Adding mode choice to a multi-agent transport simulation

2008-Nov-15

Marcel Rieser
Transport Systems Planning and Transport Telematics,
Berlin Institute of Technology, D-10587 Berlin,
and
Institute for Transport Planning and Systems, ETH Zurich, CH-8093 Zurich
phone: +49-30-314 25258
fax: +49-30-314 26269
rieser@vsp.tu-berlin.de, rieser@ivt.baug.ethz.ch

Dominik Grether
Transport Systems Planning and Transport Telematics,
Berlin Institute of Technology, D-10587 Berlin
phone: +49-30-314 29521
fax: +49-30-314 26269
grether@vsp.tu-berlin.de

Kai Nagel
Transport Systems Planning and Transport Telematics,
Berlin Institute of Technology, D-10587 Berlin
phone: +49-30-314 23308
fax: +49-30-314 26269
nagel@vsp.tu-berlin.de

Words: 7500 + 8 figures + 2 table

ABSTRACT

It has been shown in the past that so-called agent-based traffic micro-simulations can be used for dynamic traffic assignment, i.e. iterative route adjustment until either a Nash equilibrium or some steady state distribution between alternatives is found. It has also been shown that the same approach can be extended to (departure) time adjustment, i.e. time adjustment and route adjustment can exist in the same iterative approach.

In this paper, it is shown that the approach can be extended to mode choice. This is implemented by forcing every synthetic traveller to consider every available model. The implementation is verified with a test case for which an approximate solution can be analytically derived, and for which it is shown that simulation and theory are consistent. It is then applied to a large-scale real-world example, the metropolitan Zurich area, with about 1 million inhabitants. For this example, it is shown that the adaptive scheme, albeit seemingly simple, is able to outperform a more traditional approach that first computes mode choice based on aggregate data, and then runs the assignment for car traffic only. Sensitivity tests show that the model reacts in meaningful ways, in particular concerning the interaction between the time structure of activities and mode choice.

INTRODUCTION

The still increasing volume on traffic asks for different, mature measures. It is generally accepted that just building new roads is not a sustainable way out of this problem, but that the amount of traffic needs to regulated, or alternative modes of transportation be used. This makes traffic forecasts more and more complex, as the proposed measures also gain in complexity. As an example, time- and vehicle-dependent roadpricing schemes could be mentioned (e.g. *(1, 2)*).

The traditional four step process (see, e.g., *(3)*) has some shortcomings with respect to such questions, since neither time-dependent (such as time-variable toll) nor mode choice problems are adequatly addressed. Mode choice is traditionally approached by so-called trip-end (after the trip generation step) or trip-interchange (after the trip distribution step) models. Trip-end models suffer from the obvious shortcoming that the accessibility of the trip destination by mode is completely irrelevant. Trip-interchange models are better, but they neglect the possible correlation between destination and mode choice. In consequence, models of simultaneous mode and destination choice were developed (e.g. *(4, 5, 6)*), sometimes on the basis of more traditional trip modelling, sometimes in the context of activity-based demand modelling. A software version of a simultaneous destination and mode choice model is the software VISEVA *(7)*.

However, all these models have in common that they eventually come up with origin-destination (OD) matrices, which are then fed into the assignment procedure. There is at least one OD matrix for the car mode, and another matrix for the non-car mode. Often these days, these matrices are time-dependent, i.e. there are different such matrices for different time slices. These OD matrices are then assigned on to the network, where quite sufficient public transit assignment routines have been developed (e.g. *(8, 9)*).

Unfortunately, however, these assignment models at the downstream end of the procedure seem to be a bit removed from the demand generating behavioral framework further upstream. For example, it seems that congestion effects need to be manually integrated, by taking an impedance matrix from the assignment and using it for the generalized cost functions in the mode choice model. Similarly, small-scale effects such as local accessibility cannot be represented. Any time-dependent reaction, such as possible earlier departure in the morning because of a reduced service frequency in the early evening, seem to be difficult to represent.

In this situation, microscopic, behavior-based simulations may be applied to research the outcome of proposed measures (e.g. *(10)*). Yet, such models usually are limited to small scenarios for performance reasons. But as the environmental aspects gain higher attentions, the demand rises for behavior-based simulations that support large-scale scenarios as well as alternative transportation modes besides private cars. In this paper we describe how we extend a large-scale microscopic car-only simulation to also handle non-car modes, such as public transit. Especially, we compare the mode-choice reaction of the simulated agents to the mathematically to-be-expected reactions, to verify the correct functioning of our extended simulation. Afterwards, the mode choice model is applied to a large-scale application to test its feasibility in a real-world context.

The first section of the paper details the structure of our existing car-only simulation. The section following then explains how the simulation was extended to also handle non-car trips. After that, a section introduces the scenario that is used to test the behavior of the mode choice model, which is also mathematically verified in a section named “Theoretical Calculations”. We then give a short example how the mode choice model was applied to a large scale simulation and conclude the report by reflecting the achievements and giving a brief outlook, what could be done with the new features in the simulation.

SIMULATION STRUCTURE

Overview

Our simulation is constructed around the notion of agents that make independent decisions about their actions. Each traveler of the real system is modeled as an individual agent in our simulation. The overall approach consists of three important pieces:

- Each agent independently generates a so-called *plan*, which encodes its intentions during a certain time period, typically a day.
- All agents' plans are simultaneously executed in the simulation of the physical system. This is also called the *traffic flow simulation* or *mobility simulation*.
- There is a mechanism that allows agents to *learn*. In our implementation, the system iterates between plans generation and traffic flow simulation. The system remembers several plans per agent, and scores the performance of each plan. Agents normally choose the plan with the highest score, sometimes re-evaluate plans with bad scores, and sometimes obtain new plans by modifying copies of existing plans.

The simulation approach is the same as in many of our previous papers *(*e.g. *11).* The following exposition is a shortened and simplified description of key elements to limit the length of this paper. The results of this paper are based on a re-implementation of the MATSim framework in Java *(12)*.

A **plan** contains the itinerary of activities that the agent wants to perform during the day, plus the intervening trip legs the agent must take to travel between activities. An agent's plan details the order, type, location, duration and other time constraints of each activity, and the mode, route and expected departure and travel times of each leg. In the test scenario, “home" and “work" are the only activities, while the large-scale application makes use of more activity types. In both cases “car” and “non-car” are the only modes.

A plan can be modified by various **modules**. In the test scenario, only the *Time Adaptation* module (sometimes also referred to as Activity Times Generator) is used, while the large-scale application additionally uses a *Router* module. The Time Adaptation module is called to change the timing of an agent's plan. At this point, a very simple approach is used which just applies a random “mutation” to the duration attributes of the agent's activities. Although this approach is not very sophisticated, it is sufficient in order to obtain useful results. This is consistent with our overall assumption that, to a certain extent, simple modules can be used in conjunction with a large number of learning iterations *(*e.g. *13).* The router is a time-dependent best path algorithm *(14)*, normally using as link costs the link travel times from the previous iteration.

Mode choice will not be simulated by a module per se, but instead by giving every agent both a “car” and a “non-car” plan. Further details will be described later.

One of the plans is is marked as “selected”. The **traffic flow simulation** executes all agents' selected plans simultaneously on the network, and provides output describing what happened to each individual agent during the execution of its plan. The traffic flow simulation is implemented as a queue simulation, which means that each street (link) is represented as a FIFO (first-in first-out) queue with two restrictions *(15, 16)*. First, each agent has to remain for a certain time on the link, corresponding to the free speed travel time. Second, a link storage capacity is defined which limits the number of agents on the link. If it is filled up, no more agents can enter this link.

The outcome of the traffic flow simulation (e.g. congestion) depends on the planning decisions made by the decision-making modules. However, those modules can base their decisions on the output of the traffic flow simulation (e.g. knowledge of congestion) using **feedback** from the multi-agent simulation structure *(17, 18)*. This sets up an **iteration cycle** which runs the traffic flow simulation with specific plans for the agents, then uses the planning modules to update the plans; these changed plans are again fed into the traffic flow simulation, etc, until consistency between modules is reached. The feedback cycle is controlled by the **agent database**, which also keeps track of multiple plans generated by each agent.

