# TESTING AND BENCHMARKING OF

# MICROSCOPIC TRAFFIC FLOW SIMULATION MODELS

**Elmar Brockfeld ***
Institute of Transport Research, German Aerospace Center
Rutherfordstrasse 2, 12489 Berlin, Germany
phone: +49 30 67055 231, fax: +49 30 67055 202
email: elmar.brockfeld@dlr.de

**Peter Wagner**
Institute of Transport Research, German Aerospace Center
Rutherfordstrasse 2, 12489 Berlin, Germany
phone: +49 30 67055 237, fax: +49 30 67055 202
email: peter.wagner@dlr.de

For Presentation at the
WCTR 2004 Conference
July 4th-7th 2004
Istanbul, Turkey

Submission date: February 10th, 2004

**Corresponding author*

**Abstract.** Microscopic simulation models are becoming increasingly important tools in modelling transport systems. There are a large number of available models used in many countries. The important difficult stage in the development and use of such models is the calibration and validation of the microscopic sub-models describing the traffic flow, such as the car following models for example. The aim of this paper is to present recent progress in calibrating more than a dozen microscopic traffic flow models with very different data sets conducted by DGPS-equipped cars (Differential Global Positioning System), loop detectors and human observers. Different approaches to measure the errors the models produce in comparison to reality are compared. It can be stated that from a microscopic point of view errors of about 15-20% in headway- and travel time-estimation and about 2-7% in speed-estimation of individual vehicles in the car following process seem to be the minimal reachable level. Furthermore, the larger the simulation horizon is, the smaller the diversity of the analyzed models become in comparison to the diversity in the driver behaviour. Most interesting, no model cold be denoted to be the best and especially highly sophisticated models did not produce better results than very simple ones.

## INTRODUCTION

For the simulation of traffic flow various macroscopic and microscopic models exist (see (Chowdhury2000) and (Helbing2001) for an overview) and there are a large number of available models used in many countries. Nowadays, in the time of computer processors increasing rapidly in speed, especially microscopic models become very important tools in modeling transport systems. In the development of these models it is important to check the models against reality, namely to calibrate and validate them. Usually the developers of the models do this on their own using some data sets they have access to and publish the results obtained (see (Brackstone1999) for an overview on model calibration). This way every model is calibrated and sometimes validated with other data sets, but if a user has to decide which model to take for some special application, he is not really able to compare them and to choose the best one. So there seems to be a lack in the field of benchmarking these models.

This paper attempts to integrate and broaden some recently published approaches (Brockfeld2002, Brockfeld2003, Brockfeld2004, Ranjitkar2004, Schober2003, Schober2004) to calibrate microscopic traffic flow models and tries to identify the minimum errors the models produce in comparison to reality. In all approaches various models have been tested with the same data sets, which makes it possible to compare them directly to each other. This can be seen as a first step in developing a transparent benchmark for these kind of models.

In the following, five approaches for calibration of the models with various data sets are described. At first, the (four) used data sets are briefly sketched, followed by a definition of the procedure how the calibration errors were basically measured, describing the errors in the time series of gaps, velocities and/or flows. Subsequently, the analyzed models are listed and briefly described. The calibration results obtained for a specific data set conducted in Hokkaido, Japan, are presented in detail, while the results of the other calibration approaches - following a similar procedure - are briefly described. At last, an overall conclusion is given, which centralizes the results obtained.

## THE DATA SETS

The four data sets used for the calibration and validation of the models are of very different kind. The first two are vehicle trajectory data especially consisting of the vehicle positions and speeds which are stored in 0.1 second intervals. This allows a calibration and validation of the models in a very microscopic way by analyzing the car following behavior in time series of speeds and those of headways to leading vehicles. The other two data sets contain microscopic data, two, but are analyzed in a bit more macroscopic way analyzing the travel times of individual vehicles and the time series in flow and speed over a bunch of vehicles. All data sets consist of various subsets covering many different traffic situations such as congested situations with backward jam propagation, free driving and signalized intersections. Two data sets regard one lane roads and two multilane roads.

### Car following data measured by radar sensors on road ways in the USA (ICC FOT)

This data set is from the intelligent cruise control field operational test (ICC FOT) conducted in the late 1990s in the USA. A large amount of drivers took part in the test, having their cars equipped with radar sensors measuring the distances to adjacent vehicles in 0.1 second intervals. Furthermore, the vehicle positions and speeds have been stored in the same time interval. To analyze the car following behavior and compare it to the models, special short traces were taken for the calibration approach. These are only traces where one of the equipped cars followed another car over a sufficiently long time on the same lane. The idea of the data collectors to perform the test was to investigate the performance of a prototype Adaptive Cruise Control system (ACC). Thus, drivers could choose between manually driving and the cruise control. The data used for the analyses contain only manual driving. For the simulations the data of about 100 drivers of the first week of the test were taken. In total, about 300 traces (the longest three traces of each driver) with time lengths of about one to a few minutes were taken for the calibration procedure.

### DGPS (Differential Global Positioning System) data measured on a test track in Hokkaido, Japan

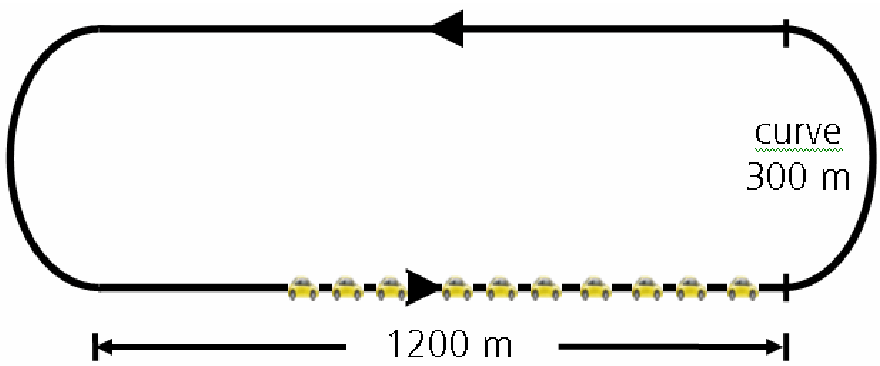



Figure 1: Sketch of the test track in Hokkaido, Japan, with ten cars driving on the course.

This data set has been recorded on a test track in Japan in October 2001 (Gurusinghe2003). Eight experiments have been conducted, where nine cars drove on a 3 km test track (2 x 1.2 km straight segments and 2 x 0.3 km curves; see figure 1) for about 15-30 minutes in each experiment following a lead car, which performed some driving patterns. These are for example driving with constant speeds of 20, 40, 60 and 80 km/h for some time, varying speeds (regularly increasing/decreasing speed) and emulating many accelerations/decelerations as they are typical at intersections. The regularly increasing/decreasing of speed is done performing half, single, double and triple waves on the two straight segments of length 1.2 km on the test track. That means for example a half wave is starting with 40 km/h, accelerating to 60 km/h on the middle of a segment and decelerating to 40 km/h at the end of it. A single wave is accelerating from 40 to 60 at the first quarter of the segment, decreasing to 40 km/h in the second quarter and 20 km/h in the third quarter, and accelerating to 40 km/h in the fourth quarter.

