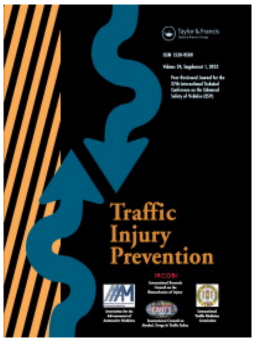

Traffic Injury Prevention

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gcpi20

How certain are we that our automated driving system is safe?

Erwin de Gelder & Olaf Op den Camp

To cite this article: Erwin de Gelder & Olaf Op den Camp (2023) How certain are we that our automated driving system is safe?, Traffic Injury Prevention, 24:sup1, S131-S140, DOI: 10.1080/15389588.2023.2186733

To link to this article: https://doi.org/10.1080/15389588.2023.2186733

Published online: 02 Jun 2023.

TRAFFIC INJURY PREVENTION
2023, VOL. 24, NO. S1, S131–S140
https://doi.org/10.1080/15389588.2023.2186733

How certain are we that our automated driving system is safe?

Erwin de Gelder and Olaf Op den Camp

Integrated Vehicle Safety, TNO, Helmond, The Netherlands

ABSTRACT

Objective: Regulations are currently being drafted by the European Commission for the safe introduction of automated driving systems (ADSs) with conditional or higher automation (SAE level 3 and above). One of the main challenges for complying with the drafted regulations is proving that the residual risk of an ADS is lower than the existing state of the art without the ADS and that the current safety state of European roads is not compromised. Therefore, much research has been conducted to estimate the safety risk of ADS. One proposed method for estimating the risk is data-driven, scenario-based assessment, where tests are partially automatically generated based on recorded traffic data. Although this is a promising method, uncertainties in the estimated risk arise from, among others, the limited number of tests that are conducted and the limited data that have been used to generate the tests. This work addresses the following question: "Given the limitations of the data and the number of tests, what is the uncertainty of the estimated safety risk of the ADS?"
Methods: To compute the safety risk, parameterized test scenarios are based on large-scale collections of road scenarios that are stored in a scenario database. The exposure of the scenarios and the parameter distributions are estimated using the data as well as confidence bounds of these estimates. Next, virtual simulations are conducted of the scenarios for a variety of parameter values. Using a probabilistic framework, all results are combined to estimate the residual risk as well as the uncertainty of this estimation.
Results: The results are used to provide confidence bounds on the calculated fatality rate in case an ADS is implemented in the vehicle. For example, using the proposed probabilistic framework, it is possible to claim with 95% certainty that the fatality rate is less than 10^{-7} fatalities per hour of driving. The proposed method is illustrated with a case study in which the risk and its uncertainty are quantified for a longitudinal controller in 3 different types of scenarios. The case study code is publicly available.
Conclusions: If results show that the uncertainty is too high, the proposed method allows answering questions like "How much more data do we need?" or "How many more (virtual) simulations must be conducted?" Therefore, the method can be used to set requirements on the amount of data and the number of (virtual) simulations. For a reliable risk estimate, though, much more data are needed than those used in the case study. Furthermore, because the method relies on (virtual) simulations, the reliability of the result depends on the validity of the models used in the simulations. The presented case study illustrates that the proposed method is able to quantify the uncertainty of the estimated safety risk of an ADS. Future work involves incorporating the proposed method into the type approval framework for future ADSs of SAE levels 3, 4, and 5, as proposed in the upcoming European Union implementing regulation for ADS.

ARTICLE HISTORY
Received 19 August 2022
Accepted 23 February 2023

KEYWORDS
Safety; assessment; automated driving system; data completeness; uncertainty

Introduction

Automated driving systems (ADSs) have the potential of making traffic safer by eliminating human errors, enabling more comfortable rides, and reducing traffic congestion (Chan 2017). Lower levels of automation systems, such as adaptive cruise control (ACC) and lane keeping assist systems, are already widely deployed in modern cars and trucks. Such systems are considered as driver assistance (level 1) or partial driving automation (level 2) systems (SAE J3016 2021). Here, the driver is still responsible for intervening if the ADS fails, so the human driver is accountable in case of any damage caused by the vehicle. For conditional driving automation (level 3), high driving automation (level 4), and full driving automation (level 5), however, the driver is neither required nor expected to intervene immediately when the ADS fails. Hence, for ADSs of SAE level 3 and above, the accountability shifts away from the driver (Luetge 2017). Therefore, for such systems, it is even more important to ensure that the risk of ADS is lower than existing state-of-the-art vehicles with no ADS or an ADS of SAE level 2 and below. In fact, new regulations are currently being drafted by the European Commission for the safe introduction of ADS of SAE level 4 (European

CONTACT Erwin de Gelder erwin.degelder@tno.nl Integrated Vehicle Safety, TNO, Automotive Campus 30, 5708JZ, Helmond, The Netherlands.
Associate Editor Cody Stolle oversaw the review of this article.

Commission 2022) that require that the ADS be free from unreasonable risks. Hence, much research has been conducted to estimating the risk of an ADS.

A scenario-based approach for prospective risk assessment of an ADS is broadly supported by the automotive field (Roesener et al. 2017; Pütz et al. 2017; Elrofai et al. 2018; Ploeg et al. 2018; Menzel et al. 2018; Thorn et al. 2018; Antona-Makoshi et al. 2019; Riedmaier et al. 2020; de Gelder et al. 2021; Scholtes et al. 2021). For the scenario-based approach, it is important that the scenarios provide a representation of the real world and that the scenarios used for the risk assessment cover the same variety that is found in real life (Riedmaier et al. 2020). A popular approach is to use real-world data collected from vehicles operating in traffic to extract scenarios that occur in real life (Pütz et al. 2017; Elrofai et al. 2018; Ploeg et al. 2018; Antona-Makoshi et al. 2019). Therefore, in this work, we follow the data-driven, scenario-based risk quantification of an ADS as proposed by de Gelder et al. (2021). In de Gelder et al. 2021, scenarios were extracted from a data set. These scenarios were used to generate new scenarios for which the response of the ADS under test was tested in virtual simulations. Using the data, the exposure of the scenarios in real-world traffic was estimated. By combining the estimated exposure with the calculated crash probability and injury rate, how frequent crashes and severe injuries are when the ADS is activated was estimated. The proposed method for quantifying the risk can be applied to assess risks based on ISO 26262 (2018) and ISO 21448 (2022), the leading standards in automotive safety.

