



ClairCity: Citizen-led air pollution reduction in cities

Deliverable 5.7: City Impact Analysis Report - Methodology

February 2020

Document Details

Authors	Vera Rodrigues, Myriam Lopes, Sandra Rafael, Sílvia Coelho, Kevin de Oliveira (UAVR) Kris Vanherle (TML) Peter Papics (TML) Willem Himpe (TML) Iason Daifas (PBL) Angreine Kewo (DTU) Carlo Trozzi (Techne Consulting) Enzo Piscitello (Techne Consulting) Joana Soares (NILU)
Contact	kris.vanherle@tmleuven.be vera.rodrigues@ua.pt
Creation Date	08/05/2019
Date of Last Revision	28/02/2020
Description	The Deliverable 5.7 collects all results of the full impact analysis, methodological approach and data sources.

Version History

Version	Updated By	Date	Changes / Comments
V1.0	Kris Vanherle	08/05/2019	Outline
V1.1	Kris Vanherle	12/11/2019	Added content by Peter (TML) and Angreine (DTU)
V2.0	Kris Vanherle	05/02/2020	Further inputs TECHNE, DTU, PBL
V2.1	Kris Vanherle	18/02/2020	Updated inputs Aveiro and updated as stand-alone document
V3.0	Vera Rodrigues	28/02/2020	Update and final revision

Contributions and Acknowledgements

Project Director	Hans Bolscher (Trinomics)
Project Manager	Irati Artola (Trinomics)
Project Technical Director	Enda Hayes (UWE)
WP Manager	Kris Vanherle (TML)

Table of Content

Document Details	2
Version History	2
Contributions and Acknowledgements.....	3
Introduction	6
1 Integrated model	8
1.1 Synthetic population	8
1.2 Generation of activity chains including temporal dimensions	13
1.3 Spatial allocation of activities.....	18
2 Temporal dimension of emissions	20
2.1 Data Pre-processing.....	20
2.2 Temporal Load profiles.....	22
2.3 Implementation.....	24
3 Transport.....	32
3.1 Transport demand.....	32
3.2 Emission factors.....	37
3.3 Fleet dynamics	39
3.4 Mode choice.....	49
4 Industrial, Residential, Commercial and Institutional emissions (IRCI).....	51
4.1 Industrial sources	51
4.2 Residential, commercial and institutional sector	51
4.3 Classification of activities and fuels for emission estimates	56
4.4 Data evaluation on domains and subdomains	57
4.5 Future emissions calculation	58
5 Carbon footprint	59

5.1	Specifications of modelling tool set.....	59
5.2	Territorial domains and subdomains.....	63
5.3	Classification of activities and fuels for carbon footprint estimates.....	64
5.4	Data evaluation at domains and subdomains	65
6	Air quality	67
6.1	Air quality and source apportionment modelling at mesoscale	67
6.2	Air quality modelling at urban scale	69
7	Health.....	71

Introduction

This report concludes all the modelling activities in the ClairCity project. We developed a full envelope for the assessment of policy measures on local air quality, starting from activity data to generate air quality, health and carbon footprint impact. This report summarizes the work of several project partners: TML focussed on the transport side, PBL on land-use and citizen activity, DTU integrated the practices/behaviour elements of citizens in typical energy consumption temporal profiles, TECHNE developed the industrial emission as well as the carbon footprint calculations, UAVEIRO used all emission calculations from the previous partners as input to produce air quality estimates and NILU estimated health impact from these air quality estimates.

All these steps require a different modelling approach, yet in a clear-cut sequential approach. In this report, we elaborate first on the general approach taken for each of these steps and secondly elaborate on the detailed results for each of the 6 cases, in separate chapters.

Given the size of the document, we split D5.7 in several documents:

1. **General methodology of each module** (this document): For all 6 cases and for the simplified approach in the generic model, we rely on a common methodology insofar possible for the cases. Some cases are data rich (e.g. Amsterdam, Bristol), others are data-poor (e.g. Aveiro, Sosnowiec). Also, not all data is available in the same format for all cases, so some specific adaptations are needed. In this first chapter, we elaborate on the **common** methodology applied for all cases.
2. **Document/Chapter per city case**: As indicated, we made some changes to the general methodology for the cases when required. In separate documents, we elaborate for each case on:
 - a. City specific particularities (e.g. method, data)
 - b. Scenario description and implementation in the modelling
 - c. Detailed results, for the baseline, the business as usual scenario (BAU) and for the scenario's, including intermediate results after each step in the assessment procedure.

The chapter per city case is available into the ClairCity community in Zenodo through the generic DOI <https://doi.org/10.5281/zenodo.3690724>.

Air quality and health impact assessment follows a fairly standard sequential approach. Emission sources are estimated at sector level. Spatial and temporal detail is added. These detailed emission estimates serve as input for an air quality model which provides results in terms of air quality level to the health assessment. We thus come to the following steps:

1. Integrated land-use for the assessment of energy consumption of the residential sector.
2. Temporal distribution to derive intra-day and seasonal variation of residential emissions
3. Estimating passenger and freight transport energy consumption and emissions.

4. Adding Industrial, Residential, Commercial and Institutional emission estimates, from (detailed) energy consumption for the residential sector and available emission assessments for other sectors
5. Carbon Footprint assessment to derive carbon footprint from emission inventories
6. The air quality modelling, using the emission data the estimate detailed air quality maps
7. Convert population exposure and air quality results to estimate exposure and health impacts

In this chapter, we elaborate in detail on the general methodology of each of these steps

1 Integrated model

The methodology for developing the integrated module for the city of Amsterdam and served as a template for subsequent cases, aiming to simulate households' practice/activity behaviour, encompasses three steps:

- 1) Creation of a synthetic population of the entire city
- 2) Generation of activity chains including temporal dimensions
- 3) Spatial allocation of trips, conditional on the activity chains from step 2

1.1 Synthetic population

Modelling phenomena at an individual, household and geographical scale simultaneously requires spatially explicit data of all individuals in the relevant geographical area. Such data is generally not available: individual-level activity data such as travel diaries, are typically collected through surveys, covering only a part of the population. In the best case, the sample is representative on a national level, meaning that it is not representative for smaller areas. Furthermore, individual-level data on income and expenditure are likely to be unavailable for the public for privacy reasons. Finally, individual-level data may not exist at all.

In Clair-City all these challenges were present. Overcoming them has necessitated the creation of synthetic spatial micro data. This technique is sometimes referred to as spatial microsimulation and it involves the creation of a synthetic micro population for small administrative areas by combining a spatial microdata with geographically aggregated zone-level data. The key elements of the approach can be summarized as follows:

- Spatial microsimulation is inherently concerned with how things vary over space, not just between individuals, groups or periods of time.
- Spatial microsimulation explores issues at the individual-level, as implied by the word micro.
- Spatial microsimulation involves the creation of fictitious data for modelling purposes, captured by the word simulation.

In general, the population synthesis entails distributing, or replicating, individual entries in national micro-level household survey such that the aggregate matches key figures in aggregate data on small areas. This requires specifying relevant variables which are available in both micro data as well as in the aggregate data. In addition to availability, the choice should obviously reflect the expected explanatory power of each variable for the problem at hand.

For the synthesized population of Amsterdam, the data came from three sources.

1.1.1 Dutch Mobility Survey (OVIN)

The Dutch Mobility Survey (OVIN) is collected yearly by Statistics Netherlands (CBS). The goal of the survey is to provide adequate information on the daily mobility of the Dutch population. To this end, the movement behavior of the Dutch population is described by place of origin and destination, time of the carriage, means of transport used and reasons for travel. In addition, considerable attention is paid to the background variables for a particular movement pattern and choice of transport.

Each version of OVIN contains around 40,000 respondents, and CBS assumes that the survey covers 1.9% of total travelled kilometres in the Netherlands. Respondents are individual persons and not households, although the data set contains information about household-level variables for each respondent.

We used 7 versions of OVIN (2010-1016) to generate a seed of individuals, ignoring the location of residence of each respondent.

1.1.2 Netherlands' Household Survey (WOON)

The goal of the Netherlands' Household Survey (WOON) is to gather information about the housing situation of the Dutch population and their living requirements and needs. The survey includes information about the composition of households, the dwelling and living environment, housing costs, living requirements and housing re-locations. Frequency is three-yearly. A minimum of 60,000 respondents have to be reached, as the survey should also provide reliable information on small geo- graphical areas. We used two versions of WOON (2012 and 2015).

WOON contains a large number of variables about households and dwellings, and we have to make choices on which household characteristics to include. Firstly, included variables should cover basic demographic information and they must be associated with the outcome(s) of interest (Burden & Steel, 2016). Secondly, the variables need to exist in both the survey data as well as in the aggre- gate geographical data. WOON contains additional information about the number of household members by age.

1.1.3 The Neighbourhood and District maps (Buurt en wijk kaarten)

The Neighbourhood and District maps from Statistics Netherlands contain digital geographical information as well as key figures for municipalities, neighbourhoods and districts (gemeente, buurt, wijk) in the Netherlands. Key figures include data on population and its composition; firms; housing stock; energy use; income and transport. For the microsimulation we used the 2014 version of the data.

Table 1 is a list of household- and dwelling variables. These are selected on the basis of the explanatory power in models for energy use, gas use and car ownership. Note that we use gender and ethnicity and labour market status of the household head as representative for the entire household.

Table 1: Household variables and description

Variable name	Variable description/unit	Note
hh_id	household ID	
bu_code	neighbourhood code	
pc4	Postcode	
dw_type1	rental or owner-occupied	the number of dwellings within a zone is equal to the number of households
dw_type2	singlefamily or multistorey	
dw_byear	building year (before or after 2000)	

hh_type	Composition	
hh_gender	gender of household head	
hh_income	lower 40 %, mid, upper 20%	
hh_ethn	ethnicity of household head	
hh_age	age of a randomly selected household member	
hh_lmstatus	labour market status of household head	
elec	electricity usage (kwh/year)	Predicted
gas	gas usage (m3/year)	Predicted
car ownership	Boolean	Predicted
long	long coordinate	Simulated
lat	lat coordinate	Simulated

Table 2 lists the individual-level variables. These are primarily chosen on the basis of their explanatory power in predicting time individuals spend on daily activities. One potentially serious problem here is that the predictive power of any statistical model drops when the number of things being predicted increases. To overcome this issue we followed common practice in the literature, distinguishing five categories of activities. Categories 1-5 encompass all outside-home activities, while category 6 captures all home activities. Time spent at home was simply calculated as 24 minus hours spent on other activities.

1. Work
2. Study
3. Shopping and non-work related provision of services
4. Recreational
5. Discretionary (visits and other)
6. Home

Table 2: Individual variables

Variable name	Variable description/unit	Note
bu_code	neighbourhood code	
ind_id	individual ID	
hh_id	household ID	
ind_gender	gender	
int_ethn	ethnicity	
ind_lmstatus	labour market status	
ind_work	time (h/year) spent on work	Predicted
ind_recreational	time (h/year) spent on recreational activities	Predicted
ind_discretionary	time (h/year) spent on discretionary activities	Predicted
ind_home	time (h/year) spent at home	Predicted

The most commonly used method for spatial microsimulation is the reweighting method Iterative Proportional Fitting (IPF). Reweighting methods create a synthetic population by reweighting a current national micro-data file to small area benchmarks from another data

source. As such IPF entails calculating a weight for each individual observation: individuals which are highly representative of a zone receive a higher weight.

In general, the main output of a spatial microsimulation exercise is a table representing the most probable configuration of the population of small areas. The table is generated using a 'seed' of individuals or households from a survey and a number of constraints from geographically aggregated data. One of the most frequently used algorithms to generate the table is IPF. IPF is simple, computationally efficient, rigorously founded and it is in practice identical to maximum entropy (Thissen&Löfgren, 1998).

The output of IPF, when used in spatial microsimulation, is a series of non-integer weight matrices of a survey sample. Each cell in the matrix indicates how representative one survey respondent is of the real population within each geographical area. The weight matrix thus gives the most probable configuration of individuals in these areas.

The standard population synthesis using IPF involves two steps: the first step is to generate a joint multiway distribution of all relevant attributes of households or individuals. Next, persons or households are drawn from a seed of individual records in order to satisfy the distribution of attributes. The last step involves creating a list representing the synthetic population of individuals or households, using the weights generated by the IPF procedure. Since the weights are fractional, this step also involves integration of the weight matrix (Lovelace & Ballas, 2013).

Due to the popularity of IPF, there are several accessible and ready-to-use implementations of IPF in software programs such as R (Lovelace & Dumont, 2016). For the project's needs, the synthetic population of individuals has been generated in R using the IPF procedure from the IPF package (Blocker, 2016).

As input for the synthetic population of individuals we used the respondent from the OVIN data set to generate the seed. The output of IPF in this case is a respondent-neighbourhood weight matrix consistent with the individual-level constraints for each neighbourhood. Constraints are the labour market status (*ind_lmstatus*), ethnicity (*ind_ethnicity*), gender (*ind_gender*) and age (*ind_age*) of each individual. The Buurt en Wijk Kaarten contains data for all these variables in percentages. To obtain the actual number we simply multiplied the percentage with the number of individuals in the neighbourhood.

1.1.4 Allocating individuals to a household

As indicated above, the spatial micro data in Clair-City needed to capture individuals and households simultaneously. This is, however, far from straight forward. Reweighting algorithms such as IPF generally break down when trying to match individual and household marginals at the same time. Our solution is to generate household- and individual-level data simultaneously and combining them afterwards. The combination can be conceptualized as a sorting problem, where individuals are allocated to a household where they are 'likely to belong'.

Our spatial micro data consists of both individuals and the household to which they belong. In travel diaries, individuals are usually the respondents, while housing demand surveys are generally filled in by a respondent acting on behalf of the household. As such, our spatial micro data consists of both household-level and individual-level data. In a final step these two are combined such that individuals are allocated to a household. This means we chose both individual-level constraints as well as household/dwelling-constraints.

In our case, the individual and household-level data come from independent sources, potentially causing inconsistency: the number of individuals allocated to a household may be much higher than the actual number of members in that household, or some individuals are not allocated to a household at all.

To overcome these challenges, we chose the following strategy:

1. Reweighted the number of household members for each household in a particular zone with the number of individuals within that zone.
2. Used the updated number of household members as well as the share of household members by type within a household to predict the number of household members by type.
3. Solved the individual-household allocation problem as an entropy-maximization problem, with the number of household members by type as prior.

1.1.5 Adding geographical coordinates to each dwelling

A spatial micro data is 'spatial' in the sense that individuals and households are allocated to a small geographical area. For air pollution source apportionment, a dwelling is considered a point-source. We therefore needed to allocate households to a more fine-grained level, i.e., we wanted to find unique geographical coordinates of each dwelling (note that the number of households are equal to the number of dwellings).

Here we implemented the following strategy:

1. Carried out the IPF procedure, obtaining a spatial micro population of each neighbourhood.
2. Found the coordinates of each neighbourhood-polygon.
3. Using data on within-city land use, we found the subset of land within each neighbourhood which is currently residential.
4. Sampled a number of spatial points, equal to the number of households within the neighbourhood, within the subset of residential land.

Figure 1 shows a small subset of households as well as their predicted coordinates from the case-study of Amsterdam. As the figure clearly suggests, the algorithm works fairly well: most households are allocated to residential areas, however some households are allocated within a park. Small discrepancies between the land-use data and spatial data sets on neighbourhoods are likely to have caused the few errors.

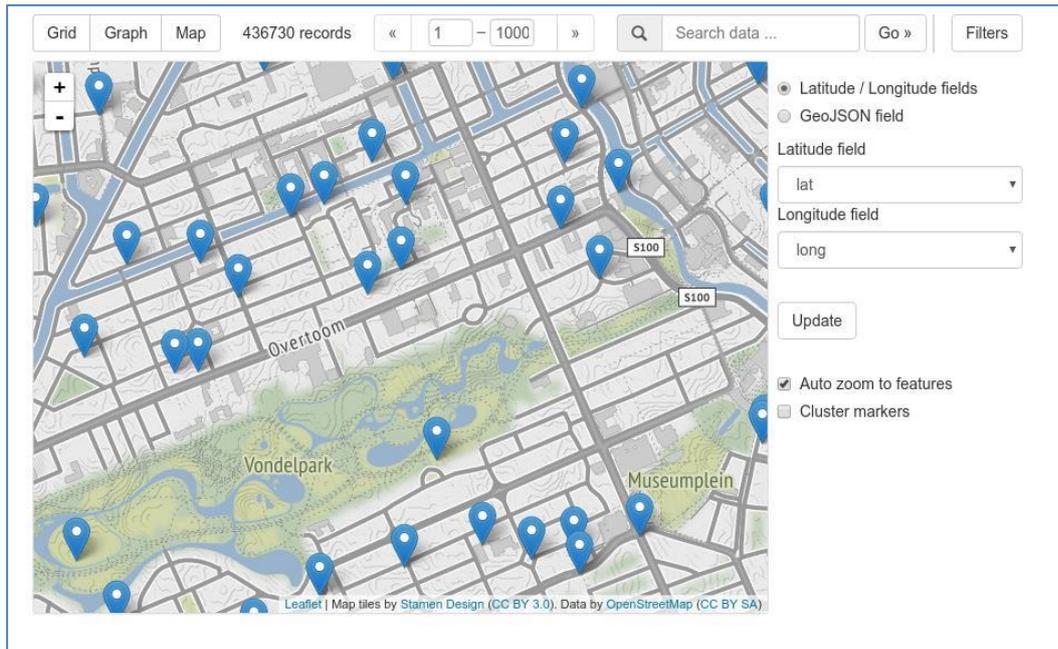


Figure 1: An example of geolocated households in Amsterdam

1.2 Generation of activity chains including temporal dimensions

Activity chains were generated using a fairly simple probabilistic sampling approach. For each individual-type we collected unique activity chains from the mobility survey data. We gathered these possible chains for each individual type in two different lists: one for weekdays and one for weekends. Each possible chain was then weighted with its frequency in the survey data. Sampling was then carried out using the normalised weights as probabilities.

Time spent on different activities varies between days and time of the day. The Mobility Survey Netherlands offers estimates of how time is spent per activity. In this survey one observation represents one unique trip, possibly part of a larger trip chain, for one respondent. Figure 2 shows the average time in minutes per activity. It shows that the relative amount of time spent on studying and working tends to diminish if the departure time is later than 08 and it starts increasing when the departure time is around 16. We also see that the distribution of time over activities does not differ significantly between weekdays and weekends, i.e. the average amount of time spent on working or studying for those respondents who did study or work is relatively stable across days.

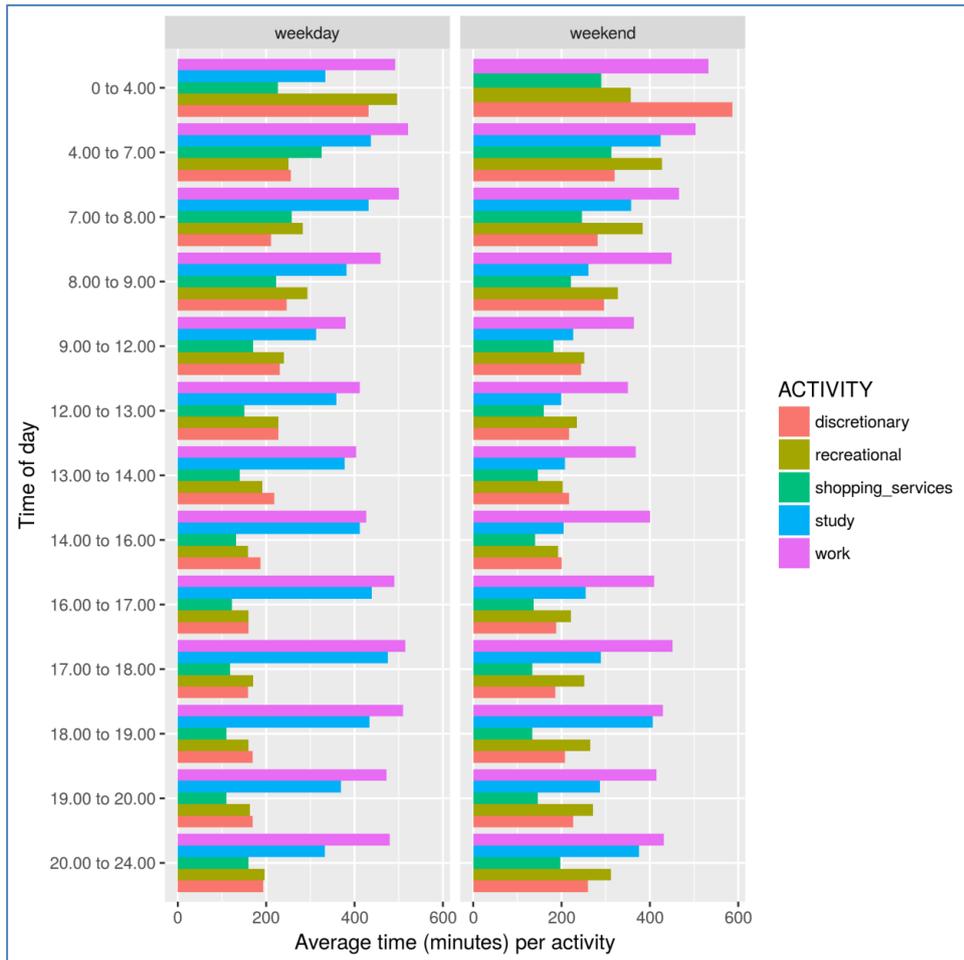


Figure 2: Activity time across day for different activities

However, Figure 2 only shows time use among respondents who actually participated in the activity. Figure 3 shows participation in activities among respondents. This does indeed suggest that there are large differences between both time of day and between weekday and weekend in terms of participation. For example, it is readily seen that hardly any activities were initiated between 0 and 4, suggesting that the average time spent on activities in Figure 2 might be affected by outliers. We also see that studying and working are less frequent in the weekends and for departure times before 9 or after 16 on weekdays.

