Capacity Reductions at Incidents Sites on Motorways

Victor Knoop[1], Serge Hoogendoorn[2] and Koen Adams[3]

Transport and Planning, TRAIL Research School, Delft University of Technology

Incidents cause a large part of the delays in road networks. This is caused by a decrease of the capacity at the incident site. A detailed knowledge of the queue discharge rate can improve for instance the traffic prediction and thereby improve delay information or routing advice. Therefore, this study determines the queue discharge rate for many incident locations during an incident situation and these are compared with the queue discharge rate at the same location in normal conditions. Ninety incidents meet the requirements to apply the proposed methodology. It is found in case a driving lane is blocked, the queue discharge rate for each available lane is reduced by 50%. In case the driving lanes are open but there is a distraction of an incident at the emergency lane or on the roadway for the opposite direction, the queue discharge rate is reduced by 30%.

Keywords: Incidents; Accidents; Road capacity

1. Introduction

A large part of the delays is caused by incidents. Kwon et al. (2006) show that this amount is about 25% in the USA. A similar number is found for the Netherlands (Dutch Road Authority (2005)). The amount of delay depends on the demand and the supply on the road. In fact, if the supply (the remaining capacity) reduces to a level lower than the demand, it causes a traffic jam. Therefore, to compute the delays it is essential to know the capacity of the road at the location of an incident. Obviously, one reason for a decreased capacity at the incident site is the decrease of the number of available lanes. We hypothesize that, in addition, the capacity decreases because the remaining lanes are used less efficient because people are distracted.

This contribution investigates the maximum throughput at the position of an incident when a jam has occurred. We consider both the incident-direction and the non-incident direction (i.e., the direction in which no physical obstruction occurs). Due to limited data availability and the required computation power to perform all analyses, we restrict the study to freeways with 3 lanes in each direction. This has the advantage that the road layout is the same for all measurements and the results can be directly compared.

---

[1] Stevinweg 1, 2628 CN Delft, the Netherlands, T: +31152781723, F: +31152873179, E: v.l.knoop@tudelft.nl
[2] Stevinweg 1, 2628 CN Delft, the Netherlands, T: +31152785475, F: +31152873179, E: s.p.hoogendoorn@tudelft.nl
[3] Groen van Prinsterersingel 43b, 2805 TD Gouda, the Netherlands, T: +31182555030, F: +31182555039, E: k.adams@arane.nl

Incidents are categorized in two groups. One of these is an accident situation where at least one of the lanes normally available for traffic, which we will call "driving lanes", is blocked. The other the situation where a car has broken down and is stopped at the shoulder lane of the freeway. With shoulder lane we mean a paved lane at the right of the driving lanes which is only meant for emergencies and which is not one of the driving lanes. In the Netherlands, there is no shoulder lane at the left of the driving lanes. In both situations, a change in driving behaviour can reduce the capacity of the remaining lanes. This study will show the magnitude of this behavioural effect.

The objective of this article is to describe the maximum flows at incidents for traffic in both the incident-direction (with a partially blocked roadway) and in the non-incident direction. We expect a lower capacity at that point due to distraction also at the other side of the guardrail, an effect which is called "rubbernecking". To this end, data from a large set of incidents is used, 90 of which were suitable to use for the computation of the capacity reduction. These incident capacities, for both directions, have never been studied on this scale before. There are handbook values, but those often give just a number for the incident capacities. This article adds the methodology that is used. Another contribution of the article is that it adds a distribution of the capacities which is possible because of the large number of incident capacities analysed. Section 2 will show the available literature on incident capacity; there, the gap is explained in more detail.

The capacity values are used to quantify the effects of shortening the incident handling time. The article only states how large the effects of a time reduction in a specific phase are which can be used to assess the impact of specific measures and specific measures, are not named or assessed within this article because there are many measures. However, because the findings are general, the findings proposed here can for instance be used in assessing the effectiveness of *new* measures which do not exist yet. For each specific measure, the costs will vary. This way, the method presented can be used to get an insight of the benefits, which then have to be weighted against the costs.

The remainder of the article is set up as follows. In section 2, an overview is given of the research that has been carried out on the incident capacity reductions over the past years, both in the US and in the Netherlands, since the values can differ for different countries. It also indicates how this study fits into the existing literature. Then, we discuss which data are used for analyzing flows. Section 4 discusses how the data were processed. The actual reductions of queue discharge rates in case of incidents are then shown in section 5. Then, the an example is presented which shows the practical relevance of this research. Finally, section 7 gives the discussion and the concluding remarks.

2. Literature review

Goolsby (1971) published an early overview of incident reductions due to incidents. The findings of his work and all other work mentioned in this section are stated in table 1. In nearly 40 years the traffic operations are likely to have changed. Apart from that, he determined the reduction compared to a reference capacity which he assumed the same for all locations.

Handbook values are useful for practitioners. However, often a description of the origin of the values lacks. This is also the case for an early publication of the Transportation Research Board by Blumentritt et al. (1981). Nowadays, the Highway Capacity Manual by the Transportation Research Board (2000) is the most referred source. Both values can be found in table 1.

The capacity values at incident sites might be dependent on the country since the capacities also vary per country. For the Netherlands there is also a handbook comparable to the Highway Capacity Manual, Nieuwe Ontwerprichtlijn Autosnelwegen, by the Dutch Road Authority

(2007b). It does not state a capacity for incident sites separately. However, there is an entry on the capacity for partially closed lanes (used mainly for road works), which is given in table 1. In general, the road capacity in a work zone is also lower (e.g., Dixon et al. (1996) or Kim et al. (2001)). That might lead to a conclusion that work zones and incident situations are comparable. However, Al-Kaisy and Hall (2001), and Heaslip et al. (2008) argue that driver familiarity also is important in the capacity of the work zone. Clearly, drivers can get used to a work zone, but drivers are never familiar to a incident situation. Another difference between an incident situation and a work zone is that there are proper markings for the work zones which increase the capacity which lack at incident sites.

