Citation:
Brilon, W., Geistefeldt, J., Regler, M.: *Reliability of Freeway Traffic Flow: A stochastic Concept of Capacity*, Proceedings of the 16th International Symposium on Transportation and Traffic Theory, pp. 125 – 144, College Park, Maryland, July 2005, ISBN 0-08-044680-9

RELIABILITY OF FREEWAY TRAFFIC FLOW: A STOCHASTIC CONCEPT OF CAPACITY

Werner Brilon, Justin Geistefeldt, and Matthias Regler,
Institute for Transportation and Traffic Engineering, Ruhr-University Bochum, Germany

ABSTRACT

The paper introduces a new understanding of freeway capacity. Here capacity is understood as the traffic volume below which traffic still flows and above which the flow breaks down into stop-and-go or even standing traffic. It is easy to understand that a capacity in this sense is by no means a constant value. Empirical analysis of traffic flow patterns, counted at 5-minute intervals over several months and at many sites, clearly shows that this type of capacity is Weibull-distributed with a nearly constant shape parameter, which represents the variance. This was identified using the so-called Product Limit Method, which is based on the statistics of lifetime data analysis. It is demonstrated that this method is applicable to all types of freeways.

The stochastic methodology allows for a derivation of a theoretical transformation between capacities identified for different interval durations. The technique can also be used to identify effects of different external conditions like speed limits or weather on the capacity of a freeway.

The statistical distribution of capacity directly indicates the reliability of the freeway section under investigation. This distribution for one section is then transformed into statistical measures of reliability for larger parts of a network composed of sections of different capacity. Thus, the stochastic concept is also expanded into reliabilities of freeway networks. It is found that a freeway operates at the highest expected efficiency if it is only loaded to 90 % of the conventionally estimated (constant-value) capacity.

On the one hand, the paper quotes some real world results from German freeways. On the other hand, the reliability-based analysis leads to a new sophisticated concept for highway traffic engineering.

1 INTRODUCTION

Conventional measures of effectiveness for freeway facilities usually reflect travel time in the form of e.g. travel velocity or delay. Recently it has been becoming more and more obvious that these parameters are not sufficient for freeway traffic performance assessment since they place great emphasis on smaller differences in travel time, whereas the more significant difference between flowing traffic and congestion is not adequately represented. In addition, traditional quality assessment fails whenever demand exceeds capacity because it will simply attest a failure in this case. However, temporary freeway overloads are quite common. This is why quality assessments for different degrees of freeway congestion are required as well (cf. Shaw, 2003).

The capacity of a freeway is traditionally treated as a constant value in traffic engineering guidelines around the world, such as the HCM (2000). Doubts about this nature of capacities as constant values were raised by Ponzlet (1996) who demonstrated that capacities vary according to external conditions like dry or wet road surfaces, daylight or darkness, and prevailing purpose of the freeway (long distance or metropolitan commuter traffic). Moreover, several authors affirmed that even under constant external conditions, different capacities can be observed on freeways in reality (Elefteriadou et al., 1995; Minderhoud et al., 1997; Persaud et al., 1998; Kuehne and Anstett, 1999; Lorenz and Elefteriadou, 2000; Okamura et al., 2000). Most of these authors only observed traffic breakdowns at different flow rates to demonstrate the variability of flows preceding a breakdown. For a more systematic analysis, however, a comprehensive theoretical concept is required.

2 BASIC CONCEPT OF STOCHASTIC CAPACITY

Corresponding to the HCM (2000), the capacity of a freeway is defined as the maximum flow rate that can reasonably be expected to traverse a facility under prevailing roadway, traffic, and control conditions. With an identical meaning, the term “capacity”, i.e. the maximum flow rate, could also be defined as the traffic volume below which the performance of the facility is acceptable and above which – in case of greater demand – proper operation fails. The transition between proper operation and non-acceptable flow conditions is called “breakdown”. On a freeway, such a breakdown occurs when the average travel velocity is reduced from an acceptable speed level to a much lower value of congested conditions. These transitions usually involve a rather sudden speed reduction. The suddenness of this breakdown, however, may differ from one country to another depending on the general driving culture.

From this definition, it is clear that capacity in this sense is by no means a constant value. A constant value would mean that, given a capacity of e.g. 3,600 veh/h, the traffic should be fluent at a demand of 3,599 veh/h and be congested at a demand of 3,601 veh/h. This clearly indicates that the demand volume that causes breakdown varies in real traffic flow and that the flow rate of a breakdown depends on the behavior of several drivers combined with the specific local constellation on the freeway. Thus, it is plausible that the breakdown volume should have all properties of a random variable.

For using this concept of freeway capacity randomness it is necessary to know more about the capacity distribution function. Its determination is, however, not a trivial task. It is clear that any analytical approach must be supported by a broad empirical investigation.

Observations of traffic flow on freeways deliver pairs of values of traffic flow rates and average speeds during predetermined observation intervals (index i). According to the definition of capacity, the observed volume will be below capacity if the average speed exceeds a certain threshold value (e.g. about 70 km/h for German freeway conditions). With an average speed lower than the threshold value, traffic flow is congested. Thus, the flow must have exceeded capacity during the time between two such observations. The capacity itself, however, can not be measured directly. In addition, higher demand volumes are less likely to be observed in the field since there is also a higher probability that, before they occur, the breakdown has already happened during preceding intervals at lower volumes. Both effects make it difficult to estimate the capacity distribution function, which is defined as:

$$F_c(q) = p(c \le q) \tag{1}$$

where

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

A practicable estimation method was first presented by van Toorenburg (1986) and discussed by Minderhoud et al. (1997). The investigations presented here are based on this idea. However, it seemed advisable to modify some of the basic assumptions of the method.

The method proposed by van Toorenburg (1986) is based on an analogy to the statistics of lifetime data analysis. This statistics, in its basic formulation, serves to describe the statistical properties of the duration of human life. Moreover, it is usually applied to analyze the durability of technical components. In this context, the lifetime distribution function is:

$$F(t) = 1 - S(t) \tag{2}$$

where

F(t)	=	distribution function of lifetime $= p(T \le t)$
T	=	lifetime
S(t)	=	survival function $= p(T > t)$

Lifetime distributions are often estimated on the basis of experiments of limited duration. Consequently, the lifetimes of several individuals in the sample exceed the duration of the experiment and therefore can not be measured. It is only possible to state that these lifetimes are longer than the duration of the experiment. However, even this information is valuable. These data are called “censored data” (cf. e.g. Lawless, 2003).

