Assessing Traffic Safety of Dutch Weaving Sections
Validation of the Surrogate Safety Assessment Model combined with VISSIM

Oude Vrielink, Ilse; Broeren, Patrick; Farah, Haneen; Hoogendoorn, Serge

Publication date
2018

Document Version
Accepted author manuscript

Published in
Transportation Research Board Conference Proceedings 2018

Citation (APA)
Oude Vrielink, I., Broeren, P., Farah, H., & Hoogendoorn, S. (2018). Assessing Traffic Safety of Dutch Weaving Sections: Validation of the Surrogate Safety Assessment Model combined with VISSIM. In *Transportation Research Board Conference Proceedings 2018* Transportation Research Board (TRB).

ASSESSING TRAFFIC SAFETY OF DUTCH WEAVING SECTIONS
Validation of the Surrogate Safety Assessment Model combined with VISSIM

Ilse Oude Vrielink, Corresponding Author
Arcadis Nederland B.V.
Piet Mondriaanlaan 26, 3812 GV Amersfoort, The Netherlands
Tel: +31 (0)6 2706 1179, Email: ilse.oudevrielink@arcadis.com

Patrick Broeren
Arcadis Nederland B.V.
Beaulieustraat 22, 6814 DV Arnhem, The Netherlands
Tel: +31 (0)6 466 47 203, Email: patrick.broeren@arcadis.com

Haneen Farah
Department of Transport and Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, 2628 CN Delft, The Netherlands
Tel: +31 (0)15 278 3401, Fax: +31 (0)15 278 7956, Email: h.farah@tudelft.nl

Serge Hoogendoorn
Department of Transport and Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, 2628 CN Delft, The Netherlands
Tel: +31 (0)15 278 5475, Fax: +31 (0)15 278 7956, Email: s.hoogendoorn@tudelft.nl

Word count: 250 words abstract + 5553 words text + 6 tables/figures x 250 words (each) = 7303 words

Submission Date: August 1, 2017

ABSTRACT

Dutch road designers and safety experts are searching for more quantitative methods to evaluate the safety of a (proposed) weaving section than the traditional methods of expert judgement and use of accident records. An alternative would be to determine safety using VISSIM micro-simulation models in combination with the Surrogate Safety Assessment Model (SSAM).

SSAM calculates the number of conflicts (i.e. observable situations in which two or more road users approach each-other resulting in a potential collision risk) that occurred in a micro-simulation model using surrogate safety measures. This study evaluates this method for Dutch weaving sections by comparing the number of conflicts observed from VISSIM microsimulation models combined with SSAM with the crash rate, and other criteria.

Nine Dutch weaving sections were selected and ranked based on four criteria: (I) crash rate, (II) conflict rate calculated from VISSIM-simulations using SSAM, (III) number of crashes expected based on a previously developed crash prediction model, and (IV) road safety experts judgement.

To examine the correlation between the different rankings, the Spearman Rank Correlation Coefficient was calculated between each two rankings. The correlation of 0.567 ($\rho_s = 0.112$) between the crash and conflict rate ranking suggests a reasonable, but insignificant correlation.

In a sensitivity analysis the effects of some micro-simulation settings, conflict analysis thresholds, and the calibration method were assessed. Although different than expected, extending the calibration process resulted in a weaker correlation. Hence care should be taken when using VISSIM and SSAM to evaluate the conflict rates as (only) safety-predictor of Dutch weaving sections.

Keywords: Freeway Weaving Sections, SSAM, Conflicts, Surrogate Safety Measures, Safety, Crash prediction

Introduction

On highways so-called weaving sections are applied when the point of convergence and point of divergence of two merging and splitting traffic streams are within a short distance. In these weaving sections many vehicles in close proximity switch lanes, which results in a complex driving task, disturbances and conflicts (2). In general, the number of accidents showed to be higher on weaving sections than on other regular freeway sections (3).

The Dutch national road authority Rijkswaterstaat has guidelines on how to design weaving sections (4,5). However, there are multiple variations in the design, in number of lanes, (as)symmetry, and length. Traffic characteristics such as intensity and vehicle composition also influence the traffic flow (5). Due to the high density of the Dutch road network, weaving sections are applied relatively frequently in the Netherlands, but due to lack of space it is often difficult to design the weaving sections exactly according to the guidelines. Therefore, it is important to understand the implications of deviating from the guidelines on traffic safety.

When designing weaving sections often multiple design options are possible for one location, and currently in most cases a choice is made based on expert judgement. Therefore, it would be valuable to determine the degree of road safety from a microscopic simulation already during the design process. Some attempts to develop crash modification factors (CMFs) and crash prediction models (CPMs) as alternatives have taken place. For example, Iliadi et al. (1) researched the effects of design elements and traffic flow characteristics on the safety of symmetrical weaving sections in the Netherlands by developing such CPM. However, the derivation of such factors and models is a complex task and not all relevant factors can be included, and also such studies do not focus on vehicles interactions influencing the origin of the accident.

Another possibility is to derive the safety of a road section from a micro-simulation using surrogate safety measures. These surrogate safety measures might be a replacement (or addition) for crashes when evaluating the safety of a weaving section. The US Federal Highway Administration (FHWA) developed the Surrogate Safety Assessment Model (SSAM) which can calculate surrogate safety measures from micro-simulations. Thus, micro-simulations can be a good alternative when for a weaving section no accident frequencies are available. A simulation is especially useful when designing a new weaving section as there is no accident data record available at that time.