When quantifying the risk using a data-driven, scenario-based approach, the calculated risk will be an estimation of the actual risk. Uncertainties in the estimated risk arise from the limited data and the limited number of simulation runs. The limited data lead to uncertainty in the estimated exposure of the scenarios. Furthermore, the generated scenarios may not fully represent the variations in real traffic due to the limited data.

This resulted in an uncertainty of the estimated crash probability. Also, the limited number of simulation runs contributed to the uncertainty of the crash probability. In the literature, several metrics have been proposed to quantify the degree of completeness of the data (Wang et al. 2017; de Gelder et al. 2019; Hauer et al. 2019). However, to the best of our knowledge, there is no method that quantifies how the limited data affect the uncertainty in the estimated risk of an ADS. This work presents a method to estimate the uncertainty of the quantified risk as a result of the limited data and the limited number of simulation runs. To do this, we first estimate the uncertainty of the estimated exposure based on how often a type of scenario is seen in an hour of driving and how this varies from hour to hour. Second, the uncertainty of the crash probability as a result of the limited data is estimated using bootstrapping (Efron 1979) of the original data. Third, the uncertainty of the crash probability as a result of the limited number of simulations is estimated using a standard formula used for importance sampling with Monte Carlo simulations (Owen 2013). Finally, the 3 uncertainties are combined to calculate the uncertainty of the estimated risk.

To illustrate the proposed method, a case study is performed in which the risk and its uncertainty are estimated while considering 3 different types of scenarios. In this case study, the uncertainty of the exposure, the crash probability, and the risk are estimated while considering a varying amount of data and a varying number of simulation runs. The case study demonstrates that more data and a larger number of simulation runs lead to a lower uncertainty of the estimated risk. The case study code is publicly available (https://github.com/ErwindeGelder/ScenarioRiskQuantification).

This article is organized as follows. The next section describes the problem in more detail. Then, a method is proposed as a solution to this problem. Next is a case study in which the proposed method is illustrated. A discussion follows, including our conclusions.

Problem definition

An ADS is designed for a specific operational design domain (ODD), where the ODD refers to the operating conditions under which the ADS is specifically designed to function (SAE J3016 2021). Therefore, the ODD is used to confine the risk analysis (Gyllenhammar et al. 2020). Still, it is expected that there will be a very large variety of scenarios within a specific ODD and that an ADS must act appropriately in the majority of those scenarios in order to be considered safe enough. In addition, an ADS must safely handle scenarios in which the ADS threatens to unintentionally exit its ODD. Here, we define a scenario as a "quantitative description of the relevant characteristics and activities and/or goals of the ego vehicle(s), the static environment, the dynamic environment, and all events that are relevant to the ego vehicle(s) within the time interval between the first and the last relevant event" (de Gelder et al. 2022, p.303). To deal with this very large variety of scenarios, we group the scenarios into so-called scenario categories. A scenario category refers to a qualitative description of a scenario (de Gelder et al. 2022).

By distinguishing between scenarios and scenario categories, the question of how certain we are that our ADS is safe can be split into two. First, are all relevant scenario categories considered during the risk assessment? Second, is the variability of the scenarios within a scenario category sufficiently considered during the risk assessment? This work focuses on the second question. We will come back to the first question in the discussion.

In de Gelder et al. 2021, a method was provided to estimate the probability that an ADS cannot avoid a crash in a scenario from a specific scenario category. To estimate this probability, (virtual) simulations of the scenarios are used. The scenarios themselves are generated based on observed scenarios in real-world data. Because neither infinite data are available nor an infinite number of simulations is performed, the estimated probability of a crash approximates the real crash probability. Therefore, this work aims to answer the following question: How does the limited data and the limited number of simulations influence the uncertainty of the estimated crash probability of an ADS in scenarios of scenario category C?

Method

For the risk quantification, this work adopts the approach presented by de Gelder et al. (2021). This section first explains how the risk is estimated by combining the exposure and the crash probability. Next, we propose methods to estimate the uncertainty of the exposure and the uncertainty of the crash probability, where the latter uncertainty is caused by the limited data and the limited number of simulation runs. Finally, this section shows how the uncertainties are combined to estimate the uncertainty of the estimated risk.

Risk quantification

This section summarizes the method for quantifying the risk described by de Gelder et al. (2021). For more details, we refer the reader to (de Gelder et al. 2021). Let us assume that n hours of driving data have been collected. Let m_i denote the number of scenarios of C that have been observed in the i-th hour with $i \in \{1, ..., n\}$. The exposure of scenario category C is expressed as the expectation of m, which is estimated as follows:

$$\mathbb{E}[m] = \frac{1}{n}\sum_{i=1}^{n} m_i = \frac{N}{n} \tag{1}$$

with $N = \sum_{i=1}^{n} m_i$ denoting the total number of scenarios that were encountered during the n hours of driving data.

The N scenarios of scenario category C are parameterized such that the j-th scenario is described by the vector $x_j \in \mathbb{R}^d$, where d denotes the number of parameters that are used to describe the scenarios from scenario category C. The probability density function (PDF) of the scenario parameters is estimated using kernel density estimation (KDE; Rosenblatt 1956; Parzen 1962):

$$\hat{f}(x) = \frac{1}{Nh^d}\sum_{j=1}^{N} K\left(\frac{1}{h}(x - x_j)\right), \tag{2}$$

with $K(\cdot)$ and h denoting the kernel function and bandwidth, respectively. Because the choice of the kernel function is not as important as the choice of the bandwidth (Chen 2017; Turlach 1993), the often-used Gaussian kernel is adopted. Note that our method also applies with alternative kernels. The Gaussian kernel is given by

$$K(u) = \frac{1}{(2\pi)^{\frac{d}{2}}}\exp\left\{-\frac{1}{2}\|u\|_2^2\right\}, \tag{3}$$

where $\|u\|_2^2 = u^{\mathrm{T}}u$ denotes the squared 2-norm of u. The bandwidth $h > 0$ controls the smoothness of the estimated PDF. Choosing the appropriate value is a trade-off because larger values result in a smoother PDF but choosing h too large may result in loss of details in the PDF. As shown in Turlach (1993), using leave-one-out cross-validation to determine the bandwidth minimizes the discrepancy between the estimated PDF and the real, unknown PDF according to the Kullback-Leibler divergence. Thus, this work uses leave-one-out cross-validation to determine h.