To minimize driver-dependent correlations between the data sets, the drivers were exchanged between the cars after each experiment. Having all cars equipped with the differential global positioning system DGPS, the position of each car is stored in 0.1 second intervals throughout each experiment. From these data other important variables like the speed, the acceleration and the headway between the cars were extracted for simulation purposes. The accuracy of the DGPS is about 1 cm and the appointment of the speeds has got an error of less than 0.2 km/h as described in (Gurusinghe2003). Thus, the data sets have got such a high resolution that they are adequate for the analysis of car-following behavior and calibration of car-following models.

## Data recorded on a one lane road in California, USA (San Pablo Dam Road)

The data set has been recorded in 1997 by C. Daganzo et al. (DAGANZO1997) on a one-lane road in California, USA, with a total of 4 miles as described in (Smilowitz2000). In one direction eight human observers were positioned at the road with distances to each other decreasing towards a traffic light at the end of the road as shown in figure 2. The traffic light causes congested states on the road, which propagate backwards up to the first segment “1”. 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. This was primarily done to minimize data errors with an adjustment of the data sets ex post.

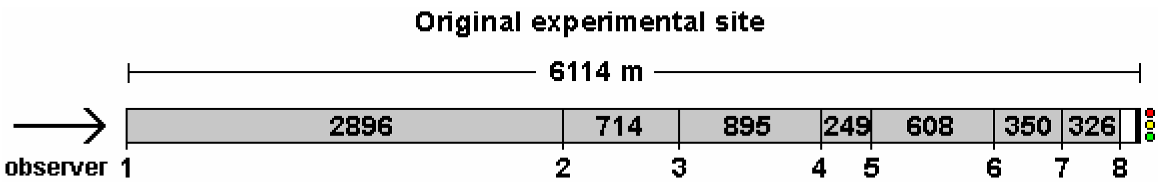



Figure 2 Sketch of the experimental site.

Therefore, the data sets consist of the arrival times at the observers of all the cars that passed the lane on the road. So this data set explicitly contains the travel times between the observers of each singular vehicle.
Daganzo’s team recorded data sets on Tuesday, November 18, 1997 and on Thursday, November 20, 1997, each from about 6:45 AM to 9:00 AM containing information of about 2298 and 2293 vehicles.

### Loop detector data measured on the I-80 highway in Berkeley, USA, Berkeley Highway Laboratory (BHL)

This data set is from double-loop detectors, which measure flows and speeds on the I-80 highway in the USA, in direction from San Francisco to Sacramento. For the calibration approach three loop detectors were taken as shown in figure 3, covering a segment of about 1067 meters without an on-ramp, but one less used off-ramp. The used data set in from one complete day in 2002 (BHL2003).

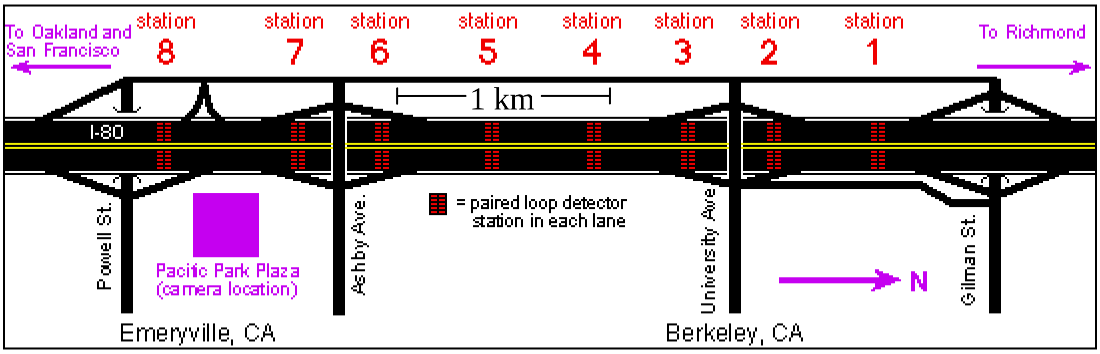



Figure 3: Sketch of the I-80 highway nearby Berkeley. The 1km-segment between station 4 and 6 was used for simulation purposes.

## ERROR MEASUREMENT AND OPTIMAL PARAMETER FINDING

In all the simulation approaches the absolute error a model produces in comparison to a measured data set is calculated via the simple distance between a recorded time series and a simulated time series. To get a percentage error it is additionally related to the average value of the time series in each particular data set:

$$e = \frac{\frac{1}{T}\sum_{t=0}^{T}\left|x^{(sim)}(t) - x^{(obs)}(t)\right|}{\frac{1}{T}\sum_{t=0}^{T}x^{(obs)}(t)}, \qquad (1)$$

where $x^{(sim)}$ and $x^{(obs)}$ are a simulated and an observed traffic flow variable, which can be travel time of singular cars, gap between two cars or speed. $T$ is the time series over the total time of each experiment. Figure 4 shows an example for an error measurement using gaps, which produced an error of about 18 %.

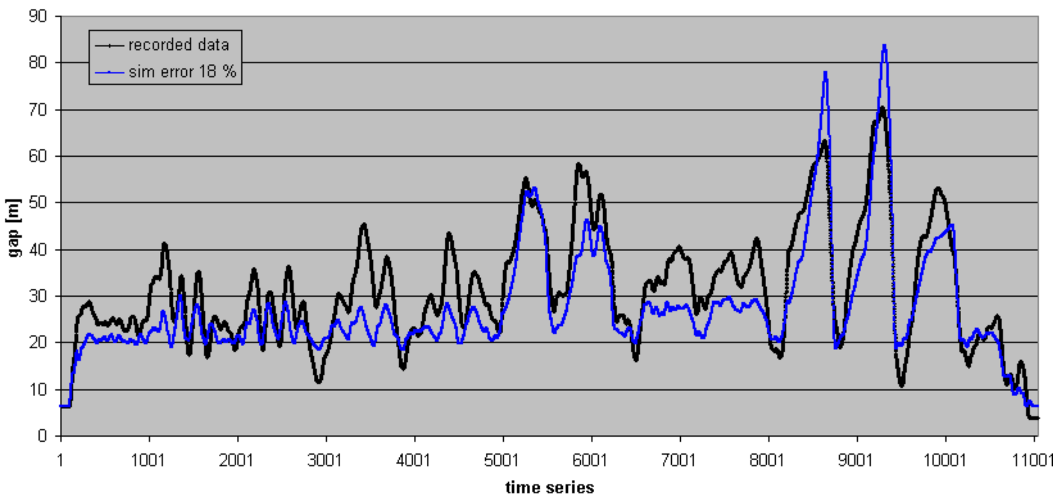



Figure 4: Example for error measurement using the time series of gaps.

To find the optimal parameter constellations in the calibration procedures a gradient-free optimization method known as the "downhill simplex method" (Press1992) was used by four of the approaches (Brockfeld2002, Brockfeld2003, Brockfeld2004, Schober2003, Schober2004, approach with the BHL data) and started many times with different initialization values for each "model-data set" pair. The variation in initialization is done to avoid sticking with a local minimum, which of course can occur because getting a global minimum can not be guaranteed by those optimization algorithms. The approach of Ranjitkar (Ranjitkar2003), which calibrates models with the data conducted in Japan, uses a genetic algorithm GENECOP III, basically defined by Z. Michalewicz (Michalewicz1992).

