DATASET2050

"Data driven approach for a Seamless Efficient Travelling in 2050"

Coordination and Support Action funded by the European Commission
H2020, Mobility for Growth 2014-2015: MG.1.7-2015, T.1
"Aviation Research and Innovation Policy"
Grant Agreement no: 640353

Deliverable 2.2
"Data Driven Model"

- Work-package: 2
- Deliverable number within work-package: 2
- Delivery date: M19, July 2016
- Dissemination level: PU
- Status: Final version
- Deliverable type: R
- Deliverable leader: The Innaxis Foundation and Research Institute
- Reviewed by: DATASET2050 consortium
- Approved by: The Innaxis Foundation and Research Institute

1. Abstract

The purpose of this document, Deliverable 2.2, is to build the structure and specifications of the DATASET2050 data driven model.

The door-to-door process is complex and therefore direct performance measurement of the process cannot be done due to availability of data and the high number of involved stakeholders. There are additional phenomena that cannot be measured, so the approach to assess performance is by collecting sample data and/or model the different elements of the mobility processes.

This modelling exercise, documented in D2.2, is a powerful tool that assesses how the process performs in the current scenario, and beyond that, identify bottlenecks how modelling paradigms can be improved to take Europe to the 4-hour door-to-door target. The model utilises data that has been sourced, analysed and documented thus far (D2.1), as well as the numerous inputs from the demand and supply profiles (WP3 and WP4, namely: D3.1 and D4.1).

This deliverable documents how the model is to be built, along with its scope and the development strategy.

Table of contents:

Table of figures

2. Introduction

DATASET2050 introduction

DATASET2050, "DATA-driven Approach for Seamless Efficient Travelling in 2050" is a Coordination and Support Action (CSA) funded by the European Commission, under H2020 Call MG.1.7-2014 "Support to European Aviation Research and Innovation Policy", Grant Agreement no: 640353. The Coordination and Support Action is coordinated by Innaxis, with EUROCONTROL, the University of Westminster and Bauhaus Luftfahrt as partners. DATASET2050 was launched in December 2014 and will last 36 Months. The key highlights of DATASET2050 are the following:

- The objective of DATASET2050 is to provide **insights into the door-to-door** European travel paradigm for the current, 2035 and 2050 transport scenarios, through a data-driven methodology.

- DATASET2050 puts the **passenger at the centre**, paving the way for a seamless, efficient door-to-door travelling experience. The main focus to analyse how the **European transport supply profile** (capacity, connections, business models, regulations, intermodality, processes, infrastructure) could adapt to the **evolution of the demand profile** (customers, demographics, passenger expectations, requirements).

- DATASET2050 addresses the **main transport mobility goal** stated in the FlightPath2050: 90% of travellers within Europe are able to complete their journey, door-to-door within 4 hours. Through application of statistical analyses, multi-modal mobility modelling and predictive analytics, DATASET2050 will compute the **current status** of air transport mobility across Europe.

- The analyses will enable the identification of **transport bottlenecks** in the current scenario and across different future scenarios. These findings will serve as a basis for the development of **intermodal transport concepts**; identifying possible solutions for current and predicted shortcomings. The insights gained will highlight **research needs and requirements towards the four-hour door-to-door goal** formulated by ACARE. Due to the multi-dimensionality of the problem, DATASET2050 will use visualisation techniques, to ease the consumption of the results.

- An Advisory Board, formed by **European transport stakeholders** (which includes: Siemens, Schiphol Airport, DLR and OpenStreetMaps), supports the DATASET2050 partners.

- The dissemination and communication plans ensure efficient circulation of results amoung key European transport policy makers and stakeholders. The plans also incorporate their valuable input and perspectives during the project workshops.

WP2 and Deliverable 2.1 and 2.2 context

Work package 2 is tasked to deliver a data-driven model that provides **mobility metrics**.

The metrics on mobility shall meet certain requirements:

- The metrics need to be **quantitative**, and based on existing datasets and/or facts.
- The metrics on mobility have to offer enough **granularity** to assess the situation in Europe across different geographical regions.
- The metrics should cover the different **phases** of the door-to-door paradigm.
- The metrics should allow an analysis that help to understand the limitations and **bottlenecks**, providing information on the phases of flight that are more limiting
- Last but not least, the metrics should reflect **real requirements** coming from operational stakeholders: metrics further usability is crucial

Some of the pointed datasets in D2.1 do not offer enough resolution or coverage to measure the entire passenger mobility in Europe through simple observation. The door-to-door process is complex and involves many different stakeholders that may not collect the data of their processes, may not open their dataset to the public due to its commercial value, may not share the data due to legal/technical constraints, or may store the data in disconnected silos. These data-access challenges are common when assessing the performance of any complex system with a large number of stakeholders and elements involved. Therefore, to allow the assessment of mobility in Europe in a way that supports policymaking, an overall data-driven model needs to be provided.

The data model uses the available datasets available in Europe and explains how DATASET2050 models the five phases of the door-to-door paradigm, including:

1. **Door-to-Kerb:** multi-modal, public/private transport;
2. **Kerb-to-Gate:** includes airport processes, check-in, baggage drop-off, security, immigration and boarding;
3. **Gate-to-Gate:** covers boarding, including off-block, taxiing-out, takeoff, route, landing, taxiing-in and in-block of the flight. This is the "air side" of the process. In the case of connections, the whole process (n flights and corresponding n-1 connecting times) will be modelled in this stage. Or in other words, this is the firstAirport-gate to the lastAirport-gate proccess.
4. **Gate-to-Kerb:** from alighting to luggage reclaim, immigration and customs;
5. **Kerb-to-Door:** multi-modal, public/private transport.

WP2 consists of the dataset collection (already completed in D2.1), along with the data model (as documented in current D2.2). Particularly, D2.2 has been structured as described below.

Deliverable D2.2 structure and contents

- Introduction - current WP2 project along with D2.1 & D2.2 context
- Scope of the data driven model - scenarios, data available, timeframe
- Model requirements and specifications - inputs and output parameters, functions, and requirements capture
- Theoretical approach - trips distributions, parameters, components, dependencies
- Model architecture - explaining how the different trip stages are modelled; door-kerb-gate-gate-kerb-door processes
- Implementation plan - building blocks, model wireframe and verification plan
- Execution - implementation procedure of the data-driven model
- Acronyms, abbreviations, references

3. Model Scope and Overview

This section presents the **model scop**e (passengers included, journey steps assessed, geographic scope), **model time frames** chosen (with the associated heterogeneous levels of granularity) and details of the **available data** that will directly feed the model. The current section expands upon the data model previously reported in D2.1 "Scope" and covers other aspects of the model beyond the data-related aspects.

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:

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.

2. Journey

The DATASET2050 model reveals insights into the current situation of the holistic, European mobility. The model identifies gaps and bottlenecks that hinder the system (*at any of its elements)* from achieving its potential. For these reasons, the **complete journey** (and not just the air leg) is modelled:

- 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, takeoff, 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. Geographical Coverage

Passenger journeys in the model include those within 32 European countries: current EU member states (EU-28) 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.

4. Timeframe, Granularity and Resolution

The project model considers three **time frames**: current time frame (2015), plus two future scenarios aligned with the roadmap of the Strategic Research and Innovation Agenda

(SRIA): mid-term 2035 and long-term 2050.

In particular:

- **Current scenario** (2015): The DATASET2050 model has been designed to realistically represent current mobility paradigm in Europe following current rules, status and procedures. The model will produce highly granulated quantitative outcomes using real, current data. Hence, the model quantitative results will be considered fully reliable in terms of providing mobility conclusions.

- **Mid-term scenario** (2035): For this model a hybrid approach will be used:
    1. Quantitative: The model provides quantitative outcomes (mid granularity), derived from running the model with 2035 demand-supply forecasts. These forecasts will be provided by the future demand and supply profiles analyses (documented in D3.2 and D4.2 accordingly).
    2. Qualitative: This involves an external qualitative assessment, derived also from the D3.2 and D4.2 exercises.

       Rationale: For the mid-term 2035 scenario we chose to use a mixed approach, taking into consideration many aspects. First, data availability for capacity and demand are forecast with unknown uncertainty levels. This is critical as knowing the specific uncertainty levels enables more precise model outputs. Second, the quantitative results from the model will be enriched with qualitative expert assessments on how the different model elements will most likely evolve. These supplemental assessments provide a holistic perspective. Third, for longer time spans, mobility may be influenced by social, political, technological and environmental factors, provoking structural changes in the mobility model paradigm. This extra source of uncertainty needs also to be considered when exploring longer-term mobility scenarios.

- **Long-term scenario** (2050) Due to the uncertainty in technology evolution, new policies implementation and unexpected socio-political factors, we take for granted that the mobility scenario by that time cannot be simply extrapolated from current mobility model/metrics/paradigm. Some questions remain open regarding the evolution of air transport vehicles, control systems (e.g. unmanned vehicles), transport management operational concepts and even new competition from new forms of surface transport that may start operating at longer distances. Quantitative mobility forecasts aiming to be accurate by that time frame are difficult to construct. For this reason, the model is expected to produce only high scale mobility figures in terms of mobility demand and supply metrics. Those figures will be completed by qualitative assessments based on existing future forecasts on how the mobility paradigm may evolve.

5. Data Availability

The datasets previously identified in D2.1 will be used to strengthen the model, either in terms of actual current data or future forecasts. The datasets were classified in the

following 9 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/.

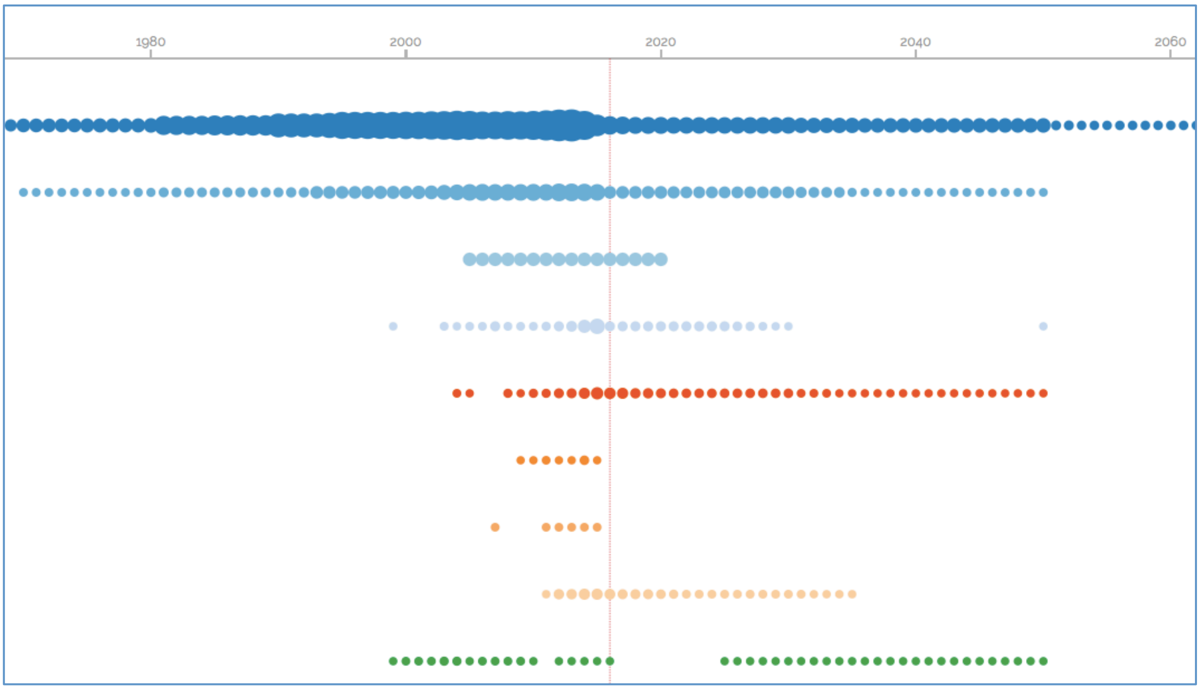


Figure 1. Datasets available (via D2.1)

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.

The following screenshot illustrates an example of the Passenger-demand group of datasets. In the case of the image: the EC EUROSTAT Passenger demand dataset. Equivalent information is displayed when the user selects any of the hundreds of datasets documented in D2.1.

Figure 2. Passenger demand datasets (example)

3.1 Model Requirements and Capture Methodology

The DATASET2050 door-to-door mobility model responds to the need to holistically assess the door-to-door mobility in Europe involving air transportation. The current, present status along with different future potential scenarios (mid term 2030 and long term 2050) is modelled, including their corresponding levels of technology evolution, policy implementation and forecast traffic growth scenarios.

The **current mobility assessment** must analyse the current status of mobility as well as the current limitations for mobility in Europe. This analysis will have **different dimensions**: phase of transport, regions, origin-destination, traveller type, etc. Air transport stakeholders and policy makers will use the assessment and the goal is to be used as a guide to drive technology programmes and appropriate policies to improve air mobility across Europe. The details of DATASET mobility model implementation will capture the multi-dimensionality of the challenge.

It is key to specify the **key functionalities** that potential users (e.g. transport stakeholders or policy makers) would need implemented as part of the DATASET2050 model and performance framework. A solid **requirements assessment** is guarantee of usability and usefulness of the outcomes of the project.

The following is a non-exclusive table that summarises the potential use cases of the DATASET2050 model and performance framework, along with stakeholders that would make use of the functionality:

Use Case	Stakeholders	Main objectives and functionalities
Assessment of current mobility across Europe, across phases	Transport policy makers - DG MOVE, Directorate C.	Development and follow-up of the implementation of Innovative and Sustainable Mobility Policy.
Evaluation of technology programmes in mobility metrics	Research policy makers - DG RTD	Assessment of SESAR technologies on mobility metrics
Evaluation of technology programmes in mobility metrics	Research policy makers - DG RTD	Assessment of CleanSky technologies on mobility metrics
Understanding of main barriers to mobility	SESAR deployment (focusing on gate-to-gate)	Assessment of main barriers and opportunities to impact on mobility in the gate to gate processes
Assessment of current mobility within countries or regions	National departments of transport; regional associations; 9 TEN-T corridors (DG MOVE)	Development and follow-up of the implementation of Innovative and Sustainable Mobility Policy.
Assessment of airport processes across Europe	Airports (perhaps through Airport associations). Airport technology service providers. Airlines. SESAR (A-CDM)	Provide an assessment and best practices on what would be the ideal airport processes in the context of door-to-door mobility.
Assessment of accessibility to airports	National & regional transport authorities.	Implementation of tools/ policies/practices that increase the airport accessibility at local levels
Assessment of safety-security in the context of mobility	EASA, Safety-security regulations, airports	Better understanding the safety aspects under door-to-door mobility paradigms. Mobility datasets tentatively within the scope of safety and security research
Understanding, reporting and forecasting EU mobility	EUROCONTROL STATFOR, EUROCONTROL PRU. Air and surface mobility stakeholders	Widening the scope and business opportunities of stakeholders at any transportation stage

The requirements for the model are identified through the following actions:

- Internal consultation with the DATASET2050 partners, leveraging the work already completed during the DATASET proposal preparation, maximising time and effort already allocated in the project.
- Internal consultation with the DATASET2050 advisory board members and consortium network of entities involved in mobility.
- Inputs and feedback received during DATASET2050 Mobility Workshops (First workshop: 12 July 2016, London; subsequent workshop to be held late 2016/2017)
- Consultation with the ACARE WG1 (Mobility) Group.
- Direct conversations with stakeholders identified in the second column of the table above (SESAR, CleanSky, EUROCONTROL, EASA, EC)

The detail of the requirements will be documented during the model implementation, in WP5.