Let $R(x)$ denote the outcome of a simulation of a scenario that is parameterized by x. If $R(x) = 1$ denotes a simulation run that ends in a crash and $R(x) = 0$ otherwise, then $\mathbb{E}[R(x)]$ is the expected probability of a crash. A straightforward way to estimate $\mathbb{E}[R(x)]$ is using a crude Monte Carlo simulation:

$$\mathbb{E}[R(x)] \approx \frac{1}{N_{\mathrm{MC}}}\sum_{k=1}^{N_{\mathrm{MC}}} R(x_k),\ \ x_k \sim \hat{f}, \tag{4}$$

where N_{MC} denotes the number of simulation runs. Following Zhang (1996), nonparametric importance sampling is used to accelerate the evaluation. With importance sampling, instead of sampling the scenario parameter values from $\hat{f}$, the scenario parameter values are sampled from the so-called importance density, g. With nonparametric importance sampling, the importance density is calculated in a similar manner as $\hat{f}$ in Eq. (2), but instead of using the scenario parameter values from the data, the parameter values of the most $N_{\mathrm{C}} < N_{\mathrm{MC}}$ critical scenarios that are simulated during the crude Monte Carlo simulation are used. Many metrics exist for determining the criticality of a simulated scenario (Mullakkal-Babu et al. 2017; C. Wang et al. 2021; Westhofen et al. 2022; de Gelder et al. 2023), and choosing appropriate metrics depends on the scenarios that are simulated. In this work, the criticality of a simulated scenario is measured using the minimum time to collision (Hayward 1972), where a lower value indicates a more critical scenario. Note that the time to collision is an appropriate metric for the scenarios considered in our case study (Mullakkal-Babu et al. 2017), but this might not be an appropriate metric to measure the criticality in simulations of other type of scenarios. To acquire an unbiased estimate of $\mathbb{E}[R(x)]$, the simulation results are weighted to correct for the fact that we sample from g instead of $\hat{f}$:

$$\mathbb{E}[R(x)] \approx \mu_{\mathrm{NIS}} = \frac{1}{N_{\mathrm{NIS}}}\sum_{k=1}^{N_{\mathrm{NIS}}} R(x_k)\frac{\hat{f}(x_k)}{g(x_k)},\ \ x_k \sim g, \tag{5}$$

where N_{NIS} denotes the number of simulation runs with importance sampling.

To compute the risk of a crash given a scenario category C, the exposure and the crash probability are multiplied:

$$Risk(C) = \mathbb{E}[m] \cdot \mathbb{E}[R(x)]. \tag{6}$$

Uncertainty of the exposure

Assuming that m_i is uncorrelated with m_j for $i \neq j$, the unbiased estimator of the variance of m is

$$\mathbb{V}[m] \approx \frac{1}{n-1}\sum_{i=1}^{n}\left(m_i - \frac{N}{n}\right)^2. \tag{7}$$

Here, $\mathbb{V}[\cdot]$ denotes the variance. Thus, the variance of the estimated exposure of Eq. (1) is

$$\mathbb{V}\left[\frac{N}{n}\right] \approx \hat{\sigma}^2_{\mathrm{exposure}} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(m_i - \frac{N}{n}\right)^2. \tag{8}$$

We will express the uncertainty of the exposure as the estimated standard deviation of the expected exposure; that is, $\hat{\sigma}_{\text{exposure}}$.

Remark 1. It is not uncommon to assume that m is distributed according to the Poisson distribution. In that case, the variance of m equals the exposure of Eq. (1), so Eq. (8) simplifies to

$$\mathbb{V}\left[\frac{N}{n}\right] = \frac{N}{n^2}. \tag{9}$$

Uncertainty because of limited data

The underlying distribution of the scenario parameters is estimated using $\hat{f}(\cdot)$ of Eq. (2). Assuming that the real underlying distribution is smooth, the estimated PDF converges to the real PDF as $h \to 0$ and $Nh^d \to \infty$ (Wasserman 2006). In reality, N will be finite, so $\hat{f}(\cdot)$ will deviate from the real underlying PDF. According to Chen (2017), a method to estimate the asymptotic variance of $\hat{f}(\cdot)$ is using bootstrapping (Efron 1979). In a similar manner, this work uses bootstrapping to estimate the uncertainty of μ_{NIS} as a result of the variance of $\hat{f}(\cdot)$.

The bootstrapping works as follows. We select N scenario parameter vectors from the set $\{x_j\}_{j=1}^{N}$ with replacement. Note that some scenario parameter vectors will be selected more than once, whereas some other scenario parameter vectors will not be selected at all. Using the selected scenario parameter vectors, a new estimate of the PDF is constructed using KDE, similar to $\hat{f}(\cdot)$ in Eq. (2). Let us denote this estimated PDF by $\hat{f}_l^*(\cdot)$, where l denotes the index of the bootstrap and the asterisk highlights the fact that this PDF is estimated by resampling $\{x_j\}_{j=1}^{N}$. This procedure is repeated B times, so $l = 1, \ldots, B$. Next, based on the PDF $\hat{f}_l^*(\cdot)$, the crash probability μ_{NIS} is evaluated. It would be time-consuming to redo the Monte Carlo simulation for each $\hat{f}_l^*(\cdot)$. Instead, μ_{NIS} is directly evaluated by substituting $\hat{f}_l^*(\cdot)$ for $\hat{f}(\cdot)$ in Eq. (5):

$$\mu_{\text{NIS},l}^* = \frac{1}{N_{\text{NIS}}} \sum_{k=1}^{N_{\text{NIS}}} R(x_k) \frac{\hat{f}_l^*(x_k)}{g(x_k)}, \tag{10}$$

where the same set of scenario parameter values $\{x_k\}_{k=1}^{N_{\text{NIS}}}$ is used as for evaluating Eq. (5).

Evaluating Eq. (10) B times leads to B estimates of $\mathbb{E}[R(x)]$. The variance of μ_{NIS} resulting from the variance in $\hat{f}(\cdot)$ is approximated using

$$\mathbb{V}[\mu_{\text{NIS}}] \approx \sigma_{\mu,\text{data}}^2 = \frac{1}{B-1} \sum_{l=1}^{B} \left(\mu_{\text{NIS},l}^* - \sum_{l'=1}^{B} \mu_{\text{NIS},l'}^* \right)^2. \tag{11}$$

The estimated standard deviation $\sigma_{\mu,\text{data}}$ is used as a measure of the uncertainty of the crash probability as a result of the limited number of scenarios.