However, such micro-simulations are developed for analysing traffic performance, and not for safety purposes which requires more details regarding vehicle interactions and the inclusion of human factors in the mathematical models. Hence it is uncertain whether SSAM can be used to assess the safety of Dutch weaving sections and how accurate the outcomes of SSAM in combination with VISSIM are. Therefore, the main goal of this research was to determine if combining the micro-simulation model VISSIM and SSAM is a reliable method to predict the traffic safety of Dutch weaving sections.

Literature

In general, the likelihood of being involved in a crash on a freeway is larger on weaving sections than on regular freeway sections (3). The extra high number of lane change manoeuvres in a weaving section with high traffic volume and speeds' variability often results in an increased number of unsafe situations (3). This is due to the weaving traffic resulting in more conflicts between the vehicles entering and exiting, leading to a more complicated driving task (6).

There are two main methods to determine the safety of a weaving section. The most straightforward method is by analysing historical accident data. However, it is a well-known problem that official accident statistics are incomplete and biased (7). Of all accidents that occur, some are not reported to the police. And if the accident is reported, the data is often incomplete or incorrect. Next to accident counts, also surrogate safety measures can be used to assess safety. The use of surrogate safety measures and conflicts has become common in recent years.

Conflicts and surrogate measures

Surrogate safety measures are measures other than actual crash frequency that represent the degree of safety (8). Most surrogate safety measures use conflicts for that. A conflict is defined by (9) as:

'*an observable situation in which two or more road users approach each other in time and space for such an extent that there is a risk of collision if their movements remain unchanged'*.

As illustrated by the pyramid of traffic events (10) shown in Figure 1 conflicts also include events that do not lead to real crashes. The area of the layer describes the frequency, while the distance of the layer from the base represents the severity of the events (10). The benefit of analysing conflicts is that they are observed more frequently than crashes. However, the disadvantage of using conflicts as a measure for safety is that conflicts do not directly give the number of accidents that occurs. However, there are some attempts to relate the number of conflicts and crashes by a formula (11).

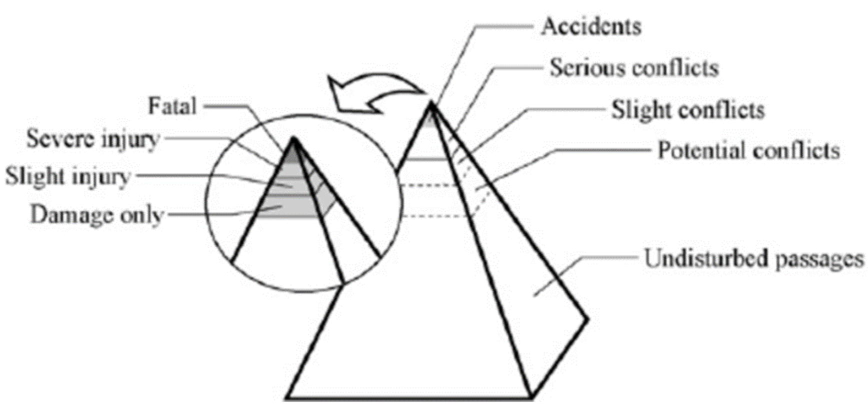

Figure 1: Pyramid of traffic events (adapted from (10))

Conflicts can be determined from simulated vehicle trajectories by calculating surrogate safety measures and assessing whether the calculated value exceeds a certain threshold value or not. There are many surrogate safety measures available. Two frequently used surrogate safety measures are the Time to Collision (TTC) and Post Encroachment Time (PET). The TTC indicates the time span left before two vehicles collide, if nobody takes evasive action. The PET represents the difference in time between the passage of the 'offending' and 'conflicting' road users over a common area of potential conflict (12).

Safety of Weaving Sections

Some research is available that focusses on the safety of weaving sections in the Netherlands. Already in 1975, Brouwer (13) concluded that the likelihood of a crash increases strongly for shorter weaving sections and weaving sections with a high traffic flow. The latter is not surprising as more vehicles lead to more conflicts and thus more crashes.
Iliadi et al. (1) included a sample of 110 symmetric weaving sections distributed over the motorway network in the Netherlands to develop a crash prediction model. Several factors were investigated and were included in the crash prediction model. Factors that were found significant and hence were included in the final model were: the length of the weaving section, AADT, the number of lanes on the main freeway, the share of weaving cars and the location of the weaving section relative to an interchange. Factors that were investigated but not included in the final model were the share of trucks, share of weaving trucks, the interchange type (i.e. cloverleaf, clover-turbine, etc.) and symmetry.

A larger part of the available literature focuses on data analysis and comparing weaving section types, or on developing crash modification factors and formulae to predict the number of crashes based on certain road design and traffic flow characteristics. However, also simulation models were proposed for assessing traffic safety.

Bared (11) emphasized that a major benefit of using simulation models is that there is no need for having a sufficient large accident data base. Moreover, the analysis of accident data is a slow process and results are influenced by the infrequent and random nature of crashes. Bared (11) found a relationship between conflicts per hour and crashes per year. However, there is a need for further research on the interpretation and comparison of such surrogate safety measures.

Yang et al. (14) also pointed on the benefit of using micro-simulations over other methods which have limitations due to data availability.

