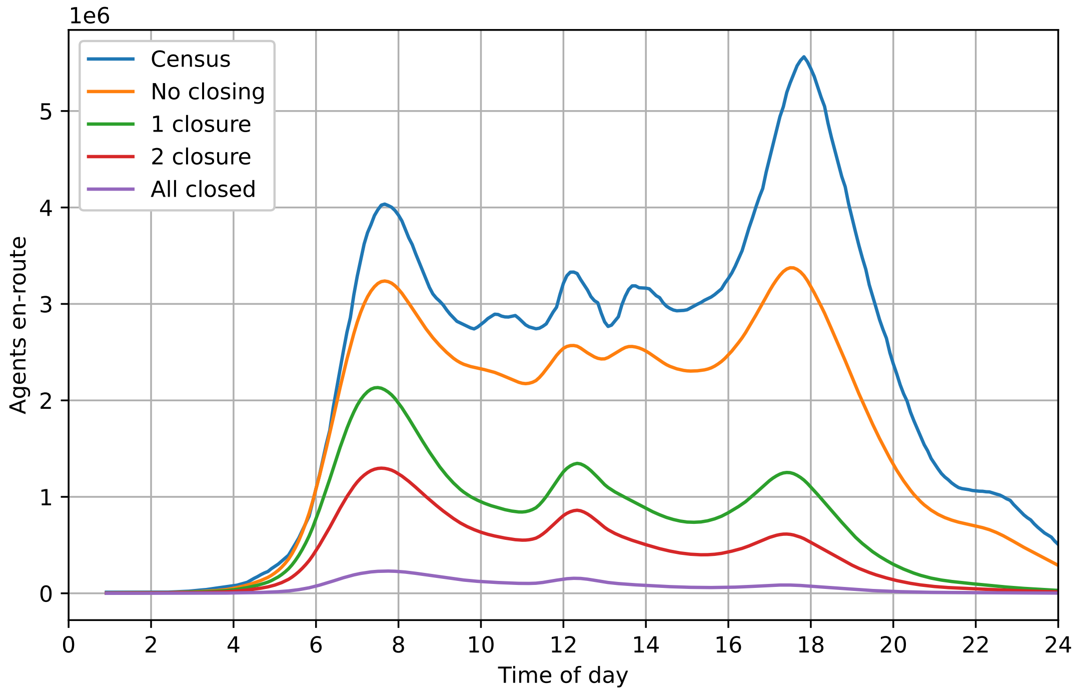

Holistic, integrated generation of daily-activity plans for Switzerland: from population synthesis to trip generation

Aleksandr Saprykin

Marcello Marini

Ndaona Chokani

Reza S. Abhari

ETH Zurich

May 2020

STRC | 20th Swiss Transport Research Conference
Monte Verità / Ascona, May 13 – 15, 2020

ETH Zurich

Holistic, integrated generation of daily-activity plans for Switzerland: from population synthesis to trip generation

Aleksandr Saprykin
Laboratory for Energy Conversion
ETH Zurich
Sonneggstrasse 3, 8092 Zürich
phone: +41 44 632-35-63
saprykin@lec.mavt.ethz.ch

Marcello Marini
Laboratory for Energy Conversion
ETH Zurich
Sonneggstrasse 3, 8092 Zürich
phone: +41 44 632-79-72
marini@lec.mavt.ethz.ch

Ndaona Chokani
Laboratory for Energy Conversion
ETH Zurich
Sonneggstrasse 3, 8092 Zürich
phone: +41 44 632-23-87
chokani@lec.mavt.ethz.ch

Reza S. Abhari
Laboratory for Energy Conversion
ETH Zurich
Sonneggstrasse 3, 8092 Zürich
phone: +41 44 632-26-91
abhari@lec.mavt.ethz.ch

May 2020

Abstract

Detailed and realistic input travel demand are required in agent-based transportation simulators in order to generate high-fidelity outputs. However, most prior work has focused separately on either population synthesis, the process of travel demand generation, or modal choice models. This paper presents a holistic and integrated approach that integrates the different elements – population synthesis, job assignment, car ownership and mode choice models, and the trip generation procedure – in order to generate a realistic synthetic population and microscopic daily-activity plans for the large-scale scenario for the entire population of Switzerland. The generated Swiss travel demand (about 5 million agents) is then executed for the entire transportation network of Switzerland using our in-house developed, GPU-accelerated mobility simulator, GEMSim, which is integrated within our agent-based simulation framework, EnerPol, to simultaneously address the impacts of transformations in mobility on energy infrastructure, land use and urban development. The insights from the simulated scenarios, and the needs for what additional data may be useful to gather are presented. It is also demonstrated how the outcomes from the high resolution framework can be used by policy makers and planners at the city, cantonal, regional and federal levels.

Keywords

agent-based modelling; population synthesis; daily-activity model; covid-19; discrete choice

1 Introduction

Agent-based models are increasingly applied in many sectors over the last decades, including transportation sector (Bonabeau, 2002). One of the main drivers for the increased usage of agent-based models is their ability to model complex behaviour on a disaggregated level of detail. This is in contrast to four-step models which have been used in the transportation sector since the 1950s. Four-step models, while being simple and computationally efficient, are unable to account for the dynamics of individual behaviour, as well as person-person interactions in transportation systems. With the development of new transportation modes such as car-sharing or ride hailing, these limitations are even more important. Moreover, it is a formidable challenge to apply four-step models in scenarios with mobility-as-a-service (MaaS) platforms. While agent-based models are well suited for these emerging challenges in the transportation sector, the development and application of agent-based models requires even bigger datasets and more detailed information about the actors and environment that are to simulated. Typically, the accuracy of an agent-based simulation directly depends on both the quality and the quantity of available data. Moreover, as the operation of transportation systems involves also the complex behaviour of people, the more widespread application of agent-based models is not trivial.

As demand and supply drive any transportation system, the synthetic population is a cornerstone in any agent-based transportation model. Both individual attributes as well as behavioural patterns, or daily activities, must be described. Typically, the synthetic population is derived from a variety of datasets, ranging from mobility census to statistical data of the local area to be studied. These datasets are often in incompatible formats, with different levels of resolution, and may refer to different years. Moreover, these raw datasets must often be pre-processed and/or converted into a customized representation in order to be analysed or visualized. Furthermore, the generation of agent-based travel demand is typically divided into the two steps of synthesis of the population and generation of trip chains, which requires additional effort to make the two steps compatible. Another challenge arises when something has to be changed either in the input data or in the steps, for example in the evaluation of scenarios, and the whole process of generating the travel demand must be repeated in a reproducible way.

In this paper, we propose a unified modelling pipeline (that is, a sequence of executed models) in order to automatize the generation of agent-based travel demand. The modelling pipeline, when executed, reads the required input data, and transforms this into an agent-based travel demand that is ready-to-use in an agent-based mobility simulator. This approach both simplifies the maintenance of agent-based models, and allows one to easily re-generate travel demand for specific situations, such as the spread of a viral disease which is very closely linked to the behaviour of people. Furthermore, this approach makes the use of agent-based models easier for non-experts, such as policy makers, who may want to just specify a set of input parameters for a scenario, and simulate the scenario, without going into the details of the generation of travel demand.

The unified modelling pipeline is implemented within our agent-based EnerPol simulation framework. The EnerPol simulation framework provides a platform to run "what-if" scenarios on a scale from either a single country to a continent. The framework includes a set of integrated and coupled models which allow a user to evaluate demographics and migrations flows (Marini *et al.,* 2019), transmission and distribution power grids (Plagowski *et al.,* 2018), gas pipelines, impacts of policies on urban transformations and mobility behaviour (Pagani *et al.,* 2019). Currently, the proposed modelling pipeline is implemented for a number of regions including Europe, North America, East Asia and sub-Saharan Africa. This paper focuses on the case of Switzerland, in order to demonstrate how the whole process of agent-based travel demand generation is automated and applied to run a large-scale mobility scenario, including people behaviour, during an epidemic.

2 Background

The synthesis of population agents that realistically represent the actual population in the area of interest is the first requirement for the generation of travel demand and the subsequent mobility simulations.

Müller and Axhausen (2010) summarise different approaches that are used. The most common approach (Bowman, 2004) is comprised of two main stages: (i) population fitting, and (ii) allocation. In the first stage, fitting, a reference sample of agents, typically derived from census data, is fitted to aggregated constraints, for example, the total population. The most common fitting approach is the iterative proportional fitting (IPF) procedure that was first described by Deming and Stephan (1940) and that estimates the distribution of control variables based on the reference sample. This method is used in recent population synthesizers, such as ALBATROSS (Arentze *et al.,* 2007) and PopGen (Ye *et al.,* 2009). The second stage, allocation, disaggregates the fitted population amongst individual population agents and households, which are selected from the reference sample. In some cases, the geographic placement of the households is also refined in this stage. The different methods of allocation include an altered selection probability (Auld *et al.,* 2010), a conditional Monte Carlo approach (Pritchard and Miller, 2009), or a deterministic selection (Srinivasan and Yathindra, 2008). The main limitation of these approaches is that the quality of the resultant synthetic population depends directly on the quality of the reference sample; this sample is very often of limited size and detail for reasons of personal privacy in census data. In the present work we propose a different approach that combines registers for dwellings and commercial activities with aggregated population registers. The population agents are synthesized through a series of models with the goal of realistically capturing the characteristics of the actual population, but without the need for using a reference population sample. Thus, this approach is more advantageous as it can be applied in many geographic locations, where the data required to generate a reference population sample are unavailable.