3.2 Expected Output

Different potential uses of the mobility model will require different desired outputs. The potential outputs will therefore range from single pan-European metrics to more specific indicators, focusing on the multi-dimensionality of the problem. The DATASET2050 model adapts to the different uses, thus providing a bottom-up explanation on how micro and macro metrics are constructed and how different input parameters may affect the outputs. Given the complexity of the model a user interface layer is provided in front of the simulation engine. Interactive visualizations, e.g. using D3.js or similar tools, would allow potential users to use the model without knowing the details of its implementation/technicalities.

Once the user requirements are captured and defined, the expected outputs and assessment will be fully documented within the following project deliverables:

- D51 **Mobility assessment**- as the mobility metric would be tailored to the user requirements
- D52 **Assessment execution**- for the implementation of the model and outputs
- D53 **Novel concept foundations for European mobility**- to include in the report using the previously defined model outputs

4. Theoretical Approach

This section establishes the theoretical framework of the DATASET2050 door-to-door mobility model. This section presents a general model of mobility and the stochastic approach. More specifically, key elements of the stochastic approach are outlined, such as as stability, time frame and predictability. Main limitations are also highlighted.

Furthermore, two fundamental tools are discussed: parametrisation using conditional probabilities, and journey components separation using Markov chains.

These elements will drive the development of the final DATASET door-to-door mobility model (DDMM), and its advantages and limitations are exemplified through various scenarios when possible.

Global door-to-door journey length distribution

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.

FlightPath2050 sets a goal for this door-to-door journey length, asking European mobility stakeholders to research new ways of managing mobility in the mid-term (next 35 years) so that 90% of travellers within Europe are able to complete their journey, door-to-door, within 4 hours.

However, it is clear that the air transportation system, just as many other socio-technological systems, evolves and changes with time. The system changes in both longer time scales, for instance due to new technologies, procedures, business models or infrastructures; and shorter time scales due to situations such as unforeseen strikes, accidents or seasonality-driven issues.

The air transportation system as the **supply** changes, but also because the **demand** (passenger behaviour) evolves with time, with new types of passengers and travel preferences. Changes in the demand stimulate subsequent changes in supply, for instance closing routes due to poor demand or economic activity. Alternatively, changes in supply such as new flight routes or even a new airport, can also provoke changes in the demand as passengers may have access to new destinations.

Exogenous factors have a systemic impact as well, with fuel prices being the best example of this.

Therefore, 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**.

Moreover, the shape of the distribution of the various door-to-door journey lengths evolves with time. For instance new technologies, new procedures or an exogenous factor may shorten the average trip length (such as an airline operating a new direct OD route, allowing passengers to skip connections in a hub). In other occasions these systemic changes could prolong the travel time (for instance when an airport terminal is partially closed due to works).

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.

The first property implies that the domain of the probability distribution function is the positive real numbers and the second property implies that the distribution can not be symmetrical, whilst the third implies certain conditions over the modes (local maxima) of the distribution and hence the mode.

These properties are somewhat vague and insufficient to determine the distribution type for the process. Different distributions showing those properties exist, like the Weibull distribution showed in **Figure 3** - note that the values on x and y-axis have been deliberately omitted, as there is no numerical evidence in the previous properties.

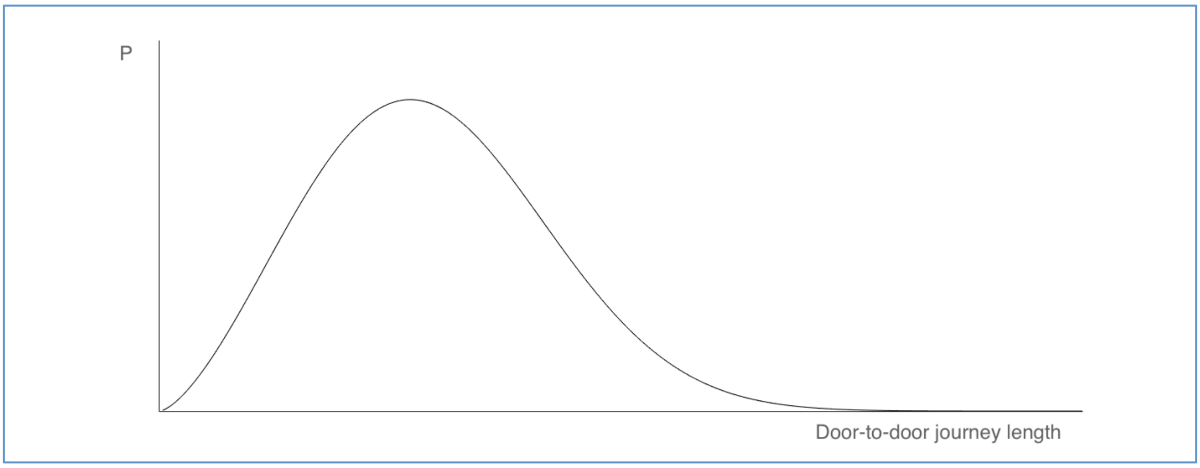


Figure 3 Example shape of a Weibull prob. distribution function

If we knew the distributions and its specifics, then the Flight Path 2050 four hour door-to-door challenge would be implemented through new technology and processes that could modify the current distribution so the 90% percentile is on the four hour range, as shown in **figure 4.**

To be more precise in the Flight Path 2050 goal, it is assumed that this goal has to be achieved in every cycle of study. Although it seems reasonable to assume that cycles affect only numerical parameters of the model (e.g. averages, variances, etc.) and not the family of distributions used, the Flight Path 2050 goal is actually set for 2050, which is too long to consider the stochastic process cyclostationary. That is, no statistical study with current or past data can be ensured to be representative so far in the future.

In any case, the DATASET2050 model focuses on analysing the current processes, which is up to date mostly unknown from a statistical point of view. Then the model takes into account future scenarios- around 2030 and 2050. An accurate understanding of the current process will help define indicators and highlight areas in which to focus development or the major bottlenecks to improve the process.

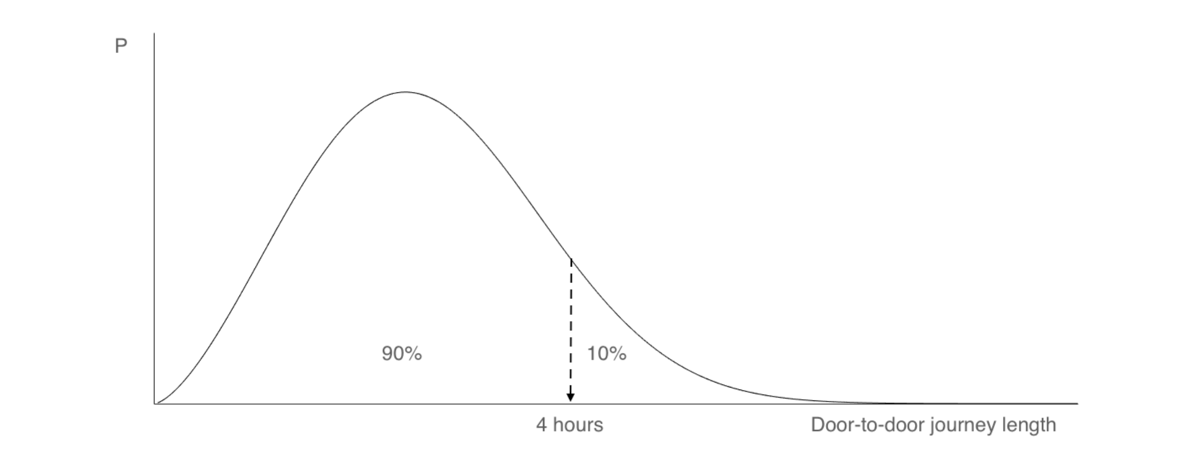


Figure 4 4HD2D (Four hour door to door)

Every modification of the air transportation system will affect the previous distribution in a unique way. For instance, when looking at the previous graph shape, a modification (technical, regulation, inter-modal) may lead to:

- Distribution curve with a similar shape but horizontally moving to the left / right: correspondingly towards / against the 4 hours mobility objective;
- Curve changing shape: increasing the peak height, width and/or modifying the shape of the tail;
- Any combination of the previous points.

This is connected to the following fact: there are several ways to achieve the 4-hour door-to-door goal for 90% of travellers, inherent to the vague definition of the metric. In each strategy the individual passenger experience is different but it would fulfill the global 90%/4h metric on a global scale. To illustrate this, some examples in the case of 100 pax from various EU countries travelling:

- 90pax doing a **3h 50minutes** trip plus 10 doing a **4h 10 minutes** trip
- 90pax doing a **1hour** trip, plus 10 doing **8hours** trips
- 50 pax doing a **2hours** trip, 30 pax doing a **3hours** trip, 5 pax doing a **4hours** 10 minutes trip and 5 pax doing a **5hours** trip
- 90 pax doing the trip below 4 hours, 10 pax (**all of them a specific EU country/region**) requiring more than 4 hours

As it can be extracted from the example, different distributions shape achieves the door-to-door goal. The following graphics show some Weibull distributions (in terms of probability density functions with different eta and beta values):

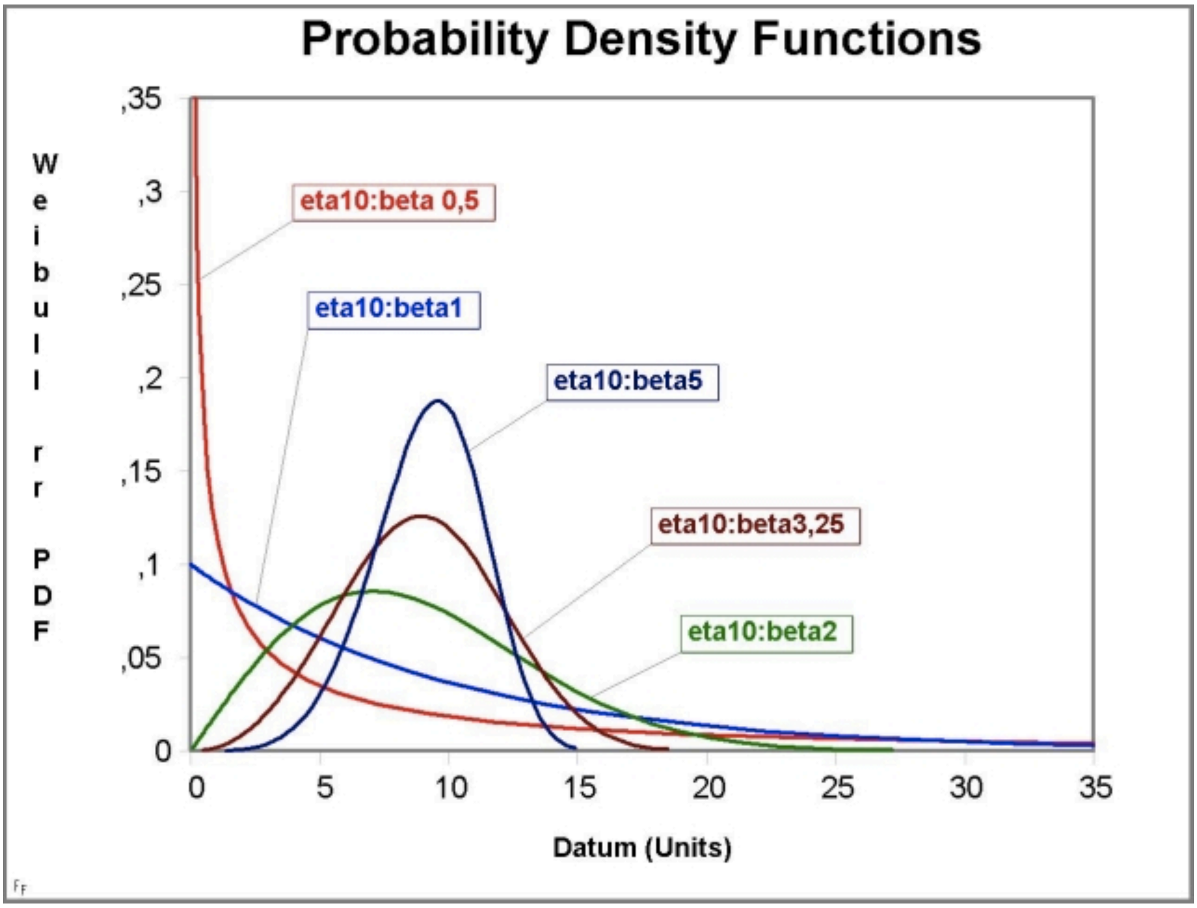


Figure 5 Weibull prob. density functions examples

Parametrization 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 $f_B(x)=f(x\wedge B)/f(B)$. 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.

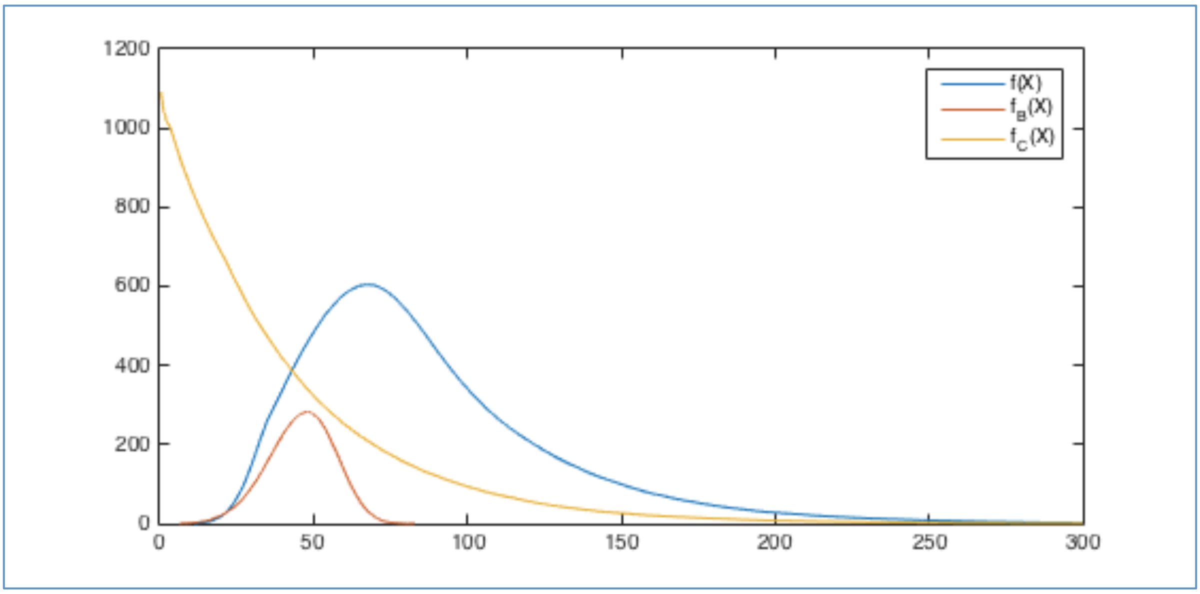


Figure 6 Components. blue distribution is the aggregation of red and yellow distributions

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 $f(x\wedge B)=f_B(x)*f(B)$ to generate the whole "journey length" distribution.

This formula helps to understand how a particular subset of passengers contributes 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 $f(x\wedge B)=f_B(x)*f(B)$.

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

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

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

Only when $X_1 \dots X_n$ are independent random variables, the joint distribution become the product: $f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f_{X_1}(x_1)\dots f_{X_n}(x_n)$, 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 is 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.