Gettman et al. (8) did an evaluation study of SSAM for the FHWA. They performed theoretical tests which compared pairs of simulated design alternatives and a field validation exercises which compared output from the real world to the simulation output. The comparison of design alternatives did not always lead to a clear design preference but rather a trade-off of surrogate safety measures. The simulation conflicts were found to correlate weakly but significantly to the field crash data ($\rho_s = 0.463$). After a sensitivity analysis it was concluded that volume-based prediction models provide a better correlation to field data.

Also the research by Huang et al. (15) had as an objective to identify if a combination of the VISSIM simulation model and the SSAM approach provides reasonable estimates for traffic conflicts. Their focus was on signalised intersections. Results of the data analysis showed a goodness-of-fit ($\rho = 0.916$) that indicates a reasonable agreement between safety ranks based simulated and field observed traffic conflicts. However, the simulated conflicts were not good indicators for conflicts that occurred due to unexpected driving manoeuvres such as illegal lane-changes in the real world.

El-Basyouny (16) also performed a field validation of SSAM by comparing predictive safety performance capabilities of SSAM with actual accident experience at Canadian signalised intersections. A poor relation was found, and it was concluded that traffic volumes can explain more variation in occurrence of accidents than simulated conflicts obtained from SSAM. The poor relation could be associated to how an intersection was modelled in VISSIM as changing model parameters resulted in considerable variations in the number of conflicts.

Essa and Sayed (17) investigated the transferability of calibrated parameters in VISSIM for safety analysis between different sites. Six parameters were identified as important for the safety analysis. Two of them (headway time and desired deceleration) were directly transferable, three (standstill distance, reduction factor for safety distance closed to stop line and start upstream of stop line) were transferable in some degree and one (negative and positive following thresholds) was not transferable. They also mention that first calibrating on delay times and thereafter calibration of driving behaviour parameters results in a stronger correlation between field-measured conflicts and simulated conflicts. By transferring calibrated parameters this calibration procedure can be shortened.

Most studies thus applied simulation models and surrogate safety measures on controlled intersections, and not on freeways or weaving sections in particular. Some studies show a reasonable fit between simulated conflicts and observed crashes or conflicts, others indicate that the quality of the simulation model is important and indicate that more research is required.

Research methodology

From the literature it becomes clear that using micro-simulations in combination with calculating the number of conflicts might be a good alternative for using crash rates, as it takes a long time to obtain a sufficient reliable and large crash database. SSAM is already applied in some studies as was shown in the literature, however there was no specific focus on the Dutch situations, and on weaving sections in the Netherlands in particular. To investigate whether SSAM in combination with VISSIM microsimulations is sufficient accurate to determine the safety of Dutch weaving sections the following research question is defined:

How representative are surrogate safety measures calculated from VISSIM micro-simulations with SSAM for predicting the safety of Dutch weaving sections?

The research approach is visualized in Figure 2 and further explained in the remainder of this section.