The synthetic population is the basis for disaggregated travel demand generation models. Activity-based models (Kitamura, 1988, Algers *et al.,* 1995, Damm, 1983) derive travel demand based on an integrated overview of people and households. These models are based on the demand for socio-economic activities, where individual agents have spatio-temporal constraints (Torsten, 1970), and move between different locations incurring time and cost of travel. Activity-based models place emphasis on the fact that participation in certain activities generates travel demand and complex interactions. Agents typically try to maximize their utility function (Habib, 2011) by building and scheduling a sequence of trips between the locations of activities. Many daily-activity models are tour-based, whereby an agent starts and ends his/her day at the same location, usually home. Rule-based algorithms (Pendyala *et al.,* 1998, Arentze *et al.,* 2000) are most widely applied in activity-based demand generation, but it is noted that these algorithms tend to approximate the process of activity scheduling (Roorda *et al.,* 2008). However, as activity-based models are used to model the socio-economic behaviour of individuals, these models are difficult to implement and calibrate; hence many such models are applied only at a regional or local level (Davidson *et al.,* 2007, Hatzopoulou and Miller, 2010). Another approach to generate travel demand for agent-based transport simulation models is to sample trip chains directly from microcensus data (Viegas and Martínez, 2010). In this approach, trip chains are sampled from statistically similar respondents and adjusted in time and space. The main advantage of this approach is its simplicity, as few assumptions are made regarding the process of activity scheduling, and the approach can be applied at the scale of census data. This paper integrates the latter approach to generate trip chains for synthetic agents, thus avoiding the need to develop and calibrate econometric models of people's behaviour. Only mobility microcensus data is required to apply the unified modelling pipeline that we describe below to another geographic area. Given that many countries conduct a mobility census on a regular basis, the unified modelling pipeline can be applied to a wide range of cases.

3 Methodology

3.1 Modelling pipeline

The structure of the modelling pipeline is shown in Figure 1. The pipeline is organized as a sequence of steps, which are executed in order to transform raw input data into daily-activity plans of the agents for a mobility simulator. The pipeline can be split into two main parts: a population synthesis model and a daily-activity model, where the daily-activity model takes the output of the population synthesis model as an input.

The whole modelling pipeline is implemented in Python, and, for reasons of improved performance, parts of the population model are accelerated with GPU using the *Numba* framework (Lam *et al.,* 2015).

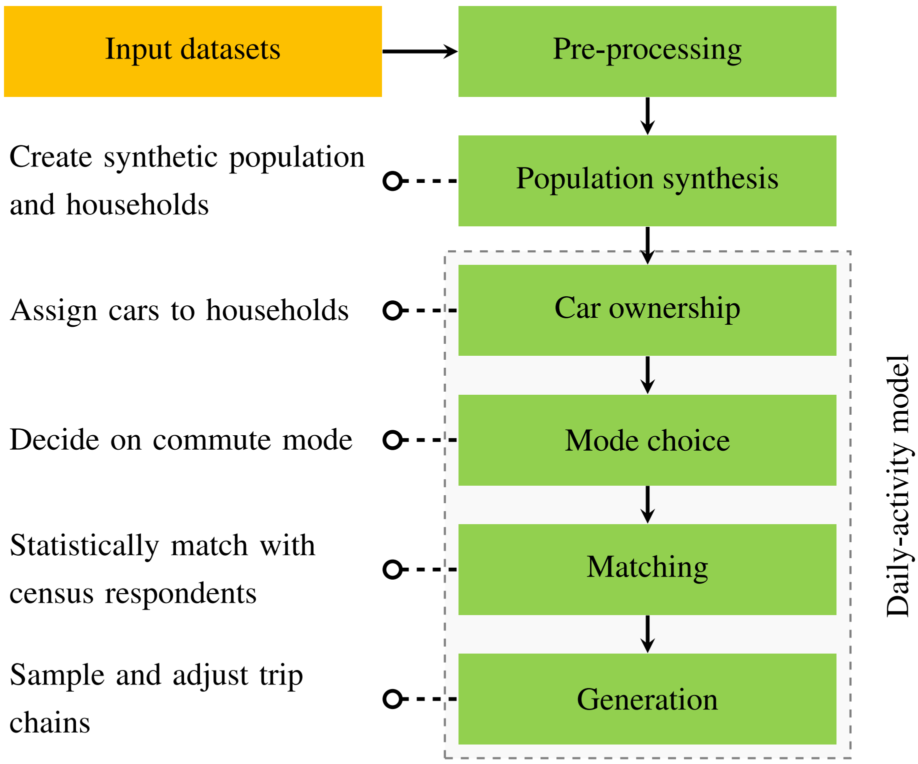

Figure 1: Structure of the main modelling pipeline.

While the modelling pipeline was developed specifically for Switzerland, the main structure is flexible and can be adapted for other geographic areas.

3.2 Population synthesis

The structure and data flow of the population synthesis is presented in Figure 2. In our in-house agent-based simulation framework, EnerPol, four typologies of geo-referenced agents are modelled. The agent typologies are:

- Population agents, representing individual persons who perform their daily activities, and travel between their locations of activities;
- Household agents, representing individuals who are grouped into a family unit, and who live in the same dwelling;
- Job agents, representing available and occupied jobs that are assigned to agents that have a matching skillset; and
- Dwelling agents, representing available and occupied residential locations, in which households live.

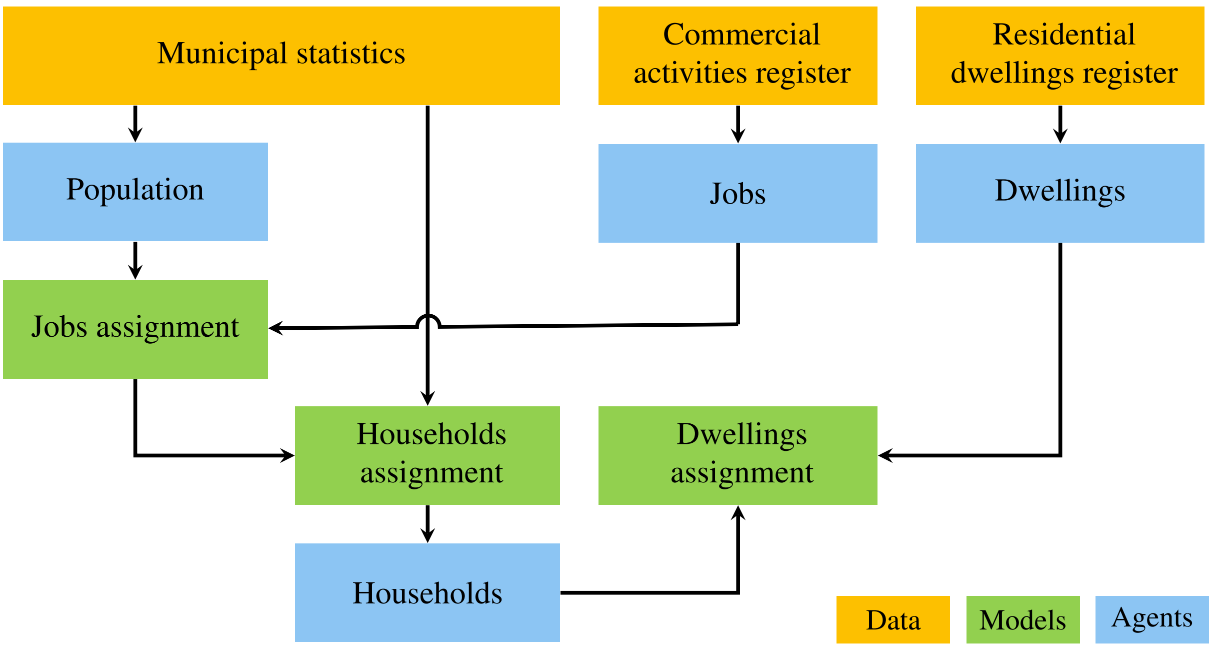

Figure 2: Structure of the population synthesis model.

Table 1 summarises, in terms of independent and dependent variables, the characteristics of the four typologies of agents; the independent variables are derived from data sources, whereas the dependent variables are modelled as the agents are generated.

