



Conference Paper

Pairing discrete mode choice models and agent-based transport simulation with MATSim

Author(s):

Hörl, Sebastian; Balać, Milos; Axhausen, Kay W.

Publication Date:

2019-01

Permanent Link:

<https://doi.org/10.3929/ethz-b-000303667> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

1 **Pairing discrete mode choice models and agent-based transport simulation with MATSim**

2 Date of submission: 2018-08-01

3

Sebastian Hörl

IVT, ETH Zürich, 8093 Zürich, Switzerland

phone: +41-44-633-38-01

email: sebastian.hoerl@ivt.baug.ethz.ch

4 orcid: 0000-0002-9018-432X

5

Milos Balac

IVT, ETH Zürich, 8093 Zürich, Switzerland

email: milos.balac@ivt.baug.ethz.ch

6 orcid: 0000-0002-6099-7442

7

Kay W. Axhausen

IVT, ETH Zürich, 8093 Zürich, Switzerland

phone: +41-44-633-39-43

email: axhausen@ivt.baug.ethz.ch

8 orcid: 0000-0003-3331-1318

9 Words: 7249 words + 1 table = 7499 word equivalents

1 **ABSTRACT**

2 As travellers are faced with an increasing portfolio of transportation options, researchers are similarly
3 faced with increasing complexity of modelling efforts to study people's choices and behaviour.
4 While discrete-choice models and in particular mode-choice models are widely used to study how
5 people react to specific changes in the system, little published research exist that analyses the
6 possibilities and pitfalls of pairing mode-choice models with the traffic simulation inside of an
7 iterative process.

8 The work presented here describes a structured framework for using discrete choice models
9 along with microsimulation. While the outcomes are based on the MATSim framework, they can be
10 generalised. The obtained results show that the combination of a mode-choice model with MATSim
11 is a promising approach to set up a feedback-enabled transport simulation. Given well-designed
12 constraints on top of the choice model, a good fit with the reference data is achieved. While the
13 modeller loses some of the freedom he or she has within the plan modelling in MATSim, gains in
14 computation time and a reduced effort for calibration are achieved.

15 The authors find that a tour-based model formulation is to be preferred over a trip-based one
16 because by construction more consistent travel decisions are made. While a trip-based model could
17 probably be calibrated to yield a good fit with MATSim, the tour-based model bears the potential of
18 not having to perform a lot of calibration work when setting up the simulation.

1 INTRODUCTION

2 As the travellers are faced with an increasing portfolio of transportation options, researchers are
3 similarly faced with increasing complexity of modelling efforts to study people's choices and
4 behaviour. In recent decades new transportation modes like bike-sharing, car-sharing and ride-
5 hailing have widened the range of services and impose new challenges for research. The coming
6 era of automated vehicles promises to cause a new revolution in how people travel. Therefore,
7 computationally fast models to estimate and predict the behaviour of people are crucial to inform
8 practitioners and policymakers alike to shape the implementation of new transportation schemes.

9 Traditionally, travel behaviour of people is predicted using discrete choice models. Large parts
10 of the transport research community have their primary interest in discrete choice modelling, i.e.,
11 predicting a traveller's choice for a particular mode, route or destination. Usually, those are based
12 on the expected travel characteristics, such as travel time and cost, in combination with personal
13 preferences and socio-demographic factors. The primary rationale behind the approach is that
14 a statistical model, estimated from stated preference data (e.g., from a travel survey) or revealed
15 preference data (e.g., GPS tracking) can predict the individual travel decisions of people. Ultimately,
16 by modifying individual variables, such as the travel time, one can predict how the travel decisions
17 of people would change. However, these changes in the traffic system are imposed as changes in the
18 exogenous variables of the model and usually do not allow for the analysis of feedback effects.

19 Only recently, activity- and agent-based transport simulators have emerged. Some of them do
20 allow for these feedback effects in different forms. One of these agent-based models is the transport
21 simulation framework MATSim (1). Unlike some of the other agent-based models, MATSim does
22 currently not allow for the direct inclusion of discrete mode choice models. To the best knowledge
23 of the authors, there has been only one documented effort in the MATSim community to investigate
24 the potential of a combination of the two modelling approaches (2).

25 The authors of the paper at hand want to further the research presented in (2) and provide a
26 framework for pairing mode-choice models with microsimulation in MATSim. Furthermore, a case
27 study based on a multinomial logit model is presented for the city of Zurich, Switzerland. Here, the
28 choices of agents for different model complexities are analysed, and the benefits and possible pitfalls
29 are discussed.

30 BACKGROUND

31 Powerful tools for the simulation of the transport system exist. However, most of them assume
32 a static demand, i.e. a fixed set of trips or flows that need to be handled by the transport system.
33 The emergence of agent-based transport simulation has made it possible also to consider feedback
34 loops between supply and demand in more detail. For instance, if too many people use a car in
35 the system, travel times will go up, so people will try to switch to other modes of transport until
36 an equilibrium between supply and demand is found. Lately, research is ongoing on the topic of
37 combining agent-based transport simulations (which can produce network conditions given a set of
38 trips) and discrete (mode) choice models (which can produce realistic mode shares given the state of
39 the transport system). POLARIS (3) is one example of such a model, where mode choice is one
40 part of the decision process of the simulated agents. They perform mode choices on a trip level
41 and also consider vehicle availability constraints - meaning that a vehicle is only available if it was

1 taken to reach the current location. (4) incorporate a mode choice model within an agent-based
2 model to simulate carsharing operations in the city of Lisbon. Mode choice modelling is performed
3 on a trip level, and vehicle constraints are not considered. Therefore additional calibration of the
4 mode choice parameters was performed. (5) present a multi-day agent-based simulation that uses a
5 discrete choice model to obtain the modes for each of the agents' trips while considering vehicle
6 constraints. However, they constrain people's choices that if the car was used on a previous trip, it
7 needs to be used on the following, which prevents many of the complex mode decisions that could
8 be made in reality. The TASHA framework makes use of a tour-based mode choice model for the
9 Toronto area (6), which distinguishes between trip chains served entirely by the private car mode or
10 by flexibly combinable other modes.

11 Even though the above mentioned agent-based models that incorporate mode choice models are
12 not covering the entire literature, they represent the current state of modelling in this field. However,
13 there is lack of a thorough analysis of such combinations.

14 **MATSim**

15 For the use cases in this study the agent- and activity-based transport simulation framework MATSim
16 is used to perform the network simulations. While the authors are affiliated with the software, the
17 results that are obtained in this work are valid for any set-up where one has:

- 18 • A discrete choice model with specific choice dimensions (e.g. for a mode choice model the
19 expected travel times and costs for each mode), which sets up a demand given the supply.
- 20 • A network simulation model which produces realistic network conditions given a specific
21 demand, i.e. it is producing the supply characteristics.