Markov chains are a type of random processes that provide a framework when dealing with sequential processes. The whole process is divided into several states and the transition from one state to another has certain known probability. The main property for the Markov chains to work is the "memorylessness" that is the probability of the next state depending only on the current state and not on the sequence of previous states. If the states are chosen carefully then the door-to-door journey could be considered as a Markov process, e.g. if a non-connecting passenger arrives at the departure gate ten minutes late it doesn't really matter how the passenger got there or what caused the original delay, however there are different implications for a connecting passenger arriving late at the connecting gate. Hence, to ensure "memorylessness," arriving late at the departure gate needs to be modelled using two estates instead of just one.

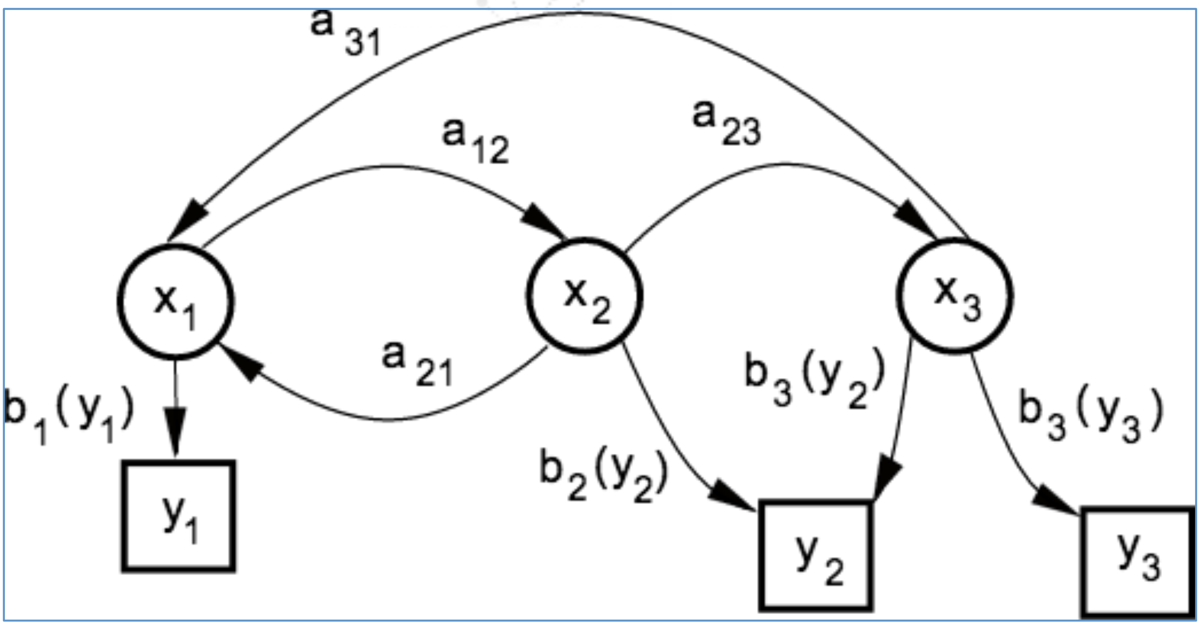


Figure 7 Example of Markov chain with six states (3 terminal)

Once the Markov chain or the event diagram has been designed and the probabilities have been estimated, it would be possible to measure how changes on the process or the length of each journey phase would affect the overall door-to-door journey length distribution.

Finally both, parametrization and journey components can be combined to check the effects of modifying the journeys of just a fraction of the total passengers on the overall distribution. This is of importance because it can help to understand priorities on how to effectively transition to a four-hour door-to-door goal by optimizing the critical/bottlenecks first.

5. Model Architecture

The DATASET door-to-door mobility model will be implemented following an **event-driven paradigm**. This was determined after examining other computing options. Specifically, in sequential computing, such as imperative programming, program flow is determined by a sequential execution of lines of code. Flow control is determined by conditional and iterative commands, but there are no bifurcations and there is only a single thread of simultaneous execution. Event-driven programming, as proposed for this model, takes a different approach.

Building blocks are events, which are small sequential algorithms (traditional computer programs), all events are independent and can be executed simultaneously. Communication through events is restricted so events do not have to relay in each other. On the contrary, communication between events is made possible by using an additional layer called environment. The environment can be read or modified by the events, and changes are instantly available to other events. Some events may be proactive and actively modify the environment, while other events may be passive and only wait until a parameter in the environment changes. In any case, events may trigger new events creating a chain of events that ultimately drives the simulation forward.

Instead of controlling the execution using conditional and iterative commands, one or more event stack(s) control the execution flow in event programming. The event stack determines which event(s) execute(s) next and provides an interface to add/modify events.

Additionally there are some supporting elements within the events and event stack elements. Specifically, there is a Data Store containing all of the static information for the scenarios, e.g. initial parameters, traffic data, declared capacity, etc. and an scenario loader that populates the event stack on launch. The Data Store does not change between simulations, and all outputs are written into a dynamic database. This dynamic database is processed once the simulations are completed to produce the final metrics and indicators.

The **figure 8** shows all these components working together and provides a view of the main architecture flow.

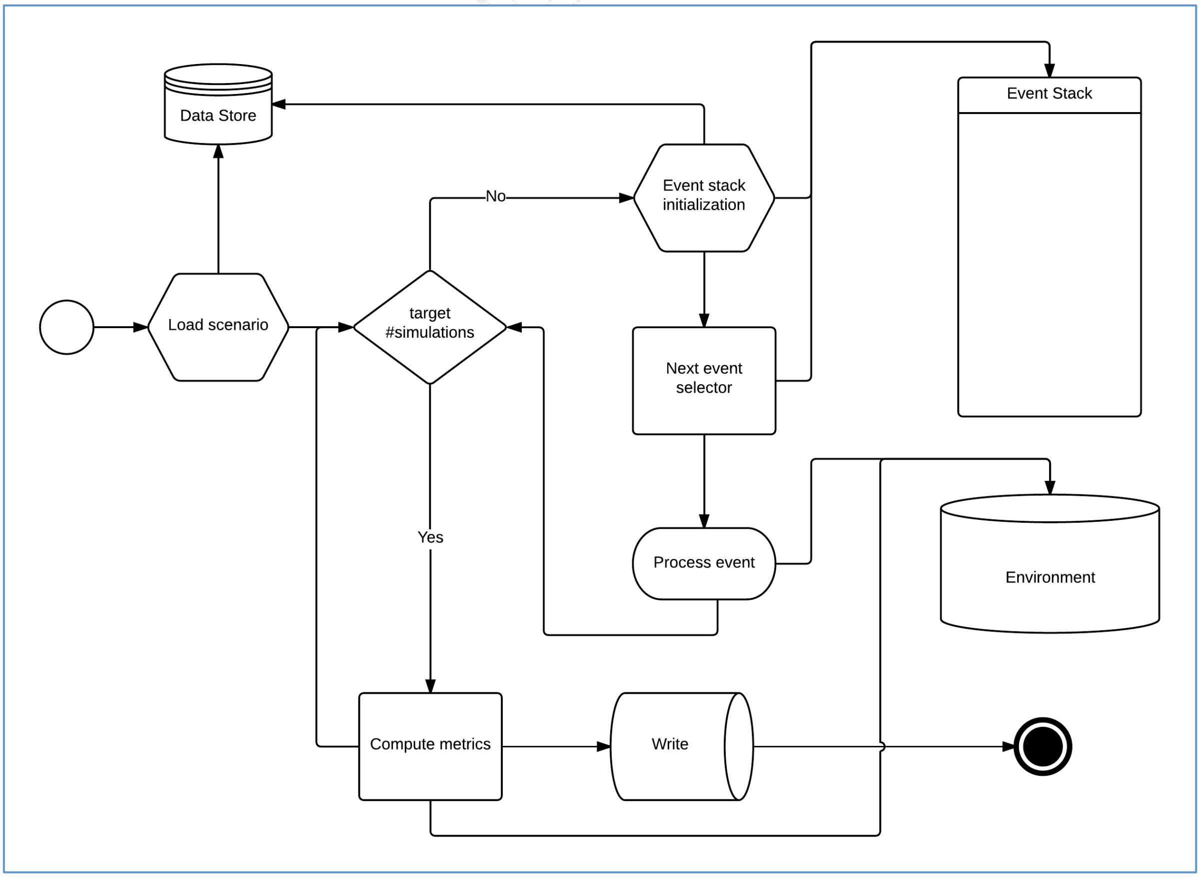


Figure 8 General architecture diagram

The design of the event-driven approach ultimately relies on how events are defined, as the other elements such as the environment should then be adapted to those. The following sections provide a description of the events considered in the DATASET door-to-door mobility model. For the sake of clarity they have been divided into the following five phases:

- Door-to-kerb, surface/underground
- Kerb-to-gate, airport side
- Gate-to-gate, air side
- Gate-to-kerb, airport side
- Kerb-to-door, surface/underground

4.1 Door-to-Kerb (D2K)

One of the most innovative approaches of the DATASET2050 model is the fact that it covers the door-to-kerb and kerb-to-door phases of passenger journeys. This part of the model should realistically simulate the different means of transport to get to/from the airport with their corresponding time spent, including buffer times estimated by the passengers. However, extra time in the door-to-kerb would be absorbed at the terminal, hence included in the kerb-to-gate phase.

Door to Kerb Process

Door-to-kerb is 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.

Kerb-to-door is the opposite process: moving from the airport's kerb to the door of the building of the destination (home, office, hotel or any other building unrelated to the trip).

It is important to note that this is valid for all the "sub-trips", including both inbound and outbound segments. For instance, in the case of someone living in Madrid and travelling to Brussels for a meeting and then returning to Madrid within the same day, she experiences:

- One "door-to-kerb" phase in Madrid (outbound trip: Madrid ->Brussels)
- One "kerb-to-door" phase in Brussels (outbound trip : Madrid ->Brussels)
- One "door-to-kerb" phase in Brussels (inbound trip: Brussels -> Madrid)
- One "kerb-to-door" phase in Madrid (inbound trip: Brussels -> Madrid)

Hence, we would split this trip into 4 distinct elements: door-to-kerb outbound, kerb-to-door outbound, door-to-kerb inbound and kerb-to-door inbound. Depending on the departure airport and the complete passenger profile it could be possible to estimate which percentage of passengers are doing inbound or outbound journeys with the data available in D3.1.

Modelling door-to-kerb

In principle, both door-to-kerb and gate-to-kerb (both inbound and outbound) processes could be modelled the same way. However there are some slight differences that need to be incorporated in the modelling, as they can impact the results :

- The **availability** of some means of transport is not always the same in door-to-kerb and kerb-to-door processes. For example, there is (almost) always a taxi stop at airports, however there may not be one at the passenger's departure door.
- **Queuing times** of some means of transport may differ in door-to-kerb and kerb-to-door processes. Following the previous example: queues at airport taxi stop may take more time than simply taking a free taxi at the street.
- Passenger perception of time and costs strongly depends on the socio-economic background, travel purpose as well as availability, comfort and frequency of different transport modes. The metrics in the model can hence be impacted depending on the particular passenger type whose journey is evaluated along with the geographical origin of the passenger. Therefore within work package 3 of this project different passenger profiles are derived applying a range of parameters (see

figure below). According to these parameters and the resulting profiles, the model output may differ.

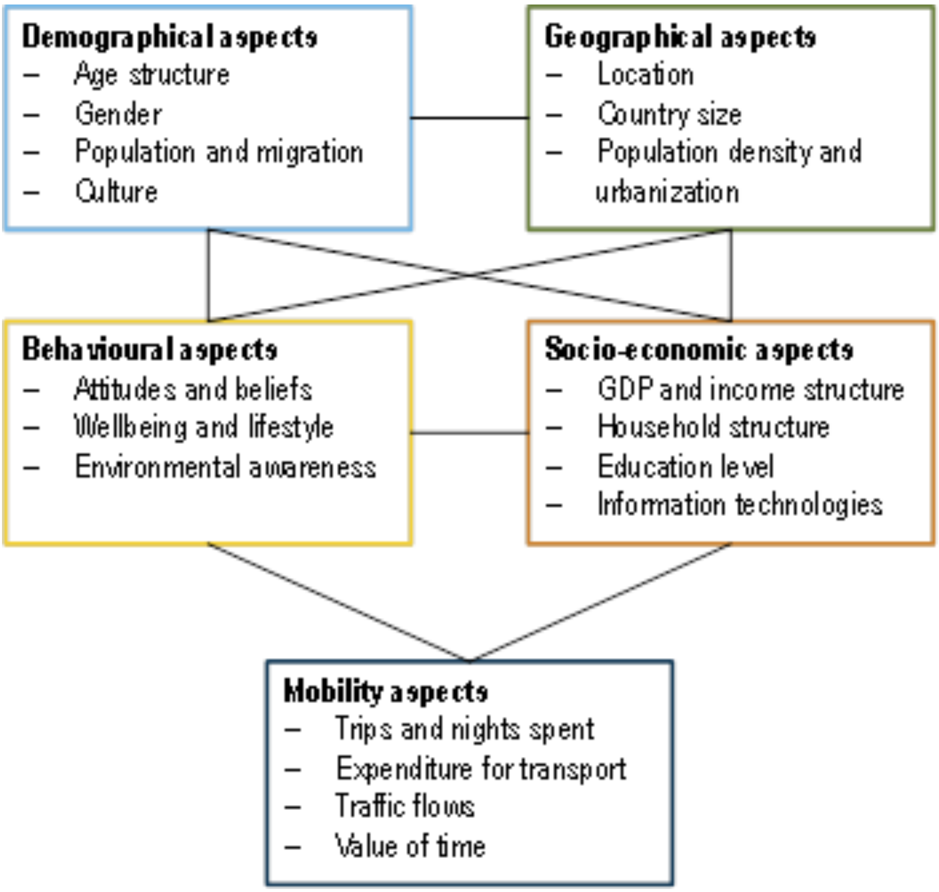


Figure 9 Passenger aspects

- As an example: passengers are more likely to hurry (higher time evaluation value) in door-to-kerb processes since they are bound by their flight's departure time (departure time plus kerb-to-gate processes) which does not apply in the reverse process. Usually kerb-to-door processes are less demanding and more flexible, as passengers are arriving to their final destination. There could be exceptions, for example when needing to punctually arrive to a meeting in an outbound trip. Hence, the model should take into account:

process step	time valuation
door-to-kerb inbound	high
door-to-kerb outbound	high
kerb-to-door inbound	low
kerb-to-door outbound	medium

The model also must take into account that passengers cost elasticity and comfort requirements may change depending on the stage of the trip (Pels et al., 2003; Loo, 2008, Tsamboulas, 2008). For example: both kerb-to-door processes and outbound processes are done after the (long, tiring) trip. The value of time Plötner (2010) and willingness to finish the trip as quickly as possible increases once the flight/business meeting/tourism experience is accomplished. So, in this context of cost elasticity, comfort and fatigue, passengers are more likely to choose faster (and more expensive) means in those processes.

process step	cost elasticity (wrt comfort)
door-to-kerb outbound	lower
kerb-to-door outbound	medium
door-to-kerb inbound	medium
kerb-to-door inbound	higher

Catchment Areas

Catchment areas are the geography locations from which an airport attracts a population (potential travellers) that uses its services as part of a door-to-door experience. The DATASET2050 model should reflect how catchment areas really work, particularly taking into account the following elements:

- Catchment areas ultimately depend on properly modelling mobility demand (demography, tourism, business activity, passenger profiling, cost elasticity) and mobility supply (means of transport available per area, duration, frequency). The model is based on D3.1 and D4.1 outcomes.
- Boundaries of catchment areas will be modelled based on Voronoi diagrams concepts. Voronoi diagrams split the European geography into regions based on "distance" to points (airports in our case). The set of points/airports are specified beforehand, and for each airport there is a corresponding region consisting of all points "closer" to that one than to any other. These regions can be modelled as a Voronoi diagram:
-

