e.g. in intermodal connections, last-mile mobility, or long-distance road, rail and air routes will be generated for a selected number of city archetypes, by studying their mobility patterns. Those data give information times required to complete a certain route, distance of routes and estimated costs for passenger to complete a route. They will be used to re-parametrise the above described stochastic distributions, leading to a more accurate door-to-door model.

6 Next steps

This deliverable outlines the methodology to use the output of the modal choice model to create data inputs that are going to be fed in the rest of the models we are developing in Modus, in particular the tactical models Mercury and R-NEST. While R-NEST is a more flight-centric model, Mercury requires data on specific passengers itineraries as input, so the high-level flow output from the modal choice model has to be translated into those so to allow the computation of passenger metrics defined in Mercury. For this purpose, a special software module is being developed and its high-level functioning has been presented in this deliverable.

In addition, a great focus is placed on discussing various data and simulation requirements of the landside part of Mercury, i.e. Modus will enrich the Mercury mobility model by providing door-to-door metrics. This part of the model is based on the prototype developed in the past project DATASET2050, led by Innaxis. In line with that, in this deliverable we discussed case studies, i.e. city archetypes which the landside model will cover, and simulation needs to properly run that model under different scenarios that are being currently defined in the project (see deliverable 3.2 "Demand and supply scenarios and performance indicators").

In short, these are the main objectives of the modelling activities planned as part of Modus that will be covered by the upcoming work:

- transformation of the high-level flow final output from the modal choice model into specific passenger itineraries
- improvement of airside part of Mercury (e.g. its forecast capabilities, in particular airspace capacity overloads)
- development of the stochastic landside model (door-to-gate/gate-to-door) of Mercury, based on the available data (historic as well as synthetic) and placing a special focus on the defined city archetypes
- better integration of the landside and airside part of Mercury
- further development of the R-NEST ATM modelling capabilities with a multimodal perspective (airport-to-airport connectivity combining air and rail)
- running simulations of the developed models on the scenarios defined as part of the work package 3

Appendix A Appendix

A.1 Airport archetypes

Table 3 Airport archetype definition [3]

Archetype description				
Archetype characteristic	**(1) Main hub**	**(2) Secondary hub**	**(3) Large/medium**	**(4) National/regional**
Number of airports (percentage of EU-28 & EFTA passengers1)	4 airports (17% of passengers)	13 airports (26% of passengers)	27 airports (26% of passengers)	156 airports (29% of passengers)
Average proportion of transfer passengers2 (range)	41% (31-57%)	17% (up to 36%)	6% (up to 28%)	4% (up to 33%)
Ratio of international:domestic passengers (range of international passengers)	0.93:0.07 (89-100%)	0.84:0.16 (63-100%)	0.79:0.21 (48-100%)	0.62:0.38 (0-100%)
Ratio of intra-EU:extra-EU passengers3 (range of intra-EU passengers)	0.50:0.50 (41-57%)	0.75:0.25 (67-84%)	0.82:0.18 (63-100%)	0.90:0.10 (40-100%)
ACI EUROPE group	Group 1 (>25m pax p.a.)	Remaining Group 1 (>25m pax p.a.); Group 2 (>10 <=25m pax p.a.)	Mainly Group 2 (>10 <=25m pax p.a.); few Group 3 (>5 <=10m pax p.a.)	Group 3 (>5 <=10m pax p.a.); Group 4 (<=5m pax p.a.)
ACI EUROPE connectivity	All 'The Majors'	Remaining 'Major'; mainly 'Secondary Hubs'; small number of 'Niche & Aspiring'/'Challenged' hubs	Most of the 'Niche & Aspiring'/'Challenged' hubs; most without connectivity classification	No connectivity classification; remaining 'Niche & Aspiring'/'Challenged' hubs

1 Top 200 airports account for 97% of EU-28 and EFTA passengers in 2015.
2 Transfer data compiled from various sources; available for approximately 50% of airports in scope.
3 Intra-/extra-EU passengers only available for EU-28 airports.

7 References

[1] D2.2 Data Driven Model, deliverable of the project Dataset2050, Deliverable leader: Innaxis, June 2016

[2] D5.2 Assessment Execution, deliverable of the project Dataset2050, Deliverable leader: Innaxis, June 2017

[3] D4.2 Future Supply profile, deliverable of the project Dataset2050, Deliverable leader: University of Westminster, April 2017

[4] D5.2 Final Assessment Report, deliverable of the project Vista, Deliverable leader: University of Westminster, November 2018