Analyzing human driving data an approach motivated by Data Science methods

Peter Wagner[a], Ronald Nippold[a], Sebastian Gabloner[b], Martin Margreiter[b]

[a]*Institute of Transportation Systems, Deutsches Zentrum für Luft- und Raumfahrt, Germany*

[b]*Engineering Faculty Construction, Geo, Environment, Department of Traffic Techniques, Technical University of Munich, Germany*

Abstract

By analyzing a large data-base of car-driving data in a generic way, a few elementary facts on car-following have been found out. The inferences stem from the application of the mutual information to detect correlations to the data. Arguably, the most interesting fact is that the acceleration of the following vehicle depends mostly on the speed-difference to the lead vehicle. This seems to be a causal relationship, since acceleration follows speed-difference with an average delay of 0.5 seconds. Furthermore, the car-following process organizes itself in such a manner that there is a strong relation between speed and distance to the vehicle in front. In most cases, this is the dominant relationship in car-following. Additionally, acceleration depends only weakly on distance, which may be surprising and is at odds to a number of simple models that state an exclusive dependency between acceleration and distance.

1. Introduction

Driving behavior models have become important tools in transportation science and engineering applications since their first appearance in the 1950s [21]. By now, a lot of models that seek to describe human driving behavior, i. e. acceleration processes (and lane changing maneuvers) have been introduced, see [16, 10, 6, 15] for albeit incomplete overviews. The complexity of these models ranges from very simple models with few parameters like Newell's lower order

Preprint submitted to Elsevier *February 5, 2016*

model [18] or the cellular automaton model [17] up to multi-regime models like Wiedemanns psycho-physical perception threshold model [26] or the model implemented in the MITSIM-lab [1] open source simulator with lots of thresholds and different conditions describing an assumptive behavior in specific situations.

It is commonly believed, that the behavior of a human driving a vehicle can be described quite generally by the two equations

$$v(t + \Delta t) = v(t) + a(\Delta v, g, v)\Delta t \tag{1}$$

$$x(t + \Delta t) = x(t) + \frac{\Delta t}{2}\left(v(t + \Delta t) + v(t)\right) \tag{2}$$

where $(x(t),\, v(t))$ are the position and speed of the vehicle at time t, $(g, \Delta v)$ are the distance and speed difference to the vehicle in front, and $a(\cdot)$ is the acceleration function. There is no need for Δt to be equi-distant, and in fact the so called action-point models [22, 26] claim that a driver changes her course of action only from time to time, based either on external forces or even on no apparent reason which may lead to an exponential distribution of the Δt–values [24].

Since a lot of empirical research has been done on this topic, too, only occasionally approaches have been made to connect empirics and models in a physicist's manner. A course of action that is often pursued in driver modeling is to do calibration and validation [8] of known models. By their very nature, these exercises often yield reasonable results (of the order of $10 - 20\,\%$ root-mean-square error [4]) even for models that are known to be a bad description of driving, like e. g. the simplest cellular automaton models [17] or the optimal velocity model [3].

Here, a different approach is adopted. A large data-set of car driving (collected in the second half of 2012 in the German simTD project) is analyzed with the help of methods from data science [9]. This research will try to look on these data and try to find some general relationships between the measured data. Within this work, the approach is limited to the four variables speed v, acceleration a, distance to the lead vehicle g, and the speed-difference Δv to the lead vehicle. However, in principle it is also possible to include other parameters that might influence driving behavior, such as the time of day when the driver is driving, the street-type, the weather conditions, or the acceleration of the lead vehicle which might be accessible to the driver by observing the lead vehicle's braking lights. Finally, this approach here can find a few general features of the function $a(g, \Delta v, v)$ in Eq. (1) above.

The present paper consists of four parts: A description of the data set under consideration is contained in section 2. The following section 3 gives an overview of the statistical measures maximal information content (MIC) and mutual information (M) used in this paper, and states the main results. Finally, the conclusions of this analysis are presented in section 4.

2. Description of the simTD data

The data to be used here have been recorded during the field test of the simTD project from July 2012 to December 2012. Altogether 120 vehicles where driving around for 98 days. Different drivers where assigned to the vehicles, but this assignment is kept confidential and therefore not part of the data-set. Because of the main goal of the project, to estimate the efficiency of vehicle-to-vehicle and vehicle-to-infrastructure communication, the vehicles drove on predefined routes only, thereby covering an area of roughly 15 × 45 km around the German city of Frankfurt.

All the data were made available by the six participating German car manufacturers via the CAN-bus in a frequency between 200 Hz and 0.5 Hz. To abstract the manufacturer-specific protocols, all the data were extracted from the internal network of the car (the CAN-bus) by a specialized VehicleAPI (VAPI) which was especially developed within this project. Therefore, all data were available in the same generic format for the project. All the signals from the CAN-bus where synchronized using the GPS time.

