URBANITE

Supporting the decision-making in urban transformation with the use of disruptive technologies

DRAFT VERSION

Deliverable D4.4

URBANITE Traffic Flow Model

Editor(s):	Maj Smerkol, Žiga Kolar, Miljana Shulajkovska, Erik Dovgan, Matjaž Gams
Responsible Partner:	Jožef Stefan Institute
Status-Version:	v1.0
Date:	19.05.2022
Distribution level (CO, PU):	PU

Project Number:	GA 870338
Project Title:	URBANITE

Title of Deliverable:	URBANITE Traffic Flow Model
Due Date of Delivery to the EC:	31. 03. 2022

Workpackage responsible for the Deliverable:	WP4 – Algorithms and simulation techniques for decision-makers
Editor(s):	Jožef Stefan Institute
Contributor(s):	JSI
Reviewer(s):	Iñaki Olabarrieta (TEC)
Approved by:	All Partners
Recommended/mandatory readers:	WP3, WP5, WP6

Abstract:	This document presents the traffic flow model developed during the project. This includes modelling the population and travel demand as well as using a microscopic traffic simulation, calibrated according to real traffic data.
Keyword List:	Traffic simulation, Population modelling, Travel demand modelling
Licensing information:	GPLv2
Disclaimer	This document reflects only the author's views and neither Agency nor the Commission are responsible for any use that may be made of the information contained therein.

DRAFT VERSION

Document Description

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
v0.1	21/02/2022	Draft ToC	JSI
v0.2	02/03/2022	Updates to Sections 3 and 5	JSI
v0.3	17/03/2022	Updates to various sections	JSI
v0.4	23/03/2022	Deliverable ready for internal review	JSI
v0.5	28/03/2022	Reviewed and sent to reviewers	TEC
v0.6	31/03/2022	Updates after review	JSI
v1.0	13/05/2022	Updates to various sections	JSI
v1.1	16/05/2022	Final edits	JSI

DRAFT VERSION

Table of Contents

DRAFT VERSION

List of Figures

DRAFT VERSION

List of Tables

DRAFT VERSION

Terms and abbreviations

EC	European Commission
KPI	Key performance indicator
SR	Synthetic reconstruction
CO	Combinatorial optimization
IPF	Iterative proportional fitting
OD	Origin-Destination
XML	Extensible markup language
HBEFA	Handbook emission factors for road transport
DSS	Decision support system
OSM	Open street maps
LTZ	Limited traffic zone
ISCED	International standard classification of education
DTA	Dynamic traffic assignment

DRAFT VERSION

Executive Summary

The deliverable D4.4 URBANITE Traffic flow model presents the developed traffic flow model and the partial models of the population and travel demand, as well as the calibration of the simulations, and the simulations developed.

This deliverable is part of the work package WP4 and the results of Task T4.3. The presented traffic simulation component is part of the URBANITE solution and provides the capability to simulate the effects of the proposed mobility policies in a synthetic reality. The results of this component are used in other components, visualisations and the URBANITE DSS, described in deliverable D4.3.

The document is split into introduction, providing basic information about the deliverable and the detailed document structure. Following are the sections describing the traffic simulation component, including the population model, travel demand model, calibration, and simulation. Each of these sections focuses on one of the steps that compose the traffic flow mode. The document also provides information on code delivery and usage.

The developed models differ between the four cities as the data available by different use cases is not equivalent. Therefore, models for each of the cities have been considered separately and several methods for modelling have been utilised to develop all the required models.

This document presents the current status of the developed traffic simulation component that will be further improved until the end of task T4.3 in M30. The final version of the URBANITE components, including the traffic simulation component, will be described in deliverable D4.6 Final implementation of the recommendation system for policy design due in M30.

DRAFT VERSION

1 Introduction

Deliverable D4.4 presents the overview of the URBANITE Traffic simulation component. It includes the descriptions of the developed population and travel demand models, calibration methodology and simulation methodology used.

This document is part of the work package WP4 "Algorithms and simulation techniques for decision-makers" and is the outcome of Task T4.3 "Policy simulation and validation engine".

It provides an overview of the traffic simulation component, population model developed, travel demand models developed, calibration methodology, and simulation capabilities. Developed simulations for all use cases are presented.

1.1 About this deliverable

This deliverable introduces the URBANITE Traffic Flow Model, including the population modelling, travel demand, calibration and traffic simulation methodology. The traffic simulation component is responsible for simulating the proposed policies in order to evaluate them and use the results in the URBANITE Decision Support System (DSS), visualisations and data analysis. The traffic flow model is composed of the population model, travel demand model, simulation calibration tool and the simulations themselves.

This deliverable is an intermediate deliverable of Task T4.3. Thus, the final implementation of the traffic simulation component will be available in Month 30 and described in Deliverable D4.6. This deliverable provides the current version of the traffic simulation component, which will be further developed in the future.

1.2 Document structure

This document is organised into sections:

- Introduction: This section explains the rationale of this document and the document structure in more detail.
- Overview of Traffic Simulation Component: This section describes the components of the system and presents the system architecture.
- Population model: This section describes three different developed population models: the basic model used in initial simulations, the census based model, utilising the census microdata to generate the population, and the household based model, which models the population at household level in addition to the representations of the individual agents.
- Travel demand model: This section describes the process of modelling travel demand and presents two different models, the facilities-based model that is appropriate when OD matrix data is not available and the OD matrix based model.
- Calibration: The section describes the methodology used to calibrate traffic simulations. It details the dataset creation and preprocessing and the MATSim co-evolutionary algorithm that is used to calibrate the simulations.
- Simulation: This section describes the simulation capabilities developed and covers in more detail the scenarios, considered for each use case, and the simulations of these scenarios.

DRAFT VERSION

- Delivery and usage: This section provides information on code delivery and usage.
- Conclusion: The last section concludes the document with final thought and outlines the future steps.

DRAFT VERSION

2 Overview of the Traffic Simulation Component

The URBANITE system consists of several components, including a data platform, AI-based tools including the mobility policy simulation, and tools for stakeholder engagement, including a forum and a social policy laboratory. We focus on the architecture of the traffic simulation component, shown in Figure 1.

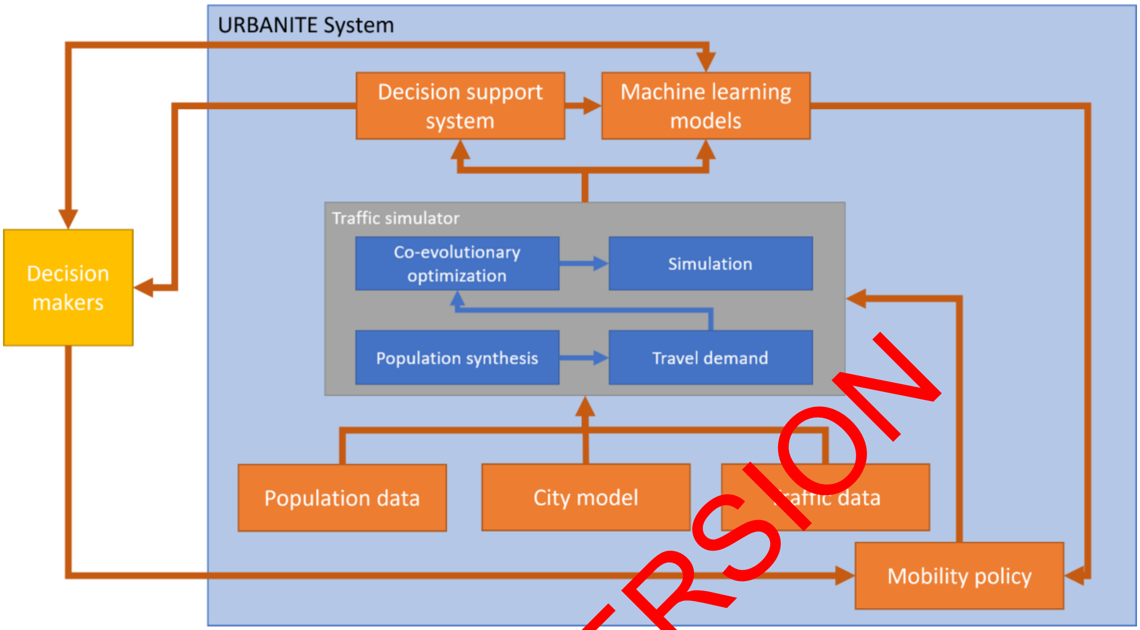


Figure1 - The traffic simulation component architecture.

DRAFT VERSION

The decision-makers work with the URBANITE system in an interactive way by evaluating and improving policy proposals in an iterative fashion, i.e., by defining the mobility policy proposals, which are simulated and evaluated by the system. In each iteration, they use the insight gained to modify the policy proposals. However, the system is also able to search for mobility policy proposals within user-provided constraints. These proposals are automatically simulated and evaluated, and the decision-makers are presented with a selection of the best ones according to the selected KPIs.

The main inputs to the traffic simulation component are the proposed mobility policy and the data required for the simulation: the population data, the city model, and the traffic data.

The traffic simulation component executes the following steps to create and run the simulation:

- Population model: It uses the real citizen population data to create the population of artificial agents. In this document the word population commonly denotes the real citizen population as opposed to the (artificial) agent population.
- Travel demand model: It is built by taking the generated population of artificial agents and their activities, and by generating the trips that the agent will take to arrive at the locations of their activities. The resulting trips are the initial plan pool, available to agents. The plan pool is modified with new and improved plans during the next step, i.e., calibration. Initially, one plan is generated for each agent. During the calibration step, however, more plans are created.
- Calibration: Before the simulation is run, it is calibrated by optimising the trips to fit them to the known traffic data. This is done in several iterations with a co-evolutionary

optimization algorithm. In each iteration of the co-evolutionary optimization, a subset of agents is selected and the plans of these agents are modified (see Section 5 for details). During optimization the activities that the agents have to perform stay unchanged, the activities' times and the way they move between them are modified.

- Simulation: Finally, the simulation run is performed with the MATSim framework for traffic microsimulation [1], and the simulation recorded for further analysis and visualisation. In addition, the simulation results are used by the decision support system to calculate the KPIs, evaluate the mobility policy proposals and provide multi-attribute decision analysis, as well as by the machine learning models used for policy proposals.

The main modules of the traffic simulation component are presented in the following sections.

3 Population model

In agent-based microsimulation models the initial step is to define agents representing individuals, which decisions are simulated over time to model a real-world scenario. Because of privacy reasons, there is no complete dataset containing socio-demographic characteristics of individuals at a small geographic scale. Therefore, in order to perform a microsimulation, a necessary step consists of generating a synthetic agent population that is representative of the real population. During this process, the characteristics of all individuals within a given geographical area are typically inferred from the characteristics of individuals in sample data from that area, as well as from aggregated data about the given area. The resulting agent population is a simplified microscopic representation of the population and consists of only those variables that are of interest.