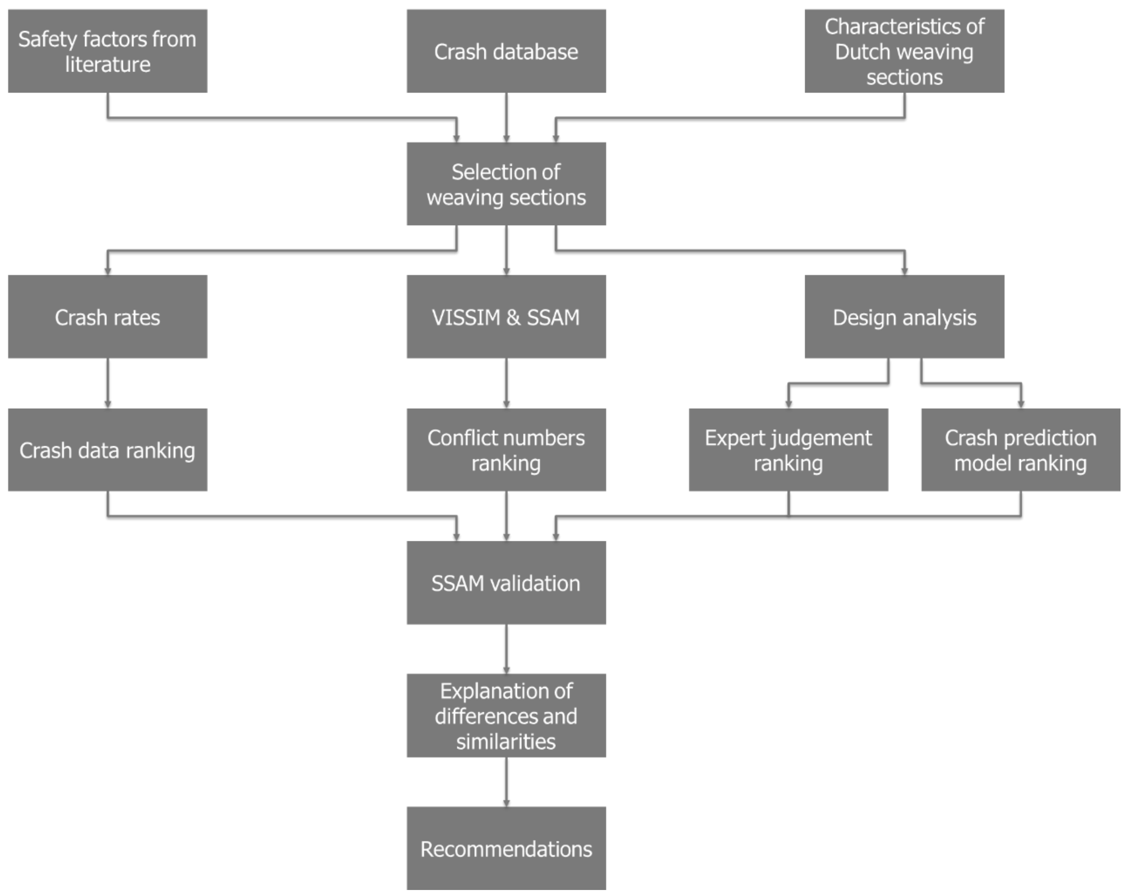

Figure 2: The research methodology

There are some hundreds of weaving sections in the Netherlands. First a weaving section database was created, including characteristics of these weaving sections such as the configuration and crash rate (based on number of crashes that occurred between 2012 and 2015). Thereafter a selection of Dutch weaving sections was made. For selecting the weaving sections the factors that influence the number of crashes were considered. These factors were obtained from the literature research. Iliadi et al. (1) included the following factors in the CPM developed for the Dutch situation: length of the weaving section, AADT, weaving width, share of weaving vehicles and location with respect to the interchange. These factors were considered when selecting the weaving sections. However, also other factors were considered. At first, it was important that the design of the weaving section did not change between 2012 and 2015 as then crashes are assigned to a design that did not exist at that time. Another consideration is that the safety of the weaving sections was determined by the weaving section itself, and not by environmental characteristics that could not be included in the simulation model such as the presence of a bridge or height differences.

Subsequently this selection was ranked according to four different methods:

(1) Based on their crash rate:
The crash rate is calculated as the number of crashes that occurred within the influence area of the weaving section per number of vehicle kilometres. Here the influence area of

the weaving section was defined as the weaving section itself and the 150 meters upstream and downstream of the gores (18). Only the crashes that were registered in the BRON crash database between 2012 and 2015 were included. The number of vehicle kilometres is calculated by multiplying the length of the weaving section as registered in the WEGGEG database by the AADT according to INWEVA 2015;

(2) Based on the number of conflicts determined from the simulations using surrogate safety measures:

To obtain the conflict rate ranking first all weaving sections were put in VISSIM micro-simulation models. Correct length and shape of the links was obtained by building the models on construction maps of the weaving sections. The weaving section itself and some distance before and after were modelled.

A choice was made to simulate one entire workday, as a whole day gives a better indication of the safety of a weaving section than simulating only the peak-hours. Each hour was simulated as 900 seconds (i.e. one quarter of the hour).

Vehicles were assigned a speed using a desired speed distribution. For each weaving section a desired speed distribution for cars and HGVs (heavy good vehicles) was determined from field loop detector data from working days in September 2015. The traffic intensities per incoming link are derived from loop detector data as well, and are changed in the models per hour to simulate variations in flows over the day. Weaving shares (OD matrices) were implemented using static vehicle routes. The required OD matrices were calculated using a selected link analysis on the Dutch regional model (NRM) and distinguish between cars and HGVs and are available for the morning peak, evening peak and off-peak period.

The models were calibrated by comparing hourly simulated intensities m to field intensities c using the GEH-formula (19, 20):

$$GEH = \sqrt{\frac{2(m-c)^2}{m+c}} \quad (1)$$

Intensities, ramp-shares and OD-matrices were changed if the GEH statistic indicated a too large deviation (i.e. $GEH > 10.0$) for too many hours.

The required number of simulation runs was calculated based on a 95% confidence interval using the travel time on the weaving section, with a minimum of 10 simulation runs. For most weaving sections the lower bound of 10 simulation runs was set, here random seed 50 – 59 are used. For weaving section ID369 12 runs were required to obtain a statistical representative result, here seed 50 – 61 were used.

The Surrogate Safety Assessment Model (SSAM) was used to calculate the number of conflicts from the trajectory files. For that the conflict prediction thresholds in SSAM were set to 1.5 seconds for the TTC, 5.0 seconds for the PET, 30° for the rear-end angle and 85° for the crossing angle, as recommended by Gettman et al. (8). Similar values are used as maximum conflict filter values. In the micro-simulation some 'virtual' crashes occur, which are identified in SSAM with a TTC of 0 seconds. According to Gettman and Head (9) and Gettman et al. (8) these crashes should be removed before analysing the results. This is done by setting the lower bound for the TTC to 0.05 seconds using a filter. No filter is applied on the other surrogate safety measures, for the MaxS, DeltaS, DR, MaxD and MaxDeltaV values between -99 and 99 are accepted. Another filter will be applied on the conflict location, such that only the conflicts that occurred within the influence area of the weaving section are included, and not the conflicts that occurred on other locations in the simulation model. This to have a fairer comparison to the crash rate, that is determined based on crashes that occurred within the influence area.

From this the conflict numbers, conflicts rates (converted per number of vehicle kilometres) and ranking positions were obtained.

(3) Based on the judgement of selected road safety experts:

A human factors analysis is an upcoming method for safety experts to be used when predicting safety of a road section (21. 22). Hence some road safety experts were asked to give their opinion on the weaving sections and rank them. Therefore, the experts received a description of each weaving section, including traffic data such as intensities, HGV shares and weaving percentages and maps and photographs. First each expert group member made a ranking of the weaving sections. Thereafter the individual rankings were compared and discussed in the expert group, such that one final ranking that all experts agreed on was obtained.