Remark 2. Note that the importance density, $g(\cdot)$, is based on the original estimated PDF, $\hat{f}(\cdot)$. This importance density is chosen such that μ_{NIS} converges as quickly as possible to $\mathbb{E}[R(x)]$ while assuming that $\hat{f}(\cdot)$ is a good estimate of $f(\cdot)$. Determining the importance density using $\hat{f}_l^*(\cdot)$ instead of $\hat{f}(\cdot)$ is likely to result in a slightly different importance density. As a result, using $g(\cdot)$ might not be optimal in the sense that another importance density might lead to a faster convergence of μ_{NIS} toward $\mathbb{E}[R(x)]$. However, it is expected that this effect is minor, because $\hat{f}_l^*(x) \approx \hat{f}(x)$ for all x. Hence, the advantage of reusing the Monte Carlo results outweighs the disadvantage of the (potential) slower convergence of the estimation of $\mathbb{E}[R(x)]$ during bootstrapping.

Uncertainty because of limited number of simulations

We express the uncertainty resulting from the limited number of simulations using the standard deviation of μ_{NIS}. In a similar manner that Eq. (8) estimates the variance of Eq. (1), the variance of μ_{NIS} resulting from the limited number of simulations can be approximated using (Owen 2013)

$$\mathbb{V}[\mu_{\text{NIS}}] \approx \hat{\sigma}_{\mu,\text{simulations}}^2 = \frac{1}{N_{\text{NIS}}(N_{\text{NIS}}-1)} \times \sum_{k=1}^{N_{\text{NIS}}} \left(R(x_k) \frac{\hat{f}(x_k)}{g(x_k)} - \mu_{\text{NIS}} \right)^2, \quad x_k \sim g. \tag{12}$$

Combining all uncertainties

To combine the uncertainties of Eqs. (8), (11), and (12), the following is assumed:

- Because both estimations of Eqs. (1) and (5) are unbiased, the estimated exposure of Eq. (1) is uncorrelated with the estimated crash probability of Eq. (5).
- Because the estimator of the crash probability of Eq. (5) is unbiased, the influence of the number of simulations on the estimated crash probability does not depend on the number of hours that are used to collect the data.

Based on the second assumption, the combined effect of the limited data and the limited number of simulations on the variance of μ_{NIS} is simply the sum of the variances of Eqs. (11) and (12):

$$\mathbb{V}[\mu_{\text{NIS}}] \approx \hat{\sigma}_\mu^2 = \hat{\sigma}_{\mu,\text{data}}^2 + \hat{\sigma}_{\mu,\text{simulations}}^2. \tag{13}$$

Based on the first assumption, the uncertainty of the risk, $\text{Risk}(C)$, is computed by treating the estimations of $\mathbb{E}[m]$ and $\mathbb{E}[R(x)]$ as independent. Hence, the variance of $\text{Risk}(C)$ follows from the standard formula used for calculating the variance of two independent variables (Goodman 1960):

$$\mathbb{V}[\text{Risk}(C)] = \left(\mathbb{E}\left[\frac{N}{n}\right]\right)^2 \cdot \mathbb{V}[\mu_{\text{NIS}}] + \left(\mathbb{E}[\mu_{\text{NIS}}]\right)^2 \cdot \mathbb{V}\left[\frac{N}{n}\right] + \mathbb{V}\left[\frac{N}{n}\right] \cdot \mathbb{V}[\mu_{\text{NIS}}] \tag{14}$$

$$\approx \left(\tfrac{N}{n}\right)^2 \cdot \hat{\sigma}_\mu^2 + \mu_{\text{NIS}}^2 \cdot \hat{\sigma}_{\text{exposure}}^2 + \hat{\sigma}_{\text{exposure}}^2 \cdot \hat{\sigma}_\mu^2. \tag{15}$$

The influence of the number of observed scenarios, N, is shown in an illustrative way in Eq. (14):

- If many scenarios were observed, the variance of the estimated PDF will be reduced. As a result, it is expected that $\mathbb{V}[\mu_{\text{NIS}}]$ will be lower compared to the case where only a few scenarios were observed. However, the exposure, N/n, will be relatively high. These two effects counteract in the first term in Eq. (14).
- Because it is expected that $\hat{\sigma}^2_{\text{exposure}}$ is proportional with N (see Eq. (9)), the second term in Eq. (14) scales linearly with μ^2_{NIS} and the exposure, N/n.
- With a similar reasoning as for the first term of Eq. (14), a large value of N typically results in a relatively large value of $\mathbb{V}[\frac{N}{n}]$ and a relatively small value of $\mathbb{V}[\mu_{\text{NIS}}]$. These two effects will have an opposite effect on the third term of Eq. (14).

Results

This section applies the methods of the previous section in a case study. First, the 3 scenario categories that we consider are described. Next, details regarding the data set are provided. Then, a subsection is dedicated to the description of the ADS that we consider in this case study. Then the resulting estimations of the uncertainties of the exposure and the crash probability are studied. Finally, the risk and its uncertainty are presented.

Scenario categories and parameterization

This work considers 3 scenario categories: leading vehicle decelerating (LVD), cut-in, and approaching slower vehicle (ASV). The first 2 scenario categories are mentioned as possibly critical scenarios in the regulation of automated lane keeping systems (World Forum for Harmonization of Vehicle Regulations 2021). The third scenario category accounts for more than 25% of all crashes that involve 2 vehicles in the United States (Najm et al. 2007).

In an LVD scenario, the ego vehicle is following a leading vehicle that decelerates. Three parameters describe an LVD scenario: v_0, Δ_v, and $\overline{a}$. The first parameter, v_0, denotes the initial speed at which both the ego vehicle and the leading vehicle are driving. The leading vehicle decelerates with an average deceleration of $\overline{a}$ such that the final speed is $v_0 - \Delta_v$. Note that $v_0 > 0$, $0 < \Delta_v \leq v_0$, and $\overline{a} > 0$.

In a cut-in scenario, another vehicle is changing lane such that it becomes the leading vehicle of the ego vehicle. Three parameters describe a cut-in scenario: g_0, $v_{\text{e},0}$, and $v_{\text{l},0}$. The parameter g_0 denotes the gap between the ego vehicle and the other vehicle at the moment of the cut-in. The initial speeds of the ego vehicle and the other vehicle are $v_{\text{e},0}$ and $v_{\text{l},0}$, respectively. It is assumed that the other vehicle is driving at a constant speed. Note that $g_0 > 0$, $v_{\text{e},0} > 0$, and $v_{\text{l},0} > 0$.

