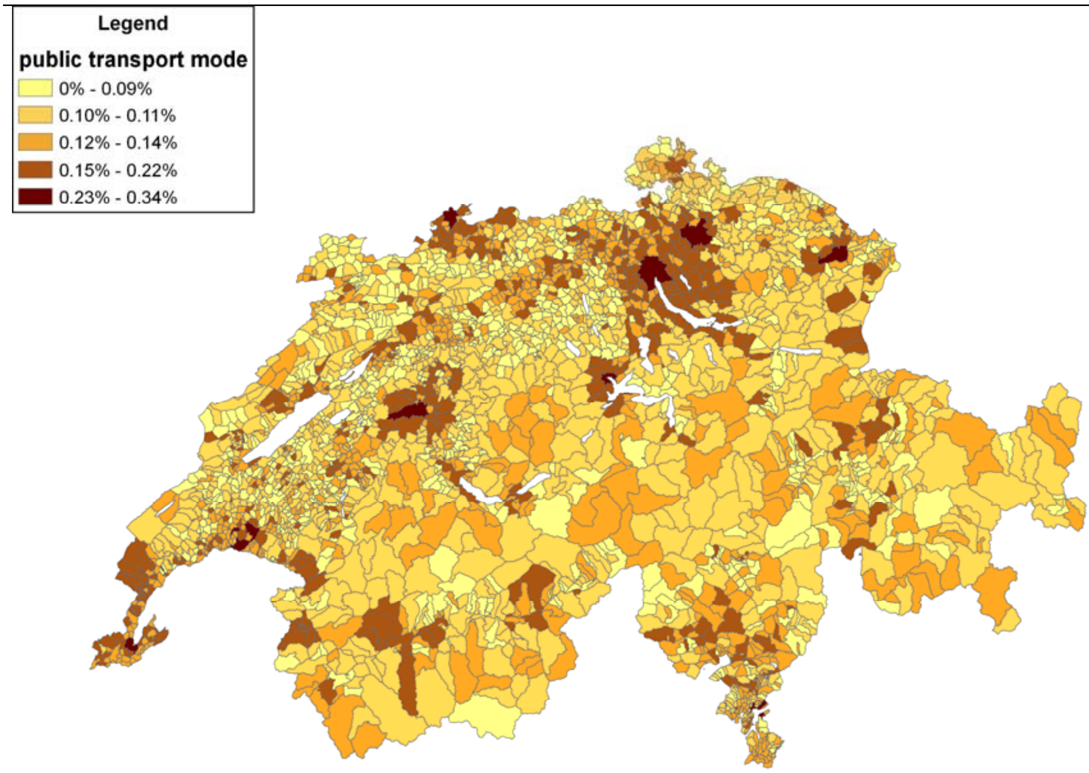


Mobility Tool Ownership and Mode Choice Decision Processes in Multi-Agent Transportation Simulation

Francesco Ciari, IVT, ETH - Zürich
Micheal Balmer, IVT, ETH - Zürich
Kay W. Axhausen, IVT, ETH - Zürich

Conference paper STRC 2007

STRC | 7th Swiss Transport Research Conference
Monte Verità / Ascona, September 12. – 14. 2007

Mobility Tool Ownership and Mode Choice Decision Processes in Multi-Agent Transportation Simulation

Francesco Ciari
IVT,ETH Zurich
ETH Hönggerberg, CH-8093 Zurich

Phone: +41-44-633 71 65
Fax:
email: ciari@ivt.baug.ethz.ch

Michael Balmer
IVT,ETH Zurich
ETH Hönggerberg, CH-8093 Zurich

Phone: +41-44-633 27 80
Fax:
email: balmer@ivt.baug.ethz.ch

Kay W. Axhausen
IVT,ETH Zurich
ETH Hönggerberg, CH-8093 Zurich

Phone: +41-44-633 39 43
Fax:
email: axhausen@ivt.baug.ethz.ch

September 2007

Abstract

This paper presents mode choice decision processes in the context of MATSim-T (Multi-Agent Transport Simulation Toolkit), a simulation toolkit which uses the concept of Evolutionary Algorithm (EA) in order to obtain individual, daily, travel demand. The module described in this paper implements logit models and is embedded in MATSim-T as a planning module. At a first stage the mobility tool ownership among the population, based on the Swiss Mikrozenzus data of 2005, is estimated with a logit model. Then a mode choice is performed at the tour level of the daily activity chain. A multiple choice among four alternatives, walk, bicycle, car, transit, is modeled. The module is integrated in the pre-process stage of MATSim-T and its goal is to produce a reasonable tour based mode choice for each individual. The attributes considered in the utility functions of alternatives are sociodemographic characteristics of the population, such as age, income, employment, car ownership, etc. Even in this simple form the module represents a substantial improvement of the toolkit, increasing its ability to simulate realistic traffic patterns. The module has been tested with different, large scale, scenarios based on Swiss data. The presented results show that the module is able to reproduce modal choices consistent with that data.

Keywords

Agent Based Simulations – MATSim-T – Initial Demand – Mode Choice – Discrete Choice Models – STRC

1. Introduction

In recent years the popularity of agent based simulations has considerably increased and their use as forecasting and planning tool in transportation is expanding. One reason is the significant and continuous increase in computer calculation power, which has made them become suitable for applications involving many millions agents or individuals. But with no doubt it is also thanks to evident advantages in comparison with traditional 4-stage models (Ortuzar and Willumsen, 2006), whereof the explicit modelling of individual decisions is the most appreciable. Among these different decision processes, the transportation mode choice is probably the most important. The share of public transport and private transport among travelers has a key role in transportation policies and its correct estimation is one of the main information that planners want to get from a transportation model.

MATSim-T (Multi-Agent Transport Simulation Toolkit) is a simulation toolkit which is able to deal with large scale scenarios (Balmer, 2005). It uses the concept of the Evolutionary Algorithm (EA) in order to generate consistent daily activity chains for each individual (agent) of a population and travel times on the network. MATSim-T works in four phases. In the first a synthetic population is generated, which is then used to generate the initial travel demand. The travel demand is then optimized and finally results are analyzed. The toolkit's most prominent application is a simulation of the travel behaviour of the whole Swiss population. In this paper the issues of generating the synthetic population and of producing the initial demand are examined, with a special focus on the transportation mode choice of individuals. A more detailed graphic representation of these processes is provided in Figure 1