Recently, two research projects are carried out in the Netherlands to determine the capacity in incident situations. Schrijver et al. (2006) use the Highway Capacity Manual as starting point. They modify these values based on experts' opinions. One of the changes is that they state that the queue discharge rate is 80% of the free flow capacity. The phenomenon that the queue discharge rate is lower than the free flow capacity is well known and discussed for instance by Hall and Agyemang-Duah (1991), Dijker et al. (1997), Cassidy and Bertini (1999) and Chung et al. (2007). Additionally, Schrijver et al. (2006) distinguish between different phases in the incident. The main distinction between the phases is the presence of emergency services (e.g, police, ambulance). They assume that the presence of workers on the roadway halves the remaining capacity.

The other project is reported by Van Toorenburg and Nijenhuis (2007). They use traffic data to check the values used in the report by Schrijver et al. (2006). They find a considerable difference for the situation in which there is an accident at the shoulder lane: 45% of the capacity is used, whereas Schrijver et al. (2006) assume 77% is still available. In their report, Van Toorenburg and Nijenhuis (2007) do not distinguish between different phases in the incident or different numbers of lanes closed.

Table 1. Overview of the remaining capacities

Type of blocking	Shoulder	1 out of 3 blocked	2 out of 3 blocked
Goolsby (1971)	0.67	0.50	0.21
Blumentritt et al. (1981)	0.84	0.53	0.22
Transportation Research Board (2000)	0.83	0.49	0.17
Schrijver et al. (2006)	0.77	0.35[1]	0.17
Dutch Road Authority (2007b)	-	0.36	0.17
Smith et al. (2003)	-	0.37	0.27
Van Toorenburg and Nijenhuis (2007)	0.67	0.50	0.21

There is no distinction for different lane closures: the value for an accident is given.

Regarding methodology, Smith et al. (2003) describe best how capacity values are found, with a more detailed description found in Qin and Smith (2001). Contrary to Goolsby (1971) and Blumentritt et al. (1981), Smith et al. (2003) describe the capacity as stochastic variable. In their research, macroscopic data of loop detectors is used to determine the maximum flow out of a queue. Thereby, their study describes the queue discharge rate, like all other quoted studies which describe a methodology. To the best of our knowledge there has been no research to the maximum possible flow in free flow state around an incident location. Our research also shows the queue discharge rate values. Section 3 will describe the methodology we applied.

Not much is known about the "rubbernecking effects", i.e. the reduction of capacity in the non-incident direction. The capacity reductions are only caused by a change of driving behaviour. Sinha et al. (2007) use several microscopic simulation tools to predict the reduction of capacity.

They use the developers' default values for the capacity reductions or rubbernecking effects. Knoop et al. (2009) presented a study based on empirical data, but for only one accident an queue discharge rate for a rubbernecking queue was found. The rubbernecking queue occurred on a freeway with 2 lanes per direction, whereas this study is restricted to freeways with 3 lanes per direction. The conclusion from Knoop et al. (2009) was that there is a considerable change in driving behaviour around incidents compared to normal data. Therefore, we also included the rubbernecking queues in this research.

This contribution fills several gaps. First of all, it is an extensive data analysis, using data of over 55,000 incidents. Even though there is a selection on the number of accidents, the remaining number of incidents (90) is still more than described in any other study. This is increases the existing knowledge on capacity, but the large number is also useful to show the variations in the capacity. Secondly, rather then giving the numbers in the handbook values, the article provides the methodology to perform this analysis. The contribution thereof is that it explicitly discusses how a bias, as often seen in existing methods, can be overcome. Thirdly, it compares the reduction of the capacity in the Netherlands, where the non-incident capacity is high, with the reduction in the USA where the non-incident capacity is lower. That shows whether the queue discharge rate at an incident are a fixed fraction of the queue discharge rate in non-incident situations. Fourthly, it provides insights into rubbernecking effects observed at traffic in the non-incident direction which are not studied on this scale before.

Additionally, this article also gives an example application for the use of these capacities, namely the assessment of possible incident management measures which reduce the incident handling time. Furthermore, it is essential to know the capacity to provide a good state estimation of future traffic conditions and these predictions form the basis for route advice as consequence of the incident.

3.   Set-Up of Data Collection

  The first part of this section describes which traffic data on during incidents are needed and which are available. The second part describes which data were used and which are the exact selection criteria. It also points out why the capacity reductions found here can only be applied in case of a traffic jam.

3.1   Type of Data

  In this contribution, we just discuss the queue discharge rate which differs from the free flow capacity (which is discussed in detail in section 4.2). The queue discharge rate is the flow at the location of a bottleneck. Bottlenecks are characterized by the condition that upstream traffic is in a congested state, and downstream traffic is in a free flowing state. To be able to find these bottlenecks, one needs to know the traffic state, which can be derived from the average speed at the road or alternatively, the occupancy along a road stretch.

The traffic states can be found by traffic flow data, which are often logged automatically, but reliable data on incidents lacks often since time and location have to be put into a database manually. Also information on the available lanes is required. It is necessary to combine the data on incidents with the data on the traffic flow properties, which makes a good registration of starting time, end time, and location essential. Also the number of lanes which are available are needed which is not recorded in the incident database. However, dedicated Variable Message Signs (VMS) in the Netherlands show every 500 meters for each lane whether it is opened or closed. The settings of these Variable Message Signs is logged automatically and therefore

accurately. From these logs the number of available lanes can be retrieved, which indicate the space and time of accidents.

3.2   Incident Selection

This contribution focusses on the capacity reduction on freeways with three lanes per direction. A large part of the Dutch freeway network is equipped with double loop detectors. At every 500 meter (about 600 yards) interval, there is a double loop detector. The aggregate data of the loop detectors is stored. This shows the average speed and the flow, both aggregated over 1 minute.

Five roadway stretches in the Netherlands were selected based on the available data and on incident frequencies. Table 2 presents these stretches.

Table 2. The used road stretches

Road number	Direction	From km	To km
A1	Eastbound	0	29
A1	Westbound	29	0
A2	Southbound	37	94
A2	Northbound	94	37
A4	Northbound	49	21

We divide the incidents in two categories: an accident and a broken down car. The requirements for both are different. We require the following:

• Upstream of the incident, the average speed is lower than 70 km/h (44 mph) and downstream of the incident, the average speed is over 70 km/h (44 mph).

• An accident needs to have a duration of at least 30 minutes and one or more lanes need to be closed for at least 30 minutes.