As two different datasets are used to replicate the population, they must contain the same set of attributes at the same geographical level (e.g., district level). Once the data is available, different techniques can be applied to allocate the individuals from the sample data to different geographical zones while maintaining the constraints given by the aggregated data. Illustration of the agent population generation approach is shown in Figure 2.

DRAFT VERSION

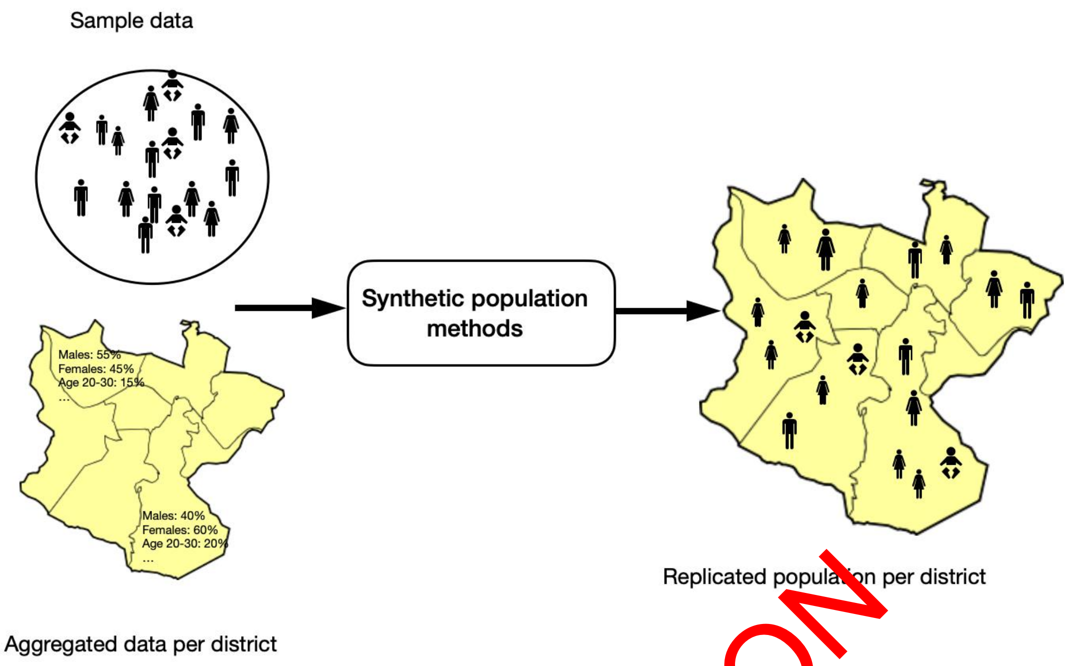


Figure2 - Agent population synthesis.

The methods can be classified into deterministic and stochastic:

- Synthetic reconstruction methods (SR) are deterministic methods that combine information from the sample and the aggregated data, and compute weights that reflect the representativeness of each individual/household in the sample within a given zone [2].
- Combinatorial optimization methods (CO) [3] are stochastic methods that, similarly to SR methods, use the sample and aggregated data to select an appropriate combination of individuals that best fits the marginals. These methods first randomly allocate individuals from sample data to zones one at a time, and then calculate the goodness-of-fit that indicates how good the sample data fits the distribution defined with the aggregated data. If the observed values differ from the expected ones, the individual is removed from that zone.

DRAFT VERSION

The most widely used algorithm for synthetic population generation is Iterative Proportional Fitting (IPF) [4]. IPF is a SR method, which involves calculating a series of non-integer weights that shows how representative each individual is of each zone. Using the sample data the weights are iteratively updated until the marginal distributions per district are satisfied. IPF has been implemented in the URBANITE synthetic population generation module.

The following sections explain several population models with increasing complexity. These are the basic model, the census-based model and the household-based model.

3.1 Basic model

The basic population model is based on the number of inhabitants in each district, and locations of facilities such as schools, hospitals and businesses. These are used as locations for work, education, shopping and, regarding hospitals, other activities. This model was initially developed for testing purposes before most of the data was harvested, and does not include population from outside the cities and trips by professionals, such as deliveries, ambulances and similar.

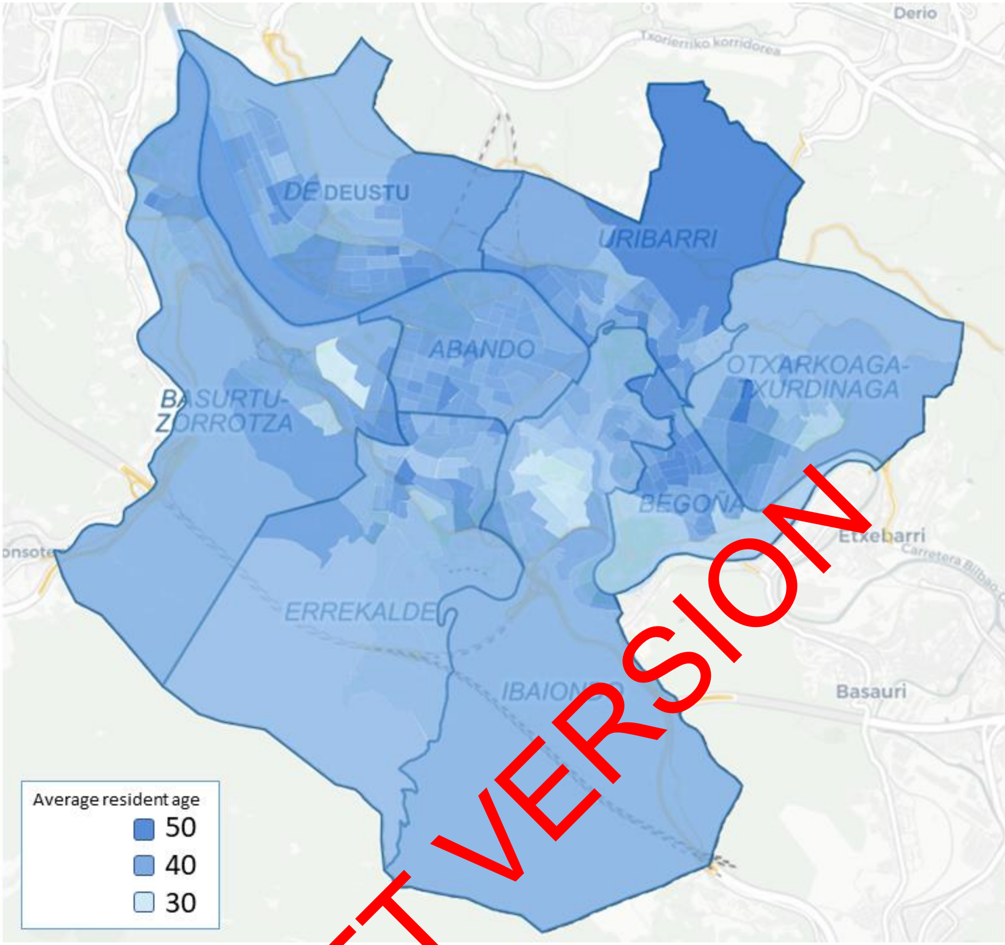


Figure3 - Mean age per district, highest 53 in Uribarri, lowest 31 in part of Bassurtu-Zorrotza.

DRAFT VERSION

Data about districts and their population was gathered from pilot cities' existing open data portals. Figure 3 shows the mean age in each district. Locations and types of facilities, shown in Figure 4, are extracted from harvested OSM data and include the locations of the facility and the type of facility. The facility type is mapped from the OSM keys amenity and building, which describe the location, shape and type of amenities and buildings. Only a few types of the many building types are used, e.g., commercial, industrial and retail. The amenity and building types are categorised into 10 categories with more than 170 unique values. The information on categories is used in the following processing step to generate the final travel demand (see Section 4).

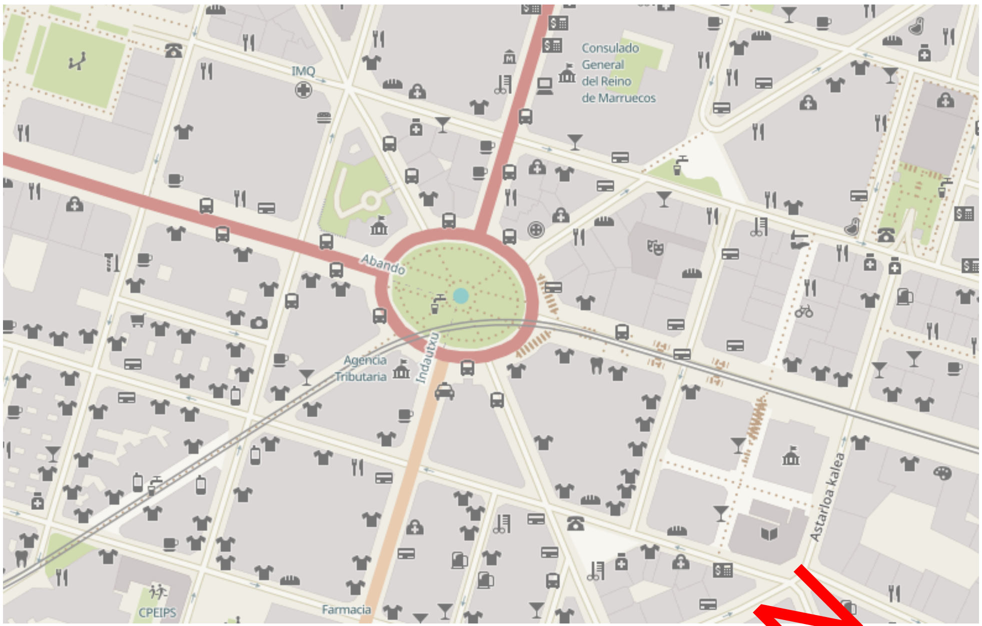


Figure 4 - Facilities near Moyua square in the centre of Bilbao.

The population model in this step generates individuals with attributes age, sex, education level, economic status, and level of leisure activity in each district based on the district's statistical properties, and selects a random location within the district as their home. These attributes are used in the next step to construct the travel demand model (see Section 4).

3.2 Census-based model

The census-based model requires the census data to model the agent population. To this end, the sample data was obtained from EU Statistics on Income and Living Conditions (EU-SILC) [5], while aggregated data was provided by each city of the URBANITE pilots. The model was built with the IPF method.

