Towards Benchmarking Microscopic Traffic Flow Models

Elmar Brockfeld, Reinhart Kühne, Peter Wagner
Institute of Transport Research
German Aerospace Center
Rutherfordstraße . 2
D-12489 Berlin, Germany

Abstract: For the simulation of traffic flows various macroscopic and microscopic models exist. Developing these models it is important to check the models against reality, whereas there are well-established means how to do this for macroscopic fluid-dynamical models. For microscopic models this work has yet to be done. This article is intended as a first step towards a common benchmarking of those models. A method is suggested to test microscopic models against single-car data sets and applied to two public available single-car data sets recorded by C. Daganzo on a one lane road in the USA in 1997. The calibration of the used models is done by analysing the deviations of the travel times on the observed street segments. As a result one gets a single number for each model giving the mean percentage error it produces in comparison to reality. This way the models can be compared directly to each other quantitatively. Furthermore, the obtained results point towards some deficiencies of the models and give ideas how to correct them.

Key Words: modeling, simulation, traffic flow, benchmark.

Introduction

A large number of macroscopic and microscopic models simulating traffic flow exists. Some of them are described in detail and some are not, see [1] and [2] for an overview on publicly available models, but all claim to reproduce the reality to a certain degree. For traffic flows this means to define equations or rules for the movement of the vehicles or the time evolution of flows, in order to reproduce phenomena like headway distributions, synchronised flow, jam-formation at bottlenecks and spontaneous jams. An important step is the calibration and validation of the models, which means to find the parameters that reproduce a given observed data set as close as possible. There are well-established means how to do this for macroscopic fluid-dynamical models, whose output can be directly matched against the aggregated data that are commonly recorded (see [3] for example). For microscopic models it is more difficult to find relevant microscopic data to test the models against, and even if such data are available, it is not entirely clear how to use these data for calibrating and validating the models.

Furthermore, another problem is that models tend to be calibrated only with a few data sets which are usually not publicly available. Some models may be performing well for special data sets, but it is very difficult to compare the models against each other for common data sets. As a result of this lack of common benchmarking, it may be roughly known which real situations a certain model is able to reproduce well, however the exact quality of real data reproduction is in general not well documented.

In this article some common microscopic traffic flow models are tested exemplarily using data recorded on a one-lane road. Furthermore, a technique for calibration is suggested how to do this.

At first the road where the data have been recorded is described including some modifications which have been performed preparing the data for use in the simulations. Further the boundary conditions used for all simulations and the simulation set-up are described.

Before the results of the simulations are presented and analysed, the method of error-measurement is defined. It analyses the deviations of travel times, which are produced by the simulations compared to real data.

Finally, some conclusions and an outlook on future work is given.

