



CIM Out-of-Home Study

Methodology

Release 2024-2

Centre d'Information sur les Médias A.S.B.L. – Centrum voor Informatie over de Media V.Z.W.

Boulevard Auguste Reyerslaan 80 – Brussel 1030 Bruxelles, België – Belgique

+32 (0)2 661 31 50 | info@cim.be | www.cim.be

TABLE OF CONTENTS

TABLE OF CONTENTS	2
PREAMBLE.....	4
1 ABOUT CIM	5
2 THE OUT-OF-HOME STUDY	6
3 SUPPLIERS.....	7
OUT-OF-HOME STUDY	8
1 PANEL INVENTORY AND VISIBILITY ADJUSTMENT.....	8
1.1 SUPPLIER	8
1.2 HARMONIZATION OF MAP INFO	8
1.3 IMS.....	9
1.4 MEDIA OWNERS	10
1.4.1 <i>Frame classification</i>	10
1.4.2 <i>Frame authentication</i>	12
1.4.3 <i>Contacts calculation</i>	12
1.4.4 <i>Indoor Metro and train stations</i>	13
1.4.5 <i>Indoor Shopping Malls & Retail Point of Sales</i>	13
2 TRAVEL DATA.....	14
2.1 SUPPLIER	14
2.2 SURVEY DATA.....	14
2.2.1 <i>Tours, trips and steps</i>	15
2.2.2 <i>Survey sources</i>	16
2.2.3 <i>Unified database</i>	19
2.3 VIRTUAL POPULATION DATABASE (VPD)	19
2.4 ACTIVITY BASED MODELLING (ABM)	20
2.4.1 <i>Tour frequency</i>	21
2.4.2 <i>Time of week</i>	22
2.4.3 <i>Destination</i>	23
2.4.4 <i>Mode choice</i>	25
2.4.5 <i>Week diary for all individuals</i>	25
2.5 TRAFFIC ASSIGNMENT	30
3 TRAFFIC DATA.....	31
3.1 DATA SOURCES.....	31
3.1.1 <i>Floating car data</i>	31
3.1.2 <i>Loop detectors and traffic counts</i>	32
3.1.3 <i>Benchmark FOD</i>	33
3.1.4 <i>Total kilometers</i>	33
3.1.5 <i>Speed on road network</i>	34
3.1.6 <i>Volumes on road network</i>	34
4 ASSIGNING TRAFFIC TO ROAD NETWORK	36
4.1 VALIDATION DATA SOURCES.....	36
4.1.1 <i>Home-work travel matrix</i>	36

4.1.2	<i>DIV total mileage</i>	36
4.1.3	<i>Federal Planning Bureau</i>	37
4.2	VALIDATION	37
4.3	CALIBRATION	39
4.3.1	<i>Number of trips per mode of transport</i>	39
4.3.2	<i>Voyager kilometers per mode</i>	39
4.3.3	<i>Car traffic volume</i>	40
4.3.4	<i>Number of passengers per train and metro station</i>	40
5	SOFTWARE DELIVERY SYSTEM	42
6	APPENDIX 1: VARIABLES USED IN THE UNIFIED DATABASE (UD)	43
6.1	PERSONAL CHARACTERISTICS.....	43
6.2	TRAVEL CHARACTERISTICS.....	45
7	APPENDIX 2 : UNIFIED DATABASE CLEANINGS	46
7.1	EDUCATION	46
7.2	PROFESSION	46
7.3	PERSONAL INCOME	46
7.4	FAMILY INCOME	46
7.5	FAMILY SIZE	46
7.6	CAR-OWNERSHIP	46
7.7	DATE OF TRIP	46
7.8	TIME	47
7.9	MODE OF TRANSPORT	47
7.10	MOTIVE.....	47
8	APPENDIX 3: ACTIVITY BASE MODEL (ABM) CHARACTERISTICS	49
8.1	NUMBER OF TRIPS / MOTIVE	49
8.1.1	<i>Work</i>	49
8.1.2	<i>School</i>	49
8.1.3	<i>Shopping</i>	49
8.1.4	<i>Picking up someone / dropping off someone</i>	50
8.1.5	<i>Visiting someone</i>	50
8.1.6	<i>Other motives (social, services ; others)</i>	50
8.2	TRANSPORT MODE	51
8.2.1	<i>Slow/Non-slow transport mode</i>	51
8.2.2	<i>Private/public transport</i>	51

PREAMBLE

The new 2017-2019 CIM Out-of-Home study has the ambition to be a landmark of future media studies.

It is based on a new, 'hybrid' methodology, integrating traditional survey data and other, 'big data' sources, among which census data, official mobility surveys and traffic counts, and independently collected road traffic data.

This study is also innovative in the way it relies on 3 independent 'donor' data sets, which are standardized in a 'unified database', used to generate the 'virtual population database' after extensive use of modelling techniques, and a recursive calibration/validation process.

The virtual population database describes the whole Belgian population aged 12+ and their demographics, on the one hand, and a full week of their travel information (roads used, day, time, mode of transport, motive) on the other hand.

Together with the classification and location of the whole Belgian panels inventory, with a precision never attained before, the demographics and travel data allow to calculate the delivery of networks or ad-hoc selections of panels.

But this study goes one step further than the metrics traditionally associated with media analytics : reach, frequency, and contacts, or GRP.

Indeed, this study aims to measure actual eye contacts, instead of the traditional potential contacts. To that end, these potential contacts have been weighted by several visual adjustment factors resulting from established, international visibility studies. Size of panel, distance of vision, angle of vision, eccentricity vs the road axis, illumination, dynamic/static frame, clutter and traffic speed are all accounted for, and result in new currencies :

- VA Reach (for Visually Adjusted reach)
- VA Frequency
- VRP

Finally, this study has been devised in a modular way, in order to make it as future proof as possible vs evolutions or disruptions caused by the advent of new data sources, new measuring techniques or new ways of marketing the panels inventory.

1 About CIM

CIM has been founded in 1971, as a result of the merging of OFADI (first body in charge of the authentication of the distribution of Belgian print media) and CEBSP (first body in charge of audience measurement). Today, the association has more than 300 members, among which media and their sales houses, and advertisers and their media agencies.

CIM is in charge of collecting audience data from different media types: television, radio, Out-Of-Home, online, print. As far as Out-of-Home is concerned, CIM organizes and controls the collection of Travel data as well as the panel inventory classification. Both parts, together with the modelling of travels, represent the foundation of the OOH survey.

2 The Out-Of-Home Study

The present methodology is specifically dedicated to the CIM Out-of-Home study. This study is managed by the Technical Commission Out-of-Home, whose members are currently :

President : Jos Van Campenhout (Outsight)

Members from the Commission :

Geoffrey Dejean (Belgian Posters)
Valérie De Coster (Clear Channel Belgium)
Lorenzo De Luca (JC Decaux Belgium)
Elodie Roland (Havas Media)
Thomas De Greef (Kinetic)
Hélène Bairiot (Dentsu)
Christophe Guisset (Rapport)

Members from the Structure Permanente :

Koenraad Deridder
Michaël Debels
Joëlle Defossez

3 Suppliers

The present study is resulting from the collaboration of the following companies:

- Be-Mobile, K. Mercierlaan 1a, 9090 Melle, Belgium, managed by :
Jan Cools & Philip Taillieu – Managing Directors
- MGE Data, U Salamounky 769/41, Prague 5 158 00, Czech Republic, managed by :
David Strnad – Managing Director

OUT-OF-HOME STUDY

1 Panel inventory and visibility adjustment

1.1 Supplier

The panel inventory software is managed by MGE Data.

MGE Data is an international out-of-home media expert, operating from the Czech Republic. Their main fields of expertise are OOH audience research, OOH visibility adjustment, mobility surveys, traffic analysis and measurement, geomatics and geo-marketing, delivery and logistics. For the CIM OOH project, MGE Data is responsible for both the inventory management software (IMS) and the delivery software (IDS).

1.2 Harmonization of map info

The main purpose of the study is to match the exact location of panels with correct traffic flows. To be able to correctly combine these two elements, it is important that every party contributing to the study works on the same map.

We decided to use the Open Street Map (OSM), an open data source made available by the OSM foundation. The November 13th 2015 version of the Open Street Map was used from the first publication (2017-1) to the end of 2021 (publication 2021-1).

For the publication 2022-1, an update of the OSM map has been done (using January 1st 2020 version), in order to reflect the latest changes in traffic conditions (changes in urban landscape, new traffic plans, development of public transport networks, etc.). This version is currently still in use.

Be-Mobile has transposed the OSM source file into a routable shape file. This means that a clear topology has been created, i.e. a well-defined network of super-nodes and links, allowing routable paths between an origin and a destination, together with transportation mode restrictions. From this routable shapefile, Be-Mobile created a base-map of super-nodes and unidirectional links between these super-nodes. Subsequently, the unidirectional links have been divided into segments, with a maximum length of 50m. The OSM source file, the routable shapefile and the Be-Mobile base-map were all fixed for use in the new OOH study. These files are then used by both MGE Data and Be-Mobile, to position the inventory and to model the traffic volume on.

1.3 IMS

MGE Data has developed the Inventory Mapping System (IMS). IMS is a web-based application that enables to load in the whole panel inventory from media owners. The frames can be adjusted to their correct position and bearing relative to their surroundings. The surroundings can be displayed by several means like Google maps, Google satellite View, Bing maps... Using these different views together with Google Street View, it is also possible to decide on the visibility of a frame. This visibility is then linked to the different nodes and links of the OSM map that is surrounding the frame.

IMS also contains the traffic volumes created by Be-Mobile (see chapter 3). These volumes are linked to the frames and their visibility to calculate contacts.

Illustration: Visualisation of the cone of vision in IMS



1.4 Media owners

Three different media owners assisted CIM in the development of the new Out-of-Home Study:

1. Belgian Posters
2. Clear Channel Belgium
3. JC Decaux Belgium

1.4.1 Frame classification

All the panels from the media owners were loaded into the inventory mapping system. Each frame was then adjusted to its correct location and bearing.

Because the panels were loaded into the system based on their coordinates, the location wasn't always 100% correct. Using a Google Satellite or Bing background, the panel was moved to its exact location. The background allowed to increase enormously the level of accuracy of the location of the panel compared to the previous study (that used a TOM-TOM map without satellite view). The use of the background also allowed to adjust the angle of the frame to the north and its surroundings more accurately, for example when a panel is mounted on a building.

Once all parameters were correct, the user had to indicate for every road segment present in the cone of vision of that panel whether the frame was visible from that road segment. This was done manually with the help of photos and Google Street View.

The road segments that were approved are then used to calculate the audience. Every time a person passes on one of the approved road segments while facing the direction of the panel, a Realistic Opportunity to See (ROTS) is counted. The sum of all the ROTs's of a panel are the possible contacts. However, these ROTs's were then corrected with the Visibility Adjustment Index (VAI).

1.4.1.1 Visibility Adjustment Index

The Visibility Adjustment Index (VAI) is based on research provided and funded by ROUTE in the UK and technically implemented in IMS by MGE Data. The index is used to determine the part of ROTs that has seen the actual advertising. It is determined by several characteristics, either from the frame or from its environment. Each of these characteristics enhances or reduces the likelihood that a frame is actually viewed by an individual passing by. The VAI is multiplied with the ROTs, resulting in Visibility Adjusted Contacts (VAC).

