Simulating traffic flow with queues

Nils Eissfeldt[a], Daniel Krajzewicz[b], Kai Nagel[c], and Peter Wagner[b]

[a] Bayer Technology Services GmbH, BTS-PMT-SCL-Supply Chain Optimization 51368 Leverkusen, Germany

[b] German Aerospace Centre (DLR), Institute of Transport Research Rutherfordstrasse 2, 12489 Berlin, Germany

[c] TU Berlin Sek SG 12; Salzufer 17-19; D-10587 Berlin; Germany

eMail: nils.eissfeldt@bayertechnology.com, daniel.krajzewicz@dlr.de, nagel@vsp.tu-berlin.de, peter.wagner@dlr.de

1. Introduction

Recently, the simulation of large study areas with microscopic simulation tools has become increasingly popular. If the model is simple enough, more than one million vehicles (update time step size 1 sec) may be simulated in real-time on moderate hardware. However, for the applications that we are currently faced with, like traffic management, planning for millions of inhabitants and the related dynamic traffic assignment, even those fast simulation models are too slow.

So, the question may be asked how to further improve simulation speed with only small losses due to the improvements in computing speed. Recently, research has followed the idea to simulate the atomic building block of a network, the link, as a waiting queue. Different from standard queueing theory, the main difference is that when a link is physically full, vehicles had to wait on the upstream links, thus allowing to model spillback. Since then, a lot of work has been performed to make this or similar models application ready [1, 4, 7, 9, 18, 11], see [8] for additional references.

It has long been known [16, 17] that the congested side of the fundamental diagram from the model in [6] is unrealistic (too steep). In [3, 5] it was pointed out that this means that jams do not travel backwards. To see that, let us recall the simulation logic of this model (the following is not exactly the definition in [6], but it is close to it and does not differ in important details):

- At time t, a vehicle enters link i which has the length L_i;
- after the minimum travel time $T = L_i / v_{\max}$, where $v_{\max}$ is the maximum speed allowed on this link (alternatively, it may be the maximum speed allowed for this vehicle), the vehicle is allowed to leave the link, provided that the flow constraint Q, which is the capacity of the link, is not violated. This can be achieved by making sure that the time difference (headway) δt between two subsequent vehicles is always larger than 1/Q.
- If there is enough space on the next link (i.e. the number of vehicles n_{i+1} on this link is smaller than the maximum number N_{i+1}) the vehicle is going to, move the vehicle to the next link.

It is the last step that leads to unrealistic consequences. This can be understood by regarding a completely occupied downstream link i+1. If the first vehicle leaves the downstream queue, it will in reality take some time until the vacancy has travelled upstream so that the car from link i can proceed into i+1. In the queue model of [6], that vacancy is immediately available at the upstream end of the link.

There is another challenge here: usually, those models do not model long links directly, but instead divide a long link into smaller pieces of a particular length. These smaller pieces will be called cells in what follows to discern them from the word "link" which has the specific meaning of a link in a network. This division into cells is done to more accurately trace the traffic flow patterns, and obviously it costs a certain amount of performance.

2. Single link dynamics

To start with, a single long link with a traditional microscopic model is considered. The model which is used here is the model invented in [18], it is a stochastic variant of the Gipps model [19]. It is run in the form of an open link of a certain length *L*. This means, that the input q_{in} have to be specified as well as the outflow which is equivalent to the capacity of the link *Q*. One possible method to control Q is to put a traffic light at the end of link with a certain cycle time *C* and a green time *G* which changes from 0 to *C*. It is obvious, that the link is free when $q_{in} < Q$ holds, while it is congested if $q_{in} > Q$. Of course, congestion can only built up until the whole link is occupied by vehicles. If the congestion reaches the upstream end of the link, it effectively shuts-down the inflow. Depending on the assumptions made in the model about what is happening further upstream (outside the simulation area) sometimes very interesting space-time patterns could be observed.

Travel time *T* is defined as the time a vehicle needs from the time of entering this link to the time where it leaves the link. It is clear, then, that travel time is a function of capacity of the link and the number of vehicles on that link, i.e. $T(n,Q)$ where $Q=q_{max}G/C$ is the maximum outflow that can emerge from the link, with q_{max} the maximum flow rate which is possible in the model. However, this is an approximation, since the travel time depends in particular on the specific pattern of vehicles on the link, and so lead to different travel times for the same number of cars on the link. A typical example could be seen in Figure 1, where the travel time is shown as function of density n and output flow rate.