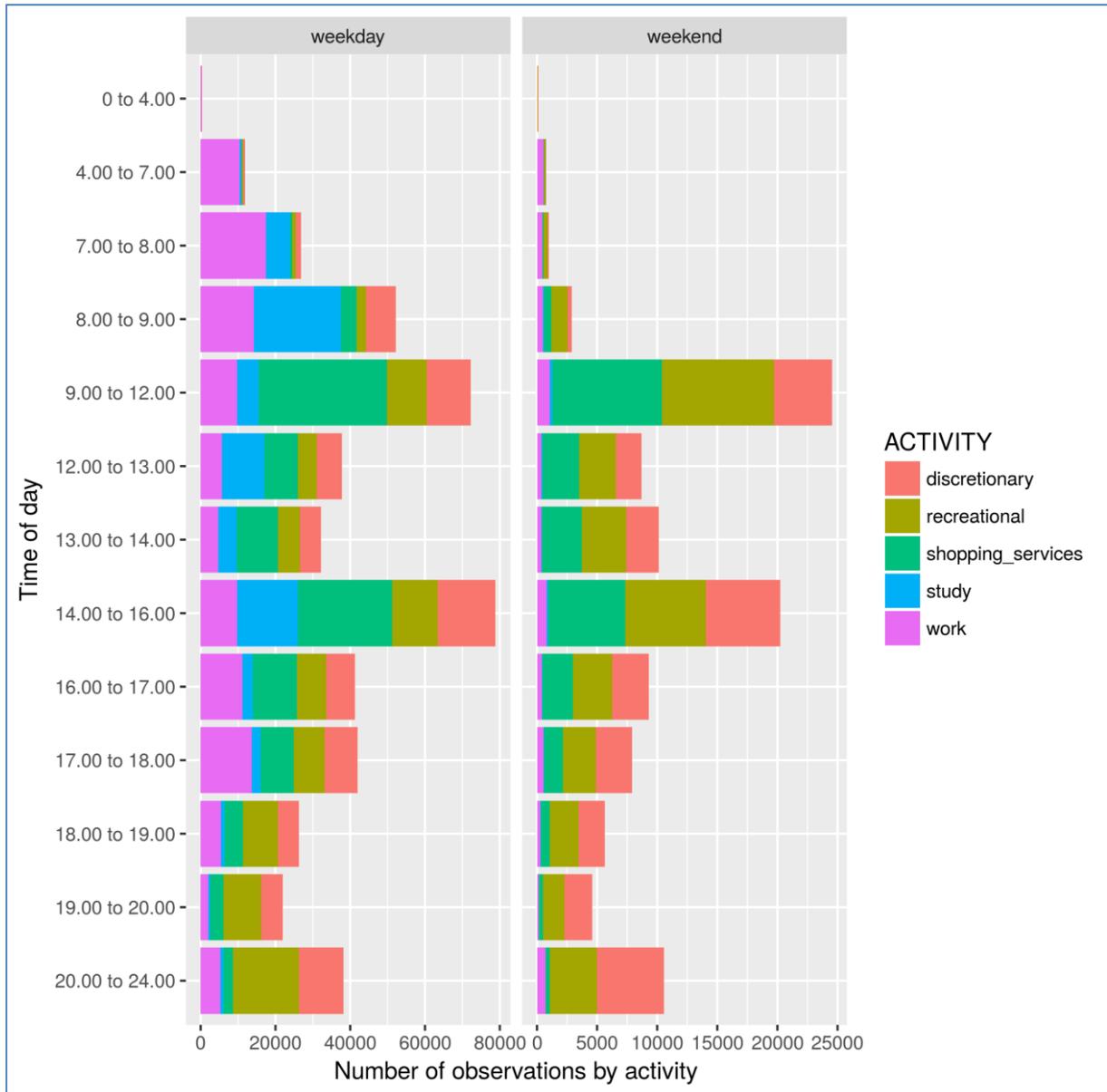


Figure 3: Participation in activities by respondents

A key element of the analysis is to predict the participation in activities by individuals. This means that, in addition to day and time of day, we would like to know how characteristics of individuals affect the choice. In order to guide the model development, we formulated a classification tree for each activity. The classification tree on the following pages was estimated using a package for recursive partitioning (rpart) in the software R. A recursive partitioning model that strives to correctly classify members of the population by splitting it into sub-populations based on several dichotomous independent variables. For example Figure 4 shows how the likelihood of participating in discretionary activities are determined primarily by day, time of day and mode. The lowest probability (18%) is obtained with observation on weekdays. The highest probability (63%) is found on weekends (the highest level), not during midday or morning (level 2), not by bike, other, public transport, scooter or walking (i.e., car) and not in the afternoon (i.e. evening). As such, the highest probability is obtained in the weekend, in the evening when a car is used.

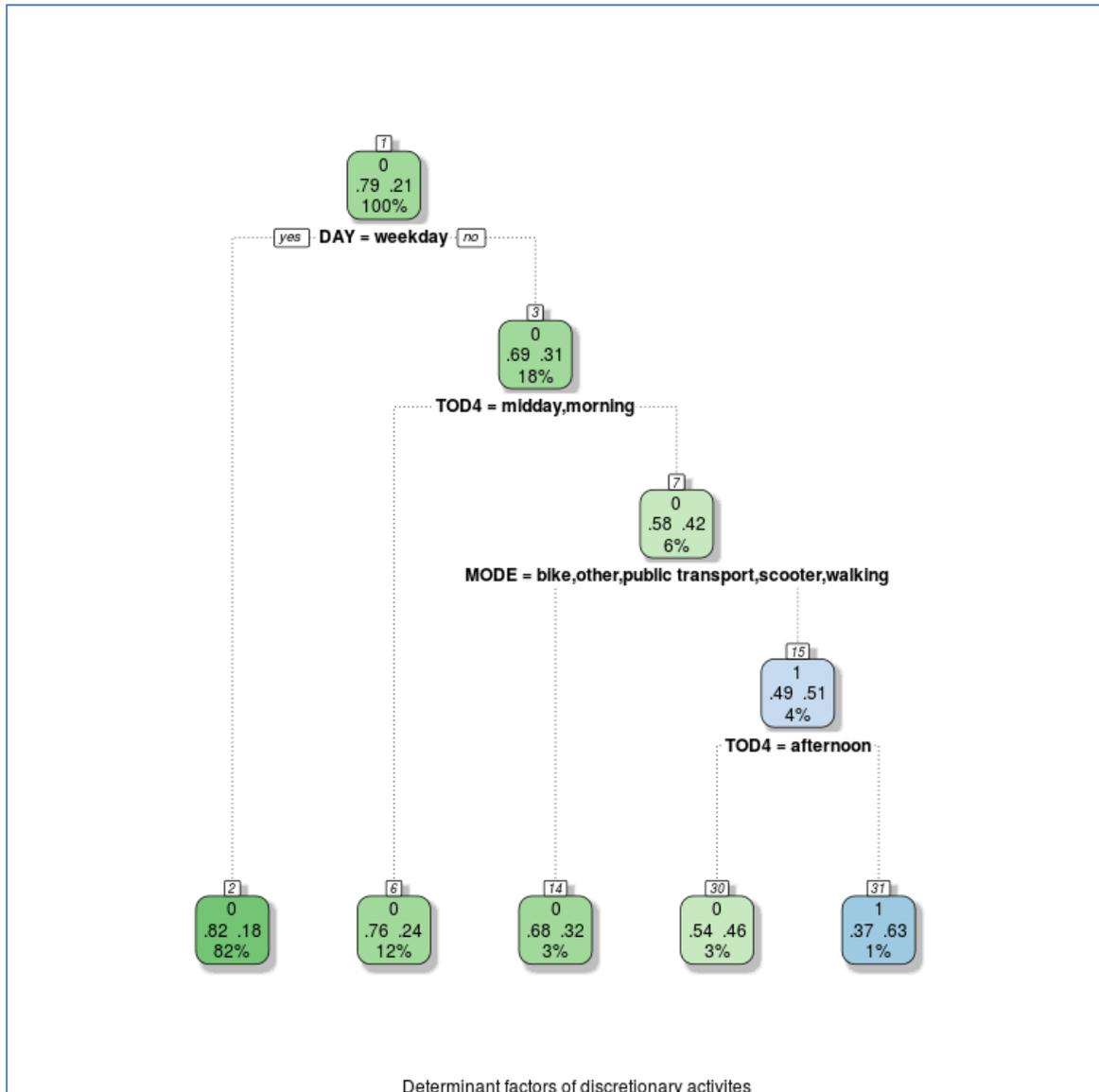


Figure 4: Decision tree for participation in activities

Similarly, a slightly more interesting picture can be found in Figure 5, which illustrates the likelihood of studying. We see immediately that the lowest probability (2%) is observations where the role of the respondent is either household head or partner/other. The highest probability occurs for observations where respondent is child, where the time of day is morning, where labour market status is non-active and when day of the week is weekday.

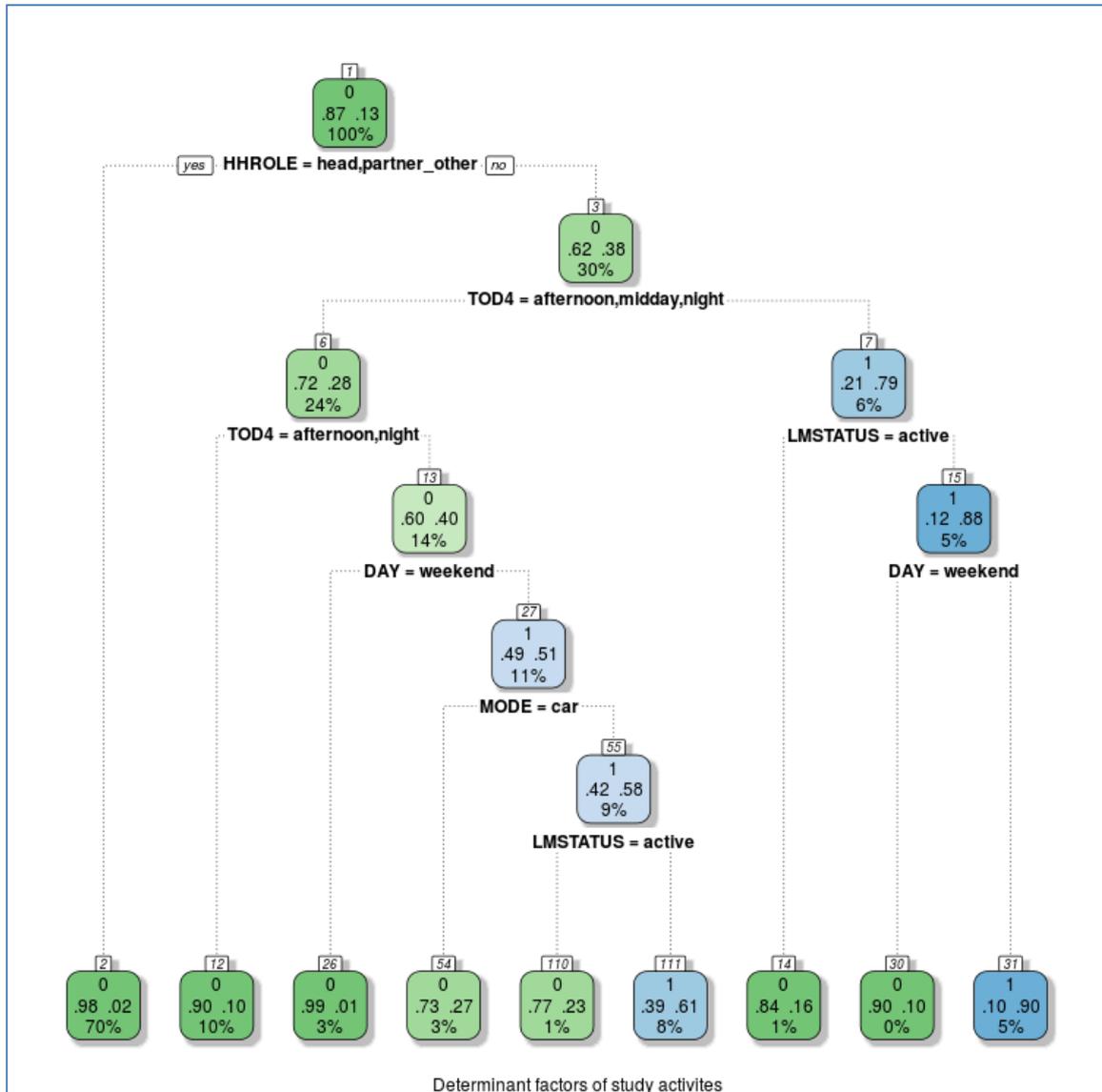


Figure 5: Decision tree for likelihood of studying

A last example, Figure 6, shows a decision-tree for work. Obviously non-active on the labour market have non-zero probability of working while the highest probability of work is in observations where departure time was in the morning, the respondent is active on the labour market and was not walking. Interestingly enough, we also see that observations with a woman as a respondent (conditional on time of day is afternoon, midday, night and day is weekday) results in lower probability of working. We furthermore see that an area variable matter, namely the WM (Dutch WoonMilieu). This is a categorical variable where small areas are categorised on a scale from 1 -5 based on density of residential addresses and services (1 represents the highest level of density). WM_VERTPC represents the WM at the origin of the trip while WM_HOME represents the WM at the location of the respondent.

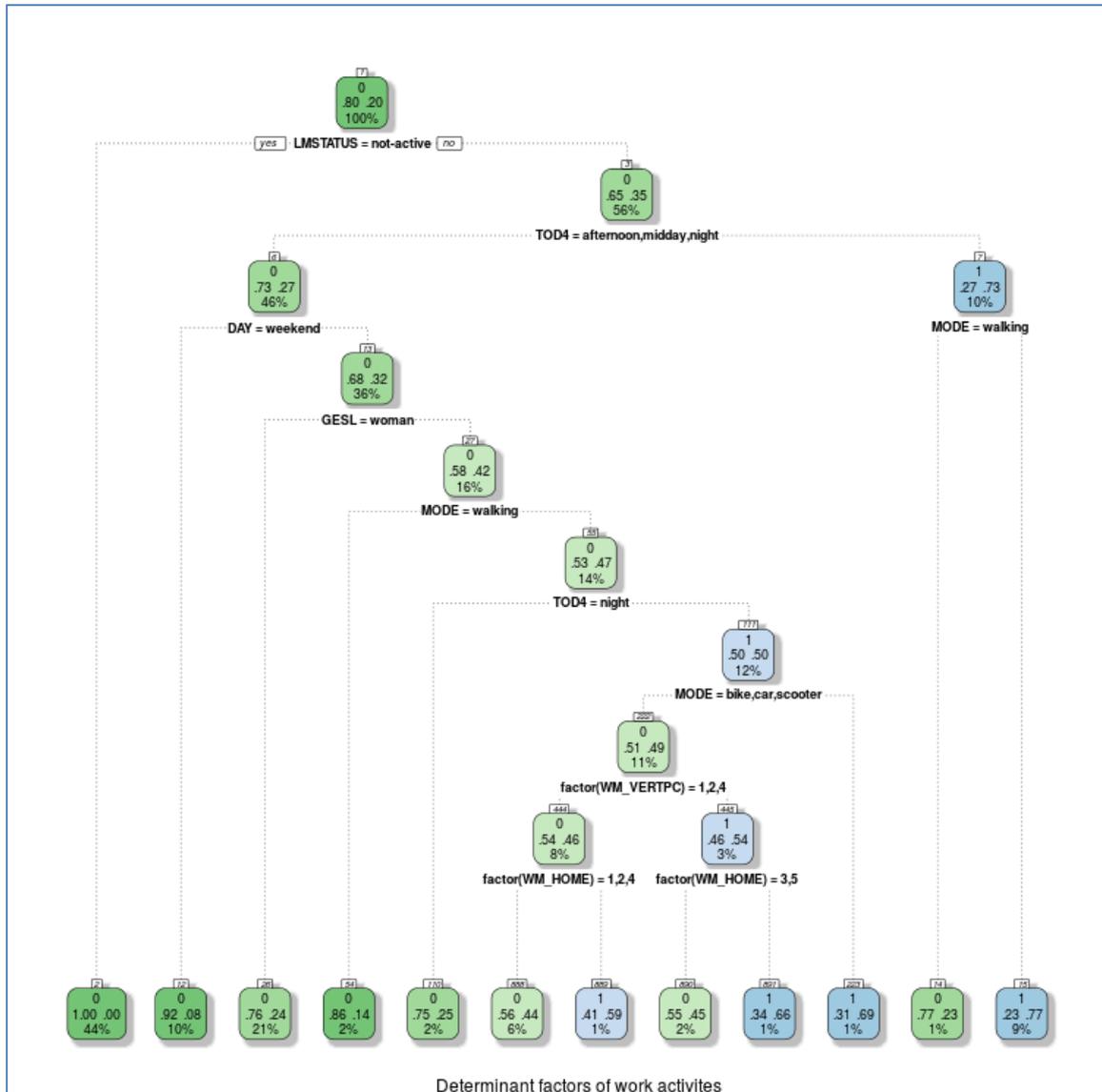


Figure 6: Decision tree for working

1.3 Spatial allocation of activities

The problem of spatial allocation of activities is formulated as a hidden Markov model and solved using a stochastic forward-backward algorithm. Hidden Markov models are used in situations where the actual sequence of states (the destination of a trip) is not directly observed. Rather, we can only observe some output generated by each state (the purpose for the trip, or the activity).

As an example of how an activity chain is constructed, consider the following example: we have the travel diary of three respondents who have reported three types of activities - namely home, work, and shopping - across two zones. To facilitate we can write this in the following manner:

$$a = h, w, s$$

$$z = 1, 2$$

Possible states in this setting are $S = h1, h2, w1, w2, s1, s2$, namely home in Zone 1 or 2, work in Zone 1 or 2, shopping in Zone 1 or 2.

Assume that a mobility survey gives us the following observations from respondents r1 through r3

$$O = h1w1_{r1}, w1s2_{r1}, s2h1_{r1}, h2s2_{r2}, s2w2_{r2}, w2h2_{r2}, h2w2_{r3}, w2s1_{r3}, s1h2_{r3}$$

For example, we know that respondent r1 lives in Zone 1, travelled to work in Zone 1, went shopping in Zone 2 and travelled back to his home in Zone 1. These observations can help us determine the sequence of activities: for example, the likelihood that an individual leaving home is going to work, or the transition probability $P(w | h)$, is $2/3$ (all respondents leave their home, but only one leaves directly for shopping before he/she goes to work). We can write the transition probabilities on a compact form as follows:

$P(ai aj)$	h	w	s
h	0	$2/3$	$1/3$
w	$1/3$	0	$2/3$
s	$2/3$	$1/3$	0

Furthermore, for each activity we also have the location of the destination, and we can determine the likelihood of each location. For example, only one respondent started a home- bound journey from Zone 1 and this respondent lives in zone two. Consequently, $P(1 | 1, h) = 0$ and $P(2 | 1, h) = 1$.

From respondents starting a home-bound journey in Zone 2, one lived in zone 1 and another in Zone 2. As such, $P(1 | 2, h) = 1/2$ and $P(2 | 2, h) = 1/2$.

$P(zi zj, h)$	1	2
1	0	1
2	$1/2$	$1/2$

Similarly, we can write the destination choices for work and shopping as

$P(zi zj, w)$	1	2
	1	0
	2	1
$P(zi zj, s)$	1	2
1	0	1
	$1/2$	$1/2$

The transition probability and the destination choice matrices contain the key parameters used to estimate the travel-activity behaviour of the synthetic population. Using knowledge of the transition probabilities between states and the probability of observing output in a certain state, we have generated a full origin-destination trip sequence conditional on the activity chain from step 1.

2 Temporal dimension of emissions

The Integrated module in the previous step, estimates energy demand annually. For air quality estimates, the seasonal and the intraday profile of the emissions are important to understand hotspots.

The proposed method consists of four phases: data collection, data pre-processing, data modelling and data clustering.



Figure 7: Block diagram for modelling emission load profiles

The nature of the datasets in the data-collection process may vary. For reasons of efficiency, we evaluate the input data with the aim of identifying the data that have a significant influence on our goal of generating emission load profiles for regions and cities. Our collected data are in the form of a panel dataset, that is, a cross-sectional data sample at a specific point in time. Therefore, in synthesizing the emission load profiles, we need related datasets dealing with temperature, a national gas supply, a national electricity supply and an emissions area. Figure 9 shows the generic block diagram, where the data input are:

- a. Monthly gas pattern
- b. Hourly local temperature
- c. Hourly national electricity load
- d. Share of fuel resources (%), especially wood emissions.

