Transportation Research Procedia
Volume 15, 2016, Pages 198–207
ISEHP 2016. International Symposium on Enhancing Highway Performance

A Stochastic Car Following Model

Andreas Kendziorra[1*], Peter Wagner[1†], and Tomer Toledo[2]
[1]Institute of Transportation Systems, German Aerospace Center (DLR), Berlin, Germany
[2]Technion, Haifa, Israel
peter.wagner@dlr.de, toledo@technion.ac.il

Abstract
This paper describes a data-driven, stochastic car-following model. From a data-base of car-following episodes, the acceleration a of the following vehicle is modeled as drawn from a distribution that is sampled directly from the data. To make this work, the input variables speed v, speed difference Δv, net space headway g (gap), and acceleration A of the lead vehicle are discretized, and in each of the resulting bins a different acceleration distribution $F_{v,\Delta v,g,A}(a)$ is estimated. In most cases, the acceleration values are distributed according to a Laplace distribution. Missing data-bins are interpolated. This model is then tested; it is found, that the resulting distributions of safety surrogate measures reproduce the ones found in reality.

Keywords: Car following, stochastic model, modelling car accidents

1 The Concept

Most existing car following models are deterministic and do not consider the uncertainty and fluctuation of human perception and behavior (Helbing, 2001), (Nagel, et al., 2003), (Treiber, 2013). Therefore, safety-relevant aspects may differ from reality in simulations which could also be seen by the fact that car following models are designed to be accident free. This constraint reduces the usability of these models for traffic safety research significantly. Nevertheless, some investigations regarding traffic safety are still possible. For example, the model of Krauß (Krauß, 1998), (Krauß, et al., 1997) allows very strong braking deceleration in critical situations, which can be used for traffic safety related investigations.

Note also, that the overwhelming majority of car-following models are essentially one-dimensional, which restricts traffic safety analyses to rear-end-crashes. This is also the case for the considerations here. However, surrogate safety measures (SSM) like TTC (time-to-collision, Eq. (1)

[*] On leave.
[†] Corresponding author

 Selection and peer-review under responsibility of the Scientific Programme Committee of ISEHP 2016

doi:10.1016/j.trpro.2016.06.017

and DRAC (deceleration rate to avoid a crash, Eq. (2)) which are used here have a similar limitation. For reference, these two SSM are defined as:

$$T := TTC = \frac{g}{\Delta v} \qquad (1)$$

$$D := DRAC = \frac{(\Delta v)^2}{2g} \qquad (2)$$

where g is the net headway between the lead and the following vehicle, and $\Delta v = V - v$ is the speed difference with V the speed of the lead and v the speed of the following (subject) car.

Here, we follow the approach of (Yang, et al., 2010), in which a non-deterministic car following model was developed based on a huge data set. We developed a stochastic car following model that enables reasonably realistic simulations of the human driving behavior, in particular with respect to traffic safety relevant aspects. That means, for example, that the distributions of the SSM reflect the ones found in reality.

In this model, the acceleration a of a following vehicle will be determined. It is assumed that a depends on the velocity v of the following vehicle, the velocity difference Δv between the subject and the proceeding vehicle, the gap g between the two, and finally, on the acceleration A of the lead car. The acceleration a will be drawn from a probability distribution, whereby this distribution depends on $v, \Delta v, g$ and A. That means, for every tuple $(v, \Delta v, g, A)$ there exists a distribution function $F_{v,\Delta v,g,A}(x)$, and a will be drawn from it. These distribution functions will be determined by the mentioned data set of the naturalistic driving study (see Section 2). For that, the data for $v, \Delta v, g$ and A will be binned. For each bin, the expected value, the standard deviation as well as the distribution class of the acceleration a will be determined. However, this will be done only if the bin contains at least 100 data-points. If the bin has a smaller number of values, the expected value will be determined by an alternative car following model. For the standard deviation, a constant will be assigned and a Laplace distribution will be chosen for this bin. The parameters of the reference model will also be determined from the data-set under consideration.

The Laplace distribution is defined as:

$$L_{\mu,\sigma}(a) = \frac{1}{2\sigma} \, exp(-|a - \mu|/\sigma) \qquad (3)$$

where μ and σ are two parameters, its center and its width, respectively. For the data-set below, the type of the distribution has been found by testing in each of the bins the empirical distribution against a several distributions via a g-test and a χ^2-test. In 86% of the cases, a Laplace distribution was found to provide the best test-result.

