

Peter Wagner¹, Ronald Nippold¹, Johannes Grötsch²

Abstract

By using a large data-base of single vehicle data from a German freeway, the Fundamental Diagram (FD) can be extended. These types of data allow to extract additional information that can be displayed as function of traffic flow Q and speed V . This is demonstrated here with a surrogate measure of safety (SMoS) and two other indices, which can be used to do more in-depth analyses of traffic flow. In case of the SMoS, it is possible to identify more and less dangerous regions in the extended FD.

Keywords

Fundamental diagram, traffic safety, surrogate measures of safety

Introduction

The Fundamental Diagram (FD in the following) describes the capacity of a road section and explains the relationship between traffic density and vehicle speed: The more vehicles enter a road section, the lower the local driving velocity. To keep traffic flow stable and avoid the occurrence of jams, no more vehicles should enter a road section than can leave it. When a critical vehicle density is reached, the corresponding vehicle speed drops and the state of the traffic flow changes from stable to unstable. For these reasons, the FD is of particular importance for the planning of traffic infrastructure and the design and adjustment of traffic management systems.

In the present paper, the FD is understood simply as a representation of measured speed data versus measured traffic flow data, which are often extracted from a local measurement of the passing of vehicles. Traditionally, and as with the data in this paper, this data collection is being done by loop detectors, but other means of measurement techniques such as video detection are possible as well. Since those local detectors cannot measure traffic density K and the relation $Q = KV$ is only valid for stationary and homogeneous flow, traffic density is not used here. Note that in the rest of this paper, capital letters Q, V, K are used for aggregated variables, while q, v denote microscopic values (single vehicle data).

The microscopic description of those data contains the following variables for each vehicle $i = 1, \dots, N$ crossing the detector:

- Time of passing t_i ,
- Speed of the vehicle v_i ,
- Length of the vehicle ℓ_i ,
- Type of the vehicle s_i ,
- Net time headway T_i ,
- Occupancy time Θ_i ,
- Gross time headway τ_i
(computed from $\tau_i = T_i + \Theta_i$).

The following approach is intended to be model- and theory-agnostic, i. e. it should be independent of the underlying preferences of the analyst, who may be in favor of one-¹, two-², three-³ or even more phases theory of traffic flow⁴.

The FD itself has been invented in the 1930s by Greenshields⁵, and it is not that long ago that it has celebrated its 75th anniversary⁶. There are many papers that dealt with it, so it seems that it has sparked the curiosity and interest of many analysts (to mention just a very few:⁷⁻¹⁰).

Typically, the FD is just displayed as it is, or it is tried to reduce its complexity, e. g. by fitting more or less fancy functions especially to the representation of traffic flow Q versus traffic density K . Doing so assumes some kind of equilibrium curve $Q(K)$ that traffic states will tend to. This may or may not be the case, it is conceivable that these equilibrium states are elusive since traffic flow, especially close to congestion, may be a deeply out of equilibrium phenomenon. What, however, is rarely done is to endow the FD with an additional analysis: E. g., the speed variance can be analyzed as a function of traffic flow and speed, or even the frequency distribution itself as function of (Q, V) .

This does not preclude the usage of such FD curves for traffic planning, as is used in handbooks as the HCM¹¹ in order to determine the right amount of highway capacity needed to meet a certain demand. Clearly, such a function can be determined through any cloud of data-points, and in this case, the analyst is free to use the most convenient function available for the purpose at hand.

¹ DLR – German Aerospace Center, Germany

² Landesbaudirektion Bayern, Zentralstelle für Verkehrsmanagement in Bayern (ZVM), Germany

Corresponding author:

Peter Wagner, DLR German Aerospace Center, Institute of Transportation Systems, Rutherfordstrasse 2, 12489 Berlin, Germany

Email: peter.wagner@dlr.de

Finally, from the field of research into traffic safety, there are a lot of approaches known that link traffic safety with traffic states, mostly with traffic flow, AADT, speed (in various aggregation grades), speed variance and geometric features (lane width, e. g.). It is difficult to summarize these approaches, they typically apply a certain generalized linear model (GLM) to a set of data and estimate the coefficients of the their model (see ¹² for the first, ¹³ for an overview of this approach). Often, the corresponding raw data are not presented, but only the fitted model data. Here, the approach is different: by defining a suitable SMOs that can be computed directly from the single vehicle data, at least such a SMOs can be presented as a function of the FD. Another approach, that is close to the one advocated here can be found in ^{14,15}. However, these approaches again do not take care of the structure of the FD. They assign a crash-rate to areas in the (Q, V) -plane that are not or only rarely visited by the system; one prominent example is the area between the two branches of the FD.