2.1 Data Pre-processing

Data pre-processing is a required step to support various data sources and formats that suitable to be used in the modelling phase. Depending on the existing data collection, a pre-processing process that consists of some tasks may require. It is due to the incompleteness, noise, ambiguity and inconsistency data in collection. Data corruption, missing values and outliers are the commonest problems in data processing [1].

In general, there four tasks of data pre-processing: Cleaning, transformation, integration and reduction [2–4]. It is depending on what input data we have, the requirement data format of the tool(s) we will use and the expected output. Cleaning task will fill in the missing values, smooth noise data, identify or remove outliers and resolve inconsistencies [3]. The data will be corrected by filling in the missing values into a data warehouse or correct the dataset by using specific a technique, for instance: interpolation [5]. While, transformation task covers normalisation and aggregation. It translates, converts and/or scales the data into the desired formats or units. Furthermore, integration task is utilising multiple databases, data cubes, or files. It also includes combining the various raw dataset into a single dataset [3,6]. Moreover, reduction task will reduce the volume and keep the same analytical

results. It includes removal the redundant records and variables [2,3,7]. The following table summarise the common problems and their solutions:

Table 3: Data pre-processing: problems and solutions [3,4]

Task	Problem/issue	Solution/Technique
Cleaning	Missing data	Ignore the record Determine and filling the missing value manually Use an expected value
	Noisy data	Binning methods Clustering Machine learning
	Inconsistent data	External reference Knowledge engineering tools
Transformation	Different format, scale or unit	Normalisation
		Aggregation
		Generalisation
Integration	Different standards among data sources	Combine data into a consistent database
Reduction	Complex analysis or infeasible	Reduce un-necessary observations, variables or values

We always start with the identification and description of the raw data we already have, which is the initial data. Then, we identify or define the desired data output. We can then make decisions regarding which data pre-processing technique, task or method will be applied. We call this approach the “*Indepth*” approach, covering successively initial data (raw or input data), desired data (the required output data) and process (technique, task or method).

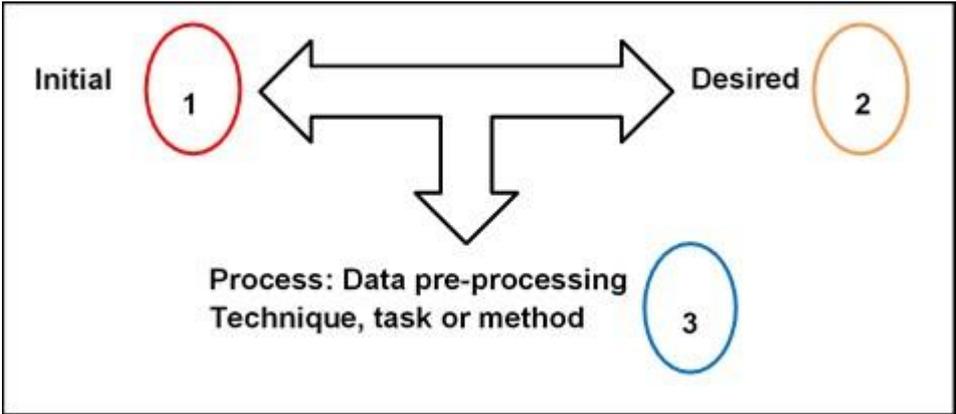


Figure 8: The Indepth approach, which starts by identifying the characteristics of the initial raw data, then defines the desired output. Based on these start and end points, the process can be computed by means of the selected technique, task or method.

2.2 Temporal Load profiles

In general, the regional emissions temporal load profile method is the same for all case cities.

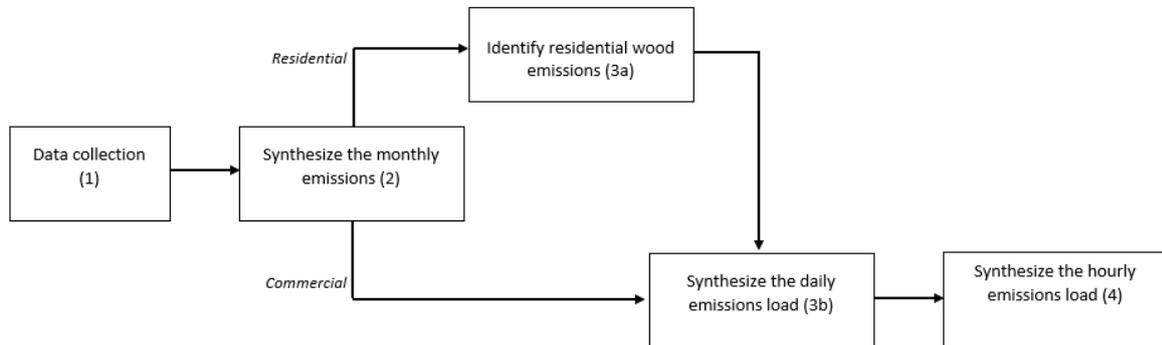


Figure 9: Block diagram for modelling emission load profiles

The steps of the generic method in the diagram in Figure 8 are the following:

1. Data collection of all required inputs: monthly gas pattern (Gas), hourly local temperature (Temp), hourly national electricity load (EI) and share of fuel resources (%), especially wood heaters.
Objective: to identify all related national or regional patterns.
2. Synthesize the monthly emissions load according to the monthly gas pattern.
Objective: to produce a monthly synthetic emissions load that has a similar monthly variation pattern to the real monthly gas pattern.
 - a. Residential: identify the residential wood emissions
Objective: to identify the share of residential wood emissions.
 - b. Commercial and residential: synthesize the daily emissions load according to the daily average temperature pattern.
3. The synthetic daily emissions are distributed from the synthetic total monthly emissions, the wood emissions only being considered for the winter period.
Objective: to produce the daily synthetic emissions load.
4. Synthesize the hourly emissions load according to the hourly electricity load. The synthetic hourly emissions are distributed from the synthetic total daily emissions.
Objective: to produce the hourly synthetic emissions load.

2.2.1 Residential model

We proposed a residential load profiles model by using the weighting proportion of some local parameters to reflect the local characteristic. We combined our model with the Pflugradt's model in Figure 10, which is applied in his tool, namely, the Load Profile Generator (LPG). Pflugradt's has developed the model with a strong focus on the behavioural aspect. The basic element to model a single household is the desire and expresses the need to do something. The model has specified: weight, threshold and decay time as desire properties [8]. Hence, our model is the determination of selecting some household profiles that are fitted with the weighting proportion.

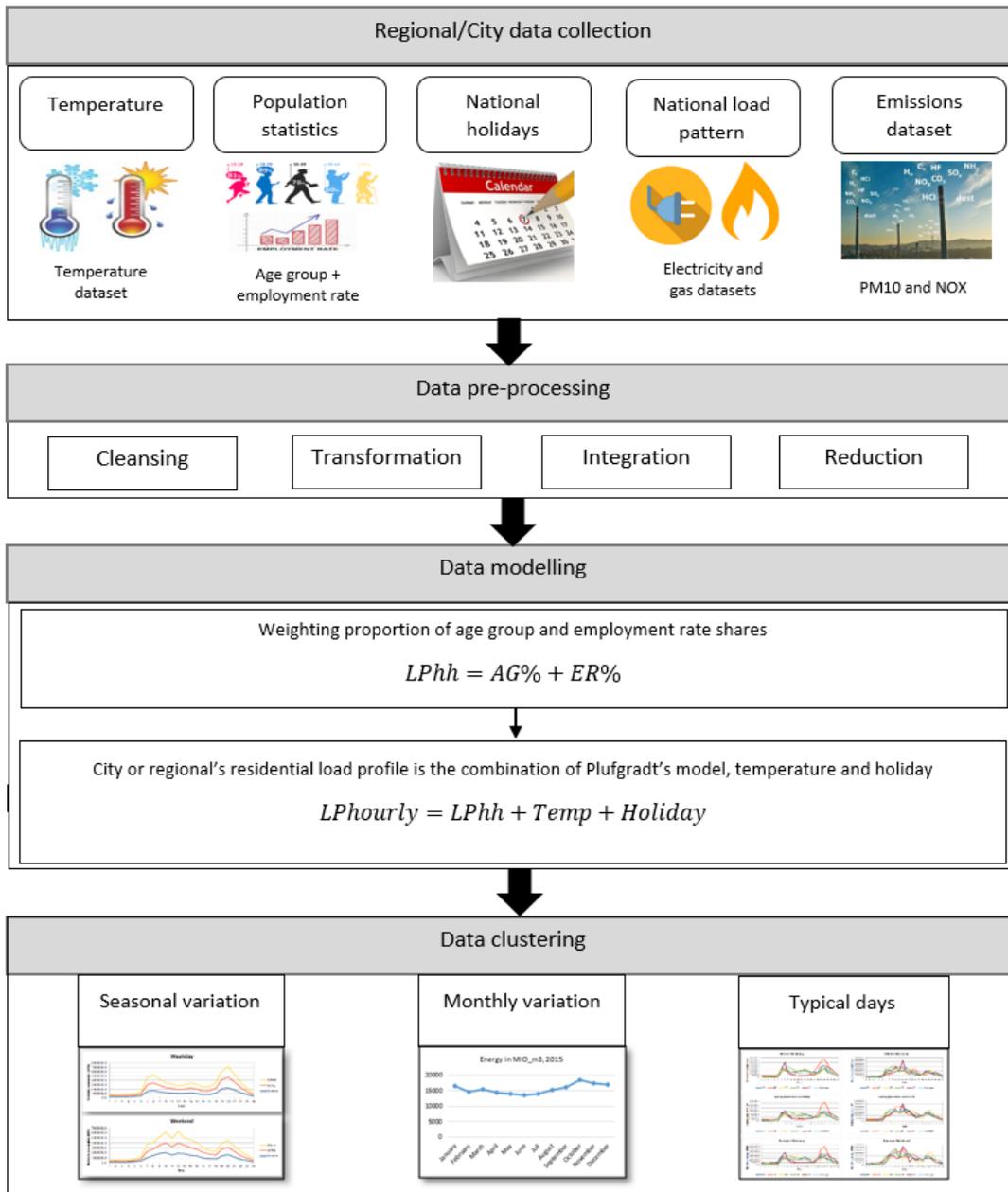


Figure 10: The framework for modelling residential's emission load profiles

2.2.2 Weighting proportion (Wepro) method

The Wepro model is an approach to model the residential load profile at the city level by emphasising on the weighting proportion of some main local parameters. First, it is required to identify the city's age groups (AG), gender (GD) information, and labour force (LF) information. In this case, the total of annual energy consumption is not required, since we only focus on providing the share of hourly energy load profiles. Second, we coupled the share of age groups and labour force, and apply this weighting proportion to the total population, as the main characteristics, which represent the city's profile. In more detail, each age group has gender information. Although, we can also identify the

gender information at the higher level of the age groups, which is the total of each gender in the city. In this model, the more detail of gender information of each age group is required.

As mentioned, our model also employed Pflugradt’s model and tool, where the early step is to select the household profiles to be modelled by LPG. The fundamental consideration is the selected household profiles in LPG should represent the city characteristic in term of age groups and labour force, as the focus of our study. Therefore, the weighting proportion of the age groups share and employment share are the fundamental input of the modelling. Consequently, the household load profiles is the combination proportion of age groups share and employment share. The weighting proportion of age groups, gender and labour force is depicted in the picture below.

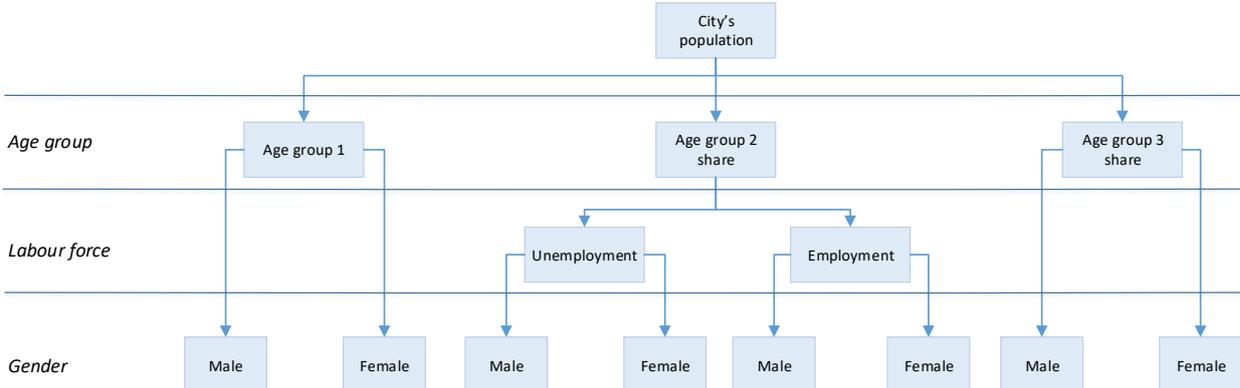


Figure 11: The Weighting proportion structure of residential sector at the city level

2.3 Implementation

2.3.1 Temperature

As an example, we elaborate on the procedure for Bristol. For this example we obtained a temperature dataset from the UK’s Meteorological Office, a national public weather service that provides weather information, including severe weather warnings. The dataset is available from the Filton station in hourly resolutions as an .xlsx file. The dataset has three variables: “date”, “hour” and “hourly temperature”. The dates are those for 2015, covering the 365 days from 1 January 2015 to 31 December 2015. The hourly variable shows the hour of each date from 0 to 23, representing the 24 hours in a day. Finally, the hourly temperature variable provides the value of hourly temperatures in Celsius to one decimal place. In this dataset, we only transformed the data from hourly values into daily average values in Python. The hourly data within the same day is determined in order to find the daily average.

The rest of the datasets applied the same procedure to all case cities: national gas consumption patterns, national electricity load and emission areas. Therefore, the results of the data pre-processing of these datasets of other case cities can refer to the following description:

2.3.2 National gas consumption patterns

Many resources are available providing national gas consumption patterns, especially annual gas consumption data. In this case, we need the higher resolution data: the higher the resolution the better, since we want to synthesize the emission load profiles that reflect local characteristics. Thus, it is essential to determine the desired output of these data. In line with the emission load profiles, we

then decided to collect data on the monthly national pattern by type of gas unit and each unit's monthly share as a percentage for 2015, both to be shown in a spreadsheet.

In all six cases, we retrieved the national gas consumption pattern from the Eurostat energy database [45]. The files are available with tab-delimited (.tsv) file extensions. We then extracted the file and open it in a spreadsheet. To visualize the monthly gas consumption profile for the selected year, it is more efficient to do filtering than reduction since the file contains all EU member states' datasets for the ten years from 2008 to 2018. In total the file has 129 variables and 815 columns. The variables consist of "unit,product,indic_nrg,geo", while the rest use the time in monthly periods from "2018M08" back to "2008M01". Since we have selected 2015 as the reference year, we selected "2015M01" to "2015M12". For each country selected, we choose all the observations marked with the country's ID, given that each country's ID is unique. For instance in the case of Poland we selected all the observations marked "PL", and similarly IT for Italy, NL for the Netherlands, etc.

In the following we will present just one example, Portugal, as the other selected cases follow the same procedure. After choosing an example in the selected monthly period, we then categorized the selection based on the type of gas unit, of which there are two: "MIO_M3" in million cubic metres (Figure 12); and "TJ_GCV" in Terajoules on the basis of gross calorific value.

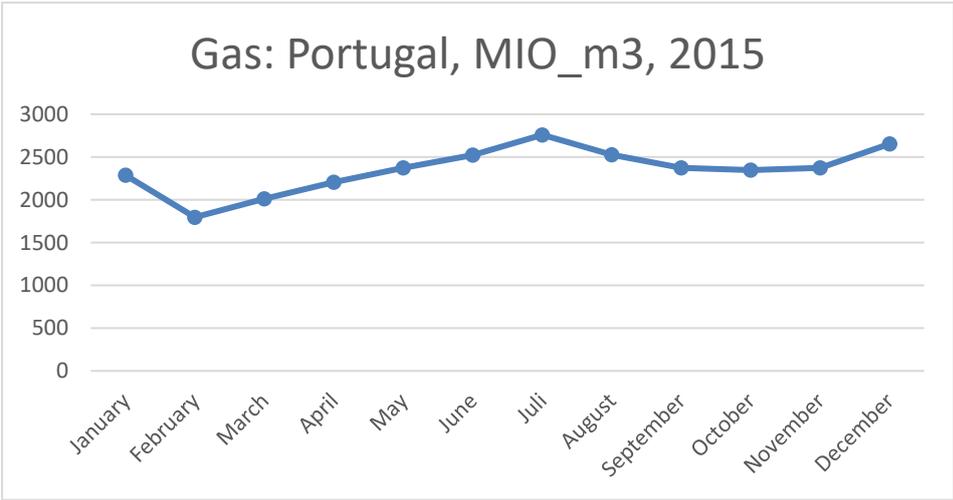


Figure 12: Portugal's monthly gas pattern, limited by unit: million cubic metres in 2015 (authors' calculation from [45])

Beside visualizing the monthly gas consumption pattern, we also identified gas consumption as a monthly share, with each month being a percentage of the annual consumption. This is the final step before synthesizing the emission load profile.

Table 2: Portugal's monthly gas pattern and share (%) in 2015 (authors' calculation from [45]).

Month in 2015	Unit (MIO_M3)	Gas monthly share (%)
January	2286	8.1
February	1796	6.4

March	2012	7.1
April	2205	7.8
May	2372	8.4
June	2523	8.9
July	2759	9.8
August	2524	8.9
September	2371	8.4
October	2346	8.3
November	2372	8.4
December	2652	9.4

Benchmarking to other sources that provide monthly gas patterns, such as data from the Joint Organisation Data Initiative (JODI) gas database [46] and a related gas study [47], is required to achieve a better synthesized result.

2.3.3 National electricity consumption load

The national electricity load datasets for all six case studies are available as .csv files from open power system data [48]. These were used as the load patterns for the commercial and industrial sectors. The dataset has four variables (Table 3). Here the desired output is the share of the hourly load resolution in percentages, to be entered in a spreadsheet.

Table 3: National power loads: various names and descriptions [48].

Field name	Type (format)	Description
utc_timestamp	Datetime	Start of time period in Coordinated Universal Time.

cet_cest_timestamp	Datetime	Start of time period in Central European (Summer) Time.
Interpolated_values	String	Marker for missing data column in source data, which has been interpolated.
United Kingdom: UK_load	Number	Total load in United Kingdom in MW
The Netherlands: NL_load		Total load in The Netherlands in MW
Poland: PL_load		Total load in Poland in MW
Slovenia: SI_load		Total load in Slovenia in MW
Portugal: PT_load		Total load in Portugal in MW
Italy: IT_load		Total load in Italy in MW

The source for load values was retrieved from the ENTSO-E data portal and power statistics. The European Network of Transmission System Operators (ENTSO-E) represents 43 electricity Transmission System Operators (TSOs) from 36 countries across Europe [49]. The desired output of this dataset is the share of hourly resolutions per country in percentages. This provides us with 8760 observations, which have values in percentages. In these datasets, the missing data have been solved by means of the interpolation technique from the data source.

Reduction must be performed by reducing the Coordinated Universal Time variable and keeping the time stamp of the Central European Time (CET) variable. We also reduced the interpolated_values field for reasons of efficiency. Therefore, the new dataset consists of two fields: CET time stamp and total load in MW; and 8760 records representing the hourly resolutions. Finally, we performed the transformation by scaling the share of the hourly resolution in percentages.

2.3.4 Residential emissions area

The emission area datasets for all six case studies are available as .csv files supplied by our ClairCity partner. The dataset for each selected example consists of eight variables: “year”, “city”, “zone”, “codvariable”, “namevariable”, “pollutant”, “emissions” and “unit”, together with hundreds of observations. Here the desired output identifies the total of emission values by sector and their share by percentage, per pollutant category, to be entered in a spreadsheet. First, we can reduce some unnecessary variables:

1. “year” is unnecessary because we have already defined the reference year as 2015.
2. “city” is unnecessary because we have also defined Sosnowiec as the sample city.
3. “zone” and “codvariable” are unnecessary at this moment, but will return when we want to make further spatial analyses.
4. “unit” is unnecessary because we have agreed to used Mg as the emissions unit, and we only need to scale the values of emissions in percentages.

The reduction phase is optional here. If we reduce the unnecessary variables, we can keep the “namevariable”, “pollutant”, and “emissions” variables. “namevariable” indicates whether the sector is residential or commercial, as well as the combustion plant and its fuel type. Nonetheless reduction is recommended for reasons of file size or storage efficiency. In fact we did not reduce the unnecessary variables in this case, since we can select the data directly in Python.