The data used in this paper have been recorded by four sensors, that were built into the vehicles: a GPS sensor, an acceleration sensor that was aligned with the car's geometry and therefore allowed the measurement of the longitudinal (in driving direction) and lateral acceleration (perpendicular to the driving direction), the distance and velocity difference to the lead vehicle by a radar/lidar sensor, and the speed from the traditional wheel sensors. The acceleration data was noisy, but not in an unreasonable manner, so it was decided to use them unfiltered as they are. This does not rule out, that the internal machinery within the cars itself does some filtering, but from what have been seen by visual inspection, this does not seem very likely.

Within this project, the GPS data have been enhanced into a differential GPS by correction signals received over UMTS and ITS-G5 (802.11p). These data have been matched on an underlying digital road network, but this has not been used in this paper but was used in the communication part of the projects. All the cars were normal cars (with all the sensors) that can be bought in exactly this form, only the data-acquisition had been added within the project.

The data were recorded asynchronously. That means, that the variables to be analyzed here (distance g, speed v, speed-difference Δv, and acceleration a) are recorded in their raw format not at the same time, and the time-difference between subsequent readings even of the same sensor is not guaranteed to be equidistant. The variables are acquired by three different sensors: a radar sensor which measures the distance and speed-difference to the vehicle in front (sometimes it also picks a tree at the border of the road), the vehicle's internal measurement of the speed, and the acceleration which is recorded by a dedicated sensor.

As explained above, the current position of the vehicle is measured by a GPS receiver. Note, that the GPS provides an additional measurement of the speed which has not been utilized here. Also, it could have been used to determine the acceleration of the vehicle, which also has not been done here for the analysis below systematically. A brief view into this, however, reveals that the data obtained from GPS are in good agreement with the recordings to be used in the following.

The typical frequency with which the data are recorded is about 10 Hz. Therefore, it was decided to force the data into a common time-basis by aggregating them to 0.1 s. This worked well, in most cases less than 4 data-points fall into one time-bin, which are then averaged to get the time-series that will be analyzed subsequently.

It is safe to assume that the data are not error-free. From visual inspection of the data, it turns out that there are a lot of points where the data-stream disconnects, i.e. there are gaps in time which are larger than 0.1 or 0.2 s, where no values are recorded. To transform this into a measure of data-quality, the one-step ahead prediction error is used in the following. This error can be computed from estimates of the gap and speed of the subject vehicle for the next step in time, which is given by:

$$\hat{g}_k = g_{k-1} + \Delta v_{k-1}(t_k - t_{k-1}), \tag{3}$$

$$\hat{v}_k = v_{k-1} + a_k(t_k - t_{k-1}). \tag{4}$$

This is compared with the actual measurements at the time-point $k+1$, thereby defining a measure of consistency of the time-series:

$$e_k^{(g)} = \hat{g}_k - g_k, \tag{5}$$

$$e_k^{(v)} = \hat{v}_k - v_k. \tag{6}$$

Note, that this works for any time-step size $\Delta t_k = t_k - t_{k-1}$, but for larger Δt_k the likelihood increases that the prediction error becomes large.

It turns out, that the two errors $(e_k^{(g)}, e_k^{(v)})$ are distributed according to a Cauchy distribution $1/(1+e^2)$, as can be seen in Figure 1. In the following, two thresholds $\theta_v = 0.5$ m/s and $\theta_g = 1$ m have been chosen: whenever both prediction errors are smaller than their respective threshold, then the data-point is considered valid. It is dropped from all the subsequent analyses, if one (or both) of the prediction errors is larger than its threshold. This eliminates about 10% of the data-points.

No additional attempts have been made to clear the data-base from errors, with one exception. It is only looked at those sequences of the data that had at least 300 data-points (about 30 seconds) in one uninterrupted sequence, defined as the time-difference between two subsequent points in not larger than 1 second.

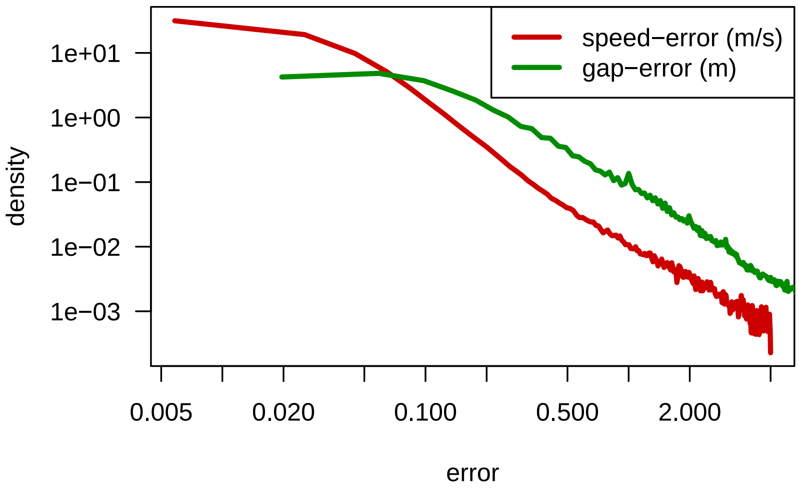

Figure 1: The distributions of the prediction error in speed (red) and distance (green). The slope in the plots is close to 2. Only the distribution for the positive values of e is displayed in this double log-plot.