• A broken down car needs to be at the same spot for at least 15 minutes.

The first requirement says that only incidents which cause congestion are considered. Without congestion, there are no extra delays and the delays form the main incentive for this research. The other two requirements impose restrictions on the duration of the incident. If the duration is too short, there are too few measurements and the capacity can not be determined. There might be a few capacity measurements, but for reliable information the confidence bounds are needed. The spread of the capacities can only de derived from a series of measurements.

Incidents in the period January 1st, 2007 to July 31, 2007 are used. For further analysis, only incidents which are a bottleneck (i.e., cause a traffic jam, as explained in section 3.1) are included. These are now filtered based on the incident type (accident or car break down) and location (selected stretches). Table 3 shows the numbers of incidents which were suitable based on all criteria, including the requirement that the incident forms an bottleneck (the first requirement). The second column shows the fraction of incidents that fulfilled this requirement.

For the rubbernecking effects, the same criteria apply. Table 3 shows that the larger the disruption is, the more likely it is that an incident causes a traffic jam. For example, a rubbernecking queue occurred in just 7% of the incidents, but in 65% of the incident that blocked 2 out of the 3 lanes a queue occurred. Although it is noteworthy, this is not surprising since the capacity reduction in case of a physical blocking is much larger.

Table 3. Fraction of incidents used

Bold	Number used	Fraction passed bottleneck criterion
Shoulder lane	20	9%
1 lane closed	21	43%
2 lanes closed	20	65%
Rubbernecking	29	7%

Note furthermore that incidents included in the analysis for rubbernecking are not necessarily a subset of the cases for which the incident was valid. It is for instance possible that there is no queue in the incident direction and hence the incident is discarded for the analysis of the capacity in the incident direction. However, the same incident could create a traffic jam in the non-incident direction (if the traffic flow is high enough) and so be included for the rubbernecking analysis.

4 Data Analysis

To determine the reduction of the queue discharge rate at incident locations it is required to have good estimates of the queue discharge rate during the incident situation and in the normal situation. This section explains how the queue discharge rates are determined in case of an bottleneck during an incident situation, in section 4.1, and in normal conditions without an incident, in section 4.2. Section 4.3 presents how the queue discharge rates are compared and a road efficiency is computed.

4.1 Queue discharge rates in incident conditions

In order to determine the queue discharge rate, the incidents are selected in such a way that they form a bottleneck during the time of the incident. This means that the head of the queue is located at the incident site. Traffic flows out at the maximum flow rate possible for that location at that moment. The queue discharge rate at the incident site, the outflow out of the queue, can thus be derived from the counts at the downstream detector. This gives an average and median queue discharge rate as well as the interval bounds around it.

We use data aggregated over 1 minute, which is a relatively short time. Brilon et al. (2005) show the stochastic nature of the capacity measurement and discuss the consequences of a large or short aggregation time. The short time available for measurements, the duration of an incident, requires to make a choice on fluctuating measurements aggregated over a short time or a few more stable measurements aggregated over a longer time. Taking a short interval means that the spread of the measurements is larger, but more important, it will provide sufficient measurements to have an indication about the reliability of the value. Although the median value might slightly change due to stochastic effects, there is no bias of choosing a short aggregation interval. Therefore, we choose a short aggregation time of 1 minute.

4.2 Queue discharge rate in normal conditions

The queue discharge rate is site-specific, like normal capacities. Apart from the site-specific influences, there are likely to be behavioural influences, which can be best described if one computes the relative queue discharge rate compared to the normal queue discharge rate. This section explains how the normal queue discharge rate is computed, even for sites which are not a bottleneck in normal conditions.

There are many methods to determine the free flow capacity in the bottleneck, such as the Product Limit Method as introduced by Kaplan and Meier (1958). Other methods, like the one proposed by Brilon et al. (2005), can compute the capacity at any point, but give the free flow capacity. These is generally based on analysing the maximum flow. The location of the incident is in general not a bottleneck for non-incident situations. Therefore, it is not possible to use the same method as during an incident (described in the last paragraph) and another method has to be used. To find the queue discharge rate, we use a fit of an reverse-lambda shaped fundamental diagram, as proposed by Koshi et al. (1981), in which the intersection of the fit of the free flow branch and the congested branch is taken as queue discharge rate – see figure 1. Even though the incident locations are not a bottleneck in normal conditions, there is congestion from time to time due to growing queues caused by downstream bottlenecks. Therefore also at the incident locations there are points in the fundamental diagram in the congested branch.
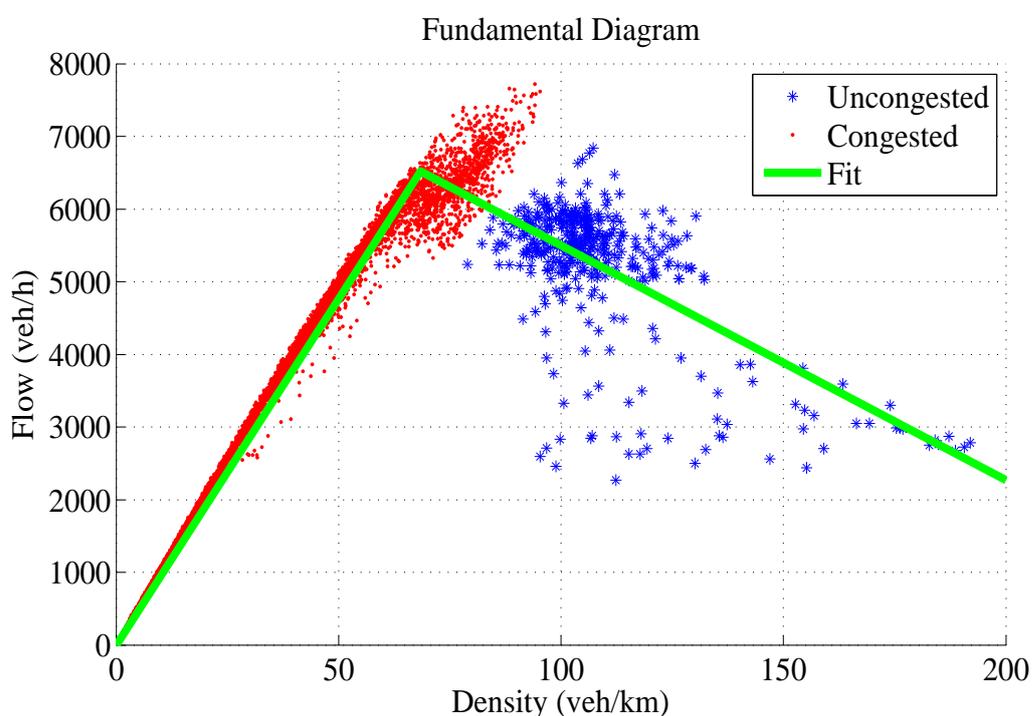