Agent	Independent variables (t=0)	Dependent variables
Person	age, sex, employment status, education, municipality of dwelling	household ID, job ID, job location, income, marital status, number of kids, job distance, county of origin
Household	number of adults, number of children	IDs of household members, years of marriage/partnership, ages of adult household members, income of household, dwelling ID
Job	municipality of activity, coordinates of activity, salary, NOGA classification, economic sector, skill level	
Dwelling	size, rooms, rental price, tax rate per household typology and income, distance from amenities	

Table 1: Characteristics of agents that are modelled in the agent-based EnerPol simulation framework.

3.2.1 Population agents

The generation and linking of the synthetic population for the reference year can be subdivided into five steps:

1. Individual population agents are generated, starting from demographic data that characterises population agents in terms of age, sex, and education level, and that are extrapolated as a function of age from the Adult Literacy and Life Skills Survey for Switzerland (Swiss Federal Office for Statistics, 2003).
2. Individual job and dwelling agents are generated based on the Register of Commercial Activities (Swiss Federal Office for Statistics, 2013) and the Register of Residential Buildings and Dwellings (Swiss Federal Office for Statistics, 2015), while the costs of dwellings are assigned based on the data taken from websites of online listings.
3. Population agents are linked to job agents in the vicinity of the municipality as a function of the distribution of skills; the commuting matrix, described in the mobility microcensus (Bundesamt für Statistik / Bundesamt für Raumentwicklung, 2015), is used to define the

probability of working in a specified district.

4. Individual agents are linked in households, starting from the distribution of household typologies in the municipality (Swiss Federal Office for Statistic, 2010). Partners are matched assuming minimization of difference, with compatible sex and sexuality, while children are assigned to parents as a function of the age of the female partner.
5. Residential dwellings are assigned to households, such that size and cost of the dwelling matches to structure and income of the household.

It is worthwhile to note that the level of detail in the modelling of population agents exceeds that which is required solely for mobility simulations, as population agents are used within the holistic EnerPol simulation framework.

3.2.2 Residential dwellings

The assignment of residential dwellings to households simulates the decision-making of a household when a dwelling is chosen. Based on statistics presented in Luzerner Zeitung (2014) that cited (i) an incompatible dwelling size, (ii) an excessive distance from the workplace, and (iii) an unsatisfactory dwelling price, as the main reasons for relocation, the characteristics of households are matched to the characteristics of available dwellings. Thus, the choice of a new dwelling is made by evaluating the following characteristics of available dwellings:

- distance from location of job;
- cost of dwelling, including rent and taxes as a function of family structure and income;
- number of rooms;
- distance from schools (for families with kids);
- distance from public transport; and
- availability of services

For each household, a sub-set of dwellings are rank-ordered based on a compatibility score that is a weighted average of the above-listed characteristics of the dwelling. The compatibility score, S (Equation 1), that is a weighted average of the scores, s_i (Equation 2), of different characteristics, i, of the dwelling available to the household. The characteristics and their different weights, w_i, are defined for each typology of households. The scores of an individual parameter s_i (Equation 2), are based on the expected and optimum values, e_i and o_i (for example, distance to public transport and 50 m, respectively), and the elasticity l_i (for example, ± 200 m) which is the variance of the Gaussian distribution. Thus, when the difference between the expected and optimum values is relatively large, the score is then relatively low.

$$S = \frac{\sum w_i * s_i}{\sum w_i} \tag{1}$$

$$s_i = exp(-\frac{(e_i - o_i)^2}{2 * l_i^2}) \tag{2}$$

The respective values are defined based on a survey of experts Fahrlaender and Partner AG. (2017). The weights account for characteristics of the households and reasons for relocation, such as household structure, number of adults, number of kids, total income, location of jobs, and average age of heads of the household. The scoring parameters are summarised in Table 2, and the weights for different household typologies are presented in Table 3.

Parameter	**Optimum value** [o_i]	**Elasticity** [l_i]
Distance from job [km]	0	30
Dwelling cost (including rent, and taxes) [-]	$1/3 * (Income - Taxes)$	$20\% * o_i$
Number of rooms [rooms]	Number of People +0.5	1
Distance from schools [m]	400	1000
Distance from public transport [m]	400	1000
Distance from services [m]	400	1000

Table 2: Summary of scoring parameters that are used in the assignment of residential dwellings for households that relocate.

In Table 2, the elasticity of the distance to job is based on a Europe-wide long distance commute, and the approximately 45 minute travel time is based on the results of the extended Europe LTD survey (SD Works, 2018) and Mobility and Transport Microcensus of Switzerland (Bundesamt für Statistik / Bundesamt für Raumentwicklung, 2015). The dwelling costs are based on the 'one-third-of-income' rule (Elkins, 2018), that is generally accepted as the share of net income that can be spent on rent, and is supported in several surveys (McCarthy, 2017, Terrazas, 2015, Hockaday, 2019, Elkins, 2018, Tagesanzeiger, 2017); thus, the elasticity can be estimated. The number of rooms is based on the optimal dwelling sizes accounted for in the Dwellings Evaluation System of the Federal Office for Housing (Federal Office for Housing / Bundemsamt für Wohnulgswesen, 2000). The other optimum values in Table 2 correspond to half of 400 m that is commonly referred as comfortable walking distance (Walker, 2011), while the elasticity of 1000 m is commonly considered a far walking distance by architects and real-estate developers (based on the pool of interviewed experts).

Household typology	**Singles and couples**	**Families**	**Retirees**
Distance from schools [%]	0	30	0
Distance from public transport [%]	50	30	20
Accessibility to services [%]	30	20	60
Quality of area [%]	20	20	20

Table 3: Summary of weights, by households' characteristics that are used in the dwelling choice model to rank-order optimum dwellings. The values are based on segmentation of demand survey of Fahrlaender and Partner AG. (2017).

3.3 Daily-activity model

The structure and data flow of the daily-activity model is shown in Figure 3. The model incorporates other sub-models including discrete choice models for car ownership and transport mode, and input datasets together with the output from the population synthesis model, which are propagated sequentially until the daily-activity plans are generated.

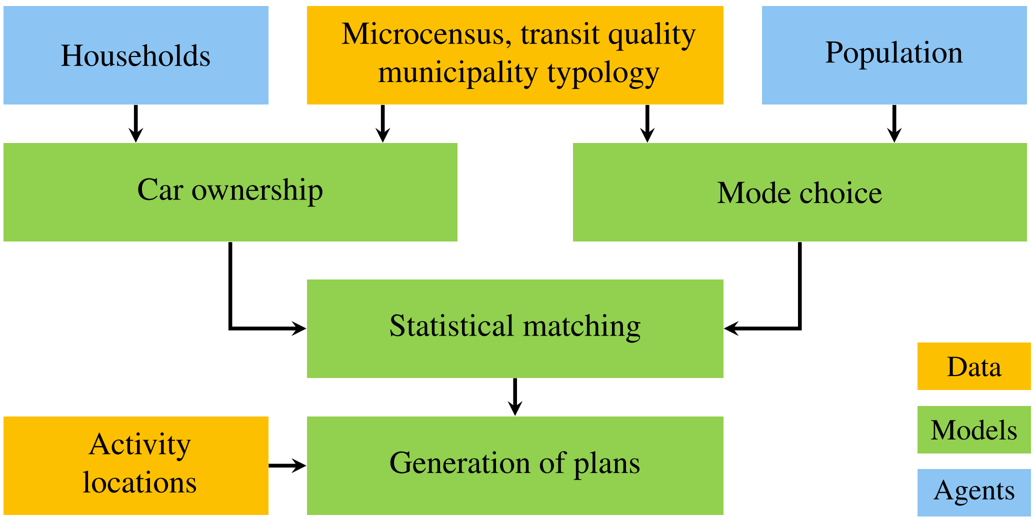

Figure 3: Structure of the daily-activity model.

3.3.1 Pre-processing

The following datasets are inputs to the daily-activity model:

- Synthetic population
- Mobility and Transport Microcensus (MTMC) (Bundesamt für Statistik / Bundesamt für Raumentwicklung, 2015)

- Administrative borders (Bundesamt für Statistik, 2018)
- Municipality typology: urban, sub-urban, rural (Bundesamt für Statistik, 2019)
- Public transit quality map (Bundesamt für Raumentwicklung, 2019)
- Postal code boundaries (Bundesamt für Landestopografie, 2019)
- Locations of places of activity from OpenStreetMap (OSM) (OpenStreetMap, 2020)
- Car register (MOFIS) (Bundesamt für Strassen, 2017)

The microcensus provides information about the personal attributes and the detailed travel behaviour of 57 090 participants in Switzerland. Both the synthetic population and microcensus datasets contain comparable attributes of agents as required in the daily-activity model:

- People: age, gender, job location
- Households: location, size, typology, income

The travel behaviour in the microcensus is described as a set of individual daily trips (193 880 in total). In the daily-activity model the following trip properties are used: start and end locations, departure and arrival times, transport mode, travel purpose, travel distance. The microcensus data can be filtered as either a typical working day or weekend; thus variations in behaviour within a week can be distinguished. In this paper, a typical working day is used as an example.

