IMPLICATIONS OF THE RANDOM CAPACITY CONCEPT FOR FREeways

Werner BRILON
Institute for Transportation and Traffic Engineering, Ruhr-University Bochum, Germany, werner.brilon@rub.de

Justin GEISTEFELDT
Hessian Road and Traffic Authority, Frankfurt, Germany, justin.geistefeldt@hsvv.hessen.de

Citation:

ABSTRACT
In this paper, the capacity of a freeway facility is described as a random variable. A methodology for the estimation of capacity distribution functions based on empirical data is introduced. Capacity analysis of freeway sections with long-term work zones is presented as an example for the application of the methodology. It is shown that a freeway operates at the highest expected efficiency if the demand volume reaches 90% of the conventional design capacity. Moreover, the probabilistic concept is included into a model for the analysis of traffic flow quality over a whole year. This technique also provides a quantitative assessment for oversaturated conditions.

1. INTRODUCTION
The Highway Capacity Manual (HCM 2000) is the most prominent document for the traditional understanding of the capacity of a highway facility. The HCM (2000) defines the capacity as the maximum possible expected throughput of the highway element under consideration. For well-defined external conditions – including the geometric layout – this capacity is treated as a constant value. This kind of definition is used in most of the traffic engineering guidelines around the world (e.g. the German HBS 2001).

The users of such manuals in practice usually believe in the precision of these guidelines. It is, however, easy to comprehend that capacities indicated in guidelines do not describe real world traffic conditions with a precision up to the last digit. Instead, the given capacities only offer a rough estimation of realistic maximum flows. The question is, however, to which extent the actual maximum traffic volumes differ from the given expectations.

Several authors (e.g. Elefteriadou et al. 1995; Minderhoud et al. 1997; Persaud et al. 1998; Lorenz and Elefteriadou 2001) have already demonstrated that the maximum traffic throughput of a freeway facility varies – even under constant external conditions. To determine the variability of capacities, most of these authors observed breakdowns of traffic flow and identified the traffic volume during the time interval preceding the breakdown. This volume, which seemed to cause the breakdown, varied rather significantly. These observations made it obvious that the maximum possible throughput could deviate from given capacities over a remarkably wide range.

In-depth investigations about these observations led to a new understanding of the capacity as a random variable. This paper summarizes these findings and the visible consequences for practical application. It mentions the mathematical background as well as empirical results based on a broad database from Germany. The whole concept is demonstrated with freeway data. The stochastic approach provides new measures of traffic flow performance based on aspects of traffic reliability.
2. STOCHASTIC CONCEPT OF CAPACITY

The HCM (2000) defines the capacity as the maximum flow rate that can reasonably be expected to traverse a facility under prevailing roadway, traffic, and control conditions. This definition can be modified without changing the original meaning too much: Capacity is the maximum flow rate that achieves acceptable traffic performance of the facility and beyond which – in case of greater demand – proper operation fails. The transition between proper operation and non-acceptable flow conditions is called a breakdown. On a freeway, such a breakdown is characterized by a reduction of the average travel speed from an acceptable level of about 70 km/h or more to a much lower value representing congested conditions. On German freeways, these transitions usually occur with a rather sudden speed reduction.

From this definition, it becomes comprehensible that a capacity in this sense is by no means a constant value. In a critical traffic situation on a homogeneous freeway section, the breakdown can be caused by specific kinds of behaviors of single drivers together with the local constellation of the traffic flow pattern. Thus, not only the macroscopic traffic flow parameters like volume or density determine the breakdown. Also unpredictable events like speed reductions of individual drivers or lane changes may cause a deceleration of following vehicles and, in consequence, a local concentration that could initiate a breakdown. The occurrence of such events has all properties of randomness. Therefore, the traffic volume at which a breakdown is initiated and which is a realization of the capacity can be treated as a random variable.

To make this concept of randomness of capacity usable, it is necessary to know more about the distribution function of the capacity. Some authors (Kuehne and Mahnke 2005) have tried to derive the type of function by theoretical considerations. This requires specific assumptions about traffic flow characteristics. However, to get a real world picture, empirical investigations are required.