The approach

A data-set of ~ 23 M data-points has been used for the following analysis. The data have been sampled at the German freeway A92 North of Munich, all data come from a three-lane road section where variable speed-limits with values from the set 60, 80, 100 and 120 km/h were applied during the day. The measurements are from three sites (with three double loop detectors on each of the three lanes), and span the range from March to July 2018. All data analysis described in the following have been performed with R ¹⁶.

Bugs

The data have bugs, which hopefully do not impede the following analysis, but should be made explicit nevertheless. Be aware that there may be other bugs that have slipped the scrutiny of the analysts.

The first three can be seen by plotting the frequency distribution of the speeds and net headways. Some values are clearly favored or reduced in frequency above or below their neighboring values. No issues can be seen in the distribution of the occupancy, while the length-distribution shows some un-natural oscillations especially for larger lengths. In addition, some of the measured vehicles have very short lengths down to $\ell = 0.4$ m, which is un-realistic, too.

Note, that the raw values are discrete. The time-stamps have 1 s resolution, the speeds 1 km/h, the headways and occupancies 0.01 s and 0.001 s, respectively, and the vehicles lengths 0.1 m. The Figure 1 was created by directly sampling these raw values into histograms, no aggregation has taken place.

In addition to the discretization, storage and processing of data also introduce certain limits to the values. Speeds and lengths are 1 Byte integers, net headways and occupancies are 2 Byte integers. Therefore, speeds are limited to 0...254 km/h (the speed 255 indicates an error), lengths to 0...25.5 m, occupancies to 0...65.535 s and net headways to 0...655.53 s. The actual limits are close to these limits, e. g. speed takes values between 5...254 km/h.

Although they should, not all times are ordered correctly, which might be due to transmission slips between the

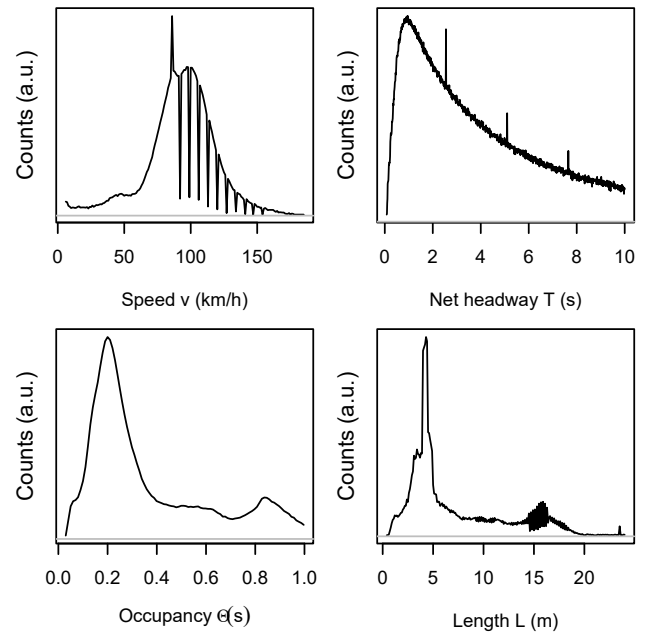


Figure 1. Frequency distributions of speed, net headway, occupancy, and length. Headway and occupancy have been restricted to values smaller than 10 or 1 s, respectively, and speeds to values below 200 km/h. The y -axis has been scaled with a squareroot to make all parts of the distribution visible.

detector and the Traffic Management Center. This happened not that often (3,600 times in the 23 M data-set); in most cases, the analysis below used the headways, and not the time-differences to compute the flows.

At least two additional tests of the internal consistency of the data are possible. The first one follows from the observation that the occupancy times Θ should read:

$$\Theta = \frac{\ell}{v}.$$

The data have these three variables separately, and in fact, there is a small difference between the expected value $\langle \ell/v \rangle$ and the one reported in the data-base (Θ), which is on average $\langle \epsilon_{\Theta} \rangle = -0.07$ s, given the errors ϵ_{Θ} are defined as follows:

$$\epsilon_{\Theta} = \frac{\ell}{v} - \Theta.$$

Another check can be done with the times. Clearly, between any two time stamps t_m and t_n , the sum of the gross headways between them should be equal to the time-difference:

$$\epsilon_t = t_n - t_m - \sum_{i=m+1}^n \tau_i.$$

This has been tested for t_m and t_n as the beginning and the end of each day, each station, and each lane, but other values can be used as well. The only condition is that start and end times must be long enough apart so that the different time resolutions do not cause problems. Here, the errors are larger, the times deviates over the course of a day by about 1000...2000 s: The sum of the gross headways is (on average) smaller than the corresponding time differences in the time stamps.