Municipalities are classified into one of three categories based on the Gemeindetypologie (Bundesamt für Statistik, 2019) classification, which accounts for spatial attributes such as population density and accessibility. This classification is used to estimate car ownership and transport mode choice models. Figure 4 shows the classification of municipalities in Switzerland that is used in the daily activity model, while the mapping of typology from the Gemeindetypologie classification is shown in Table 4.

The public transit quality map (ÖV-Güteklassen) specifies four classes (A, B, C and D) of public transit service quality depending on the proximity of stops, the variety of transport modes in the area, and the frequency of service. The public transit quality map for Zurich is shown in Figure 5. Areas outside these four classes are considered to not have public transit quality.

As agents travel to performs activities throughout a day, the locations of places of activity are required. These locations are taken from OSM (Geofabrik) and their types are mapped to the purpose of the trip from the microcensus. The mapping is presented in Table 5. Trips with a business purpose include all available types of activity. In total, 123 577 places of activity are downloaded from OSM.

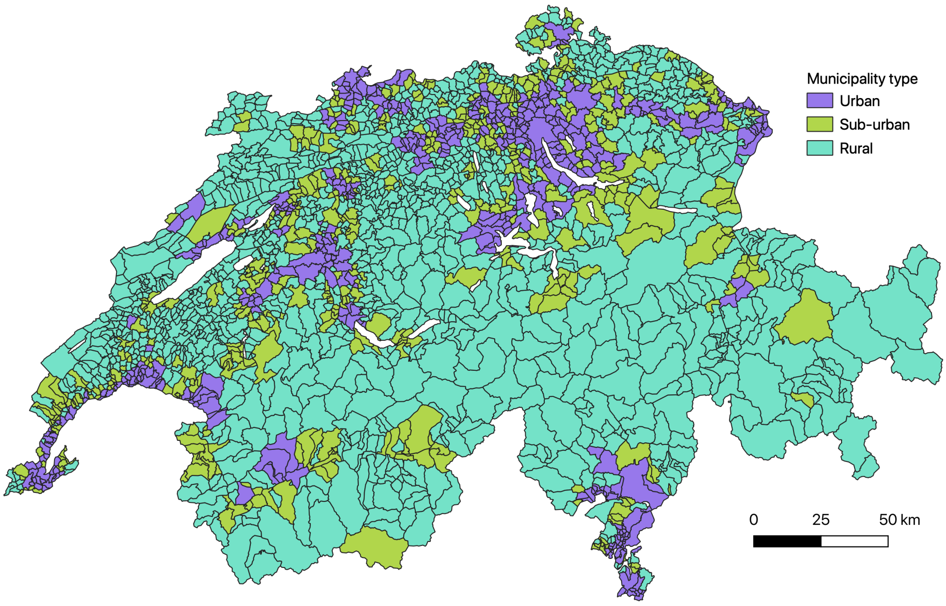

Figure 4: Municipality typology based on Gemeindetypologie classification.

Gemeindetypologie (code)	**Municipality typology**
Municipal community of a large agglomeration (11)	Urban
Municipal community of a medium-sized agglomeration (12)	Urban
Municipal community of a small or outside an agglomeration (13)	Sub-urban
Periurban high-density community (21)	Urban
Periurban medium-density community (22)	Sub-urban
Periurban low-density community (23)	Rural
Rural center (good connection) community (31)	Rural
Rural center (local) community (32)	Rural
Rural peripheral community (33)	Rural

Table 4: Mapping of municipality typology from the Gemeindetypologie classification.

The car register (MOFIS) contains detailed information about registered vehicles in Switzerland at the resolution of ZIP-code; this register is used to validate the car ownership model.

After reading, the input data is cleaned. The purpose of cleaning is to remove incomplete samples which cannot be used in discrete choice models or simply do not contain sufficient information to generate a proper daily-activity plan of an agent. Examples of removed samples include trips with loops (start and end in the same location), unknown transport mode, not starting or ending at home location, or respondents without any trips reported. In total, 34 028 trips and 15 335 respondents were removed from the microcensus dataset.

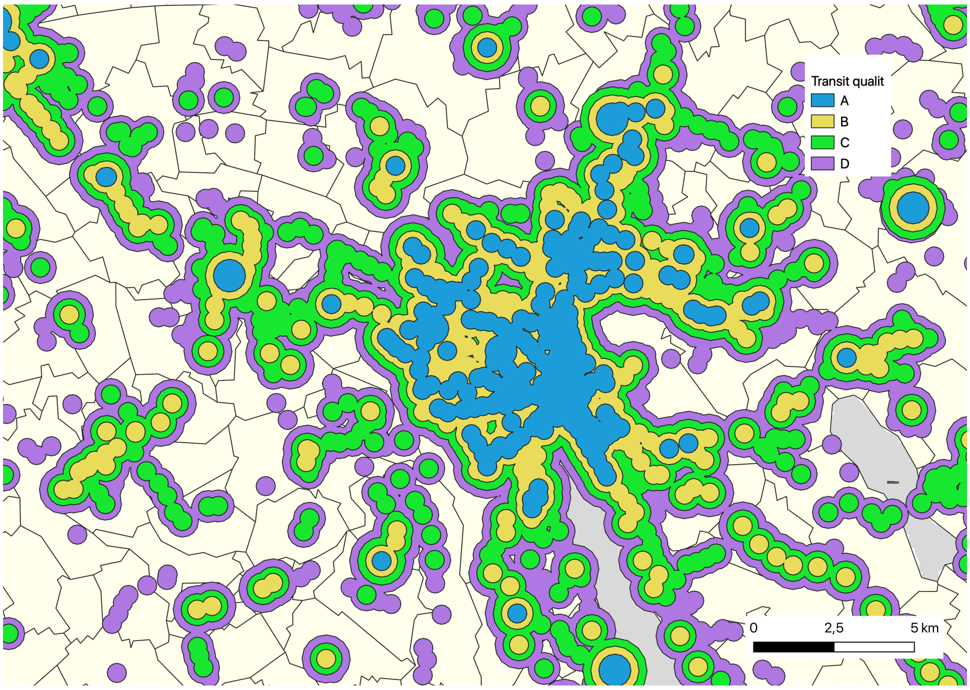

Figure 5: Map of public transit quality in the Zurich area.

Microcensus	**OSM**
Education	school, university, college
Shopping	beverages, convenience, supermarket, mall, kiosk, department_store, bookshop, clothes, outdoor_shop, doityourself, bicycle_shop, sports_shop, butcher, shoe_shop, video_shop, mobile_phone_shop, computer_shop, furniture_shop, jeweller, toy_shop, beauty_shop, gift_shop
Needs	nursing_home, courthouse, bank, post_office, recycling, recycling_metal, recycling_clothes, recycling_glass, recycling_paper, police, atm, kindergarten, hospital, town_hall, doctors, pharmacy, hairdresser, public_building, car_dealership, florist, veterinary, laundry, car_wash, dentist, optician, embassy, car_repair
Leisure	pitch, cinema, pub, fast_food, restaurant, cafe, theatre, sports_centre, park, stadium, bar, picnic_site, library, museum, swimming_pool, playground, nightclub, biergarten, garden_centre, travel_agent, theme_park, arts_centre, zoo, greengrocery, golf_course, food_court, dog_park, community_centre, ice_rink, bakery

Table 5: Mapping of OSM activity types to trip purpose in microcensus.

After cleaning, the synthetic and microcensus households data are merged with municipalities classification and public transit quality map; thus each household is assigned its attributes from municipality-related datasets and the quality of public transit service based on the location of the household.

3.3.2 Car ownership

A car ownership model is used to assign the number of cars owned by each of the synthetic households. The Multinomial Logit (MNL) discrete choice model is applied to synthetic households to estimate the number of cars. The MNL model assumes that random terms are identically and independently (iid) distributed (Gumbel distribution). The probability of choice of a given alternative i for a decision-maker n is defined as follows:

$$P_{in} = \frac{e^{V_{in}}}{\sum_j e^{V_{jn}}} \tag{3}$$

In discrete choice modelling, the probability of each alternative is based on a set of attributes which reflect the cost and benefits of an alternative, and the utility function U is defined as follows:

$$U_{in} = V_{in} + \epsilon_{in} = \beta_i x_{in} + \epsilon_{in} \tag{4}$$

where V_{in} is the deterministic part of the utility function based on the vector β_i of taste parameters and the vector x_{in} of alternative attributes, and ϵ_{in} is the non-deterministic part of the utility function.

Microcensus data is used to estimate the parameters of the MNL model. The following household variables were observed to have a strong impact on the choice: monthly income (CHF), typology, public transit quality and municipality typology. The variables are used as categorical, and the available alternatives are: no car, 1 car, 2 cars and 3 or more cars. In the microcensus, 13 712 respondents did not specify income levels and these samples were excluded from the estimate of the MNL model. The coefficients of MNL model are estimated using the maximum-likelihood method from *statsmodels* package for Python, and the results are presented in Table 6. Alternatives are estimated against the alternative of not having a car.