More cleaning is always possible, some of it happens as can be seen e.g. in Figure 2, where only the central parts of the various distributions are displayed. In effect, this means that large time headways, large speed differences, and large accelerations have not been regarded. In fact, Figure 2 displays the headway distribution $p(T)$, the histogram of the mean headways of the 258 different trips that have contributed to the headway distribution $p(T)$, the distribution of the speed differences $p(\Delta v)$ and the one of the accelerations $p(a)$.

As can be deduced from Figure 2 the distributions of Δv and a are Laplace distributions [23], i.e. they can be described by

$$p(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right) . \tag{7}$$

Here, μ is the mean value and σ is a measure for the width of the distribution. The mean-value of these two distributions is very close to zero, which is not very surprising – all trips start and end with standing vehicles, which can happen only if the sum of all the acceleration values is zero. Both distributions are also roughly symmetric, which is caused to a certain degree by the fact that the data contain episodes where a lead vehicle is at least present. In this regime, the data point out that the drivers do a fairly good job at staying in a smooth mode with moderate acceleration values. To quantify this, just 9.5 % of the acceleration values are smaller than -1 m/s^2 and 7.5 % were larger than $+1$ m/s^2, displaying a slight asymmetry between acceleration and deceleration.

The form of the time headway distribution is still open [7, 14, 12], a large amount of different basic and composed models have been presented in the literature with no consensus reached so far.

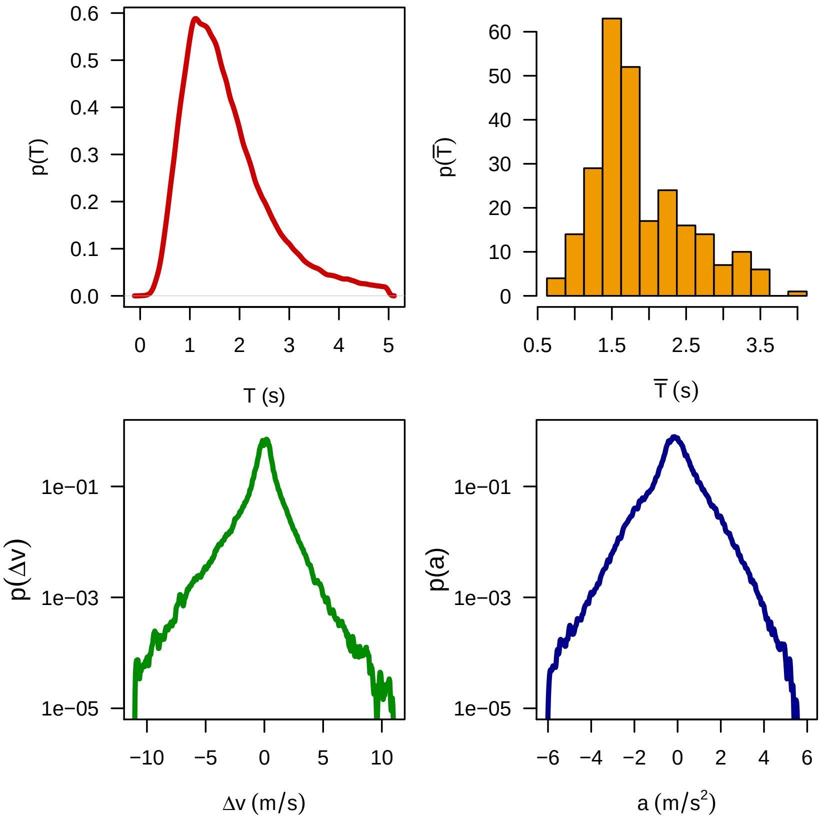

Figure 2: The distributions of time headway (top-left), the distribution of the mean-values of the individual trips (top-right), speed difference (bottom-left), and the acceleration (bottom-right) for five days randomly picked from July and August 2012. The estimation of the distribution functions has been done by a kernel density estimate as has been implemented in R's [19] density() function.

3. Results

3.1. Applying MIC, M, and C

The main tool to be used in this analysis is the maximal information content (MIC) [20, 2], and its close relative, the mutual information M. For reference, the Pearson correlation C is used, too. MIC is described in [20]. It is based on the mutual information content between two variables x and y, which can be computed by:

$$M = \int p(x,y) \log \frac{p(x,y)}{p(x)p(y)}\, dx\, dy \tag{8}$$

The advantage of MIC and M over C is that they can detect relationships between variables even if this relation is highly non-linear, where the Pearson correlation very often displays small or zero values. The advantage of MIC over M is that it is independent of the form of the division of the variable-space and it is normalized to the interval $[0, 1]$. However, MIC solves an optimization problem to actually find the maximum (mutual) information, and this make it roughly an order of magnitude slower to compute than M – which is the reason, why in most cases M is used. Let us note in passing, that C can be computed even faster, so it might be taken into consideration of the analyst as well.