If a traffic breakdown is regarded as a failure event, the methods for lifetime data analysis can be used to estimate the capacity c, which is the analogon of the lifetime T. The whole analogy between capacity analysis and lifetime data analysis is given in Table 1.

Table 1: Analogy between lifetime data analysis and capacity analysis

	Analysis of Lifetime Data	**Capacity Analysis**
Parameter	Time t	Traffic volume q
Failure event	Death at time t	Breakdown at volume q
Lifetime variable	Lifetime T	Capacity c
Censoring	Lifetime T is longer than the duration of the experiment	Capacity c is greater than traffic demand
Survival function	$S(t) = 1 - F(t)$	$S_c(q) = 1 - F_c(q)$
Probability density function	$f(t)$	$f_c(q)$
Probability distribution function	$F(t)$	$F_c(q)$

The statistics of lifetime data analysis can be used to estimate distribution functions based on samples that include censored data. A non-parametric method to estimate the survival function is the so-called “Product Limit Method” (PLM) by Kaplan and Meier (1958):

$$\hat{S}(t) = \prod_{j:t_j<t} \frac{n_j - d_j}{n_j} \tag{3}$$

where

$\hat{S}(t)$	=	estimated survival function	[-]
n_j	=	number of individuals with a lifetime $T \geq t_j$	[-]
d_j	=	number of deaths at time t_j	[-]

Usually, each observed lifetime is used as one t_j-value. In this case, d_j in Eq. 3 is always equal to 1.

Transferred to capacity analysis, Eq. 3 together with Eq. 2 can be written as:

$$F_c(q) = 1 - \prod_{i:q_i \leq q} \frac{k_i - d_i}{k_i}; \; i \in \{B\} \tag{4}$$

where

$F_c(q)$	=	distribution function of capacity c	[-]
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 (see below)	

Using this equation, each observed traffic volume q is classified according to

B: Traffic is fluent in time interval i, but the observed volume causes a breakdown, i.e. the average speed drops below the threshold speed in the next time interval i + 1.

F: Traffic is fluent in interval i and in the following interval i + 1. This interval i contains a censored value. Its information is that the actual capacity in interval i is greater than the observed volume q_i.

C1: Traffic is congested in interval i, i.e. the average speed is below the threshold value. As this interval i provides no information about the capacity, it is disregarded.

C2: Traffic is fluent in interval i, but the observed volume causes a breakdown. However, in contrast to classification B, traffic is congested at a downstream cross section during interval i or i – 1. In this case, the breakdown at the observation point is supposed to be due to a tailback from downstream. As this interval i does not contain any information for the capacity assessment at the observation point, it is disregarded.

The Product Limit Method does not require the assumption of a specific type of the distribution function. However, the maximum value of the capacity distribution function will only reach 1 if the maximum observed volume q was a B-value (i.e. followed by a breakdown). Only in this case, the product in Eq. 4 will be 0. Otherwise the distribution function will terminate at a value of $F_c(q) < 1$ at its upper end.

Eq. 4 is a useful solution for estimating the capacity distribution function of a freeway from traffic observations. For practical application, two items remain to be defined:

- **Duration Δt of observation intervals**

  For analysis, only rather short observation intervals are useful. Otherwise the causal relationship between traffic volume and breakdown would be too weak. 1-hour counts, for example, are not adequate for this reason. Ideally the observation period should be 1 minute or even less. Considering both the availability of reliable data from loop detectors and the usefulness of the results, Brilon and Zurlinden (2003), after experiments with different Δt's, came to the conclusion that $\Delta t = 5$ minutes was the best compromise. Consequently, the analyses below are all based on 5-minute volume and speed values.

- **Exact understanding of a breakdown**

  The definition of a breakdown mentioned above (e.g. in Eq. 4) is a decisive aspect of the whole methodology. Van Toorenburg (1986; see also Minderhoud et al., 1997) defined breakdown capacity as the volume measured downstream of a queue at a bottleneck. In consequence, each congested flow volume is regarded as a B-value (see comments to Eq. 4). Within the meaning of the analogy (Table 1) this would be equivalent to including in a lifetime analysis an individual who died a while ago. This does not seem to be reasonable. Instead, only those intervals i that cause a breakdown are treated as B-intervals. As a breakdown of traffic flow usually involves a significant speed reduction, breakdown events can be detected using a time series containing both traffic volumes and average space mean speeds. This is done by using a constant threshold speed value. If the speed falls below the threshold value in the next interval i + 1, the traffic volume in interval i is regarded as a B-value. A threshold speed of 70 km/h was found to be fairly representative for German freeways but may be different for other road types. In some cases, different or more detailed criteria may be required to reliably identify traffic breakdowns – e.g. a criterion that considers the minimum speed difference between the intervals i and i + 1.

The capacity of a freeway section (one direction) can be analyzed most precisely if observations are made at a clearly distinguishable bottleneck, as Fig. 1a shows. At such a bottleneck, breakdowns should only be caused by oversaturation of the bottleneck itself. Tailback from downstream should not occur as greater capacities are always available in the succeeding section. Observations are therefore made at a point slightly upstream of the bottleneck. Such observations were performed by Brilon and Zurlinden (2003) to make sure that the external conditions were clearly in harmony with theoretical assumptions.