Many characteristics were used in the calculation of the VAC, some of which are part of the ROUTE algorithm, others were added by MGE Data for the Belgian market at the request of CIM.

1.4.1.2 Size

The size of a frame, expressed in exact centimetres, determines the size of the cone of visibility. A larger frame created a larger cone and thus a larger maximum visibility distance.

1.4.1.3 Angle to road

The angle between a frame and a passer-by determines the likelihood that the passer-by will see the frame. When a person is approaching a frame head on, the VAI will be higher than when a person is

passing a frame that is parallel to him. The exact angle to the road is entered when positioning the frame.

1.4.1.4 Distance to road

The distance of a panel from the road (eccentricity) influences the likelihood that a frame is seen. The distance of the frame to the road is calculated when positioning the frame.

1.4.1.5 Construction height

In Belgium, the construction height of a frame has a limited impact on the VAI. Default heights have been applied, namely a default height of 1 meter for 2m² frames (street furniture), and 4 meters for larger frames. Exact heights have been applied only to frames higher than 8 meters.

1.4.1.6 Illumination

Illumination has been accounted for in several ways.

- First, the number of illumination hours varies during the year. The exact number of sunlit hours was calculated based on the average sunrise and sunset times, in 4 week periods.
- Second, a distinction was made between illuminated and non-illuminated frames during night time. Non-illuminated frames were attributed only 25% of their night-time contacts (this percentage is due to the high level of ambient light during night time in Belgium).
- A further distinction was made between frames that are illuminated 24 hours a day and frames that are illuminated 21 hours a day (no illumination between 2 a.m. and 5 a.m.).
- Other distinctions have been made as well between several types of lighting (standard, Led and Digital) and environments (indoor, outdoor).

1.4.1.7 Dynamic frames

Dynamic frames have been proven to attract more attention than static frames. Hence, an attraction factor is incorporated into the VAI for dynamic frames. An attraction factor of 1.2 is used for frames smaller than 12m² and a factor of 1.5 for larger frames. Because of this attraction factor, it is possible that the VAI is superior to 100%. However, the final VAC will be capped so that it doesn't exceed the ROTS.

1.4.1.8 Digital frames

Digital frames have been integrated in the 3rd publication of the OOH survey (2019-1). As for dynamic frames, they benefit from an attraction factor to reflect the uplift in attention of moving videos images compared to static posters. An average attraction factor of 1.25 is used for all digital formats. The same capping as for dynamic frames will be applied to ensure the final VAC doesn't exceed ROTS.

1.4.1.9 Speed

The likelihood that a frame is seen depends on the speed of the passer-by. The faster a passer-by is going, the shorter the time the frame will be visible to him/her, leading to a lower VAI. The average speed for each road segments was calculated and delivered by Be-Mobile (see chapter **Error! Reference source not found.**, Travel Data).

1.4.1.10 Clutter

Clutter occurs when two frames or more are visible from the same position. Each additional panel will influence the visibility of the other frames, and will result in a lower VAI. At night time, only illuminated frames will be taken into account to calculate the clutter factor. The clutter factor is capped at 85%; this means that other frames cannot take away more than 15% of the visibility of a frame.

1.4.2 Frame authentication

All panels in the database come with a picture of the frame, thus ensuring the actual existence of the frame.

Furthermore, MGE Data was responsible for the frame authentication. A large majority of the frames have been classified by MGE Data specialists, ensuring a high quality of work. A limited proportion of billboards have been classified by the media owners themselves; however, they have been controlled (and corrected if necessary) afterwards by MGE Data specialists. This control confirmed the existence of the panel, its exact size, location, angle to road, obstruction and illumination, and the visibility on the approved road segments.

1.4.3 Contacts calculation

The calculation of contacts for a panel is a two-step process.

- The first one is the calculation of ROTS (Realistic Opportunity to See). ROTS is the sum of all the matches between an individual and any road segment that is in the cone of visibility of a panel, provided the individual is moving towards the panel, and thus can see the panel. We assume that the frame must be positioned at an angle between -30° and 130° to be visible to the individual.
- Starting from these ROTS, we then calculate the VAC (visibility adjusted contacts), multiplying the ROTS obtained on each segment by the Visibility Adjustment Index (VAI) of this segment.

1.4.3.1 Repeated contacts.

In general, it is very frequent that a person had repeated contacts with the same frame from different road segments. Most of the times, these road segments will be adjacent to each other and the individuals passes them one by one. This is for example the case when a person approaches a panel head-on on a straight road and passes multiple road segments while approaching the panel.

However, it is also possible that a person sees the same frame from two segments that are not next to one another, for example when a person makes a loop. In this case, some time passes between the contacts. Because we have no idea how much time passes between these two encounters, it was decided that a repeated contact within the same etape, counts as 1 ROTS. However, when the contacts are separated by a different etape or trip, we count 2 ROTS.

1.4.3.2 Dynamic & digital frames

To calculate the VAC per face of dynamic frames, the share of time of each individual face will be taken into account. In other words, the total VAC obtained above, before any capping is applied, will be divided by the number of scrolling faces.

For digital frames, as the number of screens in rotation can vary by campaign, the total VAC will be reduced by an index taking into consideration the dwell time of the audience (duration of the passage in the cone of vision), the spot duration and the SOV (share of voice) of the campaign. The formula used is : $((\text{Dwell time}/\text{spot duration})+1) \times \text{SOV}\%$.

1.4.4 Indoor Metro and train stations

Metro stations were added in the second edition (2018-1) of this study and train stations in the fifth edition (2020-1).

Since no OSM maps are available for railway or metro stations, media owners (Clear Channel/Publifer and JC Decaux) provided detailed plans for each floor of each individual train/metro station, with the exact position (latitude/longitude) of their access points (entries/exits), their interior characteristics (walls, gates, platforms, stairways/escalators, obstructions) and the positioning of their frames.

The access points were used to connect each station to the Open Street Map, and to 'collect' the traffic volumes generated in the Virtual Population Database (see further).

MGE Data then took care of the digitization of the train/metro maps to be used in IMS, created networks of links to simulate flows of people inside each station and positioned frames on their correct position.

1.4.4.1 Indoor modelling

In order to calculate indoor contacts, MGE had to simulate traffic flows inside the stations.

The modeling of these flows is also based on indoor methodology provided by ROUTE in the UK:

- Flow modelling, calculating all possible routes from entries to exits, entries to platforms and platforms to platforms and assigning to each route a probability based on shortest route length
- Platform modelling, distributing traffic (waiting travellers) on platforms according to platform length and distance to train doors.

For the metro, a transit model has also been used to take into account views of platform frames out of metro carriages (driving and still).

1.4.5 Indoor Shopping Malls & Retail Point of Sales



The audience of digital screens located in shopping Malls and Retail POS is available in the study since September 2024.

This extension required an ad-hoc methodology since specific POS addresses were not available in the original travel database as destination of shopping trips.

First, we used external data from WIFI sensors placed in digital screens at the entries or inside the POS as reference data (provider What The Shop). These sensors are counting passers-by with WIFI enabled in a radius of maximum 25 meters around the frames.

These sensors allowed us to collect essential data for each day of the week:

- Number of visitors of shopping malls & retail POS (net volumes)
- Net volumes per frame (reach)
- Gross volumes per frame (gross contacts)
- Duration distribution of passage in the measured zone (dwell time)
- Hourly distribution of Net volumes per frame.

Using the WIFI data as objectives, we then extended the existing travel database by creating virtual bridges between the end of shopping trips and specific POS. The technical approach was the following:

1. In the travel database, isolate trips with motive shopping around Point of Sales;
2. Within these trips, select the number of individuals in the existing Virtual population (VPD), for each day of the week, to match the Net volumes per POS ;
3. Extend the shopping trip of the selected individuals, by creating a virtual bridge between the end of the trip and POS;
4. Distribute the selected individuals per day, inside the POS, to recreate the net & gross volumes per frame.

This process was handled by the provider Movias SA. The output is a new Shopping Malls & Retail database including the number of ROTS per frame per day & hour.

As a last step, the provider MGE Data calculated the VAC (visibility adjusted contacts) of each frame, applying a standard VAI based on the frames characteristics (size, illumination, angle, motion & dwell time) & their position (at entrances, in parking, in hall/main dwell area). To do so, the media owner (JC Decaux) had to provide plans of the POS with position of frames and pictures.

2 Travel data

2.1 Supplier

All travel data processing was carried out by Be-Mobile. Be-Mobile is a Belgian traffic information service provider that entered business in 2006. The activities of Be-Mobile are focused on traffic data, aggregation of traffic data into traffic information and integration of traffic information services.

2.2 Survey data

The creation of the travel data used for this study was a complicated job.

- First of all, we collected several data sources of travel data, both proprietary and external.
- We then combined these data sources into one homogenous data source.

From this data source, we created two entirely new data sets, calibrated on other external sources of socio-demographic data.

- The first data set includes the socio-demographics of the Belgian population.

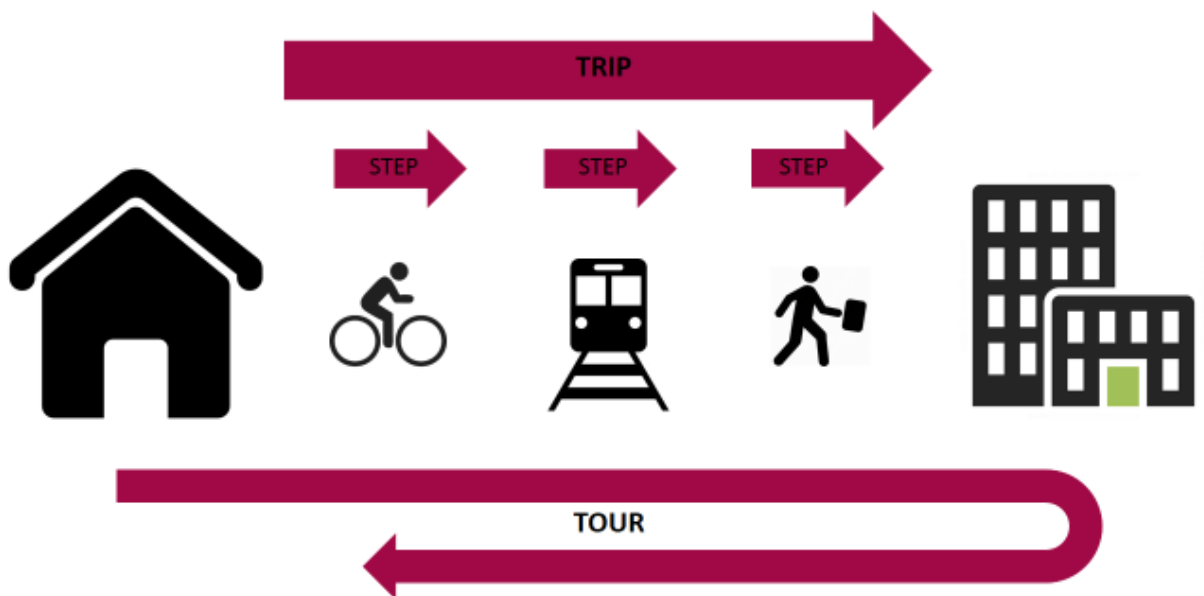
- The second set aggregates the origins and destinations of all the trips taken by the individuals during 1 week.