22 In the case of MATSim, the network conditions are produced by simulating thousands of agents
23 (travellers) in the same capacitated network. If too many people want to use a specific road segment
24 in the network they will be slowed down because traffic jams emerge. In MATSim this is possible
25 because each link in the network is simulated as a queue with limited capacity. Therefore even
26 spillback effects can be simulated using the framework.

27 On the demand side, MATSim relies on a so-called "co-evolutionary algorithm". Each agent
28 in this framework has a daily plan with several activities with desired start and end times. Those
29 activities are connected by trips, for which a specific mode of transport is planned. The selected
30 plans of all agents are iteratively simulated in the network simulation as described above. After one
31 iteration (usually equal to one day) all activities and trips that have been performed are translated
32 into a score. Performing activities at the right time for the right duration gives positive scores while
33 delays or spending time in traffic give negative scores. The cumulative score is then a measure for
34 the usefulness of the plan. After each iteration, a fraction of agents slightly modify their selected
35 plan and add it back to the set of known plans. Such modifications may be the change of mode of
36 transport or the route of one or several trips, or the adjustment of the departure time from an activity.
37 Then, the plan will be simulated in the next iteration, and a new score is obtained. If an agent does
38 not modify his selected plan, he chooses between the known ones (which are usually limited to a
39 fixed number) according to their score. So plans with a high score are more probable to be chosen
40 and also more probable to be modified and improved. Letting this algorithm run for a large number
41 of iterations has the following effects: Since all agents perform this process at the same time, they

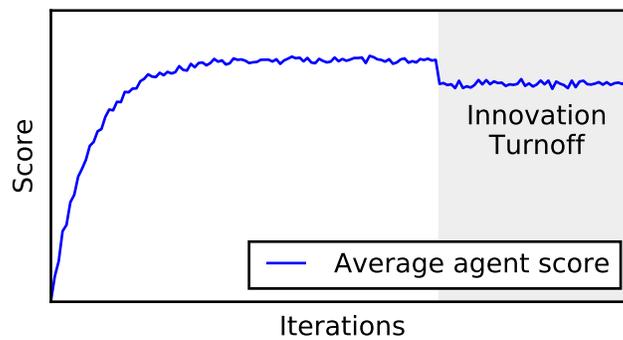


FIGURE 1 Schematic visualization of the procession of scores in a typical MATSim run.

1 react to the system state (travel times) that are produced by all of their plans. Therefore, there is
 2 a feedback loop between supply and demand, i.e. the travel times in the system may fluctuate,
 3 but stabilise once gets close to equilibrium. Ultimately, the only fluctuations in system state are
 4 produced by the random modifications that are applied to a fraction of plans after each iteration.
 5 Figure 1 shows this process from the perspective of the average agent score over many MATSim
 6 iterations.

7 The algorithm has one significant advantage: It is very versatile. Basically, as a modeller one
 8 can propose any changes to the plans of the agents and as long as one has a consistent scoring
 9 framework one will arrive at a stochastic user equilibrium. Those changes can be very simple, e.g.
 10 just random changes of transport modes for randomly selected trips in an agent's plan. A thorough
 11 scoring framework would then penalise plans that are poorly suited (e.g. walking for 50km) or
 12 merely infeasible (e.g. by giving a high penalty to plans where the car is left behind somewhere in
 13 the network without returning home). The two downsides are that (1) the process takes very long
 14 because also pointless realisations of the plan need to be scored out (e.g. walking for 50km) and (2)
 15 each plan need to be tested at least once in the network simulation. So if around 5% of the agents
 16 perform replanning after each iteration, there are potentially 5% of all agents that perform nonsense
 17 alternatives. This has the potential to slow down the convergence extremely (because it leads to
 18 large unnecessary fluctuations in the network conditions) and forbids a consistent estimation of
 19 the network state for the actual analysis that one wants to perform. To counteract the latter point
 20 the replanning ability is turned off for some iterations after the mean of the average scores in the
 21 population has stabilised (i.e. there is no upward trend anymore). This makes sure that only the
 22 available plans are then selected at random, and hopefully plans with bad scores do not happen
 23 frequently. Furthermore, this "innovation turnoff", as depicted in Figure 1, leads to a jump in
 24 scores, which represents a strong non-linearity. Therefore, any optimisation algorithm that takes
 25 into account the score of early iterations to improve some scenario parameters (e.g. road pricing)
 26 becomes almost useless, because the step response can go in any direction.

27 Furthermore, MATSim simulations are notoriously hard to calibrate: Running a simulation with
 28 millions of agents can take minutes to hours (for one iteration), so a whole run until equilibrium can
 29 last days. When trying various values for the scoring of the travel dimensions, one needs to run
 30 a multitude of those simulations, which is usually not feasible and hence only allows for a rough
 31 calibration of the models.

1 Discrete Choice Models

2 Discrete choice models are statistical models that assign probabilities to specific alternatives for a
 3 decision task. In a mode choice model, these alternatives would be the different modes of transport.
 4 Each of the alternatives has some choice dimensions assigned, e.g. the travel time for each mode or
 5 the costs. By using a utility function, similar to the scoring in MATSim, a utility for each alternative
 6 is calculated, for instance:

$$\begin{aligned}
 7 \quad V_{\text{car}} &= \alpha_{\text{car}} + \beta_{\text{car}} \cdot t_{\text{car}} + \beta_{\text{cost}} \cdot c_{\text{car}} \\
 V_{\text{walk}} &= \alpha_{\text{car}} + \beta_{\text{walk}} \cdot t_{\text{walk}}
 \end{aligned}
 \tag{1}$$

8 Here the t would stand for the (estimated / expected) travel time for a trip, and c would be the
 9 cost. The β s are linear coefficients that translate all choice dimensions into a generalised utility, and
 10 the α s are mode-specific constants that capture the unobserved components in the decision making.
 11 Using these utilities statistical models can be set up. In the case of a simple Multinomial Logit
 12 Model, one would have:

$$13 \quad \mathbb{P}(\text{car}) = \frac{\exp(V_{\text{car}})}{\exp(V_{\text{car}}) + \exp(V_{\text{walk}})}
 \tag{2}$$

14 Thus, one obtains a probability of choosing either alternative given the values for the choice
 15 dimensions in the current decision task. Through statistical inference all model parameters (α and
 16 β) can be estimated from stated choices in surveys or revealed choices in GPS traces or similar data
 17 sets.

18 The primary advantage is that those mode choice models can be estimated in a very short time
 19 compared to a MATSim run. It not only allows the modeller to estimate various parameters but even
 20 to try various formulations of the utility function. These models are backed by thorough theory,
 21 they are a standard tool in transport planning and provide, given the right model formulation has
 22 been found, a solid understanding of the decision behaviour of the population.

