Challenges in applying calibration methods to traffic models

July 15, 2015

Peter Wagner*
German Aerospace Center (DLR)
Institute of Transportation Systems
Rutherfordstrasse 2, 12489 Berlin, Germany
phone: +49 30 67055 - 237, fax: +49 30 67055 - 291
email: peter.wagner@dlr.de, web: http://www.dlr.de/ts/en/

Christine Buisson
Licit/Ifsttar-Entpe
Lyon, France
phone: +33 978 934 - 2289, fax: +33 978 934 - 3052
email: Christine.BUISSON@entpe.fr

Ronald Nippold
German Aerospace Center (DLR)
Institute of Transportation Systems
Rutherfordstrasse 2, 12489 Berlin, Germany
phone: +49 30 67055 - 263 , fax: +49 30 67055 - 291
email: ronald.nippold@dlr.de, web: http://www.dlr.de/ts/en/

For Presentation and Publication
95th Annual Meeting
Transportation Research Board
January 12 – 16, 2016
Washington, D. C.

Submission date: July 15, 2015

words:	3 732
plus 6 figures:	1 500
plus 1 table:	250
total count:	5 482
word limit:	7 500

*corresponding author

ABSTRACT

This text looks at calibration and validation as a means to understand traffic flow models better. It concentrates on the car-following part of it and demonstrates that the calibration of stochastic models under certain circumstances can become very difficult. Supported by a few synthetic examples [15] it hypothesizes that the stochasticity is the reason why calibration / validation is so far not very successful at discerning good from bad models of traffic flow.

1 Introduction

A lot of work has been conducted recently to improve on the usage of microscopic traffic flow models and especially their calibration and validation to real world data, for an overview see [5]. This has worked with an astonishingly degree of quality, i.e. traffic flow models can be calibrated to real data with an r.m.s.-error of the order of around 10%.

While good enough for daily use, it may have a weak spot, and that is its unknown quality when it comes to extending such a calibrated model to situations and data it has not yet encountered. In order to be better off in such situations, calibration / validation should be used for another purpose: to help in a better understanding of these models by showing where they fall short.

This is especially important since we are entering an area with a mix of autonomous (of various degrees, even intelligent cruise control resembles an autonomous car) and human-driven vehicles, and this interaction is only poorly understood right now. Since the controler of an autonomous vehicle can be modeled with ease, the human drivers are more challenging modeling endeavor. And it is important there to have the correct model, since so far we do not have good data that can be used to calibrate models that deal with this interaction.

Interestingly, as has already been mentioned in [4], performing a good calibration almost eliminates the differences between the models in terms of their ability to describe real-world phenomena. This result seems to be fairly robust with respect to different data, different scenarios, different calibration methodologies, and different objective functions as all the approaches where microscopic models have been calibrated show. This is surprising, and it asks for an explanation.

The present texts makes the hypothesis that this is due to a false treatment of stochastic models. It demonstrates by way of a few simple synthetic examples[14] of car-following, that under not too exotic conditions the parameter estimation of the calibration process can yield results that have nothing to do with the real parameters present, while still yielding a reasonable fit. So far, no remedy is known for work around such an result.

2 Stochasticity in data

When models face reality, a number of issues have to be considered. All data contain experimental noise. To exemplify this, two examples are considered here, data from the French project MoCoPo which are similar to data from the NGSIM project [2], and data from a large German project named simTD (Safe and Intelligent Mobility Test Field Germany [1]) which aimed at a better understanding of communication in traffic and sported a fleet of equipped vehicles that drove around in the Frankfurt region for several months.

2.1 MoCoPo

In ... data from an on-ramp in Grenoble have been collected via video from a helicopter.

2.2 simTD

The data in this project have been sampled from August to December 2012 (on 97 days) from a fleet of 125 vehicles that were driven by a few hundred different drivers. Each vehicle was equipped with at least an acceleration sensor and a speed sensor, most of the vehicles also had equipment to monitor the distance and speed-difference to a lead vehicle. In addition, they had communication devices on board and monitored their position via GPS, and vast amounts of data from their CAN bus. All the data, including the communication protocol and many more had been recorded by computers in the vehicles and subsequently transferred to a common data-base. Note however, that especially the distance and speed-differences have been recorded by the equipment that is also in use for the driver assistance system in those cars. It was not a special equipment designed for scientific experiments.

From this massive data-base (which contains in zipped format 1.3 TByte of data) a few examples have been picked to be presented here. E.g., in Figure 1 the raw data from the acceleration sensor is displayed. Since a vehicle is a heavy object, its acceleration versus time curve should be fairly smooth. However, as can be seen from the raw data, there is a considerable amount of noise even in the acceleration raw data (gray points in Figure 1).