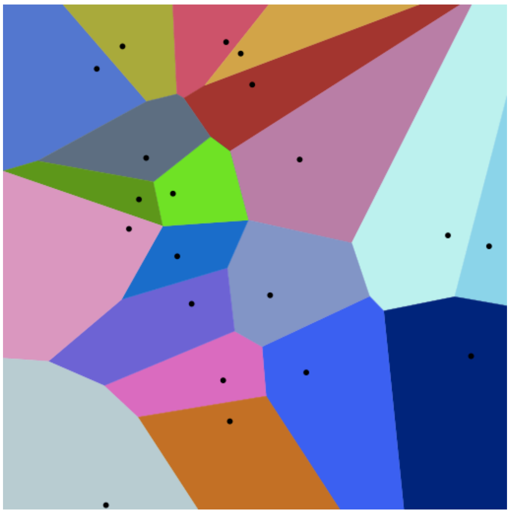
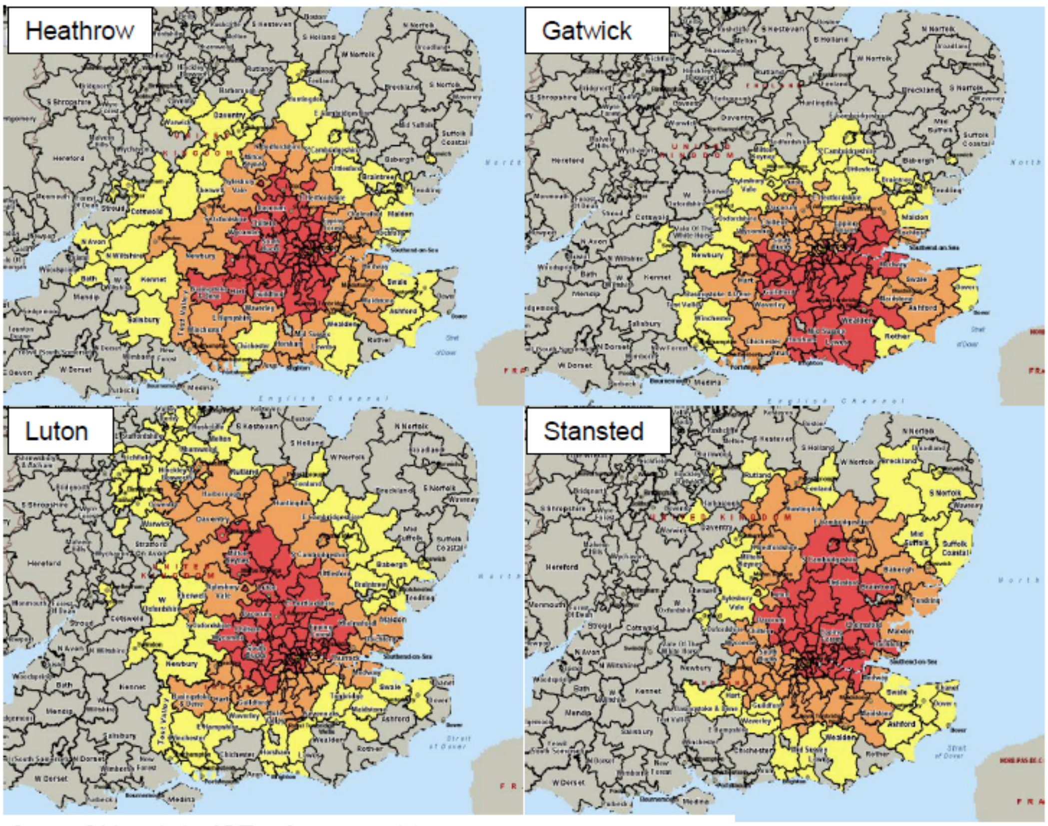


Figure 10 Voronoi concept

- **Catchment areas** (and what we will called "Voronoi simplified diagrams") are defined based on the distance to the airport. The distance (being strict: time-based distance) is measured in terms of travel time spent in the door-to-kerb process, which ultimately depends on the mean of transport chosen.
- Some major European cities are served by **several airports** (eg: London: Heathrow, Gatwick, Luton, Stansted, City), hence, the Voronoi simplified concept cannot be directly applied. For those cases, a generalized Voronoi approach will be used, taking into account several properties (distance, time, cost, comfort). The catchment areas will consist in calculating the minimum norm of the vector for each case. For instance, Luton may be preferred by low-cost tourists flying EasyJet despite living in southeast London neighbourhoods (physically closer to Gatwick).
- A similar case is experienced by those travellers, who despite being within a specific Voronoi cell, ("Airport1"), ultimately choose a different airport ("Airport2). This could be the case where the final destination cannot be reached by a direct flight from "Airport1", but there are available direct flights from "Airport2". Some travellers may prefer to extend the door-to-kerb phase in order to have a shorter gate-to-gate phase.
- The following graphic (source: CAA, surface access data) is illustrates two points. First, it shows the overlapping catchment areas of airports at a major city. Second, due to the transport options available some regions that are geographically farther from the airport are actually closer in terms of time required to reach the airport.



Source: CAA analysis of DfT surface access data

Red: 60 minutes; Orange: 90 minutes: Yellow: 120 minutes

Figure 11 London catchment areas

For example, Lieshout (2012) investigates the size of an airport's catchment area by considering access costs, airlines' airfares, as well as airside time costs, i.e. the flight time and possible connecting time between flights. The results imply that leisure passengers are willing to take on longer airport access times in order to save money on airline fares, for example. Furthermore, access time depends on the stage length of the flight and the overall duration of trip. Passengers undertaking a two week holiday to a long-distance location are willing to spend more time in airport access, for example, than a traveller conducting a business trip within one day, e.g. Madrid to London and back.

Here, the share of the door-to-kerb time and vice versa in overall travel time determines passengers' value of travel time and ultimately how much they are willing to spend within this process. Other studies also address the differences in access times by passenger profile including Pels et al. (2003), Loo (2008), or Tsamboulas and Nikoleris (2008). Plötner (2010) outlines different value of time studies differentiating between business and leisure passengers. Here, values range between about USD10 for leisure and USD86 for business passengers. However, values of time depend on the year investigated as well as the regional sample taken into consideration. It is true for all studies though, that business passengers have a higher value of time than leisure passengers. From this it can be deduced that business passengers want to spend less time in the door-to-kerb phase than private travellers.

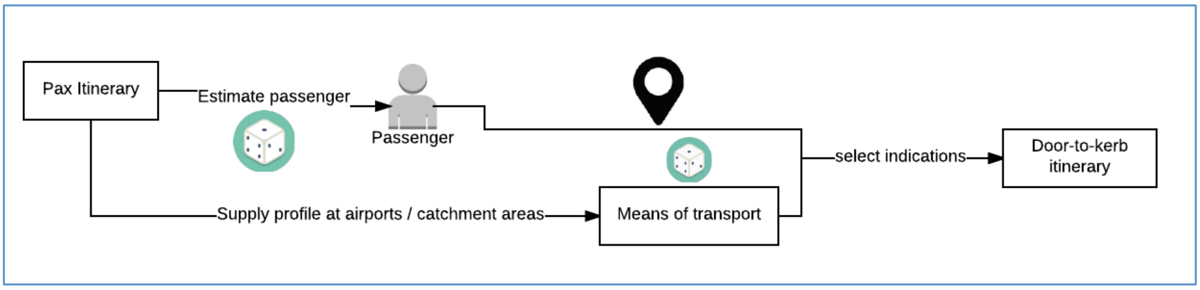


Figure 12 Passenger position and available options

Means of transport

The model needs to incorporate each possible means of transport used in the door-to-kerb phase for each particular airport. The purpose of splitting between the different sub-groups is to properly model their speed/duration, frequency, punctually, comfort and price. In some cases modes can be aggregated into clusters with similar characteristics.

- **No transport required**- hotels or buildings just in front of airports that do not require additional travel besides walking. For instance, the Sheraton Hotel in Brussels Airport is just 15 meters away from the airport entrance.
- **Cycling-** Traveling to-from the airport using bicycles (own or rental). This is not frequent, especially if traveler has luggage.
- **Riding motorbikes**- Going to-from the airport using motorbikes (own or rental). This is also not as frequent, again considering if the traveler has luggage.
- **Airport shuttles**- This would include buses and trams from hotels and companies close to the airport. Shuttles for hotels around airports are common in Europe. In

Dubai, Emirates Airline headquarters are directly connected to the airport with a direct-tram line.

- **Metro, trams and short-range trains**- Most major European cities have a subway/metro/short distance train that connects the airport terminal(s) to the city center.
- **Trains**- Long and medium distance train stations are sometimes located close or even within major airports catchment areas. This is the case for instance for the Cologne-Bonn Airport and most German airports.
- **Cars-** This pertains to vehicles driven by the passengers themselves. Both private cars (left in the short/long-term parking) and rental cars will be incorporated in the model for this subcategory.
- Passenger taken by car to the airport-
  - **Taxis** are one of the most widely used means of transport to/from the airport in the world, especially if there are no other public transport options (such as at night).
  - Other **car-based transport options** (such as Limos, Uber or Blablacar) differ from taxis in terms of the price associated but usually take similar times. Additionally, this group is also composed of passengers taken to the airport by someone not getting paid for the service: e.g.: spouse, relative, friends, relative. This group, together with the taxi group, differs from the passenger driving its own car considering there is no time allocated for parking or car rental return.
- **Boat**- Ferries and boats are sometimes used as means of transport when traveling. This is true in the case of Istanbul or Greek Islands such as Santorini.
- **Multi-modal journeys**- this combines two or more of the following means of transport to/from the airport.

Airports with similar access modes and catchment areas would be modelled as generic airports. The final door-to-kerb model would have between 5 to 10 detailed airports, plus 2 to 4 generic airport profiles. The final list of available options and methodology will be reported in D4.1 as part of the current supply profile.

Choosing the means of transport

As derived from D3.1, each passenger decides to use one specific door-to-kerb option depending on their profile (price sensitivity, length sensitivity, etc.), along with the availability of door-to-kerb alternatives (D4.1). The model for each of the previously listed means of transport will incorporate the following elements:

- Duration and frequency- In this case total duration is often a decisive parameter.
- Punctuality- Or more precisely, perception of punctuality, drives passenger expectations and therefore voluntarily extending their door-to-kerb time by adding an extra, buffer time.
- Comfort and price- Depending on the passenger profile some passengers are more price-driven others.

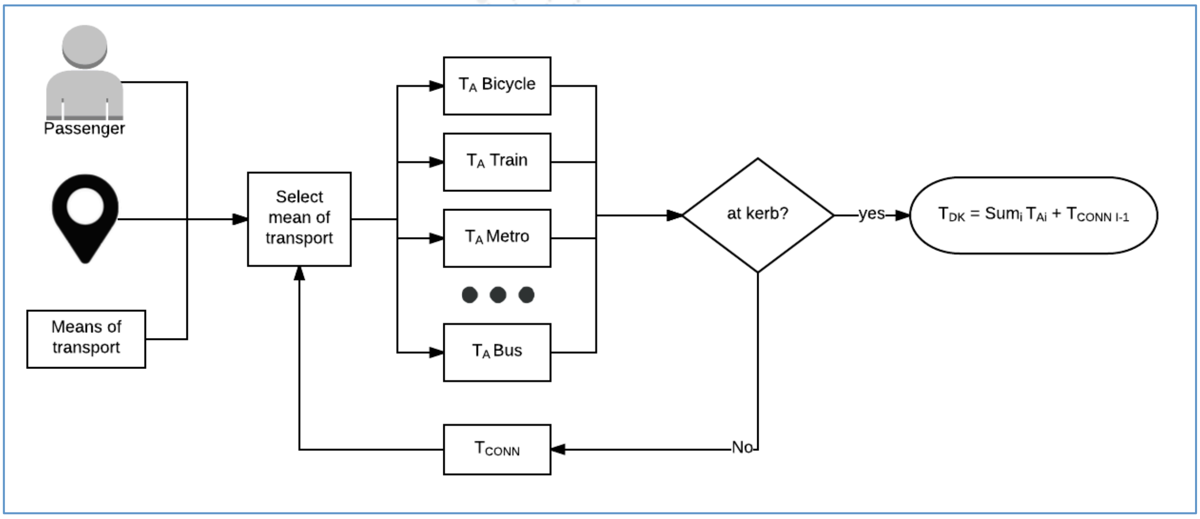


Figure 13 Pax modes. Computation of D2K time

4.2 Kerb-to-Gate (K2G)

The kerb-to-gate process 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 involves the following list of processes, all of which needs to be incorporated in the model:

From kerb to the terminal

As explained in section 4.1, the "kerb" is understood as the latest point of the door-to-kerb process; the drop-off location for passengers at the airport by the means of transport chosen: taxi stop, parking lot or airport metro station. Some European major airports have more than one terminal and/or large and complex geometries. In that case, passengers first need to ensure the correct terminal for their flight, by checking the ticket, looking at the information screens or asking someone in the airport/airline offices. Secondly, passengers need to make their way to the appropriate terminal, an action that may take a considerable amount of time, depending on the airport. Also, depending on the layout and pax-flow distribution within the airport, usually traveling to the airport terminal is done before check-in, at security check-in or directly before boarding gate. In each of these cases, the size and complexity of the airport needs to be adequately modelled, given its influence in the kerb-to-gate duration.

Boarding pass

If passengers have not obtained/printed their boarding pass in advance then they can obtain their boarding pass at self check-in machines available at the airline check-in areas. Nowadays, most airlines offer electronic boarding passes. This development has decreased the number of passengers requiring this additional step, and they usually carry their boarding pass with them in their electronic devices (PDFs, BiDi codes, Apple Wallet).

Luggage drop-off and check-in desks

Passengers leave their checked-in luggage at the airline check-in desks, after queuing (if required). Currently, both automatic "self check-in" luggage belts and standard luggage check-ins supported by airline/airport staff options are included in the model. In order to confirm the passenger identity corresponds with the boarding pass, the following documentation must be provided:

Type of flight	Type of passenger	Documentation required
Domestic	Domestic	National Identity Card or passport, not necessarily current (in some EU countries). Driving license issued in the country is also valid in some EU countries. Children (<14 years) exempt in most countries, under the responsibility of the person accompanying them on the flight.
	EU/Schengen	Valid passport or Identity Document, not necessarily current. Current residency permit for one of the Schengen states.
	Other countries	Current valid passport or travel document. Current residency permit for one of the Schengen states.
EU/Schengen	National	Current passport or ID.
	EU/Schengen	Current valid passport or identity document.
	Other countries	Current valid passport or travel document.
Other countries	Domestic	Current passport.
	EU/Schengen	Current valid passport or travel document.
	Other countries	Current valid passport or travel document.

After the documentation is provided and checked, luggage is weighed and processed. An extra fee may be charged if the luggage weight exceeds airline limit. Afterwards, luggage is labelled and directly placed in the luggage system. The model needs to realistically include this step for all passengers checking-in.