23 The first downside of discrete choice models is that they are usually estimated on a trip basis.
 24 Therefore, all parameters that are available are estimated for an “average” trip during the day of
 25 a traveller, regardless of the context (e.g. the presence of a daily ticket for the public transport).
 26 The second downside of discrete choice models is that they do not take into account a dynamic
 27 demand situation: One needs to assume a specific travel time (but can do sensitivity analyses).
 28 However, given a large set of travellers who are confronted with similar decisions the utilisation of
 29 the infrastructure may change the travel times and therefore the likelihood of an agent to choose a
 30 specific mode. Being able to insert the equilibrium travel characteristics into the equations would
 31 produce the “actual” expected choice probabilities for a use case, not only the ones under fixed
 32 assumptions about these values.

33 Therefore, the study at hand tries to establish a loop between a detailed dynamic network
 34 simulation (in this case MATSim) and a discrete mode choice model. However, this integration is
 35 not as straightforward as it might seem. In the following chapter, the authors propose one approach
 36 of integration and outline the experienced problems and difficulties while doing so.

1 APPROACH

2 Given the advantages and disadvantages of MATSim and discrete choice models alike the task of
3 the study at hand is twofold:

- 4 1. To set up a framework that makes it easy to integrate mode choice models into MATSim
- 5 2. To show the results that we obtain by using an initially trip-based model for the whole
6 plan-based choice process

7 The basic structure of the framework that has been set up can be seen in Figure 2. On the right
8 side is the simulator, in this case MATSim. There is an agent plan that goes through the choice
9 process, such that a new updated plan is generated at the end. In the updated plan the modes of all
10 trips may have changed. Afterwards, all plans are sent to the simulator to perform the next iteration.
11 After the iteration has finished, the new travel times, service costs, congestion levels are ready to be
12 consumed by the choice model.

13 Choice Model Components

14 The different stages of the mode choice model itself (middle part in Figure 2) are described in the
15 next sections. The model exists in two different versions: A trip-based model and a tour-based
16 model. The former one sequentially assigns new modes to the trips of an agent's plan (and is
17 therefore close to the original formulation of trip-based mode choice models), whereas the latter one
18 considers whole chains of modes to choose from.

19 *Alternative Generator*

20 For each choice situation, multiple alternatives are generated. In the trip-based case the choice
21 situations are the single trips of a plan, while for a tour-based model these choice situations are tours,
22 i.e. sequences of trips. The two concepts are visualized in Figure 3: In the trip-based model, each
23 mode selection for the plan would be performed only based on the utility of the modal alternatives
24 for the current trip. The set of alternatives is always the set of feasible modes for the agent.

25 On the other hand, the plan in Figure 3 contains two home-based tours, one that consists of
26 the first four activities and one that is comprised by the last three ones. Contrary to the trip-based
27 version, where five independent choices are to be made, here only two choices are performed. The
28 choice set in those two stages is much larger because it consists of all possible permutations of
29 modes that are available for the respective tour.

30 In either case, the modes that are available are determined first for each agent through the *Mode*
31 *Availability* component. It merely acts as an early filtering of modes to avoid the processing of
32 irrelevant modes for the sake of computational speed. An example for such a filtering would be the
33 removal of the *car* mode from the set of feasible modes because the agent is under age or does not
34 have a driving license.

35 To determine how a tour is defined there is a *Tour Finder*. The standard implementation examines
36 the locations of the activities to cut the plans into home-based tours, but any more complex approach
37 is possible. Note that a pure “plan-based” model falls into this framework. If a tour is defined as
38 starting at the first activity and ending at the last activity of a plan there will be only one choice
39 situation with a choice set consisting of all feasible realisations of the agent's plans.

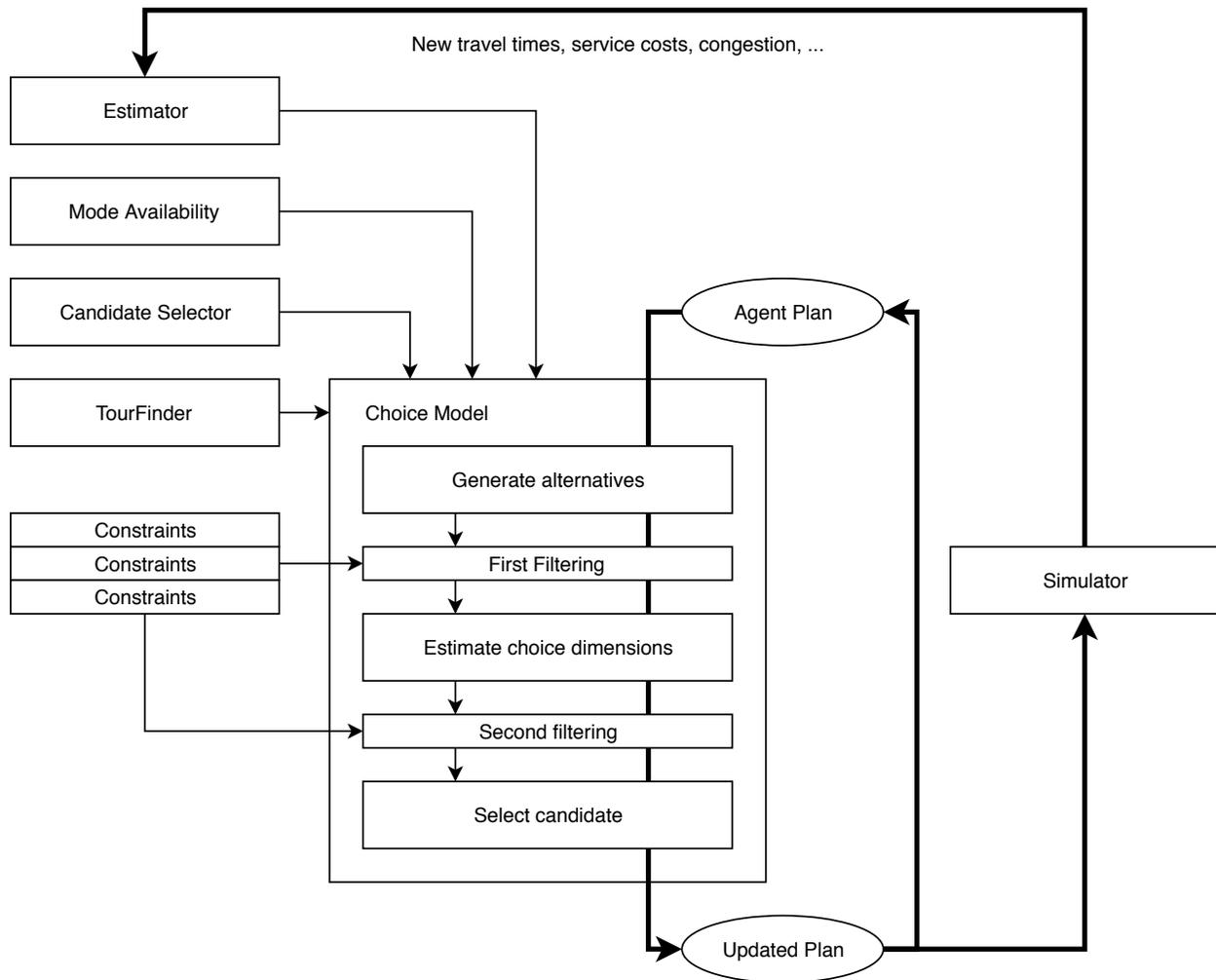


FIGURE 2 Schematic visualization of the mode choice framework.