The EU-SILC aims to collect multidimensional microdata and is used in URBANITE as the main source of population data for microsimulation purposes. The survey provides data related to people and households. The available information, i.e., target variables are stored in the form of four different files:

- Household register file (D-file) contains every selected household, including those where the address could not be found, and also households which could not be interviewed, i.e., all households from the sample.
- Household data file (H-file) contains only households which have been contacted, have completed the interview, and where at least one member aged 16 or over has completed the personal interview.
- Personal register file (R-file) contains all the people currently living in the households and those who are temporarily absent.
- Personal data file (P-file) contains every eligible person (persons aged 16 and over) for whom the information was collected from interview and/or registers.

Each file contains a predefined set of variables which covers different topics, and a set of survey units. To enable data merging and aggregation, each file also contains some of the key variables: year of the survey, country, and ID, as shown in Figure 5. Each person has an ID

DRAFT VERSION

containing the household ID of which is part, and a personal number. In that way people are grouped in households.

Figure 5 - Merging Personal data (P-file) with Personal register (R-file).

The EU_SILC microdata files (sample data) are available publicly in order to explore the content and complexity of the data. This kind of data is synthetically generated using actual data, but the structure is the same as confidential microdata which can be requested for further improvement of the modes and represents real data.

The demographic data are provided by each of the pilot cities but lack the information about persons grouped in households. Therefore, the generation of agent population is made at individual level and the statistical attributes are replicated at district level, meaning that the agent population of each district has the same statistical properties as the population of the district. Table 1 shows the summary of statistical properties of each pilot city's population.

DRAFT VERSION

Table 1. Summarization of the generated agent populations and included attributes.

City	Number of districts	Age	Sex	Education level	Income level	Leisure participation	Number of inhabitants
Bilbao	8	Yes	Yes	Yes	No	No	317694
Amsterdam	23	Yes	Yes	No	No	No	836143
Helsinki	8	Yes	Yes	Yes	Yes	Yes	639189
Messina	6	Yes	Yes	No	No	No	492482

Having the needed data at the same geographical level of interest is of crucial importance in the process of creating the census model. Individuals represented by age groups, sex and district were replicated while maintaining the demographics given by each city. Before applying any algorithm, the data needs to be preprocessed, which includes removing excess information, re-categorization of individual level attributes in order to match the aggregated level constraints, and 'flattening' the individual level data. By 'flattening' each variable is expanded to all its categories. Aggregated flattened data is shown in Figure 6.

Once the data is ready, the IPF method is applied to compute the weights that quantify how much each individual is representative to each zone. Since the computed weights are floats, the TRS (Truncate, Replicate, Sample) method is used to convert the weights into integers with a minimum loss of information [6]. High weights must be sampled proportionally more frequently than those with low weights. The following example illustrates that process:

$$v = (0.333, 0.667, 3), int(v) = (0,1,3)$$

The remaining step to create the agent population is the expansion process [7], where each weight is rounded to a whole number and corresponds to the number of needed repetitions for each individual. Examples of the obtained agents/individuals are shown in Figure 7.

```
##    a0_49 a50+ m f
## 1      8    4 6 6
## 2      2    8 4 6
```

Figure 6 - Aggregated constraints showing individual's age category (here simplified into 0-49 years old and 50+ years old) and sex for each district. Each line represents a district, showing the number of people in different age brackets and gender.

```
##   ind$agea0_49 ind$agea50+ ind$sexm ind$sexf
## 1            0           1        1        0
## 2            0           1        1        0
## 3            1           0        1        0
## 4            0           1        0        1
## 5            1           0        0        1
```

Figure 7 - Individuals'/Agents' attributes recategorized and flattened to match the aggregated constraints. Each line represents an agent/individual. The values and attributes shown are simplified.

Thus, the population of the city is represented. This agent population is the basis for travel demand generation as described in Section 4. Nevertheless, travel demand generation also considers several other types of travelling, including professional trips (such as deliveries and goods transport) and traffic originating outside of the city.

DRAFT VERSION

3.3 Household-based model

Household-based model is an advanced model that upgrades the census-based model by including the data about households, such as number of adults and children in a household, household income bracket, and number of cars a household owns.

The data used to reconstruct the households is included in EU-SILC, besides the personal registry and personal data file. The household register file (D-file) is merged with the household data file (H-file), as shown in Figure 8.

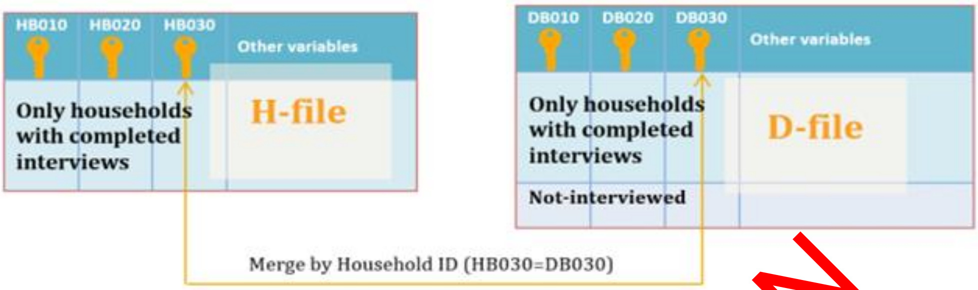


Figure 8 - Merging Household data (H-file) with Household register (D-file).

The population generation methods used are similar to the census-based model. The methods include two steps:

1. Generate households using IPF from the household data. Each household has attributes such as number of adults, number of children, number of cars owned and income bracket.
2. Generate the persons using IPF. Besides the personal data, households are used as a template to be filled with persons, e.g., a household with one adult and one child produces one person with age in the adult range (18-65 years old) and one person with age in the child range (less than 18 years old).

The resulting population model is considered more representative than the census-based model, and it also includes the household data, relevant for travel demand modelling (see Section 4) to estimate shared rides, e.g., parents taking children to school. Table 2 summarises the census data used for each city, and Table 3 presents the relevant data types used across all use cases.

DRAFT VERSION

Table 2. Summarization of census data per city.

City	Number of districts	Age	Sex	Number of households	Number of inhabitants
Bilbao	8	Yes	Yes	68913	317694
Amsterdam	23	Yes	Yes	383483	836143
Helsinki	8	Yes	Yes	330933	639189

Messina	6	Yes	Yes	120166	492482

Table 3. Attributes used from different sources. The personal microdata and the household microdata also include personal and household IDs, not mentioned in the table

Data source	Attributes
Census	Age group
	Sex
	Income group
	Car availability
	Education level
Microdata - personal	Age
	Sex
	Status in employment
	ISCED level achieved
	Regularly participate in a leisure activity
Microdata - household	Total household income

DRAFT VERSION

Like the census-based model, this model also represents only the population in the city. Other trips, such as professional trips and out of city trips are considered separately in the travel demand model, detailed in Section 4.

4 Travel demand model

The travel demand is the core part of running a simulation. Traditionally the generation of travel demand is a four-step process that refers to: trip generation, trip distribution, mode choice and traffic assignment [8], as shown in Figure 9. In other words, the model produces plans with attributes departure time, origin, destination and travel mode, but not details such as routes between activities. The result of the travel demand process is the initial plans pool. This pool is then further modified and improved in the next step, i.e., calibration (see Section 5), which uses several mutation methods to generate new plans.

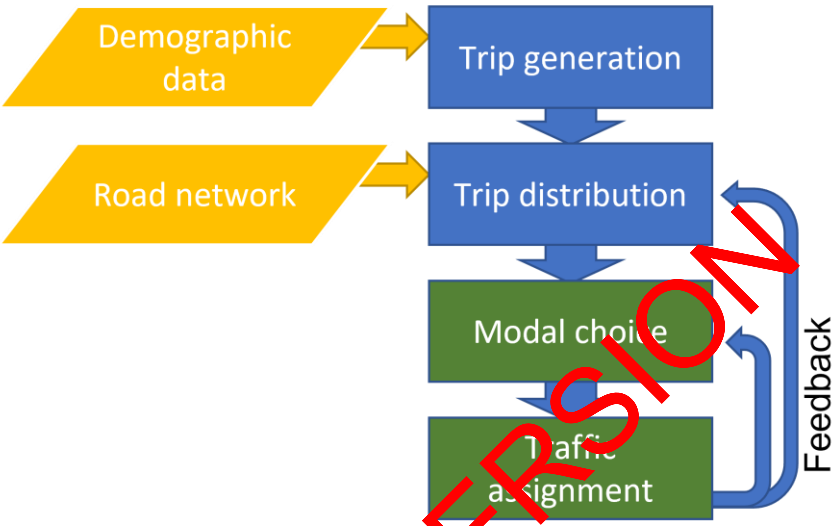


Figure 9 - The four-step travel demand calculation process.

DRAFT VERSION

The four steps of the travel demand calculation process are as follows. In the first step, demographic data is used to generate trips. In the second step, the road network is used to distribute the trips over the city. The third and fourth steps, modal choice and traffic assignment are performed as part of the MATSim loop [1].

The first step, trip generation, is based on the data from the population model, and focuses on generating the trips that should be performed in a day. The population model provides the data about agents, such as their age, which is then used to probabilistically determine whether an agent's primary activity should be education, work or something else. The result of this step are the activities that should be performed by each agent, i.e., individual. Based on the locations of activities, districts are assigned the trip production and trip attraction numbers. The trip production number determines the number of trips that start in a district, while the trip attraction number determines the number of trips that end in a district. These are used in the next step to distribute the trips among districts.

The second step, trip distribution, takes the activities generated during the trip generation step, and combines them into trips. The numbers of trip production and trip attractions are used to generate the sequences of activities. The results of this step are the origin and destination of each trip.

The modal choice selection is the third step. In this step, trips are processed into sequences of trip legs, one or more of which make up a trip. There are many different options on how to get from point A to point B using different trip modes, e.g. drive using a car from A to B, walk from A to the nearest bus station, take the bus, walk from the bus station to point B, etc. To choose the travel modes, modality data need to be provided, such as modal split OD-matrices or modal split data from traffic counters, capable of vehicle classification.