In the case of the variables here, M is computed by dividing the space of each of the two variables x and y by selecting boundary values x_i, y_j so, that each bin receives the same number of data-points. Then, the marginal distributions $p(x)$, $p(y)$ are very simple and given by $1/n$ and $1/m$ respectively (in each bin), where (n, m) are the number of bins of each variable. The argument of the logarithm of equation (8) is then the number of data-points in each of the boxes defined in this manner, divided by the total number of data-points and divided by $1/(mn)$. Typically, the data are divided into $n = m = 11$ bins to have good statistics for the computation of M.

The Figure 3, then, displays MIC for all of the six possible pairs of variables as function of the index of the vehicle running from one to the maximum number of vehicle-trips that have been recorded within this data-set.

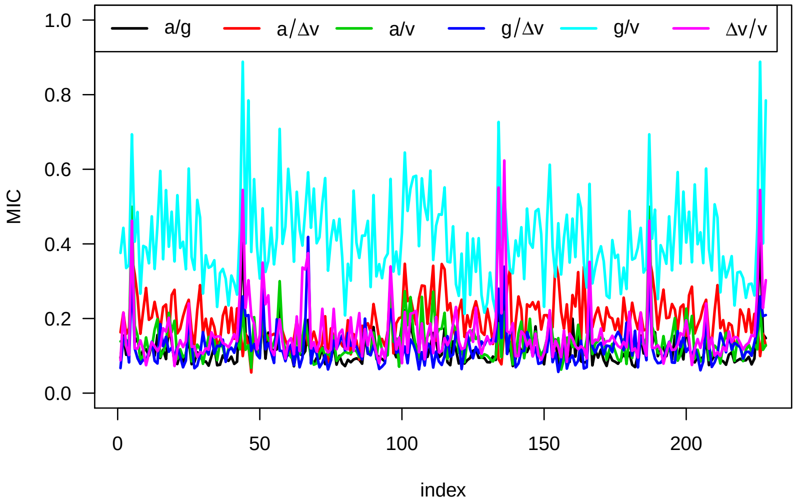

Figure 3: The six different MIC's between the four variables, as function of the index of the trips. MIC has been computed with the R-package minerva() [19, 2].

It can be seen, that there are mostly two pairs of variables that are important, which is in this case the combination (g, v) and $(a, \Delta v)$. This becomes a bit clearer when looking at the distribution of the MIC–values in Figure 4:

For these data it turns out, that with the exception of the pair $v, \Delta v$, MIC, M, and C are well in line.

However, an additional step is needed. From the computation of a time-shifted $M(\tau)$ which is defined exactly as the cross-correlation function $C(\tau)$

$$M(\tau) = M(x(t), y(t - \tau))$$

(and similar for MIC as well) it has been found, that the maximum of $M(\tau)$ is not always at $\tau = 0$. In fact, especially for the pair $(a, \Delta v)$ it turns out, that the maximum in M is roughly at $\tau = 0.5$ s. For the other combinations, the dependency of M on τ is much weaker. Therefore, to complete the picture in Figure 3, the maximum value over the interval $\tau \in [-3, 3]$ is to be plotted instead in Fig. 5. In addition, this Figure contain also the distribution of τ–values, where this maximum is achieved. Again, a clear signal can only be seen for the two combinations (g, v) and $(a, \Delta v)$.

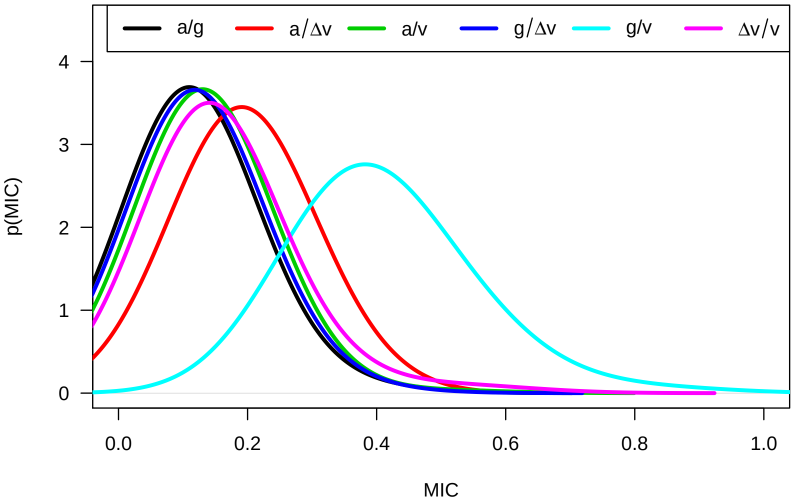

Figure 4: The distribution of the MIC–values for the six different combination of parameters. Note the average ranking: The combination of g, v has usually the largest MIC, followed by the one between $a, \Delta v$. The rest of them are weaker.