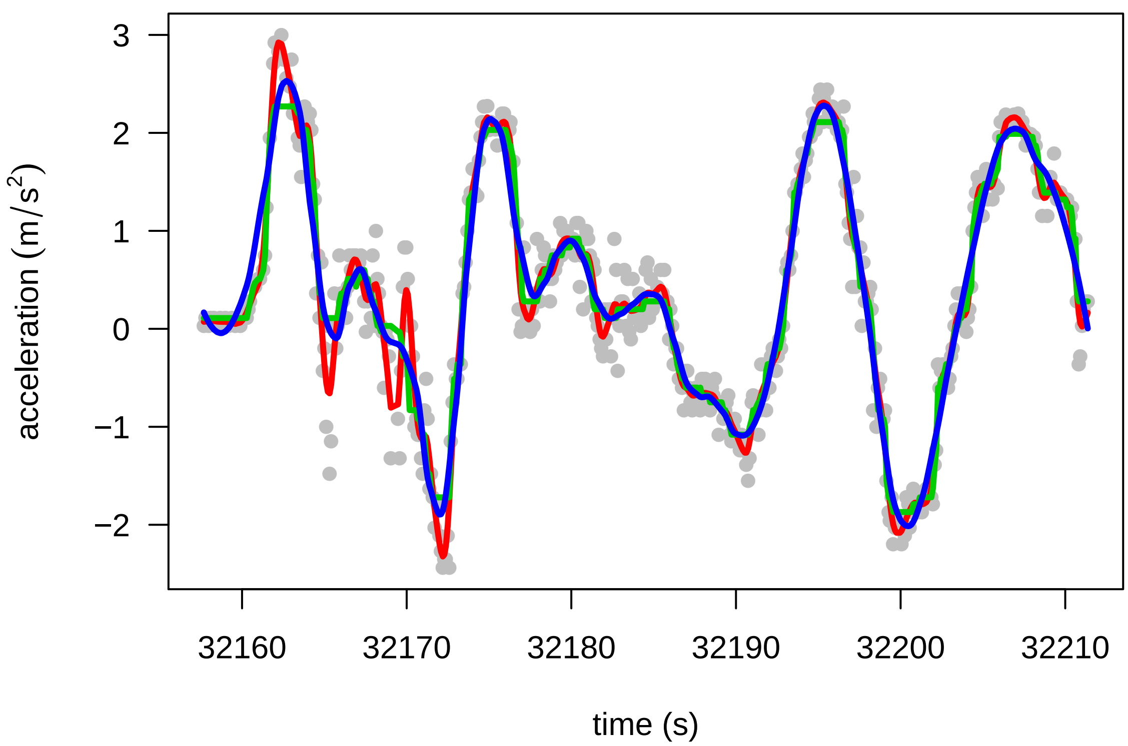

FIGURE 1 A short piece of acceleration as function of time for the vehicle 501 from the simTD database on 27.8.2012. The gray points are the raw data. The three curves represents different smoothing methods: the red curve is a spline smoothing, the green one from a median filter, and the blue curve is local polynomial regression of order two to the data. All these methods are implemented in R [16].

In Figure 1, different methods for smoothing the acceleration data from this data-base have been compared. Although they yield smoothed approximations to the raw data, they have the disadvantage of many smoothing method: they make assumptions about the underlying true process, and if one of these assumptions is wrong, they fail. Nevertheless, from the different smoothed curves it can be concluded that the empirical noise in these data is between 0.1 and 0.3 m/s^2. These numbers have been obtained by computing the root-mean-square (rms) distance between the raw data and the smoothed approximations. The 0.1 is for the spline interpolation, which follows the data quite closely, while the 0.3 is for the local polynomial regression, which in the eyes of these authors is the more reasonable choice.

3 Stochastic components in models

In addition to the noise from empirical data, the models with which traffic flow is described microscopically, can also have stochastic components. The noise can either be in the dynamic itself, or it is in the parameters: different drivers have different sets of parameters to describe their driving style. This is called driver heterogeneity. Even more confusingly, the parameters of one and the same driver may be subject to temporal changes, even during a short time-span.

Most of the models that have been described so far, however, are deterministic models. To clarify what that means, the following model (which is modeled after [3]) will be considered: using$\dot{v}$ as the acceleration a of the subject vehicle with speed v and g as the net headway to the vehicle in front, it's basic version reads:

$$\dot{v} = a = B(\omega^2(g - g^*(v, \Delta v))). \tag{1}$$

Here, the function $B()$ limits the acceleration to fall in the interval $[-\beta, \alpha(v)]$, where $\alpha(v)$ describes how the maximum acceleration decreases with speed. For this, any model might be acceptable, to be specific $\alpha(v) = \gamma(v_{\max} - v)$ is often used, thereby introducing the parameters $v_{\max}$ and γ. See also Figure 2 for an example how the acceleration values in real data are distributed as function of speed.