Finally, we assigned all the trips to the Belgian road network. All this work resulted in traffic data for the full Belgian road network for 1 week, with corresponding socio-demographic profiles.

2.2.1 Tours, trips and steps

In this document, we will refer to travel data in 3 different ways: *tours*, *trips* and *steps*. The distinction between these 3 terms is important.

- A **tour** is a combination of several *trips*. A *tour* always starts and ends at home. For example, someone leaving home to go to work and coming back home afterwards, takes what we call a *tour*. In this case, the *tour* consists of 2 separate *trips* (home-work and work-home) and has, thus, two different motives ('work' and 'returning home').
- A **trip** is a travel from one place to another with one motive; for example, going from home to work. A *trip* can include different transportation modes. For example, someone leaving home, driving his bike to the train station, taking the train to the next city, and finally walking from the train station to the office, has used 3 different transport modes ('bike', 'train', 'pedestrian'), but all for the same motive ('work').
- Each part of a *trip* with a different transport mode is called a **step**. In the example above, the trip would include a 'bike' *step*, a 'train' *step* and a 'pedestrian' *step*.



It is worth mentioning that the distinction between these 3 travel types isn't always made in the same way across the different data sources that have been used: although all surveys refer to trips, some

really meant a step while others rather meant a tour. Hence, the three data sources were cleaned and made coherent to the above-mentioned terminology before the start of the actual data processing.

2.2.2 Survey sources

In the first editions of the OOH survey, three different survey data sources were used and combined into a basis for the modelling, i.e. the Unified Database. For the 2020 modelling, another travel survey was available and added to the combined databases. These data sources were not identical and originated from different sources. We will describe each one of them separately.

2.2.2.1 CIM OOH Data

The CIM OOH dataset was collected in the context of the previous CIM OOH study (2012-2015). In this survey, 12.000 people were interviewed about their socio-demographic profile and travels. The interviews were spread out over a 4-year period. Unlike in the other 3 databases, people interviewed in the CIM OOH study reported all their travels during 1 week (7 days). The travel data collected amounted to a total of 200.000 trips (260.000 steps).

The respondents were recruited from a almost perfectly representative sample of the Belgian population 12+, with an oversampling in the 48 main cities of Belgium. These 48 cities were the research universe of the previous CIM OOH study, and only out-of-home networks present within these 48 cities were included in the study. The following socio-demographic variables were included in the questionnaire: gender, age, education, profession, home address and household size. All necessary information about trips were reported as well: date (day of the week), time of departure, time of arrival, origin, destination, distance, motive, and transport mode. A few variables were not present in the CIM OOH dataset -personal income, family income and car ownership- but were needed for further modelling.

2.2.2.2 BELDAM

The BELDAM study (BELgian DAily Mobility) was carried out by the Belgian federal government. All respondents were asked to write down all their travels of the previous day and to fill out their socio-demographic profile.

The BELDAM dataset contains a representative sample of interviews for the Belgian population. Moreover, certain regions were oversampled due to financing agreements. The regions where proportionally more interviews were collected, were the Brussels-Capital region, Liège, Charleroi and the northern part of the Walloon Brabant province. The data source contains socio-demographic information as well as travel data. The data were collected between December 2009 and December 2012. Some 8.500 households were surveyed, amounting to 15.821 individuals aged 6 years and over. Together, they reported a total of 38.000 trips (54.000 steps). For the needs of the CIM OOH study, only the individuals aged 12 years and older were included in the modelling.

The following socio-demographic variables were included in the questionnaire: gender, age, education, profession, and home address. The following variables were gathered at the household level: household size, family income and car ownership. The only socio-demographic variable that was missing in this dataset was personal income. All necessary travel characteristics were included as well:

date (day of the week), time of departure, time of arrival, origin, destination, travelling time, distance, motive, and transport mode.

2.2.2.3 OVG (Onderzoek VerplaatsingsGedrag Vlaanderen)

The OVG study was carried out by the Flemish government in Belgium. All respondents were asked to write down all their travels of the previous day and to fill out their socio-demographic profile.

The OVG dataset contains only data from respondents from the Flemish part of Belgium. The surveys were carried out in the provinces Antwerp, East Flanders, West Flanders, Limburg and Flemish Brabant, with a slight oversampling in the first two provinces. The data source contains socio-demographic as well as travel data. The data were collected in several waves between 2007 and 2013. A total of 17.000 individuals were interviewed in two waves, amounting to 51.000 trips (56.500 steps).

All basic socio-demographic information on personal level were included in the questionnaire: gender, age, education, profession, personal income, and home address. Other socio-demographics were collected at household level: household size, family income and car ownership. All major travel characteristics were included as well: date (day of the week), time of departure, time of arrival, origin, destination, travelling time, distance, motive, and transport mode. This survey included all the questions needed for the modelling phase.

2.2.2.4 MONITOR

The MONITOR study is a national survey on Belgians mobility, carried out by SPF Mobility and Vias institute. All respondents were asked to write down all their travels of the previous day and to fill out their socio-demographic profile.

The survey was carried out online, between March 2016 and June 2017, among a sample of 10.632 individuals of 6 years and more. The dataset was weighted to be representative of the Belgian population (on factors age, gender, region of residence and level of education). For the need of the CIM OOH study, only the individuals aged 12 years and older were included in the combined dataset.

All basic socio-demographic variables were collected: gender, age, education, profession, personal income, and home address. Other variables were available at household level: household size, family income and car ownership.

The following travel characteristics were also collected: date, origin, destination, motive, time of departure and arrival, destination, duration and distance, transport mode.

2.2.2.5 Cleaning of data sources

Although the most important variables were included in all data sources, some cleaning still needed to be done.

- First of all, missing variables -that were needed for further modelling- had to be created.
- After that, the response categories for each variable had to be standardized across datasets.
- Finally, as previously mentioned, all travels had to be reported in accordance with the tour-trip-step taxonomy. In many cases, this means that trips had to be split into multiple steps. Afterwards, trips were combined into tours.

In order to create a unified database, nine individual characteristics were included in the cleaning procedure. The same variables were also used later on in the activity based model.

Below, the list of variables used, with the number of classes between brackets (full details in Appendix 6 (Variables used in Unified Database)):

1. Gender (2)
2. Age (13)
3. Education (4)
4. Professional activity (9)
5. Personal income (4)
6. Family income (4)
7. Family size (5)
8. Car ownership (4)
9. Post code home address

Furthermore, 6 travel characteristics were integrated in the unified database:

1. Date of trip (2)
2. Departure time (2)
3. Arrival time (2)
4. Travel time in minutes
5. Travel time in kilometers
6. Mode of transport (9): Car/Motorcycle driver, Car passenger, Pedestrian, Bike/Scooter, Bus, Tram, Metro, Train, Other
7. Motive (10): Work, Business, School, Shopping, Visiting somebody, Social activity, Pick-up/drop-off, Go home, Services, Other.

Unfortunately, the same variables were not always available in each dataset.

In the CIM OOH dataset, no information was available on personal income, family income and car-ownership.

In the BELDAM dataset, personal income was the only missing variable.

For each individual of the CIM OOH and BELDAM surveys, the missing variables were completed by means of a prediction model created on the basis of the OVG dataset. Basically, the model predicted the missing variables based on other socio-demographic variables in 75% of the OVG dataset. The prediction model was then validated on the remaining 25%.

Decision trees were used to calibrate the classification/prediction model. Multiple decision trees were created, each one applied on a different set of OVG data. In the end, all the decision trees were aggregated into one decision tree model, also called the classification model. Confusion matrices were used to validate each prediction matrix. The use of confusion matrices allows to know how many responses were correctly predicted, or if incorrectly predicted, by how much the prediction deviated from the correct response category.

The dataset that was used to build the classification model had to be cleaned and outliers had to be removed. The reason for this is that a statistical prediction model is meant to make reliable predictions for large datasets based on general patterns, and not to predict outliers.

The following outliers were removed from the dataset:

- Respondents that are students and belonging to the income category 3 or higher
- Respondents aged below 21 and belonging to the income category 4
- Respondents belonging to the non-active category and the income category 4
- Respondents with no higher secondary education and belonging to the income category 4
- Respondents with a personal income of category 4 (> 2.000 €) and a family income belonging to category 1 or 2 (0 € – 2.000 €).

The classification model identifies patterns and correlation between independent (or explanatory) variables and dependent variables (which need be predicted). The classification model was then applied to the CIM OOH and BELDAM datasets to generate the missing variables (personal income, family income and car-ownership).

Finally, response categories on all variables were standardized in the three datasets. To estimate missing response categories (i.e. splitting reduced response categories), detailed classification models were used. The exact details of these cleanings can be found in Appendix 7 (Unified Database cleanings).

2.2.3 Unified database

The result is a unified database consisting in two separate datasets: the unified socio-demographic profiles database, and the unified travel database.

The unified socio-demographic database comprises 42.537 individuals aged 12+ and their socio-demographic characteristics.

The unified travel database includes all the tours, trips and steps made by these individuals. Together, they account for a total of 137.149 tours, 332.518 trips and 388.918 steps. On average, each individual of the unified travel database takes 1,34 tours per day; an average tour consists of 2,42 trips and an average trip consists of 1,17 steps.

2.3 Virtual Population Database (VPD)

As a next step, a Virtual population of Belgians 12+, with a complete socio-demographic profile, was created.

To do so, the 9.885.365 individuals 12+ were extracted from the CIM Golden Standard VP 2020. The CIM VP is based on a combination of official census data and the data of all CIM fieldwork interviews and includes nearly all necessary variables (gender, age, education, geography, professional activity, social groups, ...).

A complete methodology of the creation of CIM VP can be found here: <https://www.cim.be/fr/golden-standard>.

Some socio-demographic characteristics, which were missing in the CIM VP, were then added by Be-Mobile to the individuals of the VPD:

- Family income
- Car-ownership
- Statistical sector.

These additional socio-demographic characteristics are coming from the UD individuals and have been duplicated a certain number of times by allocating them to individuals of the VP. These characteristics have been allocated only to persons sharing the same source-characteristics as the “donor” in the UD, making sure that no ‘unrealistic’ combination (e.g. student + highest personal income scale) of characteristics could occur.

2.4 Activity Based Modelling (ABM)

In the next phase, the travel part of the Virtual Population Database was created, based on the Unified Database (UD): for each individual, a mobility pattern was calculated for a typical week. The output is a list of trips for 7 days, with for each trip a point of origin and destination, time and day of the week, motive and transportation mode.

The creation of trips was done in 4 separate modelling steps.

- In the first step, a number of trips per week is created, for each motive separately.
- Secondly, the time of the trip is determined, per trip and motive.
- Thirdly, a destination is modelled for the trip.
- And finally, the main transportation mode is assigned to each trip.

This modelling is based on activities and required the development of a series of discrete choice models (or logit models). Every step of the modeling can be considered as a discrete choice between different options. E.g. : the choice between a certain number of trips, the choice between different transport modes.

All the different discrete choices can be organized hierarchically into different ‘nests’. Any choice, e.g. transportation mode, is then the result of a series of separate choices. E.g. : transportation modes can be subdivided into slow modes and fast modes. Slow modes could be further divided into bike and pedestrian modes. Fast modes, on the other hand, can be subdivided into private modes and public transportation modes, and so on. The final choice of transportation mode is made by working down the hierarchical ‘nested’ choices.