(4) Based on crash prediction model (CPM) for symmetrical Dutch weaving sections developed by Iliadi et al. (1):

The CPM model was developed to predict the number of crashes to occur on a symmetrical weaving section in the Netherlands in a three-year period N:

$$N = 4.46 \cdot 10^{-5} \cdot lenght^{0.46} \cdot AADT^{0.88} \cdot e^{0.35\,lanes+1.05\,share-1.667\,loc} \quad (2)$$

In here, $length$ is the distance between the convergence and divergence gore, $AADT$ is the annual average daily traffic expressed as vehicles per day, $lanes$ is the number of lanes on the main freeway, $share$ is the percentage of car that is weaving during rush hours and $location$ is the location of the weaving section related to the interchange (0 if inside and 1 if outside).

Note that the formula is for symmetrical weaving sections, and that the weaving section ID454 is asymmetrical. However, due to absence of a Dutch CPM for asymmetrical weaving sections the formula by Iliadi et al. was applied.

This resulted in the expected number of crashes. To have a fairer comparison to the other rankings that are based on ratios, also the expected number of crashes was expressed as ratio of the number of vehicle kilometres.

The strength of the relation between the rankings will be used to conclude whether micro-simulations can be used in future for assessing safety of weaving sections. The relation between safety estimations from micro-simulations and registered accidents will be assessed by comparing the two safety rankings.

Results

This chapter describes the results of the corresponding rankings.

The final selection (Table 1) consists of nine weaving sections that have some common design characteristics, but also have some differences as that is required to assess whether differences result in different crash and conflict rates.

Table 1: Selected weaving sections

\	Location	Between	Road	Configuration	Length	Weaving	HGV
068	IC Heerenveen	Cloverleaf loops	Main	2+1	188.75	25%	14%

077	De Bilt – Maarssen	Junctions	Main	3+1	607.58	31%	9%
156	De Baars – Tilburg Noord	Interch. & junct.	Main	2+1	595.58	9%	14%
173	Kralingen – Terbregseplein	Junct. & interch.	Main	3+2	888.19	43%	8%
256	IC Hoogeveen	Cloverleaf loops	Main	2+1	152.18	12%	26%
269	IC Hattemerbroek	Cloverleaf loops	C/D	1+1	171.06	100%	13%
369	IC Zaandam	Cloverleaf loops	Main	2+1	136.71	28%	7%
412	Voorthuizen – Barneveld	Junctions	Main	2+1	1306.10	27%	14%
454	Rotterdam – Kleinpolderplein	Junct. & interch.	Main	3+1>2+2	468.16	67%	9%

Rankings and Spearman Rank Correlation Coefficient

The rankings based on the four different methods explained in the previous section are presented in table 2. Here ranking position 1 indicates an unsafe weaving section and position 9 indicates a relatively safe weaving section.

Figure 3 visualises the ranking positions of the selected weaving sections based on the four methods. It can be seen that all methods agree on that ID068 is safer than ID454, as an example. However, there are large differences in judgement of for example ID256 and ID269.

Table 2: Scores of the weaving sections on the four methods

ID	Crash			VISSIM/SSAM			CPM			Experts
	Number	Ratio	Rank	Number	Ratio	Rank	Number	Ratio	Rank	Rank
068	5	1.3	6	5	4.2	7	2	0.4	9	5
077	34	1.1	7	112	13.2	4	46	1.5	2	3
156	12	0.7	9	1	0.1	9	15	0.9	4	7
173	128	1.6	5	30823	1409.4	1	107	1.3	3	2
256	19	9.5	1	4	5.8	5	1	0.4	8	8
269	4	3.7	3	2	5.6	6	1	0.7	6	9
369	31	7.3	2	147	116.7	2	2	0.5	7	6
412	52	1.1	8	3	0.3	8	36	0.8	5	4
454	94	2.6	4	678	69.5	3	61	1.7		1