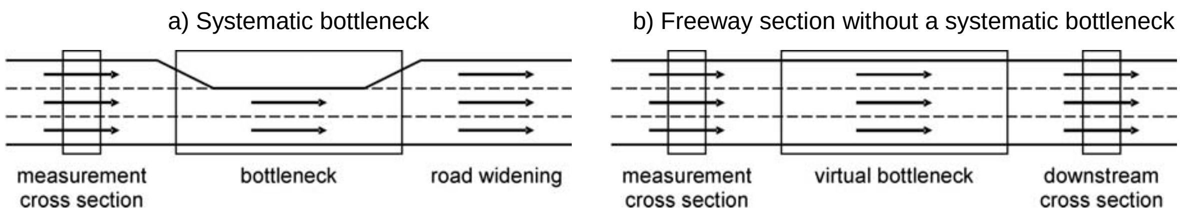

Fig. 1: Two cases of freeway sections under investigation

As an example, Fig. 2 shows speed-flow diagrams from two observation sections along the ring of freeways around the city of Cologne. Both sites are geometric bottlenecks with the road widening downstream of the observation point. The figure shows speed-flow data for 5-minute counts across all lanes obtained from automatic loop detectors throughout the year 2000. Due to frequent oversaturation of both freeway sections, many congested intervals were observed. With the conventional approach (estimation of a regression model in the k-v diagram plus deriving the maximum flow rate from $q = k \cdot v$), a capacity of 4284 veh/h for the 2-lane example and 6720 veh/h for the 3-lane case was determined based on the speed-flow relationship proposed by van Aerde (1995).

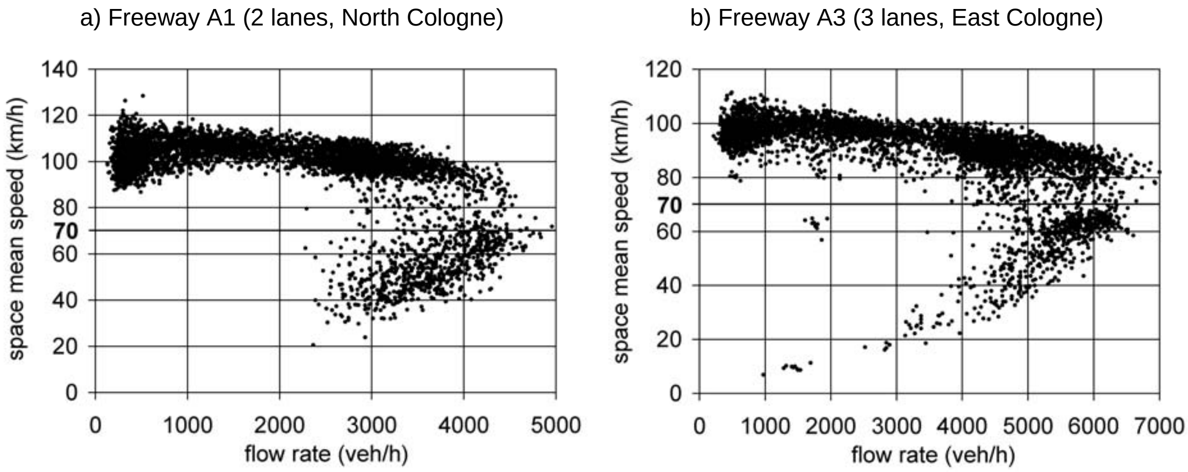

Fig. 2: Speed-flow diagrams for two freeway sections (5-minute intervals)

The Product Limit technique (Eq. 4) was used to estimate the capacity distribution function for both examples (see black lines in Fig. 3). 933 intervals with a breakdown (classification B) were identified on the A1 freeway and 834 on the A3 freeway. It appears that, despite the large size of a one-year sample, no complete distribution function could be estimated since the highest q-values observed were not followed by a breakdown. This effect makes it difficult to find an appropriate estimate for the whole capacity distribution function.

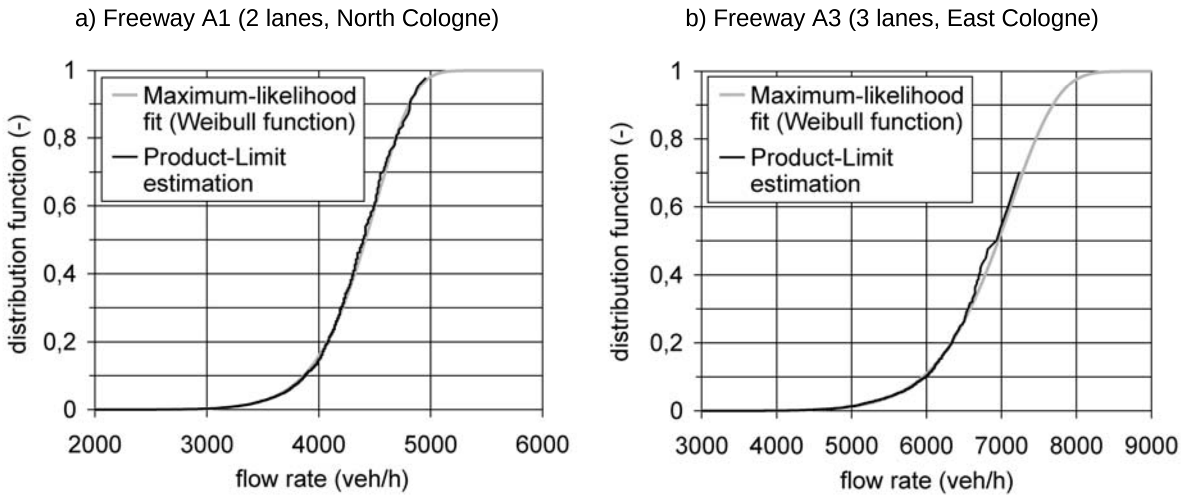

Fig. 3: Estimated capacity distribution functions for two freeway sections according to Fig. 1a (5-minute intervals, dry roadway conditions)

To overcome this problem, it is necessary to know more about the mathematical type of the distribution function $F_c(x)$, which did not have to be defined for Eq. 4. Various plausible function types like Weibull, Normal and Gamma distribution were tested (Brilon and Zurlinden, 2003). To estimate the parameters of the distribution functions, a maximum likelihood technique was used. The likelihood function is given by (cf. Lawless, 2003):

$$L = \prod_{i=1}^{n} f_c(q_i)^{\delta_i} \cdot [1 - F_c(q_i)]^{1-\delta_i} \qquad (5)$$

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	[-]
δ_i	=	1, if uncensored (breakdown of classification B)	
δ_i	=	0, elsewhere	

The likelihood function or its natural logarithm (log-likelihood) has to be maximized to calibrate the parameters of the distribution function (cf. e.g. Lawless, 2003). By comparing different types of functions based on the value of the likelihood function, the Weibull distribution turned out to be the function that best fitted the observations at all freeway sections under investigation. The Weibull distribution function is:

$$F(x) = 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}} \quad \text{for} \quad x \geq 0 \qquad (6)$$

where

α	=	shape parameter	[-]
β	=	scale parameter	[-]

The two examples in Fig. 3 show that the Weibull distribution fits very well into the PLM estimation.