The “pollutant” variable consists of two categories: PM10 and NOX. The “emissions” variable gives the value of each pollutant in the energy sector area. Furthermore, we selected the observations based on the energy-sector categories in “namevariable”: residential or commercial. Lastly, we quantified the total emissions values in the residential and commercial sectors by category of pollutant: NOX or PM10. The code is executed in Python, where we determine the emission value per pollutant in each sector (Figure below).

```
## Select residential

query = 'SELECT zone, codvariable, sum(emission) FROM {} WHERE namevariable
LIKE"Residential%".format(pstable)

c.execute(query)

pm10 = c.fetchall()

dfp = pd.DataFrame(pm10)

dfp.head()

## Select commercial

query = 'SELECT zone, codvariable, sum(emission) FROM {} WHERE namevariable
LIKE"Commercial%".format(pstable)

c.execute(query)

nox = c.fetchall()

dfn = pd.DataFrame(nox)

dfn.head()
```

Figure 13: The selection coding of the emission values by pollutant: PM10 or NOX, and by sector: residential or commercial, in Python

Furthermore, we ran a program in Python to determine the emissions per pollutant of residential fireplaces (Figure below). This is important in order to specify the local characteristics to be considered when modelling the emission load profiles.

```

## Select residential fireplaces of PM10

query = 'SELECT zone, codvariable, sum(emission) FROM {} WHERE pollutant LIKE"pm10" and namevariable
LIKE"Residential Fireplaces%".format(ptable)

c.execute(query)

pm10 = c.fetchall()

dfp = pd.DataFrame(pm10)

dfp.head()

## Select residential fireplaces of NOX

query = 'SELECT zone, codvariable, sum(emission) FROM {} WHERE pollutant LIKE"nox" and namevariable
LIKE"Residential Fireplaces%".format(ptable)

c.execute(query)

nox = c.fetchall()

```

Figure 14: The selection coding of residential fireplace values per pollutant: PM10 or NOX, in the residential sector

Table 4: Summary of regional emission values by sector (authors’ calculations from emission area dataset)

Sector	PM 10		NOX	
	Emissions (Mg)	%	Emissions (Mg)	%
Residential	1802,610	100,000	2511,290	100,000
Residential fireplaces	0,526	0,029	0,031	0,001
Commercial	279,326	100,000	608,955	100,000

The descriptions of the emissions data of the residential and commercial sectors were presented and scaled in percentages based on each pollutant. In modelling our emissions load profiles, we assumed that residential fireplaces are used according to the local energy consumption characteristics.

Table below summarizes all the pre-processing techniques that we applied to each of our datasets by case study.

Table 5: Summary of data pre-processing techniques applied to each dataset by case study

Data collection per city or region	Property of	Nature of data collected	Pre-processing techniques	Tool or computer application
Temperature				
Bristol	Met Office UK https://www.metoffice.gov.uk/	Hourly resolutions	Transformation	Programming language: Python
Amsterdam	Meteorology office: KNMI https://data.knmi.nl/datasets	Daily resolutions	Reduction Transformation	Spreadsheet Spreadsheet
Ljubljana and Sosnowiec	Commercial weather services https://www.wunderground.com/ by IBM	Thirty-minute resolutions	Reduction Integration Transformation Cleaning	Spreadsheet Command prompt Programming language: Python Programming language: Python
Aveiro	ClairCity partner, shared by UAVR, Portugal	Ten-minute resolutions	Transformation	Programming language: Python
Liguria	Local link shared by ClairCity partner, Techne, Italy	Hourly resolutions	Reduction Transformation	Spreadsheet Programming language: Python
Gas				
Bristol Amsterdam Ljubljana	Eurostat database	Monthly resolutions	Extraction Filtering Selection	Spreadsheet Spreadsheet Spreadsheet

Sosnowiec					Transformation	Spreadsheet
Aveiro						
Liguria						

Electricity

Bristol	European open data platform	Hourly	Reduction	Spreadsheet
Amsterdam	ENTSO-E	resolutions	Transformation	Spreadsheet
Ljubljana				
Sosnowiec				
Aveiro				
Liguria				

Emission

area

Bristol	ClairCity partner: Techne	Area	Selection	Programming language:
Amsterdam		dataset		Python
Ljubljana				Spreadsheet
Sosnowiec			Transformation	
Aveiro				
Liguria				

3 Transport

The objective of the transport module, is to estimate the emission of transport, primarily road transport at link level and enable the assessment of policy measures.

3.1 Transport demand

We use a distinct different approach for Bristol & Amsterdam vs. Aveiro, Ljubljana and Liguria. In the case of the former, we had access to results of a traffic model, generating transport demand on a network as a starting point to estimate traffic emissions.

For the latter 3, such a model was missing. we developed a simplified approach to estimate road transport emissions at link level for a selected area, including smaller cities. We use OpenStreetMap (OSM) data for road network topology and production/attraction to generate transport demand and assignment to a network. The approach includes a method to calibrate and rescale resulting transport volumes with a limited amount of traffic counting data points. Post-processing includes an intra-day temporal allocation.

The approach involves the following steps:

1. Based on OpenStreetMaps (OSM) we generate a connected graph that represents the transportation network. OSM allows all roads, including tertiary roads, to be included in a directed graph. The level of detail can be adjusted to the needs of the air quality model. The technique we used for ClairCity went up to tertiary road level. This is certainly possible at the level of a city, but it is possibly a challenge for a large territory (Flanders/Belgium), which can be met with zoning/segmentation.
2. Next, based on land-use data from OSM, we estimate the demand for transport (HB matrix) - we can also take the demand generation of the dynamic traffic model for Antwerp (of secondary importance). In this project, we used the OD-matrix from the Flemish strategical transport model V4.
3. a simple routing algorithm assigns the transport demand to the network
4. We fine-tune the traffic volumes by calibration with a (limited) number of measurement points, if available.

3.1.1 Step 1: Network generation with OSM

OSMnx is a Python package that lets you download spatial geometries and model, project, visualize, and analyze street networks from OpenStreetMap's APIs (<https://github.com/gboeing/osmnx>).

The procedure to extract road network and construct a connected graph for the road network is presented below. After running the program, the network is ready for calculating routes between locations.

OSMnx allows to export data for two shapefiles. One with road data & one with node data. QGIS an open-source GIS visualization platform is used to first import the shapefiles and next to export 2 .csv files called "edges.csv" and "nodes.csv" (EPSG:4326).

The result is a noded network, example for Ljubljana is shown below:

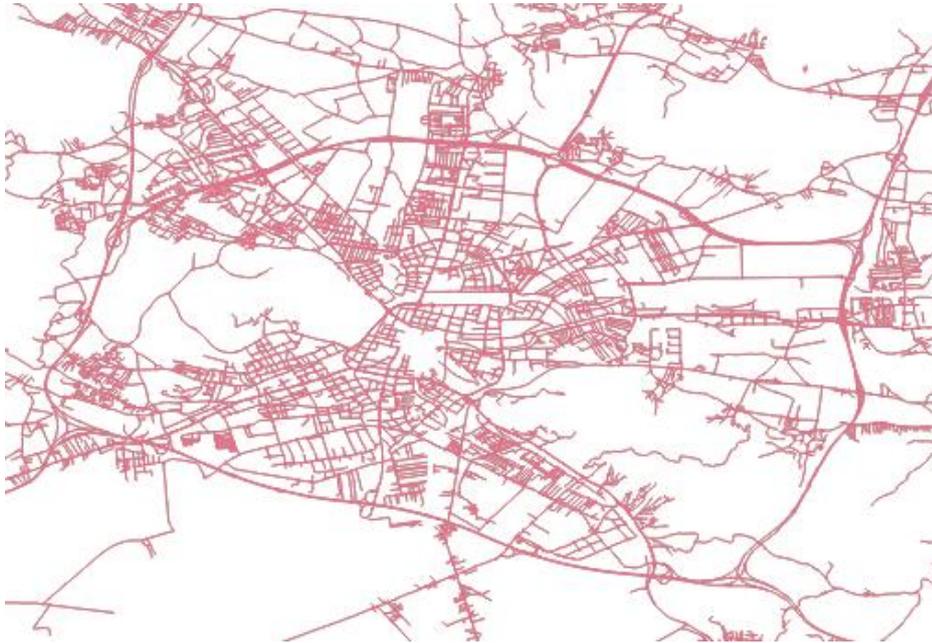


Figure 13: Example of resulting road network for Ljubljana

3.1.2 Step 2: Generation of transport demand - production/attraction

Classical transport models are composed of 4 connected and integrated stages:

- Production & Attraction
- Distribution
- Mode choice
- Assignment

Only the first two stages are important for demand generation.

Production factors define the generation of demand for a zone. The factors feed into a function that describes the total amount of trips being generated in a zone. In most cases the trip generation function is a multi-variable regression model based on socio-economic variables such as population density, age distribution, income levels, etc...

The attractiveness of a zone as a trip end is mostly defined by infrastructural/spatial characteristics. The total amount of trips that dissipate in a zone is also described by a multi-variable regression model. It is based on number of available workplaces, schools, quantity and quality of shopping locations, availability of leisure activities, etc...

In the second stage total amount of trips generated and dissipated for all zones are combined into a trip matrix connecting each production zones with attraction zones. The relation between zones described in function of the distance or travel times between zones. A gravitational weight function is used to describe decreasing attractiveness of distant zones.

The first two stages lead to a very rough estimation of the number of trips in a network. Further calibration of the trip matrix is required to match the trip matrix with actual demand patterns. This typically based on observations such as link volumes or aggregated trip statistics.

The traditional approach introduces artificial links called connectors to link the demand of a zone to the network. Each zone is represented by a centroid at the center of its shape and a handful of connectors are drawn between this center point and the import roads of the zone.



figure 14: zone centroid connection approach

3.1.3 Step 3: Assignment

The main idea of assigning demand to the network is based on equilibrium principles. These state that drivers will keep on looking for shorter routes until all drivers unilateral perceive the least resistance.

Resistance is defined as a generalized cost which is the sum of link specific cost functions. The cost function used is a BPR-curve (described by the US Bureau of Public Roads) that is defined in function of free flow travel time, capacity and link flow.

$$Cost = t_0 \left(1 + \alpha \left(\frac{Flow}{Capacity} \right)^\beta \right)$$

Algorithmically the equilibrium process translates into an iterative process where drivers are repeatedly loaded onto the network until no driver unilaterally can switches routes towards a less costly path. For efficiency reasons drivers originating from the same source are loaded according to similar routes. This allows for an implicit formulation of the route tree rooted in each origin. At every node in the network only the set of efficient outbound links per origin are required. These are the links for which the starting point is topologically closer to the origin than the end point. The efficient links are computed based on a Dijkstra shortest path tree evaluated in free flow conditions.

The distribution of traffic flows over the different routes is performed with Dail's procedure. The procedure assigns demand stochastically to different routes such that it coincides with a logit process. The algorithm avoids explicit path enumeration to save memory.

The assignment is for a full average day and capacities for the roads are adjusted accordingly. It is assumed that the maximum hourly road capacity is increased to a full day and that this factor is a parameter to control for responsiveness of drivers with respect to busy roads. The factor is set to 10 which introduces mild responsiveness and a quick convergence of the algorithm. The implications are that for most OD-pairs only a small set of paths is considered centered around the shortest path.

The stochastic distribution of flows over the different route alternatives is determined by a parameter which coincides with the mu-parameter of logit models. Mu is set to 20 in this case. The response amounts to a reduction of almost 70% on a route if 5 additional minutes are experienced.

The demand is distributed per zone according to the following rules:

- 200 nodes are selected within a zone that is part of the case-study area
- Zones that are close to the case-study are mapped to the nearest road of the lowest OSM class (residential, road, living_street or unclassified) that cuts the cordon around the case-study area
- Zones that are further away from the case-study are mapped to the nearest road of all higher OSM classes that cut the cordon around the case-study area

Zones that are very far away from the case-study are mapped to the nearest road of the highest OSM classes (trunk, trunk_link, motorway, motorway_link) that cut the cordon around the case-study area

3.1.4 Step 4: Refinement

The resulting initial assignment to the network, ignore any network specific characteristics insofar these are not captured in the assignment algorithm. For example, disutility for specific links due to road characteristics are not accounted for and will lead to an overly optimistic attraction of a road. This may lead to serious deviation of modeled and real-world traffic volumes.

We assume the initial loading will not lead to results of sufficient quality and therefore perform corrections to the OD-matrix using counting data on key highway's.

This differs case by case, depending on data availability. An estimated selected link assignment is performed based on the shortest path trees emitted upstream and downstream for a selection of key roads (typically 10-20 highway location per case). These trees are combined with the observed splitting rate at diverge/merge points to adjust redistribution weights of the deviation with the counts.

The results of these steps, is an estimation of daily traffic volumes on the road network. We tested the approach for Genoa and simulated, as a test case, how the assignment algorithm behaves after collapse of the Morandi bridge:

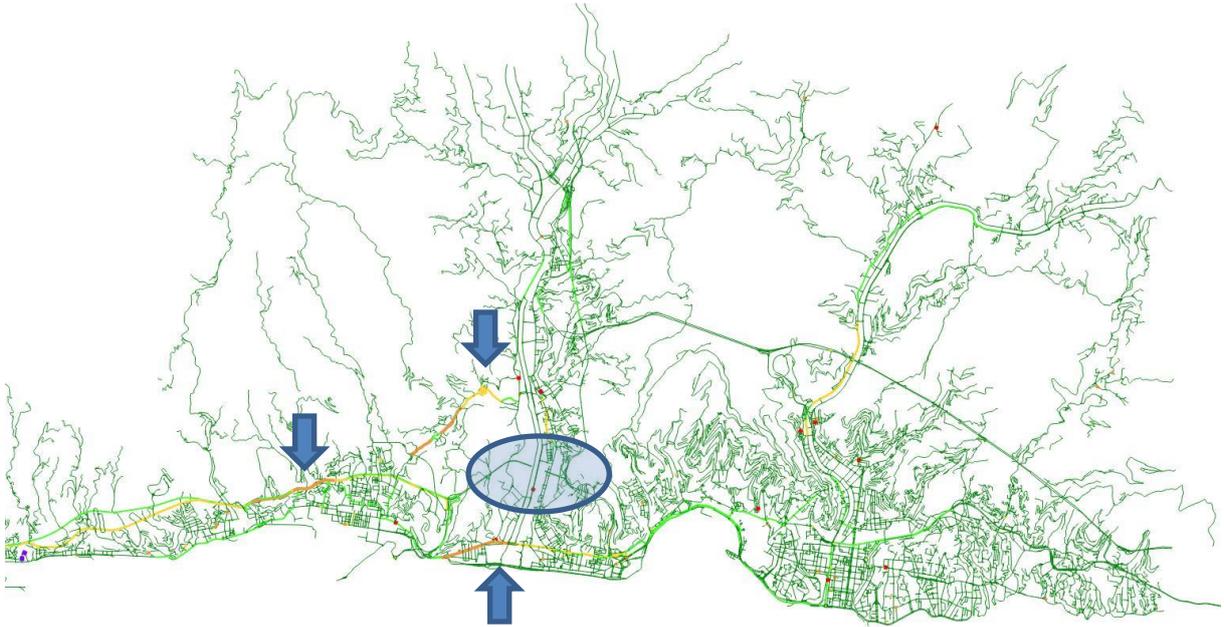


Figure 15: Typical outcome traffic estimation - Results of traffic assignment test for Genoa, in case without (top) and with (below) the Morandi bridge available.

As volumes are estimated for daily totals, a final step is needed to distribute intensity by time of day. This is fairly trivial and can be done using various data that is specific for the local situation. In table and figure below, the estimates we've used, based on observed highway traffic intensity (a good proxy for all roads), making a distinction between weekday and weekend. Note that the sum over all hours is 1 for weekday, but lower for weekend, as traffic generation an assignment is assumed for a weekday with typical peak-profiles.

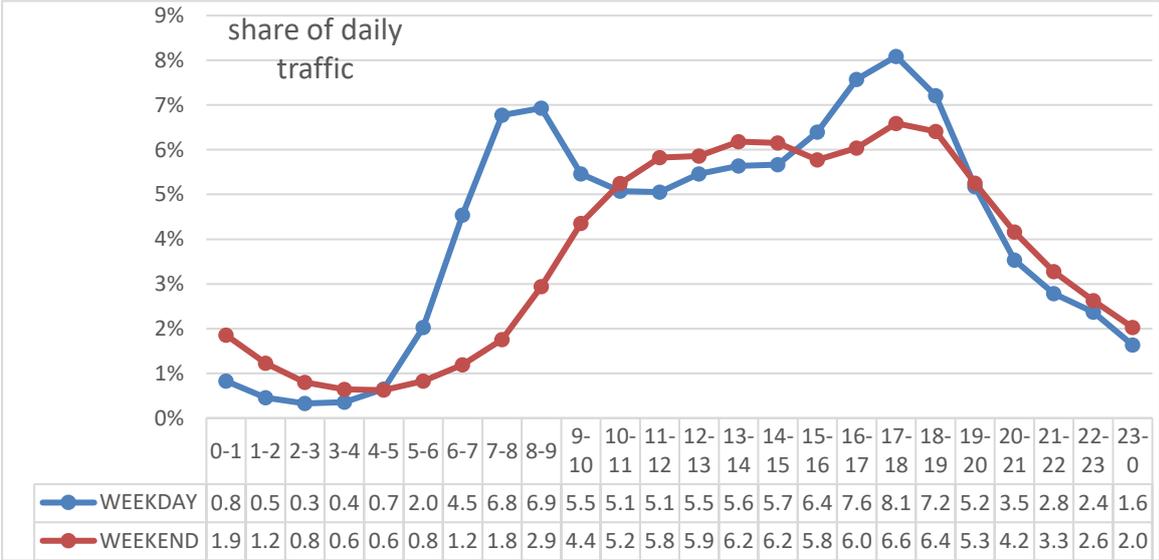


Figure 16: percentage of daily traffic by time of day (h), weekend (WE) vs. weekday (WD)

Finally, estimated traffic is split into modes, using city-specific information on modal split.

3.2 Emission factors

We use COPERT V¹ emission factors, a general methodology to estimate emission factors of all road transport modes. COPERT includes distinction between emission factors by vehicle type, fuel type, age, size,...

Table below summarizes the emission factors, linked to the year of built (YOB) of the car, fuel and size (H = engine size >2l)

¹ <https://www.emisia.com/utilities/copert/documentation/>

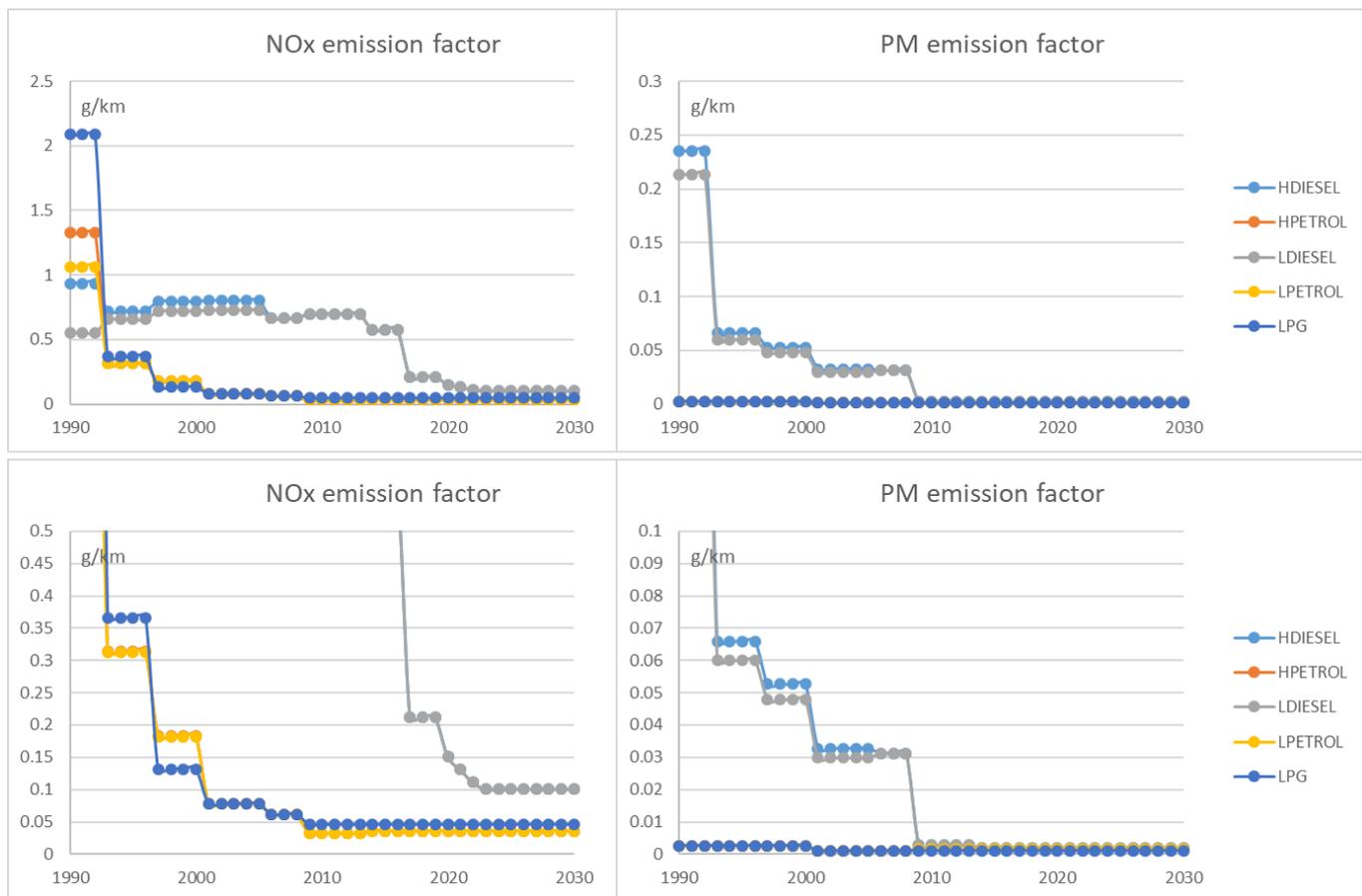


Figure 17: COPERT V NOx and PM emission factor of selected cars (diesel/petrol/LPG)

Non-exhaust emissions are assumed to remain constant over time. Emission standards for exhaust emissions don't apply to non-exhaust emissions. Table below summarizes the assumption on the non-exhaust particulate emissions.