## THE MODELS

The models used for the simulations are mainly microscopic traffic flow models, which describe the behavior of a following car in relation to a leading car. For the vehicle movement, typically equations like the following were used defining the new speed of a vehicle at time $t + \Delta t$, depending on the status of some variables at time $t$:

$$\begin{aligned} v(t+\Delta t) &= f(g(t), v(t), V(t), \{p\}) \\ g(t+\Delta t) &= V(t) - v(t), \end{aligned} \tag{2}$$

where $v$ is the speed of the following and $V$ that of the leading car, respectively, and $g$ is the headway between the cars. The symbol $\{p\}$ denotes a set of parameters of the model under consideration.

In all the calibration approaches some of the following traffic flow models of very different kind with 3 to 15 parameters have been tested. Except for the last two, which are mesoscopic models, they are all microscopic models. Some models are used in commercial simulation programs, which are popular in Germany, the USA, Great Britain and Japan, and some are scientific simulation approaches.

- 4 parameters, CA (cellular automaton model by K. Nagel, M. Schreckenberg) (Nagel1992),
- 7 p., IDM ("Intelligent Driver Model", Helbing) (Helbing2001),
- 5 p., SK (model by S. Krauss) (Krauss1997),
- 7 p., SK_STAR (model based on the SK-model by S. Krauss) (Krauss1997),
- 4 p., OVM ("Optimal Velocity Model", Bando, Hasebe) (Bando1995),
- 7 p., IDMM ("Intelligent Driver Model with Memory" by Treiber) (Treiber2003),
- 7 p., Newell (model by G. Newell, can be understood as the continuous CA with more variable acceleration and deceleration) (Newell1962, Newell2002),
- 3 p., Castello (model based on the Lighthill-Whitham theory) (Castello1994),
- 6 p., GIPPSLIKE (basic model by P.G. Gipps) (GIPPS1981),
- 6 p., Aerde (model used in the simulation package INTEGRATION) (Crowther2001),
- 13 p., FRITZSCHE (model used in the british software PARAMICS; it is similar, but not identical to what is used in the german software VISSIM by PTV) (Fritzsche1994),
- 15 p., MitSim (model by Yang and Koutsopulus, used in the software MitSim) (Iftekhar1999),
- 7 p., VDR (enhanced CA-model with Velocity Dependend Randomisation by R. Barlovic et al.) (Barlovic1998),

- 10 p., VDR++ (enhanced VDR model with additional brake light and realization of what is known as “synchronized traffic flow”) (Knopse2000),
- 11 p., caSync (further extension of the VDR model, again with realization of what is known as “synchronized traffic flow”) (Kerner2002),
- 5 p., ECS excess critical speed stimulus response model (Gurusinghe2001),
- 4 p., CT: Cell transmission model (Daganzo1997),
- 5 p., uQUEUE: mesoscopic queueing model.

The most basic parameters used by the models are the car length, the maximum speed, an acceleration rate (except for the CA0.1-model) and a deceleration rate (for most models). The acceleration and deceleration rates are specified in more detail in some models depending on the current speed or the current headway to the leading vehicle. Furthermore, some models (CA0.1, VDR, VDR++, caSync, SK_STAR and MitSim) use a parameter for random braking or another kind of stochastic parameter describing individual driver behavior. Most models use something like a reaction time of the drivers to the behavior of the leading car.

With these kinds of parameters a lot of the models are covered, except for some models with deeper conceptual design. The VDR++ and the caSync are both models which try to reproduce what is known as “synchronized traffic flow”. For that purpose a following mode is defined, which a vehicle adopts in special situations where it follows the leading vehicle with keeping the distance to it stable. The IDMM has as a special feature a memory effect. Depending on the density ahead, the cars try to hold their speeds according to a rolling horizon. The MitSim model defines two thresholds concerning the headway, which cause a switching between three different driving modes. Especially if a driver is very close to the leader the calculations become very sophisticated, depending on the headway, own speed, speed-difference and the current density. In addition to the basic simulation update equation (2) the model needs the speed of the leader one time step before as a special feature. The FRITZSCHE model provides switching to various driving modes, too. For this model the switching depends not only on the headway ($g$), but also on the speed-difference ($dV$) between the follower and the leader. Thus, a ($dV$,$g$)-car following plane is divided into different regions of free driving, approaching, emergency brake and two other driving behaviors. As a specific, differing to equation (2), the model needs the acceleration of the follower and the leader one time step before and uses some kind of “brake light” of the leader by reacting on its deceleration.

The two mesoscopic models are used only for simulation of longer road links and networks and do not implement driving behavior. As in macroscopic simulation models the road links are divided into segments containing flow or occupancy values, which are changed in time according to some differential equations.

# CALIBRATION APPROACHES

## Vehicle trace data of the ICC FOT

The calibration approach was done by Schober et al. in 2003/2004 (Schober2002, Schober2003). The 300 highway-traces with a length of one up to a few minutes have been taken to test three models (IDM, SK and SK_STAR). For simulation purposes the models were calibrated with each of the traces, taking into account a leading and a following car. So, in each case the leading vehicle is moved as calculated from the data measurements and the following car is moved according to the update rules of the model under consideration. For each of the traces the calibration was first performed using the headways (range r) and then using the velocities. Furthermore, the calibration was done using data of three different velocity classes, which means a division according to different driving situations.

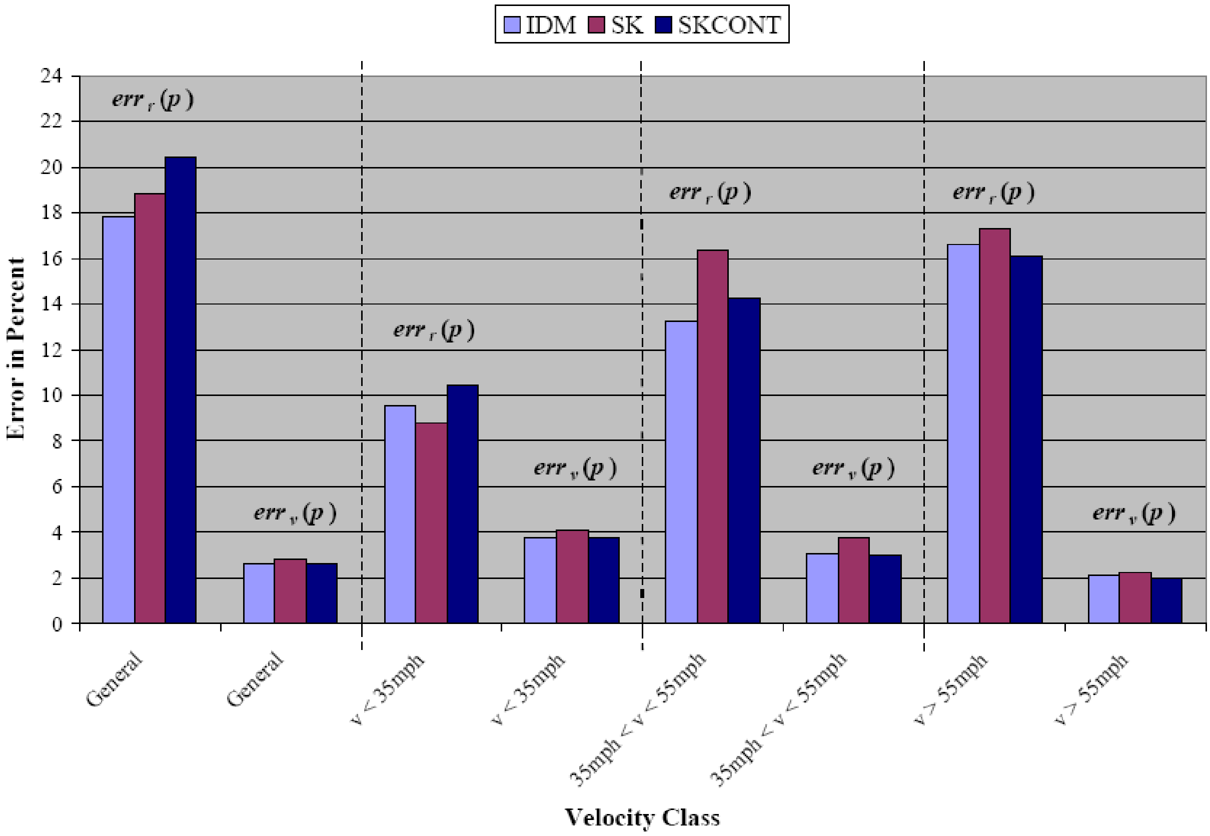



Figure 5: Calibration results for the ICC FOT data (model "SK_STAR" is named "SKCONT" in this diagram)