Of course, the Product Limit Method for capacity estimation could also be used with traffic densities k instead of volumes q. Eq. 4 and 5 remain unchanged, except that q is replaced by k. It was expected that k would be a better determinant for the occurrence of a breakdown than q and that, consequently, the analysis would reduce the variability of the resulting distribution function. This was tested by Regler (2004) using data from six 3-lane freeway sections. The median of the breakdown densities ranged from 70 to 90 veh/km with Weibull-parameters α = 8.4 through 13.2 and β = 72 through 92 veh/km for the analysis of 5-minute intervals. The resulting distribution of breakdown density and breakdown capacity tended towards larger variances compared to the analysis over the q-axis. Therefore, it was not advantageous in this context to estimate capacity distributions from densities. Moreover, densities are artificial parameters that must be calculated from measured speeds v and volumes q ($k = q / v$). The k-based analysis also needs a more complicated definition of the speed threshold between fluent and congested traffic. Therefore, using the Product Limit estimation based on densities is not recommended.

3 APPLICATION TO FREEWAYS

So far, only results from freeway bottlenecks according to Fig. 1a have been discussed. Under this assumption, the estimation technique would be restricted to rather specific geometric situations. Regler (2004) applied the methods described above to freeway sections without a distinct bottleneck (Fig. 1b) whose geometric properties included no change in the number of lanes. To make sure that intervals with a tailback from downstream bottlenecks did not impair the results, classification C2 (see above) was considered as well. With this technique, the capacity distribution of quite a variety of freeway sections could be analyzed (Table 2). The analysis was based on 5-minute counts taken over several months, with an arrangement as shown in Fig. 1b. All sites are on 3-lane freeway carriageways, mainly in level terrain. Periods of work zones were excluded from the data.

Table 2: Parameters β and α, expectation E(c) and standard deviation σ of the estimated Weibull capacity distribution at 15 freeway sections (3-lanes, 5-minute intervals)

Section	β [veh/h]	α [-]	E(c) [veh/h]	σ [veh/h]
A3-1	7441	11,31	7115	762
A5-1	6217	11,15	5941	645
A5-2	6074	13,59	5847	526
A5-4	6608	13,92	6365	559
A5-5	6392	14,16	6161	532
A5-6	6272	14,69	6053	505
A5-7	7194	13,98	6932	606
A5-8	6884	13,35	6622	606
A9-1	7937	8,85	7510	1013
A9-2	7399	13,66	7124	637
A9-3	5988	14,82	5780	478
A9-4	6141	18,86	5969	392
A9-5	6648	14,24	6409	551
A9-6	7109	9,62	6752	842
A9-7	6648	14,92	6419	528

It turns out that the shape parameter α in the Weibull distribution typically ranges from 9 to 15 with an average of 13. This magnitude seems to be characteristical for 3-lane freeways. The scale parameter β of the Weibull distribution varies over a wide range between the analyzed sections. This may be mostly due to different geometric and control conditions, different driver and vehicle populations, and diverse prevailing travel purposes (long distance travel versus metropolitan commuter traffic).

Having found that the shape parameter α seems to be almost constant, we may transform the capacity distribution function to fit different interval durations Δ. According to Eq. 1, $F_5(q)$ is the probability of a breakdown during $\Delta = 5$ minutes at flow rate q. Hence, $(1 - F_5(q))$ is the probability of no breakdown occurring in this interval. If we assume that breakdowns occurring in succeeding intervals are independent of each other, then the probability of fluent traffic flow during a whole hour is

$$p_{60}(\text{fluent traffic}) = [1 - F_5(q)]^{12} \tag{7}$$

Using the Weibull distribution (Eq. 6), this is converted into

$$F_{60}(q) = 1 - p_{60}(\text{fluent traffic}) = 1 - e^{-12\cdot\left(\frac{q}{\beta_5}\right)^{\alpha}} \tag{8}$$

which is again a Weibull distribution with an unchanged shape parameter α and a scale parameter $\beta_{60} = r \cdot \beta_5$, where $r = 12^{(-1/\alpha)}$. With $\alpha = 13$ (see above) we get $r \approx 0.82 \approx 1/1.2$. This means that for 5-minute observations the expected capacity should be in a range of 1.2 times the 1-hour capacity. This factor of 1.2 seems to be typical for the transformation of capacities from 5-minute intervals into 60-minute intervals, as was pointed out by Keller and Sachse (1992) or Ponzlet (1996) based on empirical capacity estimates obtained from the fundamental diagram.

One might object to Eq. 8 that the traffic volume q usually is not constant during a whole hour. However, numerical calculations showed that volume variations during one hour did not significantly change the results. It also seems to be realistic to assume that traffic breakdowns in succeeding intervals are independent of each other since there is no imaginable reason why the opposite should be true. This question may, however, be made a subject of further research.

The new technique was used to investigate differences in performance between dry and wet road surfaces. At all sections under examination (see list in Table 2), it turned out very clearly that on a wet road surface the capacity was reduced by around 11 %. The effects of darkness were investigated as well. Contrary to Ponzlet's (1996) results, it was clearly found that darkness did not shift the capacity distributions.

The results shown in Table 2 also demonstrate differences in the capacity distribution between an uncontrolled freeway (section A9-3) and a freeway with traffic adaptive variable speed limits (section A9-4): The mean capacity of the controlled section is slightly (by 3 %) higher compared to the uncontrolled section, but the standard deviation is significantly lower (cf. Fig. 4). The two analyzed sections are the two opposite carriageways of the freeway A9 near Munich and thus have similar geometric and traffic characteristics.