Table 4: non-exhaust emission factor (in g/km) for road transport modes

	PM10	PM2.5	PM2.5/PM10 ratio
Car	0.0211983	0.01113141	52.51%
Motorcycle	0.00989182	0.00509529	51.51%
Van	0.03324282	0.01748643	52.60%
Medium truck	0.072334743	0.034446514	47.62%
Heavy truck	0.081704733	0.041005507	50.19%

Bus	0.072334743	0.0344446514	47.62%
------------	-------------	--------------	--------

Energy consumption is also linked to year of built and in line with a the progressive reduction in CO2 emission a standers tighten. We introduced the concept of biofuels explicitly in the emission model. Their share within a given energy carrier a given region is set by a time dependent parameter, meaning that we can set for each fuel type, year, and region what part of the fuel mixture is bio (while the rest is of fossil origin). The carbon content of the biofuels is assumed to be zero. This entail some specific changes to the existing emission model:

1. The energy content of biofuels was added separately, and the total energy content of the fuel mixture is calculated as a weighted average of fossil- and biofuels according to their shares.
2. Cars running on biofuel have a higher consumption (since the energy content of biodiesel and ethanol is less than of their fossil counterparts), this is taken into account by assuming that the total energy demand should still be the same, therefore cars running on fuel with lower energy content will have proportionally higher fuel consumption values assigned.

We separated tailpipe and well-to-tank CO2 emissions, which are now calculated separately.

The emission factors are dependent on the year of built. For emission per year and understanding evolution of emissions over time, we use a model that includes fleet dynamics.

3.3 Fleet dynamics

We opt for a fully quantitative assessment of a variety of city specific scenarios in the passenger car sub-sector, to assess the impact of various technology uptake rates and scrappage schemes in the fleet on the overall emissions.

To this end, we use the fleet module of the TML-owned model MOVEET, including an extensive update of the fleet-data and some modifications on the fleet dynamics side. MOVEET is an updated version of the TREMOVE model, developed by TML in the period 2015-2017 and has been aligned with the latest PRIMES-TREMOVE scenarios in terms of future transport demand, vehicle efficiency and biofuel uptake.

We use the fleet module of MOVEET exclusively for the scenarios and introduce an exogenous propulsion technology share as scenario input. This means we don't model the expected uptake of technology (based on costs, consumer preference,...) but arbitrary sale shares. We model the scenarios in such a way that we create credible futures in terms of propulsion technology uptake and realistic policy specific scrappage schemes.

In this chapter, we go in detail into the MOVEET model itself, the data-updates and modifications to the model functionality in this project.

3.3.1 The MOVEET model

MOVEET is a System Dynamic based analytical tool that addresses the policy problems related to transport and environmental impact. The tool is capable of estimating transport demand and emissions, as well as forecasting the impacts of policy and technological measures in transport-related sectors, covering all transport modes from the different regions in the world up to 2050. The model divides the world into 57 regions, many of them representing single countries, i.e. all the European countries and other world major economies. In the model, we consider all transportation modes (road, air, rail, and maritime) that interact through four interrelated modules: Transport Demand, Fleet, Environmental, and Welfare.

The MOVEET model has been developed at Transport & Mobility Leuven. MOVEET is a System Dynamic based analytical tool that addresses the policy problems related to transport and climate change. The tool is capable of estimating transport demand and emissions, as well as forecasting the impacts of policy and technological measures in transport-related sectors, covering all transport modes from the different regions in the world up to 2050. In the model, we consider all transportation modes (road, air, rail, and maritime) that interact through four interrelated modules: Transport Demand, Fleet, Environmental, and Welfare.

It is the first transport model that:

- generates and project demand endogenously for the whole world regions,
- splits this demand into the different most intensive energy consuming modes of transport,
- makes use of the existing fleet data and project the world fleet dynamic in high level of detail into the future, and
- produces global impact assessment in term of emission and welfare.

MOVEET's four modules exchange information in order to provide a consistent picture of the different aspects modelled. Within the transport demand module, motorized transport demand is endogenously generated and segmented according to several dimensions (e.g. national/international, long or short distances, etc.). In addition, the choice of mode and road type for each specific context is carried out taking into account demand-supply interaction.

Transport demand by mode is then used as input for calculating vehicle-kilometres by type and technology according to the fleet structure estimated in the fleet module. In the environmental module, fuel consumption and emissions are calculated on the basis of vehicle-kilometres (from the fleet module) as well as average speed of each transport mode (from the demand module). Finally, the welfare module takes its input from both the transport module and the environmental module in terms of consumer utility and, respectively, external costs.

The fleet module receives the following inputs from the demand module: passengers-kilometres (pkm) distinguished by zone, purpose, region, distance, urban level, time period, mode and network, tons-kilometres (tkm) distinguished by zone, car type, region, distance, urban level, time period, mode and network, and average load factor by demand segment.

The main goal of the fleet module is to convert aggregate estimations of transport demand, in terms of pkm, tkm and/or vkm, into a more detailed vehicle classification and generation (cohort) which directly relates to technology in terms of vehicle performance and characteristics, fuel use and emissions.

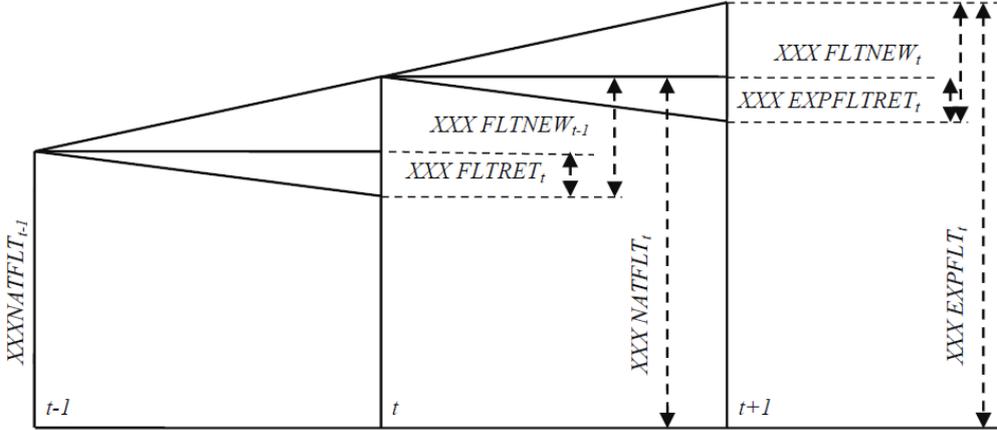


figure 18: fleet capacity planning mechanism in MOVEET fleet module

We have achieved alignment of MOVEET baseline projections on the common PRIMES-TREMOVE 2015 baseline and added changes to fuel efficiency improvements in line with recent policy developments.

In sections below we highlight the modifications done to the MOVEET model for this project. While globally we made these improvements only after simulating Bristol and Amsterdam with the original version, we have mimicked the effect of these changes in these two cases already by some specific modifications, leading to the same results (and motivating us to actually implement these modifications in a general way into our modelling tools).

One of the most significant changes in the model concerns the **fleet calculations**, and it originates from three sources:

A) In the original MOVEET model, the mileage gap created at the end of year t (as old cars are scrapped from the fleet) is filled with new (age = 0), and only new cars in year t+1. This is a standard assumption in fleet models, but to handle the particularities of the local markets more precisely, namely to include the effects of used-car import (which is many countries is a significant source of fleet-growth) and retrofitting of e.g. LPG cars (meaning that LPG cars might enter the market at an age that is not zero), it is worth moving for a more advanced description.

Therefore in this model we assume that the fleet can also grow by imported used cars (cars of age > 0), and as such the demand gap is not filled only with new cars, but with a set of cars that are distributed over a vintage range that is specific to that region. In practice both the used car dynamics and retrofitting can be addressed by introducing a growth function that describes the age distribution of cars that enter the market in a given year to fill the

mileage gap. (It can be best thought of as the inverse logic of the scrappage/survival function².) This growth function is region-specific and propulsion-type specific.

It is region specific because in regions where second hand import is minimal, and in regions where second hand imports make up a significant percentage of sales, the age distribution of cars that are introduced to the market each year will be very different. For example, in a region where second hand import is minimal all cars that are sold will be new cars (so the growth function is ~ 1 for age = 0 and ~ 0 for ages > 0), while in regions where the second hand market dominates a large amount of cars sold each year will be older (so the growth function is a small number at age = 0, as it takes the shape of a Gaussian-distribution which is centred over the dominant used car age). We define the growth function so its integral over the whole age range is 1.

The propulsion-type dependence of the growth function comes into play when we describe the dynamics, for example, of the LPG fleet, as the age distribution of LPG cars entering the fleet will most of the times be different than the age distribution of other types. This originates from the fact that a large percentage of LPG cars are conversions, therefore they are born (introduced to the market) with a non-zero age, therefore it would actually be impossible to describe the evolution of the LPG fleet with the classical approach where only age = 0 cars can be sold. In theory, all propulsion types could or should have their specific growth functions (as for example the ratio of second hand electric car import at this point is definitely going to be lower than of second hand petrol car import), but due to a lack of real-world data that could be used for proper calibration, we only differentiate between growth functions for LPG cars and the rest.

We note that due to the structure of the fleet dynamics calculations, we model conversions indirectly: in real life, when, e.g., an age = 3 petrol car is converted to LPG, an age = 3 LPG car is added to the fleet and an age = 3 petrol car is scrapped at the same time. In our model we assume an influx of LPG cars distributed over a range of ages (which is governed by the growth function), but we do not directly model the scrapped cars that are connected to this influx, instead, the excess scrappage is included in the general scrappage rate of petrol cars.

Both the growth and survival functions that were used in our model were built and calibrated post fleet reconstruction by fitting the age distribution of the fleet observed in 2016 (taking into account the observed fleet growth rate of 2005-2016). This means that assuming that the reconstructed regional fleets are true-to-reality, both the survival rates and the growth rates are quite well constrained. We note that there are some regions where we found that the average age of the reconstructed fleet is younger compared to literature values, but due to a lack of observational data we are unable to achieve a better fit. Both of these functions are used as time-independent parameters.

² In our nomenclature, the survival function works as follows: survival from age = 0 to age = 1 is calculated using the survival value at age = 1, meaning that the surviving number of cars of age $\{n\}$ in t to age $\{n+1\}$ in $t+1 = \text{cars}(\text{age}_{\{n\}}) * \text{surv}_{\{n+1\}}$.

A small but notable change is that for better precision we extended the vintage range from age 0-29 to age 0-49, even though this is not so relevant for European countries. (There is no computational limit anymore that would necessitate the trimming of the age range.)

B) For the scenarios discussed in this study we use the technology share of car-sales as a model input parameter. This in return means that we can remove uncertainty from the model. The most important consequence is that

1. The mileage gap can be filled precisely each year, and
2. there is no need for the logit calculation that governed the sale shares in the original MOVEET model.³

The calculation of the number of vehicles that are about to enter the regional market in year $t+1$ to fill the mileage gap is done by calculating the mileage supply of one virtual car distributed over the different technologies according to the know sale shares of year $t+1$ and over the different vintages following the known growth rates, and then simply dividing the demand gap with this unit car supply.

This means that the price calculations (fixed and variable costs per vehicle kilometre) in the original MOVEET model do not influence the fleet calculations in this modified model in any way. It is important to stress, that these adaptations mean that this version of the model does not make forecasts, but it simulates various scenarios based on the provided input.

C) As the fleet database of the original MOVEET model became dated since its last update, we decided to build the regional fleets from scratch. This way we could construct a homogeneous vehicle stock for each region, making sure that vehicles of all propulsion types are well represented, and various data sources are brought together in a consistent way.

We constructed new stocks for each region using various data sources and techniques that were calibrated to match available literature values as well as possible. As a result fleet numbers in 2016 match the observed totals, and the observed technology distributions, and only the validity of the age distribution varies within regions. We detail the fleet reconstruction process later on.

³ For the first two cities (Bristol and Amsterdam) we still used the original model with the logit calculations, but to reach the aimed technology market shares that were defined during the scenario definitions (instead of letting the prices in our model drive the future sales), we employed extra alternative specific (but time-variable) constants labelled as the preference parameter for each propulsion technology, therefore in practice fixing the yearly sale shares in an artificial way. (While this is admittedly artificial, it is actually a reasonable way to model the changing acceptance of various technologies, which is not only connected to parameters that such as price and range.) The end result is the same in the modified model, but without the use of the artificial alternative specific constants. Since we are not trying to predict the future, but we aim to simulate possible future scenarios, there is no need to employ a logit choice model at all.

During the modelling process when we had no exact fleet numbers for the modelled cities or regions, we scaled the country-level fleet according to the population if the city/region in question.

3.3.2 Update of the vehicle fleet data in MOVEET

We rebuilt the European car-stock from scratch, as it was easier to do this self-consistently, than trying to update the existing stock numbers. While ClairCity covers only six countries, this work was done for the whole of Europe so the resulting fleet could be used for future scenario simulations for other cities/countries too.

For this, we started from the fleet data of the TRACCS project⁴. This contains homogeneous and detailed fleet data for the 2005-2010 period, that we used as the initial age and propulsion type distribution backbone for our reconstruction process. We note here that it is clear that the TRACCS data is not perfect, and we had to treat it carefully to avoid amplifying any of its shortcomings, but it is still far the best and most complete data set available for this purpose.

We derived survival curves from the TRACCS data for each country and propulsion technology, and we let these survival rates go above unity, which – already foreshadowing the idea of the growth function – enabled us to describe the excess growth of specific vintages caused by the influx of used vehicles.

From year t to year $t+1$ these survival curves produce the complete fleet structure except for the age = 0 (in year $t+1$) cars. We filled up this vintage, the new cars for years 2010-2016, from the Monitoring of CO₂ emissions from passenger cars⁵ database of the European Environment Agency (EEA CO₂ database from hereafter), which contains a more-or-less complete record of all car registrations within Europe for the given period. We made sure that the yearly data tables are homogenised (meaning that the propulsion and weight categories had are matched with the categories in our MOVEET model), and scaled when necessary to match the total registration numbers (in a few countries the propulsion type data was not complete, here we had to scale the number of cars in a given category to match the grand total of all registrations in the country and year in question). We did not take the LPG registration numbers from this database, as the reported numbers in this category are not consistent with other data sources we consider more reliable (e.g., data from Liquid Gas Europe and EAFO).

Other data sources that we have used during the reconstruction and calibration process that is discussed below are the following: the 2017 edition of ACEA's 'Vehicles in use' report⁶, EUROSTAT tables, the European Commission's Mobility and Transport statistical Pocketbook⁷, the International Transport Forum's Key Transport Statistics documents⁸,

⁴ <https://traccs.emisia.com/index.php>

⁵ <https://www.eea.europa.eu/data-and-maps/data/co2-cars-emission-14>

⁶ <https://www.acea.be/statistics/article/vehicles-in-use-europe-2017>

⁷ https://ec.europa.eu/transport/facts-fundings/statistics/pocketbook-2018_en (and previous years)

⁸ <https://www.itf-oecd.org/key-transport-statistics-2018-2017-data> (and previous years)

Statistics Norway⁹, the Swiss Federal Statistical Office¹⁰, and the European Alternative Fuels Observatory¹¹. Most of these data sources were used to derive trustworthy total fleet numbers, the number of yearly total registrations, and the number of new-technology vehicles in countries where they are already significant. LPG vehicle counts were taken from Liquid Gas Europe and EAFO (and for a few regions estimates were made based on articles found mostly on gazeo.com). Global fleet values where necessary were taken from the International Organization of Motor Vehicle Manufacturers¹².

These are the steps in the (historic) fleet reconstruction and calibration process:

1. We derive survival rates (SR) from the TRACCS data. For some countries (mainly smaller and/or Eastern European markets) we use only the last two or three years from the 2005-2010 period, as the first years show clear signs of being incomplete or inaccurate (an artefact from the fleet construction within the TRACCS project).
2. We use these SR curves to populate the age structure (age > 0) for 2011-2016, using the EEA CO2 data for new cars (age = 0) each year.
3. We construct an initial estimate for the fleet using the yearly fleet totals per country, combined with an assumed survival rate (having a similar average scrappage age to the other propulsion types of the county) and growth curve (that was deemed a reasonable estimate given the market properties of the county in question). At this stage the age distribution of the fleet did not have to be very precise, only the yearly total numbers had to match with the numbers from EAFO and Liquid Gas Europe, as later we recalculated the fleet using more detailed and better constrained market properties.
4. To calibrate the number of vehicles in the fleet, the fleet's age distribution, and the share of different propulsion technologies to the numbers from ACEA, we adjusted the SR curves derived from the TRACCS data, influencing the import and scrappage rates (as these are not expected to stay constant in real life). At the end we scaled the fleet (excluding the LPG cars, as this part of the fleet was reconstructed in the previous step already making sure that the annual totals match the observed numbers) each year with a constant (always very close to one, typically between 0.95 and 1.05) to be fully consistent with the yearly ACEA fleet totals.
5. The SR curves derived from TRACCS were acceptable on the short term, but they were not suited to be use them for long term predictions, therefore for each county/region we calibrated a generalised survival curve and a generalised growth function from the age distribution of the 2016 fleet, based on the observed historical fleet growth percentage and an assumption (based on market data) describing the ratio of new versus used import cars entering the market each year.

⁹ <https://www.ssb.no/en>

¹⁰ <https://www.bfs.admin.ch/bfs/en/home.html>

¹¹ <https://www.eafo.eu/>

¹² <http://www.oica.net/category/vehicles-in-use/>

6. In the next step we reconstructed the (whole historical) fleet for each country/region using the corresponding general SR curve derived in the previous step, and an specific growth curve that we have constructed based on market-specific input received from Liquid Gas Europe. The main ideas behind the reconstruction process were filling the fleet-gap (from t to $t+1$) with a sum of vehicles distributed over vintages according to the growth curve, or when there was a sudden drop observed in the yearly totals (that could not be explained by simply assuming no car registrations/conversions) we temporarily strengthened the scrappage by applying an exponent to the survival rate (causing more, and even younger cars being removed from the fleet in a given year).
7. We iterated steps 5)-6) until there was no significant change in the derived survival curve (typically one iteration).

At the end of the fleet reconstruction process, we have the fleet, the survival rate, and the two growth rates (LPG and other) for each country/region (see example below for Poland).

figure 19: (top two tables) Model validation for Poland by comparing the modelled and observed age distributions, and the technology shares to the 2015 ACEA values. (lower two tables) Some statistics about the yearly evolution of fleet shares and new car sales.

Age	Total	LPETROL	HPETROL	LDIESEL	HDIESEL	ELECTRIC	FUELCCELL	HYBRID	HYBDIS	HYBH2	FCH2	CNG	LPG	E85	B30
0	297605	135474	8192	114305	13144	138	0	113	1	0	0	2309	23929	0	0
1	378729	221908	6559	107917	15130	68	0	257	15	0	0	5	26852	19	0
2	351268	194579	4758	95967	13582	65	0	687	11	0	0	12	41604	2	0
3	372694	186219	3869	98804	14445	29	0	0	0	0	0	17	69227	85	0
4	389980	168258	3570	106918	13647	123	0	0	0	0	0	28	96952	484	0
5	444041	170013	3937	127630	16110	104	0	0	0	0	0	51	125487	708	0
6	420756	162646	3446	99164	9525	18	0	0	0	0	0	22	145932	2	0
7	616802	151542	4187	275413	38661	0	0	0	0	0	0	0	146999	0	0
8	655431	162897	4726	305454	40206	0	0	0	0	0	0	0	142147	0	0
9	663290	178577	5816	291659	39944	0	0	0	0	0	0	0	147293	0	0
10	637984	179637	6344	253813	37018	0	0	0	0	0	0	0	161173	0	0
11	731892	316795	14987	200251	34345	0	0	0	0	0	0	0	165514	0	0
12	826721	342460	16004	257233	43353	0	0	0	0	0	0	0	167670	0	0
13	883357	356742	16418	297265	50651	0	0	0	0	0	0	0	162281	0	0
14	976510	379795	17331	319020	56266	0	0	0	0	0	0	0	204098	0	0
15	1029971	392685	17829	316979	58424	0	0	0	0	0	0	0	244054	0	0
16	1399311	690664	32877	384217	79378	0	0	0	0	0	0	0	212175	0	0
17	1357111	689594	32765	352246	77014	0	0	0	0	0	0	0	205492	0	0
18	1261721	683633	32436	316904	73748	0	0	0	0	0	0	0	154999	0	0
19	1178540	675770	32037	282747	70372	0	0	0	0	0	0	0	117615	0	0
20	1089184	663002	31485	247660	66238	0	0	0	0	0	0	0	80799	0	0
21	947167	602121	28568	206869	59152	0	0	0	0	0	0	0	50457	0	0
22	866938	579952	27817	175767	54061	0	0	0	0	0	0	0	29341	0	0
23	795328	553688	26999	148820	49023	0	0	0	0	0	0	0	16797	0	0
24	726022	522735	26049	122964	43562	0	0	0	0	0	0	0	10712	0	0
25	664186	492651	25133	100148	38184	0	0	0	0	0	0	0	8070	0	0
26	401932	296690	15130	59507	24025	0	0	0	0	0	0	0	6580	0	0
27	353670	270678	14160	44422	19060	0	0	0	0	0	0	0	5350	0	0
28	313904	247662	13237	33676	15116	0	0	0	0	0	0	0	4214	0	0
29	289894	233590	12702	27372	13042	0	0	0	0	0	0	0	3188	0	0

figure 20: Age structure of our reconstructed car fleet model of Poland. The scales for conventional and non-conventional fuel types are different for better visibility.