A bit more of information can be gained by determining the average $M(\tau)$ functions, along with the average cross-correlation functions. Averaging means, that these functions are computed first for each vehicle-run individually, and then averaged over the (in this case 229) trips in the data-set. If the drivers behave very different, then this averaging should yield no relevant signal. However, the Fig. 6 demonstrates a certain degree of similarity between the drivers running these cars. Again, the two combinations (v, g) and $(a, \Delta v)$ stick out, while the remaining relationships show a very weak correlation and a function that changes little over the chosen interval of lag-values.

3.2. A physicist's approach

It is interesting to remark, that in the vast amount of literature of car-following nobody has ever made a simple plot of acceleration as a function of Δv and g or something similar, at least to the knowledge of these authors. This will be done now. First of all, the $(\Delta v,\, g)$–phase-space is divided in the same manner as has been done above for the computation of M. This yields bins of unequal size, but has the advantage, that the respective bins have always

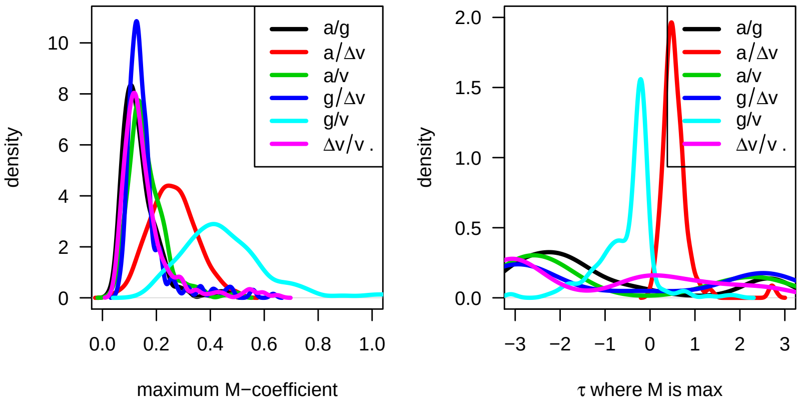

Figure 5: The distribution of the M–values for the six different combination of parameters, where now the maximum over $M(\tau)$ is used (left). The right plot shows the distribution of lags τ where the maximum is reached.

enough data-points in it and therefore the statistics is usually comparable. In each of the bins defined in this manner, the mean value of the acceleration as well as its standard deviation can be computed easily. This has been done for all the data that have been analyzed so far, the result is therefore something that may be called the average driver reaction, see Figure 7.

It could be seen, that the acceleration is a not too complicated function, and that it behaves completely in line with a naive expectation, at least in this average sense regarded here. For small g and very negative Δv, acceleration is negative, but at least on average the values are never dramatic: they stay well within the range $[-2, 1]$ m/s^2. For large g, acceleration values are more modest and not too different from zero. Only for small g and larger (positive) Δv, acceleration reaches (on average) 1 m/s^2. Note also the behavior of the standard deviation of the acceleration which may be called acceleration noise: it is in most of the areas near 0.5 m/s^2 and only for small g it becomes larger, demonstrating that in these areas individual accelerations may be well outside the range of the mean values.

In addition to this, a more quantitative approach is tried. As has been remarked [25], a linear model (originally proposed in [11]) is often a fairly good

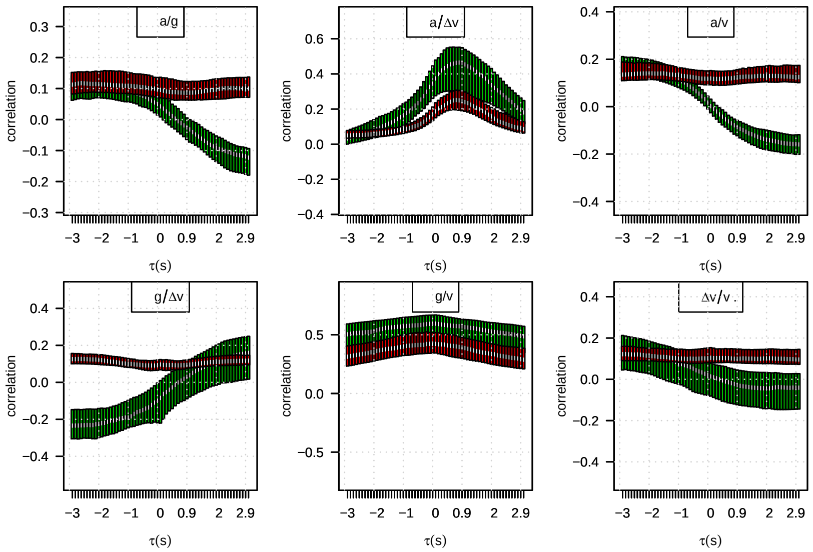

Figure 6: The six different average correlation functions. In green are the cross-correlation functions, while in red the mutual information function is displayed.