2 The data set

For determining the distribution function described in Section 1, the data of the *Intelligent Cruise Control Field Operational Test* (Fancher, et al., 1998) was used. This field operational test was conducted between 1996 and 1997 in Michigan, USA. For that, 10 vehicles were equipped with various sensors and instruments to measure driving dynamic parameters as well as the distance to the preceding vehicle. The vehicles were given to 108 volunteers for two to five weeks, in which the data was constantly recorded. Although the purpose of the experiment was to investigate the comfort of adaptive cruise control (ACC), a sufficiently large amount of recorded trips without ACC was present. (Originally, they were used to compare the performance of the ACC against natural conditions.) More precisely, there are 8,690 trips of 102 drivers adding up to 1,821 driving hours and 88,000 kilometers, in which no ACC was used. 80% of the trips were used to create the model and 20% to verify it.

The following data was extracted from the trips:

- Running time: time since ignition switch was turned on

- The velocity of the vehicle
- Distance: the distance to the detected target in front of the vehicle. If no target was detected, the distance will be assigned to NA.
- Backscatter: integer value between 0 and 1023, stating the relative amount of transmitted laser energy scattered back by the ambient conditions, and that is received by the sensor. High backscatter values indicate low quality of distance data.

These data were recorded with a frequency of 10 Hz, i.e. with a time resolution of $\Delta t = 0.1$ s, see (Fancher, et al., 1998) for a more detailed description. Furthermore, the acceleration of the vehicle and the rate of change of the distance (i.e., the velocity difference between the vehicle and the detected target) were derived from velocity and distance data, respectively. However, the quality of these derived data is often not very good. Therefore, this work used a different approach to compute these two variables which is described in the following subsection.

3 Data processing

Here, the most important steps of the data processing are presented. From now on, the following notation will be used: if t is a time point of a time series, then $t + 1$ is the next time point, i.e., the time point Δt seconds after the time point t (recall, that the data was recorded with 10 Hz). More generally, for a non-negative integer k, $t + k$ is the time point k tenths of a second after t and $t - k$ the time point k tenths of a second before t.

The following main processing steps were undertaken:

- Computing the acceleration time series based on the velocity time series
- Processing, filtering, and correcting the distance data
- Computing the velocity difference and the acceleration of the lead vehicle on the distance time series

These steps are explained in the following.

3.1 Computing acceleration time series

A Savitzky-Golay filter (Savitzky, et al., 1964) was used to compute the acceleration time series of the vehicle. This is a filter that allows to compute a smoothed version of a time-series and a smoothed derivative of the time series. For computing the acceleration time series based on the velocity time-series, a Savitzky-Golay filter with window size $m = 11$ (corresponding to 1.1 seconds) and a polynomial order $p = 1$ was chosen. It reads:

$$a_t = \sum_{k=-m}^{m} c_k v_{t+k}$$

where the c_k are the set of filter coefficients that can be precomputed, they depend only on the filter-width m and on the order p of the polynomial chosen to fit the data-points. There is no general rule how to determine these parameters to obtain the best results. While experimenting, $m = 5$ and $p = 1$ turned out to be most suitable. This is understandable, since a numerical inspection of the data indicates, that the speed is in fact mostly a linear function, at least over short time-intervals such as the window size of the filter. For the data here, there is a tendency that smaller numbers of p yield less noisy results, but again this depends on the data at hand.

3.2 Processing of distance data

These data have been used as they are, i.e. no smoothing by a Savitzky-Golay filter has been performed. However, an approximation of the speed-difference $\Delta v_t = (g_t - g_{t-5})/0.5)$ has been used and from this, the gap $\hat{g}$ for the next time-step is estimated as

$$\hat{g} = g_t + \Delta t \, \Delta v_t$$

If this estimated gap $\hat{g}$ deviates too strongly from the actually measured one g_{t+1} (this can happen if the lead vehicle changes, or if the distance sensor has lost its track), then this is noticed and the time series is divided at this point. After finding all the candidates for break points, it is tried to correct some of them. E.g., the sensor sometimes lost track only for a short period of time, after which it connected again to the same vehicle in front. This can be recognized, and then the break-point is interpolated to get a complete car following episode. Finally, only those episodes have been kept which are longer than 3 s at all. Note, that the approach put forward here is not restricted to car-following, but in general to the interaction between vehicles, even if they do not follow each other for a long time.