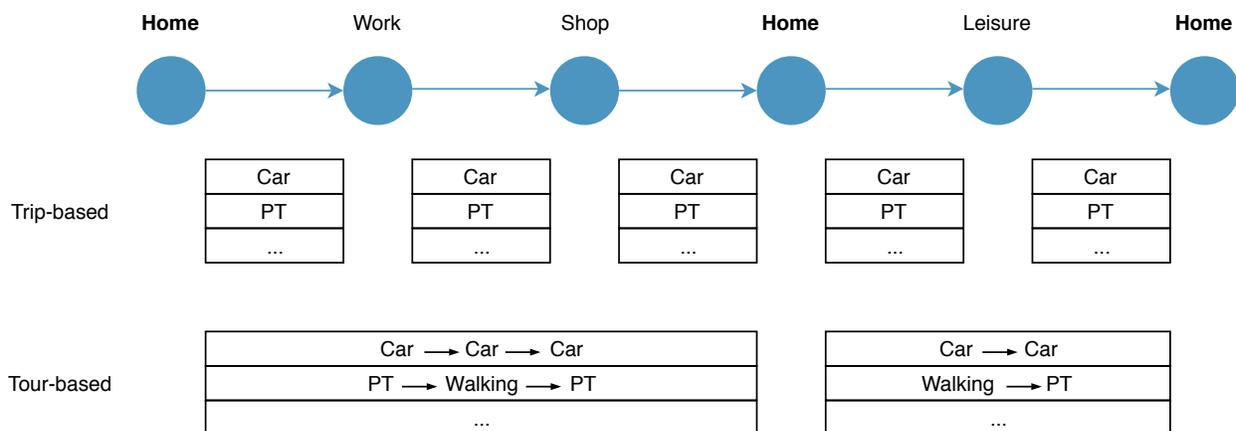


FIGURE 3 Trip-based vs. tour-based choice process

1 In a second stage, the choice situations are further filtered if they violate one of the given
 2 *constraints*. Those constraints have knowledge about the plan itself (for instance, it is known at
 3 which locations all the activities take place), and they are informed about the outcomes of the
 4 preceding choice situations. For instance, in Figure 3 the modes for the 3rd trip would be filtered
 5 *after* the choices for the first two trips have been performed. Hence, the constraints can adapt to
 6 previously taken decisions. Likewise, the second *tour* decision would have knowledge about the
 7 first one.

8 The first filtering of choices does not have knowledge about the trip characteristics themselves.
 9 For instance, if the feasibility of *public transport* is considered at this stage it cannot be filtered out
 10 because there is no service at the requested time of day. However, the framework allows for such
 11 constraints: There is a second filtering stage after all the trip and tour characteristics have been
 12 estimated, as is described in the next section.

13 *Estimation of choice dimensions*

14 The second main stage in the choice process is the estimation of the trip characteristics χ_i . Usually,
 15 it means that for each alternative i a complex routing has to be performed. Some examples of the
 16 resulting characteristics are:

- 17 • The travel time and costs by *car*.
- 18 • The waiting time, the in-vehicle time and the number of interchanges for *public transport*.
- 19 • The travel time and difference in altitude for *walk* and *bike*
- 20 • The cumulative travel costs for all *public transport* trips on a tour, considering the availability
 21 of short period tickets, single ride tickets or daily passes

22 Note that those routings are a computationally heavy task, but that in most cases a lot of caching
 23 can be performed (i.e. in most cases it is irrelevant in which tour alternative a trip is embedded to
 24 obtain the travel time). After the estimation of the trip characteristics a second filtering stage is
 25 performed as described above. Here all the information from the routing can be used.

26 What remains is a set of feasible alternatives for a choice situation with known values for the
 27 explanatory variables. Based on that data, a utility \hat{u}_i is computed for each alternative i . For that
 28 purpose any formulation of the utility function can be used. However, the idea here is to use one
 29 that has been estimated before by means of discrete choice modelling.

30 *Candidate selection*

31 Finally, given the *estimated* utilities \hat{u}_i , one of the alternatives is chosen in each choice situation.
 32 Currently, two approaches are implemented:

- 33 • **Multinomial selection** considers the estimated utilities of all alternatives, and samples one
 34 of them according to the probability $\mathbb{P}(i) = \frac{\exp(\hat{u}_i)}{\sum_{i'} \exp(\hat{u}_{i'})}$.
- 35 • **Best-response selection** selects the alternative with the highest estimated utility.

36 In this paper, only the first approach is considered because conceptually it has a higher
 37 comparability to the original discrete choice models. The main property of the approach is that
 38 decisions are stochastic: If two alternatives have roughly the same utility, they appear almost equally
 39 often in subsequent replanning attempts of the agent.

1 The best-response approach, on the other hand, makes sure that if trip characteristics have
2 converged, *always* the same option is chosen. It remains to be explored if this may be beneficial for
3 the simulation dynamics in MATSim or other simulators.

4 After the selection, the choices made are incorporated in the agent's plan.

5 **Integration with MATSim**

6 Currently, the mode choice model is integrated into MATSim in a way that only the network
7 simulation of the framework is used. Most of the other parts are disabled. For instance, the agents
8 in the test setup do not collect experienced plans and select the best among them, rather they
9 only hold one plan at a time. Furthermore, other replanning strategies, like random modifications
10 to the departure times are not taking place. The only choices that are made in the test cases are
11 mode decisions and the implicit route choices (which follow a best-response approach). Tests on
12 integrating the mode choice model into the whole MATSim environment are ongoing and pathways
13 will be discussed further below.

14 Nevertheless, some predictions can already be made: Using the framework at hand convergence
15 should be much quicker, because no irrational mode choice decisions are done at random. Likewise,
16 convergence on the supply side should be effected positively because without irrational trips taking
17 place the travel characteristics of the system can stabilise more easily.

18 While the better convergence behaviour is desirable, it also introduces a new layer of complexity
19 to the MATSim approach. Because the choice model relies heavily on reasonable values for the
20 explanatory variables, it is crucial to get these estimates right. Therefore, it is important to verify that
21 close to the equilibrium state the estimated / expected travel times and other dimensions converge
22 towards those that are actually observable in the simulation. Only then is it possible to claim that
23 the mode choice model produces consistent unbiased decisions.

