Calibration and Validation of Microscopic Traffic Flow Models

Elmar Brockfeld[1], Peter Wagner[1]

[1] Institute of Transport Research
German Aerospace Center
Rutherfordstr. 2 – 10245 Berlin – Germany
Phone: +49 30 67055 231, +49 30 67055 237, Fax: +49 30 67055 202
E-mail: Elmar.Brockfeld@dlr.de , Peter.Wagner@dlr.de

Summary. The aim of this paper is to present recent progress in calibrating ten microscopic traffic flow models. The models have been tested using data collected via DGPS-equipped cars (Differential Global Positioning System) on a test track in Japan. To calibrate the models, the data of a leading car are fed into the model under consideration and the model is used to compute the headway time series of the following car. The deviations between the measured and the simulated headways are then used to calibrate and validate the models. The calibration results agree with earlier studies as there are errors of 12 % to 17 % for all models and no model can be denoted to be the best. The differences between individual drivers are larger than the differences between different models. The validation process leads to errors from 17 % to 22 %. But for special data sets with validation errors up to 60 % the calibration process has reached what is known as “overfitting”: because of the adaptation to a particular situation, the models are not capable of generalizing to other situations.
Keywords: microscopic models, car following, DGPS, calibration, validation

1 Introduction

Microscopic simulation models are becoming increasingly important tools in modeling transport systems. There is a large number of available models used in many countries. The most 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, lane changing and gap acceptance models. This difficulty is due to the lack of suitable methods for adapting models to empirical data. The aim of this paper is to present recent progress in calibrating a number of microscopic traffic flow models. By calibrating and validating various models using the same data sets, the models are directly comparable to each other. This sets the basis for a transparent benchmarking of those models. Furthermore, the advantages and disadvantages of each model can be analyzed better to develop a more realistic behavior of the simulated vehicles.

In this work ten microscopic traffic flow models were tested from a very microscopic point of view concerning the car-following behavior and gap-acceptance. This is in contrast to a typical macroscopic analysis which compares aggregated data on links for example. The data used for calibration and validation is from car-following experiments conducted in Japan in

October 2001 [1]. The data have been collected by letting nine DGPS-equipped cars follow a lead car driving along a 3 km test track for about 15-30 minutes.

At first the experiments on the test track and the recorded data sets are briefly described and the simulation setup for testing the models is defined. In the following the measurement procedure for calculating the error differences between the recorded data and the data produced by the models is specified. After the tested models are listed and basically described, the calibration and validation results are presented leading to some conclusions.

2 Data and error measurement

2.1 The data and the simulation set-up

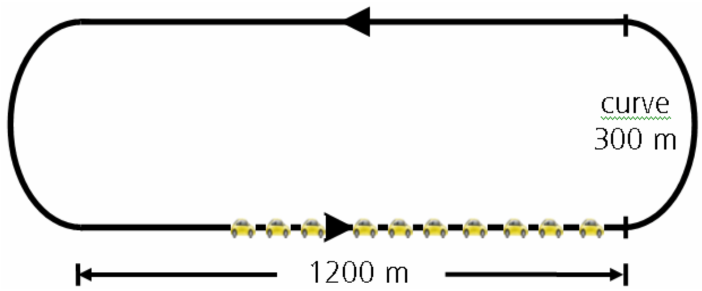

Fig. 1 Sketch of the test track with ten cars driving on the course.

The data sets have been recorded on a test track in Hokkaido, Japan in October 2001 [1]. 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 regulary increase/decrease of speed is done with different frequencies, the velocity cycles from 20 to 60 km/h being performed one to four times on the straight segments.

To minimize driver-dependent correlations between the data sets, the drivers were exchanged between the cars after each experiment. Having all cars equipped with DGPS (Differential Global Positioning System), the position of each car is stored in 0.1 second intervals throughout each experiment. From these position 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 [1]. 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.

In this paper we present analyses concerning four of the eight experiments, namely the patterns mostly with intervals of constant speeds and waveperforming. For the simulation set-up only two cars are considered at a time: the leading car is updated according to the speeds and positions in the recorded data sets and the following car is updated as defined by the equations of the used model.

2.2 Error measurement

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 of gaps. To get a percentage error it is additionally related to the average value of the time series in each particular data set:

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

where $x^{(sim)}$ and $x^{(obs)}$ are a simulated and an observed traffic flow variable, which is in this case the gap between two cars. T is the time series over the total time of each experiment.