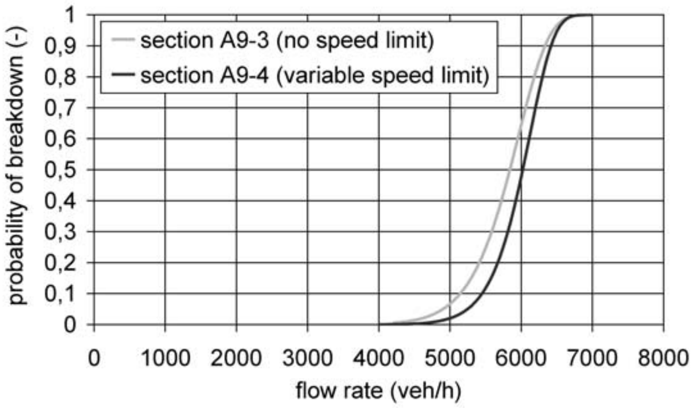

Fig. 4: Capacity distributions for a 3-lane freeway with and without variable speed control (13.5 % average truck percentage, 5-minute intervals)

All these examples demonstrate that the statistical interpretation of freeway capacity together with the corresponding estimation technique provides a better understanding of freeway traffic operation. It improves the methodology of investigating differences between various external conditions.

4 FREEWAY TRAFFIC DYNAMICS

So far, only capacities in the upper branch of the speed-flow diagram, i.e. under fluent traffic conditions, have been analyzed. However, the lower branch of the speed-flow diagram representing congested conditions must be considered as well.

It is well known that dynamics in the speed-flow diagram (i.e. the sequence of v-q points over time) follow specific patterns. The first to report typical hysteresis phenomena within these dynamics were Treiterer and Myers (1974). More recently, Kim and Keller (2001) came to the conclusion that six different typical traffic states should be distinguished within the speed-flow diagram (or the fundamental diagram). These dynamics were analyzed by Regler (2004) based on 5-minute data for the freeway sections A3-1 and A5-7 (cf. Table 2) extending over 4 months and 10 months, respectively. More than 120 breakdowns from fluent traffic to congested traffic were observed.

The analysis came to the conclusion that there are three different states of traffic conditions to be distinguished:

1. Fluent traffic at high speeds (i.e. $v > 70$ km/h) and low densities. In this state, volumes q may range from 0 to the maximum flow rate. This is the ascending branch of the q-k diagram.
2. A transient state with an average velocity of around 60 km/h and rather high volumes. We like to call this state “synchronized flow”, knowing that this term is used by other authors (Kerner and Rehborn, 1996) with a slightly different meaning. In this state, vehicles are forced to travel at fairly similar speeds on all lanes.

3. Congested traffic with low speeds and low traffic volumes.

To illustrate typical dynamic patterns, Fig. 5a shows observations from the freeway A5 with only states 1 and 2 involved. In all examples, the transition from fluent to synchronized flow began at a rather high volume. Traffic flow stabilized at slightly lower volumes at an average speed of about 60 km/h. From here, the traffic flow recovered to fluent traffic conditions. All recoveries involved much lower traffic volumes than the preceding breakdown. This hysteresis phenomenon seems to be a characteristic of traffic dynamics. In addition to these 2-state sequences, Fig. 5b shows those cases from the A5 where a breakdown from fluent to synchronized flow was followed by a subsequent transition into congested traffic with very low speeds. The recovery back to fluent traffic did never happened directly. Instead, each recovery process passed through the transient state of synchronized flow.

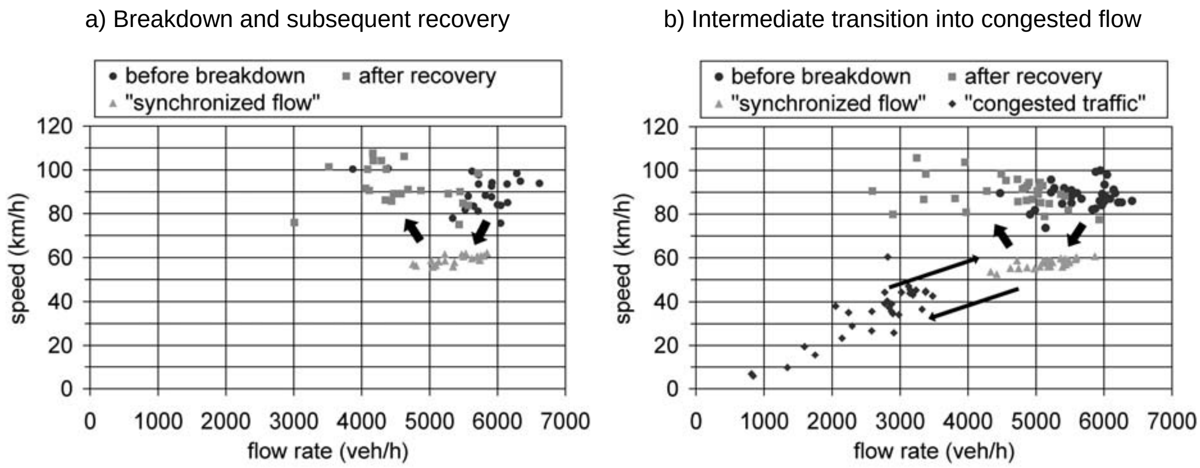

Fig. 5: Two typical patterns of traffic dynamics during breakdown and recovery (freeway section A5-7, 5-minute flow rates)

Empirical analysis of freeway traffic dynamics showed that:

- Transitions between traffic states usually happen suddenly, i.e. within rather short times and distances.
- Breakdowns from fluent traffic are first followed by the synchronized traffic state. From there, speed and flow rate may decline further to heavy congestion.
- All recoveries pass through the synchronized state. No recovery process jumps directly from congestion back to fluent traffic. Recovery from synchronized to fluent traffic always involves much lower volumes than the breakdown. The difference between breakdown volume and recovery volume (fluent traffic after a recovery) on the observed 3-lane freeway sections ranged from 500 to 1500 veh/h. Volumes in the synchronized state are always lower than the maximum flows in fluent traffic (= capacity). Observed differences on 3-lane freeways ranged from 200 to 600 veh/h, measured in 5-minute-intervals. This effect is called “capacity drop” (cf. section 5).

Because of the large amount of data that was analyzed and because of the remarkable analogy between all cases observed, we may say that these properties are typical for the dynamics of freeway traffic flow. It should, however, be admitted that these dynamic effects are predominantly due to driver behavior, concerning headways at high speeds in dense fluent traffic, together with braking and accelerating behavior. Therefore, different typical dynamics might be found for other driving cultures than in Germany. Even here, under variable speed control, a few traffic breakdowns were observed where the average flow in synchronized traffic was higher than the flow rate before breakdown.