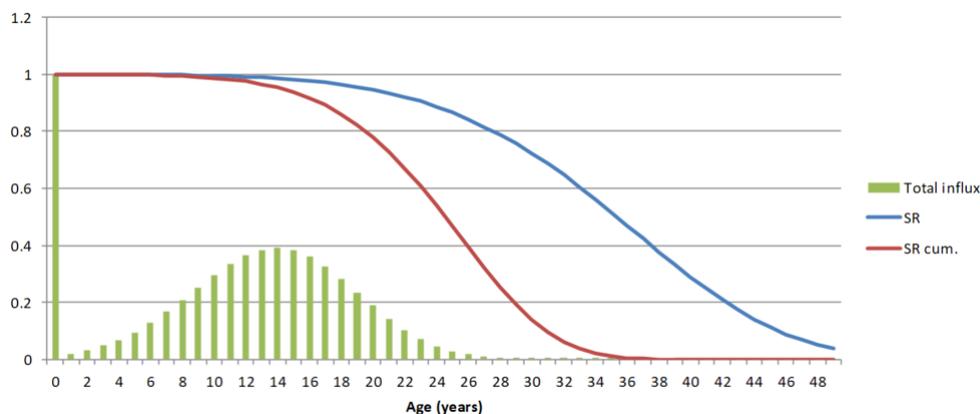


figure 21: The growth curve, the survival curve, and the cumulative survival curve for Polish non-LPG cars. It is clear, that a dominant percentage of cars are arriving to the Polish fleet as used cars.

Note for the case of Bristol and Amsterdam: the process detailed above was only carried out after we finished the modelling of Bristol and Amsterdam. Therefore for these cities we simply updated the existing MOVEET fleet using new registration numbers from the EEA CO2 database, which – by keeping the original scrappage curves of MOVEET – was sufficient to reconstruct the changes in the fleets that have occurred in the past few years (and introduce for example electric and hybrid vehicles to these fleets).

3.3.3 Effect of policies on the fleet in the scenarios

Later on we will discuss the various policies and their effects, but here we would like to give a general impression how policy scenarios affect the vehicle fleet itself. In a lot of scenarios cities wanted to get rid of the most polluting vehicles, which resulted in policies ranging from banning the most polluting vehicles (old Euro standard diesels) in low emission zones to completely banning cars with an internal combustion engine. These kind of measures were all modelled by changing the survival rates and the growth rates in the fleet models.

For example if a city decides to introduce a stepwise ban on diesel cars, so that, e.g., by 2030 only Euro 6 diesel cars are allowed on the roads, then we model this by making sure that all Euro 5 or worse diesels are scrapped out of the fleet by the end of 2029 (for example scrapping all Euro 3 by end 2025 and Euro 4 by end 2027 as a lead up to put smaller pressure on the new sales that still need to fill the suddenly larger demand-gap). Such scrapping schemes are very easy to model by setting the survival rates for all to be fully scrapped vintages to 0.

There are a few possible complications with these kind of measures:

1. Let's assume a full ban on diesel by 2030. This means that for example it is unrealistic to assume that the market share¹³ of diesel would stay unchanged up until 2030 when it suddenly drops to zero, so along with (step-by-step) scrapping newer and newer vintages of diesels leading up to the ban, we also need to revise downwards the sale shares of diesel vehicles to reflect the fact that users are (hopefully) aware that a diesel car purchased in, e.g., 2025 would only stay on the road for 5 years, which is well below the expected life of a diesel car, historically. Since the expected utility of a new car gets less and less, we need to reflect this change in the sale shares. In these cases, after dropping the diesel sale shares, either all other technology shares need to be pumped up to fill the created temporary mileage gap, or potential diesel buyers might go for petrol vehicles. Since most policies were pushing people towards clean technologies, we went for the earlier solution in most cases. When only old (Euro 5 and worse) diesel cars are banned then these changes are not necessary, as the sales of modern diesel vehicles are not limited.
2. Bans in countries with used car influx (so where growth rates for age > 0 are not zero): assuming that we implement a ban on, e.g., age 5 and older diesel cars in by 2025, besides scrapping all cars up to age 4 at the end of 2024, we also need to stop sales of cars of age 5 and older in 2025. We do this by setting the growth rates to zero for these vintages in 2025, and for the corresponding vintages in the following years. Now we need to make sure for each year that the remaining growth rates add up to 1, otherwise we create an artificial lack of sales in this propulsion technology, and – due to the specifications of our model – the mileage gap will not be filled fully. This is easy, we simply multiply the remaining non zero growth rates up with a constant so that their sum adds up to 1. It is clear that when there are no used car

¹³ By market share we always refer to the share in sales in a given year, a.k.a. sale share.

sales and only the age 0 growth rate is not zero (it is 1), then this problem never happens as we do not need to modify growth rates only the survival rates.

We will discuss in more detail the city-specific, local policy-driven modelling assumptions that went into modelling the various sections later in the city-specific sections.

3.4 Mode choice

The aim of a mode choice model in the context of this project is to set up a tool that we can use to predict the modal shift (and the mileage change) caused by various policy decisions or scenario-targets. Typical measures/scenario goals are:

- Changes in public transport pricing
- Infrastructure changes causing a change in the typical speed of the vehicles, or even the distances travelled between points A and B.
- Modal shift goals of a given percentage trip share of active modes or public transport.

What a mode choice model does is that it predicts for a given traveller who has some socioeconomic background (income, car ownership, age, gender, etc.), who wants to travel for a given reason (work, education, shopping, etc.), between points A and B that are a given distance apart, what the percentage chances are that this traveller chooses a given transport mode. So for example: a young middle class female who lives in a household without a car wants to travel from their home to work in the morning, then what is the percentage chance that she would take the car/public transport/bike/walk/etc. To build such a model, we need a lot of observational data from various trips and their properties (including distance, duration, and price), and then we can construct a model that predicts the observed choices of various users as best as possible.

Then with such choice model in hand, we can examine how the predicted mode choices would be influenced if we assume, e.g., that the costs associated with one of the modes are changed (e.g., as a result of a policy). We might (and will) see that, e.g., if the prices of public transport are lowered, then more people will choose public transport compared to what we observed before. Moreover, since the distance distribution of trips over different modes is not the same (e.g., bike trips are on average shorter than car trips, and so on), a percentage shift in mode share (trip share) does not translate linearly to a change in the mileages covered by different modes. (This would only be the case if the trip length distributions over all modes would be identical.) By evaluating a derived mode choice model on the observed trip data using, e.g., a change in pricing, we can calculate the change not only in the percentage of trips over various modes, but also the change in total mileage per mode. This is important for energy and emission calculations that are directly related to the vehicle kilometres.

In theory, in an optimal case, such a model should be built for each individual city or region. The reason for this is that the cities differ on many levels: e.g., population structure, socioeconomic parameters, geography (and therefore size), distribution of various activity centres, such as locations of housing, shopping, education, financial properties, differences in transport systems (composition, density, accessibility), etc. The most convenient way to get all the necessary observational data that could support a model that is specific to a given location, is to have access to the data of a travel diary style traffic survey.

In such a survey a carefully selected representative sample of the population is asked to journal their daily trips – from simple walks to airplane trips – over a given time period, which along with the personal and household level properties of the respondents can be translated to a complete picture about the various movements that happen within the surveyed region. One can use these trips – weighted usually with carefully derived weighting factors to make the sample fully representative of the studied region – as a starting point of building a mode choice model.

There are challenges though: these surveys take a lot of time and effort to carry out, and the resulting database is in many cases not publicly available. Even if it is available, it is most of the time a survey of a full country, and not only of a city in question. We cannot just filter the dataset to respondents with an address (if this data is not excluded from the database for privacy reasons) that is within our region of interest, because a) we would likely exclude so many trips from the database that it would not be complete anymore (various types of trips would be significantly over or under represented), or would not provide a significantly high number of observations for a detailed model, b) we would exclude traffic generated by people that do not live inside the city/region.

Considering all the data and methodological difficulties, it is – at the end – not such a bad option to assume that cities/regions of a typical ClairCity size within Europe are relatively similar, and try to gather the best available data from one given location, build a model using that dataset, and then use the resulting mode choice model in other regions too, after calibrating some parameters of that general model to some of the known properties of these other cities/regions. This is what we settled with after some initial trials. Given the often very arbitrarily defined policies or scenario-goals on the city level, we believe that the overall precision is not lowered by using such a general (but locally adjusted) mode choice model.

4 Industrial, Residential, Commercial and Institutional emissions (IRCI)

The tool set module produces a specific emissions inventory for Industrial, Residential/ Commercial and & Other services sources.

Emission sources are generally classified as point, area and line sources. Point sources are stationary sources whose emissions exceed some fixed thresholds, fixed depending on the aim of the inventory. In order to compile city/region inventory in ClairCity, stationary emission sources with total annual emissions with at least one pollutant emissions exceeding 100 Mg/year, are considered as Large Point Sources (LPS).

Emissions from LPS can be conveyed in stacks or fugitive. When emissions are emitted by stacks, emission characteristics (gas temperature and volume and stack height) are necessary for air quality modelling.

4.1 Industrial sources

For industrial sources emissions, a specific tool has been developed to include all the emissions from the European Pollutant Release and Transfer Register (E-PRTR)¹⁴ facilities and national and local Registers or emissions inventories (national, regional and local scale). Where specific facilities are individuated that don't have known emissions data, ad hoc estimates are obtained using available information and emission factors from EMEP/EEA Guidebook¹⁵. The emissions have been geographically allocated by coordinate of emission source. When data on single facility are not know emissions are evaluated from statistical sources as area sources and allocated using land cover maps.

Large Point Sources have been individuated and reported, source by source, with emissions and characteristics of emissions. Large Point Sources are defined as sources that emits more than 100 Mg of NO_x or PM₁₀.

4.2 Residential, commercial and institutional sector

The IRCI tool evaluates emissions at most detailed administrative territorial units' level, uses emission factors from EMEP/EEA Guidebook and calculates emissions as:

$$E = A_{ij} F_{ij}$$

where:

- A_{ij} is the indicator of the activity i in the territorial unit j ,
- F_{ik} is the emission factor of pollutant k for activity i (expressed in grams per unit of activity).

¹⁴ [European Pollutant Release and Transfer Register \(E-PRTR\)](#)

¹⁵ EMEP/EEA(2016), [Air pollutant emission inventory guidebook](#)

For small combustion in the residential, commercial and institutional sector, emissions have been evaluated using emission factors from EMEP/EEA Guidebook and specific activity level.

4.2.1 Emission factors

In Table 5 the fuel combustion air pollutants emission factors used are reported¹⁶.

Table 5: Residential, Commercial & Institutional Air pollutant Emission Factors

Fuel	Nitrogen oxides [gNO _x /GJ]	Particle Matter with diameter less than 10μ [gPM ₁₀ /GJ]
Boilers – Natural gas	42	0,2
Boilers – LPG	40	2
Boilers – Gas/Diesel Oil	69	1,5
Boilers – Wood	80	480
Fireplaces – Wood	50	840
Stoves – Wood	50	760
Energy Efficient Fireplaces – Wood	80	380
Energy Efficient Stoves – Wood	80	380
Advanced Fireplaces – Wood	95	95
Advanced Stoves – Wood	95	95
Pellets Stoves – Wood	80	29
Boilers - Hard Coal	158	225
District Heating Boilers - Natural gas	89	0.89
District Heating Boilers – Wood	81	155
District Heating Boilers – Hard Coal	209	7.7

4.2.2 Territorial domains

The Air Pollutant Emissions are evaluated in the city domains defined as follow for the territorial units' classes reported in Table 6.

¹⁶ EMEP/EEA(2016), [Air pollutant emission inventory guidebook,1.A.4 Small combustion](#)

Table 6: Territorial domains

City/region partner	Lower level subdivision
Bristol	LSOA
Amsterdam	Buurt
Ljubljana	Naselje
Sosnowiec	Gminas
Genoa	Census Sections
Aveiro	Freguesia

The Bristol modelling domain (red box) is reported in Figure 22, with boundary of LSOA belonging to the domain (yellow areas), Amsterdam modelling domain (red box) is reported in Figure 23, with boundary of *Buurt* belonging to the domain (yellow areas), Sosnowice modelling domain (red box) is reported in Figure 24 with boundary of *Gminas* belonging to the domain (yellow areas), Ljubljana modelling domain (red box) is reported in Figure 25 with boundary of *Občine* belonging to the domain (yellow areas), Região de Aveiro modelling domain (red box) is reported in Figure 26 with boundary of *Freguesia* subdivision belonging to the domain (yellow areas), and Liguria Region (Genoa area) modelling domain (red box) is reported in Figure 27 with boundary of Census Sections subdivision belonging to the domain (yellow areas).