The calibration results are shown in figure 5 showing the errors produced by calibration of headways and by calibration of speeds. For the headways errors of about 18 % to 21 % are found for general driving conditions, while depending on the velocity classes the errors rapidly increase with increasing speed. The errors obtained by calibration of speeds are generally on a very low level. In the general case there are errors of about 2.5 %. For the speed classes it is the opposite of the errors in headway, because the errors get bigger the lower the speed is. Generally it can be stated that the errors measuring the speeds are very much lower than the errors obtained by measuring the headways. This can be explained by the fact that a follower every time tries to adjust its speed to the leader and for the models it is relatively easy to reproduce this. In the headway case the problem will be that if an error is developing in the simulation, the error is propagated and influences a part of the forthcoming simulation sequence.

## DGPS test track data from Hokkaido

Two calibration approaches were performed using data from the test track in Hokkaido. As the time step for the models should be 0.1 seconds according to the accuracy in the recorded data, some models with a traditional time step of 1 second – as for example used for simple cellular automatons - have been modified to adopt for an arbitrarily small time-step. Thus, every model is simulated with a time step of 0.1 seconds.

**Approach by Ranjitkar et al.**

In a first approach, Ranjitkar et al. analyzed short sequences on the 1.2km straight sections of the test track and discards the data of the curve sections (Ranjitkar2004). Thus, the calibration is done for about 47 traces of about 1-2 minutes length. For simulation set up they moved the first driver as the recorded speeds in the data sets tell and let the following nine drivers move according to the used model, respectively. For the error measurement they took the time series of speeds and then those of the headways. The calibration results are as listed in table 1.

| Model No. | Models | Percentile error | | |
|---|---|---|---|---|
| | | Mean | Standard deviation | Coefficient of variation |
| **Case I: Using speed data** | | | | |
| 1 | Krauss model | 4.19% | 1.18% | 28.19% |
| 2 | Gipps model | 4.30% | 1.09% | 25.26% |
| 3 | Newell model | 4.71% | 1.54% | 32.76% |
| 4 | ECS model | 3.87% | 1.64% | 42.41% |
| 5 | Bando model | 8.46% | 3.11% | 36.73% |
| 6 | Castello model | 4.42% | 1.48% | 33.60% |
| **Case II: Using headway data** | | | | |
| 1 | Krauss model | 12.04% | 4.53% | 37.60% |
| 2 | Gipps model | 12.20% | 4.65% | 38.09% |
| 3 | Newell model | 12.91% | 5.13% | 39.72% |
| 4 | ECS model | 15.38% | 7.34% | 47.69% |
| 5 | Bando model | 18.47% | 9.20% | 49.80% |
| 6 | Castello model | 20.99% | 12.19% | 58.05% |
| | Data error (discrepancy between speed and headway data) | 0.87% | 0.52% | 60.26% |

Table 1: Mean error, standard deviation and coefficient of variation of the calibrated models using speed and headway for error measurement. ("OVM" is denoted as "Bando" and "SK" as "Krauss" here.)

The results show that using the speed time series for error measurement five models have errors of about 4-5 %. Only the OVM (model by Bando) does not perform well with errors of about 8.5 %. For the headway data errors of 12-21% are obtained with a group of three models (SK, Gipps and Newell) performing well with about 12-13 %.

Concerning the headway measurement figure 6 shows an example of some more detailed results for scheme A (certain order of the ten drivers). Obviously the driver behavior is very different producing error ranges of about 9-17 % for different drivers for the best models Gipps, Krauss and Newell. Calibrating the driver behavior one more step into detail it can be seen in figure 7 that optimizing the parameters for special driving patterns may produce better results with an error of less than 10 % concerning the headways.

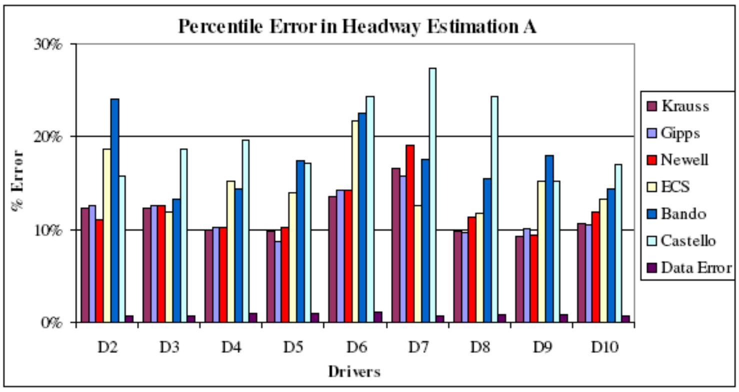
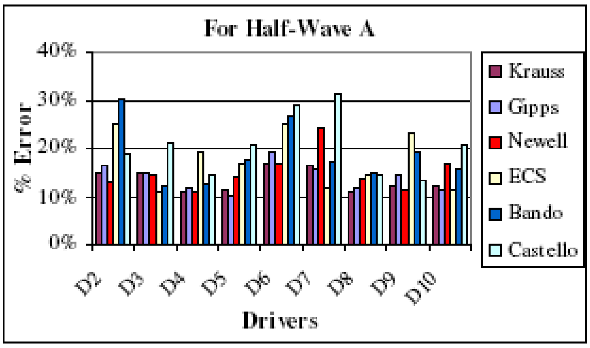
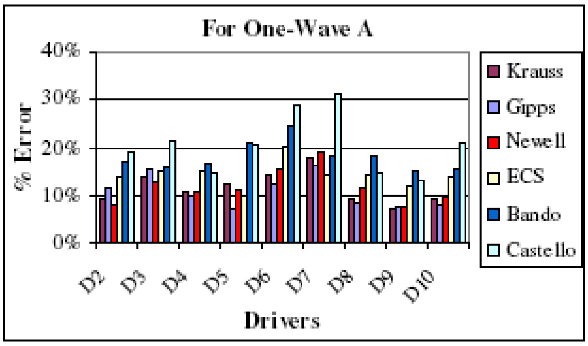


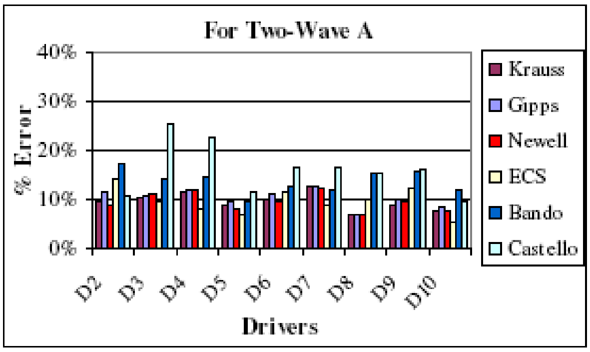

Figure 6: Calibration errors of the models for individual drivers using headway data.