24 **CASE STUDY**

25 In the following, a case study for a simulation scenario of the city of Zurich shall be presented.
26 The synthesis of the agent population, the network and all the locations in the MATSim scenario is
27 described in (7). Due to the complexity of the model with various activity types and four modes of
28 transport (*car, public transport, bike, and walking*) it used to be difficult to properly calibrate the
29 model. In the study at hand a discrete mode choice model is tested against the scenario to check how
30 well different configurations replicate the reference data. For comparison the modal split by distance
31 class is used. The reference data has been obtained from the Swiss Microcensus on Transport and
32 Mobility for 2015 (8).

33 **Model Definition**

34 To set up the mode choice model one needs to define utility functions for different modes. A mode
35 choice model for the city of Zurich is available from an ongoing project (Table 1). Based on this

1 model the utility functions for the four modes are defined as follows:

$$\begin{aligned}
 u_{\text{car}}(\chi) &= \alpha_{\text{car}} \\
 &+ \beta_{\text{travelTime,car}} \cdot \chi_{\text{travelTime,car}} \\
 &+ \beta_{\text{travelTime,car}} \cdot \theta_{\text{parkingSeachPenalty}} \\
 &+ \beta_{\text{travelTime,walk}} \cdot \theta_{\text{accessEgressWalkTime}} \\
 &+ \beta_{\text{cost}} \cdot \left(\frac{\chi_{\text{crowflyDistance}}}{\theta_{\text{averageDistance}}} \right)^\lambda \cdot \chi_{\text{cost,car}}
 \end{aligned} \tag{3}$$

$$\begin{aligned}
 u_{\text{pt}}(\chi) &= \alpha_{\text{pt}} \\
 &+ \beta_{\text{numberOfTransfers}} \cdot \chi_{\text{numberOfTransfers}} \\
 &+ \beta_{\text{inVehicleTime}} \cdot \chi_{\text{inVehicleTime}} \\
 &+ \beta_{\text{transferTime}} \cdot \chi_{\text{transferTime}} \\
 &+ \beta_{\text{accessEgressTime}} \cdot \chi_{\text{accessEgressTime}} \\
 &+ \beta_{\text{cost}} \cdot \left(\frac{\chi_{\text{crowflyDistance}}}{\theta_{\text{averageDistance}}} \right)^\lambda \cdot \chi_{\text{cost,pt}}
 \end{aligned} \tag{4}$$

$$u_{\text{bike}}(\chi) = \alpha_{\text{bike}} + \beta_{\text{travelTime,bike}} \cdot \chi_{\text{travelTime,bike}} + \beta_{\text{age,bike}} \cdot \max(0, a_{\text{age}} - 18) \tag{5}$$

$$u_{\text{walk}}(\chi) = \alpha_{\text{walk}} + \beta_{\text{travelTime,walk}} \cdot \chi_{\text{travelTime,walk}} \tag{6}$$

6 While the α , β and λ variables stem from the choice model, the θ represent calibration parameters
 7 that needed to be included to allow for a fair competition of the four modes, or constants that
 8 are used in the original mode choice model formulation. The a paramters are agent-specific
 9 socio-demographic attributes. All χ are explanatory variables that need to be estimated from the
 10 simulation.

11 The costs $\chi_{\text{cost,car}}$ and $\chi_{\text{cost,pt}}$ are backed by rather detailed cost calculations, which are omitted
 12 here for brevity. The values used in the study at hand are summarized in Table 1.

13 For the tour-based experiments home-based tours are chosen, i.e. all trips on a travel sequence
 14 from and to home are jointly assigned a utility. To do so the *estimated* utilities \hat{u}_j of the single trips
 15 j on tour alternative i are summed up:

$$\hat{u}_i = \sum_j \hat{u}_{i,j} \tag{7}$$

17 Note that this “total utility approach” is not an obvious choice. In fact, it is one formulation
 18 among many, some of which have been examined in more detail in (2).

Car	α_{car}	0.827	
	$\beta_{travelTime,car}$	-0.0667	[min ⁻¹]
Public Transport	α_{pt}	0.0	
	$\beta_{numberOfTransfers}$	-0.17	
	$\beta_{inVehicleTime}$	-0.0192	[min ⁻¹]
	$\beta_{transferTime}$	-0.0384	[min ⁻¹]
	$\beta_{accessEgressTime}$	-0.0804	[min ⁻¹]
Bike	α_{bike}	-0.1	
	$\beta_{travelTime,bike}$	-0.0805	[min ⁻¹]
	$\beta_{age,bike}$	-0.0496	[a]
Walking	α_{walk}	0.63	
	$\beta_{travelTime,walk}$	-0.141	[min ⁻¹]
Others	β_{cost}	-0.126	[CHF ⁻¹]
	λ	-0.4	
	$\theta_{averageCrowflyDistance}$	40	[km]
Calibration	$\theta_{parkingSearchPenalty}$	6	[min]
	$\theta_{accessEgressWalkTime}$	5	[min]

TABLE 1 Parameters of the discrete mode choice model

1 Constraints

2 In order to allow for a realistic travel behaviour vehicle constraints need to be introduced. They
3 make sure that a vehicle can only be used where it has been moved to before. Furthermore, it should
4 be made sure that the vehicle arrives back home after each day, here we require that it does so after
5 each tour.

6 For the tour-based model, such a constraint is quite easy to establish. Because a tour is considered
7 it is already clear that the start- and endpoint of the sequence under consideration is at the same spot.
8 Therefore, it remains to check the following constraint on the trips included in a tour:

ALGORITHM 1: Tour-based vehicle continuity constraint

Initialize: $vehicleLocation \leftarrow$ Agent's home location

For each trip with mode car **do:**

If the origin of the current trip is **not** $vehicleLocation$ **Then**

Return Constraint is violated

End If

$vehicleLocation \leftarrow$ destination of current trip

End For

If $vehicleLocation$ is **not** agent's home location **Then**

Return Constraint is violated

End If

9 Note that in this constraint one never needs to know about the previous element, i.e. the tour that
10 comes before the one in question. For a trip-based model this is necessary. A very simple version

1 would be only one check:

ALGORITHM 2: Simple trip-based vehicle continuity constraint

If *car* has been used in a preceeding trip **Then**
 vehicleLocation ← destination of the last preceeding trip with mode *car*
Else
 vehicleLocation ← Agent's home location
End If
If *car* shall be used in current trip **Then**
 If origin of current trip is **not** *vehicleLocation* **Then**
 Return Constraint is violated
 End If
End If

2 Here, one makes sure that whatever happened in the trips before is consistent with the agent now
 3 attempting to use his car at the current location. However, this doesn't make sure that the car arrives
 4 back at home. In fact, to achieve this, the constraint becomes

ALGORITHM 3: Advanced trip-based vehicle continuity constraint

If *car* shall be used for the current trip **Then**
 Execute Algorithm 2
Else
 Obtain *vehicleLocation* as in 2
 If all of
 vehicleLocation is at current origin
 vehicleLocation is **not** at agent's home location
 current origin is **not** visited again later in the tour
 Then
 If any mode except *car* shall be used for current trip **Then**
 Return Constraint is violated
 End If
 End If
End If

5 The first part of this advanced trip constraint is exactly as in the simple one: To make sure that
 6 the vehicle is present if one wants to depart with it. The second part makes sure that the vehicle
 7 returns home: If an agent shall perform a trip from a location (maybe walking), but by doing so it is
 8 clear that there is no chance of returning the car home, the agent *must* use his car for this trip, i.e. all
 9 other modes are forbidden.