Figure 1.The fitted fundamental diagram for one incident

Previously collected data were obtained for the periods of 10 days before and 10 days after the incident. If the average speed was under 70 km/h (44 mph), the traffic state was classified as congested traffic; if the average speed exceeded 70 km/h, the traffic state was classified as uncongested. For both traffic states (both branches of the fundamental diagram), we made a linear fit in the density-flow diagram. The queue discharge rate is found at the point where these two lines intersect. Note that the free capacity can be found on the free flow branch which is usually higher than the intersection point, see also for instance Hall and Agyemang-Duah (1991), Dijker et al. (1997), Cassidy and Bertini (1999) or Chung et al. (2007); this is called the capacity drop.

4.3   Comparing efficiency

This section uses the variables as used as shown in Table 4.

Table 4. The used variables

Variable	Unit	Meaning
$n_{incident}$	-	The number of lanes available for traffic in one direction during the incident
$N_{non\text{-}incident}$	-	The number of lanes available for traffic in one direction in normal conditions
$C_{incident}$	veh/h	The queue discharge rate during the incident
$C_{non\text{-}incident}$	veh/h	The queue discharge rate in normal conditions
F	-	The fraction of queue discharge rate that remains available
R	-	The reduction of the queue discharge rate
η	-	The efficiency of the road use in incident situations

C is the total hourly queue discharge rate for all lanes. Under normal conditions the queue discharge rate is $C_{incident}$, and during an incident this is reduced to $C_{non\text{-}incident}$. The quotient of the two queue discharge rates is the fraction of the capacity that remains, F:

$$F = \frac{C_{incident}}{C_{non-incident}} \qquad (1)$$

The reduction of the queue discharge rate is a combined effect of the reduction of the number of lanes and the less efficient use of the remaining lanes. We therefore expressed the efficiency η of the use of the remaining lanes by dividing the capacity factor by the fraction of the roadway that is available:

$$\eta = \frac{F}{\dfrac{n_{incident}}{n_{non-incident}}} = \frac{C_{incident}}{\dfrac{n_{incident}}{n_{non-incident}} C_{non-incident}} \qquad (2)$$

In this formula, n is the number of lanes that is available. The use of the formula is best explained by an example, which is a (fictitious) situation where the capacity reduces to 40% if 1 out of 3 lanes is closed. The remaining 2 lanes could have provided 67% of the original capacity with unchanged traffic behaviour, but the actual flow is only 40% of the original capacity. So, the road is used at 40/67=60% of the original efficiency. Note that this is an example and the real outcomes of the study can be found in section 5

In the methodology applied in this study (and in any other) is a possible bias towards higher capacity reductions (capacity reduction R is 1-F). Incident locations with a higher capacity reduction are a bottleneck at lower demand levels. This means that incidents with large capacity reductions are more likely to be incorporated in the study, which biases the results. Assuming a uniform demand distribution over the time (e.g., day), the probability to find an incident blocking the road is proportional to the capacity reduction R. This is expressed in Figure 2. One could check how biassed the set of incidents is. To this end, the incidents are first grouped in bins with similar reductions. Then, one can divide the found number of incidents by the probability for each bin of capacity reduction. However, this might give unbalanced results since one incident can get a very high weighting factor and this way the distribution might depend heavily on one or two incidents that are included. This is not desirable since the it is a stochastic process which incidents are included. We therefore choose to check the corrected distribution, ignoring some incidental cases, and to present the more homogeneous set of uncorrected (but possibly biassed) capacity reductions. Note that this bias is only a serious concern if the capacity reduction varies much. In the limit where no variation of the capacity reduction exists, there is no bias at all.
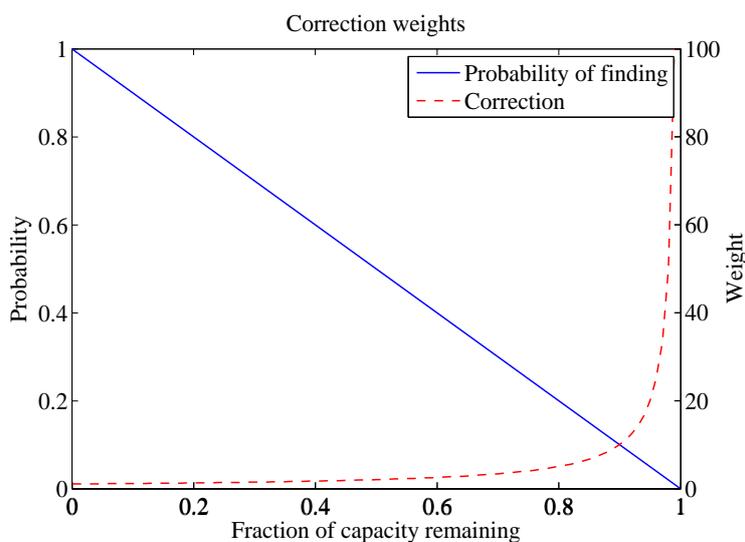
Figure 2. The probability that an incident causes a queue and the correction factor