Security checks

Airport security steps attempts to prevent any threats or potentially dangerous situations from arising or entering the airport country. If airport security does succeed in this, then the chances of any dangerous situations, illegal items or threats entering into the aircraft, airport or country are greatly reduced. For this reason a number of security processes needs to be accomplished in order to pass the security control. The EU regulation no 2320/2002 from 2002 introduced the requirement to have security checks for all passenger flights, including domestic flights. Regulation 300/2008 establishes the common rules. Particularly passengers need to go through the following process (that will need to be modelled with the corresponding queue model):

- Passengers passing the metal detector: preventing passengers from carrying weapons or tools that could potentially be used as weapons.
- Carry baggage (including bags, back-packs, clothes etc.) is required to go through backscatter X-ray
- All carried metal parts and devices (belt, coins, cell phone, earphones) placed in the X-ray machine
- Shoe removal is usually required (specially if covering the ankle or with thick soles)
- Explosives and drugs trace-detection processes in carry baggage and passenger hands or clothes
- Additional security checks may be required upon request in specific circumstances

Immigration - customs

Flights within the scope of DATASET are between EU countries, and most are also part of Schengen or similar agreements. Hence, countries usually do not require additional customs and/or passport controls than those previously mentioned. DATASET model however incorporates a buffer in case an additional process is required at this stage for whatever reason (safety, security, legal, environmental or political affairs).

Boarding gates and boarding

Airlines request the passenger to be in the boarding gate at "boarding time"; some time before the gate actually opens. Boarding times depend on the priority class of the tickets, passengers with priority boarding can board first. After boarding and "last call" the boarding gate closes. Passengers must provide a valid boarding pass (electronic or physical) along with their corresponding documentation card (see table above).

The kerb-gate process is considered to be finished when the passenger crosses the boarding door and shows their boarding pass and documentation. In the slight case there is a boarding gate change prior to providing the boarding pass and documentation, then the time spent moving around the terminal and/or waiting is still considered part of the kerb-gate process.

Buffer time

Once the flight ticket has been purchased then the flight time is fixed. Due the uncertainty of some of the door-to-kerb and kerb-to-gate processes, passengers usually plan some buffer time in advance for both processes. Most passengers would estimate a new door-to-kerb and kerb-to-gate times based on their perception on the system efficiency and reliability, and add additional time for each phase.

Passengers are in general risk-averse. For some passenger profiles arriving late to the gate and missing the flight has a severe impact and it is usually avoided by adding a buffer of several minutes to their door-to-kerb and kerb-to-gate phases. Several factors contribute to this risk-prevention behaviour, including departing from a well-known city, age-range, use of new technologies, etc. In general these concepts can be well captured using convex utility functions.

The difference between the estimated time (door-to-gate) and the actual time, if positive, is spent at the boarding gate or around the terminal.

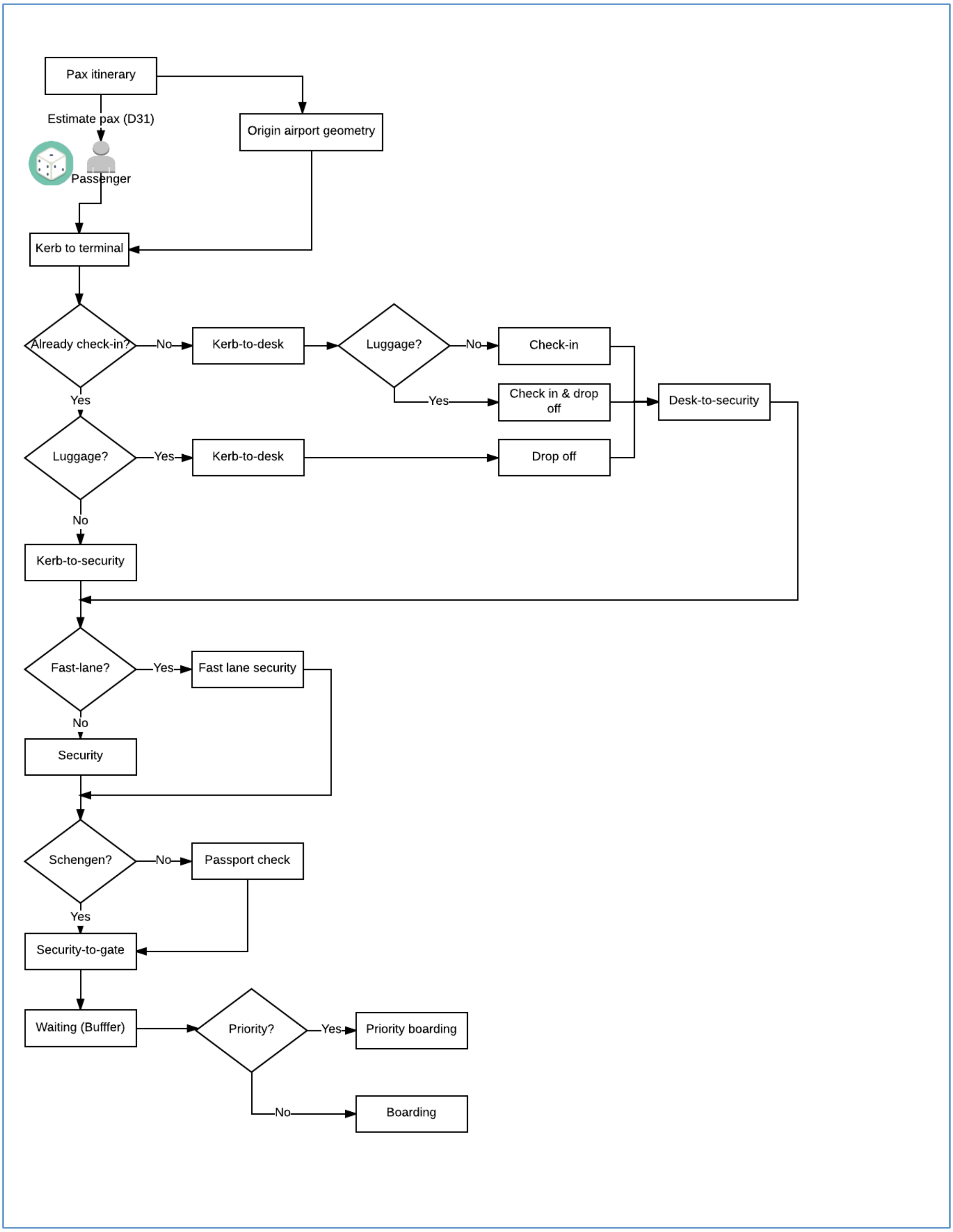


Figure 14 Kerb to gate process (K2G)

4.3 Gate-to-Gate (G2G)

Gate to gate process covers the time spent by passengers in between the following events:

- Crossing the boarding gate of the departure airport (for the first time).
- Crossing the boarding gate of the final arrival airport. This includes when the arrival airport is not the original destination (e.g. leg cancellation, flight re-routing, etc.).

Hence, in the case of passenger journeys with multiple legs, the gate-to-gate process in the model includes the connection durations and time spend in the terminal area between flights. As previously mentioned in DATASET2050 documentation, the gate-to-gate model will be based on the model already developed by two DATASET2050 consortium members (University of Westminster and Innaxis) in the SESAR- project "POEM" (Passenger Oriented Enhanced Metrics), which received the SESAR outstanding award in 2014.

In DATASET, the gate-to-gate process is comprised of the following sub-processes:

- **Boarding-** Depending on the airport, passengers either directly enter the aircraft going through a finger, or reaching the apron surface (shuttle bus or directly walking), along with escalators. The precise duration of these processes heavily depends on the airline , airport and the precise aircraft stand. Sometimes passengers have to wait additional time for missing passenger and/or crew member(s).
- **Departure and DMAN-** DMAN, also known as the departure manager, is a planning tool available at most mid-size and large EU airports developed to improve the departure flows at airports, thus increasing the predictability of flight scheduling. DMAN takes into account constraints and preferences, and also provides a planned departure flow that ensures the optimal throughput at the runway by transmitting the information to all airport stakeholders. In principle, individual passengers are benefited, although at times delayed as well, by DMAN in terms of their 4HD2D experience.
- **ATFM slots-** Too many aircraft in the air at the same time and place can lead to unsafe situations. One of the tools used by the Network Manager Operations Centre (EUROCONTROL) to prevent this from happening is to apply CTOTs (calculated takeoff times), which is also known as issuing ATFM slots or simply slot. The slot is actually a period of time within which takeoff must take place. In Europe, a slot is defined as the period between 5 minutes before and 10 minutes after the CTOT. The aircraft is required to be at the runway, ready for departure at its CTOT. The leeway is to allow air traffic control to integrate the aircraft into the other traffic.
- **Push-back; taxi out and taxi in-** In this process, the aircraft moves from the stand, to the ramp/apron and finally to the taxiway and runway. Taxiing refers to the movement of an aircraft on the ground under its own power. Sometimes aircraft queuing is required pre-threshold, in holding positions, such as de-icing areas or taxiways crosses.
- **En-route-** Climb, cruise descent, and final approach are estimated for historical flight plans. Wind and delay recovery are estimated using a stochastic function derived from historical data.

- **Arrival and AMAN-** Arrival Manager (AMAN) systems and tools assist air navigation service providers with aircraft arrivals, particularly during challenging periods such as bad weather, high demand or runway closures. Numerous products and systems have been introduced with great benefit across European airports, including Arrival Management systems and Extended Arrival systems: SESAR IP1.
- **De-boarding-** Depending on the airport, passengers enter the terminal building by either going through a finger, reaching the apron surface (via shuttle bus or directly walking), or mobile stairs. The precise duration of these processes heavily depends on the airline, airport and aircraft stand. Sometimes passengers have to wait additional time for passengers leaving the aircraft slowly due to disability, children, etc.
- **Waiting plus connecting time-** For this, the model will repeat previous steps N-1 additional times if the number of flights (N) required to reach the final destination is N>1.

Overview of the gate-to-gate model

The gate-to-gate model is a discrete-event simulation (DES). In a DES, the system is described by a discrete sequence of *Events*. Each *Event* of the sequence is a process that changes the state of the system; no change is assumed between two consecutive *Events*. One or more stack managers determine the order in which the events are triggered. Events can potentially introduce other new events during its execution. This generates a cascade effect on the number of *Events* controlling the flow of the model, which differentiates it from the more traditional sequentially programmed models.

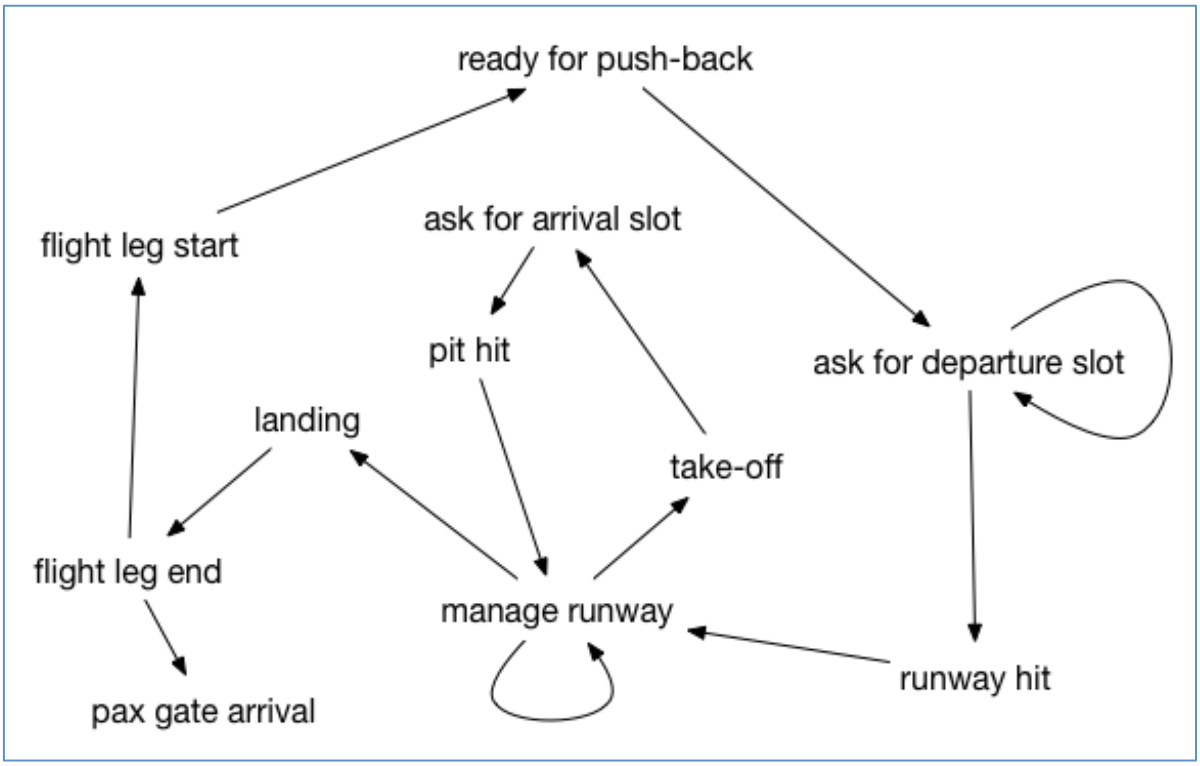


Figure 15 G2G events sequence

The "flight leg start" event is the first event in the flight processes and also the stack's initializing event. First, the flight could have been cancelled for some reason. If so then passengers are re-accommodated according to the airlines policies and a new "ready for push-back" event is introduced into the stack for the next flight, if any, in the aircraft sequence. For non-cancelled flights the location and status of the previous aircraft should be determined. If there is no previous aircraft, the Aircraft Ready Time is set and a new "ready for push-back" event is introduced in the stack. Alternatively, if there is a previous aircraft, then the Actual Ready time is calculated using the Minimum Turn Around time and the Calculated In-block time and adding a new "ready for push-back" event into the stack.

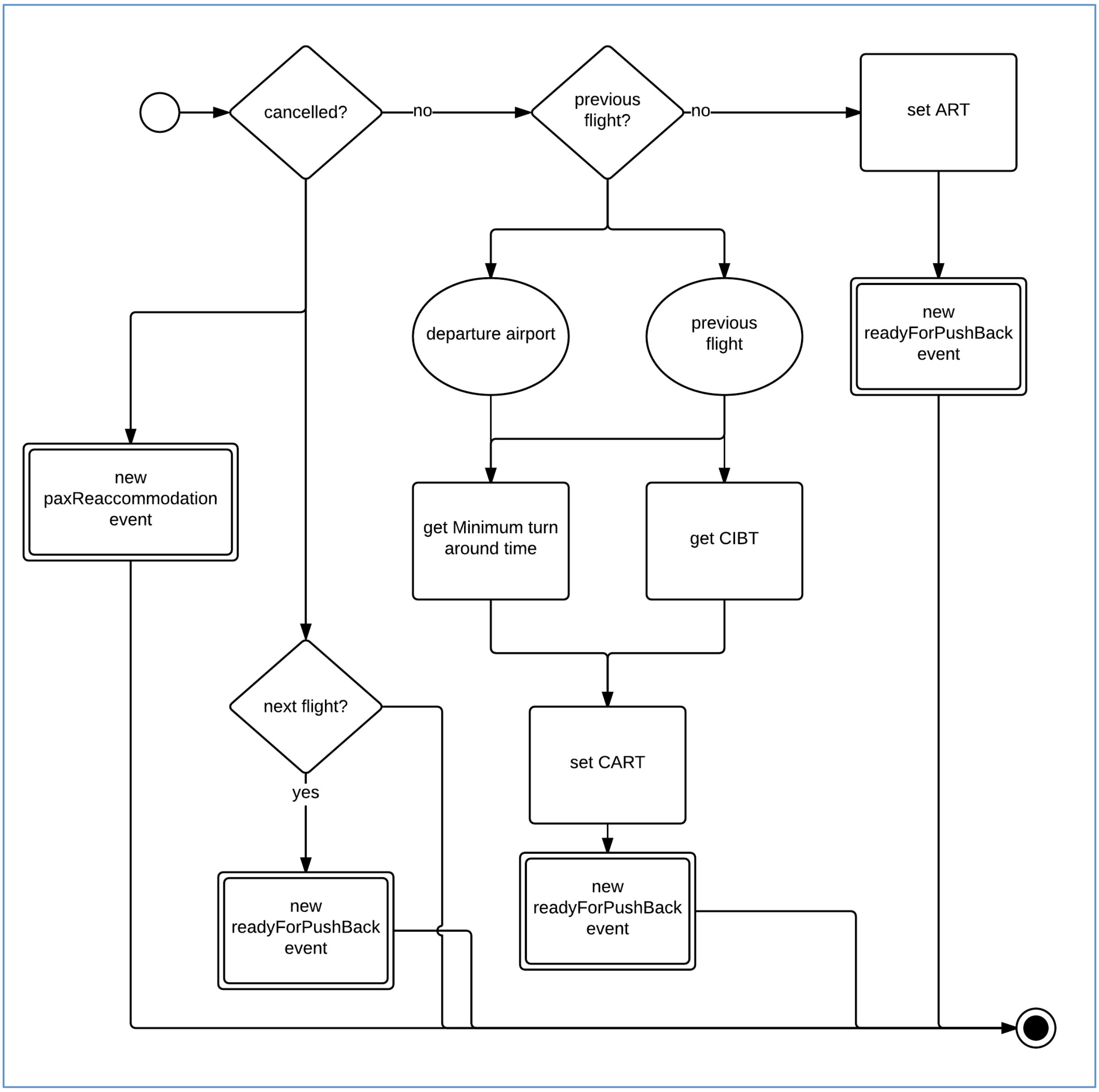


Figure 16 Flow diagram for the flight leg start event

The "ready for push-back" event first checks for any late passengers. If the flight decides to wait for those passengers, then the Off-Block Time is estimated using their Calculated Pax Gate Arrival Time, and a new "ready for push-back" event is introduced into the stack. If passengers are left behind then they are re-accommodated. Finally, an attempt is made to re-accommodate waiting passengers at the airport in any of the flight's free seats and a new "ask for departure" slot is added to the stack.