1 Modeling and simulation set-up

1.1 The experimental site and the data sets

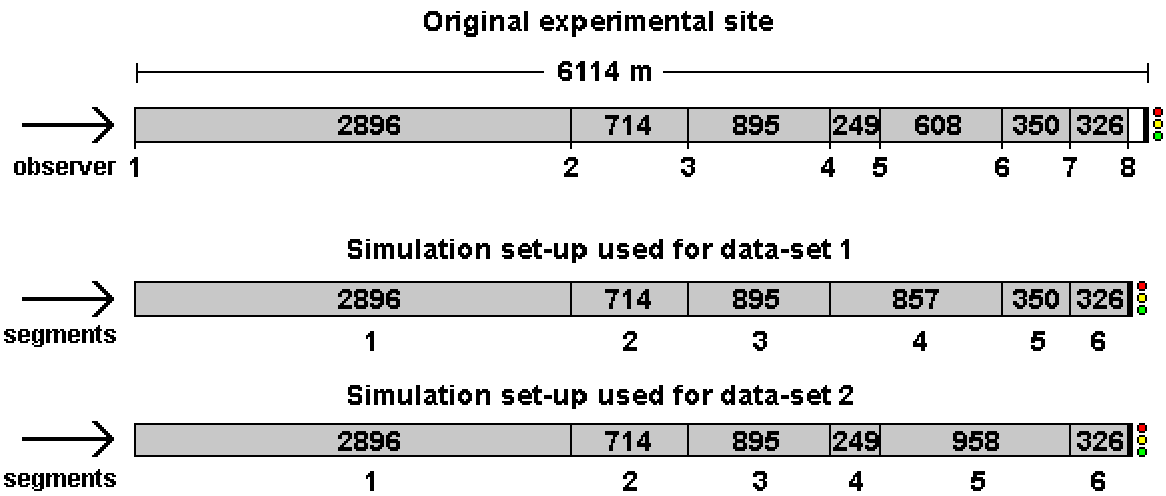

Fig. 1: At the top a sketch of the experimental site in reality is pictured, showing the observer positions on the road. The sketch in the middle shows the site as used for the simulations reproducing the first data set and at the bottom for the second data set (segment lengths in [meter]).

The data sets used for calibration ([4]) have been recorded by C. Daganzo et al. on a long one-lane road with a total of 4 miles as described in [5]. In one direction eight observers were positioned on the road with distances to each other decreasing towards the traffic light at the end of the road as shown at the top of figure 1. The traffic light causes congested states on the road. Each observer was a person who clicked a key on a laptop each time a vehicle passed the observer. From time to time a special car drove along the road, which defined the first car in the sequence of cumulative arrivals.

Therefore, the data sets consist of the arrival times of all the cars that passed the lane on the road. So this data set explicitly contains the travel times between the observers, and the number of cars that are on the stretch of road between the observers, in the form of the so called *N(t)*-curves.

Daganzo's team recorded data sets on Tuesday, November 18, 1997 and on Thursday, November 20, 1997, each from 6:45 AM to 9:00 AM containing information of about 2300 vehicles. Not surprisingly, because of hardware malfunctions, computer failures, and human errors the raw data sets contain some errors (miscounts) that could not be corrected completely. But of course Daganzo et al. tried to identify each vehicle and produced "justified IDs", which can be found in their Excel-Sheets.

For a microscopic simulation it is useful to have a complete set of observer times for each vehicle. So - based on the "justified IDs" - the data sets have been slightly modified by eliminating doubled IDs and inserting missing IDs with the passing time of the vehicle before. The number of counts can be seen in the following table.

Location of error	Type of error	Correction	Counts
First data set	Doubled justified Ids	Last ID cancelled	37
First data set	Missing ID	ID inserted with passing time of ID-1	63
First data set	Empty data at observer 5	Complete data of observer cancelled	791
Second data set	Doubled justified Ids	Last ID cancelled	165
Second data set	Missing ID	ID inserted with passing time of ID-1	35
Second data set	Empty data at observer 6	Complete data of observer cancelled	409

Table 1: Modifications of the data sets.

Finally, a refined data set was used in our analysis. A total of 2298 vehicles for the first data set from 06:47:46 - 08:57:12 AM, and 2293 vehicles for the second data set from 06:44:35 - 08:52:46 AM. For the simulations, the data of observer 5 in the first data set and those of observer 6 in the second data set have been cancelled for the calibrations because of some missing data in the time series. The resulting changes in the simulation-set-up are shown in the middle and bottom of figure 1.

The models that have been tested against the data sets did not use the topography of the road nor used the fact that drivers have knowledge about the traffic light at the end of the road that may influence their behaviour. The traffic light settings are not part of the data set therefore some workarounds had to be made in order to set the appropriate boundary conditions for the models as described in the next section of this text. Finally, the vehicle type data - which are included in the data sets – were not used.

1.2 Boundary conditions

Very important for all traffic flow models are the boundary conditions. Since each vehicle has been recorded in the data sets, the inflow and the outflow are known, unfortunately the velocities are not. So the following assumptions have been made to get a realistic behaviour at the boundaries.