3.3 Computing velocity difference and acceleration of the lead vehicle

For the valid distance data of the step before, the speed difference is then computed again as the derivative of the gaps with the help of a Savitzky-Golay filter. The same holds also for the acceleration of the lead vehicle. The speed of the lead vehicle is of course computed from $V_t = v_t + \Delta v_t$, the acceleration A_t of the lead vehicle follows then from the (Savitzky-Golay) filtering of this speed time series.

With the speed difference at hand, additional data-cleaning have been performed. For instance, sometimes the vehicle brakes strongly, but the speed difference is positive, or the driver stops but a lead vehicle is not recorded. Those episodes are eliminated from further considerations, too.

3.4 Binning

Different binning schemes have been investigated, without a clear result which is the best one. In the following, the acceleration of the lead vehicle has been binned into broad intervals of 0.5 m/s^2, the speed and distance into 4 m/s or 4 m, respectively, while the speed difference has been binned into 1 m/s bins. In one particular example, this yields 6,054 hyper-cubes that cover most of the data in the 4D space. The boundary values of the four variables are determined from the data, typical numbers are $v \in [0{,}40]$ m/s, $g \in [0{,}120]$ m, and accelerations $a \in [-5{,}5]$ m/s^2.

Alternatively, one may came up with the idea to choose the binning non-equidistant so that each bin has roughly the same number of data-points in it, at least in one dimension. Which constitutes a quantile-based binning. This has not been tried here, but will be the topic of future research.

4 The Model

This model is updated as follows. Let the position of the vehicle at time t denoted by x_t, with a time-step size of Δt (no need for this to be constant, but usually it is). Then, the position of the vehicle at the next time step is given by

$$x_{t+1} = x_t + v_t \Delta t + \frac{1}{2} a_t \, (\Delta t)^2 = x_t + \frac{\Delta t}{2} \, (v_t + v_{t+1})$$

where the kinematic relationship $v_{t+1} = v_t + a_t \Delta t$ has been used. What is needed is the specification of a_t. Usually, this acceleration a_t is a function of the input variables $v_t, \Delta v_t, g_t$, and acceleration of the lead vehicle A_t. In the model described here, this is not directly so. Instead, the acceleration is drawn from a distribution that has been sampled from the data described above. A number of methods can be used to do this. Here, within each bin defined by a certain set of bin-widths, the distribution is parameterized by its mean and standard deviation. As stated above, the distribution is a Laplace distribution. Alternatively, but this would have been numerically more demanding, one could approximate the distribution in each of the bins by a certain discretization of the sampled acceleration values.

After drawing such an acceleration value, it is subjected to certain boundary conditions (i.e. the Laplace distribution may give an unrealistic value of the acceleration), i.e. $a_t \in [a_{\min}, a_{\max}]$; the same holds for the speed v_{t+1}, which is constrained into $v_{t+1} \in [0, v_{\max}]$. No attempts are being made in making this model crash-free.

If the dynamics hit upon a region in $v_t, \Delta v_t, g_t, A_t$ where no data had been available (which happens rarely, in less than 1% of the cases), then a conventional car-following model is used. Two are currently implemented, the Helly model (Helly, 1959), Eq. (4) and the one of Krauß (Krauß, et al., 1997), Eq. (5). They are given as:

$$a_t = h_0 + h_1 \Delta v_t + h_2 g_t \tag{4}$$

$$v_{t+1} = \min\left\{V_t + 2b\frac{g_t - V_t\tau}{v_t + V_t + 2b\tau}, v_t + a\Delta t,\ v_{\max}\right\} \tag{5}$$

where the h_i are the parameters of Helly's model, and a, b, τ are the parameters of the Krauß model.

Both models are subject to the restriction in speed and acceleration, this is not included explicitly in the equations to simplify notation. Obviously, this creates discontinuities at the boundaries between the stochastic model and the conventional models, which can be seen e.g. in the plot of Figure 1.