10% of the agents generate new plans by taking an existing plan, making a copy of it, and then modifying the copy with the Time Adaptation or the Router module. The other agents reuse one of their existing plans. The agents decide if the just performed plan should be used again, or if a random plan out of the memory should be selected for the next iteration. The probability to change the selected plan is calculated by a logit model:

$$p_{change} = min(1, \alpha \cdot e^{\beta \cdot (s_{random} - s_{current})/2}) \tag{1}$$

with:

- α: The probability to change if both plans have the same score, set to 1%
- β: A sensitivity parameter, set to 2
- $s_{\{random,current\}}$: The score of the current/random plan (see next subsection)

In the long run this model is equivalent to the following probabilistic discrete choice model:

$$p_j = \frac{e^{\beta \cdot s_j}}{\sum_i e^{\beta \cdot s_i}} \tag{2}$$

where p_i is the probability for plan i to be selected and s_i its current score. The advantage of the model described by Eq. 1 is that only small numbers of agents change from one plan to another, which results in a smoother learning curve during the iteration cycle and lets the system better convert to a steady state.

The repetition of the iteration cycle coupled with the agent database enables the agents to learn how to improve their plans over many iterations. As the number of plans that one agent may have is limited by memory constraints, the plan with the worst performance is deleted when a new plan is added to a person which already has the maximum number of plans permitted. The iteration cycle continues until the system has reached a relaxed state. At this point, there is no quantitative measure of when the system is "relaxed"; we just allow the cycle to continue until the outcome is stable.

Scoring Plans

In order to compare plans, it is necessary to assign a quantitative score to the performance of each plan. In this work, in order to be consistent with economic appraisal, a simple utility-based approach is used. The elements of our approach are as follows:

- The total score of a plan is computed as the sum of individual contributions:

  $$U_{total} = \sum_{i=1}^{n} U_{perf,i} + \sum_{i=1}^{n} U_{late,i} + \sum_{i=1}^{n} U_{tr,i} \ , \tag{3}$$

  where U_{total} is the total utility for a given plan; n is the number of activities, which equals the number of trips; $U_{perf,i}$ is the (positive) utility earned for performing activity i; $U_{late,i}$ is the (negative) utility earned for arriving late to activity i; and $U_{tr,i}$ is the (negative) utility earned for traveling during trip i. In order to work in plausible real-world units, utilities are measured in Euro.

- A logarithmic form is used for the positive utility earned by performing an activity:

$$U_{perf,i}(t_{perf,i}) = \beta_{perf} \cdot t_{*,i} \cdot \ln\left(\frac{t_{perf,i}}{t_{0,i}}\right) \tag{4}$$

where t_{perf} is the actual performed duration of the activity, t_* is the "typical" duration of an activity, and β_{perf} is the marginal utility of an activity at its typical duration. β_{perf} is the same for all activities, since in equilibrium all activities at their typical duration need to have the same marginal utility.

$t_{0,i}$ is a scaling parameter that is related both to the minimum duration and to the importance of an activity. If the actual duration falls below $t_{0,i}$, then the utility contribution of the activity becomes negative, implying that the agent should rather completely drop that activity. This paper uses $t_{0,i} = t_{*,i} \cdot \exp(-\zeta/t_{*,i})$. where ζ is a scaling constant set to 10 hours. With this specific form, $U_{perf,i}(t_{*,i}) = \beta_{perf} \cdot \zeta$ is independent of the activity type.[1]

- The (dis)utility of being late is uniformly assumed as:

$$U_{late,i} = \beta_{late} \cdot t_{late,i} \ , \tag{5}$$

where β_{late} is the marginal utility (in Euro/h) for being late, and $t_{late,i}$ is the number of hours late to activity i. β_{late} is usually negative.

- The (dis)utility of traveling is uniformly assumed as:

$$U_{tr,i} = \beta_{tr} \cdot t_{tr,i} \ , \tag{6}$$

where β_{tr} is the marginal utility (in Euro/h) for travel, and $t_{tr,i}$ is the number of hours spent traveling during trip i. β_{tr} usually assumed to be negative.

In principle, arriving early or leaving early could also be punished. There is, however, no immediate need to punish early arrival, since waiting times are already indirectly punished by foregoing the reward that could be accumulated by doing an activity instead (opportunity cost). In consequence, the effective (dis)utility of waiting is already $-\beta_{perf}\, t_{*,i}/t_{perf,i} \approx -\beta_{perf}$. Similarly, that opportunity cost has to be added to the time spent traveling, arriving at an effective (dis)utility of traveling of $-|\beta_{tr}| - \beta_{perf}\, t_{*,i}/t_{perf,i} \approx -|\beta_{tr}| - \beta_{perf}$.

No opportunity cost needs to be added to late arrivals, because the late arrival time is spent somewhere else. In consequence, the effective (dis)utility of arriving late remains at β_{late}. – These approximate values (β_{perf}, $\beta_{perf} + |\beta_{tr}|$, and $|\beta_{late}|$) are the values that would correspond to the consensus values of the parameters of the Vickrey model *(20)*.

The utility function could be extended with additional terms. For example, one could think of tolls or parking costs that could be included into the score.

[1] This "consequence" is actually the motivation for the specific mathematical form of the activity performance utility contribution. The reason for this motivation is not relevant to this paper, but is described in *(19)*.

MODE CHOICE MODEL

The basic idea behind our mode choice model is that each agent always has at least one "car" plan and one "non-car" plan. Apart from that, plans are treated as described earlier. Since this always keeps both modes in the choice set, a decision between plans according to Eq. 1 is also a choice between modes.

This requires changes in many parts of the simulation framework, namely the transport simulation, the scoring of plans as well as the replanning. These changes are described in the following.

Generating non-car plans

To generate non-car plans, an initial demand with car plans must exist already. Starting with that initial demand, the leg modes of all legs in each plan are set to "car", and the fastest routes are calculated. Then, each plan is duplicated, changing all leg modes in the duplicated plans to "non-car".

The duration of every non-car trip is assumed to take twice as long as the car mode at free speed, but no exact route is provided. This is based on the (informally stated) goal of the Berlin public transit company to generally achieve door-to-door travel times that are no longer than twice as long as car travel times. This, in turn, is based on the observation that non-captive travellers can be recruited into public transit when it is faster than this benchmark *(21)*. For the purposes of the present paper, it is assumed that all non-car modes very roughly have the shared characteristics that they are slower than the (non-congested) car mode—this will be further disaggregated in future work. In the same vein, both for car and for non-car trips there are no separate considerations of access and egress.

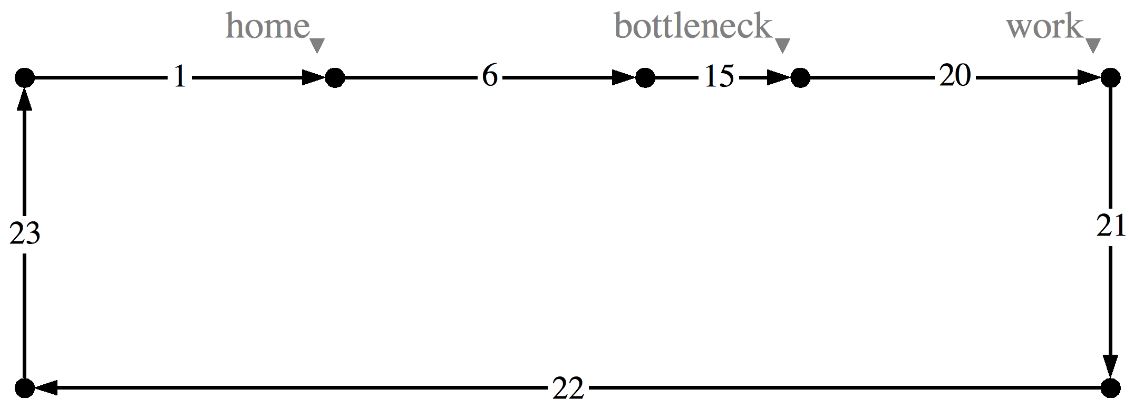
Handling non-car plans in the transport simulation

Currently, the simulation only supports a road-network, but no walk- or rail-network. Thus, only car legs can be truly simulated. Agents on non-car legs are teleportated from one location to the next. But the teleportation is not instantaneously, but takes some amount of time, which can be stored in the legs as planned travel duration. While this does not impose any transit vehicles' capacity constraints, it would still allows us to have individual travel times, depending on agents' demographics or chosen non-car mode (e.g. bike, walk, transit, …). The simulation still generates departure and arrival events for non-car legs, which can be used for analyses.