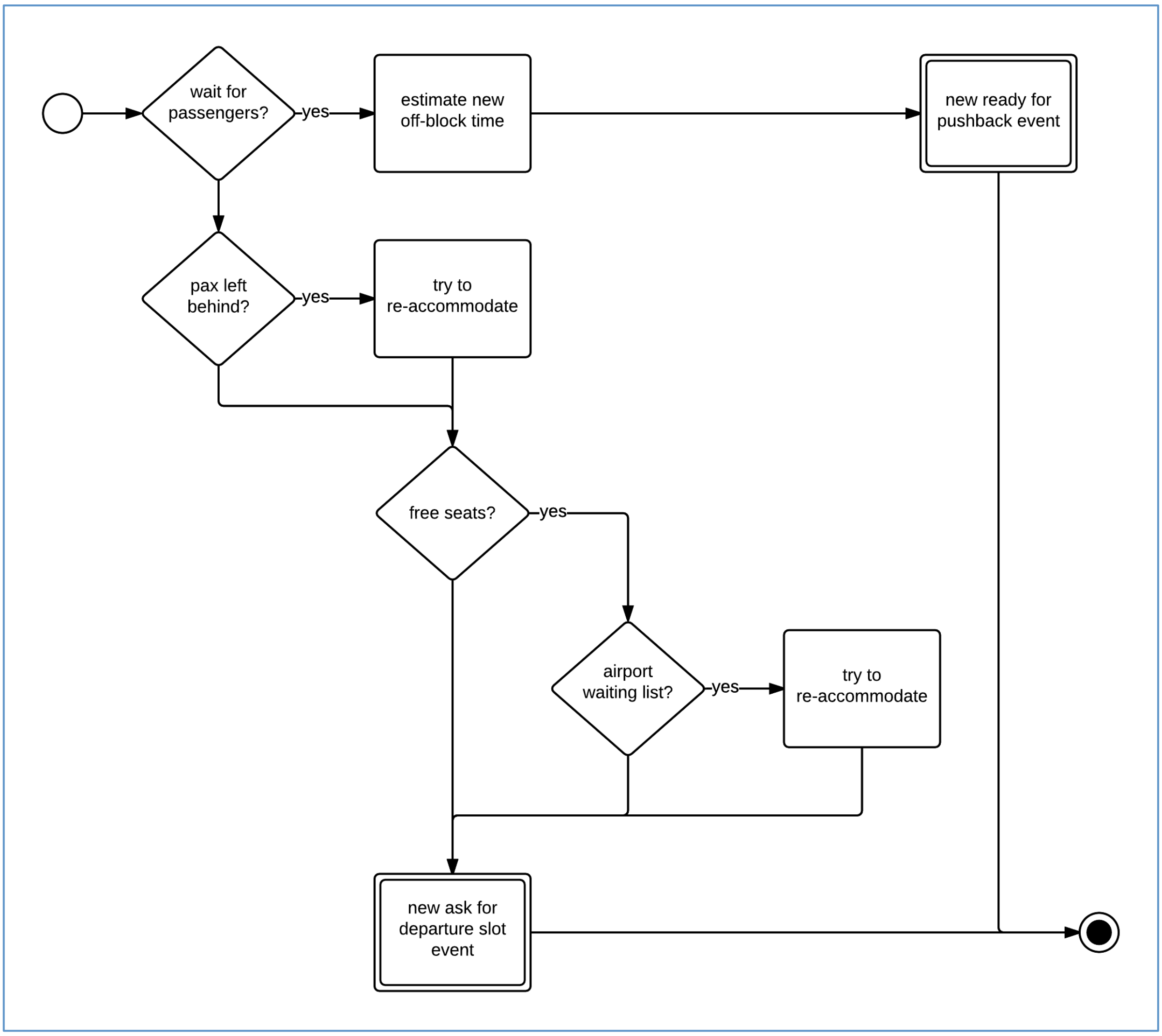


Figure 17 Ready for Pushback event

The "ask for departure slot" event first checks the current airport and checks if there is a current regulation applicable to the flight by the Network Manager. If the flight is regulated, then a new "ask for departure slot" is inserted in the stack until regulation is over. When the flight is not regulated then the departure queue is estimated. If the X length is too large then a new "ask for departure slot" is again added to the stack. If the X length is acceptable, then the Actual Off-Block Time is set and a new "runway hit" event is introduced in the stack after the taxi out time.

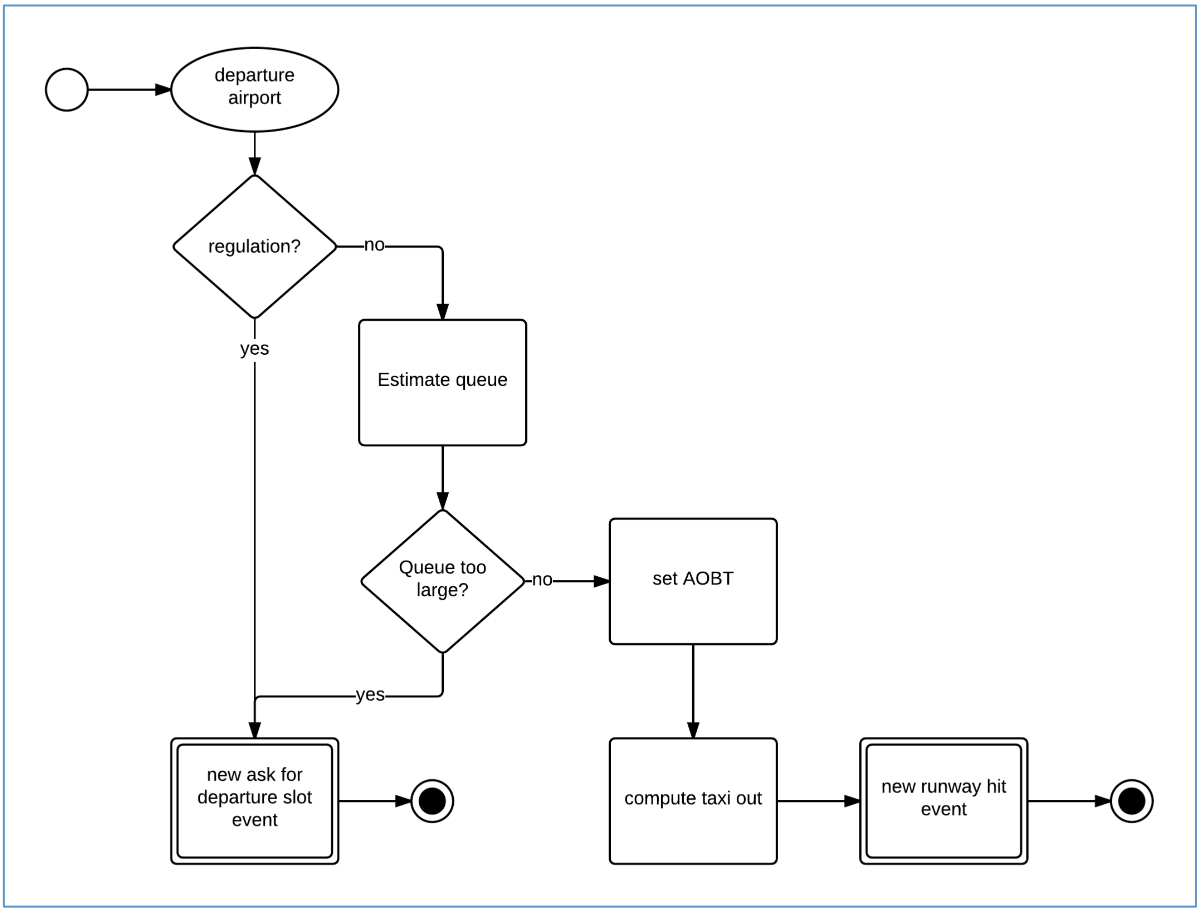


Figure 18 Ask for departure slot event flow diagram

The "runway hit" event starts by setting the Actual Runway Hit time. If the runway is clear, then a new "manage runway" event is introduced into the stack, otherwise the flight is added to the departure queue. The length of the queue is estimated and the Calculated Take-off Time is updated accordingly.

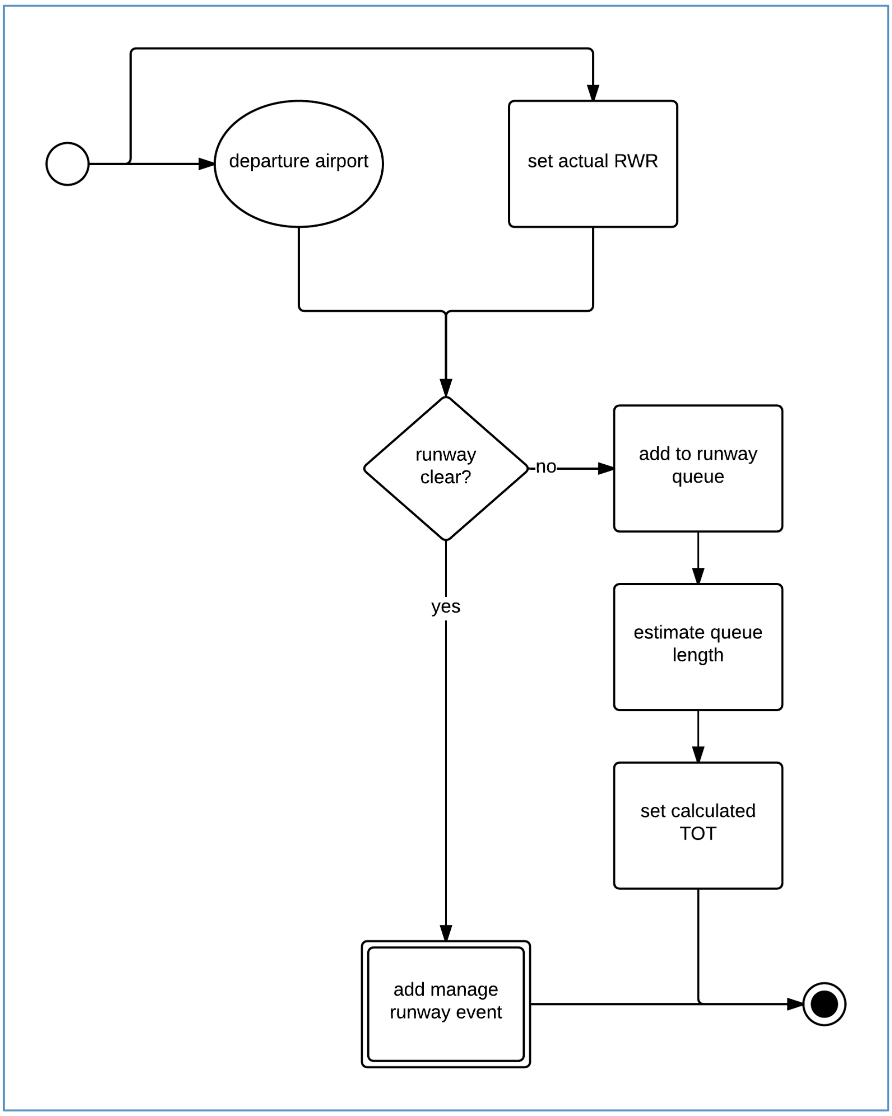


Figure 19 runway hit event flow diagram

The "runway management" or "manage runway" event starts by checking if the runway is currently ready. If it is not ready, then runway clearance is estimated and a new "runway management" event is added to the stack. If the runway is ready then the arrival and departure queues are checked and only one flight is selected next to use the runway. If a departure flight is selected, then a "take-off" event is introduced into the stack. If an arriving flight is selected, then a "landing" event is added to the stack.

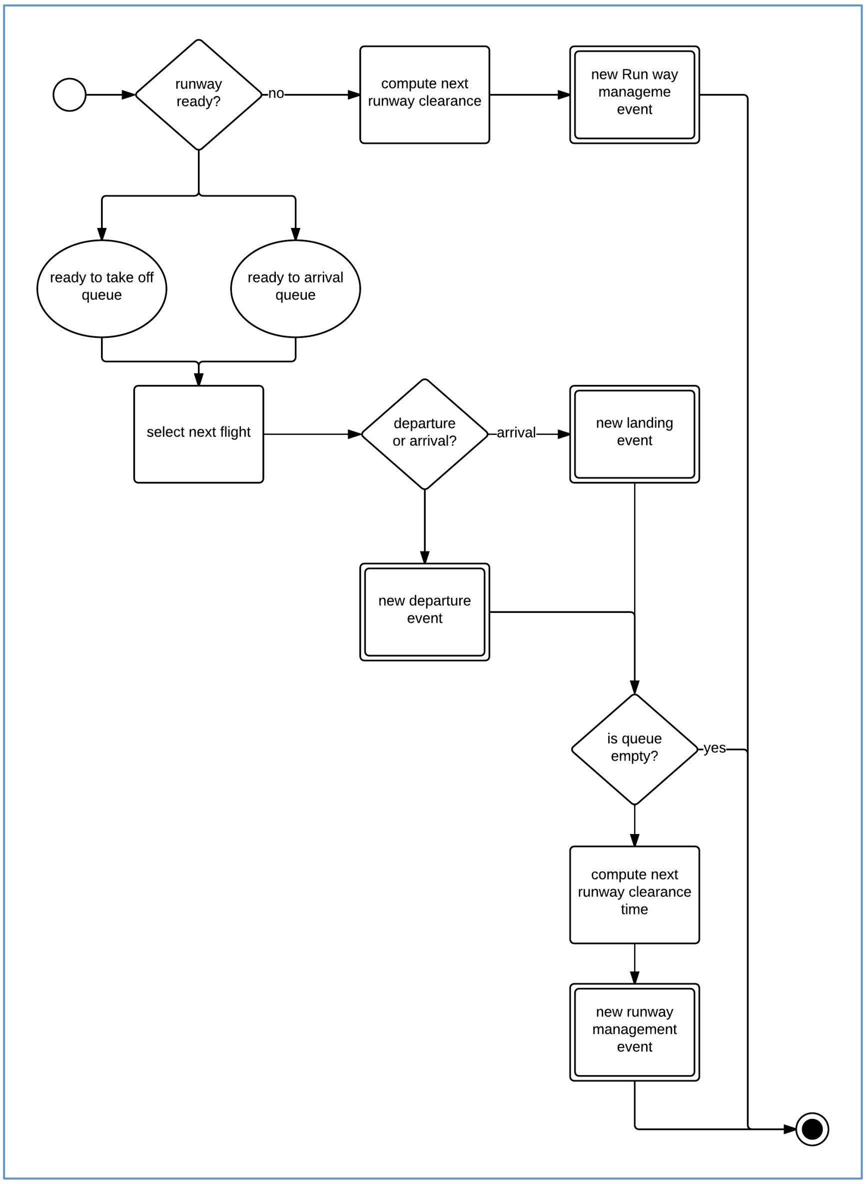


Figure 20 Flow diagram for the runway management event

In the "take-off" event the Actual take-off time is set. If there is a considerable departure delay then there could be an en-route recovery time , which would be estimated. Regardless of the case, delayed or on-time flight departure, the Arrival Time is estimated and a new "ask for arrival slot" event is introduced into the stack.