3 The models

The models used for the simulations are all 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 values 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 the calibration approach the following microscopic traffic flow models of very different kind with 3 to 15 parameters have been tested. Some models are used in commercial simulation programs, which are popular in European countries, the USA and Japan, and some are scientific simulation approaches.

Abbreviation	**Description**	**params**
CA0.1	cellular automaton model [2]	4
SK_STAR	model based on the SK-model by S. Krauss [3]	7
OVM	“Optimal Velocity Model”, Bando, Hasebe [4]	4
IDM	“Intelligent Driver Model” [5]	7
IDMM	“Intelligent Driver Model with Memory” [6]	7
Newell	can be understood as a continuous CA with more variable acceleration and deceleration [7,8]	7
GIPPSLIKE	basic model by P.G. Gipps [9]	6
Aerde	used in the simulation package INTEGRATION [10]	6
FRITZSCHE	used in the British software PARAMICS; similar to what is used in the German software VISSIM [11]	13
MitSim	model by Yang and Koutsopulus, used in the software MitSim [12]	15

Table 1 List of tested models

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, SK_STAR and MitSim) use some kind of stochastic parameters describing individual driver behavior. Most models use something like a reaction time of the drivers to the behavior of the leading car. The MitSim and the FRITZSCHE model have got a lot of more parameters defining thresholds concerning the headway and the speed difference to a leading vehicle. Depending on these various driving behaviors are realized like "free driving", "approaching" and "emergency braking" for example.

As the time step for the models is 0.1 seconds according to 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.

4 Calibration and validation

Altogether 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. To find the optimal parameter constellations a gradient-free optimization method was used [13] and started many times with different initialization values for each "model - data set" pair. This variation 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 optimizations.

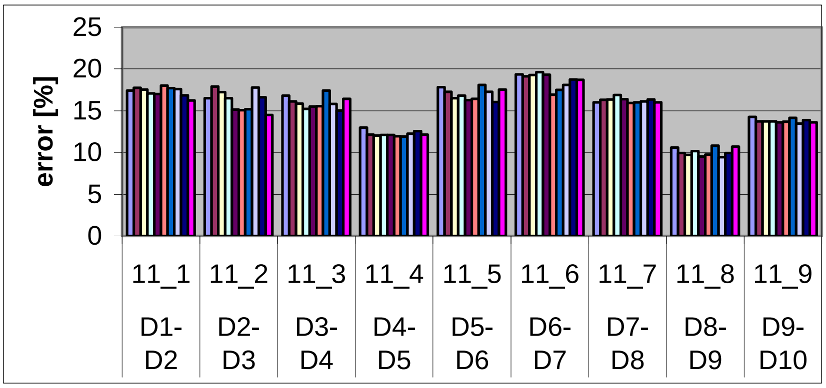

Fig. 4 Some calibration results obtained for one of the four experiments.

As an example figure 4 shows 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. It is noteworthy 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 in figure 4, too.

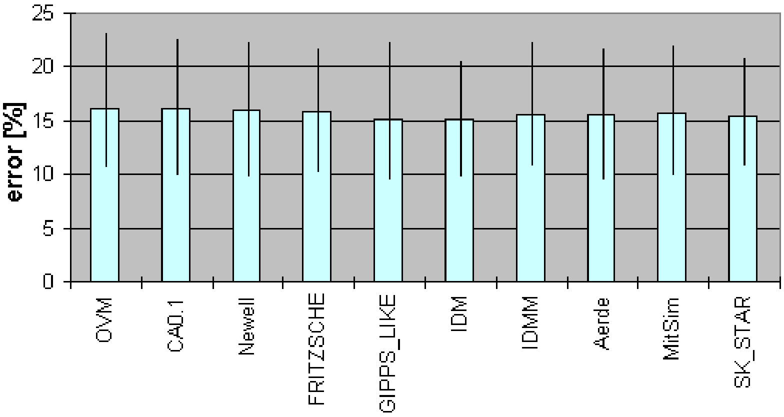

Fig. 5 Mean calibration results for all models including the total result range.

Looking at the average errors each model produces with the 36 data sets, it can be seen in figure 5, 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.