Scoring non-car plans

The scoring of non-car plans is very similar to the scoring of car plans as described in Sec. "Simulation Structure: Scoring Plans", only the disutility of traveling changes. This is expressed by using $\beta_{tr,nc}$ for the marginal utility of traveling, instead of $\beta_{tr,car}$. It is important to note once more that $\beta_{tr,car}$ and $\beta_{tr,nc}$ are *not* values of time by themselves, but they are *additional* disutilities caused by traveling, in addition to the opportunity cost of time. This is consistent with econometric approaches *(22)*.

Additional terms could be added, e.g. for changing transit lines or long waiting periods at stations, but as the simulation does not yet provide this kind of information, there is no need to add such parameters to the scoring function.



Links	Length (m)	Freespeed Velocity (m/s)	Capacity (veh/h)
1, 20, 21, 23	10000	27.78	36000
22	35000	27.78	36000
6	10000	27.78	3600
15	5000	27.78	1000

FIGURE 1 The links of the test network with their corresponding ids and attributes. Link 15 has reduced capacity. Traffic runs clockwise, i.e. agents have their home location at link 1 and work on link 20, so link 15 provides a bottleneck.

Replanning with non-car plans

During replanning, plans are duplicated and modified (see "iteration cycle" in Sec. "Simulation Structure, Overview"). This also holds for non-car plans. The only difference is that the plans deletion module makes sure that at least one plan of every mode is kept for every agent. This is to make sure that all agents keep their ability to change mode until the end of the iterations.

The above steps integrate mode choice into the replanning process that takes place iteratively with the simulation. Instead of precalculating the mode choice before the traffic assignment, as it is done in the traditional four-step process, mode choice is now treated at the same level as route choice in the traffic assignment.

TEST SCENARIO

Network

To test the mode choice model, a simple test network was used (see Fig. 1), consisting only of a cycle of one-way links.[2] The capacity of all links except links 6 and 15 are (unrealistically) high as to minimize the influence these links have on the traffic, essentially making it possible for most agents to drive with free speed. Links 6 and 15 have reduced capacity, building a bottleneck. Traffic flow starts at link 1, continuing clockwise.

Initial plans

The synthetic population consists of 2000 agents. All agents have their home activity at link 1, which they initially leave at 06:00. They drive to work (located on link 20) with a car via links

[2]It is a simplified version of another test network we use internally, which explains the numbering of the links.

Parameter	Value	Description
β_{perf}	6 Euro/h	utility of performing an activity at its typical duration
β_{late}	-18 Euro/h	disutility of coming late
$\beta_{tr,car}$	-6 Euro/h	disutility of traveling with a car
$\beta_{tr,nc}$	varied (see below)	disutility of traveling with non-car mode
$t_{*,w}$	8 hours	typical duration of work
work start time	exactly at 7:00am	
$t_{*,h}$	12 hours	typical duration of home-activity
β (existing plans)	2	constant used in binary logit model

TABLE 1 Behavioral parameters used in the test scenario.

6 and 15, where they stay for 8 hours, after which they drive back home to link 1 via links 21, 22, and 23. The free speed travel time from link 1 to link 20 is 15 minutes. The free speed travel time from link 20 to link 1 is 39 minutes. Thus the total free speed travel time driving by car is 54 minutes or 0.9 hours.

As the agents are forced to remain on that route, the scenario converts into the well-known Vickrey bottleneck scenario *(23, 20)*; also see below for more details.

In addition, each agent possesses an initially non-active plan that uses the non-car mode for both trips. These trips take twice as long as by car in an empty network, i.e. 30 minutes from link 1 to link 20, and 78 minutes from link 20 to link 1. The total non-car travel time is 108 minutes or 1.8 hours. In contrast to the car travel times, these non-car travel times are not affected by congestion. The first trip starts at 06:30, so the agents will arrive exactly at 07:00 at their work place.

Behavioral parameters

The values used for the parameters are shown in Table 1. The values can be interpreted as follows:

- "Typical" durations of 8 and 12 hours for work and home mean that work and home times have a tendency to arrange themselves with a ratio of 8:12 (i.e. 2:3): Assume a fixed travel time budget. In this situation, for optimality of the scoring function the marginal utilities of duration, $\partial U_{perf,i} / \partial t_{perf,i} = \beta_{perf}\, t_{*,i} / t_{perf,i}$, need to be equal for all activity types, resulting in

  $$\frac{t_h}{t_{*,h}} = \frac{t_w}{t_{*,w}} \tag{7}$$

  The result is only approximately correct when the overall travel time varies.
  The activity of the home activity is "wrapped around", i.e. a departure at 6am and a return at 5pm results in a home activity duration of 13 hours.
- A work start exactly at 7:00am means that (a) no utility can be accumulated from an arrival earlier than 7:00am, and (b) any late arrival is immediately punished with $\beta_{late} = -18$ Euro/h. Because of the argument made earlier regarding the opportunity cost of foregone activity time in situation (a), the *effective* marginal disutility of early arrival is $-\beta_{perf}\, t_{*,i} / t_{perf,i} \approx -\beta_{perf} = -6$ Euro/h. Since the effective marginal disutility of car trav-

eling is, by the same argument, $-\beta_{perf}\, t_{*,i}/t_{perf,i} \approx -\beta_{perf} - |\beta_{tr,car}| \approx -12$ Euro/h, the *effective* values of time of our study are approximately the same as the consensus values of $(-6, -12, -18)$ of the Vickrey scenario *(23, 20)*. The return trip has no influence since there is no congestion.

Simulation Results

The simulation in the test setup was run with different values for $\beta_{tr,nc}$, resulting in different mode shares. Each simulation was first run for 1000 iterations. In each iteration, 10% of the agents were modified by the time allocation module, while all other agents chose an existing plan. After that, the simulation was continued for 100 more iterations, but without time adaptation. This allowed agents to select their best plan, no longer being forced to execute (possibly bad) plans after replanning.

$\beta_{tr,nc}$ was varied from $+2$ to -10 in increments of -1. Fig. 2 shows the resulting car mode shares. It can clearly be seen that an increase of the disutility of traveling in the non-car mode leads to an increasing number of agents choosing car as transportation mode. In the following section, these results are validated by comparing them to the theoretical values one should expect based on the aforementioned mode choice model.

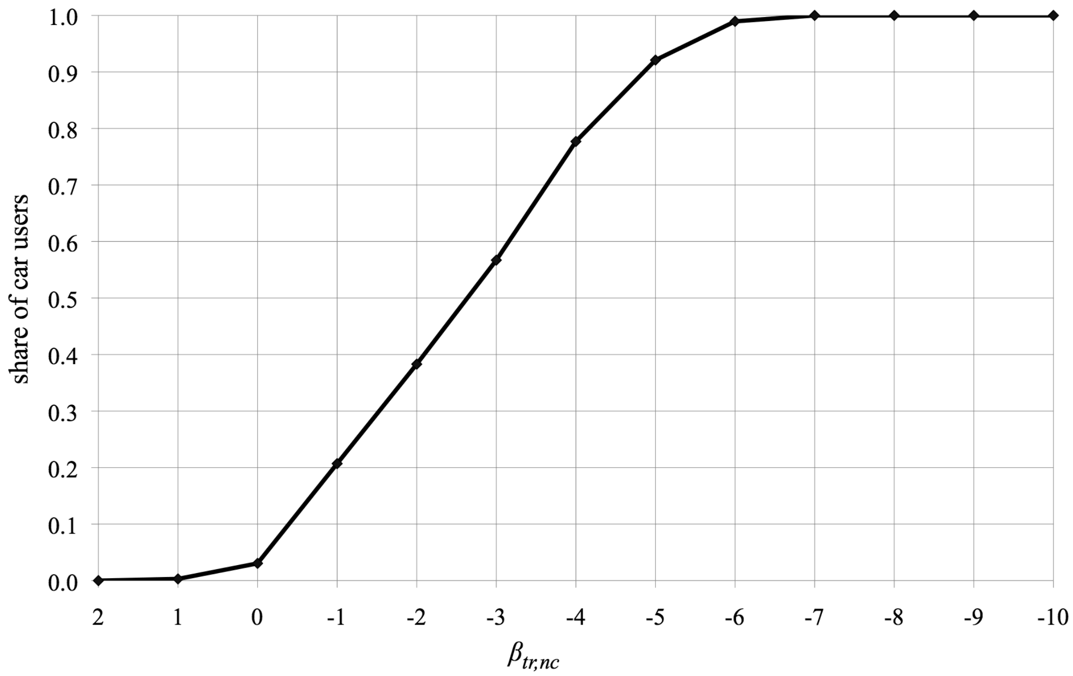


FIGURE 2 Share of car users in the simulation for different non-car travel disutilities ($\beta_{tr,nc}$).