The results show that all of the variables are statistically significant. Income does not impact a lot the alternative of having a single car, however, a higher income increases the probability of having multiple cars. Households with a single person have a lower probability of having a car, while an increase in household size (including kids) increases the probability of having one or two cars. Public transit quality in the area of the household's dwelling has a very strong impact on the decision to have a car: the higher is the quality of public transit, the more negatively the

Variable (dummy)	1 car	z	2 cars	z	3+ cars	z
Income (2 000–4 000)	0.28	2.98	0.12	0.74	0.03	0.13
Income (4 001–6 000)	0.82	8.56	0.92	5.61	1.00	3.52
Income (6 001–8 000)	1.01	10.13	1.40	8.52	1.66	5.86
Income (8 001–10 000)	1.07	9.90	1.77	10.42	2.18	7.62
Income (10 001–12 000)	1.04	9.03	1.85	10.57	2.41	8.31
Income (12 001–14 000)	1.01	7.47	1.88	9.98	2.60	8.68
Income (14 001–16 000)	0.98	6.80	1.98	10.18	3.00	9.94
Income (> 16 000 CHF)	1.06	7.80	2.22	11.83	3.50	11.82
Household (single)	0.37	3.03	-1.78	-9.63	-3.48	-11.15
Household (non-family)	0.35	2.37	0.17	0.88	-0.63	-2.05
Household (couple)	1.02	7.94	0.69	3.82	-1.03	-3.55
Household (couple+kids)	1.12	8.46	1.08	5.82	0.17	0.60
Household (single+kids)	0.67	4.85	0.08	0.44	-1.21	-3.95
PT quality (A)	-1.98	-19.94	-3.47	-31.72	-4.09	-29.57
PT quality (B)	-1.46	-14.76	-2.45	-23.19	-2.93	-23.59
PT quality (C)	-0.93	-9.59	-1.61	-15.72	-1.90	-16.67
PT quality (D)	-0.44	-4.45	-0.76	-7.37	-1.00	-8.91
Municipality (sub-urban)	0.19	3.79	0.36	6.47	0.52	7.48
Municipality (rural)	0.21	2.83	0.46	5.61	0.74	7.88

Table 6: Results of MNL estimation for car ownership in households.

alternatives of having a car are correlated. In general, households located in sub-urban and rural areas tend to have more than one vehicle.

To validate the car ownership model, car register data at ZIP-code resolution is used. Non-private cars were removed from the register, as well as all types of non-personal vehicles (trucks, agricultural, etc.). Figure 6 shows the relative error of the predicted number of cars at the spatial resolution of ZIP-code compared to the actual number of registered cars. One can note that the model captures the trend in the data very well: most of the high-density areas have a relative error below 10%, some areas have up to a 20% error, and very few areas (mostly close to the border) have a higher error. Most areas with high (>30%) relative errors are located in low-density mountainous region of Switzerland, and many of these areas have in the range of a few dozen to a few hundred registered cars. Finally, one should also account for the fact that the car register gives only an approximate spatial distribution of the actual locations of registered cars. That is, for tax reasons, cars can be registered in one canton but used in another.

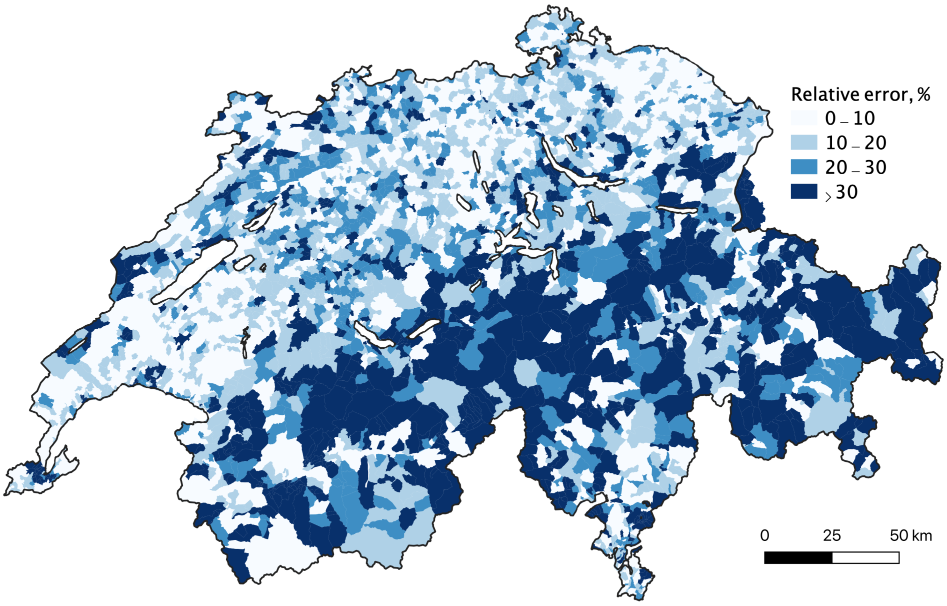

Figure 6: Relative error between predicted car ownership and car register data.

3.3.3 Mode choice

Similar to the car ownership model, a mode choice model is implemented using MNL regression with two alternatives: car or public transit (including walking). The microcensus data is additionally filtered as follows, prior to model estimation:

- Agents below the legal driving age (18 years in Switzerland) are assigned to public transit and excluded.
- Agents living in households without cars are assigned to public transit and excluded.
- Agents who are passengers, not drivers of a car, are excluded.

The coefficients of the MNL model are summarised in Table 7. The alternative of taking public transit is estimated against the alternative of using a car.

The results show that persons older than 25 tend to switch to a car rather than to public transit, while, in general, females prefer to use public transit. Interestingly, persons with high monthly income (more than 12 000 CHF) tend to prefer public transit. This can be explained by the fact that travel in first class is quite comfortable and/or the dwellings of these persons are centrally located. As expected, if there is a higher quality of public transit in an area then the probability of using public transit increases, while the availability of a car in a household reduces the probability of using public transit. Finally, in general, employed persons prefer to use a car over public transit.

Variable (dummy)	Public transit	z
Age (18–24)	-0.37	-0.65
Age (25–44)	-1.85	-3.19
Age (45–64)	-1.73	-2.99
Age (>64)	-1.79	-3.10
Gender (female)	0.55	15.155
Income (2 000–4 000)	0.02	0.16
Income (4 001–6 000)	-0.03	-0.18
Income (6 001–8 000)	0.07	0.45
Income (8 001–10 000)	0.03	0.21
Income (10 001–12 000)	0.26	1.54
Income (12 001–14 000)	0.42	2.36
Income (14 001–16 000)	0.49	2.71
Income (> 16 000 CHF)	0.43	2.44
Employed	-0.55	-10.70
PT quality (A)	1.41	17.970
PT quality (B)	0.89	12.060
PT quality (C)	0.67	9.785
PT quality (D)	0.35	5.222
Cars in household (2)	-1.4392	-32.420
Cars in household (3)	-2.1771	-25.749

Table 7: Results of MNL estimation for transport mode choice.

Figure 7 shows the predicted share of agents who select public transit in Swiss municipalities. In urban areas that have a high quality of public transit people tend to drop a car, while in mountainous and low-density regions, where the quality of public transit is insufficient, the trend is opposite.

3.3.4 Generation

The generation of the agent-based travel demand is comprised of three steps:

- Finding statistically the best donor of an activity chain for each synthetic agent derived from the microcensus dataset.
- Sampling of the locations of activity to match the activity chains relative to the location of homes.

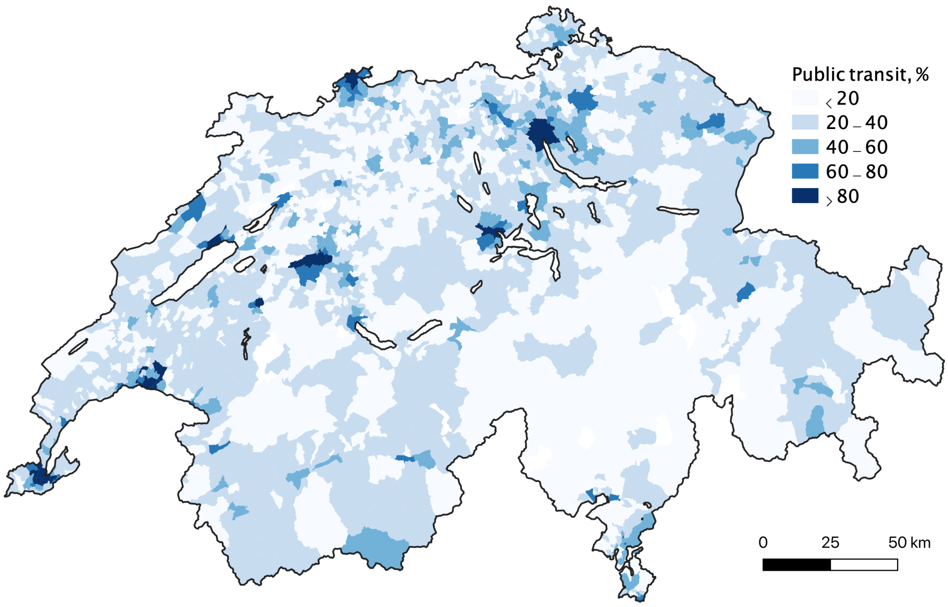

Figure 7: Predicted share of agents who take public transit for daily trips.