The final step is the traffic assignment/routing. In this step, the shortest path for each trip leg, according to the chosen mode, is selected. The traffic assignment is improved in several iterations with the co-evolutionary optimization, designed to achieve user equilibrium and calibrate the simulations [9], as described in Section 5. An example of a resulting plan is shown on Figure 10:


```
<person id="1000">
  <attributes>
    <attribute name="age" class="java.lang.Integer">65</attribute>
    <attribute name="car availability" class="java.lang.String">always</attribute>
    <attribute name="employed" class="java.lang.Boolean">false</attribute>
    <attribute name="gender" class="java.util.Collections$UnmodifiableCollection">male</attribute>
  </attributes>
  <plan selected="yes">
    <activity type="home" link="34455" x="661793.5587970266" y="962514.543380156" end_time="07:42:44" >
    </activity>
    <leg mode="car">
    </leg>
    <activity type="shopping" x="662113.9807961611" y="962018.3213639692" end_time="09:56:27" >
    </activity>
    <leg mode="car">
    </leg>
    <activity type="home" x="662125.5181128547" y="963229.9876738016" end_time="15:43:13" >
    </activity>
    <leg mode="car">
    </leg>
    <activity type="education" x="663782.1032156559" y="962112.1045874701" end_time="19:13:30" >
    </activity>
    <leg mode="car">
    </leg>
    <activity type="home" x="661793.5587970266" y="962514.543380156" >
    </activity>
  </plan>
```


Figure 10 - An example plan, built using census-based population model and facilities-based travel demand model.

DRAFT VERSION

The following sections describe two versions of the travel demand model, i.e., the facilities-based model and the OD matrix-based model. The reason we developed two different travel demand models is that the different pilot cities have different data available. The facilities-based model is simpler and less accurate but has the advantage of not requiring the origin-destination matrices that are not available for all the use cases.

4.1 Facilities-based model

As already mentioned, the travel demand model is generated based on data from the population model. The population model includes attributes, such age, sex, and home district. This section describes the creation of the travel demand model where the activity locations are selected from the appropriate group of facilities. As a result, each agent gets assigned a daily plan of activities, where each activity is described with location, start, end time and travel mode. The algorithm that produces the initial agents' plans is stochastic. The activities are selected randomly, with weights of options based on the agent's properties.

Plans contain several activities, connected with trips. All plans start with the home activity, and each plan has at least one primary activity, most commonly work or education, depending on the agent's attributes. Most agents also get assigned a secondary activity, commonly leisure or shopping, using the agent's leisure regularity status (attribute 'regularly participates in leisure activities'). At most, an agent has 7 activities per day, with a median of 3.

The work activities have start and end times assigned according to typical work times with variations. We consider morning and afternoon shifts. Similarly, are defined common education times. Leisure and shopping activities are less constrained by time, considering that shopping can only be done within typical opening times.

The plans are further modified in the calibration step. During calibration, several free parameters can be modified, including mode choice according to the population model (agent's attributes), routing, and departure times. This is further described in Section 5.

The facilities-based model uses specific locations of various facilities, such as businesses, schools and hospitals to assign the locations of activities. These facilities are grouped by category, such as work, leisure and education. Figure 11 shows locations of different facilities and their categories as provided by the Amsterdam open map portal.

Figure 11 - Locations and types of different facilities in the city, as visualised by the Amsterdam open map portal.

DRAFT VERSION

The first step, trip generation, is the same for both methods, i.e., facilities-based and OD matrix based. Activities are generated based on the population model and used in the next step, i.e., trip distribution. Types of activities and the agents that perform them are shown in Table 4 at an overview level.

Table 4. An overview of activities, the facility categories the activity can be performed at, and the corresponding agent attribute requirements.

Activity category	Facility category	Agent attributes
Work	Services, work, education	Age > 18, age < 65
Education	Education	Age < 25
Leisure	Services, other	Age > 13
Shopping	Services	Age > 13

The second step, trip distribution, presents the main difference between the facilities based model and the OD matrix based model. Activities are distributed across the facilities in the appropriate category, assigning each activity a location. The locations are assigned to activities from the list of appropriate facilities randomly, weighted by the facility size and popularity. This method is not based on reproducing known movements, but nevertheless the appropriate weights for facility size and popularity, combined with the populations of different districts, enables us to estimate the travel demand when better data is not available.

The third step is modal choice, based on the agent's attributes. Each agent can choose any travel mode available except for the car mode, which the agent needs access to. Possible values of the attribute 'has access to car' are 'always', in which case the agent's first choice will be driving, 'sometimes', meaning the first choice is selected randomly, and 'never', in which case the agent cannot drive. Because the agent's attributes are generated stochastically, the value of the attribute is based on census statistics and weighted by income level, meaning that agents with higher income will more commonly have access to a car than those with lower income according to [10]. The process is presented in Figure 12.

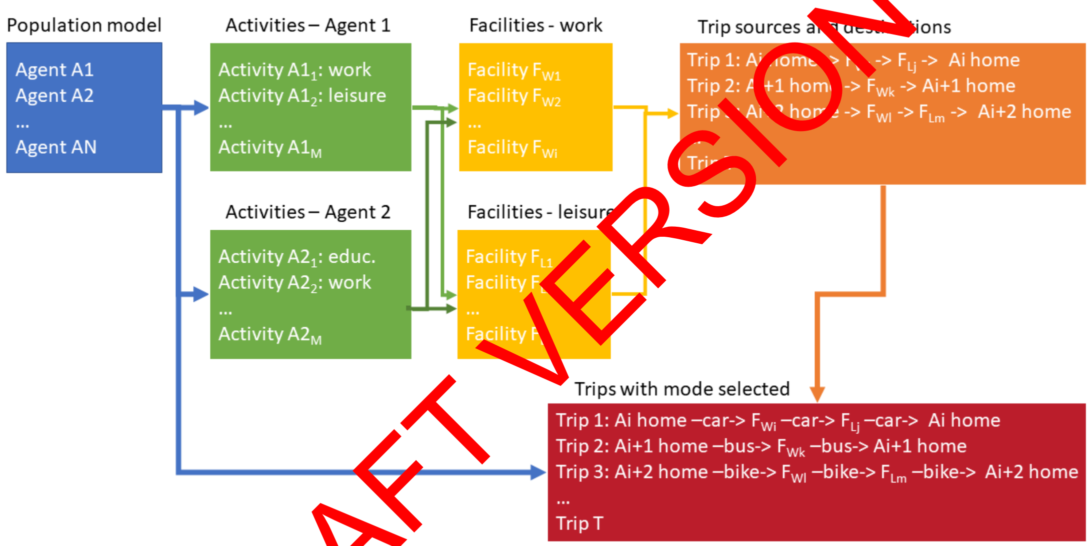


Figure 12 - Trip generation, distribution and assignment steps of the facility-based model.

DRAFT VERSION

The last step, traffic assignment, is performed as part of the MATSim simulation loop (see Section 5). The trips are optimised with a coevolutionary algorithm to reach a user equilibrium and match the calibration data set.

4.2 OD matrix-based model

Once the synthetic population is generated, the assignment of activities takes place, as already described in Section 4.1. Every person gets assigned a daily plan of activities, where each activity is described with location, start, end time, and travel mode. In order to generate the activity's location, Origin-Destination (OD) matrices are commonly used. In these matrices, the rows and columns represent the origin and destination zones respectively.

	Destination zones						
Origin zones		1	....	j	....	n	sum
	1	t_{1-1}		t_{1-j}		t_{1-n}	O_1
	. . .						
	i	t_{i-1}		t_{i-j}		t_{i-n}	O_i
	. . .						
	n	t_{n-1}		t_{n-j}		t_{n-n}	O_n
	sum	D_1		D_j		D_n	T

Figure 13 - Origin – Destination Matrix.

The OD matrix is shown in Figure 13. The cell element denoted by t_{i-j} is the number of trips originating from i and destined to j. Therefore, an OD matrix has directional meaning and its cell elements represent trip flows. The row sum is the total number of trips originating from zone i and the column sum is the total number of trips destined to zone j.

Table 5 shows representation of the OD matrix used for the trip generation for the Bilbao use case. Only origin from one district is shown, i.e., Basurto Zorrotza. "Volume" column denotes number of trips originating from district in "From location" to district in "To location" columns. Values contained in the "Time" column show the average trip time in minutes.

The main difference between the facilities-based model and the OD matrix based model, shown on Figure 14, is in the trip distribution step. This model bases the activity locations to districts based on how common trips between the districts are. Currently, the activity locations within the district are assigned randomly. The final model will use the facility-based approach to determine the activity locations within districts, selected based on OD matrices.

DRAFT VERSION

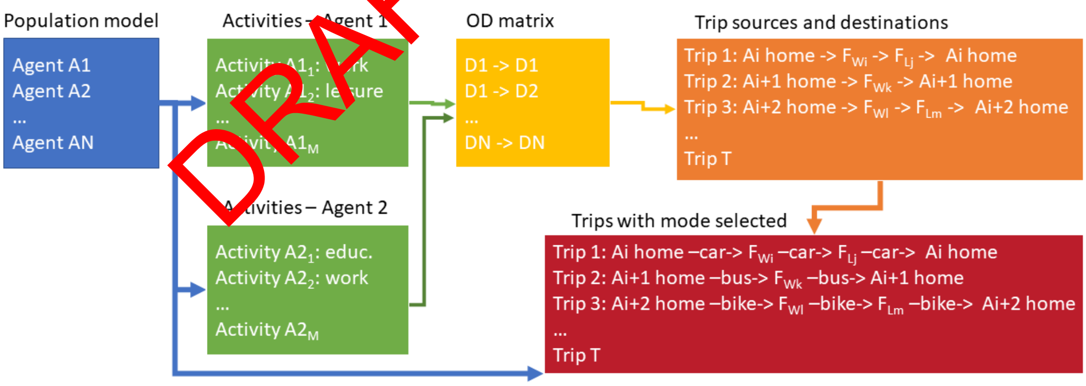


Figure 14 - Trip generation, distribution and assignment steps of the OD matrix-based model.

Table 5. OD Matrix representation (Bilbao).

From location	To location	Volume	Time (mins)
BASURTO ZORROTZA	DEUSTO	2623	48
BASURTO ZORROTZA	URIBARRI	2869	54
BASURTO ZORROTZA	OTXARKOAGA TXURDINAGA	773	68
…			
BASURTO ZORROTZA	BASURTO ZORROTZA	31771	42
BASURTO ZORROTZA	TOTAL	98381	51

DRAFT VERSION

5 Calibration

Quality of microsimulation results greatly depends on the extent to which they are calibrated to the real traffic data in the target geographical area. Calibration is defined as the process of comparing and minimising the differences between the modelling results and the real data obtained by counting and measuring traffic in the local network. The chosen traffic indicator for microsimulation calibration are the traffic counts. These counts are calibrated using the MATSim co-evolutionary algorithm for calibration. [11]

5.1 Data set preprocessing

The data about traffic volumes needs to be preprocessed in order to be used for simulation calibration. An example of the required data format in XML is shown in Figure 15. The attribute loc_id denotes the id of a MATSim link, while cs_id denotes the id of the counter sensor. Every line contains the hour and traffic counts aggregated for that hour.