Microscopic FD and aggregation

Traffic flow Q (here, aggregated variables Q and speed V are denoted by capital letters) is defined as the number of vehicles per time, so the smallest unit of it is the inverse gross time headway τ_i of one vehicle:

$$q_i = \frac{1}{\tau_i} = \frac{1}{T_i + \Theta_i}. \quad (1)$$

The speeds can be used as they are, i.e. these are the speeds of the individual vehicles. This will be named in the following the raw (microscopic) FD.

The data could have been used as they are, see Figure 2 (left panel) for an example. Here, raw measurements from the left lane of one station, and for July 2018 only are shown. This plot, as well as the next ones are density plots that display how often a certain part of the FD was found in these data. This is clearly strongly dependent on the details of the

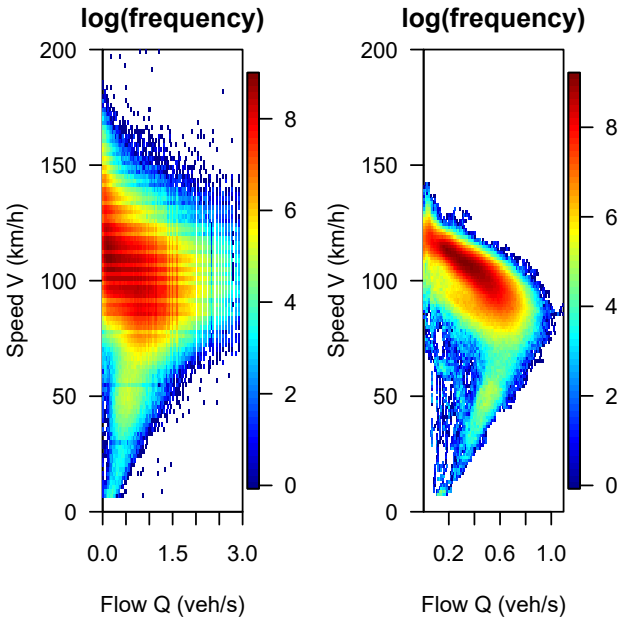


Figure 2. Conversion from the raw FD (left panel) as density plot to the aggregated FD (moving averages) as density plot (right panel). Data are from the left (fast) lane only. Flow values larger than 3 veh/s and speeds larger than 200 km/h have been omitted in the raw FD. The aggregated data have not been filtered. The color scale is logarithmic, which makes bins with small counts better visible.

demand function, i.e. the flow $Q(t)$ as function of time t over the course of a day, or the whole data range. Later on, in section , it will be shown how to get rid of this restriction.

To connect with the traditional approach, to reduce scattering, and, finally, to investigate features as a function of the FD-state, a bit of aggregation is introduced. To make full advantage of the microscopic structure of the data, a moving average approach is used. Therefore, to each microscopic data-point an aggregated speed V_i and an aggregated traffic flow Q_i is assigned, that are computed from a moving average:

$$V_i = \frac{1}{2w+1} \sum_{j=i-w}^{i+w} v_j \quad (2)$$

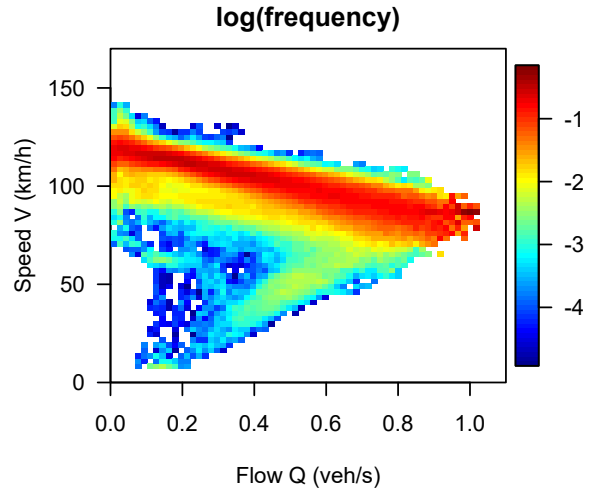


Figure 3. The normalized FD (with respect to flow); the colors are scaled logarithmically, so that boxes with small number of counts are better visible.

and

$$Q_i = \frac{2w+1}{\sum_{j=i-w}^{i+w} \tau_j}. \quad (3)$$

This smooths the FD and makes it possible to divide the (Q, V) -space into arbitrary “tilings”, and within each tile or box or bin, the data belonging to this tile can be sampled and analyzed. This yields for instance a function of the speed variance as the function of the traffic state that otherwise would be more difficult to get. Furthermore, each aggregated value of flow and speed now has roughly the same statistics, at least in terms of the number of data that went into it. This is different from the traditional approach of sampling into bins of equal duration.