At the beginning of the road the vehicles are introduced at the times corresponding to the time-recorded data of observer 1. It has been assumed that the vehicles enter the experimental site with free flow velocity that will be slightly slower than the maximum allowed velocity of $v_{\max} = 50mph \approx 22.22 \frac{m}{s}$. Of course every vehicle entering the road has to fulfil the same rules for moving as the used model defines, respectively. If it is not able to enter the first segment with free-flow velocity, it has to wait and probably enter the road in the next time step.

The outflow boundary condition is a bit more complicated to handle. The optimum would be to have the exact cycle times of the traffic light. In this case, the road could be blocked for the vehicles each time the traffic light turned to red. As mentioned above, this information is not available. Even worse, the traffic light is vehicle-actuated and so the cycle time changes permanently. First attempts using the *N(t)*-curves in order to guess the states of the traffic light were not satisfying. Finally the following procedure has been implemented to overcome this problem. The idea is to let the vehicles leave the system approximately when they passed the last observer 8 - whose real position is 75 meter apart from the traffic light - in the observed data. For that purpose a “virtual traffic light” is set exactly at the position of the last observer, which lets the vehicles pass only when they already passed it in the observed data. This way in the simulation no vehicle can pass the last observer earlier than the observed data shows.

2 Analysis

2.1 Method of measurement

Typically, comparing simulated to real data, the relation between flow and density (“fundamental diagram”) or those between velocity and density are used. For microscopic models one can go much more into detail and compare them using microscopic data.

Out of all possible measurements (see [6] for an overview) the chosen basic measurement is the travel time of vehicles on each segment of a street. So in this article the comparison was done by subtracting the observed and the simulated travel times for each vehicle. From the observed data the total amount of vehicles N and the observed passing times are available and the travel times $tobs_i^s$ for each vehicle *i* on segment *s* can be calculated. From the simulation the simulated travel times $tsim_i^s$ for every vehicle on each segment are obtained. So the mean absolute deviation over all vehicles on each segment s can be calculated:

$$f_{mean}^s = \frac{1}{N} \sum_{i=1}^{N} \left| tobs_i^s - tsim_i^s \right| .$$

In order to get a measurement not for each segment but for all segments together one has to weigh the deviations. The suggested way is to relate the mean absolute deviations to the average travel times $tobs_{mean}^s$ as calculated from the observed data for each segment (see tables 2 and 3 for details). This way one calculates as a first step the percentage mean deviation for each segment:

$$f^s_{percent} = \frac{f^s_{mean}}{tobs^s_{mean}} * 100 \text{, where } tobs^s_{mean} = \frac{1}{N}\sum_{i=1}^{N} tobs^s_i \text{ .}$$

Accumulating these values for each segment, finally the mean percentage deviation over all segments together is given by:

$$f_{percent} = \frac{1}{S}\sum_{s=1}^{S} f^s_{percent} \text{ ,}$$

where $S = 6$ is the number of segments.

These two measurements in percent will be used in the following to give an idea of the relative deviations the simulation produces in comparison to the observed data.

Segment	**1**	**2**	**3**	**4**	**5**	**6**
Length [m]	2896	714	895	858	350	326
Average travel time [mm:ss]	03:20	01:25	02:11	02:52	01:19	01:18
Average speed [km/h]	52,13	30,24	24,59	17,96	15,95	17,26

Table 2: Average values calculated from the first data set.

Segment	**1**	**2**	**3**	**4**	**5**	**6**
Length [m]	2896	714	895	249	958	326
Average travel time [mm:ss]	02:36	01:09	01:43	00:46	03:25	01:22
Average speed [km/h]	66,83	37,25	31,28	19,48	16,82	14,31

Table 3: Average values calculated from the second data set.