Each alternative, present in each choice, has a certain utility for an individual: utilities are dependent on the characteristics of the person taking the trip, or on other specific characteristics. For each individual, the chance that a certain alternative will be selected is dependent on these characteristics, and their utility.

This type of activity based modelling involves three stages:

- First, one has to take a decision about the hierarchical structure that will be used. This means that all the different alternatives have to be organized into a meaningful structure.
- Second, a utility coefficient has to be assigned to each alternative. In order to do this, one has to determine which characteristics of the individual, or which other characteristics, are relevant to the prediction of the alternative. Characteristics have then to be put in the functional form of a utility function. The coefficients of the utility function are then calculated by the method of maximum likelihood.
- Third, a probability is calculated for each person choosing a certain alternative. The actual assignment will then be done by a lottery system.

The activity based model has been used to determine the tour frequency and the choice of transportation mode. Time of week and choice of destination have been carried out differently, as we will later explain in the relevant paragraphs.

Each step of the model has been developed and tested on the UD, and then applied on the total VPD.

2.4.1 Tour frequency

The first step in the creation of the travel database is deciding how many tours each person will make during a week. The number of tours will be estimated for each motive separately: work/business, education, shopping, social activities, visiting somebody, picking up someone and services/other.

The relevant characteristics have been decided for each motive separately. Two characteristics have been assigned to the 'work' utility function: gender (2 classes) and occupation (5 classes). No other tested characteristic (age, education, income, family size, car-ownership, population density) had any marginal predictive value in forecasting the number of work trips. Therefore, the two relevant characteristics were put into the functional form of a Utility Function for number of work trips:

$$Utility(x \text{ trips}) = \alpha + \beta * Gender1 + \gamma * Occup1 + \delta * Occup2 + \varepsilon * Occup3 + \zeta * Occup4 + \eta * Occup5$$

Each coefficient of the utility function was calculated using the maximum likelihood method. Once the utility function was completed, a probability for each step of the choice model could be calculated for each individual. Finally, individuals were assigned a number of trips at random, while respecting the overall probabilities.

The characteristics used to calculate the number of work trips and the number of trips for other motives, are described in 8 (Appendix 3. Activity based model [ABM] characteristics).

The distribution tables of the number of trips per motive in the UD and in the VPD show very similar patterns. The activity based model predicted trips patterns that were very similar to those found in the UD, even for specific target groups.

For all motives, the predicted number of trips is close to the number observed in the UD:

	Observations UD	Prediction VPD
Work trips	2.48	2.55
School trips	0.51	0.55
Shopping trips	2.43	2.45
Visiting trips	1.55	1.55

Picking up/dropping off trips	1.04	1.19
Other trips	2.74	2.74

Looking at the results per day, we see an average of 2.63 trips per individual per day in the UD, versus a predicted value of 2.7 trips per individual per day in the VPD.

In the VPD results, we see less outliers with no or very low/high number of trips than in the UD. The reason is that the model is very good at predicting averages, but is less suitable to predict outliers. Add to that the fact that the prediction has been carried out independently for each motive; a person that already has very few trips for one motive, isn't necessarily attributed as few trips for other motives as well, in spite of the correlation that might be found in the UD. This results in a lower proportion of people with no trips across all motives. On average, however, the number of trips across people is correct.

2.4.2 Time of week

Once trips were attributed to every individual of the VPD, days and times had to be assigned to these trips, before trips could be combined into tours. A logit model, like the one used for the number of trips, is not suitable here. The reason is that the time choice for one trip is highly dependent on the time choice for other trips. Therefore, a process of pattern duplication has been used here. This process has been applied in 2 separate steps. The first step distributes the trips over the 7 days of the week. The second step adds the specific times of the day and tours are then composed. It is during this second step that the 'return home' part of the trips has been added as well.

As already said, all trips have been assigned to a specific day in a first instance. To do this, we looked for 'patterns' of trips in the UD, and applied the same patterns on the VPD trips. This means that days were not assigned to each trip individually, but rather per group of trips. Groups of trips were based on trips with identical motive. So, when an individual in the VPD accumulated 4 work trips during a week, we looked for individuals in the UD that had also 4 work trips, and the days these trips were taken on. The pattern found in the UD was then applied to the VPD individual. This model was executed stepwise for 6 different motives: work/business, school, shopping, visiting someone, picking up/dropping off someone and other motives.

Furthermore, only patterns from people in the same class of occupation were considered. People with different occupational statuses will probably have different patterns. E.g. a retired person might have more shopping trips during the week, while a working individual will preferably plan such trips during the weekend. Three occupational classes were used for this exercise: active people, students and non-active people.

This part of the modelling process was based on the CIM OOH part of the UD only. The reason for this is simple: the CIM OOH survey is the only source of weekly information (OVG and BELDAM respondents were surveyed on 1 day only).

In the case of individuals with 10 or more trips with the same motive during a week, the trips were labeled as 10+ trips and a 10+ pattern was searched for in the UD. This process was applied in order to prevent sparse data. As a result, we had only 0.3% sparse data in the full VPD. In order to reduce sparse data (individuals for whom no exact match was found), we looked for individuals that were a close

match within the same occupational class and with the same motive. E.g. When we couldn't find a student with 5 work trips over a week, we searched a student with 4 such trips in the UD. The times of week of the four trips were transferred. For the fifth trip, we looked at the general distribution of work trips for students across the week. Based on that distribution, we added a day of the week for the fifth trip.

After all trips were assigned a day, they had to be assigned a specific time of day. Four different times of day were used: morning rush (6 am - 10 am), mid-day (10am – 16 pm), evening rush (16 pm – 19 pm) and night-time (19pm – 6am).

To assign a time of day, all trips made on the same day by the same individual were grouped. For each group of trips (for example 1 work trip + 1 shopping trip on a Tuesday), we again looked for a similar group of trips in the UD. Similar to what we did for time of week, we looked for individuals within the same occupational class only. Whenever a match was found between the UD and the VPD, the UD dayparts were assigned to the VPD trips. Finally, the different trips of a day were transformed into tours with the addition of 'going home' trips. Here again, we found fairly little sparse data (0.4%) that were fixed in the same way as days of week.

The times of day were checked by comparing the percentage of going home trips in the UD and in the VPD, and the distribution of trips per tour. Here also the number of trips per tour are very similar between UD and VPD. The logical sequence of trips and tours had to be corrected as well: the day couldn't start with a 'going home' trip, the day had to end with a 'going home trip', every tour had to end with a 'going home' trip, and no consecutive 'going home' trips were allowed.

2.4.3 Destination

The modelling of a destination for each trip was done in three separate steps: modelling work trips, modelling school trips, modelling other trips.

2.4.3.1 Work trips

Modelling a destination for work trips is a fairly easy process. It was based on a home-work matrix on statistical sector level, which was provided by the government. Each individual in an origin zone was assigned a possible destination zone, within the mean trip radius found in the UD. The mean distance was calculated for individuals belonging to the same occupational and education class. Once every person was assigned a destination zone, this zone was attributed to all the work trips this person made.

2.4.3.2 School trips

The assignment of school locations was a bit more complex and required the use of multiple data sources.

- A data source with all Flemish speaking primary + secondary schools, together with their address and the number of students going to each school
- A data source with all French speaking primary + secondary schools, together with their address
- A list of higher education institutes, with their address on municipality level and their number of students

- A list with the number of student dorms per municipality.

Since the data source for the Flemish speaking education is very detailed, assigning a school location to the Flemish speaking students in primary or secondary schools was easy. In order to assign a school to each student, the distance distributions in the UD were replicated in the VPD.

For the French speaking part of the country, there are no public data on the number of students per school. Therefore, this was estimated, based on the population density of the municipality of the school. Once this process was completed, a school was assigned to all individuals as done for the Flemish speaking students.

For higher education, a similar method was used, with a distinction between students living on a dorm and commuting students without a dorm. For students living on a dorm, the assumption was made that they all leave the house on Sunday to go to their dorm, and return home on Friday. During the week, the dorm address was used as a home address.

To assign higher education institutes to the students, a school was assigned to commuting students, while maintaining a 58/42 ratio between commuting and non-commuting students. Again, the distance pattern found in the UD was replicated in the VPD. The school was linearly assigned to all students living on a dorm. The average length of the home-dorm trips (57.5 km) is much higher than the distance than students living home travel to go to school. This method has the logical consequence that people living closer to cities -and therefore higher institutions- have a higher change of living at home than students living further away from any higher institution.

Since only the municipality of each higher institution was available, we distributed the destinations of the student dorms over the statistical sectors of that municipality.

2.4.3.3 Other trips – Telco Matrix

In the first edition of this survey (publications 2017-1 to 2019-2), no official reference data was available for other trips. While the destinations of ‘other’ trips can be very diverse, the analysis of the UD data did show a strong correlation between these destinations and the density of the population of statistical sectors. This allowed, through the use of an attraction model, to estimate the number of destination possibilities per statistical sector and assign trips to these destinations, while replicating the distance distribution of the ‘other’ trips in the UD.

To improve this important part of the modeling (other trips represent +/- 70% of total trips), external telco data from Proximus was used as reference data for the second edition (publications 2020-1 to 2021-1) and updated in the third modelling (publication 2022-1).

The data’s delivered by Proximus, i.e. 1 billion transactions/day (based on mobile location) over a 3 months observation period, are :

- Weighted and extrapolated to the Belgian population
- And are aggregated into an origin destination matrix at
 - NIS6 level (postcode),
 - per day (7) x daypart (4),
 - for an average week

The data's are finally converted by Be-Mobile into an **OD matrix** at statistical sector level, which was used as input for the new AB model.

Compared to the first modelling, the following editions were improved by:

- taking into account the observed repetition from the UD in the destination choice model of other trips
- refining the destination to the exact address (instead of a NIS9 destination).

2.4.4 Mode choice

The attribution of a transportation mode to each trip was again based on a logit model. First, a hierarchical model had to be set up. A distinction was made between slow modes and non-slow modes. Slow modes were split between bicycle and pedestrian. Non-slow modes were further differentiated between car and public transport (meaning either bus/tram, metro or train). Car trips could be made either as a car-driver or as a car-passenger.

Then a utility function was calculated using the variables described in Appendix 8.2.

The execution of the model led to the attribution of a mode to each trip.

2.4.5 Week diary for all individuals

The above-mentioned modelling processes led to the Virtual Population Database, which really consists of two separate databases: the socio demographic database and the travels database. The composition of each database is further described here.

The Golden Standard is used to create the socio-demographic of the VPD database.