In an ASV scenario, the ego vehicle is approaching a slower vehicle. The 2 parameters that describe an ASV scenario are the initial speed of the ego vehicle, $v_{\text{e},0}$, and the speed of the vehicle in front of the ego vehicle, $v_{\text{l},0}$. Note that for an ASV scenario, $v_{\text{e},0} > 0$, and $0 < v_{\text{l},0} < v_{\text{e},0}$.

Data set

The data set described in Paardekooper et al. (2019) is used to estimate the exposure and the PDF of the scenario parameters. The data were recorded from a single vehicle in which 20 experienced drivers were asked to drive a prescribed route. Each driver drove the 50 km route 6 times, which resulted in 63 h of data. The route includes urban roads, rural roads, and highways. The surrounding traffic was measured by fusing the radar and camera data as described in Elfring et al. 2016. To extract the scenarios from the data set, the approach described by de Gelder et al. (2020) was used. In 63 h of driving, 1,300 LVD scenarios, 297 cut-in scenarios, and 291 ASV scenarios were found (de Gelder et al. 2021).

When using the KDE of Eq. (2) to estimate the PDF, the parameters are normalized such that the standard deviation is 1 for each of the parameters (Duong 2007). Parameter values that are outside the valid range of values reported in the previous subsection will have a positive probability density due to the infinite support of the Gaussian kernel of Eq. (3). To avoid sampling parameter values that are invalid, the estimated PDFs are set to 0 for all invalid parameter values. For example, the PDF for LVD scenarios is set to 0 for $v_0 \leq 0$, $\Delta_v \leq 0$, $\Delta_v > v_0$, and $\overline{a} \leq 0$. The resulting PDFs are rescaled such that they still integrate to 1.

Automated driving system under test

The ACC described by Xiao et al. (2017), which is based on the ACC proposed by Milanés and Shladover (2014), is used in this case study. The ACC maintains a safe distance from a leading vehicle while not exceeding a speed that is set by the user of the ACC. With the ACC, the acceleration of the ego vehicle is based on the speed of the ego vehicle, $v_{\text{e}}(t)$; the speed of the leading vehicle, $v_{\text{l}}(t)$; and the gap between the leading vehicle's back and the ego vehicle's front, $g(t)$, at time t. The acceleration of the ego vehicle at time t is described by the following equations (Xiao et al. 2017):

$$a_{\text{e}}(t) = \max\left(\min\left(a_{\text{ACC}}(t), a_{\text{CC}}(t)\right), -d_{\max}\right), \tag{16}$$

$$a_{\text{ACC}}(t) = \begin{cases} k_1\big(g(t) - d_0(v_{\text{e}}(t)) - \tau_{\text{h}} v_{\text{e}}(t)\big) \\ \quad + k_2\big(v_{\text{l}}(t) - v_{\text{e}}(t)\big) & \text{if } g(t) < d_{\text{ACC}}, \\ a_{\text{CC}}(t) & \text{otherwise,} \end{cases} \tag{17}$$

$$d_0(u) = \begin{cases} 5 \text{ m} & \text{if } u \geq 15 \text{ m/s}, \\ 7 \text{ m} & \text{if } u < 10.8 \text{ m/s}, \\ \dfrac{75 \text{ m}^2/\text{s}}{u} & \text{otherwise,} \end{cases} \tag{18}$$

$$a_{\text{CC}}(t) = k_{\text{CC}}(v_{\text{set}} - v_{\text{e}}(t)). \tag{19}$$

The values and descriptions of the parameters $d_{\max}$, d_{ACC}, k_1, k_2, τ_{h}, and k_{CC} are provided in Table 1. The parameter v_{set} is the desired speed, which is assumed to be the same as the initial speed of the ego vehicle in each simulation run; that is, $v_{\text{e}}(0) = v_{\text{set}}$.

Table 1. Parameters of the system under test. The values of the parameters k_1, k_2, τ_h, and k_{CC} are adopted from Xiao et al. (2017).

Parameter	Description	Value
d_{max}	Maximum deceleration	6 m/s^2
d_{ACC}	Maximum sensor range of ACC	150 m
k_1	Distance gain of ACC	0.23 s^{-2}
k_2	Speed gain of ACC	0.07 s^{-1}
τ_h	Time gap setting, also known as desired time headway	1.1 s
k_{CC}	Speed gain of cruise control	0.4 s^{-1}
v_{set}	Desired speed	Variable

Uncertainty of exposure

Table 2 summarizes the results. In $n = 63$ h of driving data, $N = 1,300$ LVD scenarios, $N = 297$ cut-in scenarios, and $N = 291$ ASV scenarios were extracted. Thus, following Eq. (1), the exposure for the corresponding scenario categories is 20.6, 4.71, and 4.62 h^{-1}, respectively. To determine the uncertainty of the estimated exposure, we assume that m_i is uncorrelated with m_j for $i \neq j$, such that we can use Eq. (8). We have performed the Ljung-Box test (Ljung and Box 1978) to check for autocorrelations up to a lag of 3, following the recommendation of Burns (2002). For the LVD and cut-in scenarios, there is no significant evidence that m_i is not independently distributed ($P > .05$). For the ASV scenarios, however, there might be an autocorrelation present, because the probability that there is no autocorrelation of lag 1 is $P = .023$ according to the Ljung-Box test. As a result, the estimated uncertainty of the exposure of the ASV scenarios might be inaccurate. Applying Eq. (8) results in the estimated uncertainty of the exposure: $\hat{\sigma}_{\text{exposure}} \approx 1.2\ h^{-1}$ for LVD scenarios, $\hat{\sigma}_{\text{exposure}} = 0.52\ h^{-1}$ for cut-in scenarios, and $\hat{\sigma}_{\text{exposure}} = 0.34\ h^{-1}$ for ASV scenarios.

Figure 1 shows the estimated exposure of Eq. (1) for an increasing number of hours of driving data. The colored area around each line represents the uncertainty; that is, the area between $N/n - \hat{\sigma}_{\text{exposure}}$ and $N/n + \hat{\sigma}_{\text{exposure}}$. Figure 1 illustrates that the uncertainty decreases when more data are available: the height of the colored area decreases for higher values of n. Furthermore, it can be observed that a higher exposure typically results in a higher uncertainty. In relative terms—that is, the ratio of $\hat{\sigma}_{\text{exposure}}$ and N/n—the uncertainty is typically lower for higher values of N/n because $\hat{\sigma}_{\text{exposure}}$ is typically proportional with the square root of N/n (cf. Eq. (9)).