approximation to car following. Such a model can be specified as:

$$a = c_0 + c_{\Delta v}\Delta v + c_g g + c_v v \tag{9}$$

Interestingly, this equation is well-known to physics, since it is just a driven, damped, harmonic oscillator [24]. Furthermore, it is also the basis for most of the ACC algorithms [5, 13] (ACC is autonomous cruise control) and to a certain degree it is even used to control an automated vehicle when it follows another car. To test whether this is a good model, the data above can be used as follows: in each of the bins defined by the procedure above, a linear fit can be performed to this model, and the results, i.e. the four parameters of the model eq. (9) can then displayed as a function of (Δv, g). The result of this procedure, again performed on all the data, is displayed in Figure 8. Note, that the fitting yields in addition to the mean values of the parameters also its standard error and therefore a statistical quality. Therefore, in some cases some of the parameters are not significant and could be set to zero. However, this does not happen too often, so it has not been used here.

Note, that the parameters do not reflect the order that has been determined

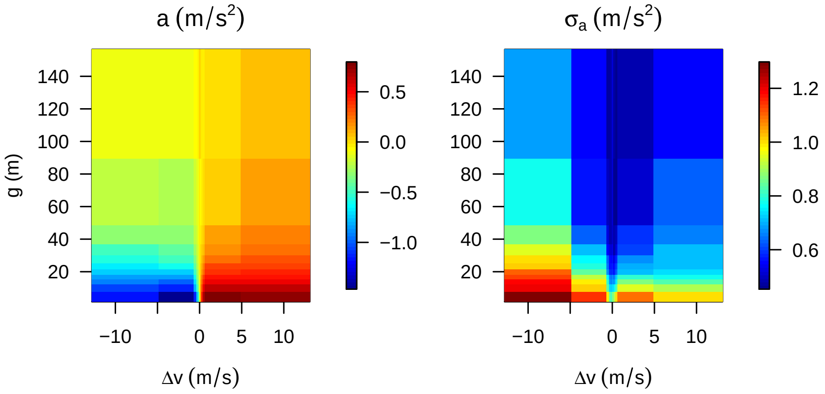

Figure 7: The acceleration, and the standard-deviation of the following vehicle as function of $(\Delta v,\, g)$.

by the mutual information approach. This is, however, one of the strengths of such a approach: the numerical value of an parameter depends on many things, e.g. the prefactor of the gap must be small since the gap itself is typically a large number, at least larger than the speed-difference. Even if these terms would be equally influential on the acceleration, this would not be reflected in their number. The mutual information provides a much more objective measure for the influence of the various parameters than the linear fit performed here.

It can be seen, that the linear model is not a too good an approximation to the modeling of car-following. If it were a perfect linear model, then the model parameters should be constants, which is obviously not the case. While this might be true to a certain degree for the parameters $(c_{\Delta v},\, c_g,\, c_v)$, it is definitely false for the parameter c_0. The function $c_0(\Delta v,\, g)$ is very similar to the acceleration itself, a fact for that so far we do not have any explanation.

3.3. Auto-correlation of the acceleration

As a final analysis, the auto-correlation function of the vehicles have been investigated. Again, this has been done both with the cross-correlation, as well as with the mutual information, see Fig. 9. And just as in Fig. 6 only the average function over all the data analyzed here is displayed. All in all, the outcomes of C and M are comparable, but it seems, that M displays a faster decrease