Note, of course, that the downstream capacity restriction can be modelled by other means as well, most notably by an off-ramp downstream, of by a speed-limit which is caused either virtually by congestion downstream or explicitly by setting a speed limit.

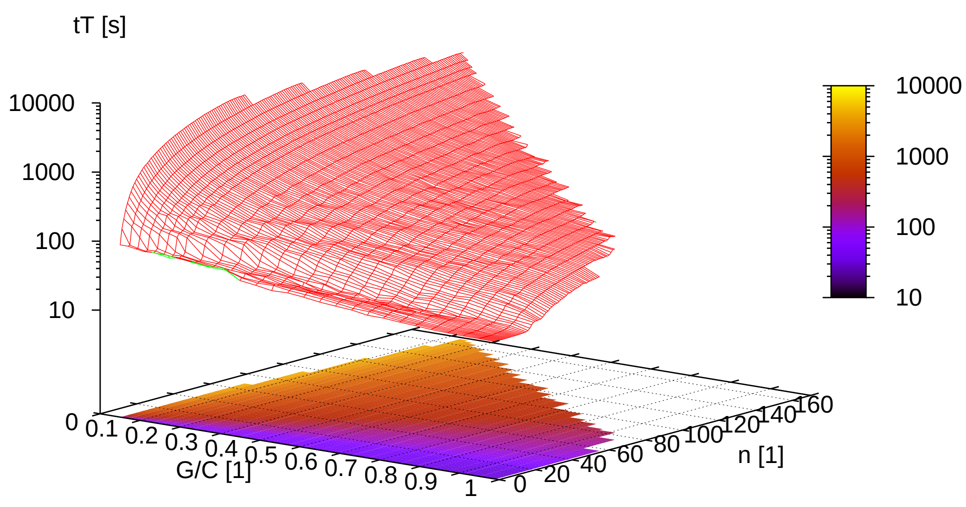

Figure 1. Travel time as function of the ratio of the green time G by cycle time C, and as function of

the number of vehicles on the link n. For small green times, the simulation time was not long enough to resolve the function, because the travel time of a vehicle becomes longer than the simulation time.

There is one surprising feature of those travel time functions. So far, we have not found any way to build a queueing model that uses such a function directly. I.e., like in planing applications it is tempting to try to set up a model where any vehicle, that enters a link, got a travel time according to the function drawn in Figure 1. Although there is no clear explanation for this, it might be due to the fact, that the large travel times are not caused by the number of vehicles on the link, but are caused by the capacity downstream. So, the travel time functions in dynamical models cannot be used as input, they are just a consequence of the underlying dynamics.

3. An improved model

Since the approach to use the link travel time function provided by a microscopic simulation model does not work, other means to improve the model in [6] have to be looked for. The example of the vacancy travelling upstream helps in developing such a better model. It needs the introduction of a state of each cell, which can be either jammed or in the free flow state $s_i = \{\mathrm{f}, \mathrm{j}\}$. These states develop as indicated in Figure 2.