Gender	%	Nielsen	%
Men	48.9	Nielsen I+II	56.0
Women	51.1	Nielsen III	12.2
		Nielsen IV+V	31.8
Age	%	CIM Habitat	%
12-17 year	7.7	Antwerp	6.5
18-34 year	24.3	Brussels	13.7
35-54 year	30.9	Charleroi	2.6
55 year +	37.1	Gent	2.6
Education	%	Liege	4.2
None-Elementary	11.1	30 Cities NL	14.4
Lower secondary	16.2	13 Cities FR	7.3
Upper secondary	41.7	Small Cities NL	16.1
Bachelor-Master	31.0		

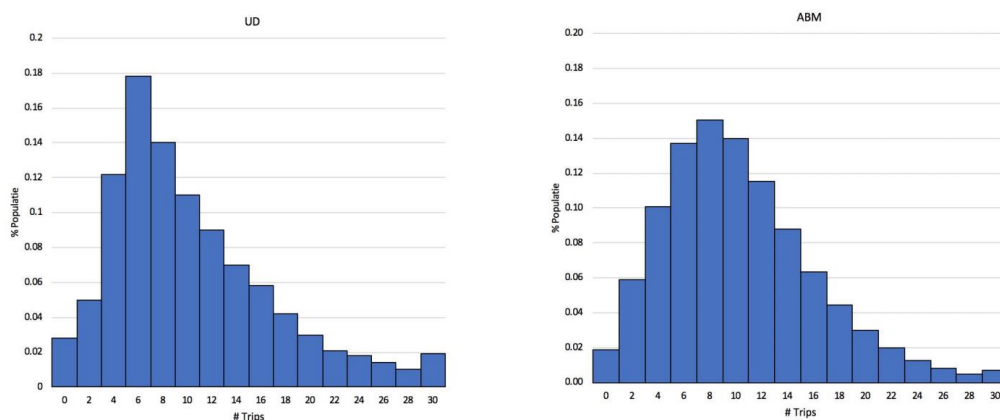
Professional Activity	%
Yes	48.8
No	51.2
Profession	%
Self employed	4.3
Manager	5.2
Employee	27.4
Worker	11.9
Retired	24.4
Student	14.4
Other non-active	12.4

Small Cities FR	7.4
Rural NL	15.6
Rural FR	9.8
Province	%
Brabant FR	3.5
Brussels 19	10.2
Antwerp	16.2
Brabant NL	10.0
West Flanders	10.7
East Flanders	13.3
Hainaut	11.7
Liege	9.7
Limburg	7.7
Luxemburg	2.5
Namur	4.3

The 9.885.365 individuals in the VPD had an average of 19.1 trips during a week, totaling 188.504.213 trips in a week's time.

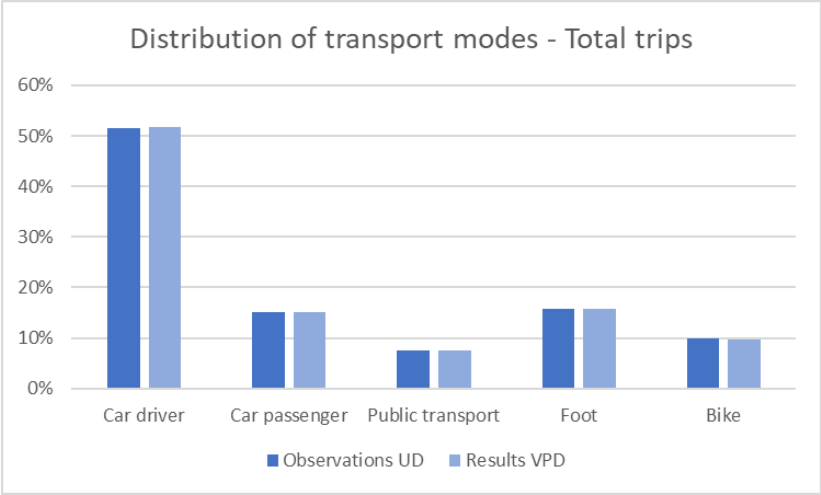
An average individual takes 2,7 trips on an average day. The average trip is 14,6 kilometres long. However, this distance was strongly correlated to the motive and the transportation mode of the trip.

Illustration: Number of trips in UD vs VPD

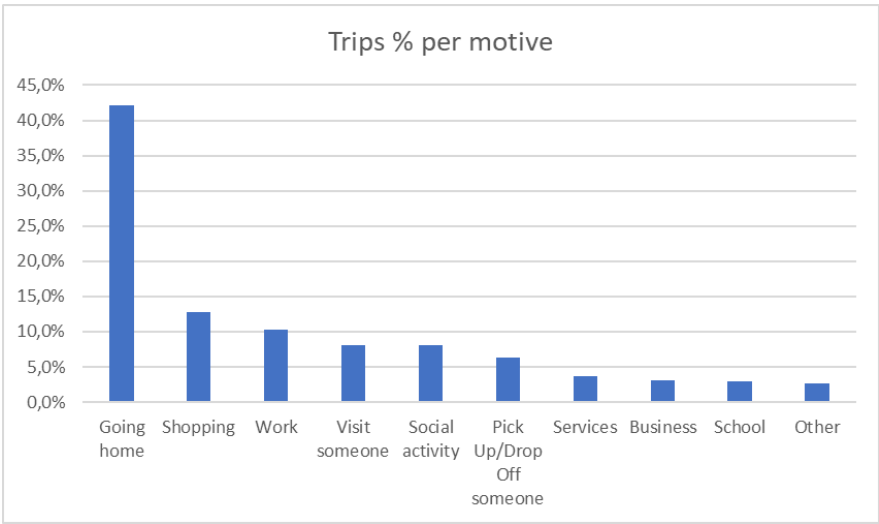


The charts below show the distribution of travels over the different transportation modes. 66.9 percent of all travels were done by car.

Illustration: Distribution of modes in UD vs VPD



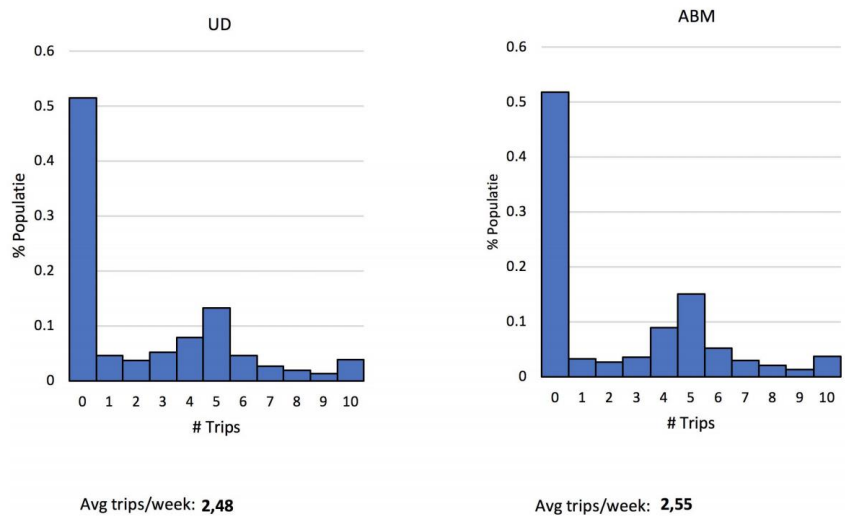
A motive was also assigned to each travel. The following graph illustrates the breakdown of trips per motive.



The following charts show the distribution of travels for each motive.

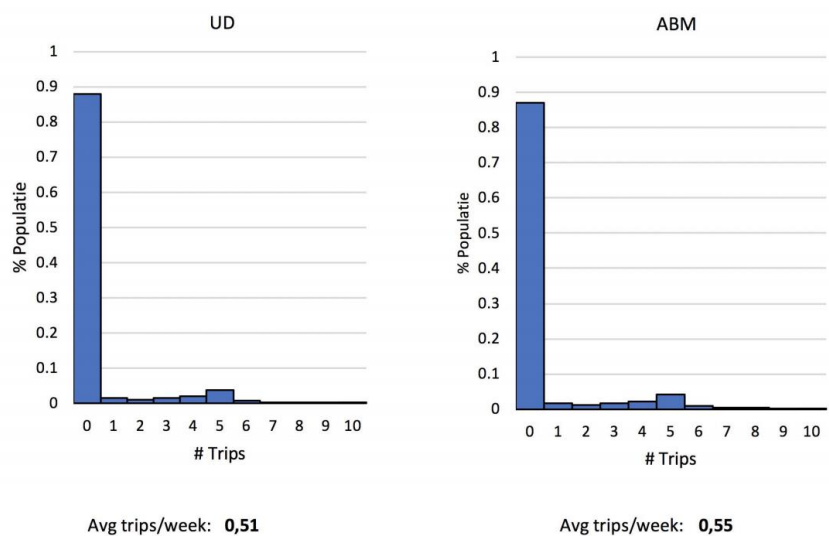
Looking at work trips into more detail, it is clear that a majority of people do not make any work trips. Among working people, most have 5 working trips a week.

Illustration: Frequency distribution of **work trips** in UD vs VPD



School trips show a similar trend; most people do not make school trips

Illustration: Frequency distribution of **school trips** in UD vs VPD



For all other motives, the distribution of number of trips is more gradual, due to the higher number of trips per motive.

Illustration: Frequency distribution of **shopping trips** in UD vs VPD

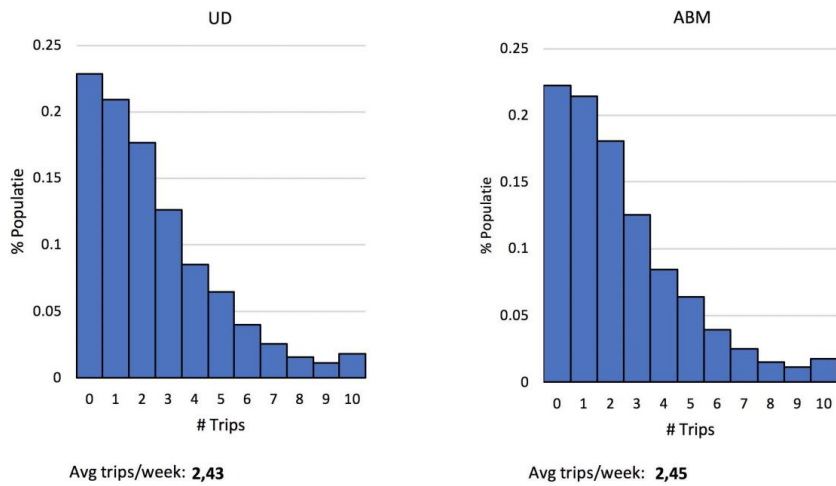


Illustration: Frequency distribution of **visiting trips** in UD vs VPD

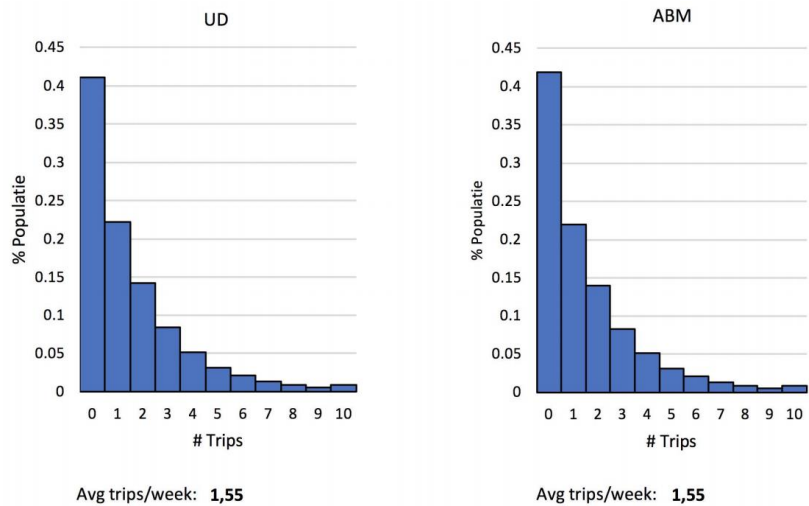
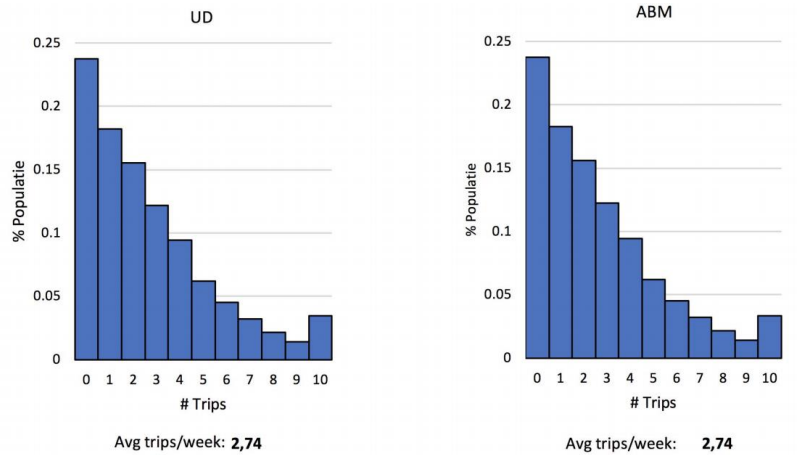


Illustration: Frequency distribution of **other trips** in UD vs VPD



2.5 Traffic assignment

The final step of the creation of travel data for the whole Belgian road network consists of the allocation of the VPD travel data to the OSM road network we discussed earlier.