5.   Results

Figure 3 shows for one incident the fluctuation of the flow over time. In this particular incident, one out of the three lanes was closed due to the incident. Consequently, two lanes remain opened. The median flow is indicated with a dotted line. The fluctuation is larger because we aggregated over short time intervals of 1 minute. For the example given here, the median is 1919 vehicles/h and a standard deviation is 209 vehicles/h.
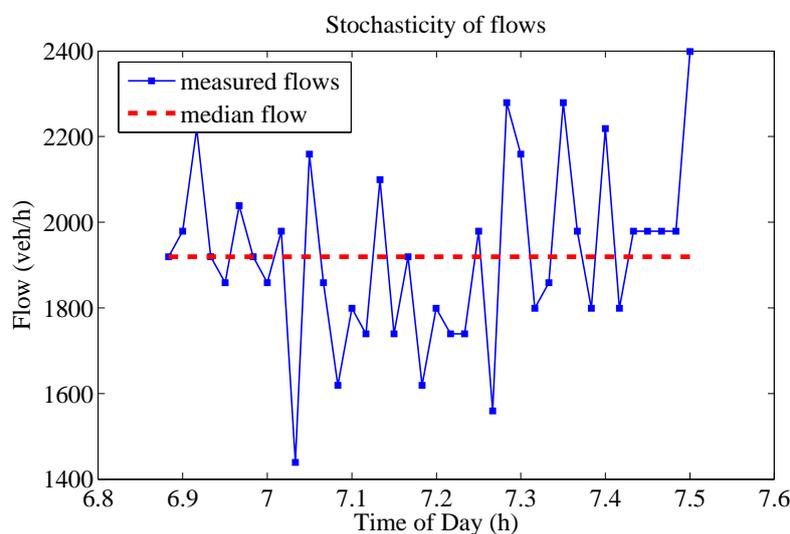

Figure 3. The fluctuation of the capacity flow for one incident

For the same incident location, Figure 1 shows the construction of the reference capacity. The free part and the congested part of the diagram are separated based on a 70 km/h threshold. This value is based on experience for the critical speed in the Netherlands. The figure shows congested and non-congested measurements, as well as linear fits thereof. Figure 1 shows that this value nicely agrees with the critical speed, the speed at the point where the flow is maximum. The flow value at the point where the fits of both branches cross is the queue discharge rate. For this

incident the reference queue discharge rate is 6500 vehicles/h over three lanes. Since the queue discharge rate is 1919 vehicles/h (see last paragraph), the resulting capacity factor F for this incident now is 1919/6500=0.30.

For all incident situations we computed the capacity factor F as mentioned in equation 1, which were grouped by category, being the shoulder lane blocked, 1 of the 3 lanes blocked, 2 of the 3 lanes blocked, and rubbernecking. In Figure 4 the full distributions of the capacity factors per group can be found. The figure shows a spread of distributions even within one group. Furthermore, it is remarkable that the group where the shoulder lane is closed and all driving lanes are available still shows a considerable capacity reduction. The same holds for the case of rubbernecking. Figure 5 combines the values of the capacity factors. In the figure, the median value is indicated by the middle line, and the 25th and 75th percentile values are indicated by the edges of the box. The line ends show the range of the capacity reductions. Some incidents cause a reduction outside the normal range; these reductions are indicated with a red cross in Figure 5. Due to a lack of detailed information, we could not analyse in detail why these situations were different.
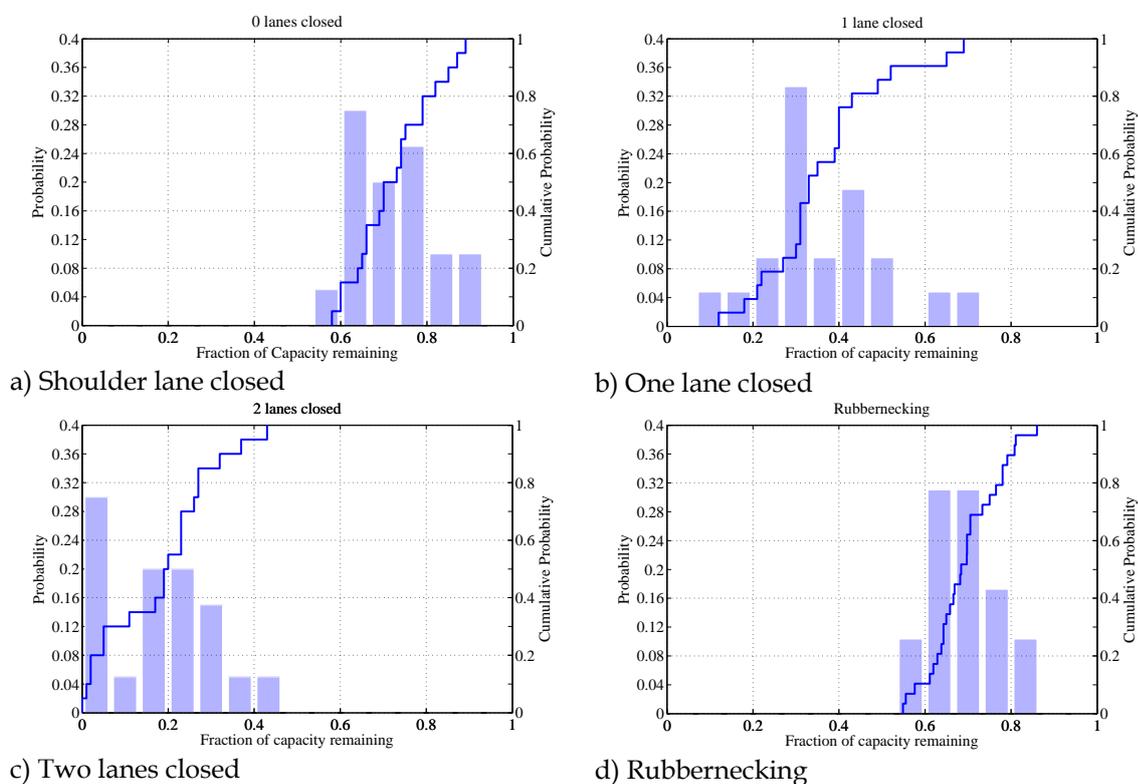

a) Shoulder lane closed b) One lane closed

c) Two lanes closed d) Rubbernecking

Figure 4. The capacity factor for the three types of incidents which block the roadway and for the rubbernecking