- Converting the activity chains into output XML format and writing to a file.

All steps are performed in parallel, whereby one thread handles a block of synthetic agents; thus the model is scalable as the size of population is increased. Each thread processes its own block of agents and writes output to a file. After all working threads are finished, the main thread concatenates the output files into a single output file.

In the first step, two sets of attributes of an agent in the synthetic population are compared with the same sets of attributes of each respondent in the microcensus dataset. The first set contains mandatory attributes (of age class, gender, employment status and commute mode) which must be exactly matched for a respondent from the microcensus in order to be accepted as a donor candidate. The second set contains optional attributes (of income class, household typology, municipality typology and distance from home to work) which should be closely matched. A minimum of 30 donors are required in order for a synthetic agent to be accepted for generation of the plan, otherwise the agent is not used. After comparison, the 30 donor candidates with the best overall match scores are selected, and a final donor is sampled uniformly at random from this pool. In total, all 5 542 305 synthetic agents have been matched using 28 328 donors from the microcensus.

As activity chains from the microcensus contain the specific coordinates of the locations that are visited, these coordinates must be re-sampled from OSM dataset for each agent from the synthetic population based on his/her household and job (if any) locations. A Monte-Carlo type re-sampling of the locations of activity is made for each trip in the chain, and the chain always starts and ends at home:

- Randomly pick a direction in the range of 0–360 degrees.
- Select a point located at the trip distance from the last place of activity in the direction picked at the previous step.
- Find the closest place of activity with the corresponding trip purpose.

When locations of activity for a whole chain are re-sampled, the total travel distance between re-sampled locations is compared with the total travel distance of the donor. If the difference is less than 200 m in the travel distance then the re-sampled chain is accepted, otherwise the re-sampling procedure is repeated for this chain. Currently, the number of re-sampling attempts is 200, as this provides a good compromise between the accuracy of re-sampling and the runtime performance of the model. If after 200 iterations a solution is not found, the best re-sampled activity chain is accepted. The timeline of a re-sampled activity chain is shifted randomly within a 30-minute interval. Finally, a re-sampled activity chain is written to an output XML file.

3.4 Performance

The runtime performance of the unified modelling pipeline is evaluated on a GPU computing node that is equipped with 4x P100 GPUs, 2x Intel Xeon E5-2680 v4 CPUs clocked at 2.4 GHz and 256 GB of RAM. Each CPU has 14 physical cores and can run up to 28 threads in parallel when a physical CPU core represents two logical cores in the system. The runtime performance of the pipeline is shown in Figure 8. Depending on the number of CPU cores used, a simulation takes 4 to 5 hours and about 10 GB of host RAM to generate travel demand for Switzerland from raw input data. When using more than 20 threads the performance of the pipeline does not improve a lot and may even degrade. One reason for the degraded performance could be the lack of multi-threading (address space is shared among threads within a single process) in Python where multi-processing (each process has dedicated address space) is used instead. Multi-processing may lead to data segmentation and less efficient utilization of CPU cache. Moreover, an operating system may schedule resources less efficiently when dealing with the large number of processes (more expensive context switch). Another explanation is that the number of physical CPU cores becomes a limiting factor because a single physical CPU core has performance drop when running more than one thread in parallel. Finally, the whole pipeline is not parallelized with CPU cores. For example, the population synthesis is parallelized by using GPUs and takes about 2.5 hours to run; this runtime does not depend on the number of available CPU cores.

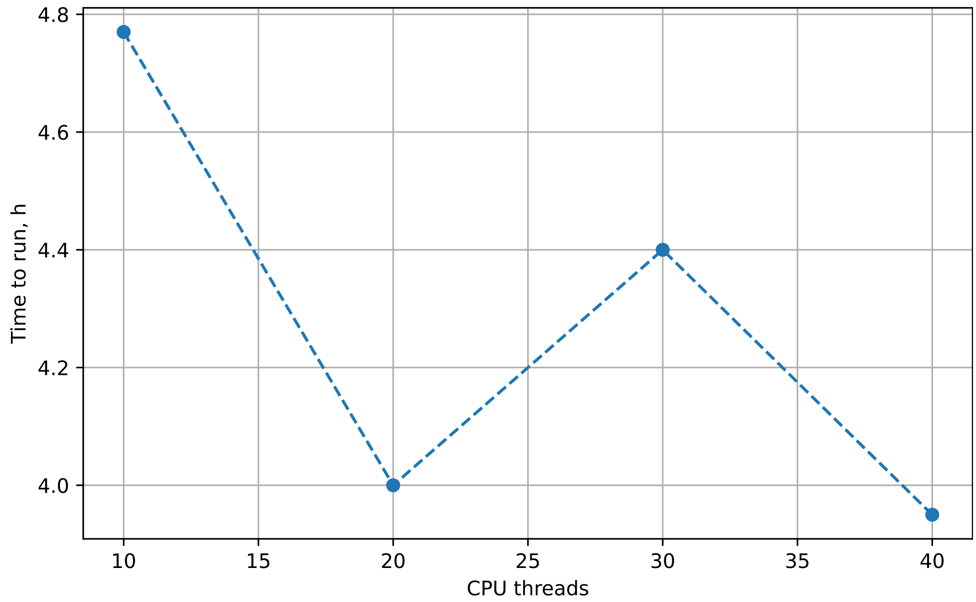

Figure 8: Runtime performance of the pipeline on multi-core CPUs.

4 Mobility scenario for COVID-19

The unified modelling pipeline is used to generate the agent-based travel demand for a large-scale mobility scenario for the whole of Switzerland. The GPU-accelerated agent-based mobility simulator, GEMSim (Saprykin *et al.,* 2019), is used to run this scenario. The scenario includes the whole population of Switzerland (3 million car drivers and 2.5 million public transit users and walkers) with their detailed travel demand, the road network (1.1 million links and 0.5 million intersections) generated from OSM, and public transit schedule (30 000 stops and 20 000 routes) from SBB (Swiss Federal Railways), where the schedule includes routes for trains, buses, trams and other means of transportation in Switzerland. GEMSim uses a co-evolutionary approach where a daily iteration is simulated repeatedly, and agents can adapt their behaviour between iterations to maximize the score of a utility function. The utility function gives a positive score when an agent performs an activity, and a negative score when an agent is travelling or arriving too late. The scenario was executed for 100 iterations, and a random population sample of 10% was re-routed between iterations with the congestion patterns known from previous iterations.

Figure 9 compares the predicted departure dynamics to the microcensus, while Figure 10 compares en-route dynamics. In both cases, the trends are very well captured, especially the dynamics of the morning and evening peak hours. The predicted generated travel demand slightly overestimates departures in the noon, while the number of agents travelling at the same time is underestimated. This can be explained by various reasons: some of the generated trips have shorter distances, traffic lights were not simulated in the scenario, or other factors such as pedestrians were not accounted for.

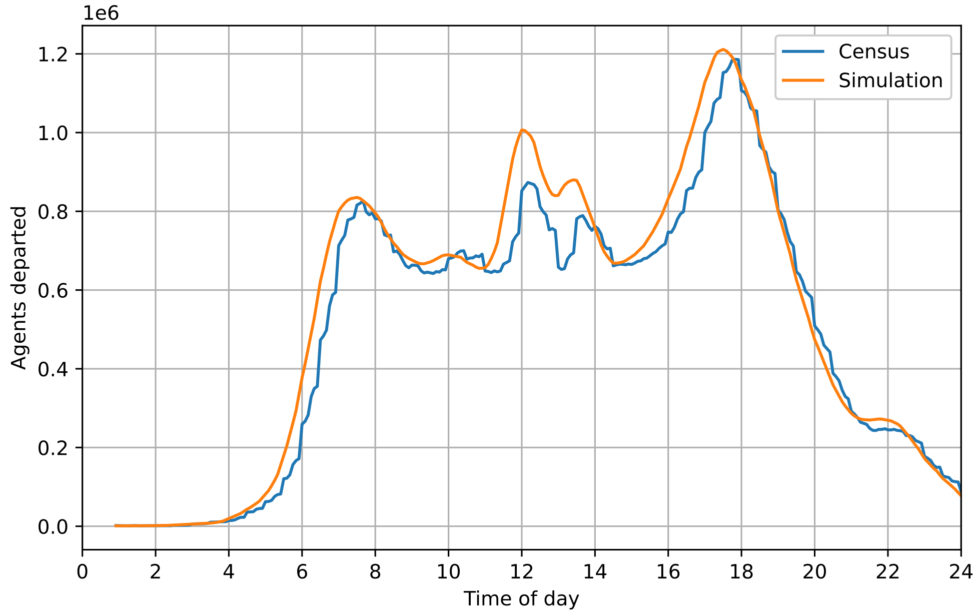

Figure 9: Departure dynamics of the agents throughout a day.