5 CAPACITY DROP

A number of investigations has proven the existence of different capacities under flowing and congested traffic conditions. Banks (1990) as well as Hall and Agyemang-Duah (1991) analyzed this "capacity drop" phenomenon for different North-American freeways. Capacity drop values of between 3 and 6 % were measured. Ponzlet (1996) analyzed traffic flow on German freeways to see whether this phenomenon existed. He determined a 6 % drop for 5-minute flow rates. Brilon and Zurlinden (2003) analyzed the capacity drop by comparing the stochastic capacity to flow rates in congested flow. They computed an average of 24 %, which is very high compared to other authors' results.

There are different hypotheses about the reasons for the capacity drop phenomenon:

- Bottleneck downstream of the study site: The flow at the point under investigation will remain fluent until the section between this point and the bottleneck is filled with congested flow. After this time, the maximum flow will be the bottleneck's capacity.
- Different driver behavior: Drivers in fluent traffic accept shorter headways since they expect to be able to pass the vehicles in front. Once they have given up this idea, they switch to a more safety-conscious style of driving and keep longer headways.
- Restricted acceleration capabilities: At the front of the congested area, drivers need to accelerate. Some vehicles, however, have limited acceleration power, which opens a larger gap in front of them.

These hypotheses and the value of the capacity drop were analyzed by Regler (2004) for the 15 freeway sections listed in Table 2, using different approaches. The main question was: What are the capacities under synchronized flow conditions as described in section 4? The conventional traffic flow model of van Aerde (1995) as enhanced by Ponzlet (1996) to account for the capacity drop phenomenon yielded an average drop of 270 veh/h in 5-minute flow rates. Using a distribution of breakdown flow rates (observed immediately prior to breakdown) and a distribution of queue discharge flow, an average drop of 250 veh/h in 5-minute flow rates was determined, which is comparable to the result from the fundamental diagram.

It is not easy to find a distribution of capacity after a breakdown that is adequate to the Product Limit estimate (Eq. 4). Brilon and Zurlinden (2003) applied a distribution of all flow rates in congested traffic and computed a very high capacity drop. Regler (2004) developed a method comparable to the Product Limit technique to obtain a capacity distribution in queue discharge flow. This method is based on the following assumptions (cf. Fig. 6):

- Flow rates during congestion do not represent the maximum possible flow of congested traffic.
- Capacity can directly be measured in queue discharge flow, i.e. in the last 5-minute interval before recovery of traffic flow (uncensored data).
- Flow rates during congestion are lower than capacity in queue discharge flow (censored data).
- Flow rates in free flow are not relevant for the capacity of queue discharge flow.

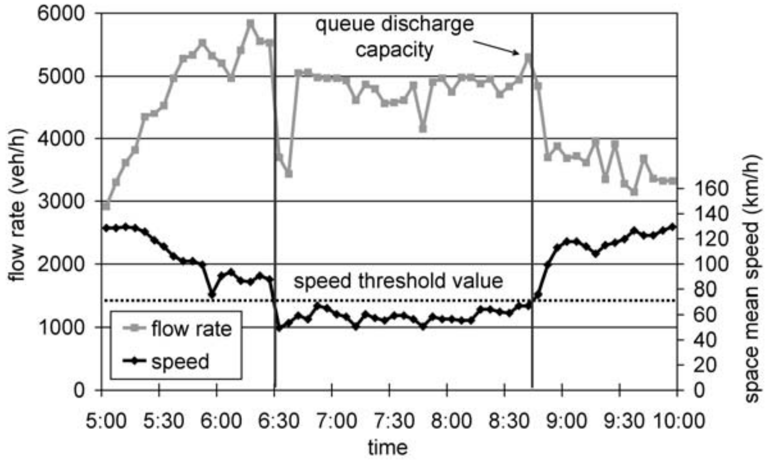

Fig. 6: Flow rate and speed time series during congestion (freeway A5, 5-minute intervals)

Based on these considerations, the observed flow rates in each time interval can be classified as follows:

B*: Traffic recovers from congestion to free flow, i.e. the average speed exceeds the threshold value from time interval i to interval i + 1

F*: Traffic is congested in intervals i and i + 1, i.e. the average speed is lower than the threshold value in both intervals. This interval i contains a censored value.

C*: Traffic is fluent in interval i, i.e. the average speed is above the threshold value. This interval is not relevant.

After this classification, a capacity distribution for queue discharge flow according to Eq. 4 (where {B} is to be replaced by {B*}) can be computed from empirical data. This capacity level turned out to be always lower than the capacity before breakdown obtained from Eq. 4.

The difference between both distributions (see Fig. 7), represented by the median value, for instance, may be regarded as the capacity drop.

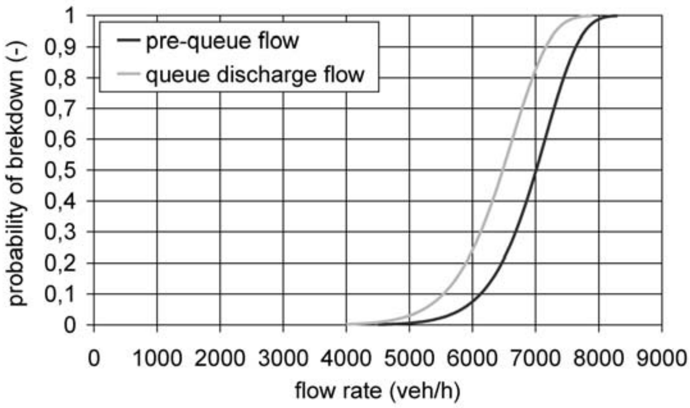

Fig. 7: Capacity distributions for pre-queue (cf. section 2) and queue discharge flow (freeway section A5-7, 11.8 % average truck percentage, 5-minute intervals)

For the 15 freeway sections listed in Table 2, an average drop of 1,180 veh/h was estimated using this algorithm. It should, however, be mentioned that the results for the capacity drop varied widely between the sites investigated. All attempts to identify regularities within the variation failed. It might be that the capacity drop shows some chaotic properties, as was indicated by other authors (Kerner, 2000) using different methodologies.