The result is displayed in Figure 2 (right panel). The main difference is that the very large flow values that result from very short headways between individual vehicles are now averaged out, and the bugs in the speed distribution have been averaged out, too.

Getting rid of the demand-dependence

From the discussion above, it makes sense to normalize the data in a different way. So, instead of plotting the raw counts, it is better to normalize them by plotting the speed density distribution for each value of the flow, independent of how often this flow value has been in the data-set. This is done in Figure 3.

Results

Having displayed just the raw FD so far, in this section functions over the FD are investigated. This is done in the following for a selection of the possible displays. The safety of the states in the FD is analyzed by using a particular SMOs, then there is the preferred headway, and the degree to which the headway and the speed-difference distributions are independent, from which something about car-following behavior can be learned.

Traffic safety: SMOs as function of the FD

There are a number of SMOs (surrogate measures of safety)¹⁷ around that can be computed directly out of these microscopic data. The bad thing is that with this type of data, only the ones related to rear-end crashes can be investigated. In addition, it is not clear how to define safety if there is a whole distribution of values available as is the case here. What comes to mind is simply to use the minimum of the chosen SMOs, but this is often a problematic number since it is not very robust and subject to the strong noise in the data.

Here, different approaches have been used to get the safety of a particular state of the FD. We start with an indicator that is based on a safety criterion that is often used in microscopic traffic flow models, where one states that driving is safe, if the following condition holds:

$$\frac{v_i^2}{2b} + v_i\sigma_i \leq \frac{v_{i-1}^2}{2b} + g_i. \quad (4)$$

In Equation (4), v_i , v_{i-1} denote the speed of the following and leading vehicle, b , B are their preferred braking decelerations, g_i represents the net space headway between them, and σ_i is preferred minimum headway of the driver of vehicle i , which is often associated with a reaction time. Note that equation (4) does not guarantee crash-freeness in all cases, there is a small loop-hole if $b > B$, see¹⁸ for more details.

By turning this equation around and solving it for σ , a safety indicator results:

$$\sigma_i = \frac{v_{i-1}^2}{2Bv_i} - \frac{v_i^2}{2bv_i} + \frac{g_i}{v_i}. \quad (5)$$

If $b \approx B$ is assumed, then the resulting value σ_i is not strongly sensitive to the precise value of b , given it is not too small. A typical value for strong decelerations that is often used in traffic safety research is $b = 3.5 \text{ m/s}^2$, and this is the value that is used in the following. The term g_i is replaced by $T_i v_{i-1}$ (which are measured by the detector), and after some re-arrangements Equation (6) results where only measured quantities appear:

$$\sigma = \frac{(v_{i-1} - v_i)(v_{i-1} + v_i)}{2bv_i} + \frac{T_i v_{i-1}}{v_i}. \quad (6)$$

Any negative value of σ implies a strong danger of a rear-end crash, values between 0 and the maximum (which is at values of $\sigma = 1.1$, 0.99, and 0.88 s for $b = 1$, 3.5, 9) might be called dangerous, and the bulk of values beyond the maximum (about 3/4 of the data) can be considered as safe. This is a bit subjective, but this is the difficulty with most of the SMOs's. In this case, the subjectivity has at least a weak relationship with features that can be detected in the data. The distribution of this SMOs shows that there is a very small part of the data in the unsafe region where the distribution "looks wrong", i.e. very negative values might be either measurement errors or from situations, where a following driver is very close because she knows that the lead driver is going to change lanes and out of the way. From this distribution, a robust indicator would be not the smallest value, but the value where the steep increase of the distribution starts.

This is difficult to determine automatically, and especially in the set-up chosen here where a level of safety is sought that can be attributed to different parts of the FD. In this case, only a comparatively small number of data-points is available in each tile of the extended FD, so that it might not be possible to find the point where the distribution frazzles out. For the lack of a better idea, the share of negative values is picked in the hope that it would yield a clear safety pattern. In addition, it has been looked at 1%-percentile of σ as an additional indicator.

To connect this with more traditional SMOs, two of them have been investigated, too. These are the time-to-collision (TTC) and the deceleration-to-avoid-crash (DRAC). There are defined as:

$$\text{TTC} = \frac{v_{i-1}T_i}{v_{i-1} - v_i} \quad \text{if } v_{i-1} < v_i. \quad (7)$$

and:

$$\text{DRAC} = \frac{(v_{i-1} - v_i)^2}{2v_{i-1}T_i}. \quad (8)$$

For these two, the shares are determined where $\text{TTC} < 1.5$ s and $\text{DRAC} > 1.75 \text{ m/s}^2$, respectively.