2.2 Finding the optimal parameter settings for each model

In order to find the optimal parameter settings for each model fitting the observed data well, basically an optimisation algorithm was used known as the downhill-simplex method in multi-dimensions [7]. This algorithm tries to minimise an objective function (the error-function $f_{percent}$ or $f^s_{percent}$ as defined above). For initialisation a simplex is set up in the multi-dimensional space of solutions by calculating the error-function for a set of parameter constellations of a used model. Once initialised, the algorithm tries to minimise the function by estimating “better” parameter constellations using heuristic methods causing the simplex to stretch, move and shrink. This process is repeated letting the simplex shrink together till it reaches a local minimum in the error-function. Of course it is only a local but not a global minimum. So for initialisation various possibilities of parameter constellations have to be tested to increase the probability of finding a global minimum. As usual, numerical minimisation cannot guarantee that a global minimum is found.

A problem using this algorithm is related to the space of solutions to be explored. For example it is obvious that the maximum velocity of vehicles – a parameter used in all models, which have been analysed – can not be negative nor having a value much greater than the allowed speed limit. The same applies for acceleration and deceleration. Therefore, the algorithm has been changed in order to force it to stay within realistic boundaries for each model parameter.

2.3 The structure of the observed data sets

Analysing the data collected by Daganzo it can be found easily, that the jams caused by the traffic light propagate against the driving direction; these congested states finally reach observer 2 and become visible in the first segment. This can clearly be seen when looking at the travel times for segment 1 in figure 2.

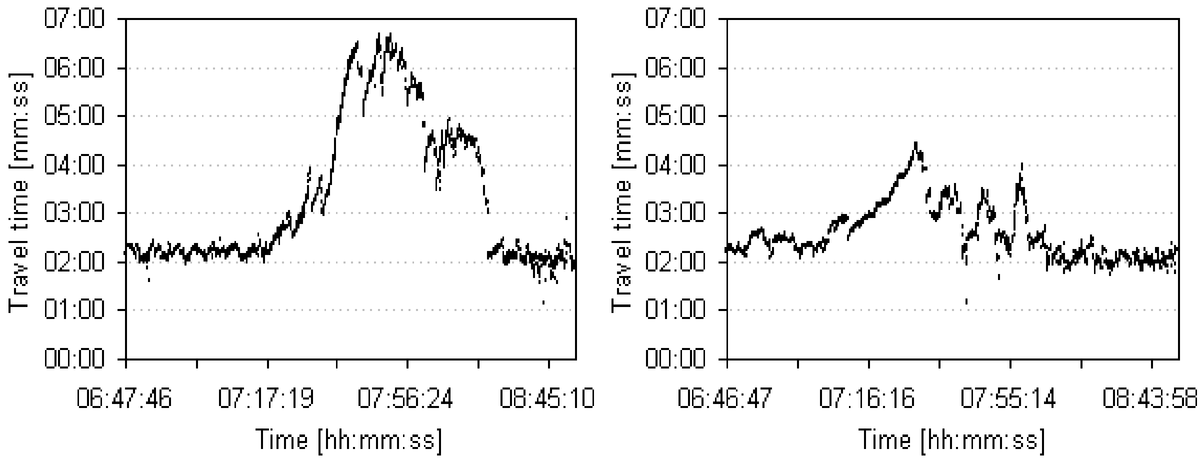

Fig. 2: Travel times on the first segment (length of 2896 m) as recorded in data set 1 (left) and data set 2 (right).

For the first data set there is a free-flow phase at the beginning and the end with approx. 2:10 minutes travel time, which means an average velocity of $22.28\frac{m}{s}(\approx 80.2km/h)$. Jammed states occur in the middle with travel times up to 6:30 minutes and an average speed of $7.43\frac{m}{s}(\approx 26.74km/h)$. In the second data set the same structure can be found except for the fact that the jammed states are obviously not as pronounced as in the first. The travel times reach a maximum of only 4:30 minutes. Therefore, the important goal was to reproduce the lengths of the jams and thus the correct backward-propagation of jams.