Table 5. The resulting capacity factors F

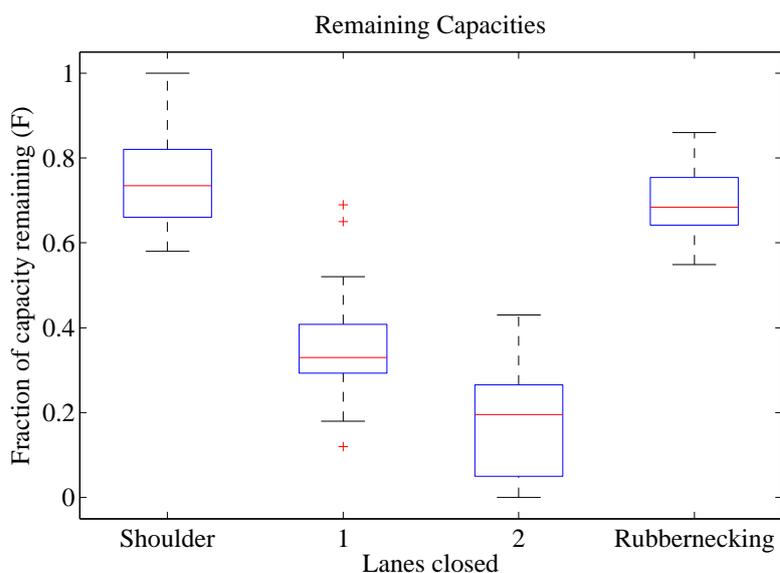| Type of blocking	Shoulder	1 out of 3	2 out of 3	Rubbernecking
Mean	0.72	0.36	0.18	0.69
Standard deviation	0.09	0.14	0.12	0.08
Efficiency of Lane Use η	0.72	0.54	0.54	0.69

Figure 5. The capacity factors found for different incidents

Table 5 shows the results in numbers. The table states the mean values of the capacity factors and the standard deviation. It also states how efficient the remaining lanes are used. This is calculated by dividing the capacity factor by the fraction of the number of lanes of the roadway that is available, see equation 2.

There were 29 rubbernecking queues that met all criteria and could be analysed using the method described in section 4.3. For these queues, the fraction of the queue discharge rate that remains, F, equals the efficiency rate η in formula 2 since no lanes are closed. This efficiency value (0.69 as shown in Table 5) is higher than stated earlier in a microscopic analysis by Knoop et al. (2009). However, that earlier work computed the quotient of the queue discharge rate during an incident and the *free* capacity. This value can be higher because queue discharge rate is lower than the free capacity (the capacity drop) and thus the queue discharge rate expressed as fraction of the changing reference value is higher. At the other hand, this decrease is much more than predicted by the simulators in using default settings by . Another interpretation for the efficiency of 69% is that the queue discharge rate at the incident site is (1-69%=) 31% lower than in normal conditions.

When one or two of the driving lanes are blocked, the efficiency of the remaining lanes reduces to 54%. Note that the efficiency factor indicates how much the driving behaviour changes compared to normal driving. That this number is the same for the case one driving lane is blocked and for the case two lanes are blocked means that both situations lead to the same behavioural effects. In general, the efficiency factor η is higher in case there is no disruption in one of the lanes normally available for traffic, so in case there is a blocking on the shoulder lane or on the carriageway in the opposite direction. We conclude that if the driving lanes are disturbed, this affects the driving behaviour more. A check as explained in 4.3 shows the effect of a possible bias. For the cases with 1 or 2 lanes blocked, this bias turned out to be negligible. For the cases of rubbernecking and a blocking of the hard shoulder, the bias in efficiency is around 0.05.

6.   Case Study: Incident Management

The capacity reductions found in this study are relevant for various purposes. One of them is explained here where it is analysed what the effect of quicker actions of incident management

would be qualitatively. First, the case is explained in section 6.1. Section 6.2 then gives the results of the case study.

6.1   Case set-up

This section discusses the possible effectiveness of shortening the clearing of an incident. It describes how it can be found how the queue length and the delay change if these processes are sped up. In this analysis we use the capacity factors found from the data analysis.

Usually, the emergency services start working on the roadway (on the lanes for through traffic) but move the wrecks as soon as possible to the shoulder lane. According to the Dutch study on incident management by the Dutch Road Authority (2007a), we distinguish three phases during the incident: (1) the time until the first emergency services arrive, (2) the time that the emergency services work on the roadway and (3) the time that the emergency services work at the shoulder lane. For this case study, it is supposed that during the first phase, one of the three lanes is blocked. When the emergency services arrive, a second lane is blocked to have a safer working space. In phase 3 only the shoulder lane is blocked.

The queue discharge rates which come out of the capacity analysis can now be used to determine the maximum flow during each of these phases. The incident management study not only determines the phases, but also gives the average time for each of these phases. To determine the delay, the demand for a non-incident day and the capacities during the incident are put into a traffic simulator. This equals a situation where people would not change their route because of the incident. The simulator now predicts the queue length and delay if the outflow was blocked for a while by the incident. Note that only the flow values for a non-incident situation can be used, otherwise flows do not indicate the demand, but are limited at the queue discharge rate. We use a vertical queuing model to compute the delay. That means that vehicles will encounter delay at the moment they pass the incident location.

We analysed the differences in queue length and total delay if the time in different phases reduces by 2 or 4 minutes. We derive two measures from the simulations: the duration of the queue and the total delay. It will show the effect of shortening a period when all other periods remain the same. It can be expected that the sensitivity is largest for the shortening of the period with the lowest capacity. In that phase, the effective inflow of vehicles into the queue, and so the queue growth, is the largest.

In the analysis we quantify how much the queue duration and the total delay of the queue can be decreased. We analyse the effectiveness of a time gain for two different, typical (but fictitious) incidents in the peak period.