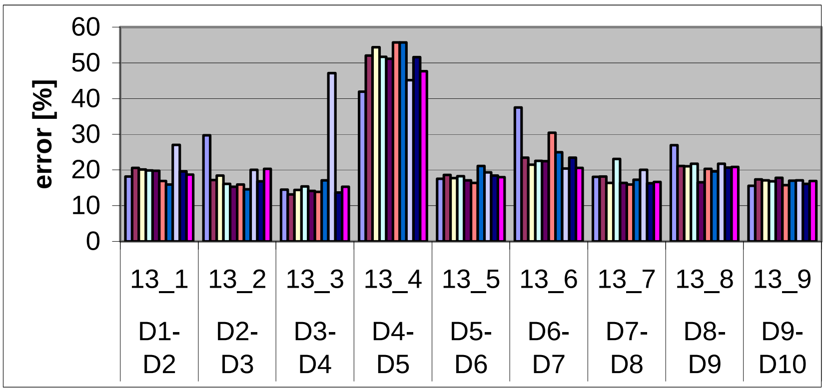

Fig. 6 Validation results using the best parameter sets of one experiment “11” and trying to reproduce the behavior of the drivers in experiment “13”.

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 6 the validation errors are shown exemplarily for the reproduction of experiment “13”. Except for some cases, where the parameter sets were not transferable due to 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.

5 Conclusions

The error rates of the models in comparison to the data sets during the calibration for each model reach from 9 % to 24 %. Surprisingly, no model appears to be significantly better than any other model and the average error rates of the models are very close to each other between 15.1 % and 16.2 %. All models share the same problems with certain data sets while other data sets can be reproduced quite well with each model. Interestingly, it can be stated that models with more parameters do not necessarily reproduce the real data better. The results of the validation process give a similar picture. The additional errors in comparison to the calibration are – apart from singular cases of "overfitting" - mainly in the area of 3 to 5 percentage points.

The results after the calibration and the validation agree with results that have been obtained before with a completely different data set taking the travel times on road segments instead of headways for the error measurement [14]. In these studies about 15 % to 27 % were found to be the minimum calibration error and additional validation-errors were found to be about 2 to 5 percentage points. It was found, too, that out of about ten models the differences are not as big as could be expected. However, the results of the validation show, that when calibrating and validating with special data sets, the parameters of a model can be "overfitted" and thus the results can be very unsatisfactory with surprisingly high errors. The calibration tends to optimize the model for a given data-set, thereby sacrificing generality.

There are two conclusions that can be drawn. First, one should call for the development of better models. Additionally, one should think about a different calibration technique which avoids "overfitting" and could produce results which stay more general. The other way to interpret the results is that – from this microscopic point of view – errors of about 15-25 % can probably not be suppressed no matter what model is used. These are due to a really stochastic component in the driver's behavior.

References

1. G. S. Gurusinghe, T. Nakatsuji, Y. Azuta, P. Ranjitkar and Y. Tanaboriboon, Multiple Car Following Data Using Real Time Kinematic Global Positioning System, Transport Research Record 1802, 166-180 (2003).
2. K. Nagel and M. Schreckenberg, J. Physique I, 2221, 2, 1992.
3. S. Krauss, P. Wagner, and C. Gawron, Phys. Rev. E 55, 5597 – 5605 (1997).
4. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, Y. Sugiyama, Physical Review E, 239-253, 51, 1999.
5. M Treiber, A Hennecke and D Helbing, "Congested traffic states in empirical observations and microscopic simulations", Physical Review E 62, 1805-1824 (2000).
6. M. Treiber, D. Helbing, "Memory effects in microscopic traffic models and wide scattering in flow-density data", Physical Review E 62, 046119 (2003). (see also http://arxiv.org/abs/cond-mat/0304337)
7. G. Newell, J. Op. Soc. Japan 5, 9 – 54 (1962).
8. G. Newell, Transportation Research B 36, 195 – 205 (2002).
9. P.G. Gipps, Transp. Res. B 15, 105 – 111 (1981).
10. Brent C. Crowther, "A Comparison of CORSIM and INTEGRATION …", Master thesis, Virginia Polytechnic Institute and State University (2001).
11. H.-T. Fritzsche, Transp. Engin. Contr. 5, 317 – 321 (1994).
12. Kazi Iftekhar Ahmed, Ph.D. thesis, MIT (1999).

13. W.H.Press, S.A.Teukolsky, W.T.Vetterling and B.P.Flannery, Numerical Recipes in C, chapter 10, Cambridge University Press and references therein (2002).
14. Brockfeld, E., Kühne, R.D., Skabardonis, A., Wagner, P., 2003. Towards a benchmarking of Microscopic Traffic Flow Models TRR No. 1852 Traffic Flow Theory and Highway Capacity 2003, Washington, D.C. (2003).