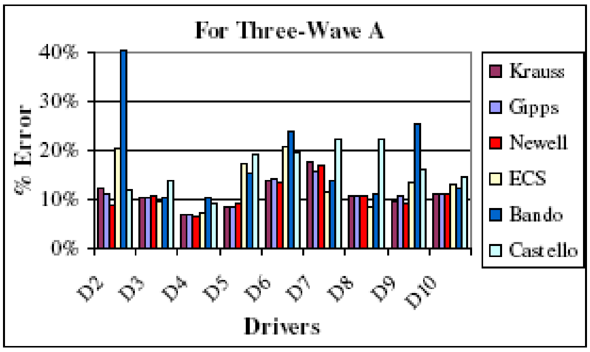

a) For half-wave speed pattern



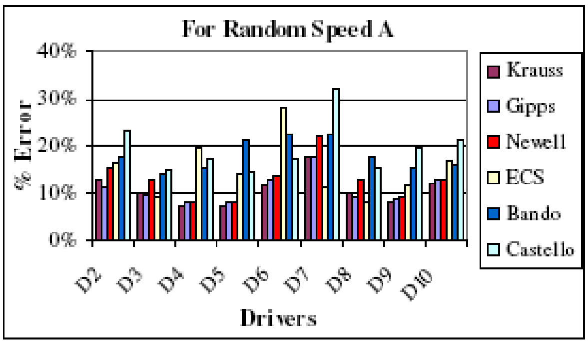

b) For one-wave speed pattern



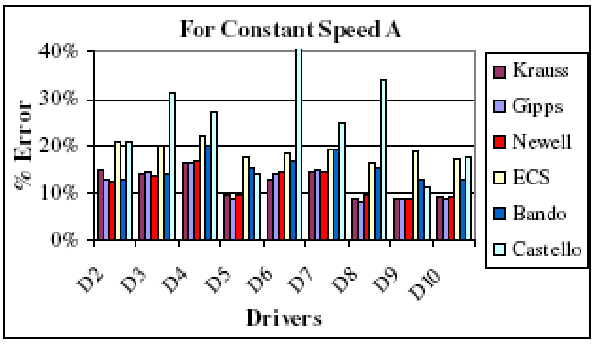

c) For two-wave speed pattern




d) For three-wave speed pattern




e) For random speed pattern




f) For constant speed pattern (20 Km/hr)

Figure 7: Calibration errors of the models for different driving patterns using headway data.

**Approach by Brockfeld et al.**
Brockfeld et al. analyzed the data set in a more general way (Brockfeld2004). They took not only the short sequences of 1-2 minutes on the straight segments. For the calibration the traces over the complete experiment times were taken from four of the eight experiments covering 26 minutes in the first experiment, 25 minutes in the second, 18 in the third and 14 in the fourth experiment. In all these experiments various driving patterns were performed. Thus, the calibration is done over more generalized data sets than in the approach of Ranjitkar et al. In small difference to Ranjitkar the calibration is done for vehicle pairs. Thus, the idea of the simulation is to move a leading car as recorded in the data and the follower as the equations of the model under consideration define.

Altogether 4*9=36 vehicle pairs (four experiments, each with nine vehicle pairs) were used as data sets for the analyses of the car following behavior. Each model has been calibrated with each of the 36 different constellations separately gaining optimal parameter sets for each “model-data set” combination. Figure 8 shows exemplarily the calibration results obtained for the first experiment (“11”). In this case one driver pair (“11_8”) can be reproduced well with errors of about 10 %. Other driver pairs like “11_6” or “11_1” are much harder to reproduce with errors up to 17-20 %. In total, for all 36 constellations, the errors mainly range from 12 % to 17 %. In nearly all cases the models do not differ so much when reproducing the behavior of a driver pair, because the average differences between the models reproducing the single driver pairs is about 2.5 percentage points. Remarkably is here that this diversity of the models is much smaller than the differences in the driver behavior (mainly about 5 percentage points), as can be seen exemplarily in figure 8, too.

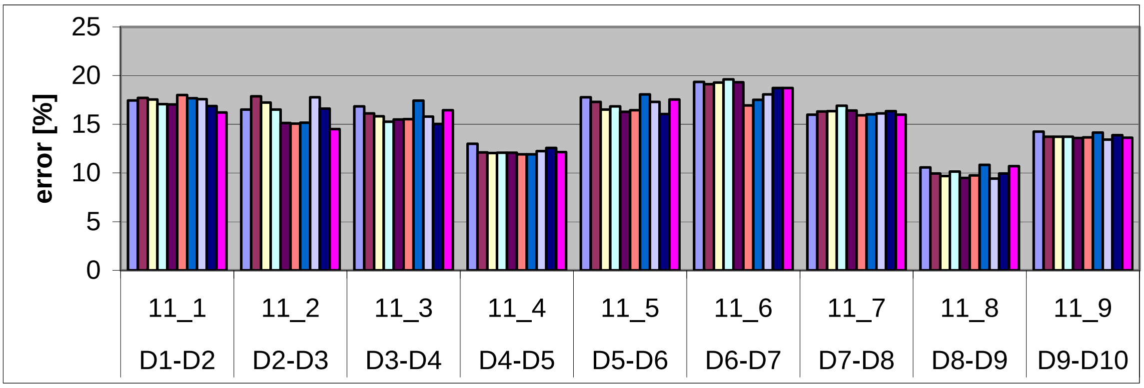



Figure 8: Example for calibration results obtained for the driver pairs in the first experiment.

Looking at the average errors each model produces with the 36 data sets, it can be seen in figure 9, that, again, the differences of the models are not very big. The best model produces an error of 15.14 %, the worst one of 16.20 %. Thus, no model can be denoted to be the best and especially complex models do not produce better results than simple models.

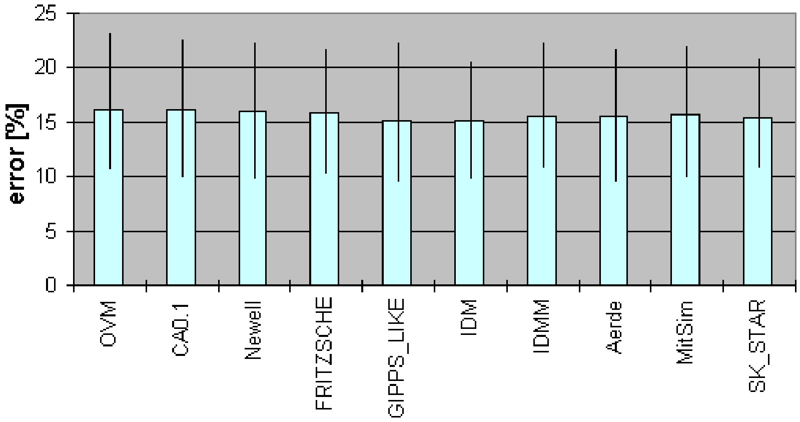



Figure 9: Mean calibration results for all models including the result range.