6.2   Results of case study

Figure 6 graphically shows the used method to compute the indicators for one set of durations per time phase. The changing capacity is shown, as well as the demand, which fluctuates at a high frequency. The queue length is computed as the cumulative sum of the difference between the demand and the capacity and is also shown in the figure. From this derivation, the indicators queue duration and delay can be computed.
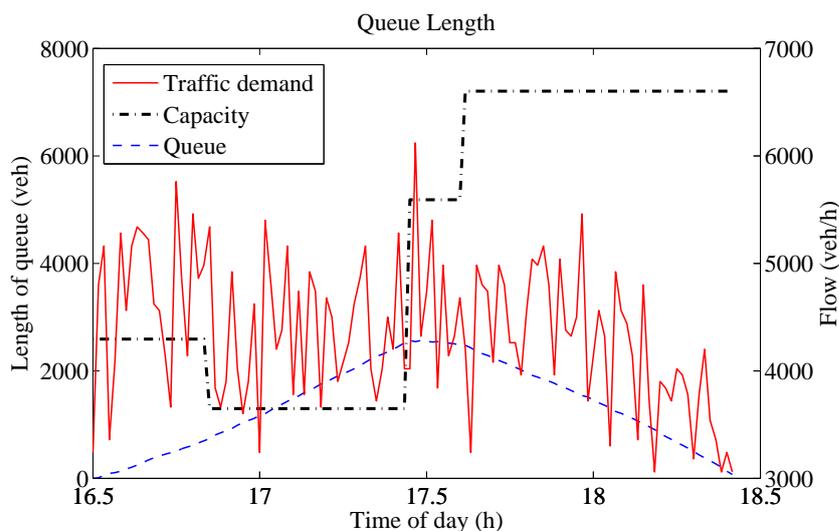
Figure 6. The length of a queue for a fictitious incident

Table 6 shows the effects of the shortening of actions. We simulated the traffic with a 2 and 4 minute shortening for each of the phases. The total effects are divided by the decrease of shortening of the phase in order to get comparable sensitivity numbers for both 2 and 4 minutes shortening. For both times, comparable results in reduction of delay and queue length are per minute shortening are found.

A error that can easily made is that one expects a large sensitivity for the earliest phases because all following vehicles profit from the total shortening. However, this reasoning does not hold. It is best explained graphically (Figure 7). In the example we show results for a case with a constant demand of 4000 veh/h and a reference capacity of 8000 veh/h. For the capacity reductions at each of the phases we use the values of the according link configuration (as explained in section 6.1) stated in Table 5. For the duration we used the results of the study of the Dutch Road Authority (2007a).

Table 6. The effectiveness of shortening the actions in Incident Management

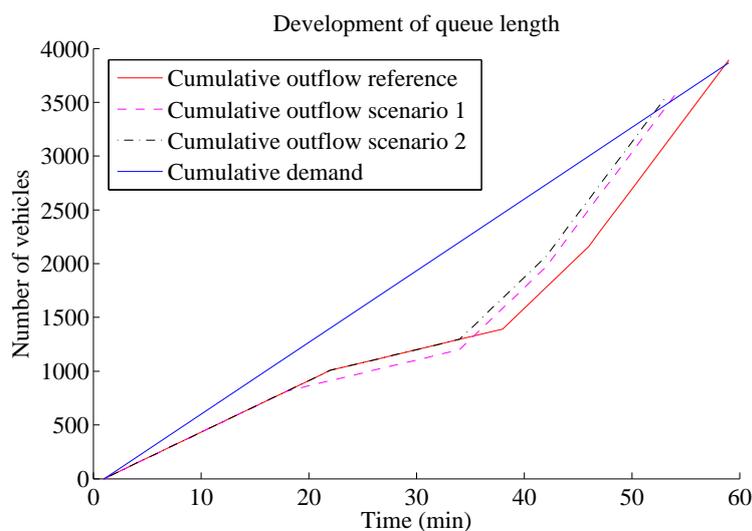|	Decrease of queue duration min queue length/min phase reduction	Decrease of delay vehicle hours/min phase reduction
Δt = 2 min		
Arriving	2.2	70
Service on roadway	2.8	101
Service on shoulder lane	0.6	21
Δt = 4 min		
Arriving	2.5	77
Service on roadway	3.2	106
Service on shoulder lane	0.85	30

Figure 7. The delays for different scenarios

The figure shows the cumulative departure and arrival curve at the (vertical) bottleneck. The area between the two curves is the total delay. Therefore it holds that the larger the area that area is, the larger the delay is. We plot the same curve for three situations: the reference situation using the actual lengths of a period, a situation with a shorter first phase (scenario 1) and a situation with a shorter second phase (scenario 2). The second phase is the phase in which emergency services are working on the freeway and the capacity is reduced most. The figure shows that the delays are most reduced if phase 2 is shortened (scenario 2). One of the reasons for this is that in both scenarios the end of phase two is at the same moment. This way it can be reasoned that the shortening the phase with the largest capacity reduction is most effective.

This is also shown in the numbers: a shortening of the period with the largest capacity reduction, the period of emergency workers at the roadway, has the most effect. If the emergency services block 2 lanes of the roadway one minute less, the duration of the queue reduces with about 3 minutes (and 100 vehicle hours of delay). For the reduction of the time until the emergency services arrive, these gains are lower, but still considerable.

7.   Conclusions and Discussion

We analysed the maximum outflow out of a jam which is caused by an incident analyzing 90 traffic jams. The most important finding is that the capacity per lane reduces significantly due to a change in driving behaviour. The size of this reduction depends on the incident type.

If one of the driving lanes is blocked, the remaining lanes are used 46% less efficient, which yields an "efficiency factor" of 54%. To compute the resulting queue discharge rate, one has to do the following. First, one has to take the proportional part of the road that is available (e.g., 33% if 2 out of 3 lanes are blocked). To compute the capacity reduction one takes the proportionally factor (33%) and multiplies this by the efficiency factor (54%). This is how much of the normal queue discharge rate remains (18%). Note that this is the reduction compared to the normal queue discharge rate; the reduction compared to the free flow capacity is even larger.

If there is an incident at the shoulder lane, the efficiency reduces by 28%. This is only due to a change in driving behaviour since all the lanes are open. A similar efficiency drop (31%) is found in case there is an incident at the other side of the guardrail, the "rubbernecking" effect.

Scientifically, the value of the article is threefold. First of all, it shows the size of the effect of distraction at the incident site. The road efficiency is several tens of percents lower than in normal conditions, which means that the drivers distraction plays an important role. Secondly, it is the first time that it is shown that how much capacity in the non-incident direction decreases because of an incident at the other side of the guardrail. This effect is solely due to a change in driving behaviour. For both directions, the article shows the bandwidth of the road efficiencies for different types of configurations. Thirdly, for scientific purposes, it is important to have the methodology of the research described, rather than just a value in a handbook. The article provides also insight in the used methodology and the possible flaws in it, being a possible bias due to a selective inclusion of capacities.