3 Simulation results

3.1 Optimisation of all segments together

The simulations have been performed for the following microscopic traffic flow models:

- SK: car following model by S. Krauss [8],
- CA: cellular automaton model by K. Nagel and M. Schreckenberg [9],
- VDR: enhanced CA-model with Velocity Dependend Randomisation by R. Barlovic et al. [10],
- OVM: Optimal Velocity Model by M. Bando et al. [11],
- IDM: Intelligent Driver Model by M. Treiber et al. [12].

While the SK-, OVM- and IDM-model could be treated as defined in the corresponding publications, the parameter settings of the CA- and the VDR-model had to be modified. Both models are fully discrete cellular automata and use normally a cell-length (= vehicle-length) of 7.5 m, which causes a free-flow velocity which is much too low for the data set at hand. For the simulations presented in this article this cell-length was increased to 8.5 m in order to obtain a more realistic free-flow-velocity. For all five models the optimisation process was started several times and the minimal obtained errors $f_{percent}$ were taken for the final results as they are shown in figure 3.

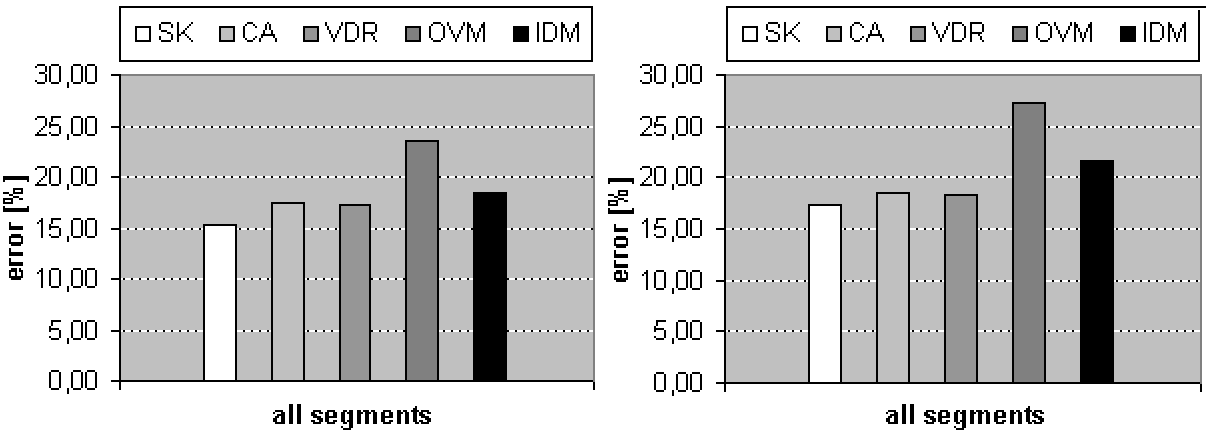

Fig. 3: Final results of the simulations minimising the error-function $f_{percent}$ (error over all segments together) for data set 1 (left) and data set 2 (right).

The best results for both data sets were found for the SK-model with errors of 15% for the first data set and 17% for the second, followed by the very simple CA- and the VDR-model with 17% and 18%, respectively. The much more sophisticated IDM-model can not reach these results with 18% resp. 21%. The largest error was produced by the OVM-model with 23% resp. 27% error.

Looking deeper into the results obtained for the individual segments in figure 4 it is found, that for both data sets the greatest errors are caused in the middle segments. The errors on the segments 1 and 6 are generally quite good with errors between 9% and 18%. For segment 1 there are three basic explanations for that. At first, this is the segment with the greatest length, so the exact driving behaviour will not play such a very important role for analysing the travel times. Second, it is the segment with the fewest states of congested flow as these states become more frequent as one approaches the traffic lights. Third, the boundary condition for the inflow will have some effect because the vehicles are forced to enter the road nearly at the same time as they do in the observed data set. For the sixth segment directly in front of the traffic light the outflow conditions play a similar role. In addition this is a segment where the traffic states are similar the whole time, namely jammed states changing regularly with starting and braking vehicles.