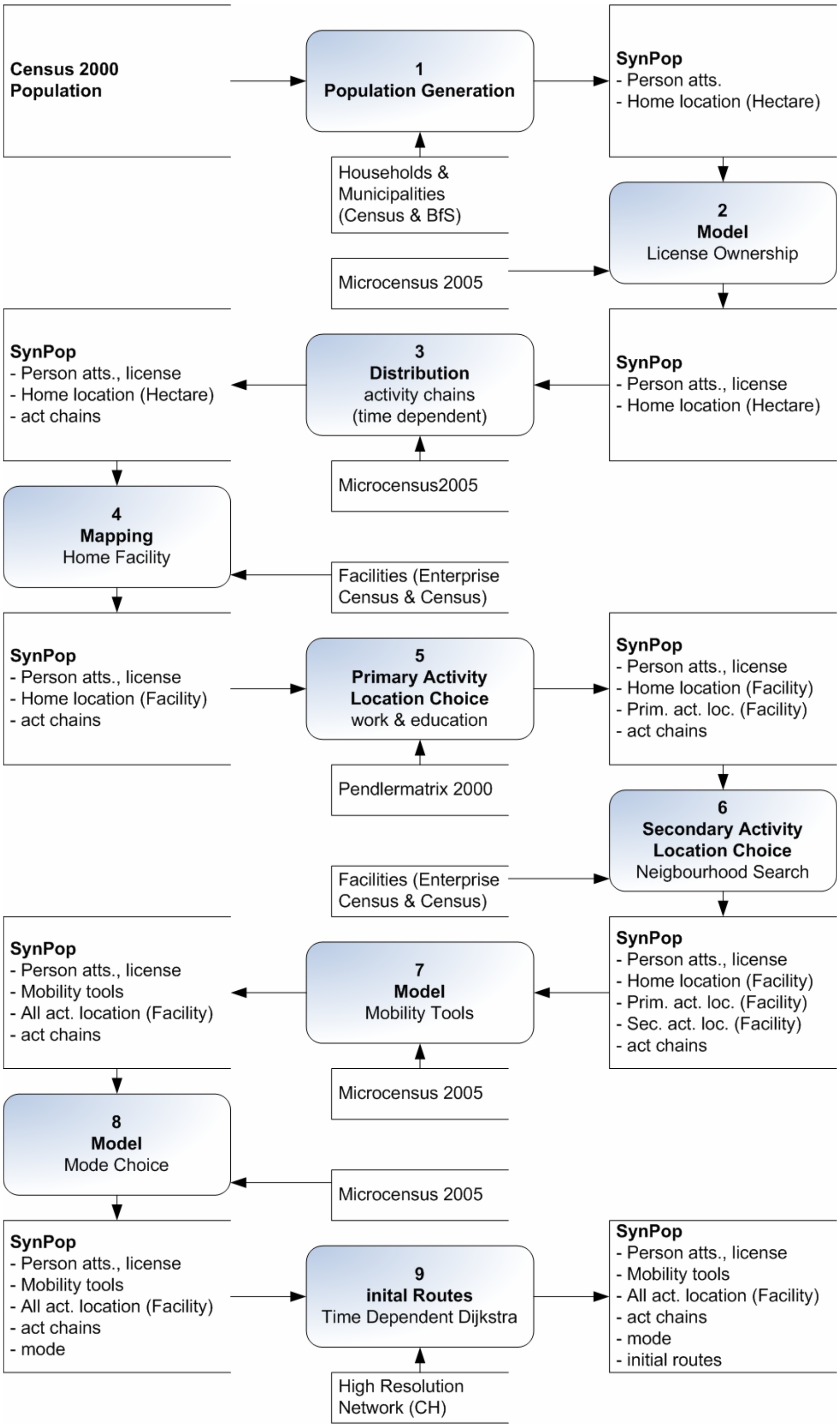
A new mode-choice module has been integrated in the pre-process stage of MATSim-T as a planning module. The module has the goal to produce a reasonable tour based transportation mode choice for each individual. The module has two stages, in each of which a logit models are implemented. At the first stage the mobility tool ownership among the population is estimated with logit models based on the Swiss Mikrozensus data of 2005 (Bundesamt für Statistik and Bundesamt für Raumentwicklung , 2006).

A logit model which imputes driving license ownership among the population is implemented. Then a second one is used to assign a car or one of the different season tickets of the Swiss public transport or a combination of them to each individual of the population. This stage of the module belongs to the population generation phase of MATSim-T.

In a second stage, belonging to the initial demand modeling, a mode choice is performed at the tour level of the daily activity chain. A multiple choice among four alternatives, walk, bicycle, car, transit, is modeled. All other alternative are not explicitly treated and a fifth alternative encompasses all of them. The attributes considered in the utility functions of alternatives are sociodemographic characteristics of the population, such as age, income, employment, car ownership, etc, plus some municipality attributes. Also at this stage it is made use of a logit model based on the Mikrozensus 2005. The module has been run on a synthetic population of more than 7 Mio individuals reproducing the whole Swiss population. This population has been obtained basing on the Swiss census 2000 data (Bundesamt für Statistik, 2001).

This paper is composed of 5 Sections. In the next section data sources are shortly described and the generation of a synthetic population is presented. Section 3 deals with the modelling of initial demand and focuses on the mode choice module. In Section 4 some results are presented. The distribution of mobility tools among the synthetic population and the transportation mode shares reproduced by the mode choice module are discussed. The last section is dedicated to conclusions and provides an outlook to further work.

Figure 1 Tasks flow chart – Population generation and



2. Synthetic Population Generation

The generation of a suitable population for a microsimulation implies the availability of data at a more disaggregated level then that needed for conventional transport models. In fact, since the main advantage of microsimulations is that each individual is modelled it is clear that also socio-demographic attributes have to be known at the individual level. Moreover, the data source should contain all the socio-demographic attributes that one will make use of, such as age, sex, income, residence, etc., and that for all individuals. However, this data does not always exist, and when it does exist usually not all desired socio-demographics are available in the same data set. More generally speaking, as Frick (2003) already pointed out, if exhaustive socio-demographic information on an individual is available the spatial resolution will be low and conversely if a high spatial resolution is there, some of the socio-demographic will be missing or will be provided only at a higher aggregation level.

That raises the issue of how to generate a suitable input synthetic population for the microsimulation tool.

The input population used here has been generated starting from the Swiss census 2000. This data set contains socio-demographic information on the whole Swiss population and has a very good spatial resolution, at the hectare level, but does not include some important information, such as income, license ownership, car ownership, etc. Therefore we create first a population which simply replicate the one of the census, and then we add the missing attributes using other data sources. Basically we need to add income, driving license ownership, and mobility tools ownership (car ownership and public transport reductions cards or season tickets ownership). This information, which can be found in the Swiss Mikrozensus, is attached to the population in successive steps since such attributes are interdependent and concatenated. That means initial population attributes are first complemented with an income estimate for each individual. Next a logit model is used to impute the driving license ownership of the agents. Thanks to another logit model the population description is further enriched with mobility tool ownership attributes, i.e. car ownership and public transport season tickets ownership.

2.1 Data Sources

For the generation of the synthetic population it was made use of two main data sources, the Swiss Census 2000 and the Swiss Mikrozensus 2005 (MZ). These data sets contain most of the attributes considered to be significant in order to describe the travel behaviour of individuals.

2.1.1 Census Data

The Swiss Census is a general demographic survey which encompasses the whole Swiss population and is undertaken every ten years. It is carried out by the Federal Office of Statistics (Bundesamt für Statistik, BFS). Each single individual of the resident population of Switzerland is surveyed and many sociodemographic are reported. This data set has also a good spatial resolution, at the hectare level, but does not have some important information, such as income, license ownership, car ownership, etc.

2.1.2 Mikrozensus Data

The Swiss Mikrozensus is a survey which describes travelling habits of the Swiss population. It is also carried out by the BFS together with the Federal Office for Spatial Development (Bundesamt für Raumentwicklung) and conducted about every 5 years. The data set contains information on both households and single individuals. Information is available for a representative sample of the population. The sample is representative at both, the household and the individual level, and different weights for cases apply. In the Mikrozensus it is possible to find most of information that is needed and which is not in the Census, but off course there is no possibility to directly integrate information from Mikrozensus into Census data. This issue will be discussed later. This data will be used also for the generation of activity chains and for the mode choice model.