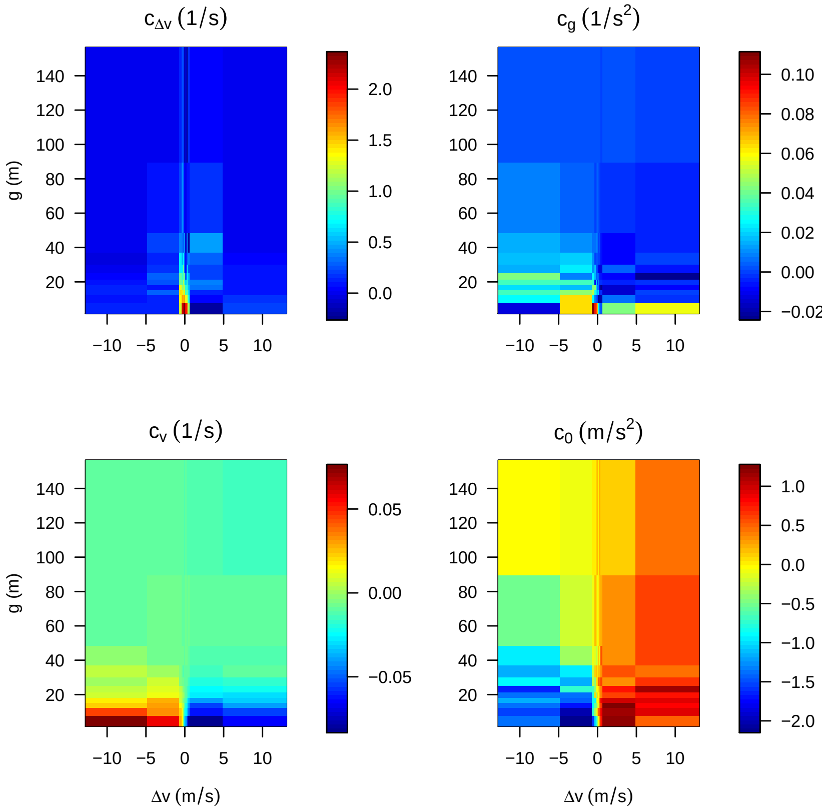

Figure 8: The parameters of the linear model eq. (9) as a function of (Δv, g).

of the correlation between $a(t)$ and $a(t-\tau)$ with respect to the lag τ. In this case, M is at the border already close to its theoretical minimum, therefore demonstrating that in this case the cross-correlation is assigning values where most likely no correlation is present anymore. Roughly, a correlation time-scale can be estimated to be 2 s, while the M-based correlation time is just one second.

4. Conclusions

These results show, that the acceleration a of a following vehicle depends strongly on the speed-difference $\Delta v(t-\tau)$, with an time-delay τ which is on average $\tau \approx 0.5$ s. Note, that this time-delay is not necessarily a reaction-time.

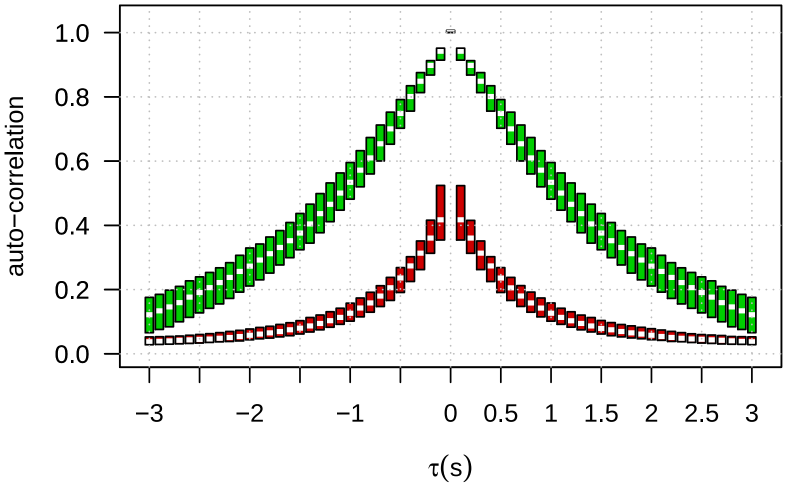

Figure 9: The autocorrelation function of the acceleration, measured by C (green) and M (red). Note, that M has been normalized so that its maximum is 1, which means dividing by $\ln n$ where n is the number of bins.

It is more likely related to the fact, that drivers cannot measure speed difference directly, they have to utilize the change of the object-size on their retina to estimate speed differences. This introduces a time-scale into the reasoning, since the driver's estimate of speed difference is given by $(g(t) - g(t - \tau))/\tau$. Note also, that this τ is not a constant either, since it depends on the variables $g, \Delta v, v$, where the dependence on v is certainly weak, and the one on g is the strongest, since the accuracy of the distance estimate δg decreases quadratically ($\delta g \propto 1/g^2$) for physical and physiological reasons. However, it was not possible to extract this type of dependency from the data.

The dependence of a on distance g and speed v is much weaker, and the results are compatible with a zero delay between a and (g, v). What is also strong is the relationship between g and v, it rivals and often exceeds the one between a and $\Delta v(t - \tau)$. However, this cannot be explained by a direct causal connection, it is a kind of indirect consequence of the driver's behavior: she uses the control which is mostly exerted via the speed difference to maintain the roughly linear relationship between v and g. Since this is an approximately linear dependency, with another time-constant, the preferred headway T is connecting the two ($v = g\,T$).

In addition to this, the results show that the acceleration itself has a memory of $T_a = 1 \ldots 2$ seconds. So, car-following can be described to a certain degree by three time constants (τ, T_r, T), and the strength of the interaction. However, it is not clear from this analysis, how the driver reaches the goal $v = g\,T$.

Note, that a forth time constant may be added to this picture, and that is the reaction time. Reaction-time, however, becomes only important when something unexpected happened, in most of cases what happens seems to be in line with the driver's forecast of the situation. This is difficult to prove, but some hint may be gained from the fact, that in this data-base (as has been analyzed so far) only quite normal deceleration values can be found. Out of the 2 million data-points only 335 (share: 1.4 10^{-4}) are below -5 m/s^2, so in fact drivers seem to handle their task quite smoothly.