In order to do this, a route-planner was developed to process the 188 million trips efficiently. On the basis of the origin and destination of each trip, together with the transportation mode chosen, the optimal route was calculated. Because the origin and destination of each trip were defined at the statistical sector level only, an exact address had to be assigned. To that end, origin and destination points were attributed at random within their statistical sector. This method will create some errors. However, there is no available source that could lead to the generation of more precise addresses.

Route calculation is a process which is dependent on various factors. The most influencing factors are the time of day and traffic. The route-planner Be-Mobile used had access to information on the average road traffic at any given time. The route-planner used this information not only to calculate the shortest route, but also the fastest route at a given time. Obviously, these two options will not generate the same result, particularly during rush hours. In other words, depending on the time of the trip, the route-planner would assign a different optimal route. All trips were not sent via the same fastest route (for example over the highway), because this method would create a big bias. In daily life, everyone will not take the same route. Some individuals will take the faster highway, while others will favor the shorter scenic route. Therefore, every viable route possibility receives a probability. The final route choice is assigned on the basis of the probabilities, leading to a realistic distribution of traffic over the full road network.

Evidently, public transport (PT) was assigned in a slightly different way. While bus and tram traffic were assigned to the same road network as car and pedestrian traffic, metro and train were not. Since there is only out-of-home advertising inside metro and train stations, not along train or metro tracks, CIM decided to model the PT traffic up to the stations only.

In the case of metro, the etape from home to the metro station will be assigned to the road network, up to the road segment with a metro access point.

The route followed inside the metro station is modelled as well all the way to the platform.

Inside the metro coach, the individual will pass along a platform in several stations, up to the final destination station, where a link with the road network will be made again.

An even simpler methodology was applied to trains, since travels inside the train station have not been modelled, and neither have been the intermediary train stations.

The result of all the modelling process is the double dataset called the Virtual Population Database. One part of this dataset comprises all the socio-demographic characteristics of the total Belgian 12+ population. The second part of the dataset comprises all the trips taken by the individuals during an average week. Trips are described down to the level of individual road segments.

3 Traffic data

3.1 Data sources

In the previous chapter, we described the models creating traffic and assigning it to the road network. This extensive modeling effort should be validated against independent benchmarks. This chapter describes how data were validated and calibrated when necessary.

3.1.1 Floating car data

Floating Car Data consist of a sample of GPS data collected through GPS systems built into vehicles. Be-Mobile collects GPS positions of a large fleet of vehicles managed by 30 different providers in Belgium. The vehicle position is collected in real-time by the providers, and sent to Be-Mobile in a pre-defined (anonymized) format. Be-Mobile processes the data on the fly in order to minimize delays. Incoming data are validated in several steps prior aggregation down to the minute level.

The sample used for the present study amounts to some 450.000 vehicles. Together, they report about 310 million positions a month, on average. This sample amounts on average to about 3-4% of the total vehicle volume in Belgium.

The sample includes professional as well as passenger vehicles. While most professional vehicles will be ignored when calculating a volume benchmark, they are nevertheless valuable to calculate average speeds since they cover large amounts of kilometers. Passenger vehicles on the other hand, are very valuable to calculate volumes on the secondary road network, where they are mostly present.

3.1.2 Loop detectors and traffic counts

Several traffic counts were collected from various sources. Traffic counts were used to benchmark the model and upscale the floating car data.

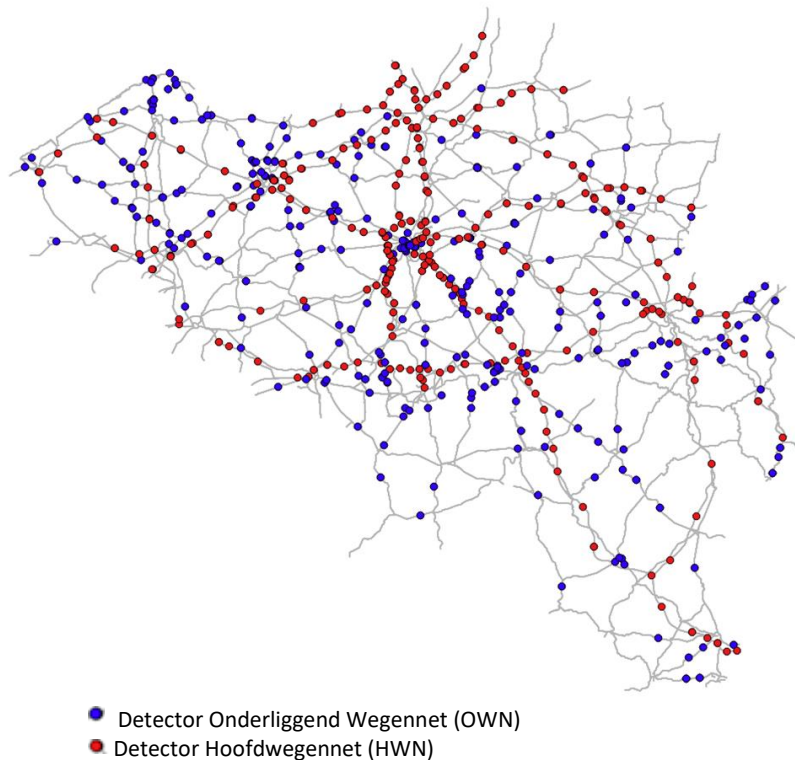
3.1.2.1 Car

The Belgian road network comprises numerous loop detectors which count how many vehicles pass at a certain location. Many loop detectors are permanently built into the structure of the road itself, both on the primary and on the secondary road network. The data from the loop detectors came from different sources, like *Het Vlaams verkeerscentrum*, *Departement Mobiliteit en Openbare werken*, *Brussel Mobiliteit* and *FOD Mobility and Transport*. Data provided by about 700 loop detectors were used for this study. About 225 detectors were located on the primary road network, and some 480 on the secondary road network.

Loop detectors are mostly found in Flanders, more than in Wallonia or Brussels.

	Primary road network	Secondary road network
Flanders	+/- 150	+/- 300
Wallonia	+/- 75	+/- 150
Brussels	0	+/- 30

Illustration: Position of loop detectors on the Belgian territory



3.1.2.2 Train

The following NMBS/SNCB data was used as benchmark: total number of boarding passengers on a week day, on a Saturday, and on a Sunday, for every railway station in Belgium. Additional data (total number of kilometers covered by public transport) were provided by the Federal Planning Bureau. These data were however not corrected for foreign passengers.

3.1.2.3 Tram/bus/metro

There are no reliable data for bus or tram passengers available. Only the total number of kilometers covered by public transport was made available by the Federal Planning Bureau as well. These data were however not corrected for foreign passengers.

As far as metro traffic is concerned, we collected information over the number of boarding and descending passengers per station during the morning (7 am – 10 am) and evening (4 pm – 19 pm) rush hours. Additionally, total numbers of travelers per station per year were provided as well.

3.1.2.4 Bike

For bicycle trips, the total amount of kilometers and number of trips covered was available as a benchmark for the regions Wallonia and the region of Brussels Capital.

3.1.2.5 Walking

As a benchmark for pedestrian trips, the total amount of kilometers walked is available for Belgium as a whole (this info is made available every 3 years by the Federal Planning Bureau).

3.1.3 Benchmark FOD

Every year the FOD Mobility and Transport reports the total amount of kilometers covered by vehicles in Belgium. This number is based on traffic counts, and broken down by region, road type, and vehicle category.

Information on the amount of kilometers covered by Belgian vehicles are also available. This data are based on car mileages collected by DIV (Directie Inschrijving Voertuigen) and complemented by Car-Pass information.

The FOD Mobility and Transport also reports the amount of kilometers covered by vehicle passengers.

3.1.4 Total kilometers

The table below gives an overview of the number of kilometers covered, broken down by transportation mode, which was used as benchmark.

	Km by transport mode per week
Car	2.335 Mio km
Pedestrian	42 Mio km

Bicycle	89 Mio km
Public Transports	239 Mio km
Total	2.705 Mio km

3.1.5 Speed on road network

Be-Mobile collects GPS positions from a large fleet of vehicles managed by some 30 different providers in Belgium. On an average month, around 310.000.000 valid GPS positions are reported by some 450.000 vehicles.

Speeds were at first calculated on the basis of 3 months data (Sept-Nov 2015), than updated for 2020 modelling, with traffic data from March-June 2019, amounting to a total of some 1.250.000.000 positions. Traveling times between two locations were deduced and attributed to the related road segments. The collected travel times were then aggregated and resulted in average speeds per hour and per day of the week. A total of 168 average speeds (7 days*24 hours) were used for the VA.

In some cases, Be-Mobile could not collect enough data to deliver reliable average speed predictions. In this case, the maximum speed allowed was used instead of the average speed. Such cases occurred mostly during night hours (not during rush hours) and predominantly in residential areas (not much traffic).

3.1.6 Volumes on road network

In the next phase, the different car benchmarks were aggregated into one coherent benchmark but first, a few small corrections had to be applied to the data.

3.1.6.1 Corrections for trucks and foreigners

The scope of this study is limited to Belgian passenger cars on the Belgian road network only. However, loop detectors are not sensitive neither to the different types of vehicles nor their origin. Therefore, it was necessary to correct the data and withdraw all trucks and foreign vehicles.

The FOD Mobility and Transport (2015) offers a benchmark for the percentage of trucks on the Belgian road network. These data are broken down by region (Flanders, Wallonia and Brussels) and by road type (primary road network, secondary road network). These percentages were applied to the data.

% Passenger Cars	Primary Road Network	Secondary Road Network
Flanders	74,98%	78,66%
Wallonia	77,27%	80,74%
Brussels	81,24%	78,34%

The same source offers a benchmark on the percentage of Belgian passenger cars. No breakdown was available per region; however, the secondary road type was further split into two categories: regional roads and secondary roads. Here again, the percentages were applied to the data.