2.2 Generation of an initial population

The Census is used as base for the synthetic population generation since information of all individuals of the Swiss population is available. In a first step the Census population is simply read and uploaded in the system. In other words, at this point, the population that was generated is a simple reproduction of the Census population, but only the relevant information is kept. Each individual has both a personal ID and a household ID and therefore it is possible to get some information at both these levels within the system. However, in the MATSim-T data structure a household layer is not yet included and the household information is integrated at the individual level.

So far we can not properly speaking of a synthetic population, since individuals are still real individual, even if not directly identifiable. We can do that only as soon as some information is added. The first is the income of the person, which is not in the Census data. This information is added using municipality income data from the BFS. To each employed person income is added simply as the average income of the municipality where the person lives. This approach is simple and highly aggregated, but a better data set was not available at the moment. A more sophisticated income imputation will be added later. Resuming, in the initial synthetic population each agent have at this point the following attributes: Age, sex, employment (yes/no), nationality (Swiss/not Swiss), income, home location (hectare level), number of persons in the household and number of kids of the household.

2.3 Using logit models to add information

The initial population does not have enough attributes in order to obtain a good initial demand. Most of all, with such attributes it would be hard to estimate modal choices of individuals. To do that it is necessary, at least, to have additional information about driving license ownership, car ownership and public transport season tickets ownership. To overcome that, this information must be added to the synthetic population. All these attributes can be found in the Swiss Mikrozensus, but as a sample the relevant information cannot be directly attached to the initial population. A possible solution would be to attach these attributes to individuals of the synthetic population according to the distribution that those attributes have among the real population. With the help of a random draw it would be simple to reproduce in the synthetic population the distribution of such attributes. However this approach would cause some problem, since the distribution would be correct only at aggregated level. In other words, it would produce, for example, the right percentage of persons owning a car, but a 10 years old scholar would have the same possibility to own a car as a 40 years old professional. Even with some correction, for example such as to assign car ownership only to persons who own a driving license, and a driving license only to persons who are at least 18 years old in turn, would definitely bring a loss of information. Using global means, information which is already in the population would be not used. To avoid that it has been made use of logit models. With this method attributes may be added to individuals according to some of their other attributes, making a better use of the information which is already in the system. To be that possible, it is necessary that the new data set contains both the information that one already have and some new information that needs to be added. This is exactly the case for the Swiss Mikrozensus where, along with attributes which agents have already, such as age, sex, nationality, etc., most of the additional information needed can be found. The models where estimated using the free software Biogeme (Bierlaire's Optimization Toolbox for GEV Model Estimation). For more details please refer to Bierlaire (2003 and 2007). The use of such an instrument implies the use of a suitable sample. For these models the only issue was the form in which some variables appear in MZ 2005, not always appropriate. New samples where generated simply recoding such variables, but in principle all cases included in MZ could be used. We will see later that the generation of the sample for the mode choice model implies some more care.

2.3.1 Driving License Ownership modelling

The first information which is added to the initial population is the driving license ownership. That is logical since the driving license ownership is in principle a prerequisite for the car ownership, and it is intuitive that this attribute will probably be a factor in a logit model of car ownership. The logit model used is estimated based on a on the Swiss MZ 2005. Basically it reproduces the model which was already estimated by Beige (2004) for a smaller sample of the Swiss population; for more details refer to cited reference. This model assume that there are two different utility for each individual, one for the "Yes" alternative (to own a driving license) and one for the "No" alternative (not to own a driving license).

Explanatory variables included in the model are of three different types, personal attributes of individuals, household attributes, and attributes of the municipality where the individual live. The first and the second are found in the synthetic population, while information of the third category may be easily added extracting municipality data from Swiss Census and mapping the residence of individuals to municipalities. In Table 1 the list of explanatory variables and the value of corresponding parameters which appear in the two utility functions is reported.

TABLE 1 Driving License Ownership Model

Alternative; Variable	Parameter Value	Robust t-test
Owning a licence		
Age in years	– 0.321	-63.886
Age in years logarithmic	13.171	65.019
Age in years per Gender: Male	0.024	26.326
Nationality: Swiss	0.791	13.033
Not owning a licence		
Number of persons in the household *	0.082	1.914
Number of children in the household	0.123	2.419
Monthly household income in 1000 CHF	-0.118	-13.532
Main centres (reference category)		
Middle and ancillary centres with railway access	-0.421	-5.133
Middle and ancillary centres without railway access	-0.467	-5.418
Agglomeration municipalities	-0.616	-10.018
Rural areas	-0.740	-11.367
Constant	+ 34.494	63.056
N		33390
L(Constants only)		-23144.2
L(Beta)		-11101
Adj. Rho squared		0.519836

The implementation of the logit model, in the specific case in Java code, is straightforward. For each individual of the synthetic population the utility of the two alternatives and associated probabilities are calculated. A random draw selects the actual choice of the individual.

2.3.2 Mobility Tools Ownership modelling

Once the driving license ownership has been added to the population, the mobility tool ownership can be estimated in turn. The idea is to estimate the utility that one individual has owning a given combination of mobility tools, where for mobility tools are intended car and public transport season tickets. In the MATSim-T data structure car availability and various different season tickets are taken into account. Therefore we need here to confront with a modelling choice since the concepts of car ownership and car availability cannot be perfectly assimilated. The car ownership concept seems to model better an individual choice and than more appropriated to be used in a logit model. This choice would raise the issue of how to impute car availability which is the information needed in the microsimulation. The simplest way would be assuming car always available for individuals owing a car. Other individuals would have the car availability consequently set to never. However, this approach would be too crude and would probably bring to an underestimation of car trips. In a household, for example, if the ownership of a car can be associated to a single person its use will be probably shared among other driving license owners. Moreover in Mikrozensus data can be verified that a not negligible number of car trips are effectuated by individuals not owning a car but having access to it. Some correction to this assumption could be made in the implementation of the model, however it is preferred to keep consistent with the MATSim approach from the beginning and directly model the car availability. Coming to season ticket ownership, the MATSim-T data structure would allow taking into account any kind of them. To limit the complexity of the model it has been taken into account only if an individual owns or not a season ticket or a reduction of any kind. On this basis 6 different alternatives are taken into account; they are listed in Table 2. The column availability indicates how the availability of an alternative for individuals is dealt by the model. One means that the alternative is available for the whole population; License means that the alternative is available only for individuals owning a driving license.

TABLE 2 List of Alternatives – Mobility Tools Ownership Model

Alternative	Characteristics	Availability
1	Car Never + No Tickets	One
2	Car Sometimes + No Tickets	License
3	Car Always + No Tickets	License
4	Car Never + Tickets	One
5	Car Sometimes + Tickets	License
6	Car Always + Tickets	License

Explanatory variables are, as for the driving license model, attributes of individuals, households and municipalities. They are listed in Table 3 along with the respective parameters estimation.