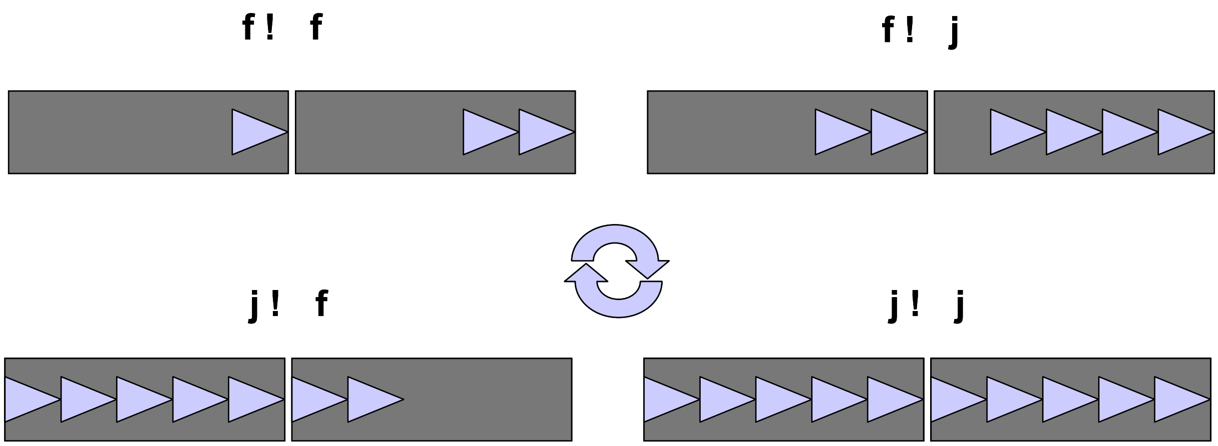

Figure 2: Sequence of possible states, starting from a completely free situation (upper left) to a jam on the downstream cell (upper right), to a totally jammed state (lower right), a dissolving jam (lower left) and back into free flow.

To develop a queueing model, the time interval (headway) δt between two vehicles leaving the upstream cell has to be specified. Four different cases may occur:

- **f → f**: both cells are free, therefore the flow between the two cells is unconstrained. That means, that $\delta t \geq \tau_{ff} = 1/Q$ should hold, where Q is the capacity of the cell and τ_{ff} is the associated minimum time headway between two cars.
- **f → j**: congestion starts in the downstream cell. Since the upstream jam front travels backward, it reaches the upstream cell only if $n_{i+1} \cong N_{i+1}$. Therefore, the flow goes unabated with $\delta t \geq \tau_{ff}$ as long as n_{i+1} is different from N_{i+1}, only for $n_{i+1} \cong N_{i+1}$ a small increase in δt might be observable which can be ignored.
- **j → j**: both cells are jammed now. This means, that the time needed for a vehicle to leave the upstream cell is, as discussed above, determined by the time a hole created at the outflow of the downstream link needs to travel upstream. This is a time $\delta t \propto \tau_{ff} n_{i+1}$, which is the earliest

moment for the front vehicle on link i to move into cell $i+1$. The time constant τ_{jf} is the time a vehicle needs to leave a jam; this number is usually larger than the free flow time headway τ_{ff}, which is the reason for the stability of jams. Except for small n_{i+1} this relation is correct, for $n_{i+1} < n_{crit}$ a different regime is entered: strictly speaking, it impossible to tell whether the cars are distributed in the cell or a located at the beginning or the end of the cell. Fortunately, this is not important, since the dissolution of the jam now tooks over to the upstream cell (see next item).

- **j → f**: In this case, vehicles leave a jammed cell. As stated above, this means that $\delta t \geq \tau_{jf}$, as long as $n_i \geq n_{crit}$. After that, the same argument as for the situation where j → j applies, i.e. without additional information it is not possible to tell what situation the cell is in. Of course, by using the information about the upstream cell $i-1$ it might be possible to get rid of this modelling error, i.e. letting $\delta t \geq \tau_{jf}$ as long as the jam has completely left this cell and only after that switch back to $\delta t \geq \tau_{ff}$. Furthermore, it only creates this (small) modelling error at the upstream front of a jam. If the upstream cell $i-1$ is in fact jammed, too, the same mechanism as described above takes over, i.e. for the interface between $i-1$ and i the j → f applies.

Putting anything together, the headway between two vehicles changing from cell i to $i+1$ can be defined as the following state-dependent function:

$$\delta t = \begin{cases} \tau_{fj} & \text{if } s_i = \{\mathrm{f}\} \text{ and } s_{i+1} = \{\mathrm{f}, \mathrm{j}\} \\ \tau_{jf} & \text{if } s_i = \{\mathrm{j}\} \text{ and } s_{i+1} = \{\mathrm{f}\} \\ \tau_{jf} n_{i+1} & \text{if } s_i = \{\mathrm{j}\} \text{ and } s_{i+1} = \{\mathrm{j}\} \end{cases} \tag{1}$$

This is a slightly simplified variant of the so called μ_4—model introduced in [5]. The final definition to be made is when a link has to be regarded jammed or free. Here, a simple threshold is used:

$$s_i = \{f\} \quad \text{if } n_i \leq n_{crit} = f_{crit} N_i \quad \text{and} \quad s_i = \{j\} \quad \text{if } n_i > n_{crit} = f_{crit} N_i \tag{2}$$

Note, that only three parameters are needed to describe any microscopic model by its corresponding queue model, namely τ_{ff}, τ_{jf}, and f_{crit}. Another parameter is hidden in the definition of N_i: it contains the jam density, which is the average distance between vehicles in a jam, and this is in fact a model depedent parameter. To realistically replicate the dynamics of the corresponding microscopic model, the correct jam density has to be inserted, which is the forth parameter of the model. This is quite similar to what is being used in the cell transmission model [21], therefore hopefully a kind of minimal description of the traffic flow dynamics on a link has been found.