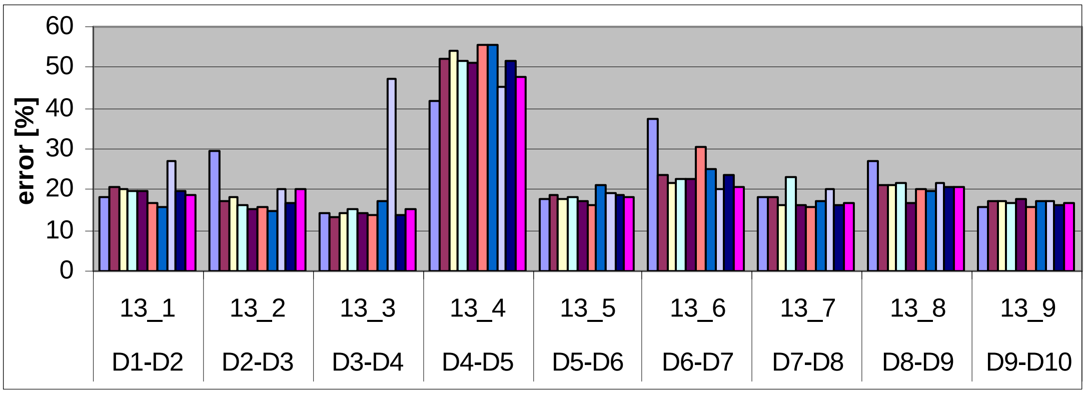



Figure 10: Example for validation results using the best parameter sets of experiment “11” and trying to reproduce the behavior of the drivers in experiment “13”.

For validation purposes the optimal parameter results for the data sets in the first experiment “11” were taken to reproduce the data sets in the other three experiments. In figure 10 the validation errors are shown exemplarily for the reproduction of experiment “13”. Besides singular cases, in which the parameter sets were not transferable with very high errors (“13_4”), the validation error over all data sets mainly ranges from 17 % to 22 %, which is for the singular models about 3.2 to 5.5 percentage points higher than in the calibration cases. The average validation errors of the models range from 19.25 % (SK_STAR) to 20.72 % (IDM). Only the model by Aerde (23.13 %) and the OVM model (22.82 %) showed slightly more problems during the validation.

## BHL loop detector data I-80

For the simulation of the 1km segment on the I-80 the multilane traffic in reality was simulated with an analogous number of parallel single lanes, each simulated independently. Of course, this way the simulation is much simplified, but it has the advantage that complicated influences of lane changing are neglected and a relatively homogenous flow can be simulated.

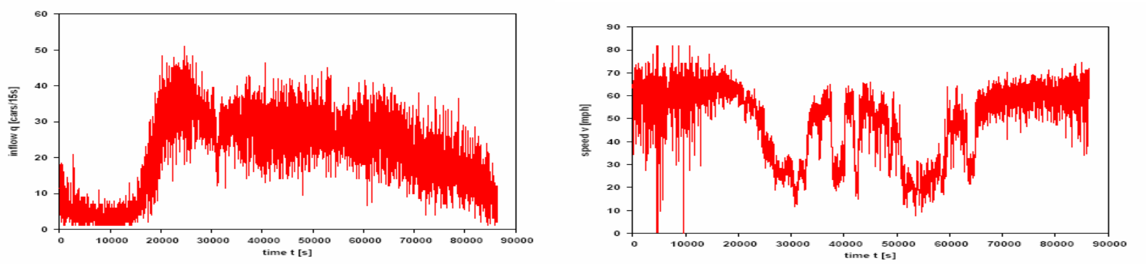



Figure 11: Time series of flow at the inflow link at station 4 (left) and speed at the outflow link at station 6.

During the simulation the vehicles are fed into the model at the time when they were recorded at the inflow station 4 (see figure 11, left) and randomly distributed on the lanes. Two calibrations have been accomplished. The models CT and SK have been calibrated then to fit the time series of speed recorded at the outflow station 6 (see figure 11, right). The errors the models produce are measured at the middle station 5. Thus, the values the models produce at station 5 are compared to the according data. For the CT model the flow was taken as measurement and the results are shown in figure 12. For the SK model the speeds are compared in figure 13. Both simulations produce best results of about 18 % error.

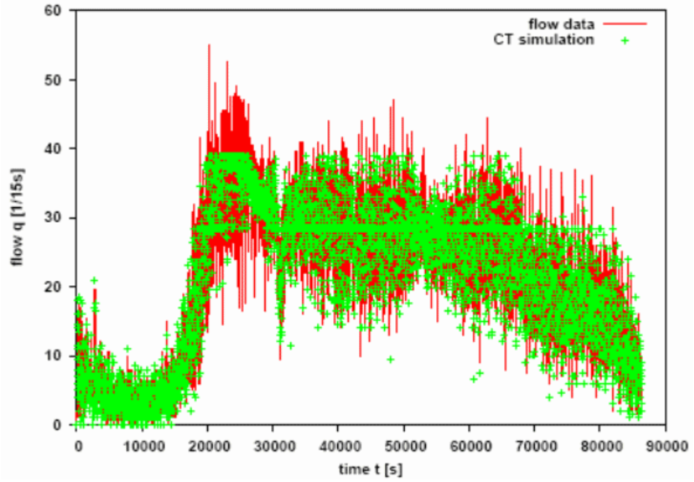



Figure 12: Time series of the flow at station 5. Calibration error is about 18 % for the CT model.

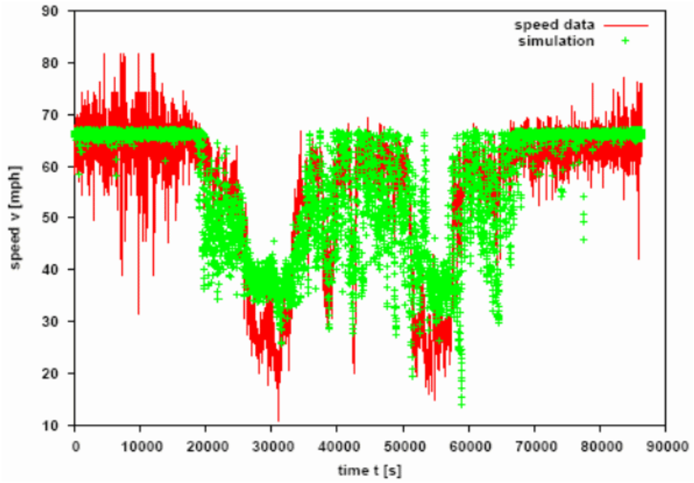



Figure 13: Time series of the speed at station 5. Calibration error is about 18 % for the SK model.

## San Pablo Dam Road

The calibration of models simulating the data on this 6km long road segment have been published in (Brockfeld2002) and (Brockfeld2003). For the simulation of the road some efforts had to be done to do this accurately. At first, the data of observer 5 in the first data set (day 1) and those of observer 6 in the second data set (day 2) 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 figure 14.