The result is displayed in Figure 4. The behavior of the two indices based on σ introduced above is consistent (shown in the two upper part plots). The danger for rear-end crashes increases with the flow and the speed, while it is fairly small in the congested region of the FD, i.e. for small speeds below 80 km/h. The result e. g. in¹⁴ indicates an u -shaped function of safety as function of speed, and an increasing function as function of flow, but there all crash-types are subsummed, and not just the rear-end crashes.

The other two safety indicators are not consistent with σ : the shares of critical TTC- and DRAC-values do not show any pattern at all, at least for the values chosen, which is clearly different from the pattern displayed by σ .

The preferred headway $\hat{T}_0(q, v)$

Next, the preferred headway is analyzed, see Figure 5. It is defined here as the mode of the net time headway distribution $p(T)$, i.e. the value where this distribution has its maximum value. Clearly, its value is to be expected to be conditional on the traffic flow conditions, i.e. in free flow the preferred value of a driver is different from the value in dense or even congested traffic. On all lanes, it decreases with increasing demand, therefore displaying the compression for those denser traffic states. As a function of speed, there is no such clear dependence.

The lanes display slightly different patterns, however the most striking feature is the different maximum speed and flow for the different lanes which is to be expected from the fact that most trucks are found on the right-most lane, and the middle and left lane are the passing, the fast lanes.

A little surprising might be the fact that for congested conditions, the preferred headway increases again – the most dense states are to be found at the tip of the FD, near capacity. The value of this preferred headway can be as small as 0.6 s, see Figure 5.