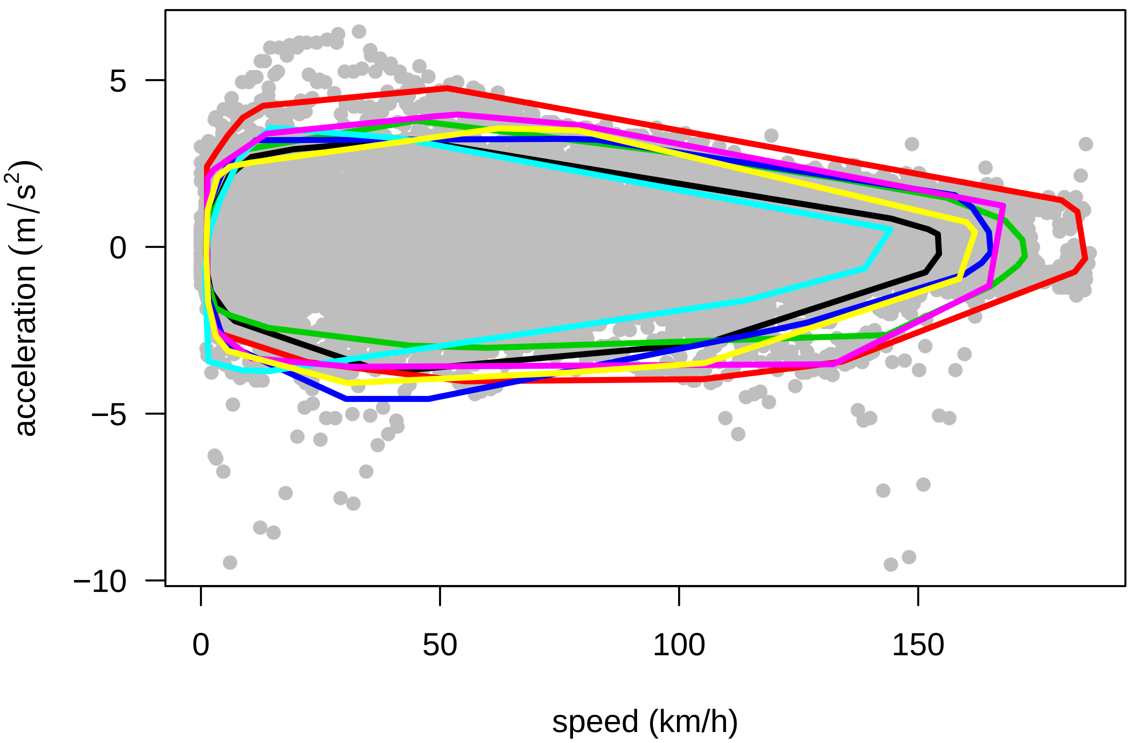

FIGURE 2 The convex hull of the acceleration and speed values of seven different drivers (colored lines) together with the raw values (gray points) of the "red" driver. The data were sampled on one day, 22. Oct. 2012. To compute the convex hull, data-points which were hit less than 5 times have been omitted. This demonstrates, that drivers occupy on average very similar regions in this (v,a)–space.

The preferred distance g^* depends on the preferred time headway T and, in addition, on the speed difference $\Delta v = V - v$ between the lead vehicle's speed V and the following vehicles speed:

$$g^*(v, \Delta v) = v\left(T + \frac{\Delta v}{b}\right) \tag{2}$$

In this equation, the parameter b is some preferred deceleration the driver typically wants to apply in a normal car-following situation. This is well different from the parameter β that limits the maximum deceleration either to the physically possible one, or to the maximum deceleration the driver applies even in a critical situation (which is typically smaller than the physical boundary of the vehicle itself).

This model has a number of more or less obvious relatives, like the Helly model [8], the Newell model [13], a cellular automaton model [12], or even a kind of brute force linearization of the Gipps and Krauß models [7, 11]. It also shares some similarity with the IDM.

The numbers γ, ω are constants (for each driver!) to make the units in the equation correct, in addition, they are inverse relaxation times. So, each driver is described by the set of six parameters $v_{\max}$, b, T, β, γ, ω. Note, that these parameters are in principle directly measurable, without getting them from a calibration process. (Clearly, the result from such a direct estimation may differ from the results of a calibration.) The acceleration bounds can be read off a time-series $a(t)$ or from Figure 2, the parameters b and ω from a plot of acceleration versus speed difference or distance, respectively, and the preferred headway T from a fit of g versus speed v. In addition to that, they can also be estimated by fitting such a model to real data that came, e.g. from car-following episodes.

As mentioned already, it is assumed here that each driver has its own set of parameters, and these parameters may change depending on external influences (weather, mood, level of stress etc.). However, as long as the changes are slow compared with typical time-scales in the model, they can be considered as constant and do not interfere with the dynamic.