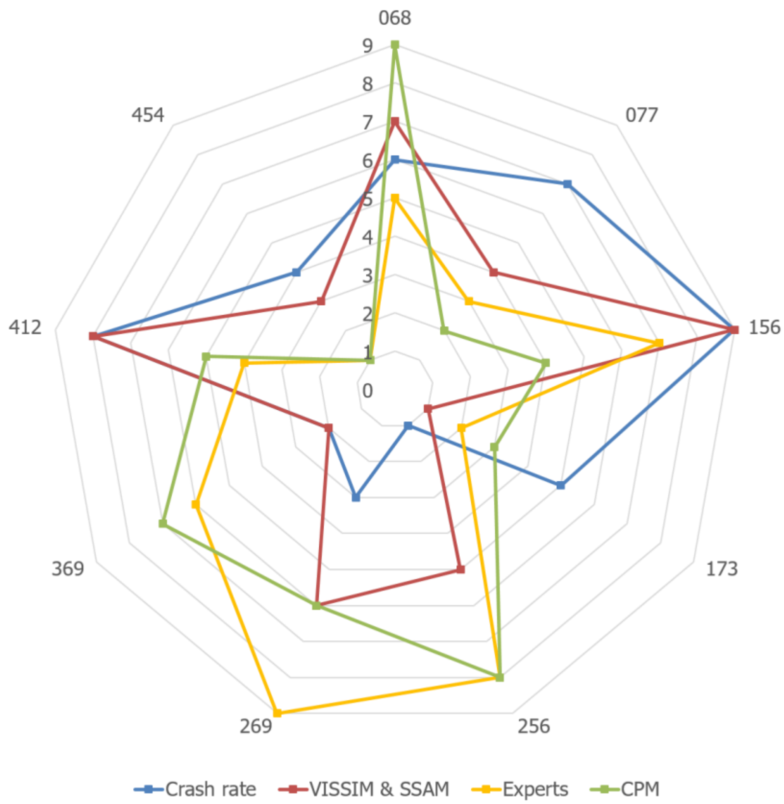

Figure 3: Ranking positions visualised

The correlation between two rankings can be evaluated using the Spearman rank correlation coefficient. This non-parametric test assesses the statistical dependence between two variables, and is often used to assess how well the relationship between two variables can be described using a monotonic function. The coefficient is calculated as

$$\rho_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad (3)$$

where d_i is the difference between ranks for observation (i.e. weaving section) i and n represent the number of observations (weaving sections) in the validation data set. Like Pearson's correlation coefficient, the closer the coefficient is to ± 1, the stronger the monotonic relationship (8, 15).

Table 3Table 3 presents the Spearman Rank Correlation Coefficients and the corresponding P-values.

Table 3: Correlation coefficients (and P-values)

	Crash rate	Experts	CPM
Experts	-0.300 (0.433)		
CPM	-0.367 (0.332)	0.683 (0.042)	
VISSIM & SSAM	0.567 (0.112)	0.467 (0.205)	0.300 (0.433)

The correlation between the VISSIM & SSAM ranking and the crash rate ranking is moderate, similarly as the correlation of the VISSIM & SSAM ranking with the experts ranking. The correlation between VISSIM & SSAM and the CPM is weak. The correlation between the experts ranking and the CPM can be classified as strong. The negative correlation between the crash rate and CPM ranking, suggesting a higher conflict rate corresponding to a lower expected number of crashes in the crash prediction model, and vice versa, is not in line with the expectations. Similarly, there is a negative correlation between the crash rate ranking and the experts ranking.
Except for the correlation between the experts and the CPM ranking, the correlations are not significant at the 5% level.

When considering rankings based on crash and conflict numbers instead of ratios, all rank correlations become stronger. This suggests that traffic intensity influences the number of crashes and number of conflicts. It might be that due to the relation to traffic intensities a correlation between VISSIM and SSAM and the crash counts is found, but that this is a spurious correlation and that the relation between crashes and conflicts is not a causal relationship.

Especially the experts ranking has a very strong correlation with other number rankings, which suggests that experts seem to focus more on the road characteristics than on exposure when judging safety of weaving sections. Although the experts were asked to consider the exposure and rank the weaving sections on the crash risk, it seems that their rankings are based more on crash numbers than on crash rates.

Sensitivity analysis

There were several factors that influenced the obtained ranking and correlations, which are discussed in this paper by means of a sensitivity analysis.

Correlation coefficient The Pearson Correlation Coefficient for ratios can be used as an alternative to the Spearman Rank Correlation Coefficient, as for the first there is no need to first rank the weaving sections and hence slight differences in ratio do not result in a large difference in ranking position. For most rankings, the differences between the Pearson correlation coefficient and Spearman Rank correlation coefficient were only minor. However, the correlation between VISSIM & SSAM and the crash rate became weaker ($\rho_s = -0.158$), which is explained by the very high conflict rate at ID173. The Pearson correlation between the CPM and the experts is significant.

Incomplete crash database As it is generally known that the used BRON crash database is incomplete, the UDLS database was used as an alternative. However, no major differences were observed for correlations with BRON crash rates and UDLS crash rates.

Sensitivity of the PET threshold The initial PET projection threshold value in SSAM is 5.0 seconds, and this value was also used as maximum for filtering conflicts. According to Gettman et al. (8) this value is selected based on a literature review. However, in literature also some lower PET threshold values were proposed. Archer (23) describes that the PET should be below a predetermined threshold value, which is typically 1 to 1.5 seconds. In his research the threshold value was set to 1.5 seconds. Kraay et al. (24) concluded that in general on roads within urban area only PET values below 1.0 seconds are perceived as possibly critical. Both have their focus in interchanges and not on weaving sections. Hence the number of conflicts is calculated for these

PET threshold values, and for one extra value in between: 3 seconds.
It is seen that for lower PET filter values the correlation between the crash rate and VISSIM & SSAM rate ranking is less strong.

Sensitivity of the TTC threshold SSAM uses an initial value of 1.5 seconds for the TTC threshold, which is also proposed by amongst other Kraay et al. (24), recommended by Gettman et al. (8) and used by Shahdah et al. (25). However, in literature also other values are proposed. Archer (23) uses a threshold TTC of 3.5 seconds, and Kuang et al. (26) mentions that the TTC varies between 1.5 and 4.0 seconds. Assessing the effect of larger TTC filter values is not possible, as then the value exceeds the projection threshold value and hence the conflict number does not change. Taking a different projection threshold results in different projected conflicts and hence would result in an unfair comparison. Hence sensitivity of the TTC filter threshold is assessed by taking maximum TTC values of 0.5, 1.0 and 1.5 seconds. A stronger correlation between crash rate ranking and conflict rate ranking is observed for lower filter values. This is in line with the expectations as the lower the TTC value, the larger the likelihood that a conflict results in a crash.