A methodology for the derivation of capacity distribution functions for freeways was implemented by Brilon et al. (2005, 2007). This method is based on mathematical solutions obtained from the statistics of lifetime data analysis (cf. e.g. Lawless 2003). It delivers a numerical solution for the cumulative distribution function of the capacity $c$:

$$ F_c(q) = p(c \leq q) \quad (1) $$

where:

- $F_c(q)$ = capacity distribution function (-)
- $c$ = capacity (veh/h)
- $q$ = traffic volume (veh/h)

Traffic flow observations on freeways deliver pairs of values of traffic volumes and average speeds during predetermined observation intervals (index $i$). For capacity analysis, the following types of intervals can be distinguished:

B: Traffic in interval $i$ is fluent, but the observed flow causes a breakdown; i.e. the average speed drops down to a level of less than a specific threshold speed in the following time interval $i + 1$. The volume in interval $i$ is regarded as a realization of the capacity $c$.

F: Traffic is fluent in interval $i$ and the following interval $i + 1$. Thus, the capacity in interval $i$ is above the observed volume $q_i$. This type of data is called “censored data”.

C: Traffic is congested in interval $i$ and the preceding interval $i - 1$, i.e. the average speed is below the threshold value. This interval $i$ provides no information about the capacity. It is not further regarded.
The differentiation of fluent and congested traffic is based on a specific threshold speed, which marks the gap between the upper and the lower branch of the speed-flow diagram. Values between 70 and 80 km/h were found to be fairly representative for German freeways. This threshold value may, however, be different for other types of roads. In some cases, more detailed approaches to identify traffic breakdowns might be required.

The intervals of classification “C” are not considered for analysis because volumes observed under congested flow conditions do not contain any information about the capacity. Other authors (van Toorenburg 1986; Minderhoud et al. 1997) defined the breakdown capacity as the traffic volume measured downstream of a queue at a bottleneck. In consequence, each interval during congestion was regarded as a “B”-interval. However, the maximum volume in fluent traffic usually differs from the maximum volume observed during congestion (so-called “capacity drop”; cf. e.g. Banks 1990; Hall and Agyemang-Duah 1991; Ponzlet 1996; Zurlinden 2003; Regler 2004). Therefore, the consideration of congested intervals seems not to be reasonable.

The statistics of lifetime data analysis provide methods to estimate distribution functions based on samples that include censored data. A non-parametric method to estimate the distribution function of lifetime variables is the so-called “Product Limit Method” (PLM) by Kaplan and Meier (1958). Based on this approach, the capacity distribution function can be estimated by:

\[ F_c(q) = 1 - \prod_{i; q_i < q} \frac{k_i - d_i}{k_i}; \quad i \in \{B\} \]  

where:  
- \( F_c(q) \) = capacity distribution function  
- \( q \) = traffic volume (veh/h)  
- \( q_i \) = traffic volume in interval \( i \) (veh/h)  
- \( k_i \) = number of intervals with a traffic volume of \( q \geq q_i \)  
- \( d_i \) = number of breakdowns at a volume of \( q_i \)  
- \( \{B\} \) = set of breakdown intervals (classification “B”, see above)

The product within this equation is calculated over all observed time intervals \( i \) with a traffic volume \( q_i < q \) that were followed by a traffic breakdown. Usually, each observed breakdown is used as one \( q \)-value, so that \( d_i \) is always equal to 1.

The distribution function will only reach a value of 1 if the maximum observed volume was a B-value (i.e. a breakdown was following). Otherwise the distribution function will be terminated at a value of \( F_c(q) < 1 \). In this case, the method does not allow to estimate the function \( F_c(q) \) completely.