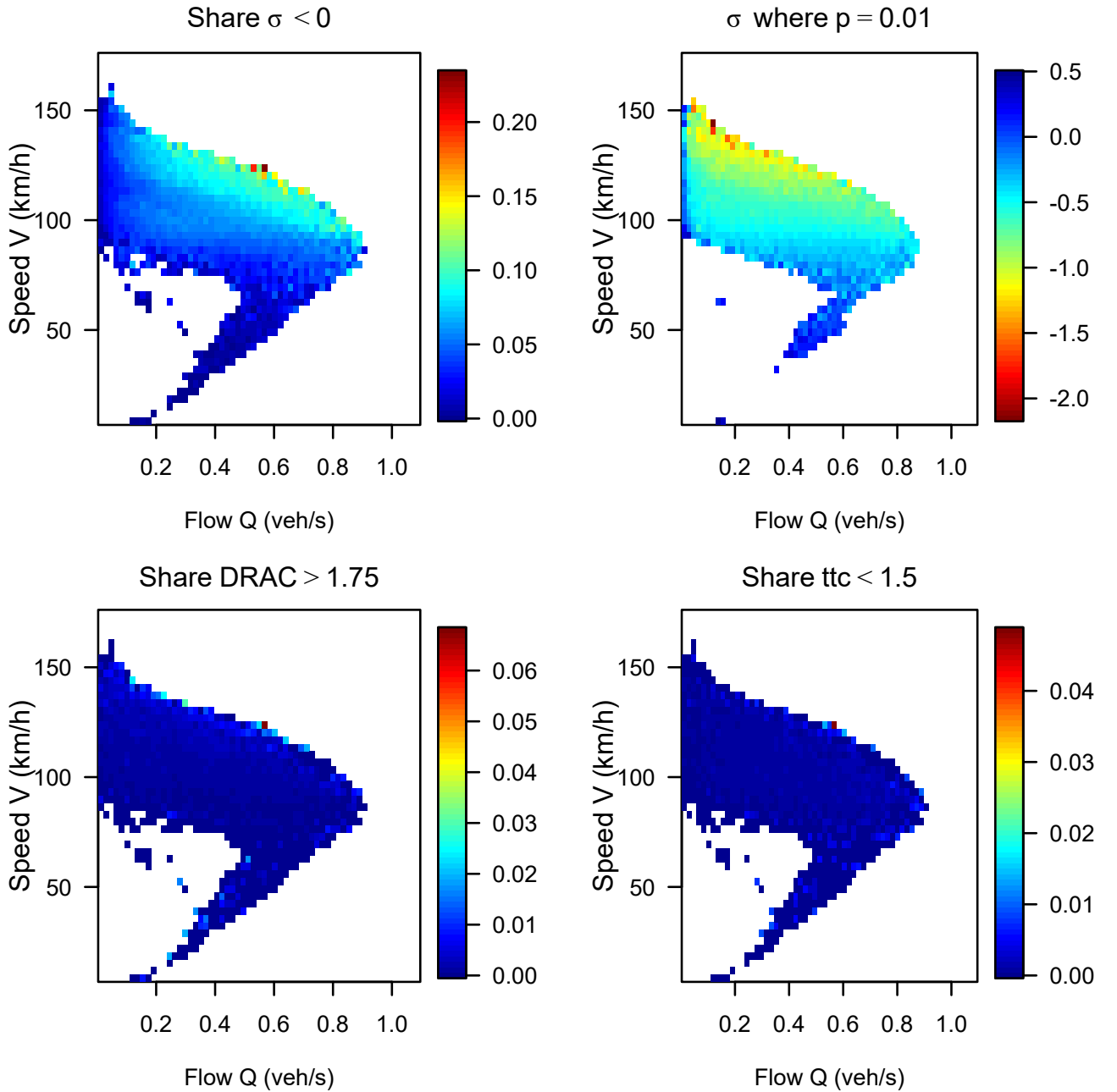


Figure 4. Behavior of four different SMoS: Upper left, the share of $\sigma < 0$ values, upper right, the σ -value belonging to the 1%-percentile, lower left the share of $\text{DRAC} > 1.75 \text{ m/s}^2$, and the share of TTC-values smaller than 1.5 s (lower right).

Segmenting the $p(\Delta v, g)$ -distribution

As a final example, a glimpse into the interactions between two following vehicles is performed. At least a static part of the interaction between two following vehicles is hidden in the $p(\Delta v, g)$ or $p(\Delta v, T)$ distribution, where $\Delta v_i = v_{i-1} - v_i$ is the speed difference between leading and following vehicle. Here, the analysis of $p(\Delta v, T)$ is preferred, since the computation of the distance between two vehicles can only be done by assuming that the lead vehicle does not change its speed v_{i-1} during the time interval $[t_{i-1}, t_i]$. The independence of the two marginal distributions $p(T)$ and $p(\Delta v)$ is tested by asking to which degree these two are independent of each other, i.e. whether $p(\Delta v, T) = p(\Delta v)p(T)$ holds. To answer this, the variables needs to be binned and counted, which results in a contingency table

n_{ij} , where i refers to Δv and j to T . Then, for each bin ij , the Pearson residuum χ_{ij} is computed from the two (binned) marginal distributions $n_{i\bullet} = \sum_{j=1}^J n_{ij}$, $n_{\bullet j} = \sum_{i=1}^I n_{ij}$ and $N = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$:

$$e_{ij} = \frac{1}{N} n_{i\bullet} n_{\bullet j}, \quad (9)$$

$$\chi_{ij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}. \quad (10)$$

This is in essence a χ^2 -test which shows that the two distributions are not independent of each other. What is more interesting is the behavior of the Pearson residuals when displayed as a function of $(\Delta v, T)$, see Figure 6.

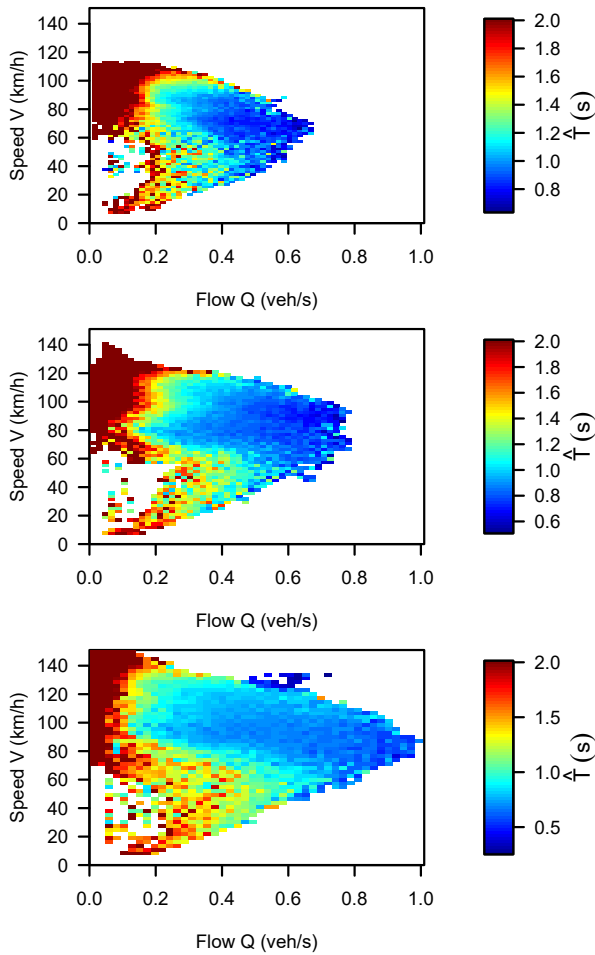


Figure 5. The mode \hat{T} of the headway distribution as function of flow and speed. Values of \hat{T} are confined to the interval 0.6...2 s, from top to bottom data are from the right, middle, and left lane.

The areas in red in Figure 6 have a strong negative residuum; this indicates that the probability to find values of $(\Delta v, T)$ in this system region is much smaller than what can be expected from the assumption of independence. And, in the same manner, the areas in green are the ones to be visited more often. All these results are highly significant, $|\chi_{ij}| > 4$ is, as a rule of thumb, on a significance level of 0.01.