So, if anything is taken together, and ignoring for the time being the true reaction time, the acceleration function in Eq. (1) might be written symbolically as:

$$a(\Delta v, g, v) = \frac{1}{T_a}\left(V\left(\frac{g(t) - g(t-\tau)}{\tau}, g - vT\right) - v\right) \qquad (10)$$

with a yet to determine function $V(\cdot)$. The estimation of the piecewise linear function above shows, that a linear function is a certain approximation, but it may in fact be not the final answer.

What may also be a bit surprising is that on this level of description, drivers share a number of similarities. E.g. the cross-correlation functions are very similar, this has in addition been confirmed by looking into more detail on the individual cross-correlation functions. These similarities do not hand over to the numerical figures, i.e. the distribution of the parameter values where they have been estimated are quite broad. Nevertheless, it seems that a kind of common driver model might be within reach, which will be an important result that may have a lot of applications.

An additional result of this analysis is that the application of the mutual information helps to sort out which correlations are real and which ones are spurious. Furthermore, it allows a kind of model-independent measurement of the strengths of the various terms in the acceleration equation. This would not have been so easily done by other means, and therefore these methods may add an important tool to the analyst's toolbox.

Acknowledgment

This work has been supported by DLR's basic research funds, especially the DLR-internal project I.MoVe. The data-acquisition within the simTD project has been funded by the German Ministry of Economy.

References

[1] K. I. Ahmed. *Modelling Drivers' Acceleration and Lane-Changing Behavior*. PhD thesis, MIT, 1999.

[2] D. Albanese, M. Filosi, R. Visintainer, S. Riccadonna, G. Jurman, and C. Furlanello. minerva and minepy: a c engine for the mine suite and its r, python and matlab wrappers. *Bioinformatics*, 29(3):407 – 408, 2012. first published online December 14, 2012.

[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] M. Campbell, M. Egerstedt, J.P. How, and R.M. Murray. Autonomous driving in urban environments: approaches, lessons, and challenges. *Philosophical Transactions of the Royal Society A*, 368:4649–4672, 2010.

[6] D. Chowdhury, L. Santen, and A. Schadschneider. Statistical physics of vehicular traffic and some related systems. *Physics Reports*, 329(4–6):199–329, 2000.

[7] R. J. Cowan. Useful headway models. *Transportation Research*, 9(6):371–375, 1976.

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

[9] V. Dhar. Data science and prediction. *Communications of the ACM*, 56(12):64 – 69, 2013.

[10] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73:1067–1141, 2001. see also arXiv.org/abs/cond-mat/0012229.

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

[12] Milan Krbalek and Dirk Helbing. Determination of interaction potentials in freeway traffic from steady-state statistics. *Physica A: Statistical Mechanics and its Applications*, 333(1-4):370–378, 2004.

[13] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J.Z. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun. Towards fully autonomous driving: Systems and algorithms. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 163–168, June 2011.

[14] T. Luttinen. Statistical properties of vehicle time headways. *Transportation Research Record*, 1365, 1992.

[15] Sven Maerivoet and Bart De Moor. Cellular automata models of road traffic. *Physics Reports*, 419(1):1 – 64, 2005.

[16] K. Nagel, P. Wagner, and R. Woesler. Still flowing: approaches to traffic flow and traffic jam modeling. *Operations Research*, 51(5):681–710, 2003.

[17] 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.

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

[19] 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.

[20] D. Reshef, Y. Reshef, H. Finucane, S. Grossman, G. McVean, P. Turnbaugh, E. Lander, M. Mitzenmacher, and P. Sabeti. Detecting novel associations in large datasets. *Science*, 334:6062, 2011.

[21] A. Reuschel. Fahrzeugbewegung in der Kolonne bei gleichförmig beschleunigtem oder verzögertem Leitfahrzeug. *Zeitschrift des österreichischen Ingenieur und Architektenvereins*, page 95, 7/8 1950. In German.

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

[23] Peter Wagner. Empirical description of car-following. In Serge P. Hoogendoorn, Stefan Luding, and Piet H. L. Bovy, editors, *Traffic and Granular Flow '03*, pages 15–28. Springer, 2005.

[24] Peter Wagner. A time-discrete harmonic oscillator model of human car-following. *European Physical Journal B*, 84(4):713–718, 2011.

[25] Peter Wagner, Gunnar Flötteröd, Ronald Nippold, and Yun-Pang Wang. Simplified car-following models. In *Proceedings of the Transportation Research Board*, pages 1–11. Transportation Research Board, January 2012. only on CD.

[26] Rainer Wiedemann. Simulation des Straßenverkehrsflußes. Technical report, Institut für Verkehrswesen, Universität Karlsruhe, 1974. Heft 8 der Schriftenreihe des IfV, in German.