Ideally, for each travel mode (except for public transport, which drives according to the lines and schedules), separate counts should be used. Currently this is not used due to limited data availability, but it is supported.

```xml
<count loc_id="11" cs_id="link_11">
    <volume h="1" val="0" />
    <volume h="2" val="18" />
    <volume h="3" val="0" />
    <volume h="4" val="55" />
    <volume h="5" val="170" />
    <volume h="6" val="213" />
    <volume h="7" val="661" />
    <volume h="8" val="817" />
    <volume h="9" val="897" />
    <volume h="10" val="1339" />
    <volume h="11" val="833" />
    <volume h="12" val="708" />
    <volume h="13" val="1084" />
    <volume h="14" val="687" />
    <volume h="15" val="653" />
    <volume h="16" val="530" />
    <volume h="17" val="1071" />
    <volume h="18" val="897" />
    <volume h="19" val="519" />
    <volume h="20" val="562" />
    <volume h="21" val="915" />
    <volume h="22" val="2[illegible]1" />
    <volume h="23" val="3[illegible]" />
    <volume h="24" val="[illegible]7" />
</count>
```

Figure 15 - Example of the XML file containing input counts needed for the calibration tool. Hourly average recorded number of passing vehicles is provided for each link where a traffic counter exists.

DRAFT VERSION

The calibration dataset is created using real traffic data from a selected time period and from all available traffic counters. The traffic counters include induction loops, traffic cameras, and radar-based systems. These provide different data attributes, such as the count of vehicles, vehicle classification, and vehicle speed. The calibration algorithm uses the vehicle hourly counts.

To achieve good calibration results, we use only the time ranges with high quality data. The data is aggregated from the 5-minute counting periods into 1-hour long periods according to the requirements of the calibration algorithm. The data with lower temporal resolution is averaged over different days to construct the dataset of a representative day. Using an average over a different selection of days, different calibration data sets can be created, e.g., rainy days, sunny days, weekends, workdays or holidays. Figure 16 shows the process for creating different data sets.

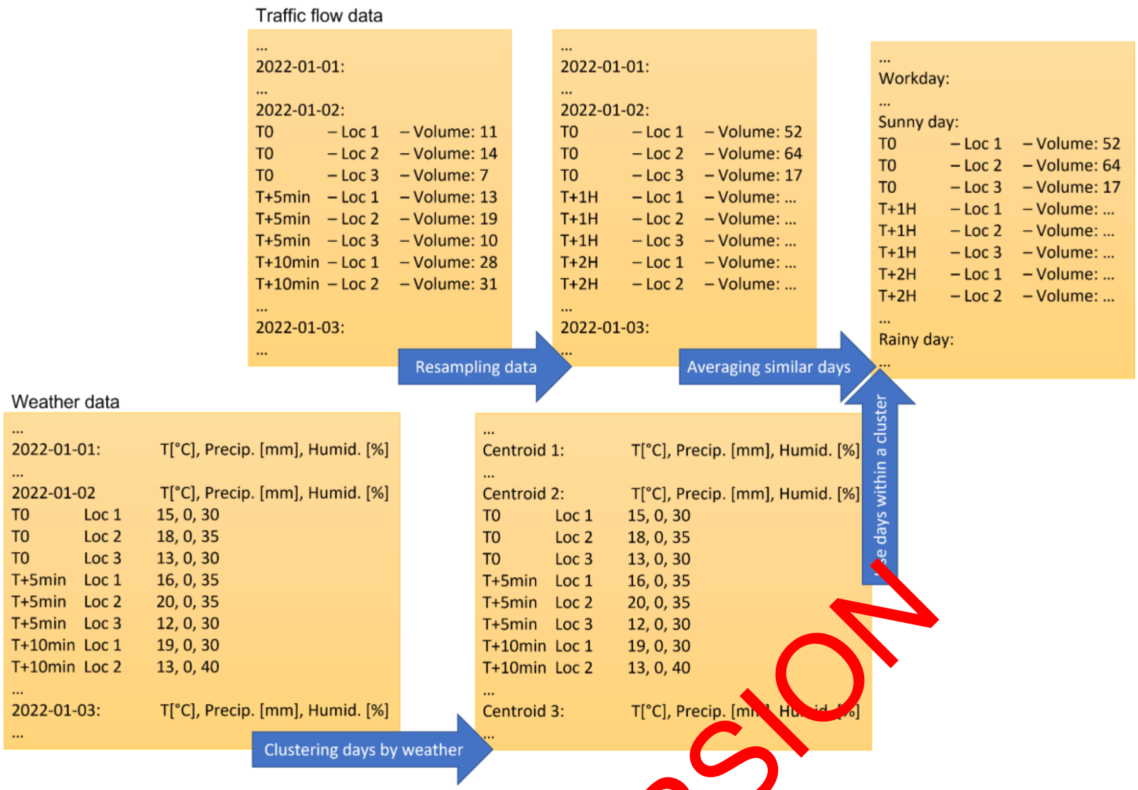


Figure 16 - Traffic flow data is resampled to temporal resolution of 1 hour. Weather data is used to cluster days, e.g. into sunny and rainy days. Similar days are averaged to produce the final calibration data set.

This method allows for calibration to different types of traffic, including different types of weather, workdays and weekends, days with football games etc. by clustering the available data by different attributes or combinations of attributes. The Traffic prediction module, developed as part of Task T4.1, can also be used to produce the calibration data. Thus new calibration datasets can be generated by specifying the desired attributes and generating the traffic count values.

5.2 MATSim co-evolutionary algorithm for calibration

The simulation calibration is performed using the Cadyts (Calibration of Dynamic Traffic Simulations) tool [11]. Cadyts calibrates disaggregate travel demand models of DTA (Dynamic Traffic Assignment) simulators (such as MATSim [1], SUMO [12] or Dracula [13]) by using sensor data such as traffic counts.

The replanning step modifies existing plans and produces new. Several strategies for plan mutation are used:

- Time allocation mutation strategy shifts the activities end times. Mutated plans will have different departure times, which may cause change in the order of activities.
- Route mutation strategy allows the agents to change the routes they move on. The route is calculated just before execution of the plan and stored to be re-used in later simulation iterations. This mutation strategy allows rerouting between iterations.
- Mode mutation strategy changes the transport mode used to travel between activities. Either all transport modes in a daily plan or just a single leg between two activities can be changed.

DRAFT VERSION

The calibration adjusts the plan choice of all agents such that the agents' behaviour in the simulated network conditions is consistent with the (real) traffic counts. The choice of a plan depends on the complete set of departure times for all trips contained in the plan. Since the plans are updated iteratively, the calibration is run jointly with the simulation until (calibrated) stationary conditions are reached. Therefore, the tool has access to the simulation in order to affect the plan choice and observes the simulated network conditions to evaluate their deviation from the traffic counts. The MATSim loop with the calibration modules is shown in Figure 17.

MATSim is a DTA simulator that tracks trip sequences of individual agents throughout the entire modelling process instead of using OD matrix-based demand representation. Since both Cadyts and MATSim are implemented in Java, they can be linked through function calls. The calibration works as follows: whenever an agent chooses a plan, it proposes this plan to the calibration, which either accepts or rejects the plan. If the plan is rejected, the agent draws again from its plan choice distribution until the new plan is accepted. The calibration makes the best effort to comply as much as possible with the calibrated choice distribution within a limited number of trials.

As shown in Figure 17, the MATSim loop consists of five components/steps.

1. The initial demand is the result of the travel demand model, described in Section 4. The initial demand is a list of each agent's plan.
2. Simulation (matsim) of the agents executes the plans in simulated reality (supply simulation). This step simulates the travel supply, which attempts to satisfy the travel demand model.
3. Scoring considers the performance of the plan in the simulated reality. Each plan is scored according to the time, distance, and mode of the trip. Scores include a penalty for being late or early to the planned activity.
4. Replanning removes plans with bad scores according to the utility function and modifies them (demand simulation). The utility function considers each trip's time spent travelling, length, travel mode, calibration data and other parameters. Afterwards, some plans are randomly selected for mutation and crossover, generating new plans. These plans are simulated and scored in the next iteration (steps 2 and 3).

Steps 2, 3, and 4 are repeated until the Nash equilibrium [14] is reached.

5. Analysis, visualisation, and postprocessing of the simulation results.

DRAFT VERSION

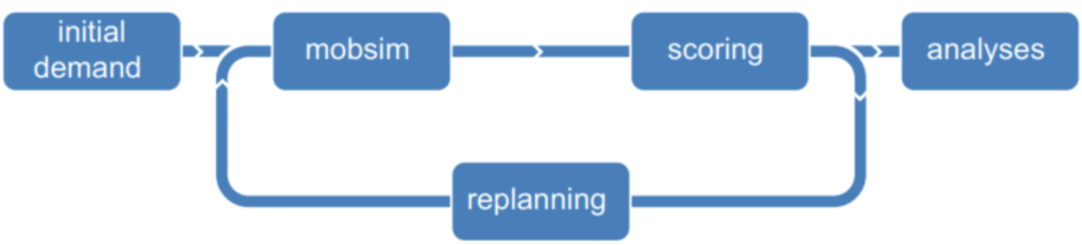


Figure 17 - The MATSim Loop. Calibration is active in the mobsim and replanning phases of the simulation.

The full explanation of the calibration algorithm is beyond the scope of this document. However, it is briefly presented here [15]. Some definitions are required as a starting point:

- **Agent** $n = 1...N$ represents one agent of the population of agents of size N.
- **Selected plan** is one of the agent's available plans $i_n \in \{i_1,...,i_N\}$, written as $\{i\}$ in the following text.
- **Supply simulator** and demand simulator are two parts of a DTA simulator. The supply simulator executes all of the agents' selected plans simultaneously in synthetic reality, resulting in dynamic network conditions $\underline{x}$, including traffic flows $\underline{d}$, densities, velocities, and other properties relevant to the agents' decision making.

  The supply simulator solves for $\underline{x}$ from a distribution of network conditions resulting from the selected plan set with probability $p(\underline{x} \mid \{i\})$. The network conditions vector $\underline{x}$ is implicit in a microscopic traffic simulation and represents the specific network conditions for the set of selected plans $\{i\}$, resulting from the simulation run, not solved algebraically.