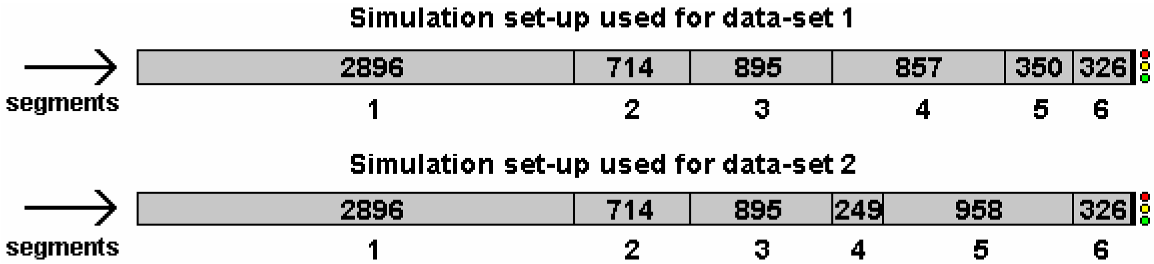



Figure 14: Simulation set up for the 6 km long road segment.

For simulation the vehicles are fed into the model in the first segment according to the recorded data. For the outflow after segment 6 a traffic signal was the crucial thing in reality. Because the exact red/green phases of this signal were not part of the data set and because it was vehicle actuated, a "virtual traffic signal" was implemented and set at the position of the last observer 8. This signal let the vehicles drive out of the system nearby the time recorded in the data of the last observer.

For all models, the simulated travel times per segment were extracted, and the percentage error was calculated per Equation (1). Finally the average error over all segments was built and used for the calibrations.

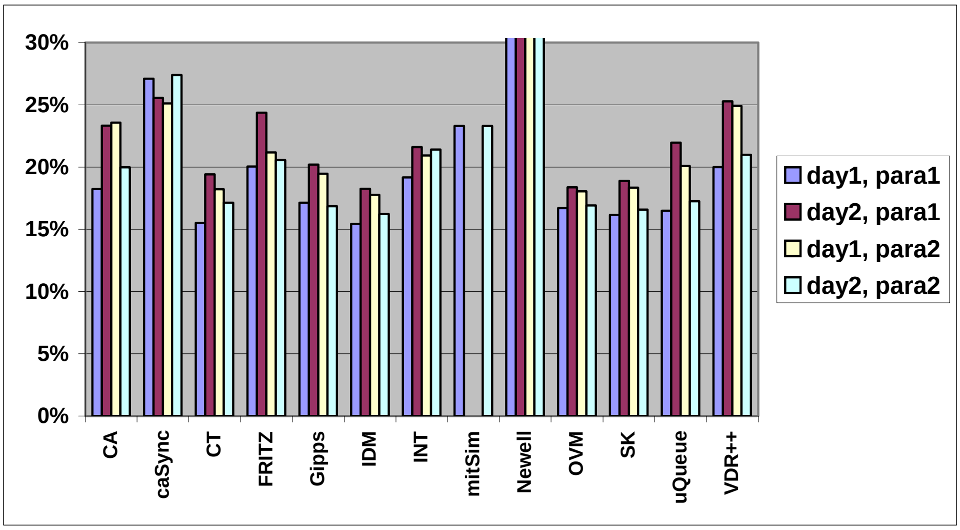



Figure 15: Calibration errors ("day1, para1", "day2, para2") and validation errors ("day2, para1", "day1, para2") for all models.

Figure 15 shows the average calibration and validation errors across the study section for each model. The results are shown for each data set with the best set of parameters. The best models give calibration errors ("day1, para1", "day2, para2") of 15.5-17% (IDM, CT, OVM, SK, queuing model, Gipps). There is a group of models that stay well below 20%

error, and another group of models that are above 20%. The difference between the top six models is hardly significant. Interestingly, the more complicated models seem not to have a better performance.

Since the data set contains two days of data, any model could be calibrated with the data from the first day, and run it with those parameters with the data of the second day (and vice versa). For this validation approach the resulting errors are about 3 % for the better and about 5 % for the worst models.

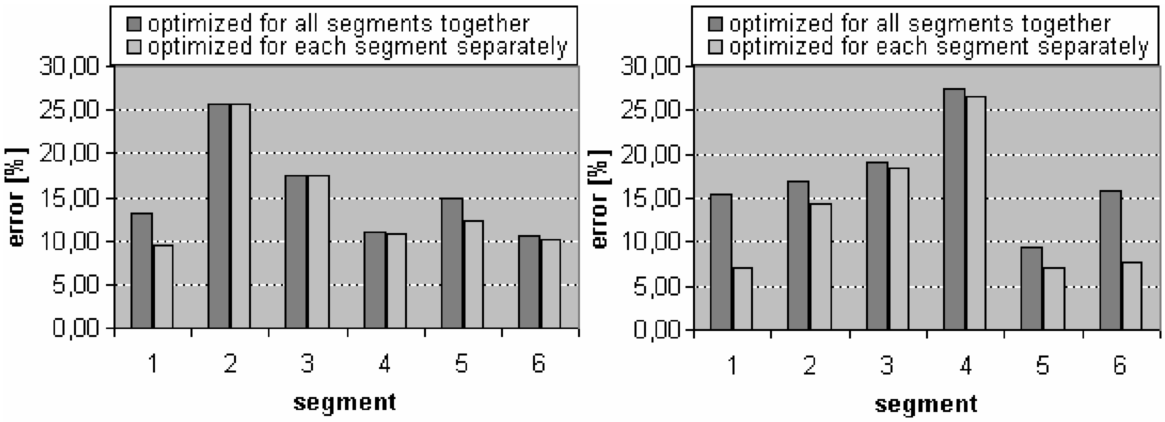


Figure 16: Errors produced by the SK model during the calibration of all segments together and each segment separately (data of day 1 on the left and of day 2 on the right).

For the SK model the errors produced on the particular segments are shown for the best parameter results obtained in figure 16. In comparison to this it is shown that when calibrating (optimizing) the parameters for each segment separately, in parts much better results can be obtained. For example on day 2 the errors of about 15 % on segments 1 and 6 can be lowered to about 7-8 % when calibrating only with the data of these smaller road segments.

## CONCLUSIONS

The error rates obtained by all the itemized approaches draw a similar picture. Taking the speed as an error measurement produces relatively low errors for the approach of Schober using the ICC data and the approach of Ranjitkar using the test track data. Schober obtains errors of about 2.5% in the general case and of 2% to 4% in different speed classes while Ranjitkar gets out errors of 4% to 5%. Taking into account the time lengths of the traces (Schober: one up to a few minutes; Ranjitkar: about 1-2 minutes) and the simulation environments (Schober: mostly multilane highways; Ranjitkar: one lane test track) the results seem to be very similar. Furthermore, the errors of 2% in Schobers approach are only obtained in the high velocity case. To get a percentage error the speed differences are related to the average speed in a data set / trace as defined in equation 1. Because of that the relative error becomes smaller and probably the absolute errors are similar in all the speed cases.

Taking the headway/gap as an error measurement the results of Schober, Ranjitkar and Brockfeld are even more consistently. While for special driving situations (or speed classes) Schober gets errors of about 10%, 13% to 17% and 16% to 17%, Ranjitkar gets in special cases (different driving patterns) errors mostly in the range of 8% to 18%. For more general situations Schober obtains errors of about 18 % to 21 %, which is in good agreement with the results of Brockfeld obtaining mostly errors in the range of 17% to 22%. The over all errors of the models analyzed by Ranjitkar range from 12 % to 13 % for three models and up to 21 for other three models. Because of the less generalized and relatively short sequences, the low errors of 12 or 13 % seem to be results for specific situations.