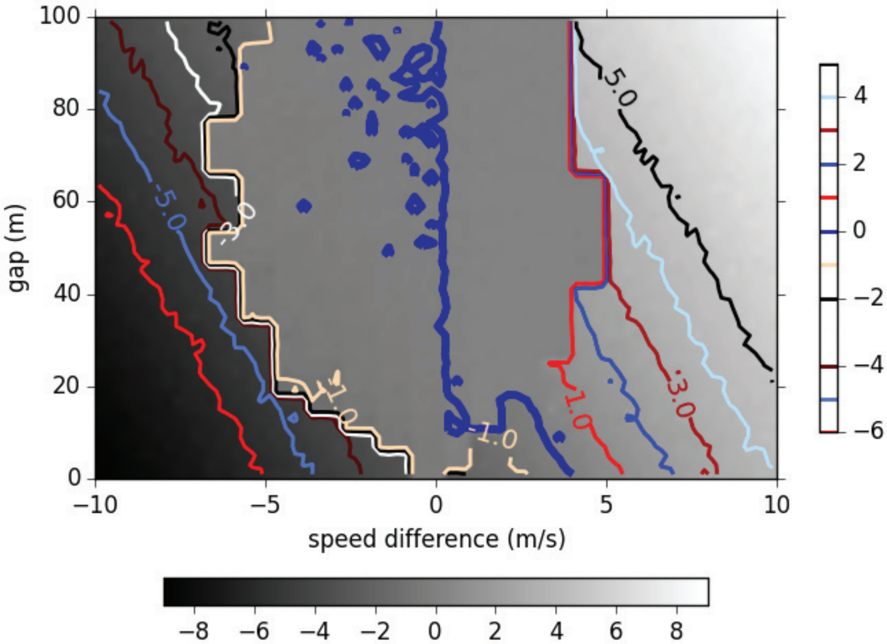

Figure 1: Acceleration function of the model. Speed has been fixed at $v = 20$ m/s and $A = 0$, respectively. Then, acceleration can be plotted as $a(\Delta v, g)$.

As mentioned, this model is not crash-free. The Figure 2 displays an example where the model produces a crash. This happens rarely, and in this case even by changing the initial condition of the random number generator may eliminate this crash or produce another one.

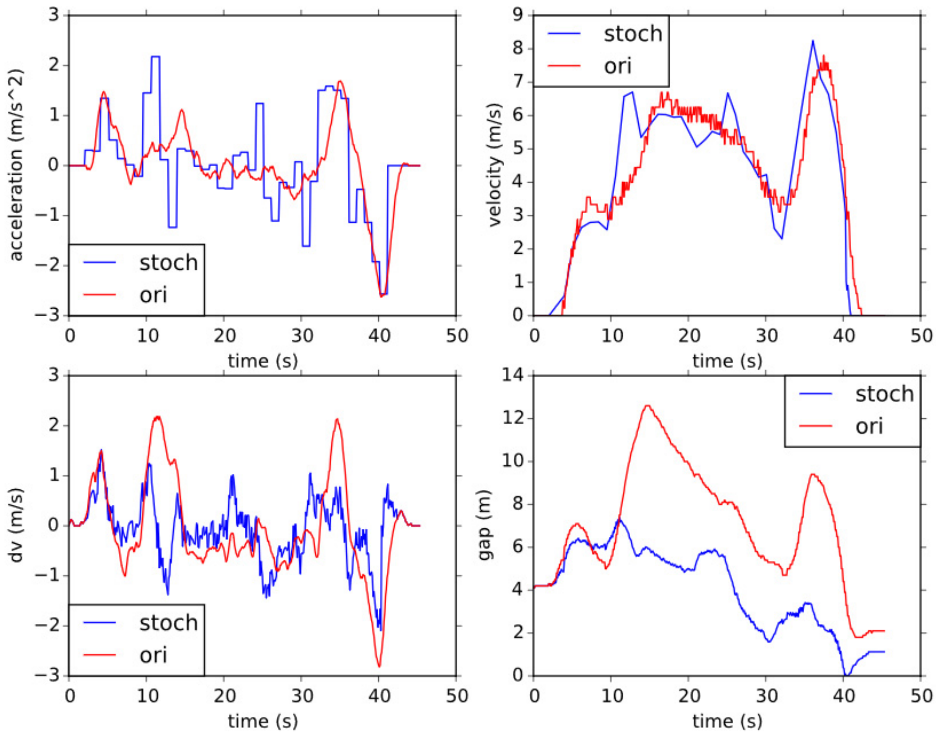

Figure 2: One example for a crash. Plotted are the acceleration (top left), speed (top right), velocity difference (bottom left) and the gap (bottom right) of the following vehicle (blue lines), together with the empirical values (in red).

5 Running the Model

The model has then tested in two different instances. In the first, the FOT data-set above is used once more. In the second test, the fundamental diagram of the model is compared to the one of the Krauß model.