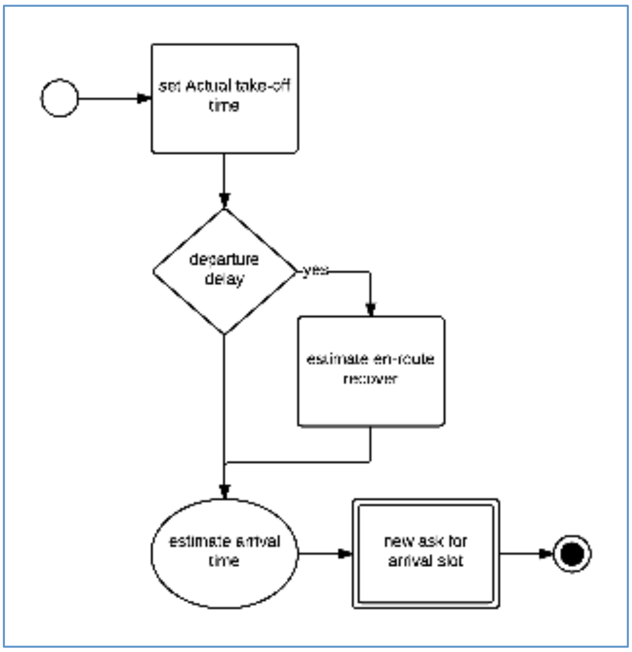


Figure 21 Take off event flow diagram

The "ask for arrival slot" event, determines when the en-route flight phase is over and the flight is making the approach. It is only after the flight has reached the Passing Time for IAF (PTI) that it can be intentionally delayed due to capacity overload at the airport. First, the distance to the IAF is asked from the airport, and the estimated time to reach it is then computed. A new "hit PTI" event is introduced into the stack at that time.

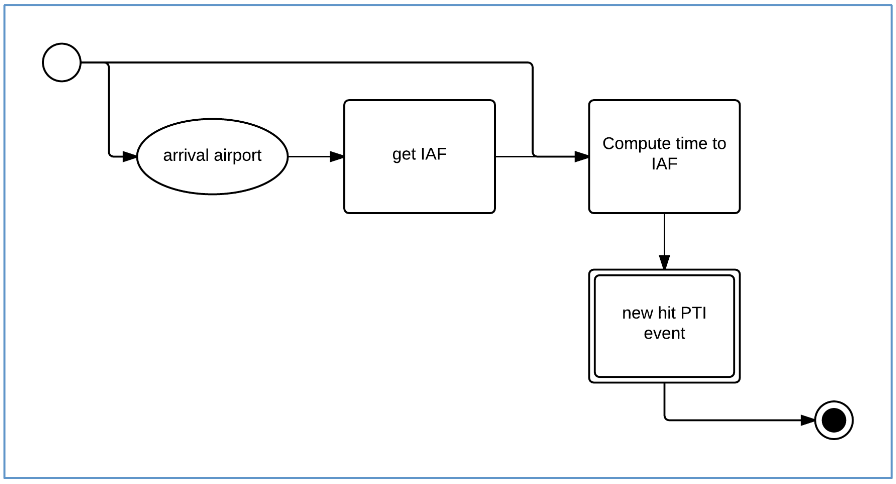


Figure 22 Flow diagram for the ask for arrival slot event

During the "hit PTI" (Passing Time to IAF) event the Actual PTI time and the Calculated Time of Arrival are computed and updated for the current flight. Next, the flight is added to the airport's arrival queue, as it waits for arrival when a slot becomes available. Finally if the runway is ready a new "manage runway" event is added into the stack. Otherwise if the runway is not ready then the flight is set to wait on holding.

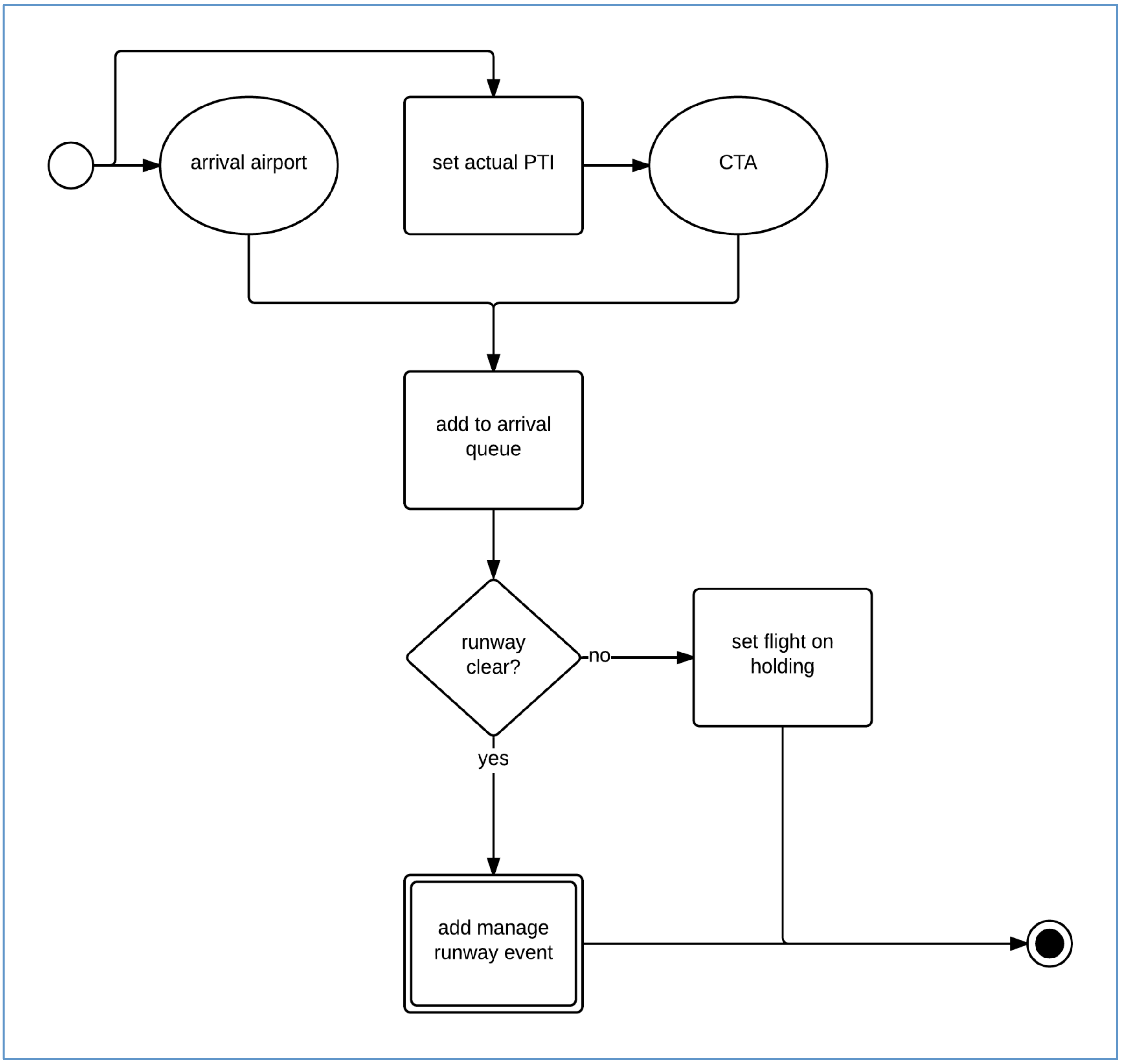


Figure 23 Flow diagram for the Hit PTI event

The "landing" event is the second to last event in the flight sequence. This event starts by setting the Actual Arrival time for the flight, and continues by calculating the In-Block time after the taxi-in time. Finally a new "flight leg end" event is introduced to the event stack.

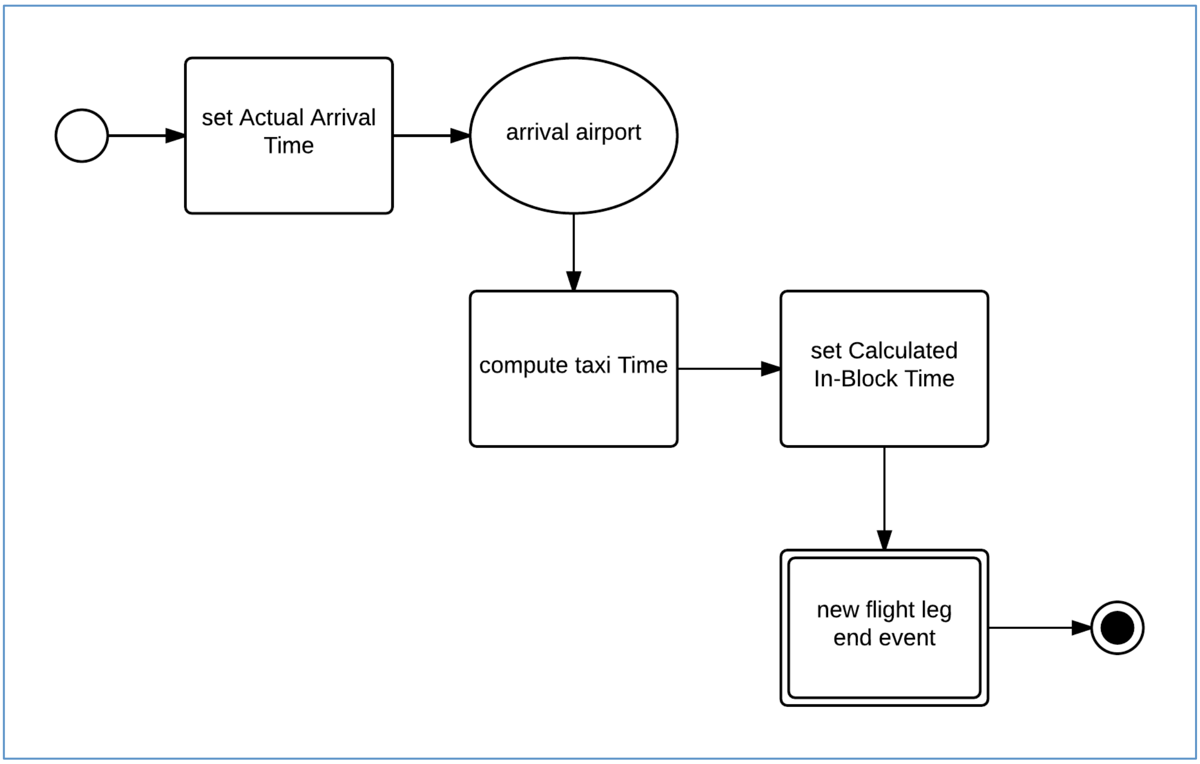


Figure 24 Landing event flow diagram

For the "flight leg end" event, first the Actual In-Block Time is set to the current time. Then, if there is a next flight scheduled for the same aircraft, then the Aircraft Ready Time is estimated using minimum turn around times and a random component, thus creating a new "flight leg start" event to be introduced into the stack. Regardless if there is a next flight or not, all of the passengers on the flight are taken into account. If passengers are reaching their final destination, then metrics are computed. If passengers are connecting then time to the next gate is estimated using the airport's minimum connecting times and a new "pax gate arrival" event is then introduced into the stack.

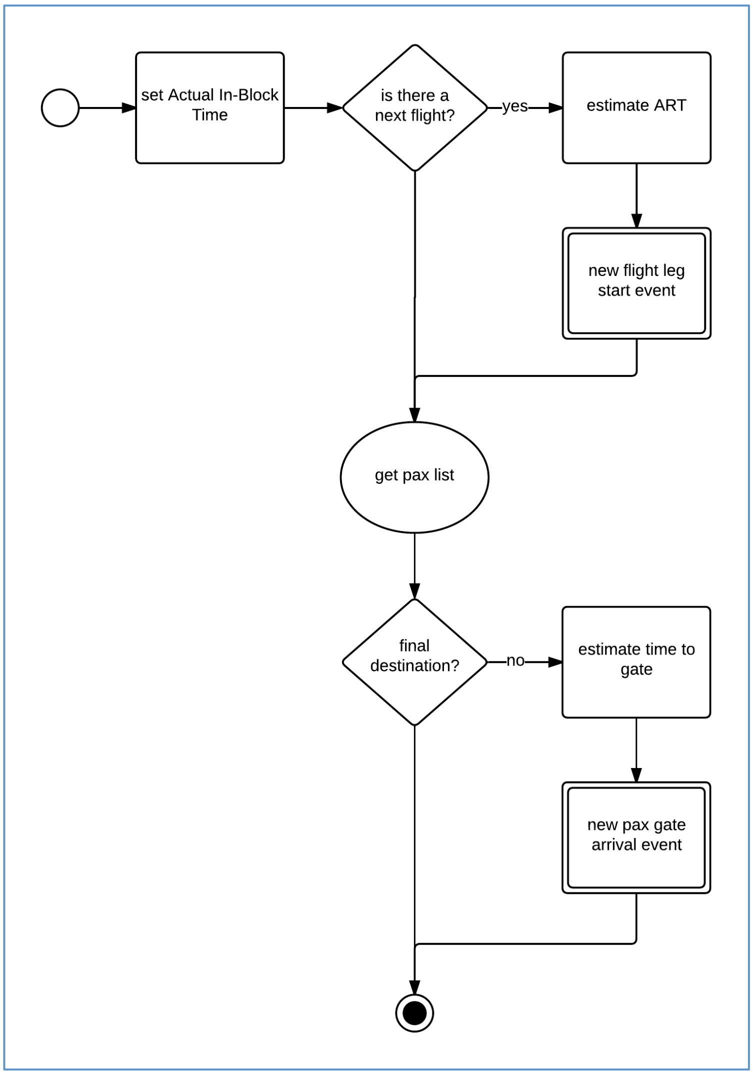


Figure 25 Flow diagram for the Flight Leg End event

The "pax gate arrival" event starts by setting the Actual Gate Arrival Time for the passengers. If boarding for the next flight is possible then passengers are boarded and the event finishes. If boarding is not possible then passengers are re-accommodated by the airline. Itineraries are updated for successfully re-accommodated passengers, while passengers that could not be re-accommodated are added to the airport's waiting list until a free seat is available in a later flight or for the first flight the following day after an overnight stay.