Practitioners can use this paper in the following ways. In the first place, the found capacity during an incident can be used by the road authority. The capacity is one of the most important road characteristics for traffic engineers. For instance rerouting of traffic in a situation after an incident will be based on the capacity of the blocked road. In case there is no other alternative, travellers can be informed about the delay. Knowing the large influence of the distraction, it is worthwhile for practitioners to try reducing the distraction and therefore increase the capacity. In the longer term, the road authority can experiment with increasing the road capacity. One of the possibilities would be to place screens in the middle of the road to prevent travellers in the non-incident direction of looking at the incident site. This could decrease the distraction and thus increase the road efficiency and capacity.

Another possibility is the quicker handling at the incident site. The impact of such a strategy is assessed quantitatively in the case study described the article. Delay can be avoided most by reducing the servicing time on the roadway and the time until the emergency services reach the incident site. These findings can provide the basis on a decision to invest in a quicker emergency response system: the benefits can now be shown and can, for individual measures, be compared with the costs.

Finally, the efficiency values found in this study for the Netherlands can be placed in a international context. Some values can be compared directly with values in literature for other countries. Comparing these shows that the reductions are similar for different countries, although there the capacity is usually lower than in the Netherlands. Therefore it is likely that the other relative reduction values which are not studied elsewhere, like the rubbernecking, are also valid in other countries, which is a value of the article for practitioners.

Acknowledgements

References

Al-Kaisy, A. and Hall, F. (2001). Examination of the Effect of Driver Population at Freeway Reconstruction Zones. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 1776, pp. 35-42.

Blumentritt, C.W., Ross, D.W., Glazer, J., Pinnell, C. and Mccasland, W.R. (1981). Guidelines for selection of ramp control systems. Report number: 232, Transportation Research Board, Washington D.C.

Brilon, W., Geistefeldt, J. and Regler, M. (2005). Reliability of Freeway Traffic Flow: A stochastic Concept of Capacity. In Proceedings of *16th International Symposium on Transportation and Traffic Theory*, College Park, Maryland.

Cassidy, M.J. and Bertini, R.L. (1999). Some traffic features at freeway bottlenecks. *Transportation Research Part B: Methodological,* Vol. 33**,** No. 1, pp. 25-42.

Chung, K., Rudjanakanoknad, J. and Cassidy, M.J. (2007). Relation between traffic density and capacity drop at three freeway bottlenecks. *Transportation Research Part B: Methodological,* Vol. 41**,** No. 1, pp. 82-95.

Dijker, T., Bovy, P.H.L. and Vermijs, R.G.M.M. (1997). Car-following under non-congested and congested conditions. Report number: VK 2206.301, Delft University of Technology, Delft.

Dixon, K.K., Hummer, J. E. and Lorscheider, A.R. (1996). Capacity for North Carolina Freeway Work Zones. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 1529, pp. 27-34.

Dutch Road Authority (2005). Nota Mobiliteit. Report number: 2004-2005, 29644 nr. 13, Ministerie van Verkeer en Waterstaat and VROM.

Dutch Road Authority (2007a). Inventarisatie beleidseffecten incidentmanagement. Report Rotterdam.

Dutch Road Authority (2007b). Nieuwe Ontwerprichtlijn Autosnelwegen. Report number: ISBN 90 3693 636 5, Rijkswaterstaat Adviesdienst Verkeer en Vervoer, Rotterdam.

Goolsby, M.E. (1971). Influence of Incidents on Freeway Quality of Service. In Proceedings of *The 50th Annual Meeting of the Transportation Research Board*, Washington D.C.

Hall, F.L. and Agyemang-Duah, K. (1991). Freeway Capacity Drop and the Definition of Capacity. *Transportation Research Record: Journal of the Transportation Research Board,* Vol.1320, pp. 91-98.

Heaslip, K., Louisell, C. and Collura, J. (2008). Driver Population Adjustment Factors for the Highway Capacity Manual Work Zone Capacity Equation. In Proceedings of *the 87th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association,* Vol. 53**,** No. 282, pp. 457-481.

Kim, T., Lovell, D.J. and Paracha, J. (2001). A New Methodology to Estimate Capacity for Freeway Work Zones. In Proceedings of *the 80th Annual Meeting of the Transportation Research Board*, Washington D.C.

Knoop, V.L., Hoogendoornm S.P. and Van Zuylen, H.J. (2009). Microscopic Traffic Behaviour near Incidents. In Proceedings of the *International Symposium of Transport and Traffic Theory*. William H.K. Lam, S.C. Wong and Hong K. Lo (eds), Springer, New York, page 75-97.

Koshi, M., Iwasaki, M. and Ohkura, I. (1981). Some findings and an overview on vehicular flow characteristics. In Proceedings of *International Symposium on Transportation and Traffic Theory*, Toronto.

Kwon, J., Mauch, M. and Varaiya, P. (2006). The components of congestion: delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. In Proceedings of *the 85th annual meeting of the Transportation Research Board*, Washington.

Qin, L. and Smith, B.L. (2001). Characterization of accident Capacity Reduction. Report number: STL-2001-02 University of Virginia, University of Virginia.

Schrijver, J., Immers, B., Snelder, M. and Jong, R.D. (2006). Effecten van de landelijke invoering van incidentmanagementmaatregelen op de voertuigverliestijd in het netwerk (Effects of nation-wide introduction of incident Management Measures for vehicle delay in the network). Report number: 06.34.15/N079/034.65116/JS/LK, TNO, Delft.

Sinha, P., Mohammed Hadi, P.E. and Amy Wang, E.I. (2007). Modeling Reductions in Freeway Capacity due to Incidents in Microscopic Simulation Models. In Proceedings of *86th Annual Meeting of the Transportation Research Board*, Washington D.C.

Smith, B.L., Qin, L. and Vankatanarayana, R. (2003). Characterization of Freeway Capacity Reduction Resulting from Traffic Accidents. *Journal of Tranportation Engineerging,* Vol. 129**,** No. 4, pp. 362 - 368.

Transportation Research Board (2000). Highway Capacity Manual. Report Washington D.C.

Van Toorenburg, J.A.C. and Nijenhuis, T.A. (2007). Capaciteitsbeperking door incidenten. Report Transpute, Amersfoort.