THEORETICAL CALCULATIONS

Because of the simulation set-up, the mode share of the car mode, f_{car}, follows a binary logit model:[3]

$$f_{car} = \frac{\exp(\beta \cdot U_{car}(f_{car}))}{\exp(\beta \cdot U_{car}(f_{car})) + \exp(\beta \cdot U_{nc})} \tag{8}$$

U_{car} and U_{nc} are the total utilities of agents traveling either with a car or using the non-car transport mode. It is really important to note that these are utilities for the full daily plan, and not partial utilities for the mode choice contribution only. These utilities are defined according to Eq. 3, with only the two activities "home" and "work":

$$U_{mode} = \beta_{perf} \cdot t_{*,h} \cdot \ln\left(\frac{t_{h,mode}}{t_{0,h}}\right) + \beta_{perf} \cdot t_{*,w} \cdot \ln\left(\frac{t_{w,mode}}{t_{0,w}}\right) + \beta_{tr,mode} \cdot t_{tr,mode} + \beta_{late} \cdot t_{late} \,. \tag{9}$$

As mentioned before, travel times depend on the transport mode.

The non-car mode

Taking the "activity duration ratio" Eq. 7 together with the time budget equation

$$t_{h,nc} + t_{w,nc} + t_{tr,nc} = 24\ h \ ,$$

one obtains for people using the non-car mode:

$$\begin{aligned}
\frac{t_{*,h} + t_{*,w}}{t_{*,h}} \cdot t_{h,nc} &= 24\ h - t_{tr,nc} \\
t_{h,nc} &= (24\ h - t_{tr,nc}) \cdot \frac{t_{*,h}}{t_{*,h} + t_{*,w}} && (10) \\
t_{w,nc} &= (24\ h - t_{tr,nc}) \cdot \frac{t_{*,w}}{t_{*,w} + t_{*,h}} && (11)
\end{aligned}$$

At this point, all variables for Eq. 9 for the non-car mode, assuming on-time arrival, are expressed in the parameters of the simulation.

The car mode

For car users, the calculation is more complex. Following *(23, 20)* we will assume that at the end of the day every agent will have experienced the same total utility: While some may spend more time traveling (by being stuck in a traffic jam) but arrive at the right time at the work place, other agents may decide to leave early, traveling the whole route with free speed but also arrive at work early, foregoing any utility by performing an activity because the work place is still closed. Other agents again may stay longer at home, traveling after the jam has disappeared, arriving late at work and receiving the schedule delay penalty for that. One can obtain results by just looking at the first and the last agent to arrive at work. When equating Eq. 9 for these two, the travel time drops out because it is the same for both, and one arrives at

[3] This statement is, in fact, only correct when the number of car plans is equal to the number of non-car plans for every agent. See the end of the section for a comment on this.

$$\beta_{perf} \cdot t_{*,h} \cdot \ln\left(\frac{t_{h,car} - \tau_h}{t_{0,h}}\right) + \beta_{perf} \cdot t_{*,w} \cdot \ln\left(\frac{t_{w,car} - \tau_w}{t_{0,w}}\right)$$

$$= \beta_{perf} \cdot t_{*,h} \cdot \ln\left(\frac{t_{h,mode}}{t_{0,h}}\right) + \beta_{perf} \cdot t_{*,w} \cdot \ln\left(\frac{t_{w,car}}{t_{0,w}}\right) + \beta_{late} \cdot t_{late} \ ,$$

where the LHS refers to the person who arrives early, and who suffers τ_h, τ_w reductions of his/her activity durations. After linearization and dropping terms that cancel out, this becomes

$$-\tau_h \cdot \beta_{perf} \cdot t_{*,h} \cdot \frac{1}{t_{h,car}} - \tau_w \cdot \beta_{perf} \cdot t_{*,w} \cdot \frac{1}{t_{w,car}} \approx \beta_{late} \cdot t_{late} \ ,$$

From the optimal time allocation, Eq. 7, one infers that also for the time deductions τ_h, τ_w one needs $\tau_h/\tau_w = t_{*,h}/t_{*,w}$ and therefore $\tau_h = t_{early} \cdot t_{*,h}/(t_{*,h} + t_{*,w})$ and $\tau_w = t_{early} \cdot t_{*,w}/(t_{*,h} + t_{*,w})$. Taking this and once more Eq. 7 directly, one obtains, after some algebra

$$t_{early} \, \beta_{perf} \, \frac{t_{*,h}}{t_{h,car}} \approx |\beta_{late}| \, t_{late} \ , \tag{12}$$

where it was also invested that β_{late} is assumed to be negative. In addition, one has the equation for the bottleneck,

$$t_{early} + t_{late} = \frac{|A| \cdot f_{car}}{C_b} \tag{13}$$

where $|A|$ is the total number of agents, C_b is the flow-capacity of the bottleneck, and f_{car} the share of car users. The equation states that the capacity of the bottleneck is exactly enough to serve all agents between the first and the last. Inserting Eq. 12, one obtains

$$t_{early} \approx \frac{|\beta_{late}| \, t_{h,car}}{|\beta_{late}| \, t_{h,car} + \beta_{perf} \, t_{*,h}} \cdot \frac{|A| \cdot f_{car}}{C_b} \tag{14}$$

The optimal activity durations for the "early" agent are, similar to Eq. 10 and 11:

$$\begin{aligned}
t_{h,car} + t_{w,car} &+ t_{tr,fs} + t_{early} = 24\ h \\
\frac{t_{*,h} + t_{*,w}}{t_{*,h}} \cdot t_{h,car} &= 24\ h - t_{tr,fs} - t_{early} \\
t_{h,car} &= (24\ h - t_{tr,fs} - t_{early}) \cdot \frac{t_{*,h}}{t_{*,h} + t_{*,w}} \qquad (15) \\
t_{w,car} &= (24\ h - t_{tr,fs} - t_{early}) \cdot \frac{t_{*,w}}{t_{*,w} + t_{*,h}} \ , \qquad (16)
\end{aligned}$$

where $t_{tr,fs}$ is the free speed travel time by car. Substituting $t_{h,car}$ from Eq. 15 into Eq. 14 leads to an equation that only contains t_{early} and f_{car} as unknowns. One can see that the resulting equation contains the square of t_{early}. Solving that resulting equation provides two solutions for t_{early}, of which only one is useful, as the other one leads to negative times for either t_{early} or t_{late} in Eq. 13.

Thus at this point one knows t_{early} and in consequence $t_{h,car}$ and $t_{w,car}$ as functions of f_{car}. The expressions can be written down, but are rather long and not easy to interpret.

The complete mode choice

Recall that we are interested in an expression that relates the mode share, f_{car}, the the additional disutility of the non-car mode, $\beta_{tr,nc}$. What we have at this point is:

- We can compute the utility of the optimal non-car plan as a function of $\beta_{tr,nc}$.
- We can compute the utility of the optimal car plan as a function of f_{car}.