Effect of the Wiedemann Car-following model Initially the Wiedemann99 car following model was used in all simulations, as this was recommended by Fan et al. (27, p. 71). However, from the VISSIM 9.0 manual (28, p. 247) it can be concluded that the Wiedemann74 model is better suitable for modelling merging areas. When only changing the car-following model and leaving all other VISSIM settings equal, it was found that for the Wiedemann99-model a correlation of 0.567 was found between the conflict rate ranking and crash rate ranking, and that this decreased to 0.300 for the Wiedemann74-model. More detailed traffic data on vehicle trajectories is required to determine which model is best representing the real traffic behaviour.

Effect of desired speed limit It was found that for the initial VISSIM simulation models congested speeds were used to determine the desired speed distributions. The desired speed distributions were updated by using only the fastest half of the speeds measured by loop detectors, which resulted in a weaker correlation of 0.433 between the crash rates and conflict rates.

Effect of calibration on speeds All previously presented results were based on a model that was calibrated by a visual inspection and calibrating on vehicle intensities. However, including also a calibration on vehicle speeds results in a model that is better representing the real traffic behaviour. For that calibration, hourly simulated vehicle speeds from three simulation runs with seed 43 - 45 are compared with average hourly field speeds measured by loop detectors for working days in September 2015. This resulted in changes in changes in the desired speed distributions and vehicle inputs, to obtain a better correspondence between field and simulated traffic flow. After this extra calibration procedure a correlation of -0.083 was found between simulated conflict rate ranking and crash rate ranking, which is lower than the initial correlation. This is contradicting to the hypothesis that a more extensive calibration leads to a stronger correlation.
Remarkable is that the conflict rate largely increased on weaving sections 077 and 412, where the amount of congestion is increased by adding vehicles, and that the conflict rate decreased on weaving section 173, where the evening peak congestion is decreased. This suggests that there is a relation between the amount of congestion and the number of conflicts calculated by SSAM.

Time, location and type It is found that the times of the conflicts are reasonable corresponding to the times of the crashes for weaving sections with a sufficient number of crashes. For weaving sections with only a few crashes no clear correspondence is found, due to the stochasticity of the occurrence of crashes. Correlation between conflict location and crash location was weak, suggesting that SSAM is not good at predicting the location. As in the BRON crash database the crash type is often registered as ‘unknown’ it was not possible to draw conclusions on the predictability of the type.

Conclusions and discussion

Nine Dutch weaving sections were selected and ranked based on four criteria. A Spearman Rank Correlation Coefficient was calculated between each of the rankings. A moderate correlation of 0.567 was observed between the conflict rate ranking and the crash rate ranking. However, this correlation is not significant at the 5% significance interval. A stronger correlation of 0.683 was observed between the CPM and experts ranking. Correlations between other rankings are weaker or even negative.

There are multiple possible explanations for the differences between the rankings. It is generally known that the BRON crash database is incomplete. However, consulting an alternative crash database (UDLS) did not result in very different crash rates and correlations. The conflict rate ranking is affected by both the VISSIM and SSAM settings. There are many input variables within the VISSIM model, such as the desired vehicle speed, the car following model, the lane change distance and many parameter settings that influence the vehicle trajectories and hence the number of conflicts. Due to unavailability of precise data sometimes as good as possible available alternatives are used. Furthermore, due to additional calibrating on speeds more congestion is simulated at some locations, resulting in simulated traffic better representing the field, but also resulting in many more conflicts at some congested locations and a weaker correlation. Hence it is desired to get deeper insight in VISSIM parameter settings. Also the TTC and PET prediction and filtering threshold values in SSAM affect the ranking. It should be noted that VISSIM and SSAM only determine vehicle to vehicle conflicts, while also single-vehicle crashes are included in the crash rates and CPM. This might lead to an unfair comparison. Furthermore, the road safety experts ranked the weaving sections in only one hour, while much more time is spent on spent on assessing the safety of the proposed design options as part of the design process. Lastly only a small sample of only 9 weaving sections was used, so one exceptional weaving section largely influences the correlation.
Furthermore, in this research weaving sections with clear variations in crash rates were selected. However another strategy could be to select weaving sections with similar crash rates and investigate whether this results in similar conflict rates. Furthermore, only one ranking of each type was created, but it can be desireable to perform cross-validation such that multiple rankings of a type are compared to multiple rankings of another type.
A possible application of VISSIM and SSAM is to compare multiple design alternatives when (re)constructing a weaving section. Hence for one site multiple designs are compared. In this research weaving sections at different types were compared, which thus differs from the proposed application.

Hence, one should be very careful with using the number of conflicts calculated using VISSIM and SSAM as (only) predictor for safety of Dutch weaving sections.

In other countries crash prediction models were developed and used frequently. Such a CPM for symmetrical weaving sections in the Netherlands resulted in a weaker correlation to crash rates than VISSIM and SSAM, but the correlation between the CPM and the experts is stronger ($\rho_s = 0.683, P = 0.042$). Hence the CPM might be more appropriate for judging safety of a proposed design for a weaving section, although it gives no details on the location and severity of the conflicts and potential crashes.