5.1 Microscopic test

Again, the car-following trajectories as found in the data-set above have been used. In order to compare it with the real data, for each of the 3,748 clean trajectories the model is set at the beginning of each episode to the empirical values of speed, gap, speed-difference and lead acceleration. From this, the model computes its acceleration, and then, together with the data of the lead vehicle (mostly lead vehicle's speed), also the updated speeds and distances. In each subsequent time step, the vehicle is fed only the speed and acceleration of the lead vehicle, the distance and its own speed are a consequence of the model. To clarify this:

$$a_t \text{ from } L_{v_t,\Delta v_t,g_t,A_t}(x)$$
$$v_{t+1} = v_t + \Delta t\, a_t$$

$$g_{t+1} = g_t + \frac{\Delta t}{2}(V_{t+1} + V_t - v_{t+1} - v_t)$$

with the initial condition v_0, g_0 taken from the data.

In each time-step the two SSM's T, D are computed, too. Their distribution is compared in Figure 3 to the empirical distributions.

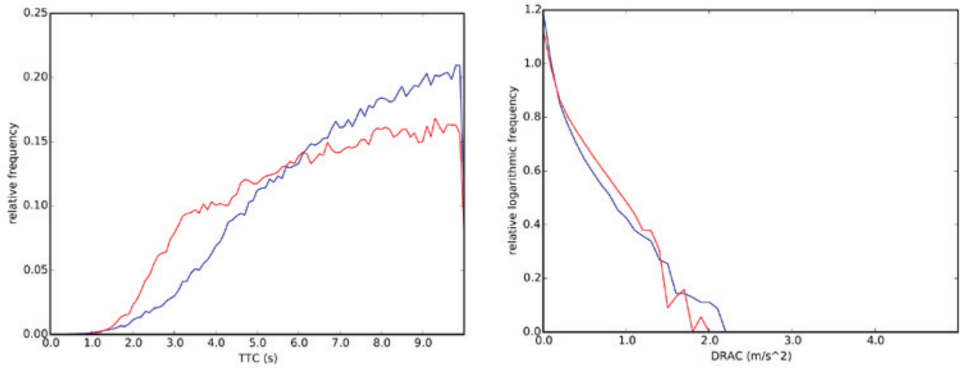

Figure 3: The distribution of TTC (left) and DRAC (right) in red, compared with the empirical distributions (in blue).

It could be seen, that the model can reproduce to a certain degree the empirical distributions. While the distribution of the DRAC values is reproduced very well, there are visible differences in the TTC distributions.

5.2 Macroscopic test – the fundamental diagram

The fundamental diagram has been computed by running a varying number of vehicles (N=40, 120, 200, 280, 360) behind each other in a circle of fixed length. Along this circle, 8 loop detectors have been placed with a distance of 500 m. They aggregate the data into 3 min averages. Note, that from these simulated loop detectors all the important variables like speed and distance can be obtained which are needed to compute the flows, densities, and average speeds. All the variables are local variables, which can be seen e.g. from the density. The global density would give just five different values, see Figure 4 for a plot.