A simulation of the model with many cells can be done in two different update schemes. In the following, a time-stepped scheme has been used, where the first vehicle in a cell is checked in any time-step of a certain size *h* whether it can leave the link or not. The model used in the application in section 4 below, however, uses a more elegant event-driven scheme by utilizing a priority queue which holds an pointer to any vehicle which stands at the head of its cell. Especially for a large number of queues, especially when large parts of the network have a low occupancy, this is much more efficient than the time-step-driven scheme.

The model has been checked at the level of single links by calibrating its parameters to real-world traffic counts, one example is shown in Figs. 3 and 4. The freeway section in Fig. 3 has been simulated by feeding flow data from the loops at station # 50 into the simulation, and restricting the outflow to the data given by loops at # 80.

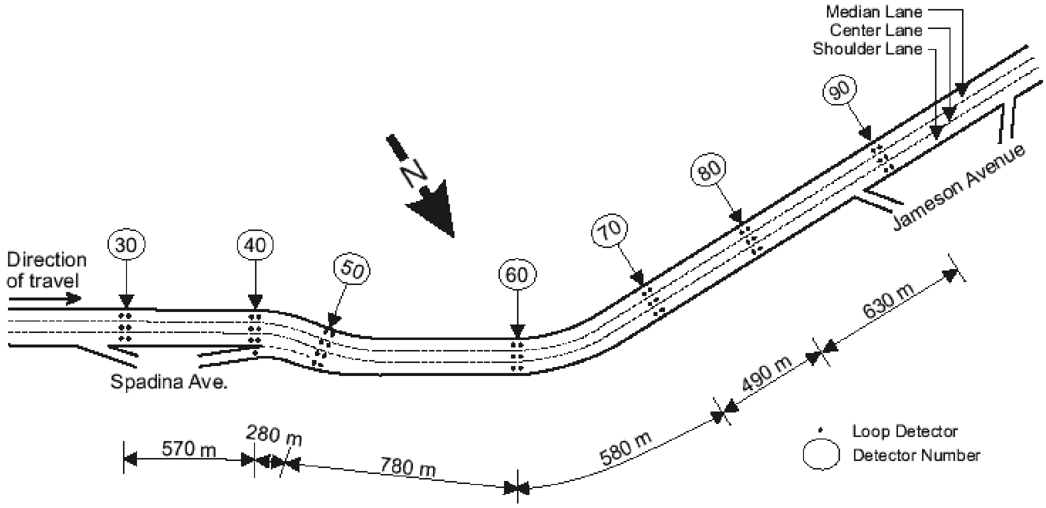

Figure 3: The freeway site used in this study is a segment of Gardiner Expressway which is located in metropolitan Toronto, Canada.

Figure 4 demonstrates, that the approach defined above is capable of describing the most prominent features of traffic flow, i.e. the emergence of traffic jams which in this case are generated upstream and travel backward into the study area.

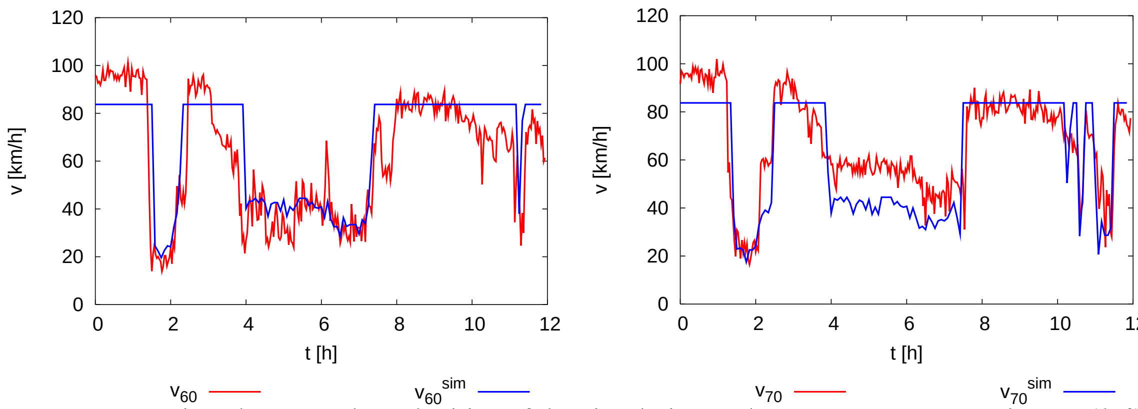

Figure 4: Comparison between the velocities of the simulation and measurements at station 60 (*left*) and 70 (*right*). Note, that the model as specified so far is a deterministic one, therefore any fluctuations imposed by the boundaries rapidly die out.