To receive a complete distribution function, a parametric estimation based on the Maximum-Likelihood technique (cf. e.g. Lawless 2003) can be applied. In this case, the type of the distribution function must be predetermined. A comparison between different mathematical types of functions revealed the best results for the Weibull distribution (Zurlinden 2003). The Weibull-type capacity distribution and density function are:

\[ F_c(q) = 1 - e^{-\left(\frac{q}{\beta}\right)^\alpha} \]  

\[ f_c(q) = \frac{\delta F}{\delta q} = \frac{\alpha}{\beta^\alpha} \cdot q^{\alpha-1} \cdot e^{-\left(\frac{q}{\beta}\right)^\alpha} \]
where: \( F_c(q) \) = distribution function of the random variable “capacity” 
\( f_c(q) \) = statistical density function of the random variable “capacity” 
\( q \) = traffic volume 
\( \alpha \) = shape parameter 
\( \beta \) = scale parameter

The expectation \( E(c) \) and variance \( \sigma^2(c) \) of the capacity distribution are given by:

\[
E(c) = \beta \cdot \Gamma\left(1 + \frac{1}{\alpha}\right) \\
\sigma^2(c) = \beta^2 \cdot \left\{ \Gamma\left(1 + \frac{2}{\alpha}\right) - \frac{1}{\Gamma\left(1 + \frac{1}{\alpha}\right)^2} \right\}
\]

where: \( \Gamma(x) \) = Gamma function at point \( x \)

The parameters \( \alpha \) and \( \beta \) of the distribution function can be estimated by maximizing the Likelihood Function \( L \) (or its natural logarithm):

\[
L = \prod_{i=1}^{n} f_c(q_i)^{\delta_i} \cdot \left[l - F_c(q_i)\right]^{1-\delta_i}
\]

where  
\( f_c(q_i) \) = statistical density function of capacity \( c \) 
\( F_c(q_i) \) = cumulative distribution function of capacity \( c \) 
\( n \) = number of intervals 
\( \delta_i \) = 1, if interval \( i \) contains an uncensored value (classification “B”) 
\( \delta_i \) = 0, if interval \( i \) contains a censored value (classification “F”)

For the stochastic analysis of freeway capacity, only rather short observation intervals are useful. Otherwise there is only little causality between the traffic volume and the breakdown. Therefore, e.g. 1-hour counts are not adequate. With respect to the availability of loop detector data and in consideration of the use of the results, particularly for application within the whole-year-analysis concept (cf. section 5), 5-minute intervals were found to be a good compromise (Zurlinden 2003).

For the estimation of capacity distribution functions based on data measured at a specific cross section, it must be ensured that all traffic breakdowns observed at this point were caused within the freeway section under observation. At clearly detectable bottlenecks as shown in Figure 1, breakdowns observed directly upstream of the bottleneck should only be generated due to oversaturation of the bottleneck itself. Spillback from downstream should not occur since a larger capacity is always available on the succeeding section. At freeway sections without a distinct bottleneck as shown in Figure 2, traffic breakdowns due to a spillback from downstream should be excluded by analyzing a second cross section downstream of the section under investigation (cf. Regler 2004). If the speed data of the downstream cross section indicates congested flow at the time when a traffic breakdown is observed at the measurement cross section, this breakdown event is not considered for the capacity analysis.
As an example, Figure 3 shows the estimated capacity distribution function for a 2-lane section of the freeway A57 near Krefeld, Germany, representing a virtual bottleneck as indicated in Figure 2. Speed-flow data in 5-minute intervals for the whole year 2002 were analyzed. A total of 182 traffic breakdowns was observed. As the highest observed volumes were not followed by a breakdown, the Product-Limit estimation already ends at a cumulative probability of 0.54. With the Maximum-Likelihood technique, a Weibull-type distribution with parameters $\alpha = 13.9$ and $\beta = 4905$ veh/h was estimated. The parametric capacity distribution fits very well into the Product-Limit estimation.

Figure 4 shows the estimated capacity distribution function for the 3-lane carriageway of the freeway A3 east of Frankfurt, Germany. This section is equipped with a variable speed limit system. The estimation was based on speed-flow-data for the year 2004, including 118 breakdowns of traffic flow. The Maximum-Likelihood technique delivers a Weibull distribution with parameters $\alpha = 21.9$ and $\beta = 6874$ veh/h. Compared to the 2-lane freeway A57, the standard deviation of the distribution function is significantly lower, which could be attributed to the impact of the variable speed limit towards a more homogeneous traffic flow at high volumes.