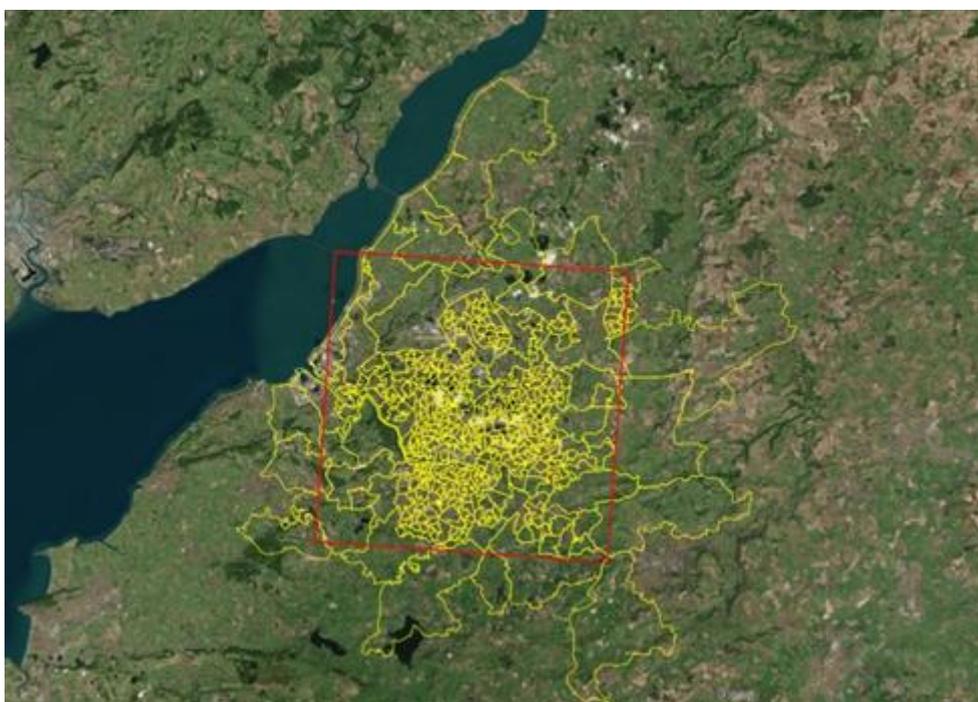


Figure 22: Bristol domain with LSOA subdivision



Figure 23: Amsterdam domain with Buurt subdivision

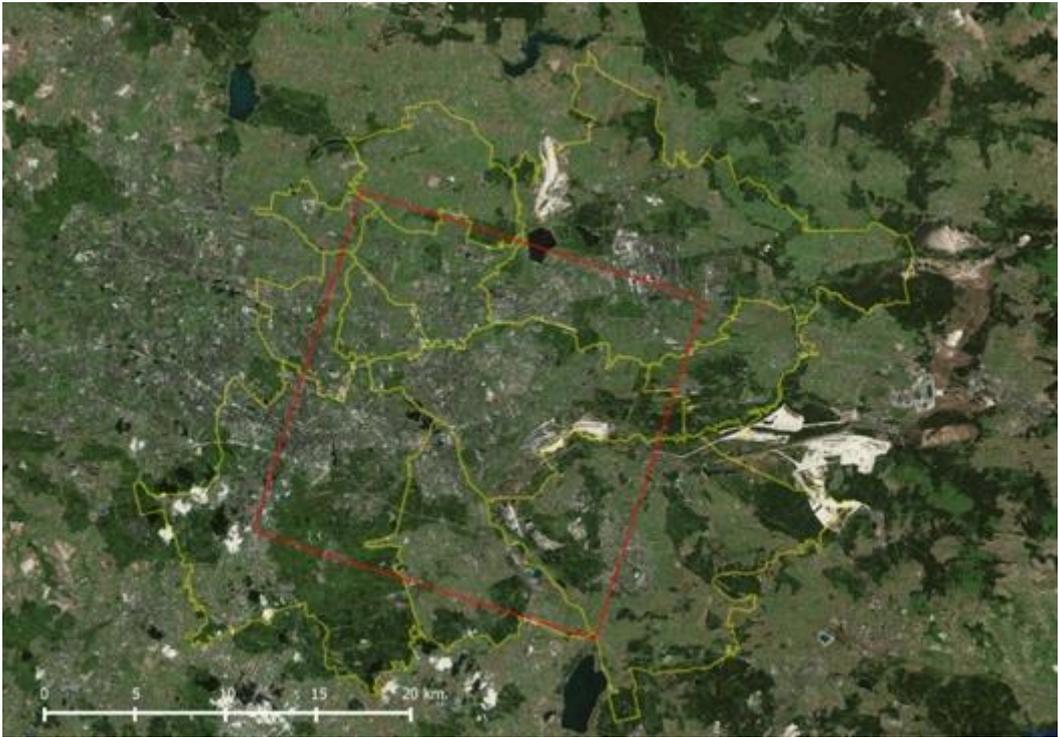


Figure 24: Sosnowiec domain with Gminas subdivision

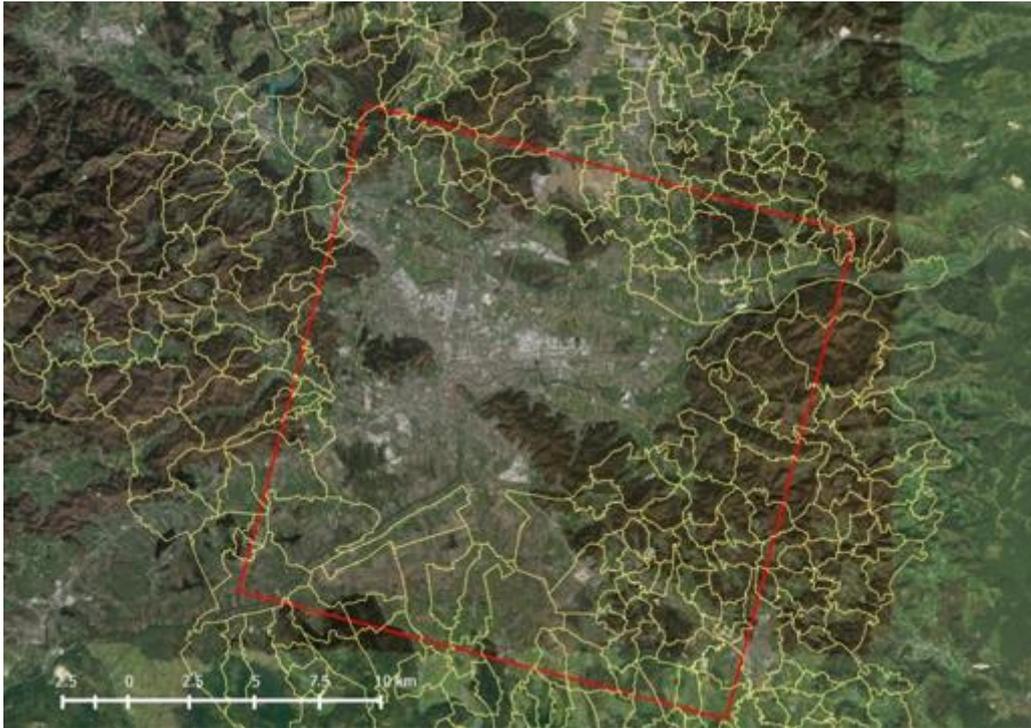


Figure 25: Ljubljana domain with Naselje subdivision

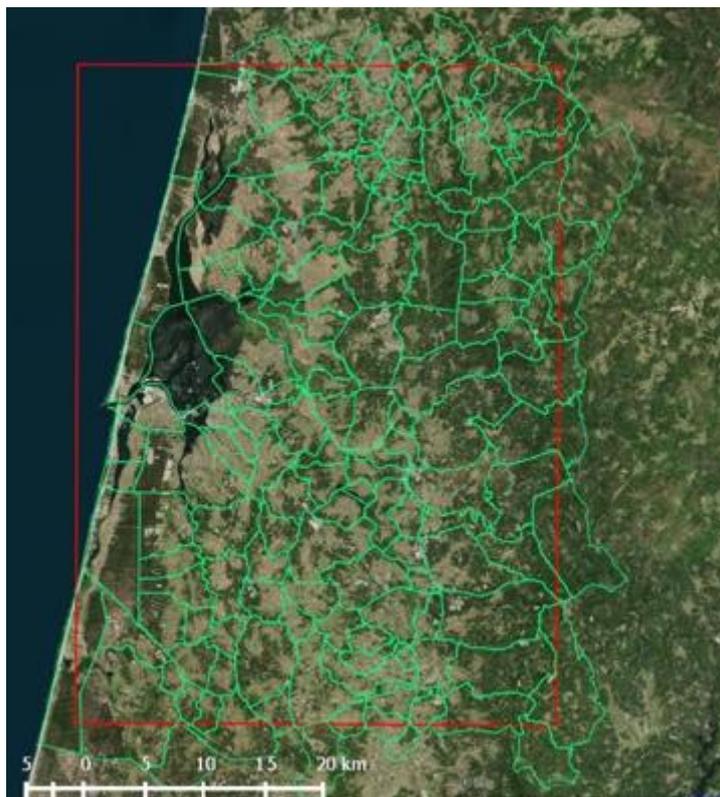


Figure 26: Região de Aveiro domain (red box) with Freguesia subdivision

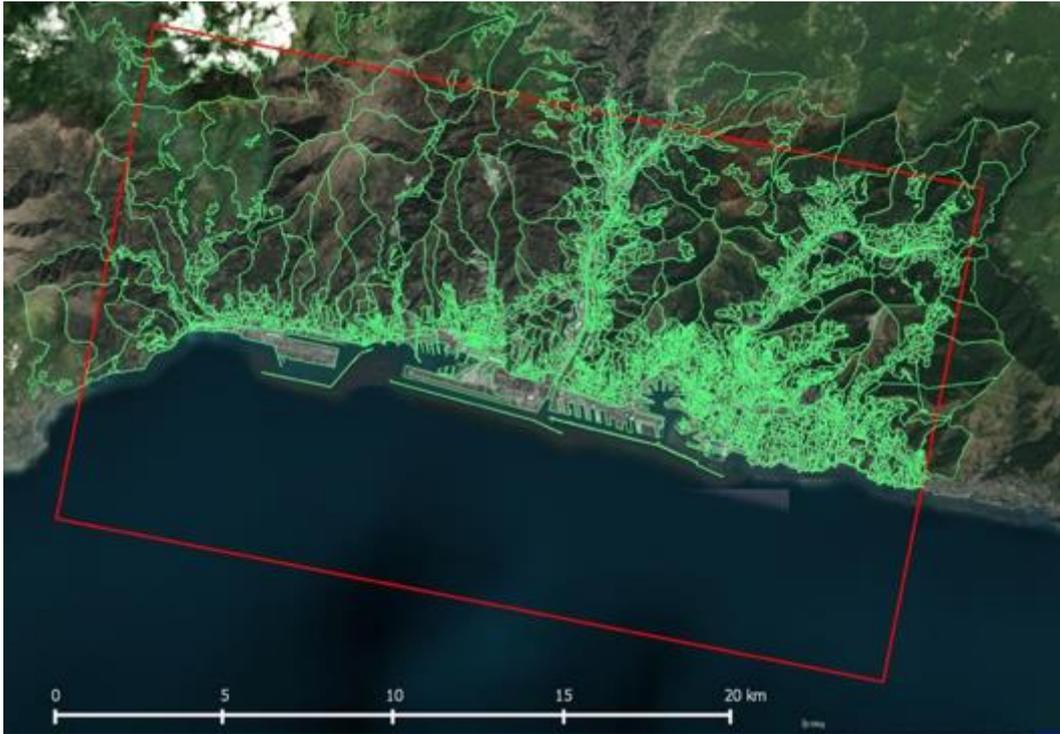


Figure 27: Liguria Region (Genoa area) domain with Census Sections subdivision

4.3 Classification of activities and fuels for emission estimates

The module needs, as an input, data on fuel consumptions in residential and commercial sectors, detailed by technologies (boilers, fireplaces, stoves, ...) and fuels as in the following Table 7

Table 7: Variable used to evaluate emissions in industrial, residential and commercial sector

Code	Name
020131A0	Commercial Combustion plants (boilers) (2110 - Hard Coal) [Mg]
020131F0	Commercial Combustion plants (boilers) (3220 - LPG) [Mg]
020131I0	Commercial Combustion plants (boilers) (3260 - Gas/Diesel Oil) [Mg]
020131M1	Commercial Combustion plants (boilers) (4100 - Natural gas) [Mm3]
020401A0	Commercial Combustion plants (district heating) (2110 -Hard Coal)
020401M1	Commercial Combustion plants (district heating) (4100 - Natural gas)
020401N0	Commercial Combustion plants (district heating) (5541 - Solid biomass) [Mg]
020220A0	Residential Combustion plants (boilers) (2110 - Hard Coal) [Mg]
020220F0	Residential Combustion plants (boilers) (3220 - LPG) [Mg]

Table 7: Variable used to evaluate emissions in industrial, residential and commercial sector

Code	Name
020220I0	Residential Combustion plants (boilers) (3260 - Gas/Diesel Oil) [Mg]
020220M1	Residential Combustion plants (boilers) (4100 - Natural gas) [Mm3]
020220N0	Residential Combustion plants (boilers) (5541 - Solid biomass) [Mg]
020221N0	Residential Fireplaces (5541 - Solid biomass) [Mg]
020222N0	Residential Advanced Fireplaces (5541 - Solid biomass) [Mg]
020223N0	Residential Conventional Stoves (5541 - Solid biomass) [Mg]
020224N0	Residential Advanced Stoves (5541 - Solid biomass) [Mg]
020226N0	Residential Energy Efficient Fireplaces (5541 - Solid biomass) [Mg]
020402A0	Residential Combustion plants (district heating) (2110 - Hard Coal)
020402M1	Residential Combustion plants (district heating) (4100 - Natural gas)
020402N0	Residential Combustion plants (district heating) (5541 - Solid biomass)

4.4 Data evaluation on domains and subdomains

Air Pollutant emissions are evaluated on defined subdomains. When data are available only in aggregate figures (overall domain) or at national level, they are allocated to subdomains using a “proxy” variable available at subdomains level. The data know at domain or national level are evaluated on subdomains using a lot of proxy variables known at the subdomains level to allocate overall values on a subdomain’s basis. Proxy variables allow obtaining information on a certain spatial resolution assuming that it is known for larger spatial resolutions.

When data are available at city/region domain, data at subdomain level is evaluated using the following equation:

$$A_i = A * P_i / \sum_i P_i$$

where: A_i and P_i are the values of variable A and proxy variable P in the subdomain i, and A is the total of variable A in the domain.

When data are available at national level, data at subdomain level is evaluated using the following equation:

$$A_i = (A * Q_d / \sum_d Q_d) * P_i / \sum_i P_i$$

where: A_i and P_i are the values of the variable A and proxy variable P in the subdomain i , Q_d is the values of the proxy variable Q in domain d , and A is the national value of proxy variable A .

4.5 Future emissions calculation

Emissions projection are estimated for area and point sources. Area emissions for future year (k) in a single territorial unit (e. g. LSOA, Buurt, ...) (u) related to a specific activity (i) are estimated starting from the base year (0) emissions and using specific projections factors (drivers) of activity level (a_{ikn}) due to activity measures n , specific drivers for emission factors (f_{ijkm}) due to emissions control measures m for a selected pollutant (j), specific projections factors (drivers) of activity level related to a selected territorial units (a_{ikun}^d) due to activity measures n and specific drivers for emission factors (f_{ijkum}^d) due to emissions control measures m and, if any, additional emissions foreseen for a selected new activity in a selected territorial units u (E_{ijk}^{new}):

$$E_{ijk}^d = E_{iju0}^d \prod_m \prod_n a_{ikn} f_{ijkm} a_{iukn}^d f_{ijkum}^d + E_{ijk}^{new}$$

Point source emissions for future year (k) for a selected production unit of a selected plant (u), related to activity (i) are estimated starting from the base year (0) emissions and using specific drivers of activity level (a_{ik}) due to activity measures n , specific drivers for emission factors (f_{ijkm}) due to emissions control measures m for a selected pollutant (j), specific drivers (a_{ijkun}^p) for activities i , related to selected production unit of a selected plant u , due to activity measures n , specific drivers for emission factors related to a selected pollutant (j) related to a selected line (f_{ijkum}^l) due to emissions control measures m , and, if any, additional emissions foreseen for a new plant u (E_{ijk}^{new}):

$$E_{ijk}^l = E_{iju0}^l \prod_m \prod_n a_{ikn} f_{ijkm} a_{iukn}^l f_{ijkum}^l + E_{ijk}^{new}$$

Drivers for activity levels and for emission factors can be related to multiple activities, plants and lines; for example, the “population” driver can be used to forecast emissions coming from glue or paint applications in household as well as drivers related to fuel consumptions in residential sector, and sulphur content of distillate oil can be used for the projections of emission factors for sulphur dioxides in the industry and service sectors. Besides, an abatement system for particulate matter can be applied to a single activity (e.g. power plants with output higher than 300 MWth) and to one or more production units related to another activity (e.g. aluminum production).

New additional emissions for point, line and area sources are used to simulate planned plants, planned units of a plant, and planned roads and so on. It is possible to manage the following groups of drivers in the policy scenario assessment:

- Drivers for activity levels related to single activities, territorial units, lines, production units of industrial plants.
- Drivers for emission factors related to single activities, territorial units, lines, production units of industrial plants.
- Additional emissions related to single territorial units, lines, production units of industrial plants.

5 Carbon footprint

The goal of this module is to compute the carbon footprint of the cities.

5.1 Specifications of modelling tool set

First of all, a complete review of existing carbon footprint methodologies was completed. Carbon footprint calculation methodologies usually follow the "emission inventory" or the "consumption-based" approaches. The review covered as the "emission inventory" approaches (such as WRI protocol for example¹⁷) than input-output models approaches.

The "emission inventory" approaches usually include generating GHG emissions activities that occur inside the city/region boundary as well as outside the city/region boundary. To distinguish between these, the approaches group emissions into three categories based on where they occur: scope 1, scope 2 or scope 3 emissions. These approaches distinguish between emissions that physically occur within the city/region i.e. from sources located within the city boundary (scope 1), from those that occur from the use of electricity, steam, and/or heating/cooling supplied by grids which may or may not cross city/region boundaries (scope 2), from those that occur outside the city/region but are driven by activities taking place within the city's/region's boundaries (scope 3). Scope 1 emissions may also be termed "territorial" emissions, because they are produced solely within the territory defined by the geographic boundary.

Cities, by virtue of their size and connectivity, inevitably give rise to GHG emissions beyond their boundaries. Measuring these emissions allows cities to take a more holistic approach to tackling climate change by assessing the GHG impact of their supply chains, and identifying areas of shared responsibility for upstream and downstream GHG emissions.

The usual methodologies for city (such as WRI Protocol) includes scope 3 accounting for a limited number of emission sources, including transmission and distribution losses associated with grid-supplied energy, and waste disposal and treatment outside the city boundary and transboundary transportation.

Inventories may optionally report other scope 3 sources associated with activity in a city-such as GHG emissions embodied in fuels, water, food and construction materials.

Consumption-based accounting is an alternative to the sector-based approach to measuring city/region emissions. This focuses on the consumption of all goods and services by residents of a city/region, and use EFs expressed as emissions for unit of good and services including all the processes (production, transport, disposal).

GHG emissions are reported by consumption category rather than the emission source categories. The consumption-based approach allocates GHG emissions to the final

¹⁷ WRI(2014), Global Protocol for Community-Scale Greenhouse Gas Emission Inventories. An Accounting and Reporting Standard for Cities

consumers of goods and services, rather than to the original producers of those GHG emissions. Consumption-based inventories typically use an input-output model, which links household consumption patterns and trade flows to energy use and GHG emissions, and their categories cut across those set out in the GPC. This approach is complementary to the GPC and provides a different insight into a city's/region's GHG emissions profile. At conclusion of review work a best methodology has been established defining the selected approach and the activities included in scope 3 to take into consideration.

In the scope of the ClairCity project, the chosen approach is to evaluate in an integrated way emissions inventory and carbon footprint. Future scenarios are produced for simultaneous reduction of air pollutants emission and carbon footprint.

Considering the general goals of the activities and to delimitate the efforts from other partners in data retrieval, we select the **scope 2 approach** considering emissions that:

- physically occur within the city/region i.e. from sources located within the city boundary;
- occur from the use of electricity, steam, and/or heating/cooling supplied by grids which may or may not cross city/region boundaries.

As the project is finalized to produce strategies for the cities the carbon footprint evaluation will be conducted following an emission inventory approach similar to the approach followed in the Covenant of Mayors¹⁸ and using both:

“Standard” emission factors (emissions will be evaluated using methodologies and emission factors from 2006 IPCC Guidelines for National Greenhouse Gas Inventories¹⁹ and specific activity level); the methodology cover all the CO₂ emissions that occur due to energy consumption within the territory of the city/region, either directly due to fuel combustion within the city/region or indirectly via fuel combustion associated with electricity and heat/cold usage within their area; the standard emission factors are based on the carbon content of each fuel, like in national greenhouse gas inventories in the context of the UNFCCC and the Kyoto protocol; in the standard approach, the CO₂ emissions from the sustainable use of biomass/biofuels, as well as emissions of certified green electricity, are considered to be zero; emission are reported as:

- CO₂ only emissions, the most important greenhouse gas,
- CO₂ equivalent emissions, including calculation of the emissions of CH₄ and N₂O with emission factor from 2006 IPCC Guidelines for National Greenhouse Gas Inventories²⁰ and reported as CO₂ using the Global Warming Potential (GWP) with 100 years - time horizon²¹:

$$1 \text{ Mg CO}_2 = 1 \text{ Mg CO}_2\text{-eq}$$

$$1 \text{ t CH}_4 = 21 \text{ Mg CO}_2\text{-eq}$$

$$1 \text{ t N}_2\text{O} = 310 \text{ Mg CO}_2\text{-eq};$$

¹⁸ [Covenant of Mayors \(2010\). How to develop a Sustainable Energy Action Plan \(SEAP\) – Guidebook Part II, Baseline emissions inventory](#)

¹⁹ [2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 2 Energy](#)

²⁰ [2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 2 Energy](#)

²¹ [IPCC, 1995. Contribution of Working Group I to the Second Assessment of the Intergovernmental Panel on Climate Change](#)

LCA (Life Cycle Assessment) emission factors, which take into consideration the overall life cycle of the energy carrier; this approach includes not only the emissions of the final combustion, but also all emissions of the supply chain; it includes emissions from exploitation, transport and processing (e.g. refinery) steps in addition to the final combustion; this hence includes also emissions that take place outside the location where the fuel is used; in this approach, the GHG emissions from the use of biomass/biofuels, as well as emissions of certified green electricity, are higher than zero; in the case of this approach, other greenhouse gases than CO₂ may play an important role, therefore the LCA approach will report emissions as CO₂ equivalent; as a default will be used the LCA emission factors given in Covenant of Mayors guidelines, based on JRC European Reference Life Cycle Database; specific national emission factors will be investigated.

Eventual local electricity production is not included in the model while district heating is included where data are available.

5.1.1 Carbon footprint modeling

The Carbon Footprint tool evaluates emissions at most detailed administrative territorial units' level. We use emission factors from Covenant of Mayors guidelines and consequently calculate emissions as:

$$E_k = A_{ij} F_{ik}$$

where:

- A_{ij} is the indicator of the activity i in the territorial unit j
- F_{ik} is the emission factor for different Carbon footprint indicators k for activity i (expressed in grams per unit of activity);
- k is the carbon footprint indicator used: CO₂, CO_{2eq}, CO_{2eq,LCA}

5.1.2 Emission factors

In Table 8 standard emission factors are reported for different energy vectors while in Table 9 electricity consumptions national emission factors are reported to use for CO₂ indirect emissions from electricity consumptions. EFs from JRC²² are used to have a comparable set of data for all the cities.

Table 8 CO2 Emission Factors

Fuel	Standard Emission Factors ²³	Standard Emission	LCA Emission

²² [Koffi, Brigitte; Cerutti, Alessandro; Duerr, Marlene; Iancu, Andreea; Kona, Albana; Janssens-Maenhout, Greet \(2017\): CoM Default Emission Factors for the Member States of the European Union - Version 2017, European Commission, Joint Research Centre \(JRC\)](#)

²³ [2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 2 Energy](#)

	[Mg CO2/MWh]	Factors ²⁴ [Mg CO2eq/MWh]	Factors ²⁵ [Mg CO2-eq/MWh]
Motor Gasoline	0.249	0.250	0.299
Gas oil, diesel	0.267	0.268	0.305
Residual Fuel Oil	0.279	0.279	0.310
Anthracite	0.354	0.356	0.393
Other Bituminous Coal	0.341	0.342	0.380
Sub-Bituminous Coal	0.346	0.348	0.385
Lignite	0.364	0.365	0.375
Natural Gas	0.202	0.202	0.237
LPG	0.227	0.227	0.281
Municipal Wastes (non-biomass fraction)	0.330	0.337	0.330
Municipal Wastes (biomass fraction)	0	0.007	0.106
Industrial Wastes	0.515	0.522	0.522
Wood [°]	0 – 0.403	0.007 – 0.410	0.017 – 0.416
Plant oil [°]	0 – 0.287	0.001 – 0.302	0.182 – 0.484
Biodiesel [°]	0 – 0.255	0.001 – 0.256	0.156 – 0.411
Biogas/Greengas [°]	0-0.197	0-0.197	0,087-0.284
Solar ^{°°}	0	0	0.04
Geothermal ^{°°}	0	0	0,05
Hydroelectric ^{°°°}	0	0	0,006
Wind ^{°°°}	0	0	0,01

[°] Lower value if fuel meet carbon neutrality criteria, higher otherwise ^{°°} Default EC/JRC Emission factors²⁶ ^{°°°} Default EC/JRC Emission factors for local electricity production⁹

²⁴ [2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 2 Energy](#)

²⁵ [ELCD \(2015\), European Reference Life Cycle Database \(ELCD\), Release 3.2. LCA data sets of key energy carriers, materials, waste and transport services of European scope](#)

²⁶ [Koffi, Brigitte; Cerutti, Alessandro; Duerr, Marlene; Iancu, Andreea; Kona, Albana; Janssens-Maenhout, Greet \(2017\): CoM Default Emission Factors for the Member States of the European Union - Version 2017, European Commission, Joint Research Centre \(JRC\)](#)

Table 9 CO2 National Electricity Emission Factors

Fuel	Standard Emission Factors [Mg CO2/MWh]	Standard Emission Factors [Mg CO2eq/MWh]	LCA Emission Factors [Mg CO2-eq/MWh]
Bristol (UK)	0,515	0,517	0,589
Amsterdam (NL)	0,429	0,430	0,486
Ljubljana (SL)	0,399	0,401	0,424
Sosnowiec (PL)	1,013	1,017	1,09
Genoa (IT)	0,343	0,344	0,424
Aveiro (PT)	0,314	0,316	0,368

5.2 Territorial domains and subdomains

The Carbon footprint is evaluated in the city domains defined in Table 10. Inside the city domains a lower level subdivision has been introduced as reported in Table 10.

Table 10 Territorial domains

City/region partner	Territory selected	Lower level subdivision
Bristol	MSOA	LSOA
Amsterdam	Geemente	Buurt
Genoa	Comune	Census Sections
Aveiro	Região	Freguesia
Ljubljana	Obcine	Naselje
Sosnowiec	Gminas	Gminas

5.3 Classification of activities and fuels for carbon footprint estimates

The module needs in input data for fuel consumptions in industrial, traffic, residential, commercial and institutional sector, detailed by subsector and, where defined, fuel as in the following Table 11.