Translation

As stated above, it is assumed that almost any microscopic model might be translated into its queueing model counterpart. In the frame work laid down so far, this can be done by measuring the various parameters directly from the corresponding microscopic simulation, if it cannot computed explicitly (which is the normal case, since most models are so complicated, that no analytical approach is possible). Here, a slightly different approach is used by explicitly searching for the best fit as a function of the parameters $\tau_{ff}, \tau_{jf}, f_{crit}$ for the difference between the space time density functions $n_q(x,t), n(x,t)$ for the queue and the normal microsimulation, respectively:

$$e = \sum_{t=0}^{T} \sum_{x=0}^{X} \left(n_q(x,t) - n(x,t) \right)^2 \qquad (3)$$

A particular example is shown in Figure 5, the left side of the space time diagram is for the normal microsimulation, while the right side is for the corresponding queue model which fits the microsimulation results best. The agreement is in fact very good.

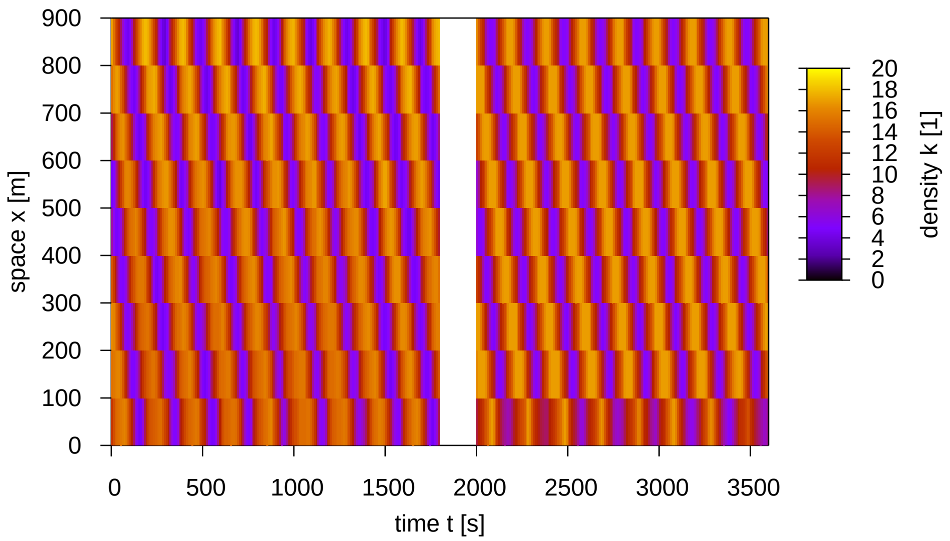

Figure 5. Left is the space time diagram for the microsimulation, the right part is for the corresponding queue model. The jams generated at the outflow end (x = 900 m) travel backwards in both models in exactly the same manner. Small difference can be seen at the feeding (input) end (x = 0) of the simulation area.

Scaling

The most interesting question is: how do the parameters scale with the length of a cell? If there is a weak dependence only, the model can very easily used to simulate large areas with very heterogeneous link lengthes. And in fact, this seems to be the case. As Figure 6 demonstrates, this time by comparing the travel time functions of queue models with different cell lengthes, there is just a small error introduced which in terms of travel times is just a few percent (on average; there are some situations, where the difference is larger).