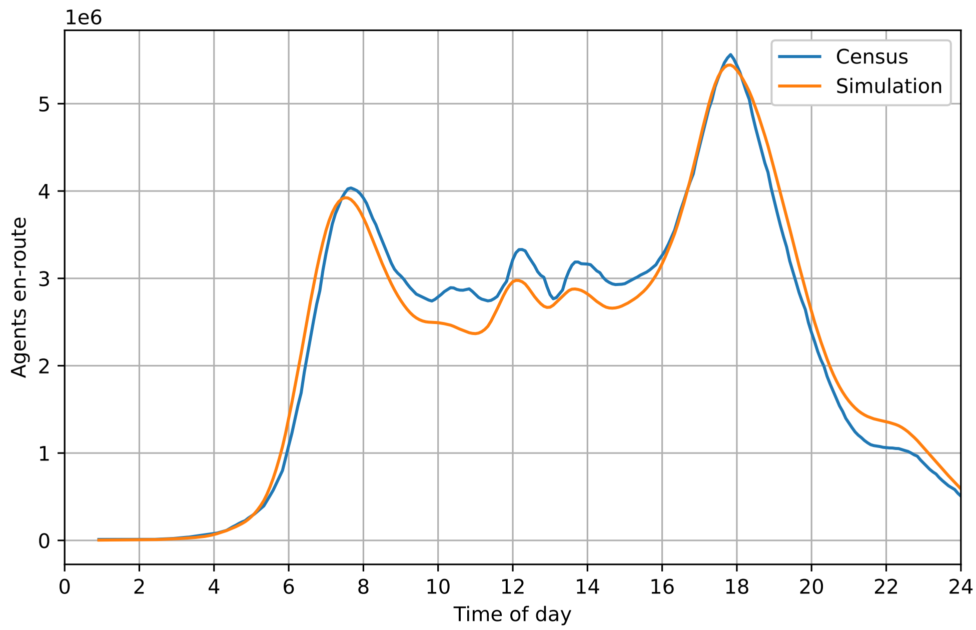

Figure 10: En-route dynamics of the agents throughout a day.

To demonstrate the flexibility of the simulated pipeline, the unified modelling pipeline was modified to simulate the behaviour of people during the public health intervention measures imposed by Swiss government to prevent the spread of the COVID-19 virus. Thus, an additional configuration that describes the transformation of people's behaviour during epidemics was added to the generation part of the pipeline. The transformation of behaviour can be defined in the following ways:

- Specify types of OSM locations which are closed and thus agents cannot go there.

- Specify NOGA codes for jobs which are unaffected by the measures, and thus employees have to travel to work (that is, these employees do not do home office).
- For each of purpose of trip in the microcensus, specify a probability that an agent will drop a trip of this purpose in his/her daily plan. The drop of a working activity means that an agent does home office.
- For agents who have a car in a household but do not use it, specify the probability of using a car instead of public transit.

Using the modified pipeline, four scenarios, based on data from (Google, 2020, Apple, 2020) and locally observed behaviour, have been simulated:

- **No closing** (February 22, 2020): no limitations of places of activity; 0.30 probability of dropping any activity; jobs in healthcare, public services and groceries are kept running; 0.1 probability of switching to a car;
- 1^{st} **partial closure** (March 13, 2020): schools, universities and leisure facilities are closed; 0.45 probability of dropping any activity (except already closed); jobs in healthcare, public services and groceries are kept running; 0.2 probability of switching to a car;
- 2^{nd} **partial closure** (March 16, 2020): schools, universities and leisure facilities are closed; 0.70 probability of dropping any activity (except already closed); jobs in healthcare, public services and groceries are kept running; 0.3 probability of switching to a car;
- **All closed** (March 20, 2020): all facilities except explicitly allowed are closed; 0.95 probability of dropping any activity (except already closed); jobs in healthcare, public services and groceries are kept running; 0.4 probability of switching to a car.

Figure 11 compares the number of travelling agents throughout a day for each COVID-19 scenario. The model clearly shows that there is the change in mobility behaviour, especially in the 1^{st} **partial closure** scenario when educational and leisure facilities were closed. Further measures, such as more strict social distancing, do not have a strong impact on the number of people travelling daily.

Table 8 summarises the reduction in average travel distance and time for the scenarios. The baseline scenario shows good match with the microcensus: 34.6 km of average distance per person and 1.5 hour of average travel time. The predicted average travel distance and time were compared to real tracking data obtained in the MOBIS project (IVT and WWZ, 2020), in which more than 3 000 participants were tracked with GPS and travel diaries. The real tracking data shows that after the introduction of the most strict measures on March 20, the reduction in average travel distance by car was about 50%. Thus: (i) the 2^{nd} **partial closure** scenario is the most realistic; and, (ii) more than expected people continued to travel even after the imposed

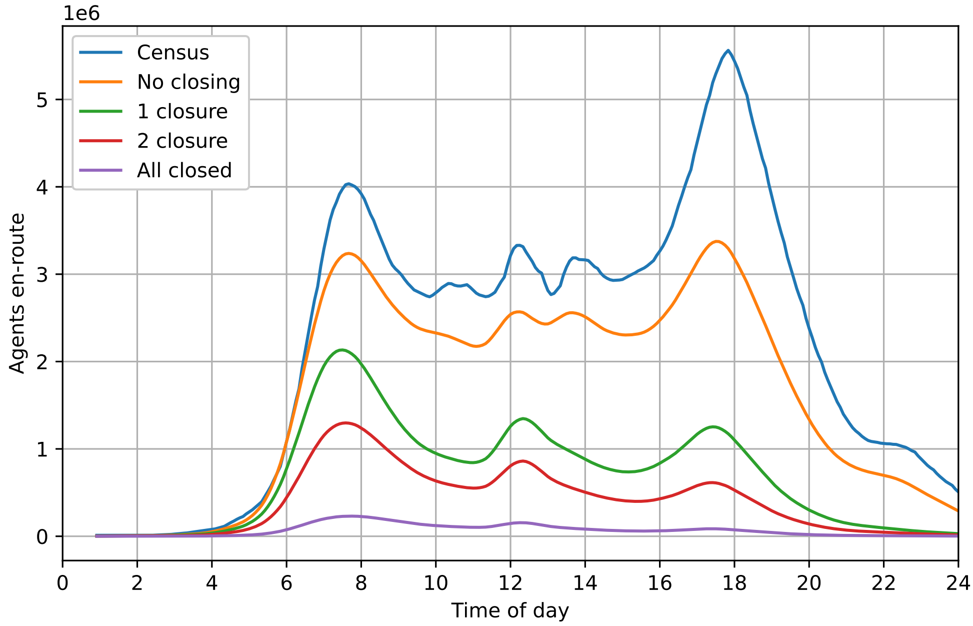

Figure 11: En-route dynamics of the agents throughout a day for COVID-19 scenarios.

restrictions.

Scenario	**Avg. dist [km]**	**Avg. time [h]**	**Reduce dist [%]**	**Reduce time [%]**
Baseline	34.62	1.42	0.00	0.00
No closing	28.49	1.12	17.70	21.13
1^{st} partial closing	19.47	0.74	43.76	47.89
2^{nd} partial closing	17.47	0.65	49.53	54.23
All closed	15.71	0.57	54.62	59.86

Table 8: Reduction in average distance and time for COVID-19 scenarios.

5 Conclusions

A unified modelling pipeline to generate agent-based travel demand of Switzerland has been presented and demonstrated. The pipeline is used to synthesize the whole population of Switzerland with personal attributes and travel behaviour. The generated travel demand is then executed in a large-scale mobility scenario using a GPU-accelerated mobility simulator, GEMSim. The predicted travel dynamics have a good match to microcensus data.

The unified modelling pipeline has been modified to account for the change in people's behaviour that was observed during the COVID-19 pandemic. Four different scenarios related to the public health measures introduced by the government on March 2020 to reduce the impact of COVID-19 pandemic, have been simulated. The simulations show that the measures imposed on March 13, 2020 had the strongest impact on travel behaviour, leading a 44% reduction in average travel distance and 48% reduction in average travel time. After the most strict measures were introduced on March 20, 2020, the reductions were 49% and 54% in distance and time respectively. These predicted average travel distance and time match well with real tracking data.