10 So where is the difference between the advanced trip constraint and the tour constraint? It lies
 11 in the order of decisions. In the trip-based model all choices are made step by step. So it may
 12 make sense for the first trip itself to be performed walking because it is just a short trip to a shop.
 13 However, once this decision is made all following decisions are dependent on it. So after the shop
 14 the agent may go to work, which is 50km away. He is now bound to use public transport although

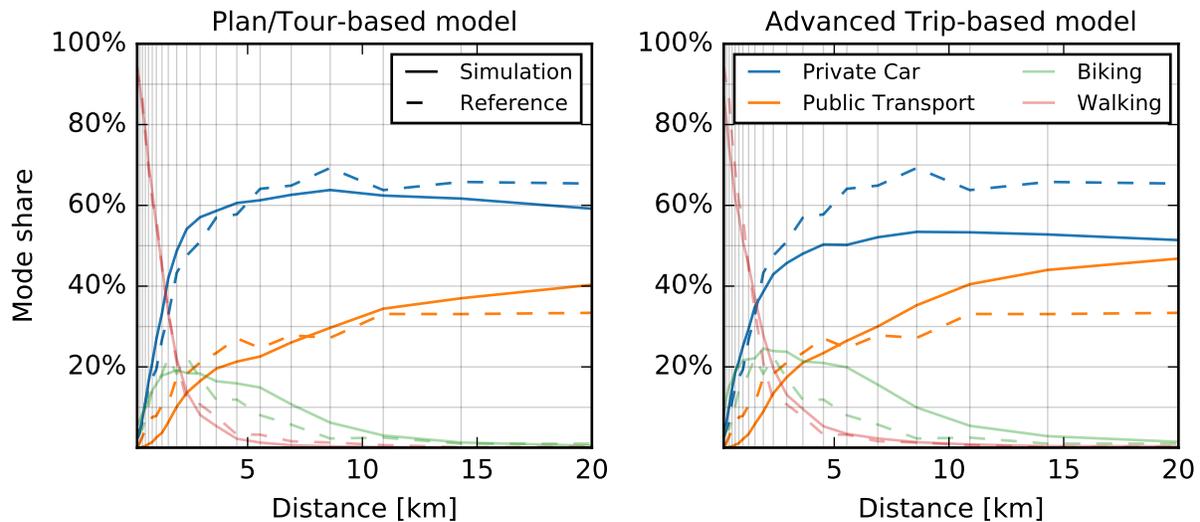


FIGURE 4 Comparison of mode shares by distance class for the plan/tour-based model and the trip-based model. The reference points are the modal shares at twenty quantiles of the reference distance distribution (vertical lines).

1 the connection might be rather bad. In the tour-based model all decisions in one tour are made
 2 jointly. Because of the large negative utility that a public transport commute would bring the score
 3 of the whole tour would be dragged down and hence it would make sense to use a car for the long
 4 trip. However, to comply with the vehicle constraint, the car would also be used for the short first
 5 trip. In general, making choices on the tour-level should produce much more reasonable decisions
 6 than making choices for each trip.

7 Results

8 The following sections are used to show the differences between the tour-based and trip-based model,
 9 as well as the differences between the two trip constraints.

10 Mode shares

11 First of all it is interesting to explore how well the models reproduce the mode shares that can be
 12 observed in reality. Figure 4 shows the comparison between the reference and two realizations
 13 of the mode choice model: On the left side the tour-based model is used while on the right side
 14 the trip-based one is shown. The first observation is that the tour-based model approximates the
 15 reference quite well while the trip-based model with the advanced trip constraint has a larger error.
 16 Especially the *car* mode is under-represented, which can be explained by the reasoning in the
 17 previous chapter. Since the first trip has a large weight in the decision process, agents for which
 18 it makes sense to use their car later during the day but have decided against it on the first trip are
 19 forced to use other modes. The same is true for agents who chose to use their bike for the first trip.

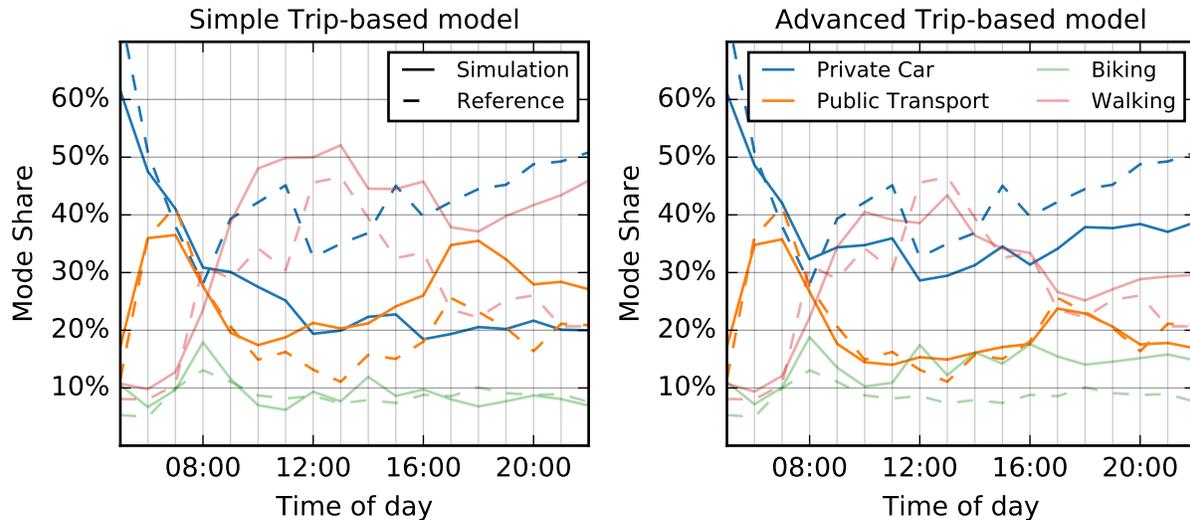


FIGURE 5 Comparison of the simple vehicle continuity constraint and the advanced one for the trip-based model

1 *Trip-based constraints*

2 The second comparison is between the two trip-based constraints. Figure 5 shows the simple one on
 3 the left side and the advanced one on the right side. In this case the mode share by time of day is
 4 shown. Clearly, the simple constraint does not consider whether a vehicle arrives at home at the
 5 end of the day. Due to the trip-by-trip decision making cars become less and less used over the
 6 course of the day and most of the vehicles are left behind somewhere in the city. The advanced
 7 constraint improves the situation. Despite the agents still using the car less than they are supposed
 8 to the constraint forces them to return home at least. Therefore, more *car* trips can be seen in the
 9 evening hours.