Uncertainty because of limited data

As reported in Table 2, the estimated crash probability is $7.32 \cdot 10^{-3}$ in case the ACC encounters an LVD scenario. The uncertainty resulting from the variance of $\hat{f}(\cdot)$ is estimated using bootstrapping according to Eq. (11) with $B = 1000$. This results in $\hat{\sigma}_{\mu,\text{data}} = 1.52 \cdot 10^{-3}$. The estimated crash probability for a cut-in scenario is about 4 times lower: $1.88 \cdot 10^{-3}$. Note, however, that the uncertainty resulting from the limited data is only slightly lower: $1.38 \cdot 10^{-3}$. This result illustrates the influence of the number of scenarios, N, that are used to compute $\hat{f}(\cdot)$: more scenarios lead to a relatively lower uncertainty. In other words, $\hat{\sigma}_{\mu,\text{data}}/\mu_{\text{NIS}}$ is generally lower for higher values of N. For the ASV scenarios, $\hat{\sigma}_{\mu,\text{data}} = 5.05 \cdot 10^{-3}$ is also relatively large compared to $\mu_{\text{NIS}} = 9.20 \cdot 10^{-3}$.

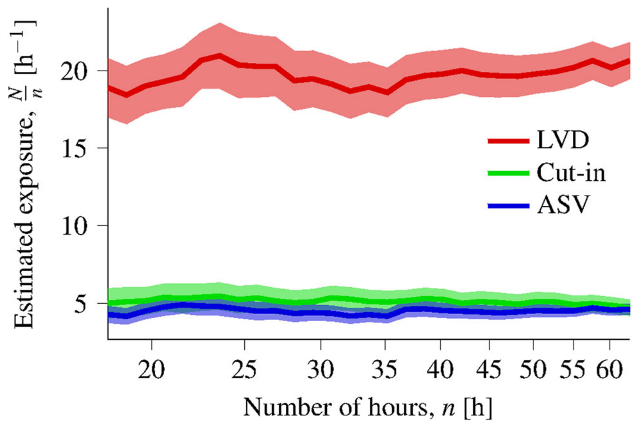

Figure 1. Estimated exposure of LVD, cut-in, and ASV scenarios based on n hours of collected data, where n varies between 18 and 63 h. The colored areas mark the estimated exposure plus or minus the uncertainty, $\hat{\sigma}_{\text{exposure}}$.

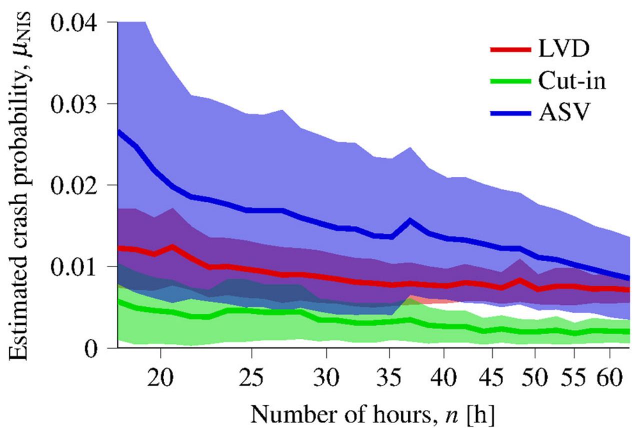

Figure 2. Estimated crash probability of LVD, cut-in, and ASV scenarios based on n hours of collected data, where n varies between 18 and 63 h. The colored areas mark the estimated exposure plus or minus the uncertainty $\hat{\sigma}_{\mu,\text{data}}$, which results from the variance of the estimated PDF; see Eq. (11).

Figure 2 shows the result of the bootstrapping for different number of hours of data. The lines represent the mean of the values $\mu^*_{\text{NIS},l}$, $l \in \{1, ..., B\}$. Note that this mean might deviate from the μ_{NIS} from Eq. (5). The colored areas denote the mean plus or minus $\hat{\sigma}_{\mu,\text{data}}$. Figure 2 clearly illustrates that the uncertainty $\hat{\sigma}_{\mu,\text{data}}$ decreases with increasing n. This is an expected result because the variance of $\hat{f}(\cdot)$ decreases if more data are used. Perhaps more surprising is the effect that the estimated crash probability itself is decreasing with the use of more data. One explanation for this is that with less data, the bandwidth used for the KDE tends to be larger. As a result, the tails of the estimated PDF, $\hat{f}(\cdot)$, tend to be larger and the crashes typically

Table 2. Summary of results with $n = 63$ h, $N_{MC} = N_{NIS} = 10,000$, and $N_C = 200$.

Scenario category	N/n (h^{-1})	$\hat{\sigma}_{\text{exposure}}$ (h^{-1})	μ_{NIS}	$\hat{\sigma}_{\mu,\text{data}}$	$\hat{\sigma}_{\mu,\text{simulations}}$	Risk(C) (h^{-1})
LVD	20.6	1.2	$7.32 \cdot 10^{-3}$	$1.52 \cdot 10^{-3}$	$1.33 \cdot 10^{-4}$	0.151
Cut-in	4.71	0.52	$1.88 \cdot 10^{-3}$	$1.38 \cdot 10^{-3}$	$9.04 \cdot 10^{-5}$	$8.85 \cdot 10^{-3}$
ASV	4.62	0.34	$9.20 \cdot 10^{-3}$	$5.05 \cdot 10^{-3}$	$1.33 \cdot 10^{-4}$	$4.25 \cdot 10^{-2}$

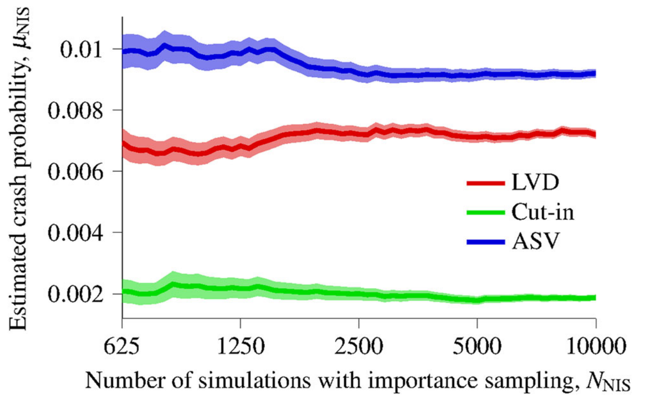

Figure 3. Estimated crash probability of LVD, cut-in, and ASV scenarios based on N_{NIS} simulations, where N_{NIS} varies between 625 and 10,000. The colored areas mark the estimated uncertainty $\hat{\sigma}_{\mu,\text{simulations}}$, which results from the limited number of simulations; see Eq. (12).