TABLE 3 Mobility Ownership Model

Alternative Variable	1 beta	2 beta	3 beta	4 beta	5 beta	6 beta
Constant	-	-25.902	-2.231	-5.244	4.891	-19.048
Age	-	-0.411	0.112	-0.069	0.223	-0.221
Age squared	-	10.487	-0.788	1.759	-3.751	6.585
Age ln	-	0.007	-0.023	0.002	0.000	0.006
Age * Gender	-	0.002	-0.001	0.000	-0.002	0.001
Gender	-	0.497	1.298	-0.029	0.084	0.235
Nationality	-	0.791	0.972	0.802	1.769	1.463
Distance Work-Home	-	0.000	0.000	0.000	0.000	0.000
Income (1000 Chf)	-	0.182	0.269	0.077	0.511	0.359
Income squared	-	0.092	-0.063	-0.050	-0.509	-0.140
Income ln	-	-0.001	-0.007	-0.001	-0.013	-0.006
Price Gasoline 95	-1.297	-	-	-	-	-
HH Dimension	0.179	-	-	-	-	-
HH Kids	-0.060		-	-	-	-
Municipality 2[1]	-	0.809	0.420	-0.622	-0.159	0.320
Municipality 3[1]	-	1.017	1.086	-0.363	0.293	0.688
Municipality 4[1]	-	1.448	1.270	-0.566	0.257	0.889
Municipality 5[1]	-	1.655	1.462	-0.896	0.098	0.771
Number of observations						33390
L(Constants only)						-48384.3
L(Beta)						-33916.7
Adj. Rho squared						0.299014

[1] Municipality 2 = Middle and ancillary centres with railway access; Municipality 3 = Middle and ancillary centres without railway access; Municipality 4 = Agglomeration municipalities, Municipality 5 = Rural areas, Main centres are the reference category).

It is important to note that a variable indicating the distance between home and work location is in the model. According to what affirmed in the introduction this variable should not be yet available since it is first calculated as activity location are assigned, which is part of the demand modelling step. However the distinction between population generation and initial demand generation is more a conceptual than a strictly hierarchical one. Actually some of initial demand modelling tasks may be performed before the population has been enriched with all desired attributes. Indeed, as in this case, some of the results may be used in order to attach these attributes to the population.

3. Modelling individual Initial demand

The first task modelling individual initial demand is the generation of activity chains for the whole population, which is performed using a set of reference chains based on the Swiss Mikrozensus. In MATSim both home-based and non-home-based tours can be simulated, but in MZ the distinction is not clearly reported and for simplicity only home-based activity chains are taken into account. Moreover, they account for the overwhelming majority of tours. Each complete activity chain will therefore start and end at the home location. Then activities are spatially assigned. Finally, in order to complete the initial demand, a third logit model is used to estimate transportation mode choice at the tour level for each individual. A further task, also belonging to initial demand modelling in MATSim, but not described in this paper, is the assignation of an initial route. At the end of this process each individual of the population will have a plan assigned which is composed of activities to be performed, the time slot in which they will be performed, the location of such activities, and a route

The modelling of initial demand has a precise goal in MATSim –T. In the optimization stage Evolutionary Algorithms are used and an initial demand is needed in order to get the simulation working. In fact whatever initial demand could be used, but if optimization algorithms are provided with a good initial solution their computation time is in general considerably shortened (Edelhoff et al., 2007).

The initial demand for an individual will be stored in a data structure called “plan” (see Meister et al., 2006). The generation of the initial demand includes two main parts, the generation of activity chains and mode choice modelling.

3.1 Activity Chains

It has been observed that most of the activity chains that are reported in Mikrozensus belong to a limited group. In fact Balmer et al. (2005) found that in Mikrozensus 2000 more than 1500 (without counting time slots) different activity chains do exist, but with 21 it is already possible to cover more then 90 percent of them.

In the present case it is made use of 509 different activity chains which become 4611 considering different departure times. These cover more than 95% of Mikrozensus activity chains.

These activity chains in MATSim are called reference plans, and they appear in the Mikrozensus with a given distribution. Therefore the first step is to assign to each individual one of these activity chain. To do that, the given reference plans are classified based on whether they contain a shopping, work and/or education activity. Then they are assigned to the persons based on their age (e.g. people with age < 6 and age > 65 do not engage in education), based on whether they are employed or not (the plan of an employed person must contain a work activity and vice versa) and according on the distribution of the reference plans. The assignation task is performed using an IPF (Iterative Proportional Fitting) technique. The resulting distribution has only a 0.1 % relative bias, which is quite satisfactory. Once an initial plan has been generated for each individual basing on Census 2000, activities need to be located. In MATSim two types of activities are distinguished, primary activities and secondary activities.

3.1.1 Primary activity location

Primary activities include Home, Work and Education. The home location is known, since it is one of the person attributes and thus it is just a matter of copying this information into the plan. For the other primary activities, work and education, is made use of the work commuter matrix and the education commuter matrix (Vrtic et al., 2005). These are origin-destination matrices at municipal resolution. It is known how many individuals living in one specific zone are commuting for work or education purposes in a second specific zone. Individuals are mapped to these zones and, based on this distribution, work and education locations are assigned with a random draw. At this point only the zone where an activity is performed is assigned. Since a higher resolution is desired, facilities have been mapped in MATSim. In order to avoid possible inconsistencies (i.e. entries pointing to zones that have no facilities that allow performing the respective activity) a fusion process between these data sources is performed. A second random draw assigns the primary activity to a specific facility of a zone, according to capacities of facilities where the considered primary activity might be performed.

3.1.2 Secondary activity location

All activities which are not classified as primary are secondary activities, and they also need to be assigned to a given facility where they may be performed. Since for such activities data sets comparable to commuter matrices are not available, that task is accomplished using a neighbourhood search. Two cases are contemplated: a chain contain only home as primary activity, a chain contain also work and education. In the first case the neighbourhood is simply a circular area where the radius is related to the dimension of the municipality. If the research of a suitable facility fails in this area the radius is doubled until a suitable facility is found. If two primary activities are in the chain, with two different locations, the neighbourhood in which the research is performed is created as the union of two circles having the two locations as centres and the radius proportional to the distance between the two locations. Also in this case if the first research is unsuccessful the radius is lengthened until a suitable facility is found.

3.2 Mode choice

The last step that will be described of the whole initial demand modelling is the transport mode choice estimation. Among individual decision processes the transportation mode choice is probably the most important. The share of public transport and private transport among travellers has a key role in transportation policies and its correct estimation is one of the main information that planners want to get from a transportation model.

A new mode-choice module has been integrated in the pre-process stage of MATSim-T as a planning module. Until now the simulation considered every trip as a car trip. In order to reasonably reproduce traffic in the simulation only a fixed percentage of these trips were actually simulated. This approach was satisfactory in terms of its ability to reproduce morning and evening peaks. However, adding a mode choice module, it is expected that simulation's reliability and flexibility are considerably increased. The following is about the introduction of a mode choice module in the pre-process stage of MATSim. It is a first step towards integrating mode choice in the optimization stage.

3.2.1 Mode choice logit model

The logit model for the transportation mode choice is based on the Swiss Mikrozensus 2005. The mode choice is performed at the tour level of the daily activity chain. That means, it is assumed that a single mode is used in a generic tour of home – out-of-home-activities – home. The choice among four alternatives, walk, bicycle, car, transit is modelled. All other alternatives, registered in the MZ, are not explicitly treated and a fifth alternative encompass all of them. To better capture differences in mode choice behaviour that individuals may have in different situations, three separated multinomial logit models were estimated, according to the main purpose of the tour. Purposes considered were, in hierarchical order, work, education, and shop/leisure. That means, since a tour, in general, will contain more than one of these activities, a tour containing work will be classified as a tour with work purpose whatever the other activities will be (and whatever the time spent for each activity). A tour with education will be classified as education, except if work is also an activity of the tour. A tour will be classified as a shop/leisure tour only if no work and no education activities are included in the tour. In order to obtain a suitable data set for the model estimation, similarly to what applied for License and Mobility Tools models, some variables needed to be recoded. However, in this case, the generation of a suitable sample involved some more drastic operation. For a matter of simplicity cases reported in MZ referring to incomplete tours where deleted (about 4% of the original sample). Also tours where all or part of the geographic information is missing where deleted (1.5% of the remaining sample). On a similar vein, since only information which is contained in the system may be used, utility functions take a form which is somewhat unusual for a mode choice model. Only socio-demographic variables are in the utility functions except for the total distance travelled for the tour and for municipality attributes. Usually alternative specific attributes should be also included, typically cost and travel time. This lack limits the power of the models, but these models are expected to be good enough for their goal: the production of meaningful initial mode choice shares. The models are shown in the tables below.