What remains is to insert these expressions into Eq. 8, which can also be written as

$$U_{car} = \frac{1}{\beta} \cdot \ln\left(\frac{f_{car}}{1-f_{car}}\right) + U_{nc} \tag{17}$$

Substituting U_{car} and U_{nc} with Eq. 9, one gets:

$$\beta_{perf} \cdot t_{*,h} \cdot \ln\left(\frac{t_{h,car}}{t_{0,h}}\right) + \beta_{perf} \cdot t_{*,w} \cdot \ln\left(\frac{t_{w,car}}{t_{0,w}}\right) + \beta_{tr,car} \cdot t_{tr,fs} = \tag{18}$$

$$\frac{1}{\beta} \cdot \ln\left(\frac{f_{car}}{1-f_{car}}\right) + \beta_{perf} \cdot t_{*,h} \cdot \ln\left(\frac{t_{h,nc}}{t_{0,h}}\right) + \beta_{perf} \cdot t_{*,w} \cdot \ln\left(\frac{t_{w,nc}}{t_{0,w}}\right) + \beta_{tr,nc} \cdot t_{tr,nc} \tag{19}$$

Recall that for the car mode we are considering the "first" (= most early) agent; the term regarding late arrival is thus dropped.

More variables can be substituted in Eq. 18 by their corresponding calculations in the previous equations. While it could still be solved analytically, it once more gets quite complex and not easily readable.

Evidently, $\beta_{tr,nc}$ can be isolated in Eq. 18, but not so f_{car} if one remembers that f_{car} is also part of t_{early} which is used to substitute $t_{h,car}$ and $t_{w,car}$ (Eq. 14–16).

Extracting $\beta_{tr,nc}$ and plotting it as a function of f_{car} ranging from 0 to 1, one gets the graph shown in Fig. 3. Comparing Fig. 3 with Fig. 2, one can see the the two curves are very similar (Fig. 3 may be rotated by 90 degrees counter clock wise for better comparison). Only small variations can be seen, likely due to the discrete size of agents in the simulation as well as the not completely predictable behavior of random numbers used in the simulation. Additionally, the calculations assume that every agent has an optimal plan, which cannot be guaranteed in the simulation.

The fact that in spite of the noise the mode choice curve is "steeper" in the simulations than in the analytical calculations is due to the learning algorithm: If for an agent one mode is clearly better than the other mode, than that mode will have more plans than the other mode. This gives an additional statistical advantage to the better mode, making the curve more steep.

Overall, one finds that the mode choice model is in excellent agreement with the theoretical calculations. This, on the one hand, verifies the implementation of the model. On the other hand, it means that, to an extent, it is possible to understand analytically what the simulation does, which will help to uncover and understand the economic and behavioral principles embedded in the implementation.

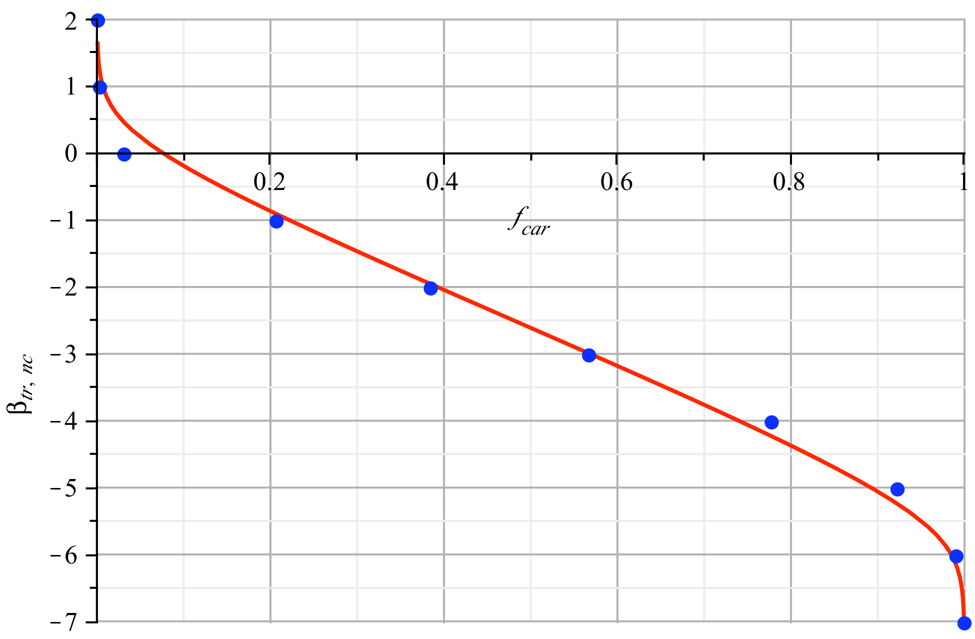


FIGURE 3 Non-car travel disutilities ($\beta_{tr,nc}$) for different car shares (f_{car}). The line refers to the analytical result, the dots to the simulation results.

LARGE-SCALE APPLICATION

The mode choice model was also applied to a large-scale, real-world scenario. We used the area of Zurich, Switzerland, for this application, which has about 1 million inhabitants. The following paragraphs give a simplified description of the scenario to limit the length of this paper. A full description of the scenario can be found in *(24)*.

The network used is a Swiss regional planning network that includes the major European transit corridors. It consists of 24 180 nodes and 60 492 links.

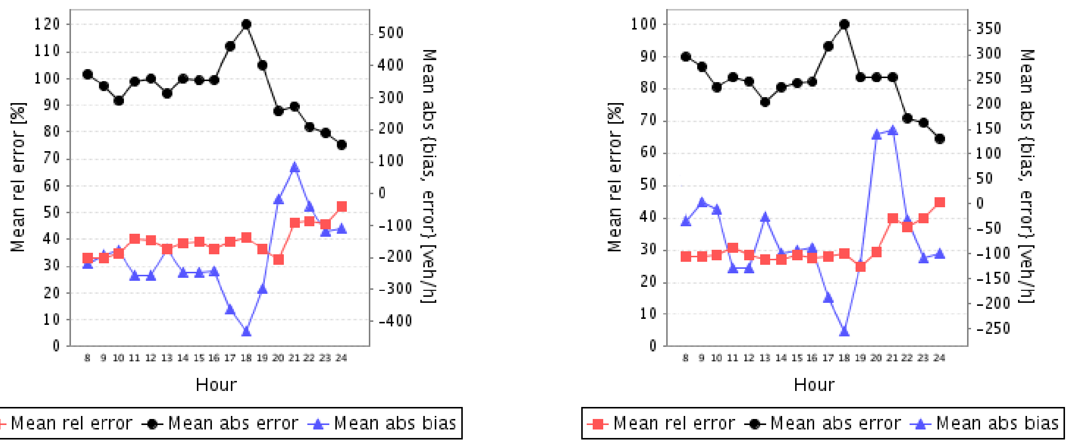
The simulated demand consists of all travelers within Switzerland that are inside an imaginary boundary around Zurich at least once during their day. This boundary is defined by a circle with a radius of 30 kilometers ($\approx$ 18.6 miles) with its center at “Bellevue”, a central place in Zurich. Additionally, through traffic was added to the traffic demand, coming from abroad but passing the Zurich area *(25)*. All agents have complete day plans with activities like *home*, *work*, *education*, *shopping*, *leisure*, based on microcensus information *(26, 27)*. The time window during which activities could be performed was limited to certain hours of the day. Table 2 shows the time restrictions for the different activity types. Unlike the sample scenario described in the sections above, there is no punishment for being late. This was not possible because agents could split their work activity into two or more parts, e.g. one in the morning and one in the afternoon. In such a case it would be complicated to specify when an agent starts an activity late or not.

To speed up computations, a random 10% sample was chosen from the synthetic population for simulation, consisting of 181 725 agents. In this large-scale application, the agents could not only perform time adaptation as described in a previous sections, but could also do route adaptation, which is essential for the car mode. For comparison, the same scenario was run with the pre-

Activity type	Opening time	Closing time
Home	00:00	24:00
Work	06:00	20:00
Education	06:00	20:00
Shop	08:00	20:00
Leisure	00:00	24:00

TABLE 2 Activity opening and closing times used in the large-scale scenario.

calculated mode choice (see *(24)*).