This is still a deterministic model, and it is formulated as a differential equation. This means that the driver is applying her control in each instant of time. Adding noise to such a model leads to a stochastic differential equation:

$$\dot{v} = a = B(\omega^2(g - g^*(v, \Delta v))) + \sigma\xi. \tag{3}$$

The size of the noise σ is of course another parameter, while ξ stands for the noise term itself. It is important to note that the noise term should be bounded (it cannot be normally distributed) and that it must have a memory, which means that acceleration cannot change in 1 ms or so, but changes slowly, which is true for vehicles with masses well above 1000 kg. A noise term with such a memory is named colored noise.

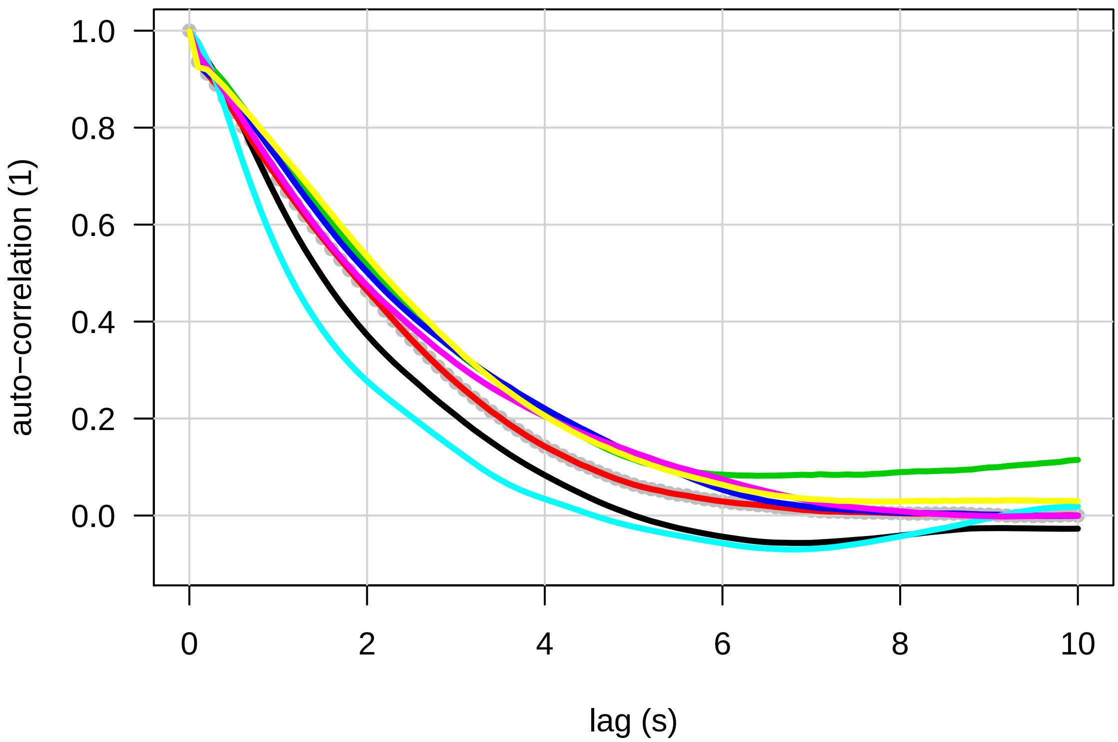

FIGURE 3 Auto-correlation function of the acceleration of seven vehicles, again taken from the simTD data.

This memory of the acceleration is an empirical feature. Real acceleration time-series have an auto-correlation function that drops from 1 for time lag 0 to $1/e$ for a time lag between two and four seconds. This can be seen in Figure 3, where the same data as in Figure 2 have been used to compute the auto-correlation function of the seven vehicles picked from the simTD data-base, also from 27.8.2012. The auto-correlation function is computed by

$$c(\tau) = \frac{1}{(n-\tau)\sigma_a^2} \sum_{0}^{n-\tau} \hat{a}(t)\, \hat{a}(t+\tau)$$

where the variable $\hat{a}(t)$ is the acceleration from which the mean value is subtracted, while σ_a is the standard deviation. Again, this is computed using R [16]. Using pure white noise gives an acceleration time series with zero correlation time.

A different model may be obtained by assuming that a driver does not react permanently, but only from time to time. These so called action-point models [17] have discrete points in time where acceleration (more precisely: driver's control of it) changes quickly to a new value that might be based on equation (1). In this case, the acceleration noise is added only at these action-points and by assuming that a driver it not very good at setting the acceleration based on equation (1) but adds an error to it. In this case, the action-point mechanism introduces the memory in the acceleration, since acceleration changes only little or not at all between two subsequent action-points:

$$a_k = B(\omega^2(g(t) - g^*(v(t), \Delta v(t)))) + \sigma\xi(t) \quad t = t_k \tag{4}$$

$$x(t) = x(t_k) + v(t_k)(t - t_k) + \frac{1}{2} a_k (t - t_k)^2 \quad t \in [t_k, t_{k+1}[\tag{5}$$

Recently, a new class of models has been introduced, that assume that the noise is not simply in the acceleration (it may be there, in addition), but is in one or all of the parameters describing the driver [10]. The main culprit here is the preferred headway T, but other parameters might do as well. From empirical data it is well-known, that the headway distance in real traffic is a very volatile variable in a wide range of numbers, typically covering a range between 0.5 and 2 times the mean value [18]. This is quite different from the fluctuations in speed difference, which are typically around a few m/s compared to speeds of 20...40 m/s.