6 References

Algers, S., A. Daly, P. Kjellman and S. Widlert (1995) Stockholm model system (sims): Application, paper presented at the *7th World Conference of Transportation Research*, 16–21.

Apple (2020) `https://www.apple.com/covid19/mobility` [Accessed: 20-April-2020].

Arentze, T., F. Hofman, H. van Mourik and H. Timmermans (2000) ALBATROSS: multiagent, rule-based model of activity pattern decisions, *Transportation Research Record*, **1706** (1) 136–144.

Arentze, T. A., H. J. P. Timmermans and F. Hofman (2007) Population synthesis for microsimulating travel behavior, *Transportation Research Record*, **2014 (11)**, 85–91.

Auld, J., A. K. Mohammadian and K. Wies (2010) An efficient methodology for generating synthetic populations with multiple control levels, *Paper presented at the the 89th Annual Meeting of the Transportation Research Board, Washington, D.C.*

Bonabeau, E. (2002) Agent-based modeling: Methods and techniques for simulating human systems, *Proceedings of the national academy of sciences*, **99** (suppl 3) 7280–7287.

Bowman, J. L. (2004) A comparison of population synthesizers used in microsimulation models of activity and travel demand.

Bundesamt für Landestopografie (2019) Amtliches Ortschaftenverzeichnis mit Postleitzahl und Perimeter, `https://www.cadastre.ch/en/services/service/registry/plz.html` [Accessed: 20-April-2020].

Bundesamt für Raumentwicklung (2019) ÖV-Güteklassen, `https://www.are.admin.ch/are/de/home/verkehr-und-infrastruktur/grundlagen-und-daten/verkehrserschliessung-in-der-schweiz.html` [Accessed: 20-April-2020].

Bundesamt für Statistik (2018) Generalisierte Gemeindegrenzen: Geodaten, https://www.are.admin.ch/are/de/home/verkehr-und-infrastruktur/grundlagen-und-daten/verkehrserschliessung-in-der-schweiz.html [Accessed: 20-April-2020].

Bundesamt für Statistik (2019) Gemeindetypologie und Stadt/Land-Typologie 2012, https://www.bfs.admin.ch/bfs/de/home/statistiken/querschnittsthemen/raeumliche-analysen.assetdetail.2543323.html [Accessed: 20-April-2020].

Bundesamt für Statistik / Bundesamt für Raumentwicklung (2015) Mikrozensus Mobilität und Verkehr (MZMV) 2015, https://www.are.admin.ch/are/en/home/transport-and-infrastructure/data/mtmc.html [Accessed: 20-April-2020].

Bundesamt für Strassen (2017) Fahrzeughalterregister, https://www.astra.admin.ch/astra/de/home/fachleute/fahrzeuge/fahrzeughalterregister.html [Accessed: 20-April-2020].

Damm, D. (1983) Theory and empirical results: a comparison of recent activity-based research, *Recent advances in travel demand analysis*, **3**, 33.

Davidson, W., R. Donnelly, P. Vovsha, J. Freedman, S. Ruegg, J. Hicks, J. Castiglione and R. Picado (2007) Synthesis of first practices and operational research approaches in activity-based travel demand modeling, *Transportation Research Part A: Policy and Practice*, **41** (5) 464–488.

Deming, W. E. and F. F. Stephan (1940) On the least squares adjustment of a sampled frequency table when the expected marginal totals are known, *Annals of Mathemtical Statistics*, **11 (4)**, 427–444.

Elkins, K. (2018) Here's how much of your income you should be spending on housing, https://www.cnbc.com/2018/06/06/how-much-of-your-income-you-should-be-spending-on-housing.html.

Fahrlaender and Partner AG. (2017) Segmentation of demand in the housing market, Bern.

Federal Office for Housing / Bundemsamt für Wohnulgswesen (2000) Wohnungbauten planen, beurteilen un vergleichen, Grenchen.

Google (2020) https://www.google.com/covid19/mobility/ [Accessed: 20-April-2020].

Habib, K. M. N. (2011) A random utility maximization (RUM) based dynamic activity scheduling model: Application in weekend activity scheduling, *Transportation*, **38** (1) 123–151.

Hatzopoulou, M. and E. J. Miller (2010) Linking an activity-based travel demand model with traffic emission and dispersion models: Transport's contribution to air pollution in Toronto, *Transportation Research Part D: Transport and Environment*, **15** (6) 315–325.

Hockaday, J. (2019) People in their 20s apend a third of their income on rent, *Metro*.

IVT, E. Z. and U. B. WWZ (2020) https://ivtmobis.ethz.ch/mobis/covid19/reports/mobis_covid19_report_2020-06-04.html [Accessed: 20-April-2020].

Kitamura, R. (1988) An evaluation of activity-based travel analysis, *Transportation*, **15** (1-2) 9–34.

Lam, S. K., A. Pitrou and S. Seibert (2015) Numba: A llvm-based python jit compiler, paper presented at the *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, 1–6.

Luzerner Zeitung (2014) Umzugsmonitoring: Das sind die graunde faur einen wohnortwechsel., https://www.luzernerzeitung.ch/zentralschweiz/luzern/umzugsmonitoring-das-sind-die-gruende-fuer-einen-wohnortwechsel-ld.10223.

Marini, M., N. Chokani and R. S. Abhari (2019) Immigration and future housing needs in switzerland: Agent-based modelling of agglomeration lausanne, *Computers, Environment and Urban Systems*, **78**, 101400.

McCarthy, N. (2017) What share of income do germans spend on rent?, https://www.statista.com/chart/11104/what-share-of-income-do-germans-spend-on-rent/.

Müller, K. and K. W. Axhausen (2010) Population synthesis for microsimulation: state of the art, paper presented at the *STRC 2010*.

OpenStreetMap (2020) https://www.openstreetmap.org [Accessed: 20-April-2020].

Pagani, M., W. Korosec, N. Chokani and R. S. Abhari (2019) User behaviour and electric vehicle charging infrastructure: An agent-based model assessment, *Applied Energy*, **254**, 113680.

Pendyala, R. M., R. Kitamura and D. P. Reddy (1998) Application of an activity-based travel-demand model incorporating a rule-based algorithm, *Environment and Planning B: Planning and Design*, **25** (5) 753–772.

Plagowski, P., N. Chokani and R. S. Abhari (2018) Energy transition on the korean peninsula: Role of thermal plants in a future smart grid, paper presented at the *2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 156–161.

Pritchard, D. R. and E. J. Miller (2009) Advances in agent population synthesis and application in an integrated land use and transportation model, *Paper presented at the the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.*

Roorda, M. J., E. J. Miller and K. M. Habib (2008) Validation of TASHA: A 24-h activity scheduling microsimulation model, *Transportation Research Part A: Policy and Practice*, **42** (2) 360–375.

Saprykin, A., N. Chokani and R. S. Abhari (2019) GEMSim: A GPU-accelerated multi-modal mobility simulator for large-scale scenarios, *Simulation Modelling Practice and Theory*, **94**, 199–214.

SD Works (2018) More than 20`https://www.sdworx.com/en/press/2018/2018-09-20-more-than-20percent-of-europeans-commute-at-least-90-minutes-daily`.

Srinivasan, L. M., S. and K. Yathindra (2008) Procedure for forecasting household characteristics for input to travel-demand models, *Final Report, TRC-FDOT-64011-2008, Transportation Research Center, University of Florida*.

Swiss Federal Office for Statistic (2010) Population and households statistics, Bern, Switzerland.

Swiss Federal Office for Statistics (2003) *Adult Literacy and Life Skills Survey*, Bern, Switzerland.

Swiss Federal Office for Statistics (2013) *Betriebs- und Unternehmensregister*, Bern, Switzerland.

Swiss Federal Office for Statistics (2015) *Eidgenoessisches Gebaeude- und Wohnungsregister*, Bern, Switzerland.

Tagesanzeiger (2017) Jeder dritte bezahlt zu viel miete, `https://www.tagesanzeiger.ch/wirtschaft/geld/jede-dritte-person-bezahlt-zu-viel-miete-in-der-schweiz/story/17796157`.

Terrazas, A. (2015) For many renters, crushing rents are intolerable, `https://www.cnbc.com/2018/06/06/how-much-of-your-income-you-should-be-spending-on-housing.html`.

Torsten, H. (1970) What about people in regional science?, *Regional Science Association*, **24** (1) 6–21.

Viegas, J. M. and L. M. Martínez (2010) Generating the universe of urban trips from a mobility survey sample with minimum recourse to behavioural assumptions, paper presented at the *Proceedings of the 12th World Conference on Transport Research*.

Walker, J. (2011) Basics: Walking distance to transit. human transit., `https://humantransit.org/2011/04/basics-walking-distance-to-transit.html`.

Ye, X., K. Konduri, R. M. Pendyala, B. Sana and P. Waddell (2009) A methodology to match distributions of both household and person attributes in the generation of synthetic populations, *Paper presented at the the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.*