(a) pre-calculated mode choice

(b) mode choice during simulation ($\beta_{tr,nc} = -3$)

FIGURE 4 Comparison of simulated traffic volumes with real-world counts. Note different scales on y-axis

Simulated traffic volumes were compared with the hourly traffic volumes from 159 real-world counting stations. Fig. 4 shows, in red, the mean relative error between hourly flows in reality and hourly flows from the simulation. The left figure contains the result from the fixed, pre-determined mode choice, the right figure the result of the new adaptive mode choice which was explained in this paper. One notices a quite distinct reduction in the average error, from about 40% to about 30%. Also the absolute bias, in blue, is reduced.

Although a mean relative error of around 30% during the day hours may seem high, one must recognize that daily fluctuations in real traffic counts can also be quite high. Also, while many models claim lower relative errors, they do so over a period of 24 hours usually, where over- and underestimations in single hours cancel each other out.

For the large-scale tests, the disutility for the car mode, $\beta_{tr,car}$, was set to -6 Euro per hour, while the disutility for the non-car mode, $\beta_{tr,nc}$, was varied between 0 and -6. An interpretation of this might be that measures are discussed that change the attractiveness of the non-car modes, leaving everything else, including the travel times, constant. An obvious concrete example would be fare changes. And the importance of the results as this point is not so much the magnitude of the response itself, but the fact that the model displays the interaction between activity timing and mode choice. Fig. 5 shows the number of agents en route with cars over the time of day. It can be clearly seen that the number of car users decreases the lower the travel disutility for the non-car mode gets. The peaks at 6am and 8pm are due to our opening time restrictions (see Table 2).

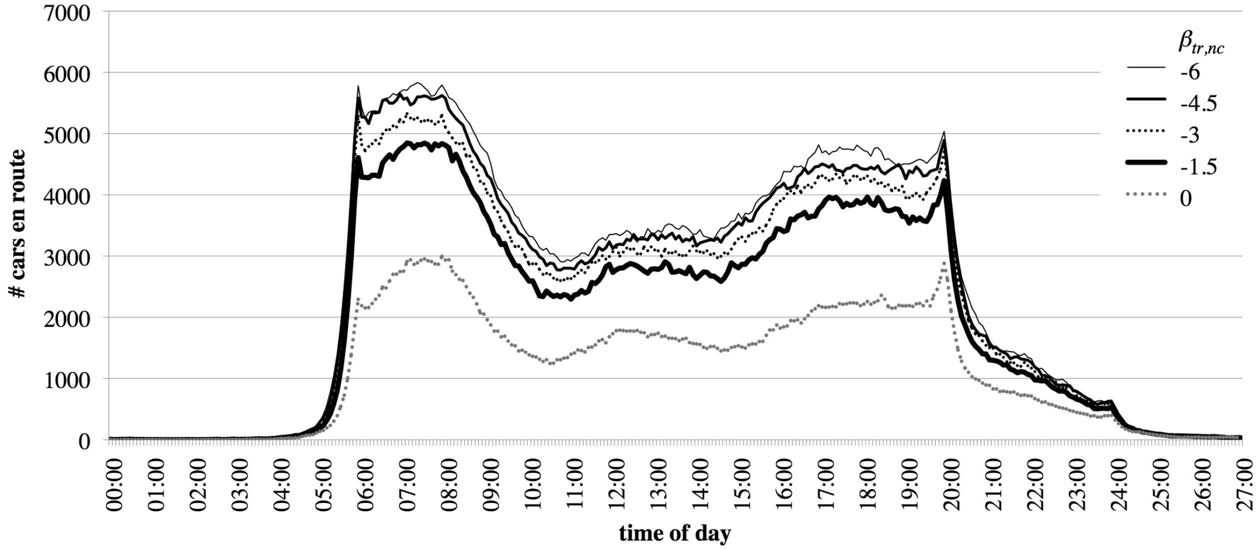


FIGURE 5 Car en route in large-scale scenario over time of day with different disutilities for traveling with non-car modes.

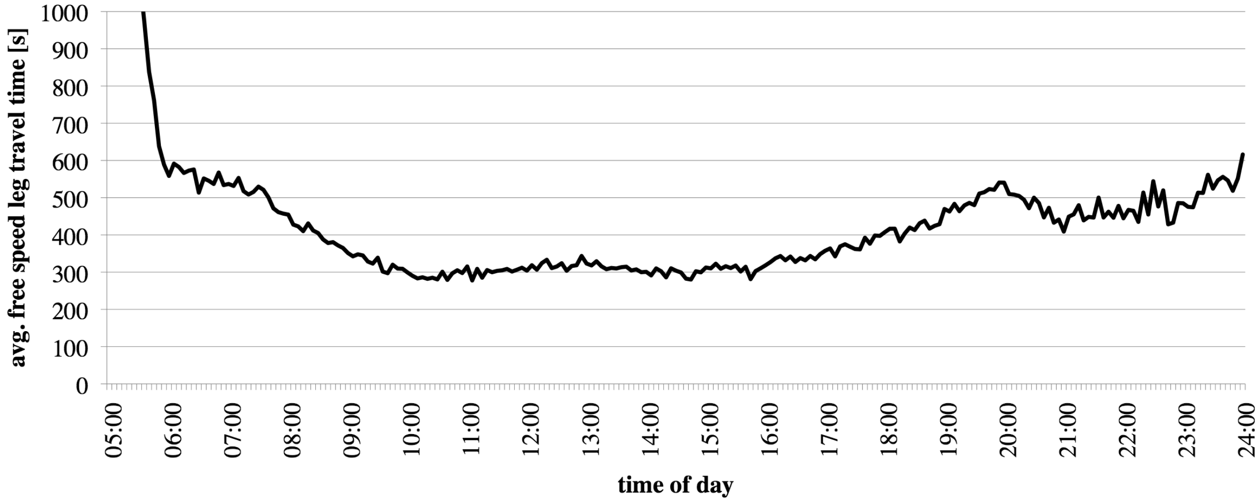
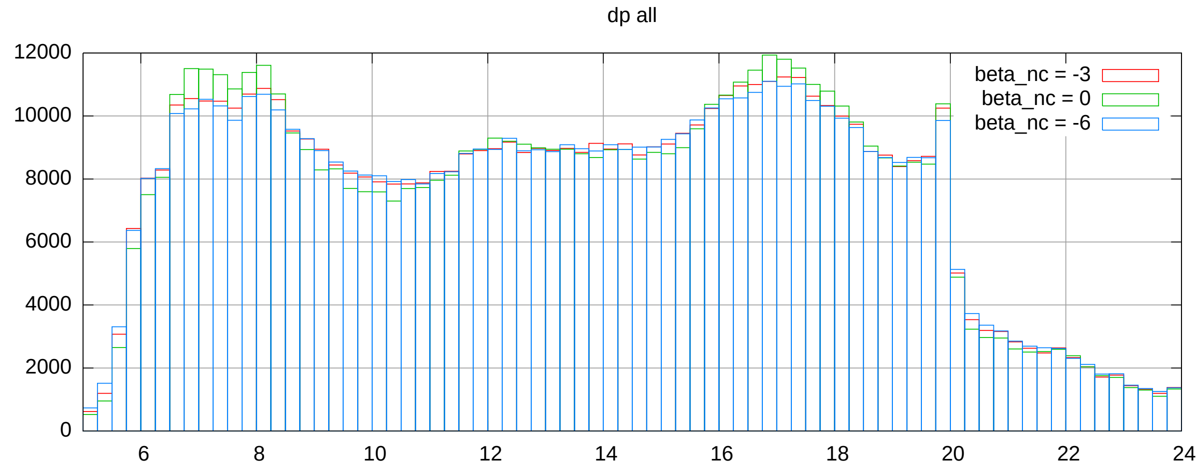
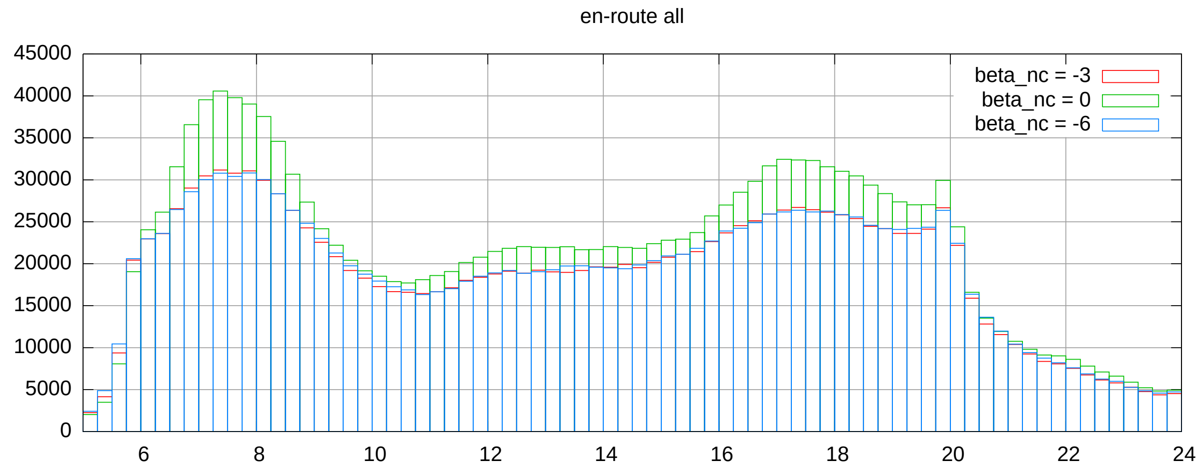