- The other part is the **demand simulator**, which models the agents' decision making. Agents select their plans based on the given network conditions $\underline{x}$ by sampling the probability distribution $P(\{i\} \mid \underline{x})$. The agent's choice set C_n is the set of possible plan alternatives. An assumption of independence is required for agents' plan choice, which can be interpreted as agents not communicating their intentions to each other, which is reasonable. Therefone, the plan choice distribution can we expressed as

$$P(\{i\}|\underline{x}) = \prod_{n=1}^{N} P_n(i_n|\underline{x})$$

- Since the traffic simulation uses **iterative simulation logic** to achieve the Nash equilibrium, both the supply and demand simulators are performed many times. The demand simulator uses the previous run's results as the network condition vector $\underline{x}$, and the supply simulator uses the newly selected results to produce the next iteration's network condition vector.
- Finally, the calibration is based on scaling the probability distribution of plan selection according to the measured traffic volumes $\boldsymbol{y}$ on specific links in the network. The calibration aims to adjust the plan selection such that the observed traffic volumes y are reasonably reproduced by identifying the plans that maximise the **posterior entry function**

$$\boldsymbol{W}(\underline{\boldsymbol{d}}|\underline{\boldsymbol{y}}) = \boldsymbol{ln\ p}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}}(\underline{\boldsymbol{d}}) + \boldsymbol{W}(\underline{\boldsymbol{d}}) \qquad \sum_{i \in C_n} \boldsymbol{d_{ni}} = \boldsymbol{d_n}\ \forall \boldsymbol{n} = \boldsymbol{1}...\boldsymbol{N}$$

  where the vector of all traffic volumes $\underline{d} = (d_{ni})$ is a function of the dynamic network condition vector $\underline{x}(\underline{d})$ and W is the posterior entropy function. The prior is considered the uncalibrated state and the posterior includes calibration data y.

Thus, the measured traffic flows are used to alter the probability distribution of agent's plan choice. This explanation does not include several details, including the logit model of the final plan choice and its parameters, but illustrates the inner workings of the calibration module.

DRAFT VERSION

6 Simulation

The final stage of the traffic flow modelling is the traffic simulation, built on top of the population model and the travel demand model. An open source solution is used to simulate the traffic, a Java framework for multi-agent microscopic traffic simulation, MATSim [1].

The population model is the basis of the travel demand model. Based on population attributes, different agents are created with specific goals and activities they have to perform.

The traffic demand model is utilised to generate initial agent plans, which are further optimised during the calibration step by introducing mutated plans and removing those with low scores. The final simulation run provides the simulation results.

The simulation requires the road network and travel demand model. The simulation tool is designed to model a single day based on co-evolutionary principles [9]. Therefore, each individual repeatedly optimises its daily activity schedule while in competition with space-time slots with all other agents on the transportation network. Each iteration starts by simulating the demand on the network, continues by scoring the executed daily plan, and finishes by choosing a plan for the next iteration. The agents learn by maintaining multiple plans that are evaluated. Based on the evaluation scores the plans are selected and modified if needed. The scores are computed based on a scoring function that takes into account the performance of activities and travel times. A typical score is calculated as follows [1]:

$$S_{plan} = \sum_{q=0}^{N-1} S_{act,q} + \sum_{q=0}^{N-1} S_{trav,mode(q)}$$

The scoring function is represented as a sum of all activity utilities ($S_{act,q}$) plus the sum of all travel utilities ($S_{trav,mode(q)}$). N represents the number of activities.

The files produced as output of the simulation include statistics for each of the interactions of the calibration step, and detailed description of the final simulation. Every action taken by agents in the simulation generates an event, which is recorded and then used for analysis. This includes events such as end of activity, departure, entering a vehicle, vehicle entering a link, arriving at the destination, starting an activity and others. Each event has one or multiple attributes, including the time when the event has occurred, the id of the agent triggering that event, and the link id where the event has occurred. In addition, statistics about distance travel per mode, number of agents travelling per mode, trip durations, etc. are also calculated.

The simulation results are further analysed, visualised and evaluated. For each of the use cases specific KPIs are calculated, as well as some common ones. For example, the amounts of many different air pollutants are calculated utilising the Handbook Emission Factors for Road Transport (HBEFA) [16]. These, along with other data, are used by the URBANITE Decision Support System (DSS), which is the topic of deliverable D4.3.

The following sections detail the simulation scenarios that are being developed for each use case.

6.1 Use case: Amsterdam

The Amsterdam use case is focused on bicycle traffic. Amsterdam is one of the cities with the largest number of bicycles per resident and already most of the traffic in the city is bike traffic. While this is a great achievement of the city and good from ecological, public health and traffic management points of view, it also brings unique problems and challenges.

DRAFT VERSION

Bicycle traffic jams and congestions are fairly common in the city and at peak times often cause delays. Bicycle traffic is spilling from bike lanes to the streets and sidewalks as well, which can endanger the cyclists, pedestrians and general road safety. In some places there is also danger of night-time petty crime targeting cyclists.

Bicycle traffic is the main focus of the Amsterdam use case. To represent the bike traffic accurately, we allow agents using bicycles to use any appropriate link, not only bike lanes. This includes walkways, streets and roads, but not stairways or motorways.

While these problems are already present in the city, there are plans to rebuild the Amsterdam Noord district to increase the population density and provide much needed housing. Increased population density is expected to exacerbate the already present problems and the city is looking for ways to improve the situation.

Two scenarios have been envisioned for the Amsterdam use case. The first introduces a limited traffic zone on one of the major collector roads, Oranje Lopper. The second represents the plans for increasing the population density in the Amsterdam Noord district. Following are the descriptions of the use cases, including the baseline scenario, i.e., the current state.

6.1.1 Baseline scenario

Current state of Amsterdam urban traffic is represented by a simulation, based on data available. This includes the population model, travel demand model and calibration data set.

The road network is extracted from the OSM data, cropped to the relevant area and preprocessed to remove disconnected roads near the edges of the simulated area. Public transport data is added to the network, enabling the simulation of public transport. The road network is somewhat simplified to increase simulation performance, e.g. circular road segments are simplified into a diamond shape. The final baseline network represents the road network and includes information about infrastructure, such as detailed road types, road segment lengths and similar.

The current population model is built using the census-based model, described in Section 3. The census data is acquired from the Eurostat web page and represents the population of the Netherlands, as the city currently does not have access to city or regional census. Generally, the national statistics differ from the statistics of the capital city, e.g. national averages may have different education levels, people work in different industries etc. As explained in Section 3.2, the census data is combined with local data about districts. This can offset the misrepresentation resulting from using the national census. The result of the population modelling is a set of agents, each representing a citizen, with attributes age, sex, home location, and work location.

The travel demand model is created from the population model using the facilities-based model. The data about the facilities in the city are extracted from the OSM map service and mapped to districts. The resulting travel demand model is a set of plans that specific agents should perform during the day. These are further optimised during the simulation.

The simulation of the baseline Amsterdam scenario is iteratively optimised using the co-evolutionary approach described in Section 5.2. The final results are the moment-by-moment locations and states of all vehicles.

Finally, to calculate the baseline values of various KPIs, the simulation results are processed to get the air pollutants emissions, bikeability, bike safety and other KPIs. These are used further

DRAFT VERSION

in the decision support system of the URBANITE solution and described in detail in deliverable D4.3.

6.1.2 Oranje Loper LTZ scenario

The first scenario, developed for the Amsterdam use case, is the closure of one of the collector streets that connects the city centre with the west part of the city. The street is to be closed for private motorised traffic, allowing for the introduction of a new cyclist avenue. The goal of this change is to alleviate some of the bike congestions and allow for easy access to the city centre. To simulate this scenario, several changes to the baseline scenario were made, as described below.

First, the road network model needs to be changed to reflect the scenario, i.e., the mobility policy. For this scenario, the road network is changed by hand according to existing initial plans of the municipality. The Oranje Loper road's attributes are changed to disallow private car traffic and the attributes related to the infrastructure are changed accordingly. The resulting road network is shown on Figure 18.

DRAFT VERSION

Figure 18 - Road network with changes. The LTZ includes streets coloured green, red and purple.

The population model is not changed, as the policy is to be implemented soon and the current population will use it. In order to estimate the expected outcomes of the policy change in the future, projections of population data can be used to build the population model.

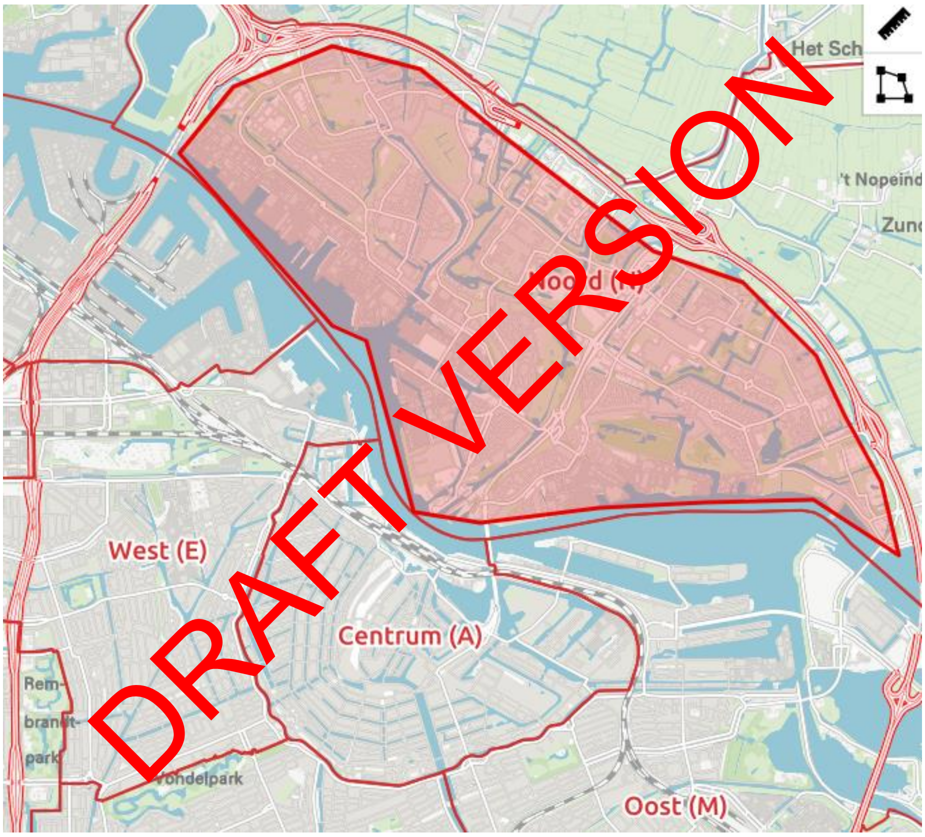
The traffic demand model is rebuilt, however the data about facilities is the same as in the baseline scenario, and only the road network is changed. This however does have an effect on the travel demand, as the locations along the Oranje Loper are no longer accessible by car. Introduction of new bike lanes only affects the mode selection in that the buildings along it are not accessible by car anymore but does not explicitly cause changes of preferred travel mode.