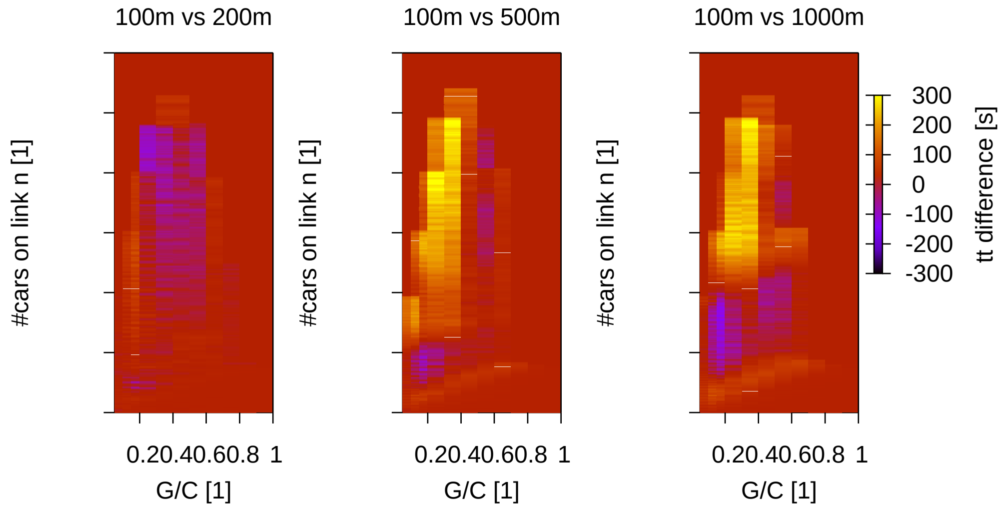

Figure 6. Differences in the travel time function tT(n,Q) between a simulation with cell size $\lambda = 100m$ compared to cell sizes $\lambda = 200m$ (left), $\lambda = 500m$ (middle), and $\lambda = 1000m$ (right).

Compute times

In principle, such a model could be much faster than a corresponding microscopic model. In the simple comparison between the microscopic model used to compute Figure 1 and the corresponding queueing model, the speed-up is around a factor of ten, by using a time step size of $h = 0.1$ s and a cell length of 100 m. For longer cells, or larger time intervals (which increase, however, also the error made by the model in comparison to the microscopic model), even larger speed-ups could be achieved. Internally, the implementation of the queueing model uses the same car object as do the fully microscopic simulation. However, in real applications with complicated networks, a lot of the compute time is not used by the update of the vehicles, but for the organization of the simulation itself. Therefore, only a part of the speed-ups are realized in a simulation model like SUMO [23], which has been used to run some of the applications described below. Another interesting feature is that such a model is more robust against simple coding errors and modelling errors when feeding the network with external data, which is interesting for simulations in its own right.

4. Applications

A number of real-world applications had been run with this or similar models, each simulating networks with more 20,000…50,000 links and more than 1 Million travellers and several millions of trips:

- Demand forecast for Zürich (a planning application, [15, 12]),
- Short-term forecasting of the traffic state in Cologne (traffic management),
- World Youth Day 2005 in Cologne [2], and the Soccer World Championship in Cologne 2006 (traffic management) [20].

While in the planning applications an accuracy similar to traditional planning tools has been achieved [15, 12], the accuracy of the traffic management applications still needs improvement. This is due to the fact that it is not only the quality of the traffic flow model used that matters, but additionally the quality of the algorithm to merge simulation and online traffic data, under the burden of real-time (see also [14, 13]). And of course, it is due to the quality of the online data itself.

Simulating Cologne

Nevertheless, the last application will be discussed here shortly, more details will be found soon elsewhere [20]. To get a complete picture of the traffic in the conurbation of Cologne, the following simulation had been set-up. The infrastructure data, especially the network, had been extracted from a commercially available data-base. The final simulation network consists of 39389 links and 16879 nodes. No additional data about the traffic lights had been available, so all the intersection had been modeled very simply according to the priority of the links that make up the intersection.

Demand modeling

In a first step, the demand for travel had to be estimated, both for the normal situation without the World Cup, and for the additional demand generated by the World Cup itself. This was not only for the visitors traveling to the stadium, but also for the visitors who went to the public viewing places. Four sources of information have been used to describe the traffic demand for the given area. A previously generated synthetic population of the urban area of Cologne was used for the inner-city area as input to a microscopic demand generation tool TAPAS [24]. These data have been used because they were well validated within several precursor projects. Because this data-set only covers the inner-city area, other data to model traffic around the city had to be used. For this purpose, the area of interest was extracted from the VALIDATE data-set by ptv which covers long-distance travel in entire Germany. To avoid having trips double – from TAPAS and VALIDATE – only those data from VALIDATE were used that were either entirely outside the city area, or were going into the inner-city area from outside, or were leaving the inner-city area. TAPAS was used to model inner-city traffic only. A third data set was generated to resemble the expected visitors' traffic. Additionally, previous data from highway-detectors were used to compute the distribution of routes over the highway network to supplement the routes generated so far.