TABLE 4 Tour mode choice model: Work

Alternative Variable	Bike	Walk	Car	Public transport
	beta	beta	beta	beta
Constant	1.2865	2.4705	0.1888	-
Tour distance [Km]	-0.0676	-0.3277	0.0183	0.0221
Driving license	-	-	0.3997	-
Car Always	-	-	1.8671	-
Car Never	-	-	-	0.5341
Season ticket	-	-	-	2.4362
Age squared	-	-	-	-0.0001
Municipality 2	-	-	0.4058	-0.3364
Municipality 3	-	-	0.4115	-0.0757
Municipality 4	-	-	0.6374	-0.2876
Municipality 5	-	-	0.5967	-0.9666
Const of “Other”				0.1411
Number of observations				11619
L(Constants only)				-19624.7
L(Beta)				-10091.8
Adj. Rho squared				0.484693

TABLE 5 Tour mode choice model: Education

Variable / Alternative	Bike	Walk	Car	Public transport
	beta	beta	beta	beta
Constant	-0.0110	1.1372	-0.9321	-
Tour distance [Km]	-0.0008	-0.0346	0.0166	0.0185
Driving license	-	-	0.8032	-
Car Always	-	-	2.2058	-
Car Never	-	-	-	-0.5345
Season ticket	-	-	-	1.6101
Age squared	-	-	-	0.0002
Municipality 2	-	-	-0.7262	-0.4462
Municipality 3	-	-	-1.1044	-0.4952
Municipality 4	-	-	0.1005	-0.1249
Municipality 5	-	-	0.0230	-0.2516
Const of “Other”				-1.4410
Number of observations				3963
L(Constants only)				-6763.87
L(Beta)				-5085.08
Adj. Rho squared				0.245095

TABLE 6 Tour mode choice model: Shop and Leisure

Variable / Alternative	Bike	Walk	Car	Public transport
	beta	beta	beta	beta
Constant	1.1708	3.4582	0.7736	-
Tour distance [Km]	-0.1235	-0.4103	0.0174	0.0183
Driving license	-		0.3074	-
Car Always	-	-	1.0021	-
Car Never		-	-	0.8039
Season ticket		-	-	1.3487
Age squared	-	-	-	0.0001
Municipality 2	-	-	0.3201	-0.7621
Municipality 3	-	-	0.3077	-0.6468
Municipality 4	-	-	0.4039	-0.9435
Municipality 5	-		0.3682	-1.5534
HH Dimension	-	-	0.1102	-
Const of “Other”				-0.7620
Number of observations				29169
L(Constants only)				-45966.2
L(Beta)				-25016.6
Adj. Rho squared				0.4553

Alternatives availability

Another issue which deserves some consideration is the availability of alternatives. It is important that model predictions are not only good if compared with global MZ shares, but also consistent at the individual level. If availability is not considered at all, inconsistencies may appear. For various reasons in the mode choice logit model all alternatives have one as availability, which means that in principle all alternatives are always accessible for everybody. Of course the reality is different and some corrections might be introduced in the implementation. However, it was decided to keep the availability to one for all alternatives except that for bicycle tours. This choice is shortly discussed here below.

For the car mode the availability was directly estimated. However this estimation cannot be directly used in the mode choice model as availability of the alternative. In fact it has been observed in MZ data that 20% of tours of individuals with no car availability are made travelling by car. This is not surprising, since also people not owning a driving license may get a ride on a car. For that reason the availability of the car option was set to one in the model. In order to keep consistent with the model, it was avoided any correction in the implementation, even if obviously that opens a new question on how to simulate those agents since they don’t generate additional traffic.

In the case of public transport, mobility tool ownership does not help to determine its availability as an alternative. On the one hand, the ownership of some sort of season ticket increases the possibility that an individual will use public transport, but does not prevent this individual to choose another alternative. On the other hand, an individual, despite not owing a season ticket, can use public transport. Also in Mikrozensus it is not possible to find how many individuals effectively have this option available. Most of all it is hard to see it as a person attribute since, unlike other alternatives, which are individual transportation means, in principle pubic transport is always available for everybody, even if not necessarily available (or meaningful) for each trip. Therefore also in this case no correction was introduced.

For the bicycle alternative none of the synthetic population attributes provide availability information. Nevertheless, it was observed that the introduction of one availability parameter in the implementation improved model performance. For that reason the bicycle alternative was made available to a given percentage of the whole population, consistently with MZ data.

The mode “walk”, which is assumed to be always available for everybody; and the mode “other”, which accounts only for a modest share of trips and includes different modes which have different individual availability in turn, had also their availability kept to one.

4. Results

This paper first described the generation of a synthetic population in the context of MATSim-T. The goal was to produce meaningful distributions of attributes among the synthetic population. Successively the modelling of the initial transportation demand was undertaken, with the generation of reasonable mode shares as main goal. In this section some results are shown in order to demonstrate that both these tasks were accomplished.

4.1 License and mobility tools ownership results

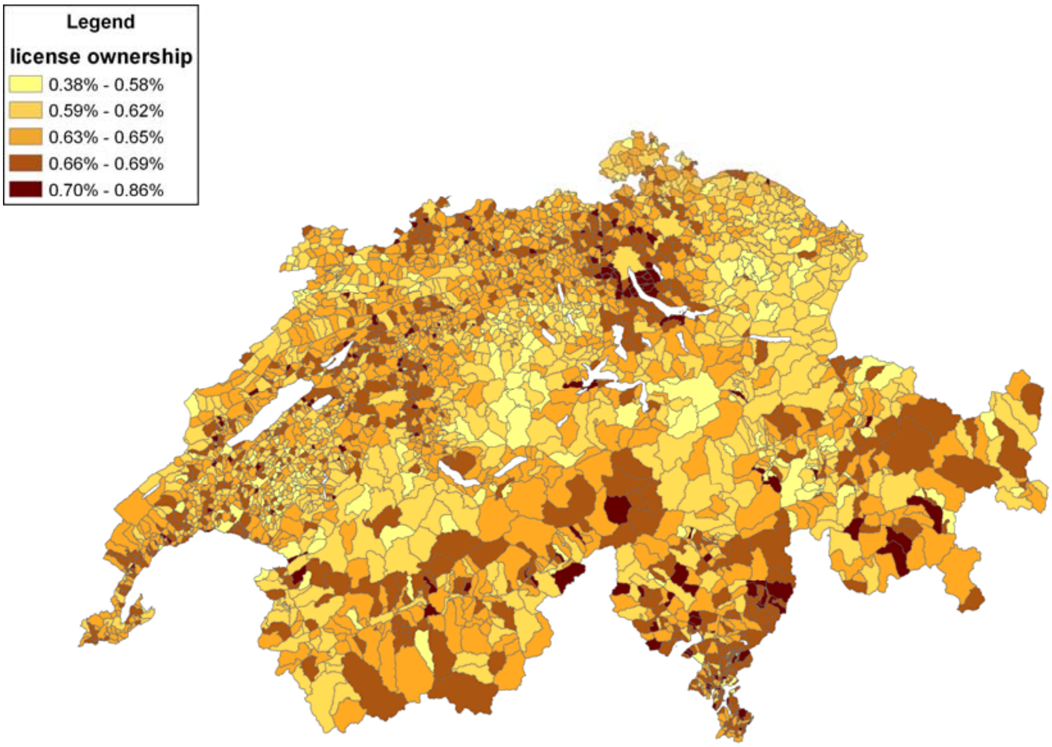
To verify that the synthetic population is reproducing well characteristics of the real Swiss population the distribution of some attributes are shown. Evidently only attributes which were attached to the population with logit models are interesting to be shown, thus license ownership, car availability and tickets ownership.