6 TRAFFIC RELIABILITY

Traffic reliability is an important factor for the assessment of the performance of highway segments and systems. In this context, the term "reliability" mainly refers to the variability of travel times. However, several definitions can be found in the literature. A comprehensive outline of these definitions is given by Shaw (2003).

Here, traffic reliability is assessed by analyzing the probability that a freeway link is not congested, i.e. that the travel time does not exceed an acceptable level. This is becoming a question of increasing importance for "just-in-time" transportation in modern logistics chains. With the stochastic concept of capacity, it is possible to assess this kind of traffic reliability for a freeway link consisting of several sections.

The overload probability for a single bottleneck (either distinct or virtual, cf. Fig. 1) is equal to the capacity distribution function $F_c(q)$ given in Eq. 1. The probability of no congestion $p_{free}(q)$ represents the complementary event:

$$p_{free}(q) = 1 - F_c(q) \tag{9}$$

For the analysis of n subsequent (quasi-) bottleneck sections, the probability of no congestion in the whole system is the product of the single probabilities for each section:

$$p_{free,1+2+...+n}(q_1, q_2, ..., q_n) = \prod_{i=1}^{n} \left[1 - F_{c,i}(q_i)\right] \tag{10}$$

where

q_i = demand at section i

$F_{c,i}(q_i)$ = capacity distribution function of section i

By applying the Weibull distribution function to each section, this can be written as:

$$p_{free,1+2+...+n}(q_1, q_2, ..., q_n) = e^{-\left(\frac{q_1{}^{\alpha}}{\beta_1{}^{\alpha}} + \frac{q_2{}^{\alpha}}{\beta_2{}^{\alpha}} + ... + \frac{q_n{}^{\alpha}}{\beta_n{}^{\alpha}}\right)} = e^{-\sum_{i=1}^{n} \frac{q_i{}^{\alpha}}{\beta_i{}^{\alpha}}} \tag{11}$$

The overload probability for the whole system, which is equal to the capacity distribution function, is:

$$F_{c,1+2+...+n}(q_1, q_2, ..., q_n) = 1 - e^{-\sum_{i=1}^{n} \frac{q_i{}^{\alpha}}{\beta_i{}^{\alpha}}} \tag{12}$$

In the special case of the traffic demand being identical in all sections ($q_1 = q_2 = \ldots = q_n = q$), the overload probability for the system is:

$$F_{c,1+2+...+n}(q) = 1 - e^{-q^{\alpha} \cdot \sum_{i=1}^{n} \frac{1}{\beta_i{}^{\alpha}}} \tag{13}$$

This is again a Weibull distribution with the scale parameter $\beta_{1+2+\ldots+n}$:

$$\beta_{1+2+...+n} = \left(\frac{1}{\beta_1{}^{\alpha}} + \frac{1}{\beta_2{}^{\alpha}} + ... + \frac{1}{\beta_n{}^{\alpha}}\right)^{-1/\alpha} = \left(\sum_{i=1}^{n} \frac{1}{\beta_i{}^{\alpha}}\right)^{-1/\alpha} \tag{14}$$

Eq. 10 through 14 are based on the assumption that breakdown events due to an overload at the bottleneck sections and thus the capacity distribution functions are statistically independent. This assumption seems to be reasonable if the length of each section is sufficiently large. However, further empirical research may be required to establish the degree to which this assumption is justified. The fact that the volume of traffic arriving at bottleneck section i is influenced by congestion incidents at preceding sections (i – 1, i – 2, …), e.g. due to the capacity drop, is not relevant because in case of an overload of one section, the whole system is regarded as overloaded according to the applied definition of traffic reliability. For the same reason, the impact of a queue spillback spreading over several sections does not affect the results.

7 TRAFFIC EFFICIENCY

Brilon (2000) has proposed to use the parameter

$$E = q \cdot v \cdot T \tag{15}$$

where

E	=	traffic efficiency	[veh · km/h]
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 of traffic flow on a freeway. This parameter describes the "production per time unit" of a freeway. The more veh · km a freeway produces per hour, the greater the efficiency with which the potential of the existing infrastructure is exploited.

By applying the concept of random capacities, each volume q has to be combined with the corresponding probability of a breakdown. Brilon and Zurlinden (2003) have derived:

$$E_{exp}(q_D) = \left[q_D \cdot v \cdot (1-\chi) + q_{dc} \cdot v_{dc} \cdot \chi\right] \cdot T \tag{16}$$

where

$E_{exp}(q_D)$	=	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]
χ	=	$\frac{1}{n} \cdot \sum_{i=1}^{h} \sum_{k=0}^{m} p_B(q_{i-k}) \cdot (1 - p_{cong,i-k-1})$	
	=	expected proportion of congested intervals with $q = q_{dc}$	[-]
n	=	number of 5-minute intervals during T	[-]
m	=	$\mathrm{Min}\{n_{cong} - 1; i - 1\}$	[-]
n_{cong}	=	average duration (number of intervals) of a congested period ≥ 1	[-]
$p_B(q)$	=	probability of a breakdown at volume q (here: $q = q_0$) (e.g. after Eq. 4: $p_B = dF_c(q)/dq$)	[-]
$p_{cong,i}$	=	$\sum_{k=0}^{m} p_B(q_{i-k}) \cdot (1 - p_{cong,i-k-1})$	
	=	probability of congested flow in 5-minute interval	[-]

If we insert real data for $p_B(q) = f_c(q)$ into this set of equations it becomes clear that the maximum expected efficiency is achieved for a demand volume q_D that is lower than the average capacity. Sample calculations show that the highest efficiency of a freeway is to be expected at a demand of approximately $0.9 \cdot c$, where c is the traditionally defined (constant-value) capacity.

8 CONCLUSIONS

As a result of a series of studies of German freeways, the concept of stochastic capacities seems to be more realistic and more useful than the traditional use of single value capacities. This probabilistic approach provides an improved understanding of both the variability of traffic flow observations and the typical dynamics in different traffic states on a freeway.