The simulation is run iteratively by optimising the agents' plans and producing the final result. Some of the simulated vehicles can be seen in Figure 18, marked as triangles pointing in the direction of travel.

Finally, the simulation results are used to calculate the values of KPIs, enabling comparison of the Oranje Loper LTZ scenario with the baseline.

6.1.3 Amsterdam Noord expansion scenario

The second scenario, developed for the Amsterdam use case, is the expansion of the Amsterdam Noord district. The city is planning to invest heavily into the area with the goal of alleviating the housing problem. After the intervention, Amsterdam Noord will have a much higher population density and population, with expectations of 150 thousand new dwellings in the city for 250 thousand residents. In the Amsterdam Noord, around 30 thousand dwellings for around 50 thousand residents are planned. The area is shown in Figure 19.



DRAFT VERSION

Figure 19 - Districts of Amsterdam, with the relevant part of Amsterdam Noord coloured red.

This scenario is developed with the goal of analysing the expected behaviour of the city traffic capabilities under enlarged travel demand. The Amsterdam Noord district's urbanistic design may be changed while rebuilding or renovating the residential buildings in yet unexpected ways. Since these plans are not yet developed, we ignore these changes and do not change the road network, instead, we reuse the baseline road network.

The population model however is the most important aspect of this simulation, since this is a large-scale change expected to happen within 20 years. Census data is of course not available for the future, however demographic data projections are available for up to 50 years from now. We use this data from Amsterdam's open data portal and change the number of residents for the Amsterdam Noord district. The population model is rebuilt with expected population numbers and projected demographics. This model is not a very accurate representation of the future population but enables the simulation and analysis of this large-scale change to the city.

The travel demand model is rebuilt using the future population model. While additional facilities are planned to be developed during the time of increasing population, these are currently unknown. While the use case is gathering more information, we are currently not considering simulating unknown changes.

6.2 Use case: Bilbao

Bilbao use case presents a more general approach to urban mobility policy modelling. Bilbao is a city in the Biscay region in the Basque country, with around 350 thousand residents and almost one million residents in the metropolitan area. The use case is interested in increasing the share of trips made using green travel modes, such as public transport, walking and cycling. The use case's KPIs include air pollutant emissions, entry capacity to the central part, and shares of pedestrian and bicycle trips made in a day.

Below are described both scenarios developed for the Bilbao use case.

6.2.1 Baseline scenario

The baseline scenario represents the current state of Bilbao traffic. The scenario includes a road network, population model, travel demand model and calibration methods used to produce the simulation. The simulation includes the detailed road network, public transit model with many lines, schedules for specific dates, multi-modality such as walking (note that pedestrians are simulated inaccurately due to MATSim's limitations, they do not follow the road network, and instead move along a direct line to the destination with a configurable speed), cycling, cars, buses and heavy goods transport with the last three including support for categorisation according to the EURO emission standard. The road network includes sidewalks, bike lanes, paths, and other cyclist and pedestrian surfaces, as well as surfaces used by several different transport modes. Due to lack of data, traffic lights are not included.

The population model is based on the Eurostat public census microdata and enriched with Bilbao's open data portal providing demographic data at district level, such as numbers of residents in different age spans, education level, income bracket and similar. Not all of the population model's attributes are used in travel demand modelling at the time of writing.

The travel demand model is currently facility based. Based on the population model, the initial plans are generated using the harvested facility data from city sources and OSM, and a calibration dataset has been created for Bilbao using data from January 2021, including only workdays.

The baseline scenario, as others, is evaluated according to the KPIs and the decision analysis model, detailed in deliverable D4.3. The KPIs in Bilbao include air pollutant emissions (CO_2, NOx, PM10), entry capacity to the central part of the city, and shares of trips made using different travel modes (bike, car, and public transport).

6.2.2 Moyúa LTZ scenario

The other scenario, developed for Bilbao use case, is the closure of the Moyúa square for traffic. The Moyua square is located in the city centre and connects four streets, as shown in Figure 20.

DRAFT VERSION

Figure 20 - The Moyúa square is shown in the middle of the image. The square is a large roundabout that connects roads into all directions.

DRAFT VERSION

The city is considering closing the square for cars, only allowing cyclists and public transport. Together with the surrounding streets, the square with its central park constitutes the proposed Moyúa LTZ.

The road network has been modified manually to reflect the proposed changes. All of the relevant road segments' attributes have been modified to reflect the change and support the simulation of the closure. Some of the connecting streets have no motorised vehicles crossing, while others are only allowed to public transport and emergency services. The changes are shown in Figure 21.

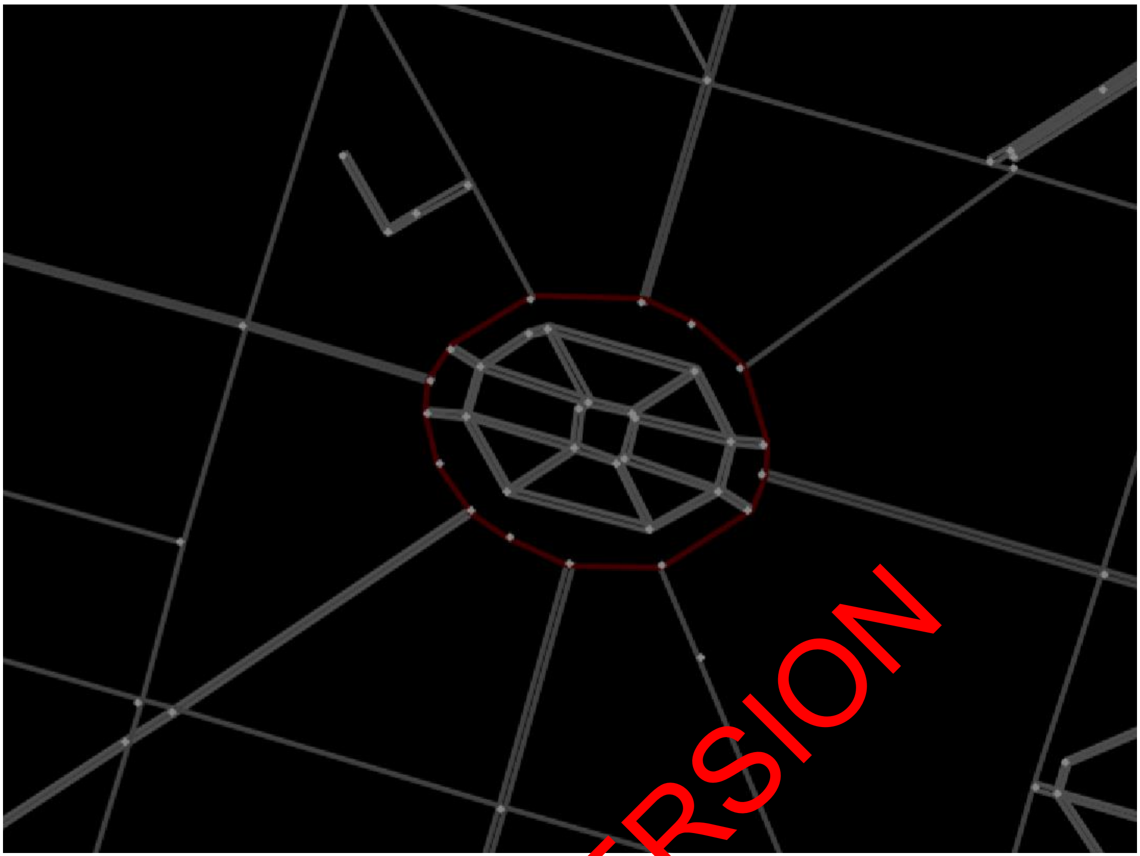


Figure 21 - The Moyuá square in the road network. The red ring is the roundabout and the links inside it are pedestrian paths across the park.

DRAFT VERSION

Since the closure of the Moyúa square is a short- to mid-term plan, we do not need to consider population changes for this scenario. Therefore, the baseline population model is reused to build the travel demand model.

Travel demand model, which is partially based on the road network data, needs to be rebuilt. Besides the network change, there are no other changes to the travel demand. The calibration uses the previously described baseline dataset.

The resulting simulation is evaluated for KPIs and available for URBANITE DSS for comparison and further analysis.

6.3 Use case: Helsinki

The Helsinki use case is focused on the area of the West harbour and Jätkäsaari. The everyday traffic is dominated by the traffic, arriving via ferries to the harbour. There is only a single road connecting Jätkäsaari to the mainland, creating a bottleneck, as shown in Figure 22. There are plans to build a tunnel connecting the harbour directly to the motorway to alleviate this problem. In order to create the simulation for this use case, both the baseline scenario and the scenario with the new tunnel are focused on the Jätkäsaari area and the harbour. This is reflected in the data used, simulation functionality supported, and the traffic model. The baseline and the harbour tunnel scenario models are detailed in the following sections.

Figure 22 - The harbour area with traffic flows shown in red. Thicker lines mean larger volumes of traffic. The location of the road, connecting the harbour to the rest of the city, is marked with a blue circle.

DRAFT VERSION

The problematic part of the road is monitored by a modern camera and radar utilising system known as the Jätkäsaari Smart Junction. It provides high quality data about vehicle types, positions, speeds as well as traffic counts, however the data has not yet been harvested.

6.3.1 Baseline scenario

The baseline scenario for Helsinki represents the current state of the city and is specifically focused on the harbour. Several specific steps are taken to simulate the harbour traffic, using the road network, population model and travel demand model.

The city of Helsinki already has a large and complex traffic flow model, i.e., the HELMET model. We are in contact with the team behind the HELMET model to share some of the data. At the time of writing, unresolved technical obstacles need to be overcome before we can utilise the HELMET model.

The road network of Helsinki is extracted from the harvested OSM data, and includes the urban area and city's islands. The network includes common water travel as well as land travel, including ferries in the city as well as the routes to the harbour. This enables us to correctly locate the traffic arriving from the sea. The road network is preprocessed to enable the simulation of public transport, cycling, cars, and heavy vehicles. The main focus is however on the cars and heavy goods vehicles that move from the harbour to the motorway and have to move through the city.

The population model is built similarly to others, based on the Eurostat data. The demographic data has been gathered and is used to enrich the population model.

The travel demand model of the Helsinki use case is based on facilities. The population model is used to generate the initial travel demand in combination with facility data, extracted from the harvested OSM data. An important additional source of travel demand in this use case is the harbour. Since the population model does not include population from outside of the city, this travel demand must be generated otherwise. We use the harvested data sources about harbour traffic which includes the schedules of ships and the number of vehicles, people and goods arriving on them. Limited OD matrix data is available for these vehicles, such as estimations of shares of vehicles departing towards the motorway. The travel demand model is therefore built by combining the population model data and other data sources of travel demand.