The model predict that about 63% of the Swiss population holds a driving license some points less than the percentage in the real Swiss population, which is 69%. This result can be considered satisfactory, but not enough to judge the goodness of the model. In principle logit models always tend to reproduce global shares of alternatives and a more interesting test is to verify if the distribution of alternatives against some other variable is also correct. In this case a good test is the spatial distribution of license ownership. The distribution of license ownership at the municipality level is shown in Figure 1.

General spatial patterns of license ownership are known, and if the model is working well we can expect the following:

- In main municipalities of bigger agglomeration (i.e. Zurich, Basel, Bern, Geneva) the percentage of persons holding a license should be in the average or slightly below.
- In other municipalities of main agglomerations this percentage should be above the average.
- In countryside municipalities the percentage should be in the average, possibly with exception both below and above according to various factors such as: the economic vocation of the municipality, the presence of a well developed public transport network, etc.

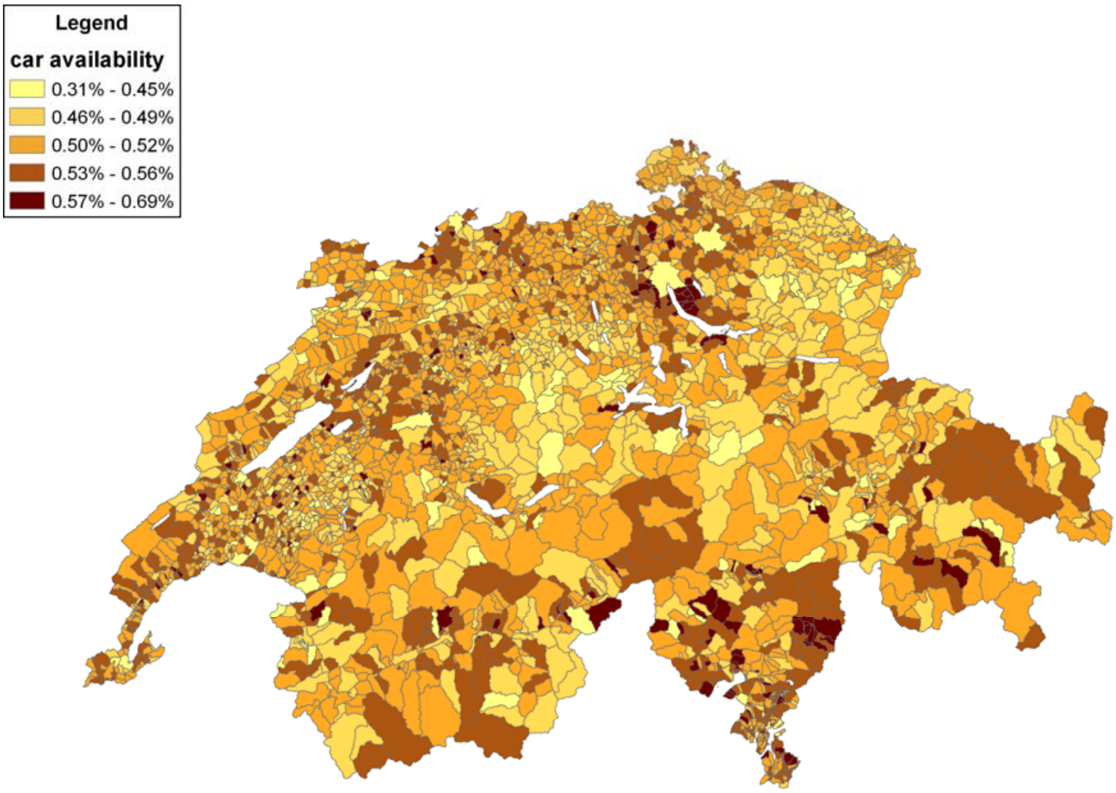
Figure 2 Share of licence holding



In and around bigger towns the expected pattern is well reproduced. Values in less inhabited zones are sometimes unexpectedly high or low; it might have happened that in same municipality with few inhabitants the sample was too small and results are less reliable. In any case, globally, the expected pattern is correctly reproduced by the model.

Car availability and tickets ownership are also presented the same way in two different figures. In Figure 2 the car availability percentage is shown. For this attribute in principle the same pattern as driving license is expected, except for main agglomeration municipalities where car availability values well below the average are expected.

Figure 3 Share of car availability

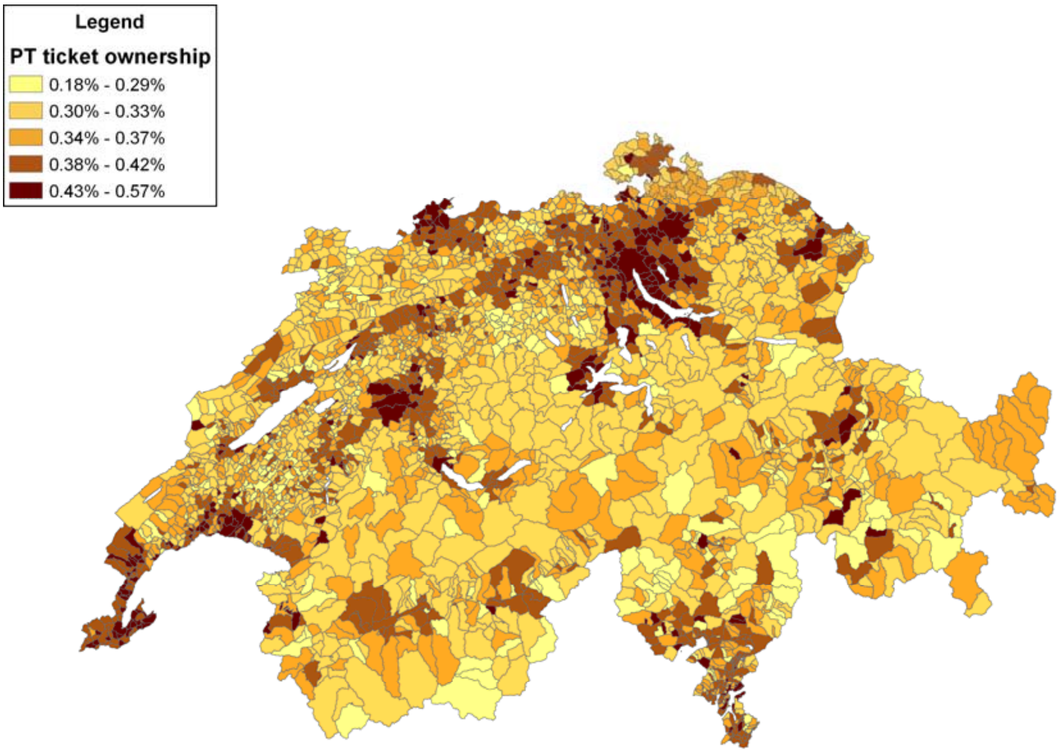


Also in this case the model seems to reproduce well the expected pattern.