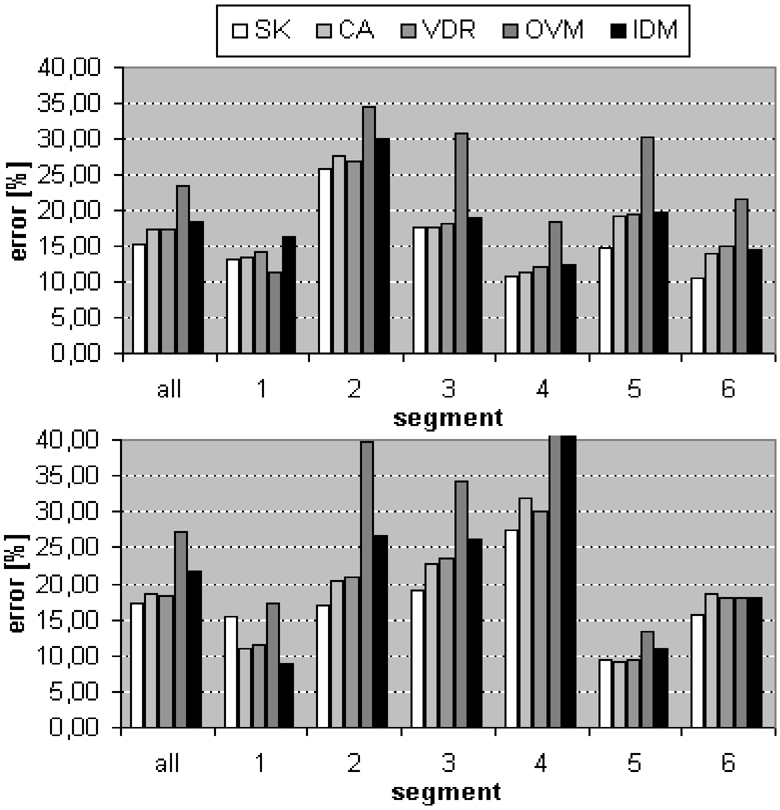

Fig. 4: Errors produced on each segment after minimising the error function $f_{percent}$ (optimisation has been performed over all segments together) for data set 1 (top) and data set 2 (bottom).

Going backwards from the traffic light, from segment 5 to segment 2, the deviations for the first data set increase. This can be explained by the propagation of jams against the driving direction. Getting far away from the traffic light, the less congested states occur and the jams are not that dominant. Furthermore, the influence of the segments in front becomes stronger, which means propagation of the errors. For the second data set the situation is similar, having in mind that segment 4 is a very small one with 249 meter length and thus very difficult to reproduce.

The SK-model outperforms the other models nearly in all segments except for the first. For data set 1 surprisingly the OVM-model produces the lowest errors in the first segment while the IDM-model does this for the second data set with an error of only 8%.