Table 11 Variable used to evaluate carbon footprint

Code	Name
CF1010A0	Final Consumptions in Industry - Hard Coal
CF1010F0	Final Consumptions in Industry – LPG
CF1010I0	Final Consumptions in Industry - Gas/ Diesel Oil
CF1010M1	Final Consumptions in Industry - Natural gas
CF1010Q0	Final Consumptions in Industry – Electricity
CF1010J0	Final Consumptions in Industry – Biogas
CF1010N0	Final Consumptions in Industry - Solid biomass
CF1010P0	Final Consumptions in Industry - Motor Gasoline
CF3040A0	Final Consumptions in Other Sectors: Residential - Hard Coal
CF3040F0	Final Consumptions in Other Sectors: Residential – LPG
CF3040I0	Final Consumptions in Other Sectors: Residential - Gas/ Diesel Oil
CF3040M1	Final Consumptions in Other Sectors: Residential - Natural gas
CF3040N0	Final Consumptions in Other Sectors: Residential – Biomass
CF3040Q0	Final Consumptions in Other Sectors: Residential – Electricity
CF3040M2	Final Consumptions in Other Sectors: Residential - Green gas
CF3040M3	Final Consumptions in Other Sectors: Residential – Biogas
CF3010A0	Final Consumptions in Services - Hard Coal
CF3010F0	Final Consumptions in Services – LPG
CF3010I0	Final Consumptions in Services - Gas/ Diesel Oil
CF3010M1	Final Consumptions in Services - Natural gas
CF3010Q0	Final Consumptions in Services – Electricity
CF3020N0	Final Consumptions in Services – Biomass
CF3010G0	Final Consumptions in Services - Motor Gasoline

Table 11 Variable used to evaluate carbon footprint

Code	Name
CF3010J0	Final Consumptions in Services – Biogas
CF3010M2	Final Consumptions in Services - Green gas
CF2010I0	Final Consumptions in Transport: Road - private car - Gas/ Diesel Oil
CF2010P0	Final Consumptions in Transport: Road - private car - Motor Gasoline
CF2010Q0	Final Consumptions in Transport: Road - private car – Electricity
CF2010R0	Final Consumptions in Transport: Road - private car – Hydrogen
CF2020P0	Final Consumptions in Transport: Road - motorcycles - Motor Gasoline
CF2030I0	Final Consumptions in Transport: Road - buses - Gas/ Diesel Oil
CF2030M1	Final Consumptions in Transport: Road - buses – CNG
CF2030Q0	Final Consumptions in Transport: Road - buses – Electricity
CF2040I0	Final Consumptions in Transport: Road - light duty vehicles - Gas/ Diesel Oil
CF2040P0	Final Consumptions in Transport: Road - light duty vehicles - Motor Gasoline
CF2050I0	Final Consumptions in Transport: Road - trucks - Gas/ Diesel Oil
CF2010M1	Final Consumptions in Transport: Road - private car - Natural Gas
CF5002N0	Final Consumptions in Other Sectors: Residential (district heating) - Solid biomass
CF5001A0	Final Consumptions in Services (district heating) - Hard Coal
CF5001M1	Final Consumptions in Services (district heating) - Natural gas
CF5001N0	Final Consumptions in Services (district heating) - Solid biomass
CF5002A0	Final Consumptions in Other Sectors: Residential (district heating) - Hard Coal
CF5002M1	Final Consumptions in Other Sectors: Residential (district heating) - Natural gas

5.4 Data evaluation at domains and subdomains

Carbon footprint is evaluated at defined subdomains. When data are available only in aggregate figures (overall domain) or at national level, it is allocated to subdomains using a “proxy” variable available at subdomains.

When data are available at city/region domain, data at subdomain level is evaluated using the following equation:

$$A_i = A * P_i / \sum_i P_i$$

where: A_i and P_i are the values of variable A and proxy variable P in the subdomain i, and A is the total of variable A in the domain.

When data are available at national level, data at subdomain level is evaluated using the following equation:

$$A_i = (N * Q_d / \sum_d Q_d) * P_i / \sum_i P_i$$

where: A_i and P_i are the values of the variable A and proxy variable P in the subdomain i, Q_d is the values of the proxy variable Q in domain d, and N is the national value of proxy variable Q.

Carbon footprint for future year (k) related to a specific activity (i), in a single territorial unit (e.g. LSOA, Buurt, ...) (u), is estimated starting from the base year (0) carbon footprint and using specific projections factors (drivers) of activity level (a_{ikn}), due to activity measures n, specific drivers for emission factors (f_{ikm}) due to emissions control measures m, specific projections factors (drivers) of activity level related to a selected territorial unit (a^d_{iku}) due to activity measures n and specific drivers for emission factor (f^d_{ikum}) due to emissions control measures m in the territorial unit u:

$$C^d_{iku} = C^d_{iu0} \prod_m \prod_n a_{ikn} f_{ikm} a^d_{iku} f^d_{ikum}$$

6 Air quality

6.1 Air quality and source apportionment modelling at mesoscale

The CAMx - Comprehensive Air Quality Model with Extensions²⁷, forced by the WRF - Weather Research and Forecasting²⁸ meteorological fields, was applied over the urban areas of the 6 ClairCity pilot case studies, for the year of 2010. Since air quality is strongly dependent on meteorological conditions, for this study it is necessary to choose a meteorological year that represents the typical weather conditions of the various case studies in the period between 2010 and 2015. Thus, the 2010 meteorological year was chosen because it presents low values of temperature and precipitation anomalies, when compared to average values.

The WRF model, from the National Center for Atmospheric Research (NCAR), version 3.7., is a next generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. CAMx is a 3D chemistry-transport model suited for the simulations of the emission, dispersion, chemical reactions, and removal of pollutants in the troposphere based on the integration of the continuity equation for each chemical species on a system of nested three-dimensional grids. The gas-phase photochemistry is resolved through the Carbon Bond (CB05 or CB6) or the SAPRC99 chemical mechanism. CAMx includes a source apportionment (SA) or attribution capability, the Particulate Source Apportionment Technology (PSAT), which chemically apportions specific pollutants to boundary conditions and emissions. It uses multiple tracer families to track the fate of primary and secondary PM and gaseous precursors to secondary PM. This approach estimates the contributions from multiple source areas and categories to the spatial and temporal distribution of PM and NO₂ in a single model run²⁹.

CAMx version 6.30, with its PSAT tool, was applied using a two-nesting approach based on a European domain with 0.25 degrees' horizontal resolution and 6 domains of interest centred in each case study, with 0.05 degrees' horizontal resolution (Table 12). Meteorological inputs to the chemical simulations were driven by the meteorological model WRF, forced by ERA-Interim reanalysis data from ECMWF (European Centre for Medium Range Weather Forecast) at 6 hours and 0.75 degrees temporal and spatial resolution respectively. Initial and boundary conditions for the first domain provided by the global chemical model MOZART (Emmons et al. 2010) with a time resolution of 6 hours. Anthropogenic emissions for both

²⁷ ENVIRON. 2016. *CAMx User Guide, Comprehensive Air Quality Model with Extensions, Version 6.30*. Novato, California: Ramboll Environ.

²⁸ Skamarock, William C, Joseph B Klemp, Jimy Dudhia, David O Gill, Dale M Barker, Michael G Duda, Xiang-Yu Huang, Wei Wang, and Jordan G Powers. 2008. "A Description of the Advanced Research WRF Version 3 NCAR/TN-475+STR." Boulder, Colorado, USA.

²⁹ Yarwood, Greg, Ralph E Morris, and Gary M Wilson. 2007. "Particulate Matter Source Apportionment Technology (PSAT) in the CAMx Photochemical Grid Model." In *Air Pollution Modeling and Its Application XVII*, edited by Carlos Borrego and Ann-Lise Norman, 478–92. Boston: Springer.

domains were taken from the TNO-MACC_II European emission inventory³⁰ available at a resolution of 0.125 by 0.0625 degrees, and were speciated into the CB6 chemical mechanism species considered in the CAMx simulation³¹

Figure 28: Map of the WRF-CAMx European domain (within green rectangle) and 6 domains of interest (within blue rectangles) for each cases study location (in red).

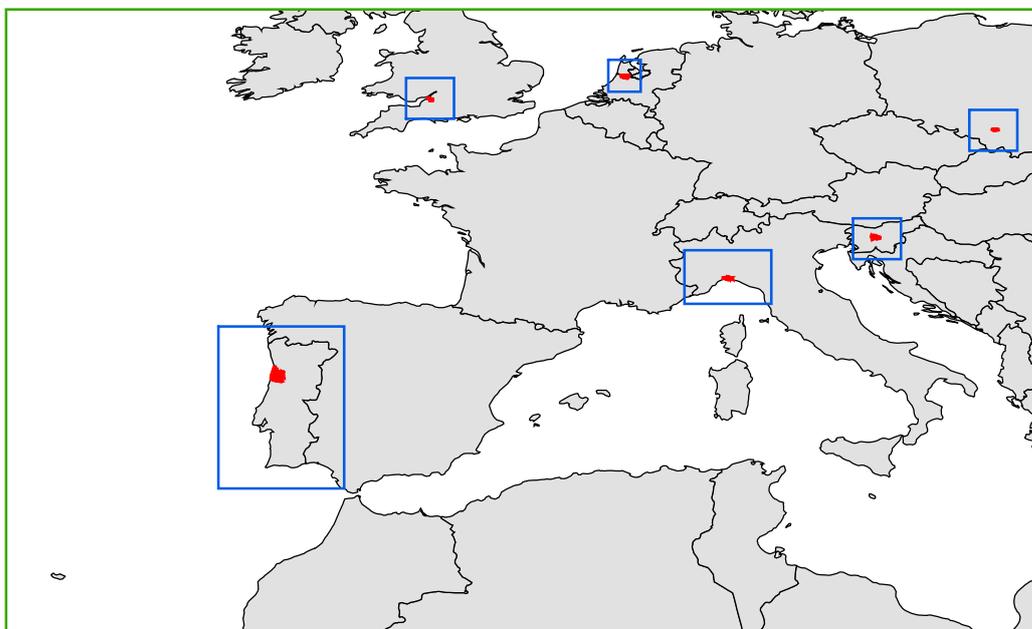


Table 12 – WRF-CAMX domain characteristics.

Case study name	Case study abbreviation	European domain (0.25° resolution)		Case study domain (0.05° resolution)	
		Lon (°)	Lat (°)	Lon (°)	Lat (°)
Bristol	BRS	-19.00 to 21.00	30.50 to 55.50	-3.50 to -1.75	50.75 to 52.25
Amsterdam	AMS			4.25 to 5.50	51.75 to 53.00
Ljubljana	LJB			13.75 to 15.50	45.25 to 46.75

³⁰ Kuenen, J. J. P., A. J. H. Visschedijk, M. Jozwicka, and H. A. C. Denier van der Gon. 2014. "TNO-MACC_II Emission Inventory; a Multi-Year (2003-2009) Consistent High-Resolution European Emission Inventory for Air Quality Modelling." *Atmospheric Chemistry and Physics* 14 (October): 10963–76. <https://doi.org/10.5194/acp-14-10963-2014>.

³¹ Yarwood, Greg, Jaegun Jung, Gary Z Whitten, Gookyoung Heo, Jocelyn Mellberg, and Mark Estes. 2010. "Updates to the Carbon Bond Mechanism for Version 6 (CB6)." In *9th Annual CMAS Conference*. Chapel Hill, NC.

Sosnowiec	SOS			18.25 to 20.00	49.50 to 51.00
Liguria	LIG			7.25 to 10.50	43.50 to 45.50
Aveiro	CIRA			-10.75 to -6.00	36.25 to 42.50

The PSAT application requires the definition of source groups to be tracked and thus the input of extra emission files for each of the groups to be considered. Based on the national emission inventories of each case study^{32,33,34,35,36,37} and on the emission sources, the main sectors contributing to PM and NO₂ emissions in the year under study (2010), that are common to all case studies, are: (i) residential and commercial combustion; (ii) road transport; and (iii) industry. In this sense, emissions were splitted into these activity sectors in order to evaluate the individual contribution of each source sector to the air quality in the selected urban areas through the PSAT application, defined as receptor areas.

6.2 Air quality modelling at urban scale

The second-generation Gaussian model URBAIR was setup and run at urban scale for the computational domains over the urban areas of the six case studies. The baseline simulations were performed for the full-year using the meteorological vertical profiles from the WRF-CAMx system and the emissions available on the ClairCity emissions database including i) point sources with the large industry emissions, as well as the point sources from ships in the case study of Amsterdam and Liguria, ii) line sources with the road-traffic emissions, and iii) area sources covering the residential, commercial and industrial emissions

³² Dębski, Bogusław, Anna Olecka, Katarzyna Bebkiewicz, Zdzisław Chłopek, Iwona Kargulewicz, Janusz Rutkowski, Damian Zasina, Magdalena Zimakowska - Laskowska, Marcin Żaczek, and Sylwia Waśniewska. 2018. "POLAND'S INFORMATIVE INVENTORY REPORT 2018." Warszawa, Poland: National Centre for Emission Management (KOBiZE) at the Institute of Environmental Protection – National Research Institute.

³³ Logar, Martina, Jože Verbič, and Tajda Mekinda Majaron. 2018. "Informative Inventory Report Slovenia 2018." Ljubljana, Slovenia: Slovenian Environment Agency.

³⁴ Pereira, Teresa Costa, Tiago Seabra, Ana Pina, Paulo Canaveira, André Amaro, and Mónica Borges. 2018. "Portuguese Informative Inventory Report 1990-2016." Amadora, Portugal: Portuguese Environmental Agency.

³⁵ Taurino, Ernesto, Antonella Bernetti, Riccardo De Lauretis, Ilaria D'Elia, Eleonora Di Cristofaro, Marco Cordella, Andrea Gagna, et al. 2018. "Italian Emission Inventory 1990-2016 - Informative Inventory Report 2018." ISPRA-Italian National Institute for Environmental Protection and Research.

³⁶ Wakeling, D, NR Passant, TP Murrells, A Misra, Y Pang, G Thistlethwaite, C Walker, et al. 2018. "UK Informative Inventory Report (1990 to 2016)." London, UK: Ricardo Energy & Environment.

³⁷ Wever, D., P.W.H.G. Coenen, R. Dröge, G.P. Geilenkirchen, M. 't Hoen, B.A. Jimmink, W.W.R. Koch, et al. 2018. "Informative Inventory Report 2018 - Emissions of Transboundary Air Pollutants in the Netherlands 1990-2016." Bilthoven, The Netherlands: National Institute for Public Health and the Environment.

from the IRCI module, as well as the shipping emissions for the case study of Amsterdam and Liguria.

For the transport sector, the ClairCity database provides hourly emissions for weekdays and weekends for each road link. Using a GIS software the links were matched to URBAIR model grid cells. From the typical weekday and weekend we were able to rebuild hourly, every day of the year. The emissions are used as input in URBAIR model, in $\text{g.m}^{-2}.\text{s}^{-1}$. For the commercial and residential sector the total annual emissions were disaggregated by applying a temporal profile. The temporal profile consists of hourly shares which differ from weekdays and weekends but also differs from summer and winter months. By applying this profile, the emissions were disaggregated into $\text{g.m}^{-2}.\text{s}^{-1}$ that will be used as an input to the URBAIR model. For the industrial sector, the annual emission rates were evenly distributed over all days of the year and hours of each day. For industrial sources (area) and industrial sources (point), respectively, into $\text{g.m}^{-2}.\text{s}^{-1}$ and g.s^{-1} . As established by ClairCity, if the annual industrial emission rate is over 100 Mg.year^{-1} it's assumed as an Industrial point sources and if lower as area sources.

The air quality simulations were performed for the entire year in an hourly basis considering the meteorological conditions for each hour. We have performed a source apportionment analysis for the baseline year, for the cells where the maximum value occurs, for the entire domain through a spatial average of the total concentrations and the concentrations from each sector. In addition, we have also performed the SA analysis for the cells corresponding to the location of the measurement points. For the SA the background was not accounted. The simulations were performed separately for each emission sector, allowing the application of a source apportionment analysis. The final air quality maps result from the simulated concentrations with URBAIR model, together with added background concentrations, and a calibration of those against the measurements through an adjustment procedure. Each case study had different data availability but also presented some methodological particularities which will be mentioned later for each city. As example, for the background, for Bristol and Amsterdam a more detailed background data was available and for the other cities the background was obtained from CAMx.

7 Health

The burden of disease associated with ambient air pollution is estimated by relating air concentrations to health outcomes. Hence, the relative risk (RR) at a given exposure level can be specified as follows:

$$RR = \exp [\beta * (C_i - C_0)] \quad [2.7.1]$$

where, C_i is the concentration level the population is exposed to in grid cell i , C_0 is the baseline concentration, and β is the coefficient of the concentration response functions (CRF). CRF: the estimated effect of a pollutant on the health outcome based on health studies and is typically described as change in incidence per unit concentration (UC) for those at risk. Table 2.7.1 shows the recommended CRFs for mortality with 95% confidence interval (CI), including the baseline concentration taken into consideration when calculating the health outcomes for each air pollutant. The baseline concentration is a concentration below which no health effects are expected.

Table 2.7.1: concentration-response coefficients for mortality with 95% confidence interval (CI)

Pollutant	Risk ratios for mortality		
	Value [per 10 $\mu\text{g}/\text{m}^3$]	Type	Reference
PM _{2.5}	1.062 (95 % CI 1.040-1.083) No threshold	All-cause (natural) mortality in ages above 30 (ICD-10 codes A00-R99).	WHO (2013)
PM ₁₀	1.04 (95% CI, 1-1.09) No threshold	All-cause (natural) mortality in ages above 30 (ICD-10 codes A00-R99).	Beelen et al. (2014)
NO ₂	1.055 (95 % CI 1.031-1.08%) Threshold 10 $\mu\text{g}/\text{m}^3$	All-cause (natural) mortality in ages above 30 (ICD-10 codes A00-R99).	WHO (2013)

Assuming an exponential behaviour, β can be estimated based on the CRF:

$$CRF \text{ per } UC = \exp(\beta * UC) \quad [2.7.2]$$

where CRFs applied are described in Table 2.7.1 and $UC = 10 \mu\text{g}/\text{m}^3$.

The contribution of a risk factor to a disease or a death can be estimated by means of population attributable fraction (PAF). PAF is defined as the proportional reduction in population disease or mortality that would occur if exposure to a risk factor were reduced to an alternative ideal exposure scenario (e.g. concentrations equal to C_0). If the population is exposed to a single concentration level over the assessed period, PAF can be calculated from the relative risk, for every grid cell i , as follows:

$$PAF_i = \frac{RR-1}{RR} \quad [2.7.3]$$

The health impact assessment focuses on estimating premature deaths (PD) and years of life lost (YLL) as mortality-related health outcomes. PD are deaths that occur before a person reaches an expected age, thus considered to be preventable if their cause is eliminated. The so-called expected age is usually the life expectancy for a country typically stratified by sex and age. the PD metric is estimated assuming the baseline incidence as the crude death rates by sex (s) and age (a) $CDR_{a,s}$:

$$PD = \sum_{a,s} CDR_{a,s} * PAF * Pop_i \quad [2.7.4]$$

Where Pop_i is the concentration at grid cell i .

To estimate the CDR for the population aged above 30, the following steps were taken:

- 1) Population pyramids data is published by the United Nations (UN) and is accessible via <https://population.un.org/wpp/Download/Standard/Population/>, Tables F15-2 and F15-3. Population pyramids data is available for 5-year age intervals from 0-4 to 100+ years old. Total population ($Tpop$) for a country is calculated by summing population pyramids data for females and males.
- 2) All-cause mortality data is published by the World Health Organisations (WHO) and is accessible via http://apps.who.int/healthinfo/statistics/mortality/causeofdeath_query/start.php. All-cause mortality data is available for each country and sex in 5-year age interval, from 0-4 to 95+, for different years. For the calculations, choose the data for the current year and the country where the city is located at. The data should be collected between age interval 30-34 to 95+ interval for each sex. Because only natural causes are taken into consideration, only ICD codes A00-R99 should be included in the data. The CDR for each age interval a and sex s will be then re-calculated by:

$$CDR_{a,s} = \frac{CDR_{a,s}}{Tpop/1000} \quad [2.7.5]$$

YLL is defined as the years of potential life loss due to premature death. It is an estimate of the average number of years that a person would have lived if the person would not have died prematurely. YLL considers the age at which death occurs and is greater for deaths at a younger age and lower for deaths at an older age (Murray, 1996). It gives, therefore, more nuanced information than the number of PD alone. YLL is determined by relating PD with life expectancy (LE) by sex and age, for every grid cell i :

$$YLL = \sum_{a,s,i} PD \cdot LE_{a,s} \quad [2.7.6]$$

$LE_{a,s}$ is the average time a person is expected to live, based on the year of their birth, their current age and sex. This statistical measure is typically available from demographic datasets.

Life expectancy data is required for estimating YLL (Equation 2.7.6). Life expectancy by age and sex must be calculated based on 'Life expectancy at exact age' and 'Average number of years lived' for the population over 30 years old. The following steps were taken:

- 1) Extract life expectancy at exact age data is published by the UN and is accessible via <https://population.un.org/wpp/Download/Standard/Mortality/>, Tables F16-2 and F16-3. Life expectancy at exact age data is available every 5 years from age 5 up to 100+ per country; each table contains sex-specific data. The data for the country should be collected between age 30 and 100 for each sex.
- 2) Average number of years lived, broken down by age cohort and sex, and country is published by the UN and is accessible via <https://population.un.org/wpp/Download/Standard/Mortality/>, Tables F17-2 and F17-3. Average number of years lived data is available for 5-years interval from interval 5-9 to 80-84, and an 85-100 interval per country; each table contains sex-specific data. The data for the country should be collected between age interval 30-34 to 85-100 interval for each sex.
- 3) Average age at death data for 5-years interval is then calculated by summing the lower limit of the age interval ($ageMin_{a,s,c}$) and UN estimate for the average number of years lived ($AvgY_{a,s,c}$) at the correspondent age interval a , sex s and country c by applying the following WHO recommendations (WHO, 2020) :

$$ageD_{a,s,c} = AvgY_{a,s,c} + ageMin_{a,s,c} \quad [2.7.7]$$

- 4) Average life expectancy corresponding to the average age of death for age interval a , for sex s and country c ($LEcal_{i,s,c}$), is determined by the following formulation:

$$LEcal_{a,s,c} = LE_a + \frac{(ageMax_{a,s,c} - ageD_{a,s,c}) * (LE_{a,s,c} - LE_{a+1,s,c})}{ageDMax_{a+1,s,c} - ageDMax_{a,s,c}} \quad [2.7.8]$$

where higher limit of the age interval ($ageMax_{a,s,c}$), $LE_{a,s,c}$ is the Life expectancy at exact age a , sex s and country c .