Figure 3 shows how the ownership of some of season tickets is distributed. In this case a high concentration of tickets owners is expected in all municipalities of main agglomerations and in general near to towns, while in countryside areas values should be considerably and consistently lower. The model seems to work well and the described pattern is well visible not only in main agglomerations but also in smaller urban centres.

Globally, it seems that implemented logit models have meaningfully added mobility attributes to the synthetic population, and results may be considered satisfactory.

Figure 4 Share of Tickets ownership



4.2 Mode choice results

In Figure 4 transportation mode shares observed in MZ 2005 are compared with the predicted shares using the new mode choice module. Only a light underestimation of the car mode share and a parallel overestimation of public transport are observed. Still, this is not enough to consider this result satisfactory. The same as for other logit models applies and a better proof of model goodness is the spatial distribution of mode shares. Bike and walk modes do not produce traffic, and the transportation means included as “other” have a negligible share and therefore their distribution is not discussed here. The percentage of tours having car as main mode is shown in Figure 5. As expected, in countryside areas this percentage is at its highest level. In towns, especially in bigger ones, is relatively low. Maybe in some cases (for example Zürich) it would have been expected a higher percentage of car tours in municipalities around the main one of the agglomeration.

Figure 5 Aggregate market shares of the modes

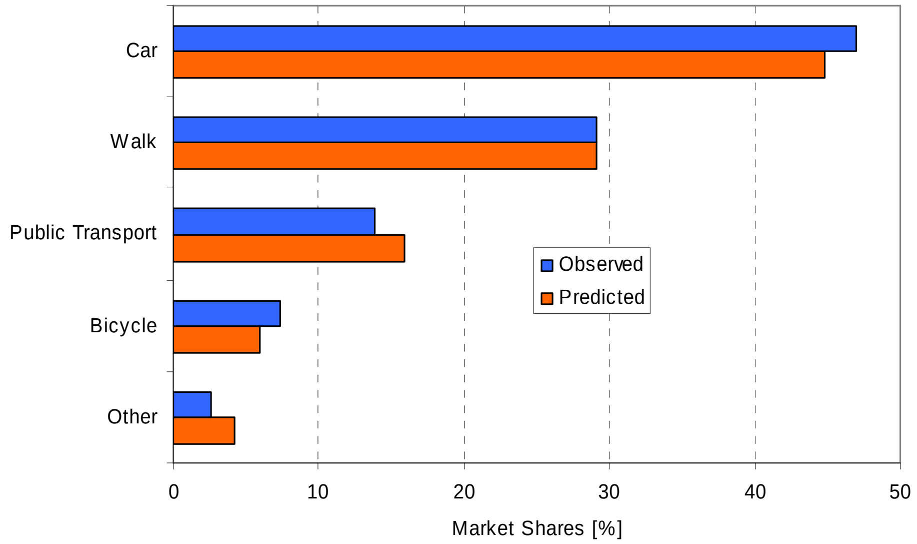


Figure 6 Share of Car as Tour main mode

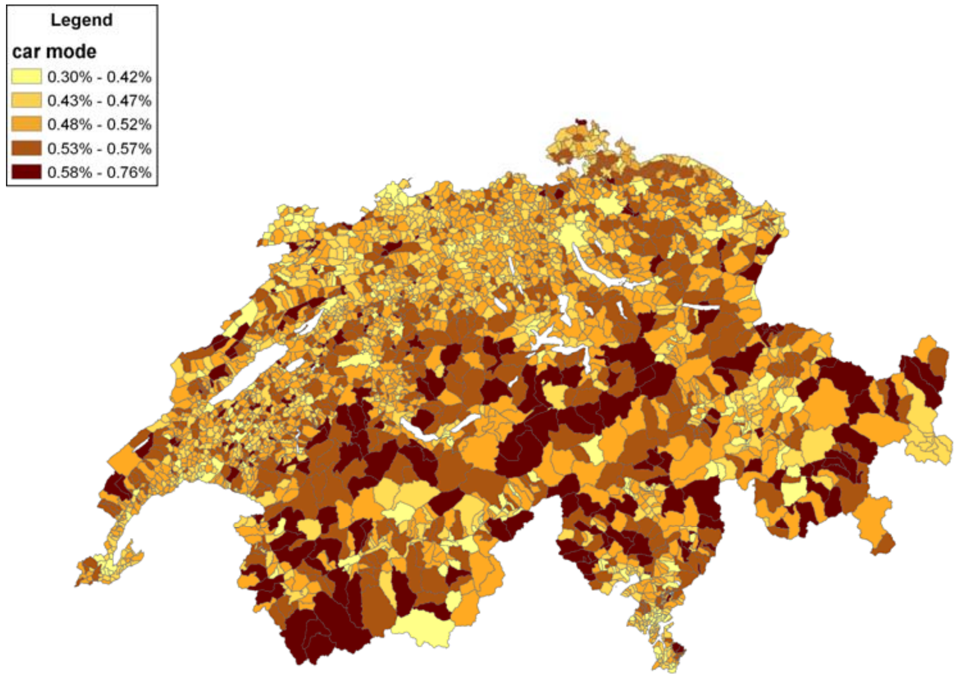
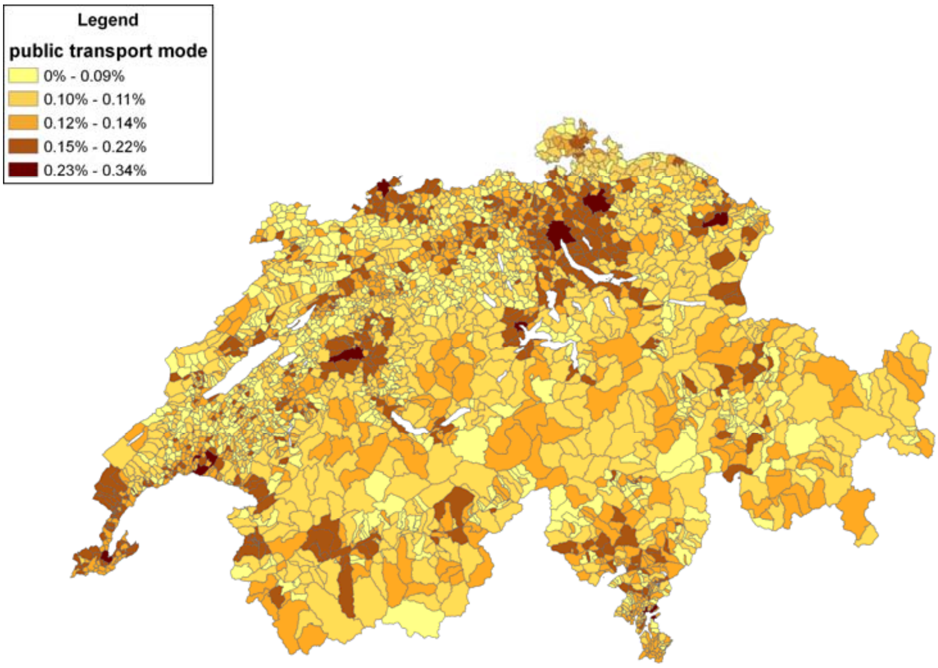