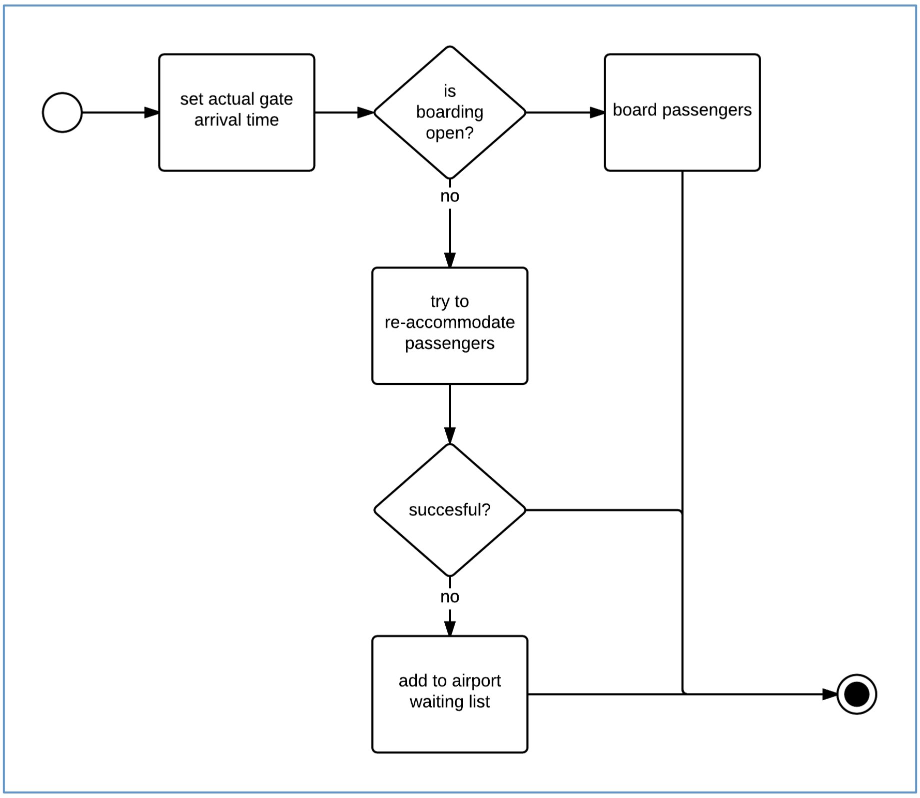


Figure 26 Flow diagram for the Pax Gate Arrival event

4.4 Gate-to-Kerb (G2K)

Gate-to-kerb is 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. The following outlines steps within the gate-to-kerb phase that a passenger experiences:

Passport- Security-Health Controls

At a certain point of the gate-to-kerb process, passengers need to provide their documentation (passport, ID) at the airport security borders. Taking into account the scope of passengers analysed in DATASET2050, the document required within EU-Schengen pax is a valid passport or a valid national identity document. Additional health and security controls may be required, as required and requested by country officials (police, military).

Gate-to-baggage claim

In airport terminals, those passengers with checked-in baggage will need to go through the baggage claim area after disembarking. The standard baggage claim area contains baggage carousels or conveyor systems that directly deliver checked bagged to the passengers. Also in this area, will usually be the airline and airport customer services for oversized baggage, special baggage (living species/weapons), or inquires on missing baggage.

Customs, for declared items, is usually located adjunct to the baggage claim area.

Gate to kerb

In addition to the security and baggage-related time spent, the time spent moving the additional distance (usually by walking) need to be taken into account within the DATASET2050 model. Depending on the airport and terminal, this can mean a significant amount of time. This process, as many others, will be driven by the passenger profile (taking for instance a reduced mobility passenger).

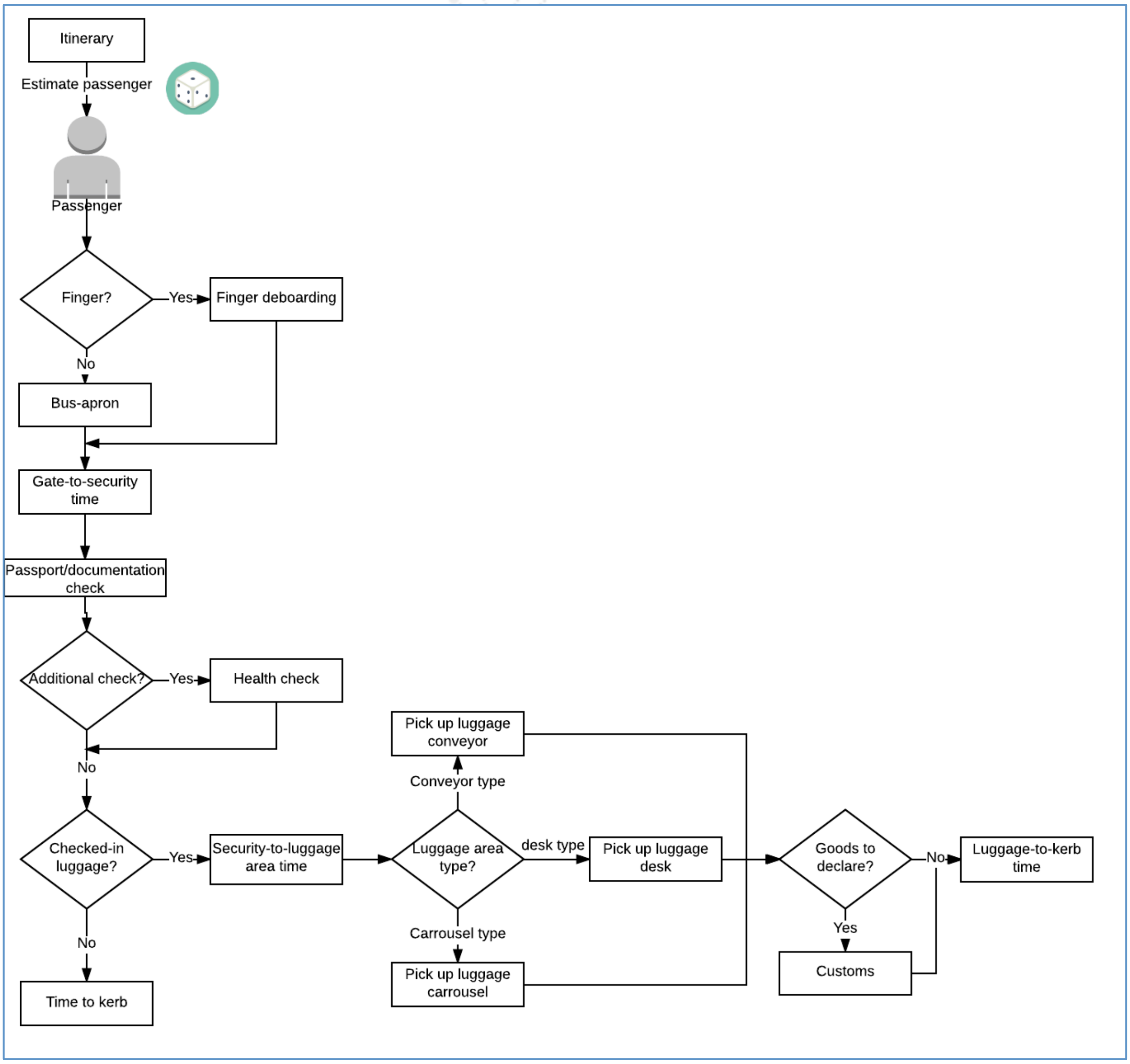


Figure 27 G2K flow diagram

4.5 Kerb-to-Door (K2D)

Specifically "kerb-to-door" is 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).

This step will be modelled similarly and symmetrically to the door-to-kerb process, taking into account the inherent differences, as documented in section 4.1 of the current document. Those elements that need to be taken into account are:

- Inbound and outbound trips
- Full list of possible means of transport: duration frequency, punctuality, comfort and price
- Means of transport availability at kerb and door points, queuing times
- Airports' catchment areas

It is important to note however that the passenger's perception, expectation, cost-elasticity and time-effort value would be different than in the kerb-to-door phase. Since there is no risk of missing a connection, passengers are no longer considering extra time buffers and queuing times may increase as most passengers will not wait at the terminal until their transport arrives, but rather rush to exit the terminal.

Lastly, it is worth mentioning that both choices of transport door-to-kerb and kerb-to-door are determined by the passenger profile and do not necessarily need to be consistent, e.g. departure airport reached by public transport and arrival airport by taxi. This especially holds true if the actual journey length to the arrival kerb was longer than expected, e.g. the passenger missed its original kerb-to-door planned connection.

6. Next Steps: Implementation and Execution Plan

This document defines the theoretical approach and the data and event-driven stochastic door-to-door air transportation mobility model for DATASET. However, the actual implementation of the software will be carried out during the next phase of the DATASET CSA. The software implementation will follow an agile software development methodology called Scrum.

Scrum is an incrementally iterative software development process, although it can be applied to model design in general. Instead of defining a final list of requirements and functionalities, the software is built in an iterative process, focusing on having short development cycles and sequentially increasing functionalities as needed. Each development cycle needs to be verified and validated, so it is necessary to not only define the next cycle functionalities but to also define a collection of input/output tests for each component as well as a collection of test cases/use cases for the whole integrated software.

Finally, inputs from the potential final user requirements will be incorporated as they become available during the requirement capture phase described in section 3

- Key elements of the overall architecture presented in section 4
- A reduced set of events and environment, so that the overall architecture can be tested
- An increase in the number of events to effectively cover all details of each journey phase
- Expand the environment to accommodate the new events.

Building Blocks

The first functional prototype incorporates the simulation architecture presented in section 4, which is composed of the following elements:

- Data Store and scenario loader: set up and design the static database as well as prepare methods and scripts to populate it.
- Event stack and initialization methods: event stack design and implementation, using OOP, includes creating a series of management methods, add/delete/edit events but more importantly the initialization methods, which loads the initial events from the Data Store to start the simulations.
- Environment design: dynamic database design and implementation, ready to work with events: add/remove/edit/access to registers
- Output metrics data-based design: dynamic database designed to store the final results after each model run.

When these components are implemented then the simulation core is ready to handle any sequence of events. The next step would be to create a simplified version of the events presented in sections 4. These events shall fully represent the door-to-door, i.e. include each of the five processes involving door, kerb, gate and forward, but not including all the functionalities. Having this first set of simplified events allows the Scrum methodology to start and also helps to verify the core components work well together.

Model Wireframe

Once the simulation core is implemented and tested using a simplified set of events it is time to fully implement the events as described in section 4. This will be done in an iterative process following the Scrum agile software development methodology. Each cycle has a typical duration of two to four weeks. In each cycle the events are refined and expanded with new functionalities. The priority of those developments, as well as the details are aligned with the potential final user requirements captured as described in section 3.

Each cycle also defines a series of tests to be run on the new elements individually. These tests are run on simplified cases; a collection of input/outputs sufficient enough to be confident on the correct implementation of the model. The new model, once the new events are integrated, will be tested using a series of test cases. A test case is a reduced and controlled input artificially created using a collection of sample outputs (not necessarily final outputs), used just for debugging and verification purposes.

The Scrum iterative process repeats itself until the final model fully incorporates all the elements in section 4, responding to all requirements captured in section 3.

Validation and Verification Plan

Where ***validation*** procedures ensure we are building the right model for the questions that DATASET2050 poses, ***verification*** ensures that the system is built correctly, without errors or software bugs.

The scrum methodology requires a validation of the functional specifications at the beginning of each development cycle, so that validation of the functional requirements will be addressed iteratively.

The verification of the model provides proof that the model has been implemented correctly, according to the specifications previously established by the project. An absolute zero bug state in software programming is a delusion as errors will always appear and unexpected behaviour is always possible. However, it is the task of verification to ensure that the number of errors is dropped to a minimum and that unexpected behaviours have the least possible impact on the overall performance.

Two verification approaches will be used. First, dynamic testing will be used to run a unit test (isolated components or class method) or integration tests (groups of classes and interacting methods) with the model.

Dynamic tests are usually divided into three categories:

- Functional tests using simplified inputs with known outputs to evaluate critical functionalities of the implementation.
- Structural test using simplified inputs to test the code structure. That is to say inputs may not correspond to a realistic scenario but rather a worst-case scenario. This is a case of extreme values testing.
- Sequential or random test is an exhaustive test tool in which inputs are logged into the tested components and outputs are analysed checking for inconsistencies.

For instance, a simplified input set can be used with minimal variation to perform a sensitivity analysis (i.e. expected solutions should be continuous and smooth).

Software management, deployment and execution
One key element of the Scrum methodology is the version control. A version control records changes in a given software, and more sophisticated tools allow change requests, comment and validation of new changes, and even the creation of new software branches that would develop independently. The **DATASET mobility model** is implemented using GitHub, a fully featured version control repository, which stores all data source files online (but on a private repository) and therefore it can be accessed by multiple collaborators so changes in code can be easily tracked, reviewed and commented.

Testing and development is done locally, using in-house computing infrastructures. However, for the final cycles of the development process, the platform will be migrated to a cloud-based environment. This would allow anyone with the right credentials to run the model and produce outputs. The scalability of the cloud-based hardware is critical when more computational power is required. Cloud-based solutions are easily configured and adapted to software requirements.

Considering the inputs and outputs are stored in databases, and all model results will be accessible online, the system will require a series of passwords and security over SQL.

7. Acronyms and Abbreviations

ACARE: Advisory Council for Aviation Research and Innovation in Europe
ACC: Area control centre
ASKs: Available seat kilometres
ATS: Air traffic services
BHL: Short name of DATASET2050 partner: Bauhaus Luftfahrt
CAA: Civil Aviation Authority
CBD: Central business district
CSA: Coordination and Support Action
D2D: door-to-door
D2K: door to kerb
DDMM: DATASET door-to-door mobility mode
DDR2: Demand Data Repository (second phase)
DLR: Deutsches Zentrum für Luft- und Raumfahrt e.V.
DX.Y: Deliverable's name (X=workpackage, Y=deliverable numbering within workpackage)
EC: European Commission
ECTL: Short name of DATASET2050 partner: EUROCONTROL
EEA: European Economic Area
EFTA: European Free Trade Association
ERTMS: European Rail Traffic Management System
EU: European Union
EU-28: European Union 28 member countries (since July 2013)
G2G: Gate to gate
G2K: Gate to Kerb
GDP: Gross domestic product
GHG: Greenhouse gas
GIS: Geographic information system
H2020: Horizon 2020 research programme
HS2: High Speed 2 (planned rail link)
HS3: High Speed 3 (proposed rail link)
ICT: Information and communication technology
IFR: Instrument flight rules
INX: Short name of DATASET2050 coordinator: Innaxis
ITS: Intelligent Transport Systems
K2D: Kerb to door
K2G: Kerb to gate
LCC: Low-cost carrier
LSSIP: Local Single Sky ImPlementation
MCT: Minimum connecting time
MPPA: Million passengers per year
NUTS: Eurostat's hierarchical classification of spatial units (NUTS1 - NUTS3)
O&D: Origin and destination
OECD: Organisation for Economic Co-operation and Development
PAV: Personal air vehicle
R&D: Research and development
RIS: River information services
RPKs: Revenue passenger kilometres
SESAR: Single European Sky ATM Research
SRIA: Strategic Research and Innovation Agenda
UNSD: United Nations Statistical Division
UOW: Short name of DATASET2050 partner: University of Westminster
USD: United States dollar
VFR: Visiting friends and relatives
WP: Workpackage