Dynamic stochastic traffic assignment

The second step consisted in the determination of the dynamic user equilibrium consistent with this demand. This step had been done in advance (offline), because it is very time-consuming. It consists of an iterated day-to-day simulation with rerouting, which is a standard procedure to determine the simulation-based user equilibrium. This equilibrium state (for different days of the week) have been used in the final application, where a forecast of the traffic state 30 min into the future was sought.

Prediction

To predict the traffic, the actual simulated state had to be corrected by the loop and other data available on line. This has been done by correcting the flows on the links which where equipped with detectors. The speeds have been used directly to fix the constant v_{max} at any link with a detector, as described in the model description above.

Note, that this does not lead to an exact match between simulation and reality, since the loop detectors have measurement errors, and since the improvement of a certain loop detector might worsen the mismatch between data and simulation on an adjacent detector. The actual match from this process between simulated and measured flows and speeds is shown in Figure 7. It could be seen, that there is still a systematic mismatch especially in the flows. This is currently under inspection.

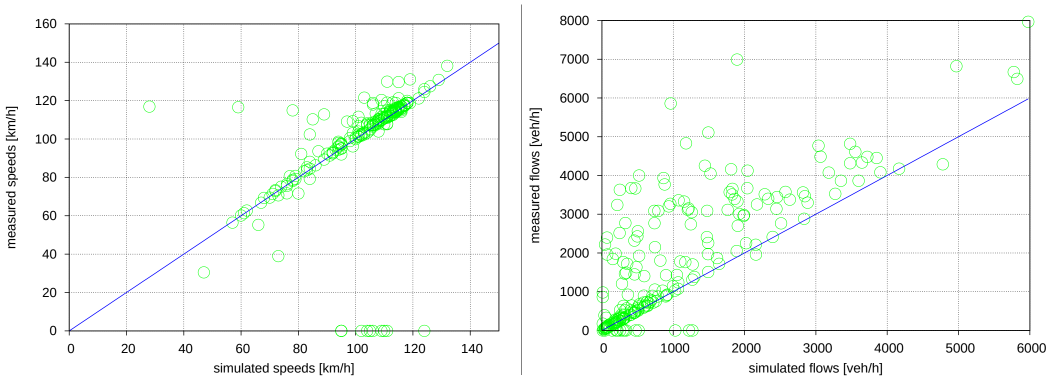

Figure 7. Comparison between simulated (x-axis) and measured data (y-axis); left are the speeds, right picture shows the flows. While the data-points with measured flow zero are due to mal-functioning detectors, the scatter above the diagonal still waits for an explanation.

5. Open questions & future developments

It has demonstrated, that queueing models can be used quite successful to run simulations of really large areas. They can describe the traffic flow of reality, and of (hopefully) most microscopic traffic flow models up to a certain level of fidelity, which may be sufficient for large study areas. Despite the fact, that performance (in comparison to reality) seems to decrease with larger cell-sizes, it is possible to use them to model real links with different sizes. Therefore, such models are ideally suited to get good enough to use them in daily works. Hopefully, it seems, that it is only a small step towards using them in daily applied work.

Obviously, there is still room for improvement. Just to mention a few:

- The modeling of inhomogeneous fleets of vehicles. It is clear, that especially the time constants depend on the type of the vehicle, with trucks having larger headways. Another difficulty arises from the different maximum speeds of the vehicles, because this sacrifices the FIFO discipline used. There are two possibilities around this obstacle: either introducing passing, or by modelling a link with more than one lane explicitly by as many queues as there are lanes, each one with a different maximum speed.
- The stochastic has to described correctly. In principle, this should be nothing more than instead of using a deterministic δt , a random number had to be added. Hopefully, not much is changed by this, but it should be made clear that this is indeed the case.
- Short links provide another challenge, since then the mechanism stated above does not apply anymore: with $N_i = 1$, n_{crit} is not a sensible variable anymore. However, it may be argued, that those links might be ignored altogether, since the vehicles simply rush over them, provided of course the downstream link is free.
- It is an open question how to model synchronized traffic flow with this model. However, according to [22], most of the microscopic simulation tools are not capable of doing that, so the approach demonstrated here suffices to approximate all the simulation tools without synchronized flow.