10 *Runtime*

11 Finally, the runtime of the different approaches is compared in Figure 6. On the left side the
 12 constrained models can be seen whereas on the right side the constraints have been deactivated
 13 (which means that the model produces inconsistent plans).

14 For the constrained models, the tour/plan-based model takes longer to execute. This is due to the
 15 higher effort in generating the more complex chains rather than simply providing a set of feasible
 16 modes. Furthermore, the advanced trip constraint takes more time to run than the simple one. This
 17 is due to the expensive checking whether a specific mode should be enforced.

18 Interestingly, the constrained models are quicker in comparison to the unconstrained ones. This
 19 may come as a surprise, but is easy to explain. The unconstrained case considers every possible
 20 alternative whereas the constrained models filter out most of them on the fly. As described above,
 21 most of these constraints already hit before any (computationally expensive) routing is performed.

22 To put the runtime in perspective it should be noted that the scenario uses a sample of the agent

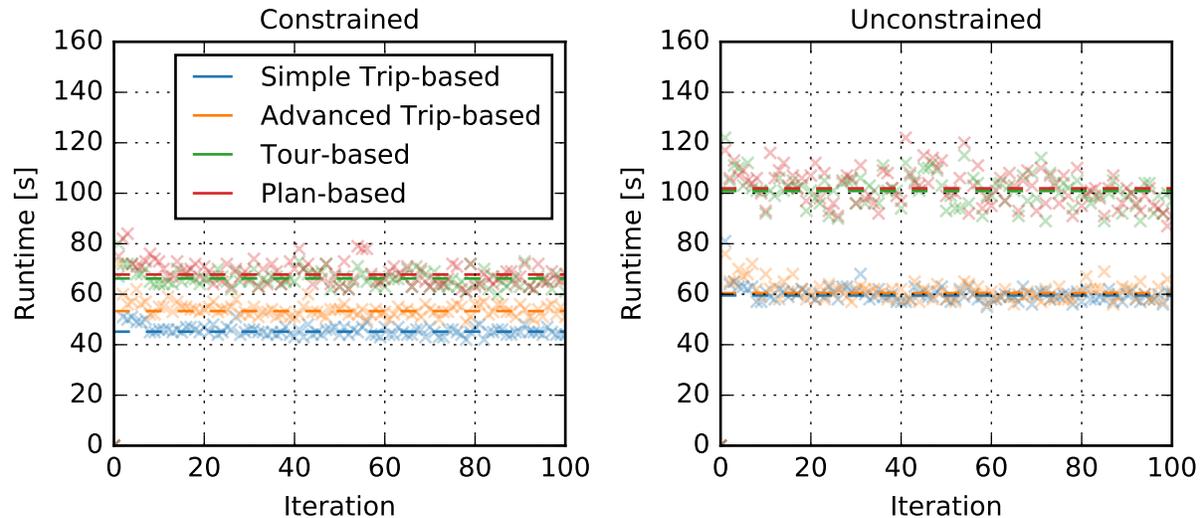


FIGURE 6 Runtimes for the mode choice models, compared with the unconstrained case.

1 population of around 200k agents. In each iteration around 22k of them perform mode choice. The
 2 simulations are performed on a slightly outdated Intel Xeon E7-4870 machine from 2011 on 80
 3 cores. It should be noted that since the decisions are made independently for all agents, the choice
 4 process can be parallelised up to the hardware limits. The obtained runtimes and prospects for
 5 scalability promise a good performance for a 100% sample or even a whole Switzerland scenario.

6 Estimation quality

7 While the fit with the reference data may seem good for the tour-based model, one needs to be
 8 careful that the choice model is operating on the right data. Here, this specifically means that one
 9 should make sure that the estimated travel times that are fed into the choice model do actually predict
 10 the travel times that are observed in the simulation.

11 For that purpose the prediction error is tracked: Whenever an agent makes the choice to use *car*
 12 for a trip, the estimated travel time, provided by MATSim, is saved. Right after, the plan is simulated
 13 in MATSim and the travel time is measured *during* the simulation. Hence, a pair of predicted and
 14 observed travel times is available for all of those trips. Figure 7 shows the average relative prediction
 15 error. Positive values mean that the travel time is overestimated while negative values mean that the
 16 predicted travel time is too low.

17 The median prediction error at 0% shows that in general travel times are predicted quite well
 18 except in the first iterations when the system is not yet close to the stochastic user equilibrium.
 19 However, one can also see that the average error is at -2.5% which means that the travel times are
 20 slightly underestimated. There are multiple reasons why this can happen. For instance, the travel
 21 times on each link in the network are averaged over five minute intervals by default in MATSim.
 22 This can lead to uncertainties and may be adjusted in future experiments.

23 While the error in the presented use case seems acceptably low and it can be assumed that the
 24 produced mode choices are almost unbiased, it is an example that shows how important it is to verify

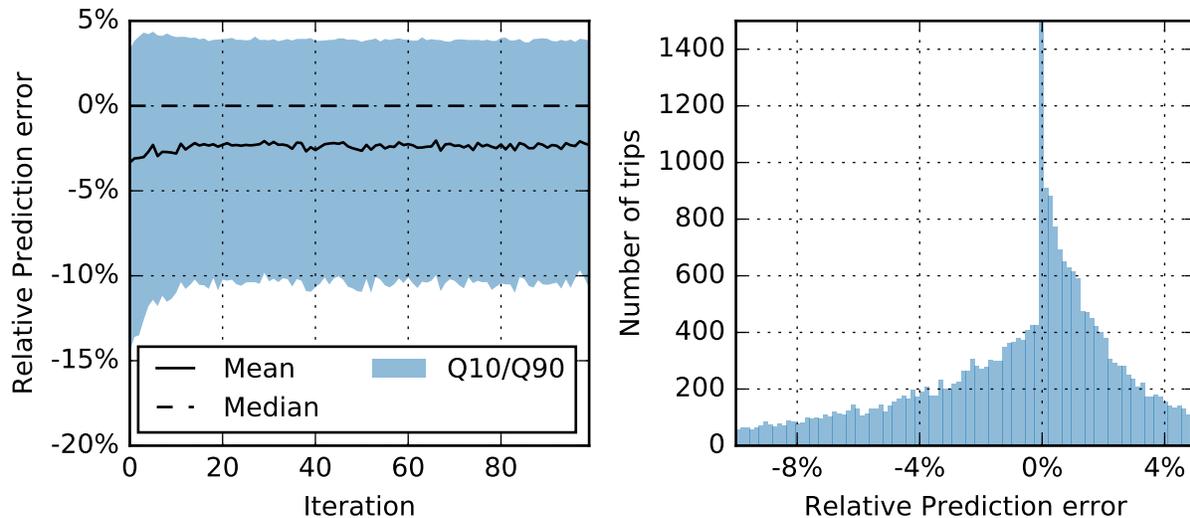


FIGURE 7 Prediction error of the tour-based mode choice model

1 the prediction quality when executing a discrete choice model in the loop.