FIGURE 6 Average free speed travel time per leg starting at different times of day.

Figure 7a shows *all* departures as a function of the time-of-day, for different values of $\beta_{tr,nc}$. Since demand itself is inelastic, the area under all the curves is the same. One notices, however, a shift towards the peak periods when the disutility of the non-car mode is reduced. Together with



(a) number of agents departing per 15 min over time of day



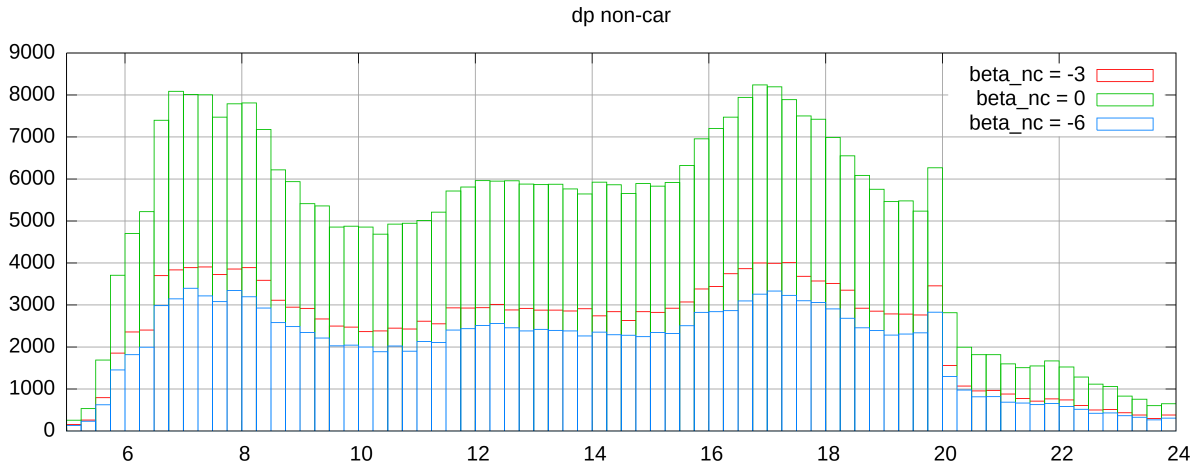
(b) number of agents en route per 15 min over time of day

FIGURE 7 Number of agents departing and en route over time of day.

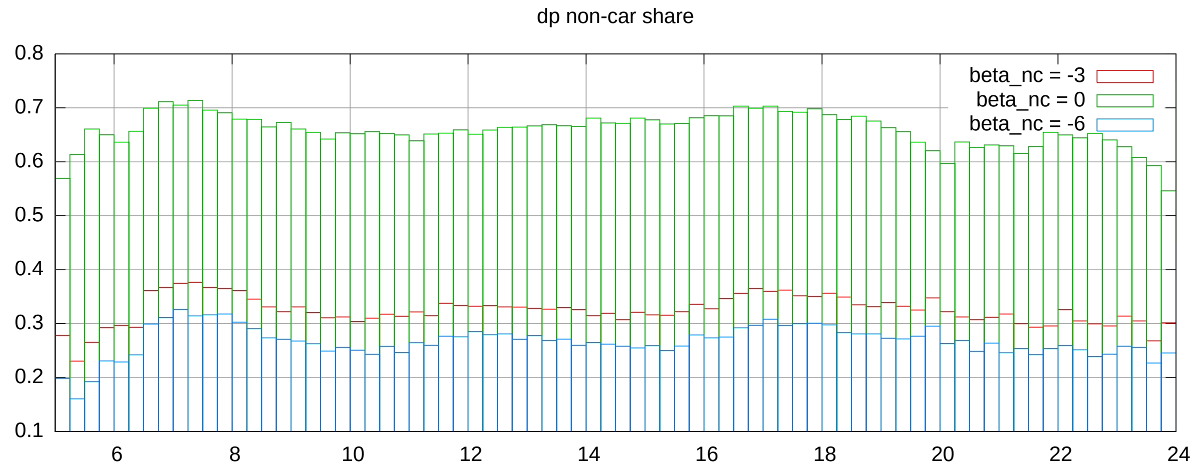
further results, discussed below, the reason is that, because of people moving away from the car, the peak period is less congested, meaning that other car-users re-adjust their schedules towards the peak hour.

Figure 7b shows 15-min averages of *all* agents that are en route. First, one notices that the variance along the time axis is much stronger than for the departures. This is due to the fact that the trips during the peak periods are longer (Fig. 6). Thus, the increased number of departures goes along with an increase of the time that travellers remain in the system. Second, one notices that the overall number of people en-route goes up with the reduced non-car disutility. Clearly, this is a consequence of the longer travel times with the non-car modes.

Non-car departures (Fig. 8a) show the expected behavior: More non-car departures at all times as a function of a reduced disutility. Somewhat surprising may be the *share* of the non-car departures: It essentially remains constant from 6am to 6pm, and it does that for all values of $\beta_{tr,nc}$



(a) number of agents departing with non-car mode per 15 min over time of day



(b) share of agents departing with non-car mode per 15 min over time of day

FIGURE 8 Absolute and relative number of agents departing with non-car mode over time of day.

(albeit at different levels). That is, one can*not* observe a maybe expected tendency to have increased non-car departures at the peak periods, in order to avoid congestion. This is probably due to the increased trip length that goes along with the peak periods; increased trip length makes the non-car mode *less* attractive because the longer travel times eat more into the daily time budget. That is, the increased congestion (moving travellers to non-car modes) and the increased trip length (moving travellers away from non-car modes) seem to approximately cancel each other out. This could be clarified with an additional simulation where the car network obtains very high capacities, thus setting congestion to zero.

Outside the time from 6am to 6pm, the non-car share is markedly lower than during the day. This is due to the fact that during those times there is little car congestion, thus making the car more attractive.

FURTHER STEPS

At the moment, only one transport mode can be used for the complete plan. That is, all trips of a given day need to be done by the same mode. While the data structures, file formats and simulation could deal with different transport mode per leg, there are some conceptual points that we want to solve first before applying mode choice to a subtour level.

The preparation of the demand could be simplified by not giving two distinct plans at initialization, but to implement mode choice similar to time or route adaptation: agents could make a copy of a plan and just replace the transport mode with whatever mode is available and they want to try out.

The simulation setup would allow to have different $\beta_{tr,nc}$ over the time of day, as every trip has a departure time. This could be used to model a changing attractivity to use the non-car modes during a day. One example might be to improve the quality of service in transit in the late evening or night hours, resulting in a lower absolute disutility during that time of day.