3. APPLICATIONS
The methodology for the stochastic capacity analysis was applied to a number of freeway sections in Germany (cf. e.g. Zurlinden 2003; Regler 2004, Geistefeldt 2007). The shape parameter $\alpha$, which determines the variance of the Weibull-type capacity distribution function, typically ranges between 12 and 22. The scale parameter $\beta$ mainly depends on the prevailing geometric and traffic conditions of the freeway (e.g. number of lanes, grade, driver population: long distance traffic versus commuters).

With the stochastic concept of capacity, a variety of specific aspects of freeway traffic flow can be investigated. The approach was e.g. used to investigate differences in performance between dry and
wet road surface (Zurlinden 2003; Regler 2004). At all studied sections it turned out very clearly that on wet road surface the capacity was reduced by around 12% compared to dry conditions. Also the effects of darkness were investigated. Contrary to Ponzlet’s (1996) results from 1-h-counts, it could be shown that darkness did not shift the capacity distributions.

As an example for the application of the stochastic concept, the capacity of freeway sections with work zones is analyzed in this paper. The stochastic method was applied to long-term work zones on the freeways A3 and A5 west of Frankfurt, Germany. At the freeway A3 with three lanes each direction, a “4 + 2” scheme was used during the construction activities, which means that one lane of the 3-lane northbound carriageway was crossed over to the opposite carriageway (now carrying 4 lanes) and two lanes were remaining in the reconstruction area. All lanes in the work zone area were significantly narrower than normal lanes. A speed limit of 80 km/h was valid in the work zone. Traffic data were obtained from a cross section on the southbound carriageway directly upstream of the beginning of the work zone for an 83 day period. At the freeway A5, two lanes of the 3-lane northbound carriageway were crossed over to the opposite carriageway with one lane remaining on the side of reconstruction. Traffic data over a 62 day period during construction work were analyzed for the southbound carriageway with three narrowed lanes.

For both freeways, traffic flow during the time period with work zone traffic was compared to time periods with undisturbed traffic flow in the same section. Only times with dry conditions were considered for the analysis. Figures 5 and 6 show the estimated capacity distribution functions in 5-minute intervals for the freeways A3 and A5, respectively.

The comparisons of the capacity distributions show the significant impact of both work zones on traffic flow. The median of the Weibull-distribution is reduced by 9.4% on the freeway A3 and by 16.7% on the freeway A5. The difference in the extent of the capacity reduction mainly arises from different capacity distributions for the time periods without work zone. As both sections have similar geometric conditions, this difference is assumed to be attributed to the higher proportion of commuters on the freeway A5. On both freeway sections, the standard deviation of the capacity distribution is significantly lower during the time period with work zone. One reason for this remarkable result might be the effect of the speed limit in the work zone area.

The results of these two examples and other freeway sections with work zones analyzed in the same way show that the median of the capacity distribution of a work zone with three (narrowed) lanes is in a range of 6000 to 6300 veh/h for 5-minute intervals. The capacity at the work zone depends on
the extent of the lane width reduction and the existence of a transition of lanes to the opposite
carriageway. Traffic flow within the work zone seems to be more homogeneous as indicated by a
reduced standard deviation of the capacity distribution function.

4. TRAFFIC EFFICIENCY
Brilon (2000) has proposed to use the parameter

\[ E = q \cdot v \cdot T \] (8)

where

- \( q \) = volume (veh/h)
- \( v \) = travel velocity over an extended section of the freeway (km/h)
- \( T \) = duration of the time period for analysis of flow (h)

as a measure to characterize the efficiency for the usage of a freeway section. This parameter
describes the product of a freeway how it is achieved per time unit. The more veh · km a freeway
produces per hour, the more efficient the existing infrastructure’s potential is exploited.