3.2 Optimisation of each segment separately for the SK-model

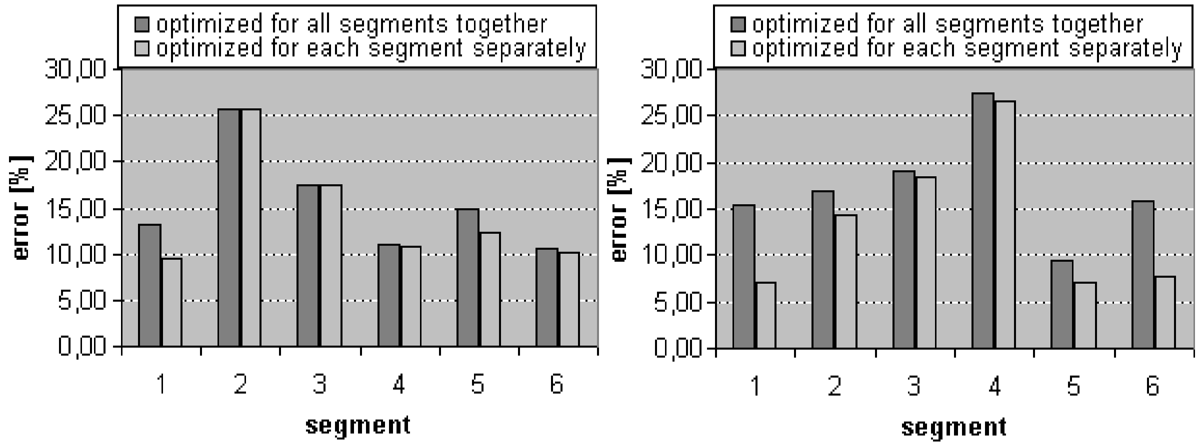

Fig. 5: Errors produced on each segment after minimising the error function $f_{percent}$ (optimisation over all segments together) in comparison to the errors produced minimising the $f^{s}_{percent}$ functions (optimising each segment separately). Results for the first data set on the left, for the second on the right.

Exemplarily for the SK-model, parameter values have been determined by optimising each segment individually, in order to have an idea about the robustness of the results. So, the functions $f^{s}_{percent}$ for each segment s have been optimised. The results for the optimal values of the functions are shown in figure 5 together with the optimisation values as obtained by the optimisation over all segments. Optimising each segment separately, especially for the second data set it turns out that the deviations can be minimised much more than with optimising over all segments together. Especially on the segments 1, 5 and 6 the error can be minimized to very small values of 7 and 8 %. For the first segment figure 6 shows the evolution of the travel times exemplarily.

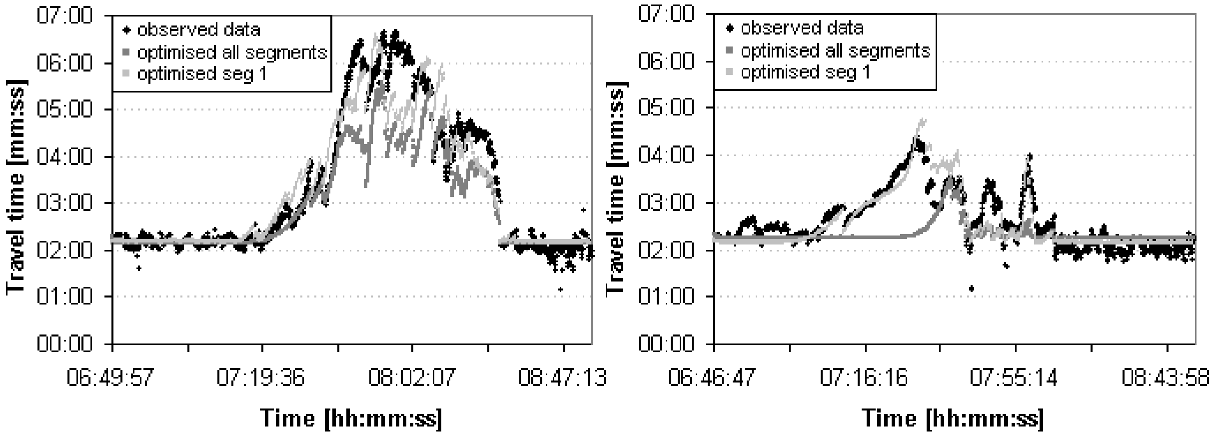

Fig. 6: Evolution of the travel times on segment 1 for data set 1 (left) and data set 2 (right). Shown are the observed travel times in comparison to the times optimising segment 1 separately and the resulting evolution on segment 1 when optimising over all segments.

Additionally analysing the obtained parameter values for the different minimisations as listed in table 4 it stands out that optimising segment 6, which is directly in front of the traffic light, the values for random braking and time delays in the reaction of the drivers are smaller than by optimizing the other segments. A possible conclusion is that the driver behaviour in front of the traffic light is much less dawdling than in the other jammed traffic situations. This seems to be realistic since the drivers in front of the traffic light have more than information about a few vehicles directly in front of them. The state of the traffic light gives the drivers a guess about the traffic situation in a much wider area so they will react much quicker than in other jammed situations.