% Belgian Passengers Cars

Primary roads	86,95%
Regional roads	96,00%
Secondary roads	98,00%

3.1.6.2 Creation of benchmark volume for the full network

In order to prepare a benchmark for car traffic volume for each segment of the Belgian road network, an algorithm was developed based on the fusion of floating car and loop detectors data.

Each data source has one vital part of the information.

- Loop detectors aggregate data on all vehicles passing by a certain point but their number is limited.
- Floating car data do aggregate information on the full road network, but they are collected by a sample of vehicles only.

The fusion of the two data sources allows to combine their strengths and create reliable traffic volumes for the whole road network.

For segments where no loop detector is present, an algorithm calculates a weighted average of volumes based on the loop detectors and on the extrapolation of the floating car data. The weighting factors result from the 'influence' of different loop detectors on the road segment at hand. The 'influence' factor of a loop detector on a given segment is defined as the proportion of vehicles which were registered both on the loop detector segment and on the segment which was only measured by the floating car data.

The algorithm has been applied to all segments of road classes 1, 2 and 3 (i.e. all segments with major traffic flows, out of a total of 6 road classes). As a result, traffic volumes have been assigned to each segment of the Belgian road network during a week. A confidence interval has been calculated per segment as well. The confidence interval indicates how much the actual volumes deviate from the estimated volumes. The size of the intervals depends mostly on the distance of the loop detectors, and vary between 5% (close to a loop detector) and 55% (far away from any loop detector).

The result of this process was used later on as a benchmark for travel volumes resulting from the activity based modeling.

4 Assigning traffic to road network

The output of the Activity Based Modelling is a list of all trips made by each individual during a week. Each trip consists of an origin and a destination down to the statistical sector level, a day of the week and time, a motive and a mode of transport.

The aim of the traffic assignment process is to allocate the trips to the road network. By using a route-planner, each trip can be allocated to a route. The route-planner takes the intensity of traffic into account, together with the availability of public transport means, broken down by day-part. The traffic assignment process results in:

- Traffic volume per road-segment, for car drivers, car passengers, pedestrians, cyclists and bus passengers.
- A total number of passengers for each metro or railway station.

4.1 Validation data sources

The output of the Activity Based Model has been compared to the available traffic benchmark data.

4.1.1 Home-work travel matrix

For home-work travels, the only benchmark available is the 2011 Federal census survey.

4.1.2 DIV total mileage

For car drivers

- The number of kilometres driven by car drivers, broken down by region, type of road (major - federal- roads, regional roads, secondary roads) and mode of transport (passenger cars, light trucks, trucks and busses, motorcycles).

- The number of kilometres accumulated by Belgian vehicles per year (same source).

For car passengers

- The number of kilometres covered by car passengers, broken down by region, type of road (major -federal- roads, regional roads, secondary roads) and mode of transport (passenger cars, light trucks, trucks and busses, motorcycles).

All of these figures were first made available by the FOD Mobility and Transport in 2015, and updated in 2018 and 2020 for the following editions of this study.

4.1.3 Federal Planning Bureau

4.1.3.1 For pedestrians/cyclists

Every 3 year, the Federal Planning Bureau publish statistics over the total number of kilometres covered annually, either as a pedestrian or on a bike. This figure was last published in 2015, and based on 2012 data. However, since this figure is aggregated, it had to be disaggregated on the basis of the UD figure. In spite of very limited differences between the 2 data sources, the decision was taken to retain the UD figure as benchmark.

4.1.3.2 For public transport

Every 3 year, the Federal Planning Bureau publish statistics over the total number of kilometres covered annually, either by bus, tram, metro or train. This figure was last published in 2015, based on 2012 data, and takes foreigners and children younger than 12 into account. Ultimately, the UD data were thought to be more reliable; hence, the decision was taken to retain the UD figure as benchmark.

4.2 Validation

The output of the Activity Based Model was compared to the following benchmark criteria:

- Number of trips per transport mode.
- Voyager kilometres per transport mode.
- Car passenger kilometres, broken down by region and road network.
- Number of segments with car traffic within the confidence interval.
- Number of passengers per railway and metro station.

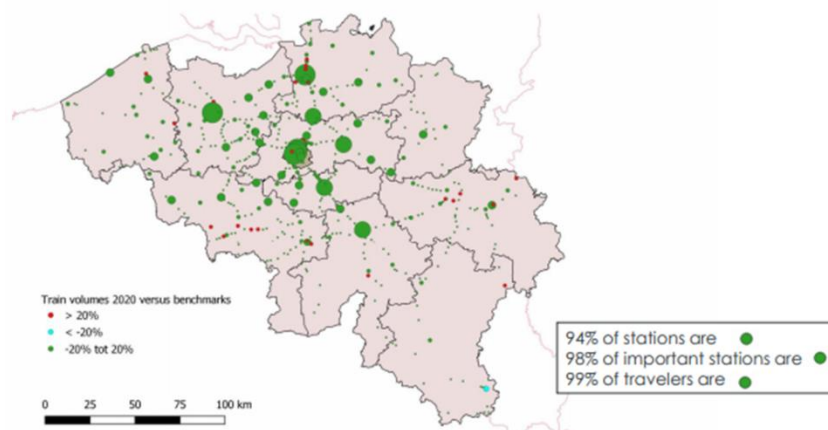
Predicted traffic volumes that were within the confidence interval of the benchmark (green) are considered valid.

Illustration: Primary roads confidence interval Secondary road confidence interval



In the case of railway and metro stations, the predictions of the Traffic Assignment Module were compared with the benchmark data, which are based on passengers' numbers per station. Foreigners, and children below the age of 12 were accounted for.

The map below shows which railway stations fell within (green), below (blue) or above (red) the benchmark.



4.3 Calibration

The traffic volumes that didn't pass the validation process were calibrated; the calibration occurred by fine-tuning the parameters of the model:

- Logit coefficients of the trip-frequency model
- For the 'other' motives, the Origin-Destination matrix of the destination model
- Logit coefficients of the mode-selection model

The calibration process was been carried out in several steps:

4.3.1 Number of trips per mode of transport

In a first process run, the number of trips was validated and calibrated at national level.

In a second instance, the methodology was improved in order to calibrate the results per 'modal-region'. The 5 'modal-regions' used were

- Brussels capital
- Flemish cities with share of public transport above 8% (Antwerpen, Gent, Leuven, Oostende)
- Rest of Flanders
- Walloon cities with share of public transport above 8% (Liège, Verviers, Namur)
- Rest of Wallonia

Through an iterative calibration process, the ABM output was aligned with the UD benchmark.

Share of Trips	Brussels cap.	Flemish Cities	Rest Flanders	Walloon Cities	Rest Wallonia
Car	41.9%	52,7%	69,2%	73,3%	79,5%
Pedestrian	32,5%	19,7%	11,8%	15,7%	13,3%
Bicycle	4,1%	18,7%	14,2%	1,2%	1,7%
Public Transport	21,4%	8,9%	4,7%	9,8%	5,5%

4.3.2 Voyager kilometers per mode

Where necessary, the calibration process occurred at 2 levels.

Distribution function of distances in the O-D matrix with trip-motive 'other'.

- The distribution function, used to produce the O-D matrix, has a big influence on the total kilometres. Since the major differences between the predicted and benchmark data were found for the 'other' motive, only this O-D matrix had to be calibrated.

Logit coefficient.

- In the case of pedestrians, cyclists and public transport passengers, the logit function is correlated with to the length of the trip. The calibration of these coefficients made it possible to modify the relative number of kilometres per mode.

The following results were obtained after calibration:

Million Kilometer / Week	Benchmark	Model
Car	2335	2385
Slow-modes (ped., bicycle)	131	128
Public Transport	239	243

4.3.3 Car traffic volume

The predicted car traffic volume on all road segments has been aggregated by road type and compared with the benchmark data.

As a result of this comparison, it appeared that the volume of 85% of the road segments fell within the confidence interval.

Network Type	Within Conf. Interval	Above	Below
Primary	76%	7%	17%
Regional	89%	3%	8%
Secondary	88%	3%	9%
Total	85%	4%	11%

However, if one aggregates several segments at random, the volume of these ‘trips’ that falls within the confidence interval increases dramatically. The table hereunder gives an estimation of the increase of the confidence interval in function of the number of analysed segments :

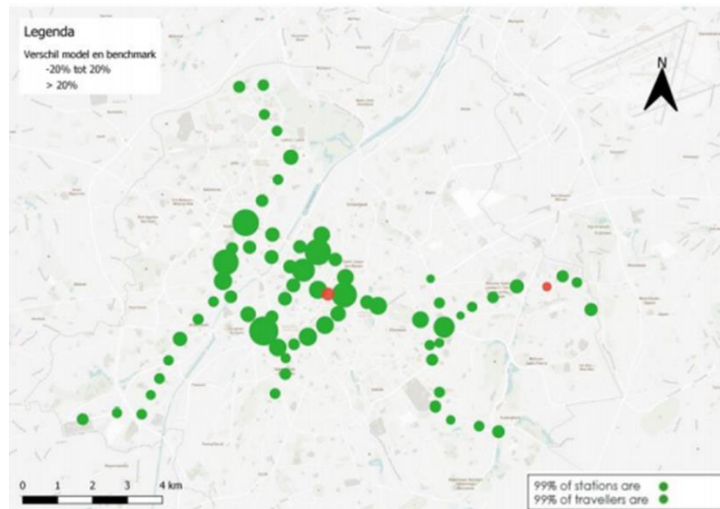
N of segments in ‘Trip’	Within Confidence Interval
1	83,0%
10	95,0%
25	99,0%
100	100,0%

4.3.4 Number of passengers per train and metro station

The predicted number of passengers per train/metro station has been compared with the available benchmark data.

This comparison takes into account the fact that foreign passengers and children below the age of 12 are included in the official figures reported by the Federal Planning Bureau, but not in the benchmark data.

After calibration and validation, we obtained the following result:



The map above shows which metro stations fell within (green), below (blue) or above (red) the benchmark.

5 Software delivery system

The Inventory Delivery Software (IDS) is managed by MGE Data. MGE Data is an international out-of-home media expert, operating out of the Czech Republic. Their main fields of expertise are OOH audience research, OOH visibility adjustment, mobility surveys, traffic analysis and measurement, geomatics and geo-marketing, delivery and logistics. For the CIM OOH project, MGE Data is responsible for both the inventory management software (IMS) and the delivery software (IDS).