2 DISCUSSION

3 The presented results show that the combination of a discrete choice model with a simulation tool
 4 like MATSim is a promising approach to set up feedback-enabled transport simulations. Given the
 5 constraints on top of the choice model, a good fit with the reference data is achieved. However, only
 6 one use case for the city of Zurich is presented so the results remain anecdotal. A more thorough
 7 analysis of different scenarios and also other simulation tools is needed in the future.

8 In the context of MATSim, the approach overcomes some disadvantages that the authors see in
 9 the standard approach:

- 10 • No irrelevant alternatives are executed during the simulation, because the choice model always
 11 produces a useful chain of modes.
- 12 • No “innovation turnoff” is needed and hence no strong non-linearity is introduced into the
 13 convergence dynamics.

14 However, these improvements come at a cost: The developer loses a lot of freedom. While in
 15 the standard MATSim approach virtually any plan modification strategy will eventually lead to the
 16 desired results, here it is crucial to verify that the data that is given to the choice model is consistent.
 17 Hence, the approach adds another level of complexity to working with MATSim.

18 From the comparison of the trip-based and the tour-based model it becomes evident that the
 19 latter is the better choice, because by construction more rational travel decisions are made. While a
 20 trip-based model *could* probably also be calibrated to yield a good fit with MATSim, the tour-based
 21 model bears the potential of *not* having to perform a lot of calibration work when setting up the
 22 simulation.

23 For future research there are many pathways to follow:

- 1 • **Full integration into the MATSim loop:** An important step will be to fully integrate the
2 presented approach into the evolutionary algorithm of MATSim. The authors even see
3 additional benefits in doing so: While the discrete choice model speeds up MATSim the
4 evolutionary mutation and selection algorithm can smoothen biases in the prediction of the
5 explanatory variables. Even if biased choices are made, MATSim would prefer the plans that
6 actually give the best score in the simulation and prefer those in the selection. Hence, the
7 discrete choice model would work as an importance sampler that does not necessarily need
8 100% accurate predictions.
- 9 • **Convergence criteria:** The implementation of discrete choice models, but more specifically
10 the comparison between observation and prediction could lead to new convergence criteria for
11 MATSim. If the prediction error for specific choice dimensions would fall under a predefined
12 threshold, the simulation would be marked as converged.
- 13 • **Choice model components:** On one side, it will be interesting to compare the performance
14 of a best-response selection with the multinomial approach. In addition, different formulations
15 of the tour-level utilities can be explored. While the “total utility” approach in this study may
16 be the obvious first choice, other approaches are possible. Specifically, this topic is important
17 when trying to find the best way to integrate more complex choice models such as nested or
18 cross-nested logit models.
- 19 • **Generalization of the framework:** As the last point it will be important to generalize the
20 framework and to open it to a larger audience. One of the next steps will be to make the tool
21 an open source executable to be used as a “multi-purpose discrete mode choice in the loop”
22 that can be combined with any other simulation software.

23 CONCLUSION

24 To conclude, the paper shows an approach of combining a discrete mode choice model with a
25 detailed network simulation. It is in line with a couple of ongoing research endeavours in that
26 direction but starts to explore systematically some of the pitfalls that can arise when executing
27 discrete mode choice models within a loop.

28 The paper describes a structured framework for using discrete mode choice models along with
29 customizable constraints and validation of the feedback loop dynamics to set up a consistent choice
30 process. For a case study it is shown that the parameters from a discrete mode choice model in
31 combination with well-designed constraints and embedding into a tour-based choice process can
32 yield good results without much calibration.

33 The authors point out that combining a network simulation and a discrete choice model is a
34 conceivably simple task, but that many problems can arise. The main problem is the introduction of
35 choice biases depending on the error that is attached to the estimates for the explanatory variables.

36 Although there is a large demand for further research, the presented results render a promising
37 picture for running discrete choice models in the loop with network simulation software.

38 AUTHOR CONTRIBUTION

39 The authors confirm contribution to the paper as follows: Sebastian Hörl and Milos Balac
40 conceptualised the study, and prepared the report. Sebastian Hörl developed the code base,

1 performed the simulations and did the analysis work. Prof Kay. W. Axhausen gave feedback on the
2 results, and made this research possible at the Institute for Transport Planning and Systems at ETH
3 Zurich. All authors reviewed the results and approved the final version of the manuscript.

4 REFERENCES

- 5 1. Horni, A., K. Nagel and K. W. Axhausen (eds.) (2016) *The Multi-Agent Transport Simulation*
6 *MATSim*, Ubiquity, London.
- 7 2. Hörl, S., M. Balac and K. W. Axhausen (2018) A first look at bridging discrete choice modeling
8 and agent-based microsimulation in MATSim, *Procedia Computer Science*, **130**, 900–907.
- 9 3. Auld, J., M. Hope, H. Ley, V. Sokolov, B. Xu and K. Zhang (2016) Polaris: Agent-based modeling
10 framework development and implementation for integrated travel demand and network and
11 operations simulations, *Transportation Research Part C: Emerging Technologies*, **64**, 101 – 116.
- 12 4. Martínez, L. M., G. H. de Almeida Correia, F. Moura and M. M. Lopes (2017) Insights into
13 carsharing demand dynamics: Outputs of an agent-based model application to lisbon, portugal,
14 *International Journal of Sustainable Transportation*, **11** (2) 148–159.
- 15 5. Heilig, M., N. Mallig, O. Schröder, M. Kagerbauer and P. Vortisch (2018) Implementation
16 of free-floating and station-based carsharing in an agent-based travel demand model, *Travel*
17 *Behaviour and Society*, **12**, 151 – 158.
- 18 6. Miller, E. J., M. J. Roorda and J. A. Carrasco (2005) A tour-based model of travel mode choice,
19 *Transportation*, **32** (4) 399–422, Jul 2005.
- 20 7. Bösch, P. M., K. Müller and F. Ciari (2016) The ivt 2015 baseline scenario, *16th Swiss Transport*
21 *Research Conference*.
- 22 8. Federal Statistical Office, Federal Office for Spatial Development (2017) Verkehrsverhalten der
23 Bevölkerung. Ergebnisse des Mikrozensus Mobilität und Verkehr 2015.