An improved router for non-car modes would improve results. Possibilities are the usage of transit schedules instead of the "double free speed travel time" assumption currently used. Currently, that assumption makes the non-car mode highly unattractive for long distance trips. This will likely change by using more realistic travel times, especially for long distance trips that are served well by fast trains.

The simulation should not only teleport agents with non-car mode, but actually simulate them as well. Different aspects of this would be important to include, say, public transport vehicle overcrowing effects, or the effect of public transport being caught in car congestion. It would require to add transit vehicles, bikes and other means of transport, together with their characteristics, schedules and so on.

A car ownership model, or arguably a life style model, could be added in the demand modelling. This would reduce the choice between car and non-car mode to travellers that actually have access to a car. A preliminary attempt to do this for the Zurich scenario did *not* lead to improved results with respect to the real world traffic counts. This was presumably due to the fact that the car ownership model was based on zonal characteristics, while the mode choice model of our simulation at least on the car side picks up very detailed accessibility issues. It becomes quite clear that the behavioral basis of all relevant decision models needs to be consistent.

CONCLUSION

It was shown how to include a non-car mode into a multi agent transport simulation with relatively few conceptual changes. The non-car mode was integrated by giving every agent *two* initial plans, one using the car for all trips, and one using the non-car mode for all trips. The non-car trips are assumed to use up twice as much travel time as the uncongested car mode. Travellers can then, in the simulation, adjust times and car routes; the performance of the resulting plans is scored after execution in the traffic flow simulation, based on a utility function that includes positive utility for performing an activity, different negative utilities for travelling by different modes, and opening times outside which no utility for performing an activity can be accumulated.

The model was first tested in a simplified scenario based on the famous Vickrey bottleneck example. The non-car mode was used as an alterative to the congested car mode. It was shown that the analytical calculation and the simulation model produce the same results when looking at the mode split as a function of the non-car mode disutility.

The model was then applied to a realistic real world example for the Zurich metropolitan area.

The reaction of users to changes in the non-car disutility was analyzed in some detail, including temporal reactions. Adding the mode choice to the large-scale scenario improved the realism of the scenario when comparing the simulated traffic volumes to data from counting stations.

ACKNOWLEDGMENTS
We would like to thank Michael Balmer from IVT, ETH Zurich, for his important contributions that led to the network and demand used for the large scale application in the Zurich area. The simulations were run on our compute cluster, set up and maintained by the Faculty for Mathematic at the Berlin Institute of Technology (TU Berlin).

REFERENCES

1. E.T. Verhoef. Marginal cost based pricing in transport. Key implementation issues from the economic perspective. See http://www.imprint-eu.org/public/EV_V2.PDF, 2001. Paper presented at Seminar One of IMPRINT-EUROPE.

2. P. Bonsall, J. Shires, D. Ngoduy, H. Link, A. Becker, P. Papaioannou, and P. Xanthopoulos. Optimal complexity of prices for transport infrastructure. Deliverable 6 of GRACE (generalisation of research on accounts and cost estimation), funded by sixth framework programme, ITS, University of Leeds, Leeds, 2007.

3. J. de D. Ortúzar and L.G. Willumsen. *Modelling transport*. Wiley, Chichester, 1995.

4. M.G. Richards and M. Ben-Akiva. A simultaneous destination and mode choice model for shopping trips. *Transportation*, pages 343–356, 1974.

5. J. Jonnalagadda, N. Freedman, W.A. Davidson, and J.D. Hunt. Development of microsimulation activity-based model for San Francisco: destination and mode choice models. *Transportation Research Record*, 1777:25–35, 2001.

6. D. Boyce and H. Bar-Gera. Validation of multiclass urban travel forecasting models combining origin-destination, mode, and route choices. *Journal of Regional Science*, 43:517–540, 2003.

7. D. Lohse, G. Bachner, B. Dugge, and H. Teicher. Ermittlung von Verkehrsströmen mit n-linearen Gleichungssystemen unter Beachtung von Nebenbedingungen einschließlich Parameterschätzung (Verkehrsnachfragemodellierung: Erzeugung, Verteilung, Aufteilung). Technical Report 5/1997, Institut für Verkehrsplanung und Straßenverkehr, Dresden, Germany, 1997.

8. O.A. Nielsen and R.D. Frederiksen. Optimisation of timetable-based, stochastic transit assignment models based on MSA. *Annals of Operations Research*, 144:263–285, 2006.

9. K. Nökel. VISUM 10 – Status and perspectives. Technical report, PTV AG, Karlsruhe, Germany, 2007. See www.ptv.de.

10. L. Zhang and D. Levinson. Agent-based approach to travel demand modelling: Exploratory analysis. *Transportat Research Record*, 1989:28–36, 2004.

11. B. Raney and K. Nagel. An improved framework for large-scale multi-agent simulations of travel behaviour. In P. Rietveld, B. Jourquin, and K. Westin, editors, *Towards better performing European Transportation Systems*, page 42. Routledge, London, 2006.

12. MATSIM www page. MultiAgent Transport SIMulation. http://matsim.org/, accessed 2008.

13. K. Nagel, M. Strauss, and M. Shubik. The importance of timescales: Simple models for economic markets. *Physica A*, 340(4):668–677, 2004.

14. N. Lefebvre and M. Balmer. Fast shortest path computation in time-dependent traffic networks. In *Proceedings of Swiss Transport Research Conference (STRC)*, Monte Verita, CH, September 2007. See www.strc.ch.

15. C. Gawron. *Simulation-based traffic assignment*. PhD thesis, University of Cologne, Cologne, Germany, 1998.

16. N. Cetin, A. Burri, and K. Nagel. A large-scale agent-based traffic microsimulation based on queue model. In *Proceedings of Swiss Transport Research Conference (STRC)*, Monte Verita, CH, 2003. See www.strc.ch. Earlier version, with inferior performance values: Transportation Research Board Annual Meeting 2003 paper number 03-4272.

17. D.E. Kaufman, K.E. Wunderlich, and R.L. Smith. An iterative routing/assignment method for anticipatory real-time route guidance. Technical Report IVHS Technical Report 91-02, University of Michigan Department of Industrial and Operations Engineering, Ann Arbor MI 48109, May 1991.

18. J.A. Bottom. *Consistent anticipatory route guidance*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2000.

19. D. Charypar and K. Nagel. Generating complete all-day activity plans with genetic algorithms. *Transportation*, 32(4):369–397, 2005.

20. R. Arnott, A. de Palma, and R. Lindsey. Economics of a bottleneck. *Journal of Urban Economics*, 27(1):111–130, 1990.

21. T. Reinhold. Konzept zur integrierten Optimierung des Berliner Nahverkehrs. In *Öffentlicher Personennahverkehr*. Springer Berlin Heidelberg, 2006.

22. S.R. Jara-Díaz and R. Guerra. Modeling activity duration and travel choice from a common microeconomic framework. In *Proceedings of the meeting of the International Association for Travel Behavior Research (IATBR)*, Lucerne, Switzerland, 2003. See www.ivt.baug.ethz.ch.

23. W. S. Vickrey. Congestion theory and transport investment. *The American Economic Review*, 59(2):251–260, 1969.

24. Y. Chen, M. Rieser, D. Grether, and K. Nagel. Improving a large-scale agent-based simulation scenario. VSP Working Paper 08-15, TU Berlin, Transport Systems Planning and Transport Telematics, 2008. See www.vsp.tu-berlin.de/publications.

25. M. Vrtic, N. Schüssler, A. Erath, K. Meister, and K. Axhausen. Tageszeitliche Fahrtenmatrizen im Personenverkehr an Werktagen im jahr 2000. Research report, Swiss Federal Department for Environment, Transport, Energy and Communication, Swiss Federal Office for Spatial Development, Swiss Federal Roads Authority and Swiss Federal Office for Transport, IVT, ETH Zürich, Zürich, 2007.

26. SFSO. Eidgenössische Volkszählung. Swiss Federal Statistical Office, Neuchatel, 2000.

27. SFSO. Ergebnisse des Mikrozensus 2005 zum Verkehr. Swiss Federal Statistical Office, Neuchatel, 2006.