If we follow the traditional concept of a deterministic \( q-v \)-relation for the average velocity
depending on volume \( q \), then with typical German \( q-v \)-diagrams (i.e. significantly reduced \( v \) with
increasing \( q \)), the maximum of \( E \) is obtained at \( q = d \cdot c \approx 0.9 \cdot c \), where \( c \) is the traditionally defined
capacity. The factor \( d \), however, tends to 1 for the more flat \( q-v \)-dependency as it is observed under
speed limit conditions like in the USA (e.g. HCM, 2000).

If we leave the deterministic \( q-v \)-curves behind us and apply the concept of randomness, then we
have to combine each volume \( q \) with its probability of a breakdown. Brilon and Zurlinden (2003)
have derived:

\[ E_{\text{exp}}(q_D) = [q_D \cdot v \cdot (1 - \chi) + q_{dc} \cdot v_{dc} \cdot \chi] \cdot T \] (9)

where

- \( E_{\text{exp}}(q_D) \) = expected efficiency at a demand volume \( q_D \) (veh · km/h)
- \( q_D \) = demand traffic volume (veh/h)
- \( v \) = average velocity in fluent traffic for \( q = q_D \) (km/h)
- \( q_{dc} \) = queue discharge volume (veh/h)
- \( v_{dc} \) = queue discharge velocity (km/h)
- \( T \) = duration of the period under investigation (h)
- \( \chi \) = expected proportion of congested intervals with \( q = q_{dc} \) (-)

\[ \chi = \sum_{i=1}^{n} \sum_{k=0}^{m} p_B(q_{i-k}) \cdot (1 - p_{\text{cong},i-k-1}) \] (10)

- \( n \) = number of 5-minute intervals during \( T \) (-)
- \( m = \min \left\{ \frac{t_{\text{cong}}}{5} - 1, i - 1 \right\} \) (-)
- \( t_{\text{cong}} \) = average duration of a congested period (min)
- \( p_B(q) \) = probability of a breakdown at volume \( q \) (here: \( q = q_0 \)) (-)
  (e.g. after eq. (2): \( p_B = dF_i(q)/dq \) or after eq. (4): \( p_B = f_i(q) \))
- \( p_{\text{cong},i} \) = probability of congested flow in 5-minute interval \( i \)
\[ p_{\text{cong},i} = \sum_{k=0}^{m} p_B(q_{i-k}) \cdot (1 - p_{\text{cong},i-k-1}) \]

If this set of equations is applied with real data for \( p_B(q) = f_c(q) \) (\( f_c(q) \): eq. (4)), then it becomes clear that a maximum of the expected efficiency is achieved for a demand volume \( q_D \) that should be lower than the average capacity. Sample calculations show that the highest efficiency of a freeway is to be expected for around \( 0.9 \cdot c \), where \( c \) is the traditionally defined capacity or the 1-hour average of the Product Limit estimated capacity.

As an example for practical application this means: A toll freeway operator should control the input flow in such a way that at no time more than 90% of the nominal freeway capacity is entering the toll road. This maximizes his expected toll revenues. Of course, this would also be a useful strategy for a public freeway as well.

5. WHOLE-YEAR-ANALYSIS

Traditionally, the technical layout of highway facilities is based on the analysis of one specific peak hour. The expected traffic demand during this single peak hour is compared with the capacity to assess the quality of traffic flow. The result is regarded as the representation of traffic performance of the facility. This basic concept is used in the HCM (2000) as well as in other comparable guidelines (e.g. HBS 2001). However, the analysis of one peak hour can not reflect the whole life cycle of a road infrastructure. In particular, this concept does not allow for a detailed assessment of overload impacts as the highest demand values arising during one year are not considered. In addition, the concept does not provide an assessment for overloaded conditions (LOS F).

In order to overcome these limitations, Brilon (2000) proposed to assess traffic flow performance over a whole year instead of the analysis for one single peak hour. Here it is useful to include also the stochastic properties of both capacity and traffic demand. A basic concept for the Whole-Year-Analysis (WYA) of freeway traffic flow was developed by Zurlinden (2003). It was further improved by the authors.