The idea of random capacity is based on the work of authors like van Toorenburg (1986) and Minderhoud et al. (1997). Compared to their approach, the so-called Product Limit Method for capacity estimation has been modified and extended. It was rather distinctly shown that the capacity of a freeway section is Weibull-distributed. For German freeways, the shape parameter seems to be in a range of 13, whereas the scale parameter can be different for specific freeway sections. One drawback of the methodology is that huge sample sizes are needed to estimate capacity distribution functions. With the application of ITS-methods, these data will become increasingly available.

The concept of randomness permits to demonstrate the capacity reducing effect of wet road surfaces (-11 %) and the capacity increasing effect of traffic adaptive variable speed limits. The Product Limit Method was also used to estimate queue discharge capacity. It was confirmed that this capacity is usually lower than the capacity in fluent traffic. The studies did clearly show that three typical states in traffic flow exist: fluent traffic, congested traffic, and a transient state that occurs in each breakdown and recovery of traffic flow. The extent of the so-called capacity drop did not show any regularities, although many freeway sections, each with a large sample size, have been analyzed. It is concluded that the capacity drop has rather chaotic properties.

The concept of random capacity reveals that the optimum degree of saturation for a freeway, based on data from Germany, ranges around 90 %. If the degree of saturation increases further, the risk of a breakdown becomes too high, so that the efficiency of freeway operation must be expected to be lower than at a saturation of 90 %. The stochastic concept can also be applied to freeways consisting of several succeeding sections.

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

REFERENCES

van Aerde, M. (1995). A Single Regime Speed-Flow-Density Relationship for Freeways and Arterials. *Proceedings of the 74th TRB Annual Meeting*. Washington D.C.

Banks, J. H. (1990). Flow Processes at a Freeway Bottleneck. *Transportation Research Record No. 1287*, Transportation Research Board, National Research Council, Washington D.C.

Brilon, W. (2000). Traffic Flow Analysis beyond Traditional Methods. *Proceedings of the 4th International Symposium on Highway Capacity*, pp. 26-41, TRB Circular E-C018, Transportation Research Board, Washington D.C.

Brilon, W. and H. Zurlinden (2003). Ueberlastungswahrscheinlichkeiten und Verkehrsleistung als Bemessungskriterium fuer Strassenverkehrsanlagen (Breakdown Probability and Traffic Efficiency as Design Criteria for Freeways). *Forschung Strassenbau und Strassenverkehrstechnik No. 870*. Bonn.

Elefteriadou, L., R.P. Roess, and W.R. McShane (1995). Probabilistic Nature of Breakdown at Freeway Merge Junctions. *Transportation Research Record 1484*, Transportation Research Board, National Research Council, Washington D.C.

Hall, F.L. and K. Agyemang-Duah (1991). Freeway Capacity Drop and the Definition of Capacity. *Transportation Research Record 1320*, Transportation Research Board, National Research Council, Washington D.C.

HCM (2000). *Highway Capacity Manual*. Transportation Research Board, Washington D.C.

Kaplan, E.L. and P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457-481.

Keller, H. and T. Sachse (1992). Einfluss des Bezugsintervalls in Fundamentaldiagrammen auf die zutreffende Beschreibung der Leistungsfaehigkeit von Straßenabschnitten (Influence of Aggregation Intervals in the Fundamental Diagram on the Appropriate Description of Roadway Capacity). *Forschung Strassenbau und Strassenverkehrstechnik No. 614*. Bonn.

Kerner, B. S. (2000). Theory of Breakdown Phenomenon at Highway Bottlenecks. *Transportation Research Record No. 1710*, pp. 136-144. Transportation Research Board, National Research Council, Washington D.C.

Kerner, B. S. and H. Rehborn (1996). Experimental Properties of Complexity in Traffic Flow. *Physical Review E*, **53** (5), 4275-4278.

Kim, Y. and H. Keller (2001): Zur Dynamik zwischen Verkehrszuständen im Fundamentaldiagramm (Dynamics between Traffic States in the Fundamental Diagram). *Strassenverkehrstechnik*, Issue 9/2001, pp. 433-442.

Kuehne, R.D. and N. Anstett (1999). Stochastic Methods for Analysis of Traffic Pattern Formation. *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*. Jerusalem.

Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

Lorenz, M. and L. Elefteriadou (2000). A Probabilistic Approach to Defining Freeway Capacity and Breakdown. *Proceedings of the 4th International Symposium on Highway Capacity*, pp. 84-95. TRB-Circular E-C018, Transportation Research Board, Washington D.C.

Minderhoud, M.M., H. Botma and P.H.L. Bovy (1997). Roadway Capacity Estimation Methods Explained and Assessed. *Transportation Research Record No. 1572*. Transportation Research Board, National Research Council, Washington D.C.

Okamura, H., S. Watanabe and T. Watanabe (2000). An Empirical Study on the Capacity of Bottlenecks on the Basic Suburban Expressway Sections in Japan. *Proc. of the 4th International Symposium on Highway Capacity,* pp. 120-129. TRB Circular E-C018, Transportation Research Board, Washington D.C.

Persaud, B., S. Yagar and R. Brownlee (1998). Exploration of the Breakdown Phenomenon in Freeway Traffic. *Transportation Research Record No. 1634*, pp. 64-69. Transportation Research Board, National Research Council, Washington D.C.

Ponzlet, M. (1996). Dynamik der Leistungsfähigkeiten von Autobahnen (Dynamics of Freeway Capacity). *Schriftenreihe des Lehrstuhls fuer Verkehrswesen der Ruhr-Universitaet Bochum, No. 16*. Bochum.

Regler, M. (2004). Verkehrsablauf und Kapazität auf Autobahnen (Freeway Traffic Flow and Capacity). *Schriftenreihe des Lehrstuhls fuer Verkehrswesen der Ruhr-Universitaet Bochum, No. 28*. Bochum.

Shaw, T. (2003). *Performance Measures of Operational Effectiveness for Highway Segments and Systems*. NCHRP Synthesis 311, Transportation Research Board, Washington, D.C.

van Toorenburg, J. (1986). *Praktijwaarden voor de capaciteit*. Rijkswaterstaat dienst Verkeerskunde, Rotterdam.

Treiterer, J. and J.A. Myers, (1974). The Hysteresis Phenomenon in Traffic Flow. *Proceedings of the 6th International Symposium on Transportation and Traffic Theory*, pp. 13-38. Reed Pty Ltd., Sydney.