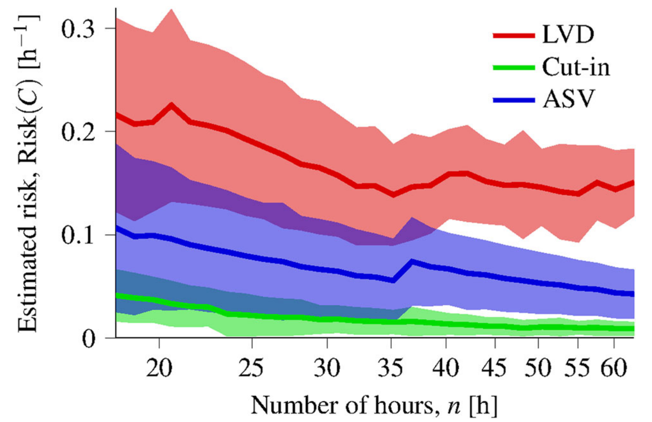

Figure 4. Estimated risk of a collision in LVD, a cut-in, and an ASV scenario per hour of driving. The colored areas mark the estimated uncertainty, $\sqrt{\mathbb{V}[\text{Risk}(C)]}$, where $\mathbb{V}[\text{Risk}(C)]$ is estimated using Eq. (15).

occur at the tails of the parameter distributions. It is expected that this effect diminishes as $Nh^d \to \infty$ (Cadre 2006).

Uncertainty because of limited number of simulations

Table 2 reports the values of $\hat{\sigma}_{\mu,\text{simulations}}$ for the 3 scenario categories in case $N_{MC} = N_{NIS} = 10,000$ and $N_C = N_{MC}/50 = 200$. The uncertainty resulting from the limited number of simulations is an order of magnitude smaller than the uncertainty resulting from the limited data; that is, $\hat{\sigma}_{\mu,\text{data}} \gg \hat{\sigma}_{\mu,\text{simulations}}$.

In Figure 3, the estimated crash probability is shown for varying number of performed simulations during the importance sampling. Figure 3 shows that $\hat{\sigma}_{\mu,\text{simulations}}$ is substantially lower than $\hat{\sigma}_{\mu,\text{data}}$, even if fewer simulation runs are performed. Note that also in case of lower values of N_{NIS}, there are still $N_{MC} = 10,000$ simulation runs performed during the crude Monte Carlo sampling. Using lower values of N_{MC} may result in wrong choices of the importance density, thus resulting in a slow convergence of the result during importance sampling. Figure 3 illustrates the effect of using more simulation runs to estimate the crash probability: The uncertainty of the estimated crash probability decreases if more simulation runs are used.

Uncertainty of the estimated risk

Using Eq. (6), the estimated exposure and crash probability are multiplied to estimate the risk. When using the complete data set of 63 h of driving data and $N_{NIS} = 10,000$, it is estimated that the evaluated ACC crashes about 0.151 times in an LVD scenario per hour of driving; see Table 2. For the cut-in and ASV scenarios, this is $8.85 \cdot 10^{-3}$ and $4.25 \cdot 10^{-2}$ times per hour, respectively. Note that these risk estimations assume that a human driver does not intervene. Note that the estimated crash probability in an ASV scenario is higher than the estimated crash probability in an LVD scenario, but due to the higher exposure of LVD scenarios, the estimated risk for LVD scenarios is higher.

Employing the uncertainties of the exposure and the crash probability, the variance of the estimated risk can be estimated using Eq. (15). The estimated uncertainty of the estimated risk is listed in Table 3. The standard deviation of the estimated risk of the LVD scenario category is $3.26 \cdot 10^{-2}$ h^{-1}, which is roughly 5 times smaller than the actual risk. For the cut-in and ASV scenario categories, the uncertainties of the estimated risks are lower: $6.64 \cdot 10^{-3}$ and $2.36 \cdot 10^{-2}$ h^{-1}, respectively. However, relatively speaking, the uncertainties are higher, because the uncertainties are 75 % and 56 %, respectively, of the estimated risk.

When looking at the 3 terms in Eq. (15) that constitute the variance of the risk estimation, the first term is an order of magnitude higher than the other 2 terms; see Table 3. In fact, the first term contributes 93 % to the total variance ($9.87 \cdot 10^{-4}$ of $1.06 \cdot 10^{-3}$), whereas the second only contributes 7 % and the third term contributes less than 1 %. In other words, in this example, the exposure and the uncertainty of the crash probability (first term in Eq. (15)) have the largest influence on the estimated uncertainty of the risk, whereas the crash probability and the uncertainty of the exposure (second term in Eq. (15)) contribute substantially less to the risk uncertainty. For the cut-in and ASV scenario categories, the first term of Eq. (15) contributes even more to the total variance: 97 % and 98 %, respectively.

Figure 4 shows the estimated risk and its estimated uncertainty when using different amounts of data. This figure illustrates that using more data leads to less uncertainty of the estimated risk. It is interesting to note the small peak of the estimated risk for the ASV scenario category at around $n = 37$ h. An explanation for this is that the data contain an ASV scenario after a bit less than 37 h of driving

Table 3. Uncertainty of the estimated risk and the contribution of each of the terms to the total variance with $n = 63$ h, $N_{MC} = N_{NIS} = 10,000$, and $N_C = 200$.

Scenario category	Uncertainty (h^{-1})	$\mathbb{V}[\text{Risk}(C)]$ (h^{-2})	$\left(\frac{N}{n}\right)^2 \cdot \hat{\sigma}_\mu^2$ (h^{-2})	$\mu_{NIS}^2 \cdot \hat{\sigma}_{\text{exposure}}^2$ (h^{-2})	$\hat{\sigma}_{\text{exposure}}^2 \cdot \hat{\sigma}_\mu^2$ (h^{-2})
LVD	$3.26 \cdot 10^{-2}$	$1.06 \cdot 10^{-3}$	$9.87 \cdot 10^{-4}$	$7.34 \cdot 10^{-5}$	$3.21 \cdot 10^{-6}$
Cut-in	$6.64 \cdot 10^{-3}$	$4.41 \cdot 10^{-5}$	$4.27 \cdot 10^{-5}$	$9.45 \cdot 10^{-7}$	$5.15 \cdot 10^{-7}$
ASV	$2.36 \cdot 10^{-2}$	$5.58 \cdot 10^{-4}$	$5.45 \cdot 10^{-4}$	$9.96 \cdot 10^{-6}$	$3.01 \cdot 10^{-6}$

in which the ego vehicle is driving at 27.9 m/s while approaching a vehicle that drives only 11.5 m/s. Adding this scenario to the data results in a change of the estimated PDF of the scenario parameters that has a large influence on the tail of the PDF. Because most crashes occur for scenario parameters within the tail of the PDF, the inclusion of the scenario just before 37 h of driving influences the estimated crash probability. Note that these effects become less substantial as more and more data are used.