So, these models assume, that the parameter T changes due to a stochastic process. The consequences of such a noise mechanism will be detailed in section 4.

4 The parameter estimation

To fix ideas, the model in equation(1) is used without bounding the values of acceleration and speed, i.e. $B(\cdot)$ is the identity function (1) and is thus written:

$$\dot{v} = \omega^2 \left(g - v \left(T + \frac{\Delta v}{b} \right) \right). \tag{6}$$

This reduces the number of parameters to three; the omitted three parameters are parameters that limit the dynamics, while the remaining three parameters describe the interaction between the vehicles. Another advantage of this reduction is that the model can be rewritten as a model that is linear in the three parameters p_i (and weakly non-linear in the dynamics itself):

$$\dot{v} = p_1 g + p_2 v + p_3 v \Delta v \tag{7}$$

A simple scenario is presented in the following. A vehicle that does not follow this model but drives its own stochastic course will be used as leading vehicle to this model. Note, that time is discretized in chunks of 0.1 s, which is a common empirical resolution. The lead vehicle's trajectory is created by drawing acceleration values A from a Laplace distribution $p(A) \propto exp(-|A|/a_0)$, which will be changed at random points in time whose distance is drawn randomly from the interval [0.5,2] s. This yields an acceleration trajectory $A(t)$ of the lead vehicle, from which the speed $V(t)$ of the lead vehicle is computed. In addition, it is made sure that the resulting $V(t)$ remains within two bounds $[V_1, V_2]$, an example is displayed in Figure 4. Such a trajectory looks similar to real-world speed functions, and it is important that it is not a simple function. If it were simple and mostly constant, the following vehicle's behavior in the case of a deterministic model would be very simple, too, and markedly different from all empirical data we have analysed so far, which always appear very noisy. Note, however, that this noisy look of this curve is only due to the random jumps at the action-points, in between the trajectory follows a simple linear function $V(t) = V_i + a_i(t - t_i) \quad t \in [t_i, t_{i+1}[$.

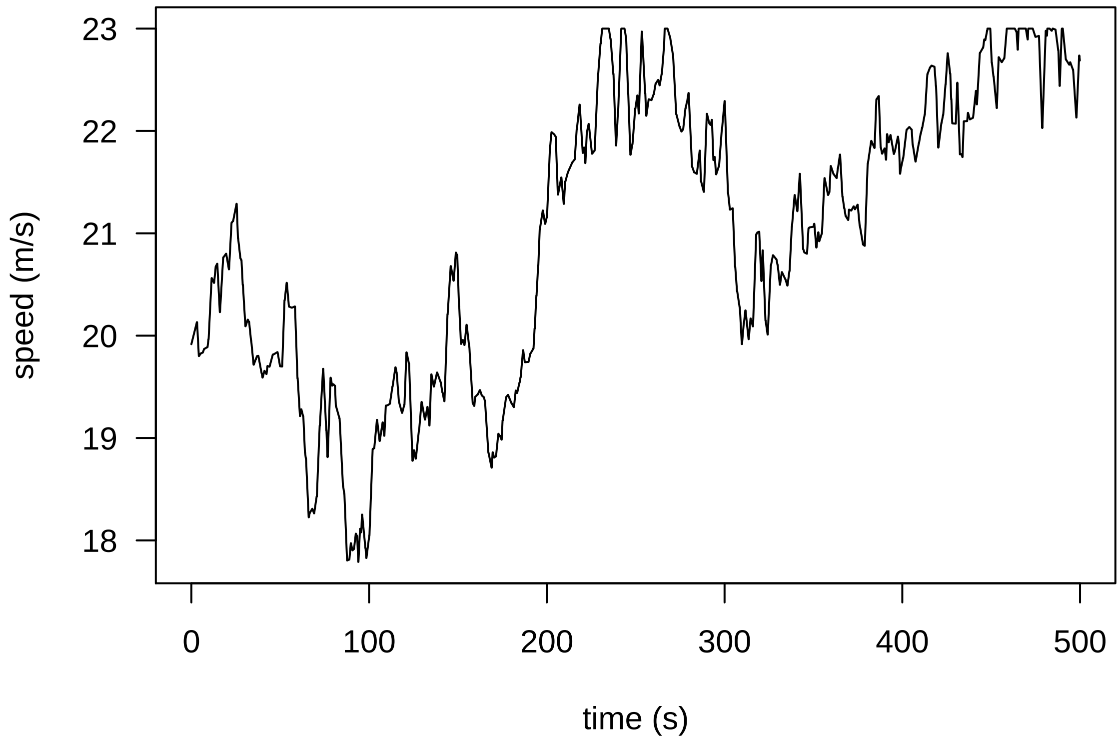

FIGURE 4 Speed versus time of a small piece of the noisy synthetic trajectory of the lead vehicle. The trajectory is bounded between 18 and 23 m/s.