There may be different reasons for this distribution: the lower left corner is clearly a very dangerous area, so this is to be avoided by the driver. Large speed differences and large headways are preferred, as well as the small headway and small speed difference. A bit surprising might be the fact that large headways and small speed differences are less likely to be visited, which could be interpreted as an area where it is difficult to stay there: human drivers are not the best to balance a small Δv when the distance is large. It will be interesting to compare this to a pattern produced by a microscopic simulation model, and to trajectory data from Naturalistic Driving Studies.

Conclusions

The methods here are generalizable to other places, the results may be not, although it is highly likely that very similar findings will turn out when using data from other

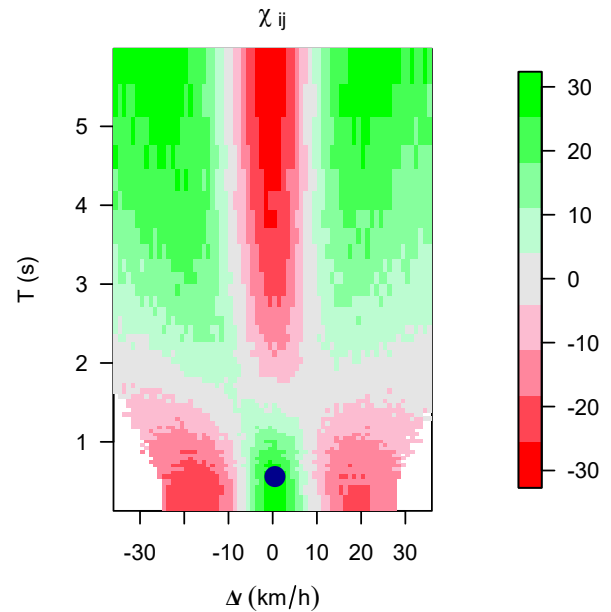


Figure 6. Pearson residuals χ_{ij} as function of speed difference Δv and headway T between two vehicles. The blue filled circle is at the maximum of the $p(\Delta v, T)$ -distribution.

places. It should be mentioned, that the results presented here are just a part of the total results obtained.

Three results should be mentioned: it was possible to assign to the FD a SMoS that displayed reasonable results, with the limitation to rear-end crashes, which is due to a limitation in the data itself. Note, however, that different SMoS display different patterns, which points to the fact, that there is research needed: a minimum requirement for SMoS should be that they have the same meaning about what is dangerous and what not.

It is difficult to see how SMoS's for other crash-types could be extracted from this type of data; one may speculate, that the analysis of the gaps on neighbouring lanes might be used to measure the difficulty of a lane-change, so something might be said about crashes related to lane-changing.

Also interesting is that there seems to be a smooth transition between free flowing states and the very dense high-flow state, and further, to the more congested states, but here the statistics is weak since this freeway has very few traffic jams. However, especially the high-flow state is a very interesting state in its own right, since vehicle movements must be well tuned to stabilize this region. We have the feeling that this area has been a little bit neglected over the research into traffic flow breakdowns.

Finally, we found it a bit surprising, that something could be learned about the microscopic interaction between two following vehicles. The tool used here for this purpose demonstrated that following drivers have certain preferred places in $(\Delta v, T)$ -plane, and there are places that drivers tend to avoid. This is on top of the general distribution $p(\Delta v, T)$ of values in the car-following phase-space, and it remains to be compared with trajectory data, and with car-following models to better understand what is going on here.

Acknowledgements

The data that have been used here are from Freeway Directorate South-Bavarian in Germany, which is gratefully acknowledged here.

Declaration of conflicting interests

The authors do not have any conflicts of interest to declare.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: P. Wagner; data collection: J. Grötsch; analysis and interpretation of results: P. Wagner, R. Nippold; draft manuscript preparation: R. Nippold, P. Wagner, J. Grötsch. All authors reviewed the results and approved the final version of the manuscript.