IDS allows to create, copy and delete groups of panels. The media analytics of groups of panels may then be calculated against target groups, composed on the grounds of the available variables

- Personal socio-demographics
- Geographical socio-demographics
- Travel socio-demographics (mode of transport, motive)

6 Appendix 1: Variables used in the Unified Database (UD)

6.1 Personal characteristics

1. Gender
 - a. Man
 - b. Women
2. Age
 - a. 12-14 years old
 - b. 15-17 years old
 - c. 18-20 years old
 - d. 21-24 years old
 - e. 25-29 years old
 - f. 30-34 years old
 - g. 35-39 years old
 - h. 40-44 years old
 - i. 45-49 years old
 - j. 50-54 years old
 - k. 55-59 years old
 - l. 60-64 years old
 - m. 65 + years old
3. Profession
 - a. Small commercial, freelance industrial 5-, artisan and farmer
 - b. Big commerce, freelance and industrial 6+, management and liberal profession
 - c. Employee
 - d. Worker
 - e. Retired
 - f. Student
 - g. Housewife
 - h. Unemployed
 - i. Other
4. Education
 - a. No education or primary education
 - b. Lower secondary education
 - c. Higher secondary education
 - d. Higher education (bachelor or master)
5. Personal Income
 - a. Category 1: 0 – 750 €
 - b. Category 2: 751 – 1.500 €
 - c. Category 3: 1.501 – 2.000 €
 - d. Category 4: + 2.000 €

6. Family income

- a. Category 1: 0 – 1.000 €
- b. Category 2: 1.001 – 2.000 €
- c. Category 3: 2.001 – 3.000 €
- d. Category 4: + 3.000 €

7. Family size

- a. 1
- b. 2
- c. 3
- d. 4
- e. 5+

8. Car ownership

- a. 0
- b. 1
- c. 2
- d. 3+

9. Home address

- a. Post code
- b. 1
- c. 2
- d. 3+

6.2 Travel characteristics

1. Date of trip
 - a. Day/month/year
 - b. Day of week
2. Time of day
 - a. Departure time
 - b. Arrival time
 - c. Part of day (6h-10h, 10h-16h, 16h-19h, 19h-6h)
3. Travel time in minutes
4. Travel time in kilometers
5. Mode of transport
 - a. Car driver / motorcycle
 - b. Car passenger
 - c. Pedestrian
 - d. Bike / scooter
 - e. Bus
 - f. Tram
 - g. Metro
 - h. Train
 - i. Other
6. Motive
 - a. Work
 - b. School
 - c. Business
 - d. Shopping
 - e. Visiting someone
 - f. Social activity
 - g. Pick-up/drop-off someone
 - h. Go home
 - i. Services
 - j. Other

7 Appendix 2 : Unified Database cleanings

7.1 Education

- Lower secondary degree was missing in the BELDAM Dataset. This means that the primary degree category had to be split up into primary and lower secondary. To do this, a relation between education and other personal characteristics was found in the other 2 datasets and that relation was then applied to the BELDAM dataset. Age was used as a hard condition. Respondents of age 12 were not allowed to have a lower secondary degree, since this is not possible in Belgium.
- Several response categories in all 3 datasets were aggregated to come to 4 categories. For example, no degree and primary degree were aggregated in the OVG dataset, and all different types of higher education degrees found in all three datasets were aggregated into 1 'higher education' category.

7.2 Profession

- No variables had to be split up
- Several professions have been aggregated to get the final 9 categories.

7.3 Personal income

- Personal income was estimated for the CIM OOH and the BELDAM datasets. To that end, the predicting variables education and occupation have been used.
- The three highest personal income categories from the OVG dataset were aggregated into 1 category (2.000 € +)

7.4 Family income

- Family income was estimated for the CIM OOH dataset. To that end, the predicting variables personal income, family size and car-ownership have been used.
- Several income categories from the OVG and BELDAM datasets were aggregated to get the final 4 response categories.

7.5 Family size

In the BELDAM, OVG and MONITOR datasets, all family sizes of 5 people and more were aggregated into one 5+ response category.

7.6 Car-ownership

Car-ownership was estimated for the CIM OOH dataset, on the basis of personal variables and travel variables.

7.7 Date of trip

While the day of the week is relevant for the activity based model, the exact date of the trip isn't. For each trip, the day of the week was derived from the exact date.

7.8 Time

While the day-part is relevant for the activity based model, the exact time of the trip isn't. For each trip, the day-part was derived from the exact time.

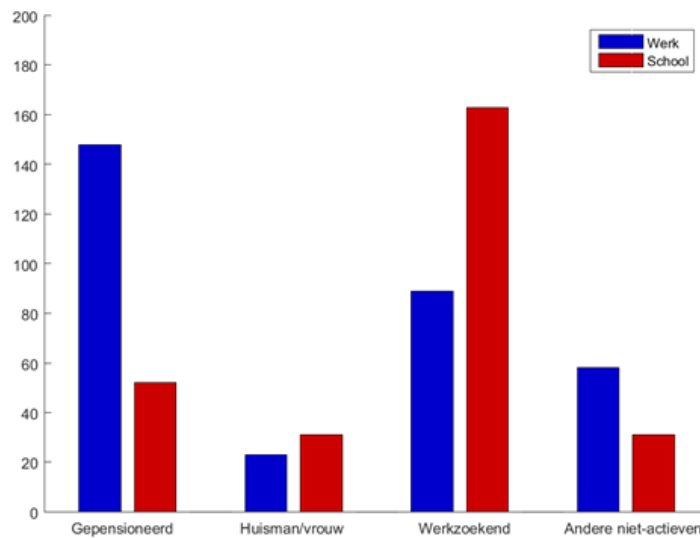
7.9 Mode of transport

- Several response categories in each dataset were aggregated to get the final 10 response categories.
- Car trips had to be split into car passengers and car drivers for the CIM OOH dataset. To that end, personal characteristics and travel characteristics were taken into account. For car drivers, a hard age limit of 18 years has been taken into account (legal driving age in Belgium).
- Motorcycle and scooter information in the BELDAM dataset had to be split. To that end, the speed of travel (travel distance / travel time) was taken into account. All 40 km/hour and over travels were assigned to motorcycle, slower travels were assigned to scooter. The distinction between the two categories is important because of the influence of speed on the VAI calculation.
- Tram/metro was considered as 1 response category in the OVG dataset. These specific travels were processed by a multimodal route planning tool, and the results of the route planning process (tram or metro) were assigned to the travels.
- The BELDAM dataset showed a significant proportion of pedestrian travels, mostly in Brussels. In the CIM OOH dataset, pedestrian travels were much less frequent. Analyses revealed that a lot of the BELDAM pedestrian travels were short trips from and to public transport stops. Based on public transport route planning of the CIM OOH dataset that was done in the previous OOH study, a lot of short pedestrian travels were identified in the CIM OOH dataset. Hence, short travels have been added to the CIM OOH dataset in order to bring the number of pedestrian travels to a more realistic level.

7.10 Motive

- In the CIM OOH dataset, work and school travel were considered as one response category. However, it was very important for the activity base model that this category be split into two categories. The disaggregation process was based on the following assumptions: students, housewives and unemployed people make school travels; while active people, retired and other non-active people make work travels. These assumptions were based on the analysis of the two other datasets.

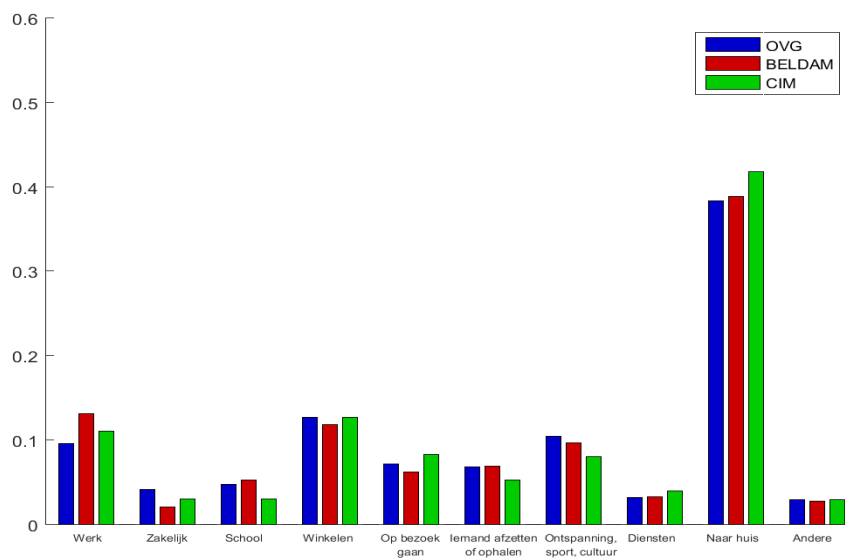
Illustration: Split work/school motive per activity



- In the OVG and BELDAM datasets, several response categories had to be aggregated to result in the final 10 response variables.

Illustration: Distribution of motives across datasets

Distributie motief



- In the OVG and BELDAM, wherever necessary, a final motive 'going home' was attributed at the end of the day.

8 Appendix 3: Activity Base Model (ABM) characteristics

8.1 Number of trips / motive

The following variables were used in the activity based model to calculate the number of trips per motive.

8.1.1 Work

1. Gender
 - a. Male
 - b. Female
2. Occupation
 - a. Student
 - b. Retired, housewife
 - c. Looking for work, other
 - d. Employees, workers, self-employed
 - e. Management, free professions

8.1.2 School

1. Occupation
 - a. Students
 - b. Not-students
2. Age
 - a. 12-17 year old
 - b. 18-24 year old
 - c. 25+ year old

8.1.3 Shopping

1. Gender
 - a. Male
 - b. Female
2. Occupation
 - a. Self-employed, management, free professions, employees, worker, student
 - b. Retired, housewife
 - c. Looking for work, other
3. Family Size
 - a. 1 person
 - b. 2 persons
 - c. 3 persons
 - d. 4+ persons

8.1.4 Picking up someone / dropping off someone

1. Gender
 - a. Male
 - b. Female
2. Occupation
 - a. Self-employed, worker
 - b. Retired, student
 - c. Other
3. Family Size
 - a. 1 person
 - b. 2 persons
 - c. 3 persons
 - d. 4+ persons

8.1.5 Visiting someone

1. Age
 - a. 12-45 years old
 - b. 45+ years old
2. Occupation
 - a. Self-employed, management
 - b. Student, looking for work
 - c. Other
3. Family size
 - a. 1-2 persons
 - b. 3-4 persons
 - c. 5+ persons

8.1.6 Other motives (social, services ; others)

1. Age
 - a. 12-45 years old
 - b. 45+ years old
2. Occupation
 - a. Self-employed, management
 - b. Student, looking for work
 - c. Other
3. Family size
 - a. 1-2 persons
 - b. 3-4 persons
 - c. 5+ persons

8.2 Transport mode

8.2.1 Slow/Non-slow transport mode

The following variables were used in the utility function to choose between slow and non-slow transport modes.

1. Occupation
 - a. Self-employed, management, employee, worker
 - b. Retired, housewife, unemployed, other
 - c. Student
2. Family car ownership
 - a. No car
 - b. 1 car
 - c. 2 cars
 - d. 3+ cars
3. Distance of trip
 - a. 0-1 km
 - b. 1-2 km
 - c. 2-3 km
 - d. 3-4 km
 - e. 4-10 km
 - f. 10+ km
4. Trip motive
 - a. Work/business, picking up someone
 - b. School
 - c. Visiting someone
 - d. Social
 - e. Shopping, service, other

8.2.2 Private/public transport

Within the non-slow modes, a choice had to be made between car and public transport. The following characteristics influenced this second choice.

1. Occupation
 - a. Self-employed, management, employee, worker
 - b. Retired, housewife, unemployed, other
 - c. Student
2. Family car ownership
 - a. No cars
 - b. 1-2 cars
 - c. 3+ cars
3. Distance of trip
 - a. 0-1 km
 - b. 1-2 km
 - c. 2+ km

4. Trip motive
 - a. Work/business, social, services
 - b. School
 - c. Visiting, shopping, picking up someone
 - d. Other
5. Travel time difference between car and public transport