The subject vehicle follows this trajectory with a certain set of parameters, and this subject vehicle is described by the model equation (7) and endowed with different noise mechanisms. In the following, the triple $p = (0.0169, -0.0239, 0.0172)$ will be used, which has been obtained from a real-word data-set by the calibration method specified below. Note, that this set corresponds to $T = 1.41$ s and $b = 0.98$ m/s^2 which seems a realistic choice of parameters. Running the model equation (7) with a trajectory generated for the lead vehicle gives a certain simulated trajectory $(g(t), v(t), a(t))$ of the following vehicle.

By fitting this trajectory to the model equation (7) as a simple (robust) linear fit [16], the parameters can be reproduced with a small error which can be found in the column labeled "rms(acc,fit)" in table 1. This remains also true if the fitted parameters are used to run the model once more and then compare the acceleration gotten by this with the acceleration generated during the first run of the model. The corresponding values can be found in the column of table 1 labeled by "rms(acc,sim)". All these results are in the second row of table 1 which is labeled "raw".

So, when the model equation (7) is following a noisy lead vehicle with speed $V(t)$ it is possible to find from the time-series the parameters that have gone into the model.

This was the case for a deterministic model which was driven by a stochastic lead vehicle. It becomes way more difficult when the stochastic variants of the model are being used. In total, the 4 different stochastic models have been used:

Model-1: Adding a white-noise term $\sigma\xi$ to equation (7), which could be named the physicist's approach since it is lend from the modeling of the Brownian motion.

Model-1a: Just like model 1, but instead of white noise coloured noise have been added to the acceleration time-series [6], [9]. Coloured noise can be understood as an exponentially smoothed white noise process, the simplest approach that has been used here is $n(t+\Delta t) = \alpha n(t) + (1-\alpha)\sigma\xi(t)$.

Model-2: An action-point type algorithm in the acceleration, which is given by equations (4) and (5).

Model-3: In the so called 2D models[10, 18], a parameter could be changed randomly instead of fiddling around with the dynamics. The method here uses a mechanism that changes the parameter $p_2$from time to time to a value that is drawn from a symmetric interval around the true value.

One realization of the three models are displayed in Figure 5. Apart from the raw model which follows the input data quite well, the different models lead to distance-versus-time curves that look different. Albeit these curves do not look too different, the parameter estimation of all these stochastic models fails. The parameters are way off, and the rms error for the acceleration becomes considerable. Obviously, it depends on the parameters chosen for the size of the noise, so no general statement about its size can be made. All the results can be found table 1.

TABLE 1 Results of the parameter estimation for the four models.

	p_1	p_2	p_3	rms(acc,fit)	rms(acc,sim)
input	0.0169	-0.0239	0.0172		
raw	0.0182	-0.0258	0.0171	0.0064	0.0010
model 1	0.0172	-0.0245	0.0144	0.2987	0.4134
model 1a	0.0097	-0.0139	0.0017	0.5109	0.5946
model 2	0.0205	-0.0306	0.0158	0.1102	0.1166
model 3	0.0064	-0.0091	0.0116	0.1246	0.1168