Another very interesting idea might to use the results here to built sensible hybrid models: short links, complicated intersections are modelled microscopically, while the large long links between those complicated areas are simulated by the queueing approach. This requires still some work to be done regarding a seamless coupling between the two very different models, however, as has been demonstrated in section 3, an algorithmic means to find the correct parameters of the queueing model corresponding to a particular micrsocopic model have been demonstrated.

References

[1] T. Akamatsu. An efficient algorithm for dynamic traffic assignment with queues. *Transportation Science*, 35(4):389–404, 2001.
[2] M. Bonert, E. Brockfeld, P. Brunke, I. Ernst, D. Krajzewicz, M. Ruhé, and P. Wagner. Airborne traffic monitoring during the XX world youth day 2005 in cologne. *Journal of Infrastructure Systems*, 2006. submitted.
[3] E. Brockfeld, R. D. Kühne, A. Skabardonis, and P. Wagner. Towards a benchmarking of microscopic traffic flow models. *Transportation Research Records*, 1852:124 – 129, 2003.
[4] C. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *TRB*, 28(4):269–287, 1994.
[5] N. Eissfeldt. *Vehicle-based modelling of traffic*. PhD thesis, University of Cologne, 2004.
[6] C. Gawron. An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model. *International Journal of Modern Physics C*, 9(3):393–407, 1998.
[7] D. Heidemann. A queueing theory model of nonstionary traffic flow. *Transportation Science*, 35(4):405–412, 2001.
[8] R. Jain and J. Smith. Modeling vehicular traffic flow using M/G/C/C state dependent queueing models. *Transportation Science*, 31(4):324–336, 1997.
[9] M. Mahut, M. Florian, and N. Tremblay. Space–time queues and dynamic traffic assignment: A model, algorithm and applications. TRB 2003 Annual Meeting CD-ROM, 2003.
[10] TRANSIMS. TRansportation ANalysis and SIMulation System. See transims.tsasa.lanl.gov.
[11] N. Vandaele, T. V. Woensel, and A. Verbruggen. A queueing based traffic flow model. *Transportation Research D*, 5(2):121–135, 2000.
[12] M. Balmer, B. Raney, and K. Nagel. Adjustment of activity timing and duration in an agent-based traffic flow simulation. In H. Timmermans, editor, *Progress in activity-based analysis*. In press. Also see www.vsp.tu-berlin.de/publications.
[13] G. Flötteröd and K. Nagel. Bayesian modeling and simulation of combined route and activity location choice. See www.vsp.tu-berlin.de/publications.
[14] G. Flötteröd and K. Nagel. Simulation and optimization of trajectories in a congested network. See www.vsp.tu-berlin.de/publications.
[15] B. Raney, N. Cetin, A. Völlmy, M. Vrtic, K. Axhausen, and K. Nagel. An agent-based microsimulation model of Swiss travel: First results. *Networks and Spatial Economics*, 3(1):23–41, 2003.
[16] P. Simon and K. Nagel. Simple queueing model applied to the city of Portland. Paper 99-0861, Transportation Research Board Annual Meeting, Washington, D.C., 1999.
[17] P. M. Simon, J. Esser, and K. Nagel. Simple queueing model applied to the city of Portland. *International Journal of Modern Physics C*, 10(5):941–960, 1999.
[18] S. Krauß, P. Wagner, and C. Gawron, Metastable states in a microscopic model of traffic flow, Phys. Rev. E 55, 5597 – 5605 (1997).
[19] P.G. Gipps, A behavioural car following model for computer simulation, Transportation Research B 15, 105 – 111 (1981).
[20] M. Bonert, E. Brockfeld, D. Krajzewicz, and P.Wagner, preprint.
[21] C.F. Daganzo, The Cell transmission model: a simple dynamical representation of highway traffic, Transp. Res. B 28, 269 – 287 (1994).
[22] B. Kerner, 2004 The Physics of Traffic (Berlin: Springer)
[23] SUMO, Simulation of Urban Mobility, open source simulation package, available on http://sumo.sourceforge.org, web page, last access 11.7.2006
[24] Hertkorn, G. and P. Wagner (2004), The application of microscopic activity based travel demand modelling in large scale simulations, paper presented at the 10th World Conference for Transport Research, Istanbul, Juli 2004.