	Vmax [m/s]	a [m/s^2]	b [m/s^2]	e_a	tau [s]	%-error
all segments	21,64	2,325	1,397	0,43	1,691	17,3
segment 1	22,60	0,463	1,024	0,46	1,534	7,1
segment 2	22,48	1,552	1,541	0,59	1,750	14,3
segment 3	20,95	3,494	1,369	0,48	1,688	18,4
segment 4	20,44	3,968	1,335	0,32	1,664	26,6
segment 5	21,12	2,833	2,163	0,57	1,565	7,1
segment 6	20,79	5,787	1,266	0,33	1,361	7,7

Table 4: Obtained parameter values for the SK-model optimising the second data set over all segments together and every segment separately. Vmax is the maximum velocity of the cars, a the acceleration and b the deceleration. e_a and tau are parameters which cause random brakings and time delays in the driver reaction.

4 Conclusions and further plans

The results obtained point towards some problems the used models have reproducing the observed data sets. On a few segments of the observed road some models produced acceptable errors of 10% or less. But especially the observed propagation and inner structure of jams is not described well enough. First detailed analyses of the SK-model give hints how models could be enhanced.

In future work on the one hand more models will be tested against the data sets. On the other hand new public available data sets will be recorded by the Institute of Transportation Research to make further steps towards benchmarking traffic flow models. All researchers are invited to test yet implemented models or their own models under the topic “clearing” on http://ivf.dlr.de/. Finally, detailed criteria should be developed concerning a benchmark for traffic flow models so that existing models can be objectively compared to each other, forcing the development of better models.

Acknowledgements

It is a pleasure to thank Carlos Daganzo for making this beautiful data set available on his web-site. Part of the work reported here has been supported by a grant from the German Ministry for Education and Research in an effort to launch a new institute for transport science.

References

[1] A. Schadschneider, D. Chowdhury, L. Santen: Statistical Physics of Vehicular Traffic and Some Related Systems, Physics Reports, 2000, vol 329, 199.

[2] D. Helbing: Verkehrsdynamik: Neue Physikalische Modellierungskonzepte, Berlin, Germany: Springer, 1996.

[3] M. Cremer, M. Papageorgiou: Parameter identification for a traffic flow model, Automatica 17, 1981, pp. 837-843.

[4] Carlos F. Daganzo: San Pablo Dam Road Data, 1998, http://www.ce.berkeley.edu/~daganzo/, Accessed July 24., 2001.

[5] Karen R. Smilowitz, Carlos F. Daganzo, Michael J. Cassidy and Robert L. Bertini: Some Observations of Highway Traffic in Long Queues, TRB 990094, 1998.

[6] Fred L. Hall: Traffic Stream Characteristics, in: Traffic Flow Theory: A State-of-the-Art Report, 1992, http://www-cta.ornl.gov/cta/research/trb/tft.htm, Accessed Nov 8., 2001.

[7] William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian Flannery: Numerical Recipes in C: The art of Scientific Computing, 2nd Edition, Cambridge University Press, 1992.

[8] Stefan Krauß: Microscopic modeling of traffic flow: Investigation of Collision Free Vehicle Dynamics, Ph.D.–Thesis, University of Cologne, 1998.

[9] K. Nagel, M. Schreckenberg, J. Physique I, 2221, vol 2,1992.

[10] R. Barlovic, L. Santen, A. Schadschneider and M. Schreckenberg: Metastable States in Celular Automata for Traffic Flow, Europe. Journ. Phys. B, 793, vol 5, 1998.

[11] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, Y. Sugiyama: Dynamical model of traffic congestion and numerical simulation, Physical Review E, 239-253, vol 51, 1999.

[12] M. Treiber et. al., Physical Review E, 1805-1824, vol 62, 2000.