The other two approaches measuring the travel times between segments on a 6km long road (minimum errors of about 15-16 %) and the speed and flow on a 1km highway (errors of 18 %) stress the error results obtained with the generalized car following approaches of Schober and Brockfeld. As a conclusion for headway estimation errors of about 15 % to 20 % may not be undercut by any approach when time horizons of more than a few minutes are simulated. Of course, for shorter time sequences or special homogenous driving situations one can obtain better results as shown by Schober and Ranjitkar.

For the speed estimation of individual vehicles in the car following process errors of about 2 - 7 % are quite good results and seem to be the minimum reachable level.

Except for the analyses with relatively short trace no big regular differences can be found in the performance of the very different models. Especially the more sophisticated models with a lot of parameters seem not to produce regularly better results than simple ones.

Finally, if one would centralize these results and take them as given reality, the recommendation would be to take the simplest model for a particular application, because complex models likely will not produce better results. The only reason a complex model could be preferred would be, if the user is very familiar with the model and knows the consequences for its behavior whatever a parameter (or set of parameters) is changed. But the results obtained should be confirmed by testing microscopic models with much more different data sets than in this contribution to get a more precise insight what the models are able to describe and which error rates probably have to be accepted.

## ACKNOWLEDGEMENTS

Thanks are to G. S. Gurusinghe, T. Nakatsuji, Y. Azuta, P. Ranjitkar and Y. Tanaboriboon for producing and analyzing the very interesting data sets in Hokkaido and giving us access to these.

**REFERENCES**

Bando, M., Hasebe, K., Nakayama, A. Shibata, A., Sugiyama, Y., 1995. Dynamical model of traffic congestion and numerical simulation, Physical Review E, **51**, pp 1035-1042.
Barlovic, R., Santen, L., Schadschneider, A., Schreckenberg, M., 1998. Metastable States in Cellular Automata for Traffic Flow, Europ. Journ. Phys. B, 793, vol 5.
Brackstone, M., McDonald, M., 1999. Car-following: a historical review, Transportation Research F **2**, pp 181-196.
Brackstone, M., Sultan, B., McDonald, M., 2002. Motorway driver behaviour: studies on car following, Transportation Research F **5**, pp 329-344.
Brockfeld, E., Kühne, R.D., Wagner, P., 2002. Towards Benchmarking Microscopic Traffic Flow Models, In: Möhlenbrink, W; Bargende, M.; Hangleiter, U.; Martin, U., FOVUS: Networks for Mobility 2002, International Symposium 18.-20. September 2002, University of Stuttgart, Band 1, S. 321-331.
Brockfeld, E., Kühne, R.D., Skabardonis, A., Wagner, P., 2003. Towards a benchmarking of Microscopic Traffic Flow Models TRB 2003 Annual Meeting (CD- ROM), TRB2003-001164. To be published in TRR 2003.
Brockfeld, E., Wagner, P., Kühne, R.D., 2004. Calibration and Validation of Microscopic Traffic Flow Models, TRB 2004 Annual Meeting (CD- ROM), TRB2004-001743.
Homepage of the Berkeley Highway Laboratory (downloadable from http://www.its.berkeley.edu/bhl/).
Castello, J.M.D., 1994. A car-following model based on the Lighthill-Whitham theory, in: Lesort J.B. (Ed.), Transportation and Traffic Flow Theory, 13 ISTTT, Pergamon, New York.
Chowdhury, D., Santen, L., Schadschneider, A., 2000. Statistical Physics of Vehicular Traffic and Some Related Systems, Physics Reports **329**, 199 (2000), (downloadable from http://arxiv.org/abs/cond-mat/0070053).
Crowther, B.C., 2001. A Comparison of CORSIM and INTEGRATION for the Modeling of Stationary Bottlenecks, Master thesis, Virginia Polytechnic Institute and State University.
Daganzo, C.F., 1994. The Cell transmission model: a simple dynamical representation of highway traffic, Transp. Res. B **28**, pp 269 - 287.
Daganzo, C.F., 1997. Homepage, (downloadable from http://www.ce.berkeley.edu/~daganzo/).
Gipps, P.G., 1981. A behavioral car following model for computer simulation, Transp. Res. B **15**, pp 105-111.
Fritzsche, H.T., 1994. A model for traffic simulation, Transp. Engin. Contr. **5**, pp 317 – 321.
Gurusinghe, G.S., Nakatsuji, T., Tanaboriboon, Takahashi, K., Suzuki, J., 2001. A Car Following Model Incorporating Excess Critical Speed Concept, Journal of the Eastern Asia Society for Transportation Studies, Vol.4, No.3, pp 171-184.
Gurusinghe, G.S., Nakatsuji, T., Azuta, Y., Ranjitkar, P., Tanaboriboon, Y., 2003. Multiple Car Following Data Using Real Time Kinematic Global Positioning System, TRB 2004 Annual Meeting (CD- ROM), TRB2003-004137.
Helbing, D., 2001. Traffic and Related Self-Driven Many-Particle Systems, Reviews of Modern Physics **73**(4), 1067-1141 (2001). [See also at http://arxiv.org/abs/cond-mat/0012229]
Kazi Iftekhar Ahmed, 1999. Modeling Drivers' Acceleration and Lane-Changing Behavior, Ph.D. thesis, MIT.

Kerner, S., Klenov, S.L., Wolf, D.E., 2002. Cellular Automata approach to the three phase theory of traffic flow, (downloadable from http://arxiv.org/abs/cond-mat/0206370).

Knospe, W., Santen, L., Schadschneider, A., Schreckenberg, M., 2000. Towards a realistic microscopic description of highway traffic, J. Phys. A **33**, L477.

Krauss, S., Wagner, P., Gawron, C., 1997. Metastable states in a microscopic model of traffic flow, Phys. Rev. E **55**, pp 5597-5605.

Michalewicz, Z., 1992. Genetic Algorithms + Data Structures = Evolution Programs, Springer.

Nagel, K., Schreckenberg, M., 1992. J. Physique I, **2**, 2221,.

Newell, G., 1962. Theories of instabilities in dense highway traffic, J. Op. Soc. Japan **5**, pp 9-54.

Newell, G., 2002. A simplified car-following theory – a lower order model, Transportation Research B **36**, pp 195 – 205.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C, chapter 10, Cambridge University Press and references therein.

Ranjitkar, P., Nakatsuji, T., Asano, M., 2004. Performance Evaluation of Microscopic Traffic Flow Models Using Test Track Data, TRB 2004 Annual Meeting (CD- ROM), TRB2004-002272.

Schober, M.: diploma thesis, University of Stuttgart, 2003.

Schober, M., Wagner, P., Kühne, R.D., Eskadarian, A., Wehlan, H., 2004. Calibration of Microscopic Car-Following Models on the Basis of the Intelligent Cruise Control Field Operational Test Data, ITS Safety and Security Conference, Miami, Florida.

Smilowitz, K.R., Daganzo, C.F., 2000. Experimental Verification of Time-Dependent Accumulation Predictions in Congested Traffic, Transportation Research Record **1710**, pp 85-95.

Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observation und numerical simulations, Physical Review E, **62**, pp 1805-1824.

Treiber, M., Helbing, D., 2003. Memory effects in microscopic traffic flow models and wide scattering in flow-density data (downloadable from http://arxiv.org/abs/cond-mat/0304337).

P. Wagner, P., Lubashevsky, I., 2003. Empirical basis for car-following theory development (downloadable from http://arxiv.org/abs/cond-mat/0311192).