References

1) Iliadi, H. Farah, P. Schepers, and S.P. Hoogendoorn. A crash prediction model for weaving sections in the Netherlands. In 95th Annual Meeting Transportation Research Board, Washington, USA, 10-14 January 2016; Authors version. Transportation Research Board, 2016.
2) T.F. Golob, W.W Recker, and V.M. Alvarez. Safety aspects of freeway weaving sections. Transportation Research Part A: Policy and Practice, 38(1):35–51, 2004.
3) Y. Qi, J. Liu, and Y. Wang. Safety performance for freeway weaving segments. Technical report, Texas Southern University, 2014.
4) Grontmij. Capaciteitswaarden Infrastructuur Autosnelwegen; Handboek, versie 4. Technical report, Rijkswaterstaat Water, Verkeer en Leefomgeving, De Bilt, 2015.
5) Rijkswaterstaat GPO and Witteveen + Bos. Richtlijn Ontwerp Autosnelwegen 2014. Technical report, Rijkswaterstaat, 2015.
6) B.J. Park, K. Fitzpatrick, and D. Lord. Evaluating the effects of freeway design elements on safety. Transportation Research Record: Journal of the Transportation Research Board, 2195:58–69, 2010.
7) R. Elvik, T. Vaa, A. Erke, and M. Sorensen. The handbook of road safety measures. Emerald Group Publishing, 2009.
8) D. Gettman, L. Pu, T. Sayed, and S.G. Shelby. Surrogate Safety Assessment Model and validation: Final report (FHWA-HRT-08-051) . Technical report, Federal Highway Administration, 2008.
9) D. Gettman and L. Head. Surrogate safety measures from traffic simulation models. Transportation Research Record: Journal of the Transportation Research Board, 1840:104–115, 2003.
10) A. Laureshyn, Å. Svensson, and C. Hydén. Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. Accident Analysis & Prevention, 42(6):1637–1646, 2010.
11) J. Bared. Techbrief: Surrogate Safety Assessment Model (SSAM). Technical summary (FHWA-HRT-08-049), 2016.
12) W. Young, A. Sobhani, M.G. Lenné, and M. Sarvi. Simulation of safety: A review of the state of the art in road safety simulation modelling. Accident Analysis & Prevention, 66: 89–103, 2014.
13) C. Brouwer. Het klaverblad en de veiligheid van het verkeer. Verkeerskunde, 26(3), 1975.

14) H. Yang, K. Ozbay, and B. Bartin. Application of simulation-based traffic conflict analysis for highway safety evaluation. Proceedings of the 12th WCTR, Lisbon, Portugal, 2010.
15) F. Huang, P. Liu, H. Yu, and W. Wang. Identifying if VISSIM simulation model and SSAM provide reasonable estimates for field measured traffic conflicts at signalized intersections. Accident Analysis & Prevention, 50:1014–1024, 2013.
16) K. El-Basyouny. Field validation for surrogate safety assessment methodology (SSAM) using a multi-purpose micro-simulation. Master thesis, University of British Columbia, 2006.
17) M. Essa and T. Sayed. Transferability of calibrated microsimulation model parameters for safety assessment using simulated conflicts. Accident Analysis & Prevention, 84:41–53, 2015.
18) R.G.M.M. Vermijs. Capaciteitswaarden symmetrische weefvakken. TU Delft, 1997.
19) C. Mai, C. McDaniel-Wilson, D. Norval, D. Upton, J. Auth, et al. Protocol for VISSIM Simulation. Oregon Department of Transportation, USA, 2011.
20) J. Smith, R. Blewitt, et al. Traffic modelling guidelines. Traffic manager and network performance best practice. Version, 3, 2010.
21) Rijkswaterstaat, Water Verkeer en Leefomgeving. Human Factors voor verkeersveiligheid in het wegontwerp (Handreiking voor Human Factors; Bijlage E bij VOA). Technical report, Rijkswaterstaat, 2016.
22) J. L. Campbell, M.G. Lichty, J.L. Brown, C. M. Richard, J.S. Graving, J. L. Graham, M. O'Laughlin, D. Torbic, and D. Harwood. NCHRP report 600: Human factors guidelines for road systems. Transportation Research Board, Washington, D.C., second edition, 2012.
23) J. Archer. Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections. PhD thesis, Royal Institute of Technology, Kungliga Tekniska Hogskolan (KTH), Stockholm, Sweden, 2005.
24) J.H. Kraay, A.R.A. van der Horst, and S. Oppe. Handleiding voor de conflictobservatietechniek DOCTOR (in Dutch) Manual for conflict observation technique DOCTOR. (Report R-86-3). Technical report, SWOV, 1986.
25) U. Shahdah, F. Saccomanno, and B. Persaud. Integrated traffic conflict model for estimating crash modification factors. Accident Analysis & Prevention, 71:228–235, 2014.
26) Y. Kuang, X. Qu, and S. Wang. A tree-structured crash surrogate measure for freeways. Accident Analysis & Prevention, 77:137–148, 2015.
27) R. Fan, H. Yu, P. Liu, and W. Wang. Using VISSIM simulation model and Surrogate Safety Assessment Model for estimating field measured traffic conflicts at freeway merge areas. IET Intelligent Transport Systems, 7(1):68–77, 2013.
28) PTV VISSIM. PTV VISSIM 9, User Manual. PTV AG, Karlsruhe, 2016.