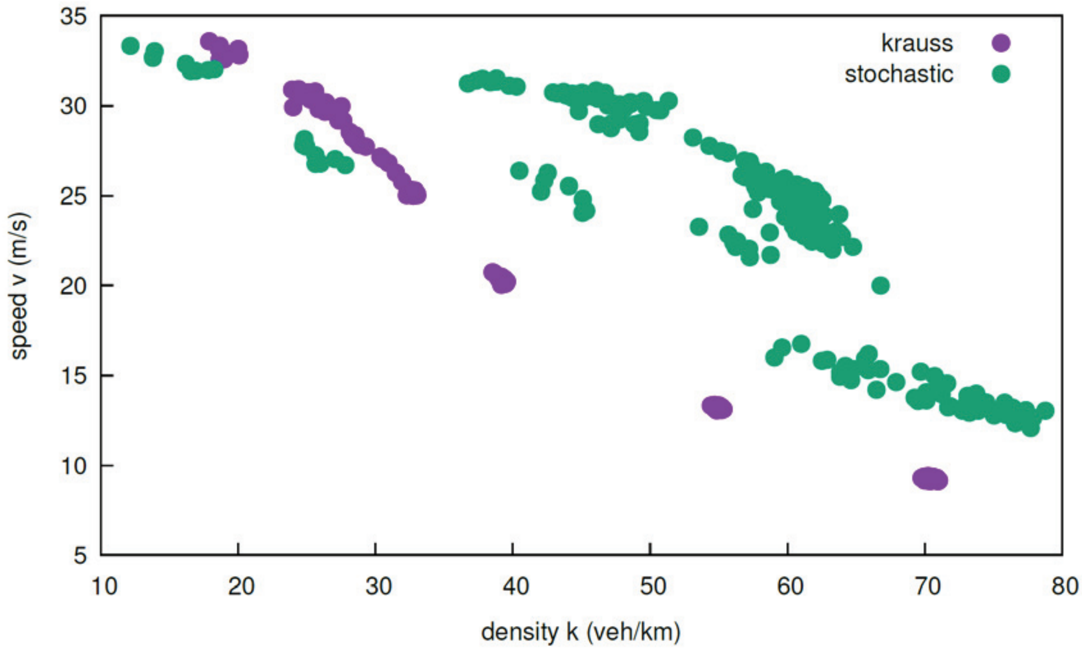

Figure 4: The fundamental diagram as average speed versus density as sampled from the synthetic loop detectors along the circle. The green points represent the stochastic model, while the purple ones are the results of the Krauß model.

While the general form of the fundamental diagram looks fine, the values obtained are way off. Especially the maximum flow, and the density where the speed starts to decline, are roughly a factor of two too big for the stochastic model. So far, there is no good explanation available why this happens.

6 Conclusions

It has been demonstrated, that a stochastic car following model that is constructed by a purely data-driven approach can at least approximately reproduce SSM's like TTC and DRAC. However, a lot of work remains to be done. Most importantly, as has been shown, this model cannot fully reproduce the fundamental diagram. To remedy this, the model has to be re-analyzed to find the point where it deviates from real drives, albeit microscopically it looks good. It seems, that human drivers so far are more carefully than what the model does. Also, it could be that the filtering conditions has been too restrictive. In addition to this, the current implementation of the model is very slow, which makes additional tests very time-consuming.

Then, it might be interesting to investigate the relationship between the distribution of the SSM and crashes in the model into more detail. For this to make sense, it is a good idea to find a good method to produce synthetic trajectories of lead vehicles (more precisely: its speed and acceleration) so that such a test is not restricted to the data recorded in reality, which are known to be simply not enough in order to compare it with real accident rates. For instance, in Germany a vehicle must drive on average 280,000 km until it experiences one crash, and this is not necessarily a rear-end crash. Therefore, a powerful method to create those lead vehicle data is needed so that a large number of test-cases can be tested.

Let us end with a warning: the approach chosen here is one that uses very few assumptions about the underlying human behavior, since in the first place it uses only data, and no assumptions. This makes it very attractive. However, to make it finally work, a considerable number of additional safeguards had to be implemented, and still not anything is as good as one could expect. And clearly, there are still a number of things that are not well dealt with, such as the acceleration memory and the fact that subsequent draws from the acceleration distributions are uncorrelated. Unfortunately, it is not so easy to say how to remedy this.

Acknowledgements

Some parts of this work had been done during a stay of one of the authors (PW) at the Technion, and with financial support of the GIF, the German-Israeli Foundation for Scientific Research and Development. Both are gratefully acknowledged.

References

Fancher, P., et al. 1998. *Intelligent Cruise Control Fiel Operational Test, Final Report.* 1998.

Helbing, Dirk. 2001. Traffic and Related Self-Driven Many-Particle Systems. *Reviews of Modern Physics.* 2001, Vol. 73, pp. 1067-1141.

Helly, W. 1959. Simulation of bottlenecks in single lane traffic flow. *Proceedings of the symposium on theory of traffic flow.* 1959, pp. 207-238.

Krauß, Stefan. 1998. *Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics (PhD thesis).* Cologne : University of Cologne , 1998.

Krauß, Stefan, Wagner, Peter and Gawron, Christian. 1997. Metastable states in a microscopic model of traffic flow. *Physical Review E.* 1997, Vol. 55, pp. 5597-5602.

Nagel, K., Wagner, P. and Woesler, R. 2003. Still flowing: approaches to traffic flow and traffic jam modeling. *Operations Research.* 2003, Vol. 51, pp. 681-710.

Savitzky, A and Golay, M. J. E. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem., Analytical Chemistry.* 1964, Vol. 36, pp. 1627-1639.

Treiber, M. and Kesting, A. 2013. *Traffic Flow Dynamics Springer.* s.l. : Springer, 2013.

Yang, H.-H. and Peng, H. 2010. Development of an errorable car-following driver model. *Vehicle System Dynamics.* 2010, Vol. 48, pp. 751-773.