The method is based on a comparison of annual patterns of traffic demand and capacity including their properties as random variables. Hence, the stochastic concept of capacity plays a key role in this approach. By applying a Monte-Carlo technique, patterns of traffic demand and capacity are generated for a whole year, divided into 5-minute intervals. All relevant systematic influences on freeway capacity, like road geometry (number of lanes, gradient), weather conditions, and incidents, are considered by varying the scale parameter \( \beta \) of the capacity distribution function. The capacity drop, thus the difference of freeway capacity before and after a breakdown, is also accounted for. Incidents and accidents are randomly generated based on typical accident and car breakdown rates, respectively. The capacity reduction in case of accidents is estimated by using the values given in the HCM (2000). Rainfall events are randomly generated based on monthly weather statistics.
The assessment of traffic flow quality within the model is based on a deterministic queuing model. The queue length at the beginning and the end of each 5-minute interval can be determined by comparing the estimated traffic demand and capacity patterns. As an example, Figure 8 shows the comparison of traffic demand and capacity patterns over one week for a 3-lane freeway. The delays due to congestion are calculated by multiplying the average queue length with the interval duration.

For practical application of the WYA, the computer program “KAPASIM” was developed. The comparison of (randomly generated) demand and capacity patterns is repeated several times. The resulting averages for the sum of delays over a whole year or the total duration of congested flow conditions during a year are used as measures of traffic flow performance. The sum of delays can be transferred into economic costs and may be applied in cost/benefit-analyses.

The WYA can be used for a variety of applications like the economic analysis of road construction projects, the estimation of the share of different congestion causes (high demand, accidents, road works, bad weather conditions) to improve road management strategies, the evaluation of improved incident management, or the optimization of construction zone scheduling. Based on the capacity distribution functions given in Figures 5 and 6, the additional time losses caused by the reduced capacity during construction work could e.g. be determined. Meanwhile, the computer tool KAPASIM is under application for the optimization of freeway construction zone planning strategies in two German Federal States (Hesse and Saar). An improved version (named MacroSim) is under test application for the whole freeway network (about 950 km) of the State of Hesse. Here the simple deterministic queueing model is replaced by a more realistic macroscopic traffic flow model.

The WYA provides an estimation of a variety of parameters to characterize the reliability of freeway operation (cf. Brilon et al. 2005): The sum of lost times due to congestion, the risk of being significantly delayed by congestion, the number of traffic breakdowns, and a whole bunch of other potential reliability measures for freeway traffic flow as they have been described in several publications (cf. e.g. Shaw 2003). Each of these parameters is given as the expectation over one year. The measures allow for an assessment of traffic performance for all possible degrees of saturation of the system. In particular, the stochastic concept provides an analytical access to a quantitative assessment of different situations within LOS F.
6. CONCLUSIONS
The concept of stochastic capacities seems to be more realistic and more useful than the traditional use of single value capacities. The probabilistic approach provides an improved understanding for the variability of freeway traffic flow.

A methodology for the estimation of distribution functions of freeway capacity based on the statistics of lifetime data analysis is presented. It is rather distinctly shown that the capacity of a freeway section can be treated as a Weibull-distributed random variable.

The stochastic concept of capacity offers a variety of applications. As an example, the analysis of the impact of work zones on the capacity of a freeway is demonstrated. It is found that a work zone has a significant effect towards a reduction of both the median and the standard deviation of the capacity distribution. This effect can be precisely quantified by new statistical analysis tools.

The concept of random capacities was included into a Monte-Carlo simulation for a Whole-Year-Analysis of freeway traffic flow. Parameters like the sum of time losses or hours with congestion are used as measures of traffic performance. In particular, different degrees of congestion (LOS F) can be specified by quantitative measures. The simulation approach can be used as a method for economic appraisal of alternative freeway planning schemes or for the assessment of traffic management strategies.

Overall, it is expected that the random interpretation of traffic facility capacities offers the potential for improved traffic engineering methodologies.

REFERENCES