References

- Lighthill MJ and Whitham JB. On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London Series A Mathematical and Physical Sciences* 1955; 229(1178): 281–345. DOI:10.1098/rspa.1955.0089. URL <http://rspa.royalsocietypublishing.org/content/229/1178/317.short>. <http://rspa.royalsocietypublishing.org/content/229/1178/317.full.pdf+html>.
- Bando M, Hasebe K, Nakayama A et al. Dynamical model of traffic congestion and numerical simulation. *Physical Review E* 1995; 51: 1035–1042.
- Kerner BS and Rehborn H. Experimental properties of phase transitions in traffic flow. *Phys Rev Lett* 1997; 79(20): 4030–4033.
- Treiber M, Kesting A and Helbing D. Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts. *Transportation Research Part B: Methodological* 2010; 44(8–9): 983–1000. DOI:10.1016/j.trb.2010.03.004. URL <http://www.sciencedirect.com/science/article/pii/S019126151000041X>.
- Greenshields B. The photographic method of studying traffic behaviour. In *Proceedings of the 13th Annual Meeting of the Highway Research Board*.
- Kühne RD and Gartner NH (eds.). *75 Years of the Fundamental Diagram for Traffic Flow Theory - Greenshields Symposium*, volume E-C149. Transportation Research Board, 2011. URL <http://onlinepubs.trb.org/onlinepubs/circulars/ec149.pdf>.
- Aerde MV. Single regime speed-flow-density relationship for congested and uncongested highways. In *74th TRB Annual Conference, Washington D.C.,*.
- Lu X, Varaiya P and Horowitz R. Fundamental diagram modelling and analysis based ngsim data. *IFAC Proceedings Volumes* 2009; 42(15): 367–374. DOI: <https://doi.org/10.3182/20090902-3-US-2007.0068>. URL <http://www.sciencedirect.com/science/article/pii/S1474667016318213>. 12th IFAC Symposium on Control in Transportation Systems.
- Li J and Zhang HM. Fundamental diagram of traffic flow new identification scheme and further evidence from empirical data. *Transportation Research Record: Journal of the Transportation Research Board*, 2011; 2260: 50–59. DOI: 10.3141/2260-06.
- Knoop VL and Daamen W. Automatic fitting procedure for the fundamental diagram. *Transportmetrica B: Transport Dynamics* 2017; 5(2): 129–144. DOI:10.1080/21680566.2016.1256239. URL <https://doi.org/10.1080/21680566.2016.1256239>. <https://doi.org/10.1080/21680566.2016.1256239>.
- Transportation Research Board. Transportation Research Board Highway Capacity Manual. Technical report, 2010.
- Veh A. Improvements to reduce traffic accidents. *Proceedings ASCE* 1937; 7: 1775–1785.
- Lord D and Mannering F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice* 2010; 44: 291–305. DOI:10.1016/j.tra.2010.02.001.
- Yu R, Quddus M, Wang X et al. Impact of data aggregation approaches on the relationships between operating speed and traffic safety. *Accident Analysis & Prevention* 2018; 120: 304–310. DOI:<https://doi.org/10.1016/j.aap.2018.06.007>. URL <http://www.sciencedirect.com/science/article/pii/S0001457518302604>.
- Imprialou MIM, Quddus M, Pitfield DE et al. Revisiting crash-speed relationships: A new perspective in crash modelling. *Accident Analysis & Prevention* 2016; 86: 173–185. DOI:<https://doi.org/10.1016/j.aap.2015.10.001>. URL <http://www.sciencedirect.com/science/article/pii/S000145751530083X>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL <https://www.R-project.org/>.
- Tarko AP. Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis & Prevention* 2012; 45: 230–240. DOI:<https://doi.org/10.1016/j.aap.2011.07.008>. URL <http://www.sciencedirect.com/science/article/pii/S000145751100193X>.
- Lücken L. Resolving collisions for the gipps car-following model, 2019. [1902.04927](https://arxiv.org/abs/1902.04927).

Figure captions

Figure 1

Frequency distributions of speed, net headway, occupancy, and length. Headway and occupancy have been restricted to values smaller than 10 or 1 s, respectively, and speeds to values below 200 km/h. The y -axis has been scaled with a squareroot to make all parts of the distribution visible.

Figure 2

Conversion from the raw FD (left panel) as density plot to the aggregated FD (moving averages) as density plot (right panel). Data are from the left (fast) lane only. Flow values larger than 3 veh/s and speeds larger than 200 km/h have been omitted in the raw FD. The aggregated data have not been filtered. The color scale is logarithmic, which makes bins with small counts better visible.

Figure 3

The normalized FD (with respect to flow); the colors are scaled logarithmically, so that boxes with small number of counts are better visible.

Figure 4

Behavior of four different SMoS: Upper left, the share of $\sigma < 0$ values, upper right, the σ -value belonging to the 5%-percentile, lower left the share of DRAC $> 1.75 \text{ m/s}^2$, and the share of TTC-values smaller than 1.5 s (lower right).

Figure 5

The mode \hat{T} of the headway distribution as function of flow and speed. Values of \hat{T} are confined to the interval 0.6...2 s, from top to bottom data are from the right, middle, and left lane.

Figure 6

Pearson residuals χ_{ij} as function of speed difference Δv and headway T between two vehicles. The blue filled circle is at the maximum of the $p(\Delta v, T)$ -distribution.