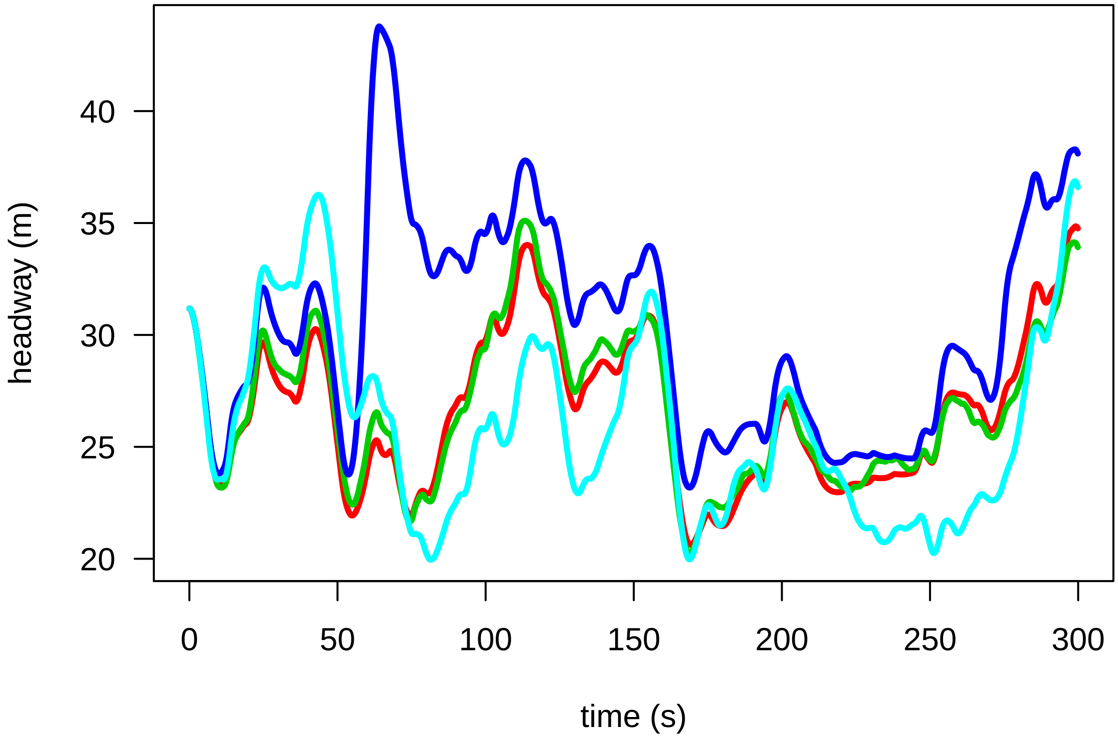

FIGURE 5 Plot of the headway versus time for the three different models, for the same lead vehicle speed function. In red (mostly hidden behind the blue line for model 1 is the original (simulated) gap that the models ought to reproduce.

To see that this is not just the effect from a single simulation, 100 realizations of the process (model 3) have been created with the same fixed parameter set. The fitting of this simulation data to the model equation then yields a different parameter set, whose distribution is displayed in Figure 6 along with the input parameters (vertical lines). The distribution is robust against changes in the lead vehicle's speed, i.e. different realizations of $V(t)$ give the same distribution of fitted parameters.

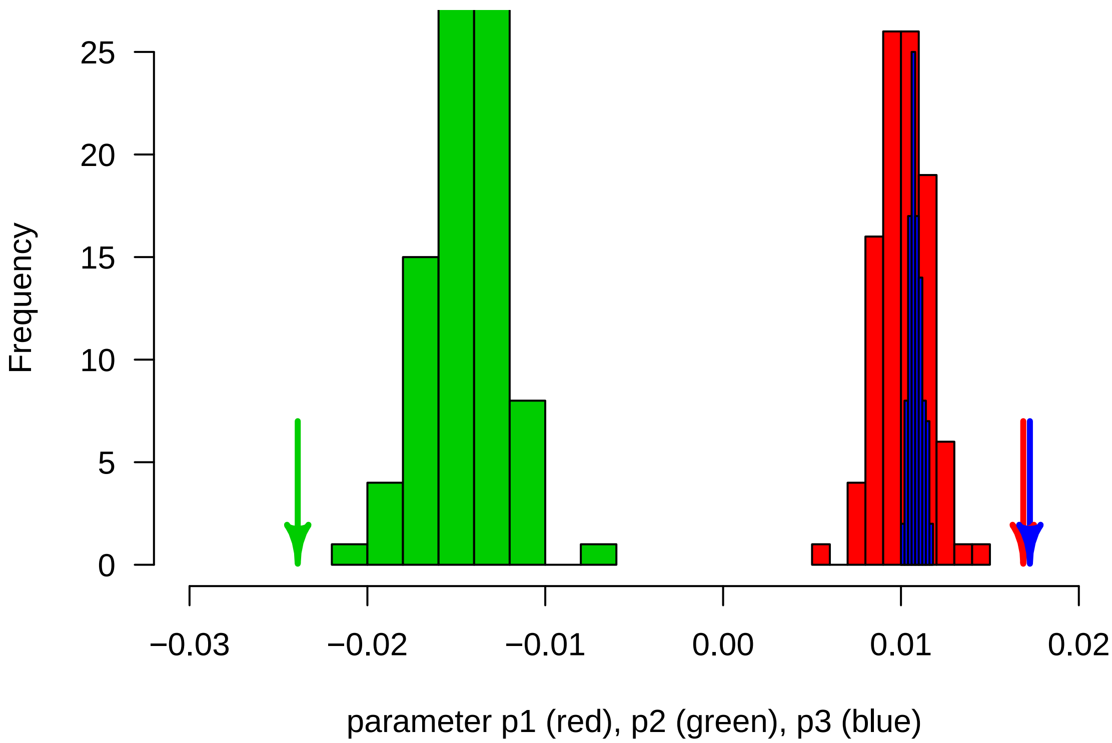

FIGURE 6 Distribution of the parameters estimated from 100 different runs of creating a following trajectory with the same set of parameters and then fitting them to the model above. The input parameters are indicated by the three arrows pointing to the *x*-axis.