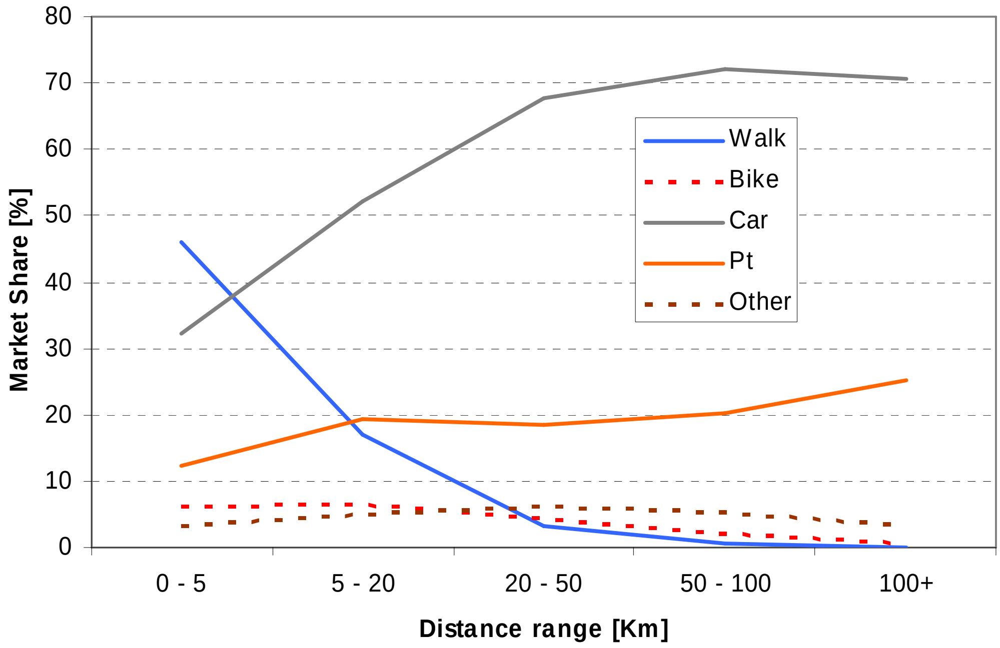
Figure 6 reproduces the share of tours having public transport as main mode. As expected the share is particularly high in urban areas with highest levels in big agglomerations main municipalities. In the countryside this share is considerably lower.

Figure 7 Share of Public Transport as Tour main mode



Another plot which can provide some interesting information is the Distance-Mode plot, where the market shares of modes is plotted against travelled distance of the tour (Figure 7)

Figure 8 Main Mode - Distance



Some observations can be made:

- The car mode is the most used for all distance ranges except for very short tours.
- Public transport has a similar behaviour compared with car, but always a smaller market share. It gains same share against car for very long tours.
- Walk is the preferred mode for short tours. As the distance increases the percentage of tours with walk as main mode decreases dramatically.
- Modes other than car and public transport are almost negligible for tours longer than 50 km, for tour of more than 100 km practically all the market is absorbed by car and public transport mode.

These observations bring us to the conclusion that also from this point of view the model prediction is quite meaningful since the same pattern can be observed also in Mikrozensus data.

From the discussion above we can therefore conclude that the model delivers results which are consistent with available survey data, and their accuracy is suitable for the pre-process stage of MATSim-T.

5. Summary and Outlook

Two main issues were the subject of this paper, the generation of synthetic population and the modelling of initial demand in micro-simulations. Both were discussed in the context of MATSim –T which is a multi-agent transport simulation tool.

At the moment MATSim –T is set to simulate transportation in Switzerland, and therefore the synthetic population that is generated need to have characteristics consistent with the Swiss population. Data which were used for that scope were presented and methodologies used in order to generate and enrich the synthetic population discussed. In particular, it was shown that multinomial logit models can be successfully used to add information to an already generated population.

The modelling of individual demand in MATSim-T was also presented. It involves the assignment of activity chains among the population consistently with individual attributes. The aspect of the choice of reference activity chains was also dealt with. Finally the transport mode choice was discussed, presenting the logit models which are used to estimate it.

All the above constitute together a coherent path in order to obtain a meaningful initial demand at the individual level. Results show that the presented methods are effectively able to generate a population which realistically reproduces the Swiss population. The transportation mode choice shares of the modelled initial demand are also consistent with Swiss MZ data.

A good model of the initial demand in the pre-process stage of MATSim-T guarantees that the optimization stage will be not too costly. Until now no mode choice module was included in MATSim, and therefore the work presented in this paper must be intended as a first step in this direction. In the future a mode choice module will be integrated also in the optimization stage. When also this tool will be integrated in MATSim-T, it will be possible to estimates if the mode choice module in the pre-process stage will need to be improved or not. One possible change, for example, will be to extend the mode choice module encompassing also sub-tours. Another innovation which is already on the way is the introduction of household data. It was pointed out that, at the moment, some of this data can be indirectly retrieved. But the introduction of an explicit household layer in the population data structure will allow for some improvement in the logit models which are now in use.

It would be also interesting to perform an Anova exercise in order to identify population groups and use different chains distributions among them measuring the impact of commuting distances on such distributions.

6. References

Balmer, M., K.W. Axhausen and K.Nagel (2005) A demand generation framework for large scale micro-simulations, paper presented at the *85th Annual Meeting of the Transportation Research Board*, Washington D.C., January 2006

Beige, S. and K.W. Axhausen (2004) Ownership of mobility tools in Switzerland, paper presented at the *4th Swiss Transport Research Conference*, Ascona, March 2004

Beige, S. and K. W. Axhausen (2006) Long-term mobility decisions during the life course: Experiences with a retrospective survey, paper presented at the *11th International Conference on Travel Behaviour Research,* Kyoto, August 2006

Bierlaire, M. (2007) An introduction to BIOGEME Version 1.5, biogeme.epfl.ch

Bierlaire, M. (2003) BIOGEME: A free package for the estimation of discrete choice models, Proceedings of the *3rd Swiss Transportation Research Conference*, Ascona, March 2003

Bundesamt für Statistik and Bundesamt für Raumentwicklung (2006) Mikrozensus Verkehrsverhalten 2005, Bundesamt für Statistik, Neuenburg.

Bundesamt für Statistik (2001) Volkszählung 2000, Bundesamt für Statistik, Neuenburg.

Edelhoff, T., H. Schilling, M. Balmer and R. Möhring (2007) Optimal route assignment in large scale micro-simulations, *Arbeitsberichte Verkehrs- und Raumplanung*, 409, IVT, ETH Zürich, Zürich

Frick, M. and K.W. Axhausen (2003) Generating Synthetic Populations using IPF and Monte Carlo Techniques: Some New Results, paper presented at the *4th Swiss Transport Research Conference*, Ascona, March 2004

Meister, K., M. Balmer, K.W. Axhausen and K. Nagel (2006) planomat: A comprehensive scheduler for a large-scale multi-agent transportation simulation, paper presented at the *11th International Conference on Travel Behaviour Research*, Kyoto, August 2006.

Ortuzar, J.D.D. and L.G. Willumsen (2006) Modelling Transports, John Wiley & Sons, Chichester

Vrtic, M., Ph. Fröhlich, N. Schüssler, K.W. Axhausen, S. Dasen, S. Erne, B. Singer, D. Lohse and C. Schiller (2005) Erzeugung neuer Quell-/Zielmatrizen im Personenverkehr, *Report to the Bundesämter für Raumentwicklung, für Strassen und für Verkehr*, IVT, ETH Zürich, Zürich