6.3.2 Harbour tunnel scenario

To tackle the problems with the harbour traffic in the city, a proposal has been made to build a tunnel connecting the harbour directly to the motorway. The expectation is that the tunnel would direct most of the heavy vehicle traffic directly to the motorway, thereby alleviating the traffic jams as well as lowering the air and noise pollution in the residential areas.

The road network has been edited to include the proposed tunnel. This includes the entry and exit ramps in the harbour area, and the connection between the tunnel and the existing motorway Länsiväylä. The resulting network is shown in Figure 23.

The population model for this scenario is the same as the baseline, as are other data sources used to generate the travel demand model, such as harbour arrivals.

The travel demand model is rebuilt to include the modifications of the road network. The facility-based model is further enhanced with additional demand, based on the harbour arrivals data. According to the schedules of ferries and data about passengers and goods on board, the number of cars and heavy goods vehicles is estimated. The harbour-based trips are distributed according to available data between the motorway and destinations across the city.

The Jätkäsaari Smart Junction data, currently in the process of harvesting, will be used for simulation calibration, after temporal resampling and aggregation.

DRAFT VERSION

Figure 23 - The modified road network with the tunnel. The main tunnel path is marked in yellow.

6.4 Use case: Messina

Messina is a city between the shore of Sicily and the hills. Due to the geographical layout, the city is very long. The use case is specifically interested in public transport analysis and simulation of new public transport lines. Similar to Helsinki, Messina is also a port city with an important harbour that contributes significantly to the city traffic with the cars and heavy goods vehicles arriving via sea. The focus is reflected in the traffic flow model by inclusion of public transport data, support for and inclusion of an extended public transport in the second scenario, and inclusion of the harbour data in travel demand modelling.

Figure 24 shows the central part of the city, including the port area. From image the layout of the city can be seen, centred between the seashore and the hills north of the city.

DRAFT VERSION

Figure 24 - A detail of the Messina road network, showing the central part of the city. Simulated vehicles are included, where pink triangles represent public transport, green ones cars and blue ones cyclists.

6.4.1 Baseline scenario

The baseline scenario represents the current state of the city traffic. The simulation includes multi-modality (bus, car, bicycle, pedestrian), population model and OD matrix based travel demand model with the addition of harbour traffic. The use case has introduced new traffic cameras across the city, located at main entry points to the city and along the more important streets in the city. The data provided by these cameras will be used to calibrate the simulations and is in the process of harvesting.

The road network is extracted from the harvested OSM data and preprocessed. Data about the public transport is included, providing correct bus lines, bus station locations, and schedules. A detail of the road network is shown in Figure 24.

The population model is based on Eurostat public census microdata and city provided demographic data at district level. The population model will be further refined, in case additional files from Eurostat data are acquired until the end of task T4.2. Current model is census based; however, the household-based approach is also considered for this use case.

The travel demand model is OD matrix based and further enhanced with the inclusion of harbour generated traffic. The harbour traffic is included similarly to the Helsinki use case by enhancing the base traffic flow model. In the case of Messina, origin-destination data is available, and the facility data is not required.

The simulation is run and the resulting simulated vehicle movements are used in visualisations, as basis to calculate KPIs, and for other analysis. The Messina KPIs, detailed in deliverable D4.3, include air pollutant emissions, daily shares of trips made using different modes (public transport, car, bicycle, pedestrian), and others.

6.4.2 Bus line extension scenario

The second scenario that has been developed for this use case is the extension of a bus line. A part of the extended line is shown in Figure 25. The included extension covers a rural area that previously had no public transport connection to the city. Generally, the city is well connected

DRAFT VERSION

by public transport, however as the city is caught between the sea shore and a mountain area, some of the more remote parts lack connectivity and are only accessible by private vehicles. These areas are also generally too remote to access the city by foot and too mountainous for everyone to use bicycles.

The road network is extracted from the harvested OSM data. The preprocessing is similar to other use case networks. The existing public transport is included and support for multi-modality is provided. The public transport extension is created and exported as GTFS using an open source GTFS editing software[1]. The extension includes several bus stops, located along existing roads. We were careful in placing the bus stops at feasible locations, where there is enough space for introduction of a new bus station. The schedules were defined for the new line including different schedules for workdays, weekends, and holidays.

The population model for this scenario is the same as the baseline population model. It is a census-based population model, as described in Section 3.

The travel demand model is rebuilt to allow the use of the newly introduced public transport line extension. Since the population model is the same as baseline, the bus line extension is the only difference with impact on the travel model. Other traffic influences, such as the harbour-based traffic, are accounted for in the same way as the baseline.

Calibration data has been identified and is in the process of harvesting. Once this data is available, several data sets for simulation calibration will be created as detailed in Section 4.

DRAFT VERSION

Figure 25 - The orange line shows an extension of an existing bus line in Messina.

[1] Available online: https://static-gtfs-manager.herokuapp.com/

7 Delivery and usage

7.1 Traffic flow model

This component preprocesses the input data, creates the required models for the simulation, and runs and calibrates the MATSim simulation.

7.1.1 Prerequisites

The MATSim open-source framework for large-scale agent-based transport simulations is required to run the simulations.

MATSim installation instructions: https://www.matsim.org/downloads/

MATSim documentation: https://www.matsim.org/docs/

7.1.2 Licensing information

The licence terms for the software are under discussion among the consortium. AGPLv3[2] are being considered.

7.1.3 Download

Code and configuration files for population modelling, travel demand modelling, simulation calibration, and use case scenario simulation can be found on the webpage:

https://git.code.tecnalia.com/urbanite/private/wp4-algorithms-and-simulation/models_and_working_files

DRAFT VERSION

[2] https://www.gnu.org/licenses/agpl-3.0.en.html

8 Conclusions

The URBANITE traffic flow model is based on microscopic traffic simulations, and is simulated using the open source framework for multi-agent traffic simulation, MATSim. Several models need to be built to feed the simulations. These include the road network, population model, travel demand model, and calibration.

Road networks are part of the harvested data, and available in the URBANITE data platform. These are preprocessed to support different simulation functionalities, required by the four use cases, such as integration of public transport simulation and support for multi modal trips.

Population model is presented in detail in Section 3. Three different types of population models are developed. The first model is a basic randomised population model, which is based only on the city provided demographic data. An improvement of this population model is the census-based population model, which uses microdata census data to improve the representative power of the model. The third model, i.e., the household-based population model, further improves the census based on by modelling the population on household level before creating the specific agents for each person. This model allows for ride sharing among agents from the same household, further improving the quality of the simulations.

Two different developed travel demand models are presented in Section 4. These are the facilities-based travel demand model and the OD matrix-based travel demand model. The facilities-based model utilises a set of different facilities, such as schools, shops etc., to provide locations of different activities. The OD matrix-based model uses OD matrices to obtain the locations of different activities in order to correctly model the traffic between different districts. While this model is more realistic, the data required to build it is not available for all pilot cities.

Calibration method and data set creation are presented in Section 5. The calibration method allows for calibration of the simulations to real data, measured by vehicle counters, traffic cameras or other equipment. The data requires some preprocessing. Different dataset can be created to represent traffic behaviour under different conditions, e.g. on rainy or sunny days.

Finally, the simulation method and the developed traffic simulations for the four use cases are presented in Section 6. For each use case, a baseline scenario is developed, representing the current state of the city traffic. In addition, specific scenarios are developed for each use case.

The simulation results are evaluated by calculating city specific KPIs and used for visualisation. The simulated vehicle movements are also further exploited in the URBANITE DSS, described in deliverable D4.3.

The traffic simulation component will be further improved until the end of task T4.2. Some of the developed models will be improved using better data and the calibration will be performed for all the use cases.

DRAFT VERSION

9 References

[1] W Axhausen, K., Horni, A., & Nagel, K. (2016). *The multi-agent transport simulation MATSim* (p. 618). Ubiquity Press.

[2] Ryan, J., Maoh, H., & Kanaroglou, P. (2009). Population synthesis: Comparing the major techniques using a small, complete population of firms. *Geographical Analysis*, *41*(2), 181-203.

[3] Hermes, K., & Poulsen, M. (2012). A review of current methods to generate synthetic spatial microdata using reweighting and future directions. *Computers, Environment and Urban Systems*, *36*(4), 281-290.

[4] Choupani, A. A., & Mamdoohi, A. R. (2016). Population synthesis using iterative proportional fitting (IPF): A review and future research. *Transportation Research Procedia*, *17*, 223-233.

[5] *European Union Statistics on Income and Living Conditions - Access to microdata - Eurostat*. Ec.europa.eu. (2022). Retrieved 13 April 2022, from https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions.

[6] Lovelace, R., & Ballas, D. (2013). 'Truncate, replicate, sample': A method for creating integer weights for spatial microsimulation. *Computers, Environment and Urban Systems*, *41*, 1-11.

[7] Smith, A., Lovelace, R., & Birkin, M. (2017). Population synthesis with quasirandom integer sampling. *Journal of Artificial Societies and Social Simulation*, *20*(4).

[8] Li, Z. (2014). *Feedback with Alternate Trip Assignment Approaches in the Four-Step Model*. University of California, Irvine.

[9] Popovici, E., Bucci, A., Wiegand, R. P., & De Jong, E. D. (2012). Coevolutionary Principles.

[10] Dargay, J., Hivert, L., & Legros, D. (2008). The dynamics of car availability in EU countries: A Comparison Based on the European Household Panel Survey. *IATSS research*, *32*(2), 44-55.

[11] Nagel, K., Zilske, M., & Flötteröd, G. (2016). *CaDyTS: Calibration of Dynamic Traffic Simulations*.

[12] Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y. P., Hilbrich, R., ... & Wießner, E. (2018, November). Microscopic traffic simulation using sumo. In *2018 21st international conference on intelligent transportation systems (ITSC)* (pp. 2575-2582). IEEE.

[13] Liu, R. (2010). Traffic simulation with DRACULA. In *Fundamentals of Traffic Simulation* (pp. 295-322). Springer, New York, NY.

[14] Holt, C. A., & Roth, A. E. (2004). The Nash equilibrium: A perspective. *Proceedings of the National Academy of Sciences*, *101*(12), 3999-4002.

[15] Flötteröd, G., Bierlaire, M., & Nagel, K. (2011). Bayesian demand calibration for dynamic traffic simulations. *Transportation Science*, *45*(4), 541-561.

[16] Notter, B., Keller, M., Althaus, H. J., Cox, B., Knörr, W., Heidt, C., ... & Jamet, M. (2019). HBEFA 4.1. *INFRAS*.

DRAFT VERSION