5 Conclusions

This work demonstrates that most stochastic processes are potent enough to let a parameter estimation process go awry. It is not only that the goodness of fit becomes worse, in addition, even the parameters do not come out correctly. In order to minimize numerical problems, the model and the fitting procedure have been simplified strongly so that a linear fit was sufficient, with the full statistical power that such a method provides (there are no false minima to be approached by this method, a linear fit also yields the optimum solution). In addition, all the statistical quality measures like t-values of the parameters and their respective significance levels displayed strong values indicating a really good fit nevertheless.

The important point here is that the bad fits do not manifest themselves. So, the researcher would be convinced that anything is good and that the parameter estimation has lead to a good result. However, all of the noise models used here had the power to let the estimated parameter values come out wrongly. So far, we do not have a remedy for this. The statistics, as well as the r.m.s. and even the visual inspection of the results look quite good, but from the numerical experiments it could be seen, that when fitting these type of models, the parameters cannot be estimated correctly.

As a final remark it may be added, that the scripts used for this research, along with the data used to produce Figures 2 and 1 can be retrieved from a public web-space, see (this is a preliminary place, a more permanent web-space will be given in the final version).

Acknowledgments

The funding for this project has been provided by DLR's project I.MoVe. Our special thanks go to the team of the TU Munich (Sebastian Gabloner, Martin Margreiter), who provided the simTD data. simTD itself was funded by the Federal Ministry of Economics and Technology, the Ministry of Education and Research as well as the Ministry of Transport, Building and Urban Development. To fit the model's parameters yet another data-set provided by T. Nakatsuji by have been used, which is gratefully acknowledged.

References

[1] Safe and intelligent mobility – test field germany, 2012. URL: http://www.simtd.de/index.dhtml/enEN/index.html, last access 7.7.2015.

[2] Vassili Alexiadis, James Colyar, John Halkias, Rob Hranac, and Gene McHale. The next generation simulation program. *ITE Journal*, 74(8):22 – 26, August 2004.

[3] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama. Dynamical model of traffic congestion and numerical simulation. *Physical Review E*, 51:1035–1042, 2 1995.

[4] E. Brockfeld, R. D. Kühne, and P. Wagner. Calibration and validation of microscopic traffic flow models. *Transportation Research Records*, 1876:62–70, 2004.

[5] Winnie Daamen, Christine Buisson, and Serge P. Hoogendoorn, editors. *Traffic Simulation and Data – Validation Methods and Applications*. Taylor and Francis, CRC Press, 2014.

[6] Ronald F. Fox, Ian R. Gatland, Rajarshi Roy, and Gautam Vemuri. Fast, accurate algorithm for numerical simulation of exponentially correlated colored noise. *Phys. Rev. A*, 38:5938–5940, Dec 1988.

[7] P.G. Gipps. A behavioural car-following model for computer simulation. *Transportation Research Part B*, 15(2):105–111, 1981.

[8] W. Helly. Simulation of bottlenecks in single lane traffic flow. In *Proceedings of the symposium on theory of traffic flow.*, pages 207–238, 1959.

[9] J. Honerkamp. *Stochastic Dynamical Systems*. VCH Weinheim, 1994.

[10] Rui Jiang, Mao-Bin Hu, Gao Zi-You Zhang, H. M., Bin Jia, Qing-Song Wu, Bing Wang, and Ming Yang. Traffic experiment reveals the nature of car-following. *PLoS ONE*, 9(4):e94351, 2014.

[11] Stefan Krauss, Peter Wagner, and Christian Gawron. Metastable states in a microscopic model of traffic flow. *Physical Review E*, 55(5):5597–5602, 1997.

[12] Kai Nagel and Michael Schreckenberg. A cellular automaton model for freeway traffic. *Journal de Physique I France*, 2:2221 – 2229, 1992. According to the web page of J Phys. I FRance, this paper has so far (12-9-2012) 829 cross references.

[13] G.F. Newell. A simplified car-following theory: a lower order model. *Transportation Research Part B*, 36(3):195–205, 2002.

[14] S. Ossen. *Longitudinal Driving Behavior: Theory and Empirics*. PhD thesis, TU Delft, 2008.

[15] S. Ossen and S. P. Hoogendoorn. Validity of trajectory-based calibration approach of car-following models in the presence of measurement errors. *Transportation Research Records*, 2088:117–125, 2008.

[16] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.

[17] E. P. Todosiev and L. C. Barbosa. A proposed model for the driver-vehicle-system. *Traffic Engineering*, 34:17–20, 1963/64.

[18] Peter Wagner. Analyzing fluctuations in car-following. *Transportation Research Part B*, 46:1384 – 1392, October 2012.