Discussion

In this work, we have presented a method to estimate the uncertainty of the estimated risk of ADS as a result of lack of data. This lack of data concerns the limited scenarios that were observed in the data and the limited simulations that were carried out. Note that uncertainties of the estimated risk may also result from the following:

- Inaccuracies in the acquired data or inaccuracies in the simulation of the scenarios.
- Simplification of the scenarios; for example, by assuming that the scenario of a specific scenario category can be described by a finite number of parameters.
- A misspecification of the ODD of the ADS. For example, if the actual ODD of an ADS is substantially different from the operating conditions under which the data were acquired, the data may not be representative of the actual ODD. This may result in inaccuracies of the estimated exposure and the estimated PDF of the scenario parameters.

As mentioned in the problem definition, we distinguish between scenarios and scenario categories. In this work, a method is provided to determine whether the variability of the scenarios within a scenario category is sufficiently considered. Another important question is whether all relevant scenario categories are considered during the risk assessment. To tackle this question, another approach might be taken. Whereas the scenarios—described with parameter values—within a scenario category are uncountable, scenario categories may be treated as countable. In Hauer et al. (2019), the problem of estimating whether all scenario categories are observed in real traffic is treated as the coupon collector's problem. Another way to look at this problem is to consider it as the so-called unseen species problem (Bunge and Fitzpatrick 1993). In case of the unseen species problem, the entire population is partitioned into M classes and the objective is to estimate M given only a part of the entire population.

The uncertainty of the estimated risk can be used as a metric to quantify whether more data have to be collected or more simulations have to be conducted, because more data and more simulations will generally result in a lower uncertainty of the estimated risk. In de Gelder et al. 2019, an alternative metric to quantify the completeness of the data was proposed. The metric of de Gelder et al. (2019) estimates the similarity between the estimated PDF of the scenario parameters and the real underlying, unknown PDF. To do this, the whole PDF is considered in de Gelder et al. 2019. In this work, however, only the part of the PDF where the scenario parameters describe a scenario that results in a crash—that is, $R(x) = 1$—is considered. If the purpose of the data is to assess the risk of an ADS, then the accuracy of the estimated PDF for scenario parameters that do not lead to scenarios in which the vehicle with the ADS crashes is less relevant. Therefore, in that case, the method provided in this work may be better to answer the question of whether enough data were collected. Another consideration is that compared to the metric in de Gelder et al. (2019), this work's method to quantify the uncertainty of the estimated risk requires (a model of) an ADS.

When using the uncertainty of the estimated risk to determine whether more data have to be collected or more simulations have to be conducted, a threshold needs to be chosen. Only in case of an infinite set of data and an infinite number of simulations does the uncertainty approach 0. When choosing a threshold, this might be a threshold for the uncertainty itself, but it is also possible to require enough certainty that the risk is below a certain threshold. In the latter case, this implies that the uncertainty may be larger as long as the estimated risk is equivalently lower.

We used virtual simulations to estimate the risk. Note that the simulations we conducted are deterministic. High-fidelity simulators typically also include stochastic behavior of, for example, the sensor models, to replicate the imperfect sensors that are used in real life. Modeling the stochastic nature is a whole research topic on its own and is outside the scope of this work. We refer the interested reader to Rosique et al. (2019) and Kaur et al. (2021) for an overview of simulators used for the assessment of ADSs.

The case study illustrates the application of the provided method for estimating the risk and its uncertainty. Note, however, that the case study comes with a few limitations. First, the actual data set considers only 63 h of driving. To do a complete risk assessment of an ADS, much more data are required. Second, as a proof of concept, we considered a simplified ACC controller as our ADS under test. The results of the ACC controller may not represent state-of-the-art ADSs. In particular, the fact that the risk of a crash in an LVD scenario is estimated to be 0.151 per hour is likely to be unacceptable. We chose to use the simplified ACC controller for several reasons. First, using this simple controller contributes to the explainability of the results, ensures short simulation run times, and facilitates the reproducibility of the results. Second, though the actual ACC model is not the focus of the article, the proposed method for the estimation of the risk uncertainty can also be applied to state-of-the-art ADSs. Third, the ACC controller studied in this work is often used in the literature (Milanés and Shladover 2014; Xiao et al. 2017). Fourth, for a state-of-the-art ADS with a much lower crash probability, it is expected that much more data will be required to obtain enough certainty. Therefore, for the purpose of demonstration, the current studied ACC might be more suitable. Future work involves applying the method on a larger data set and considering a more mature ADS.

For ADSs with conditional or higher automation level, the responsibility and accountability shift from the human driver to the ADS when the ADS is activated. Therefore, for the safe introduction of such ADSs, it must be assured that the safety risk is lower than the existing state of the art and that the current safety state is not compromised. One proposed method for estimating the risk is through a data-driven, scenario-based assessment. This work has summarized this method and uses scenarios extracted from existing data to derive tests that are used to evaluate the exposure of scenarios and the crash probability. The risk is defined as the product of the exposure and the crash probability. The resulting risk is directly influenced by the data set used for the risk assessment. The fact that the data set is finite introduces an uncertainty in the estimated exposure and crash probability. In this work, we show how these uncertainties can be estimated and how this leads to the uncertainty of the estimated risk itself. This, in turn, can be used to quantify how certain we are that our ADS is safe.

To illustrate the method, we applied the method to estimate the risk and its uncertainty in a case study that is publicly available. This case study considered 3 types of scenarios: scenarios with a leading vehicle decelerating, cut-in scenarios, and scenarios in which the ego vehicle approaches a slower leading vehicle. For each of these 3 scenario categories, the uncertainty of the exposure, the crash probability, and the risk were estimated while considering a varying amount of data. This illustrates that the uncertainty is expected to decrease if more data are available. Furthermore, when more simulation runs are conducted when estimating the crash probability, the uncertainty decreases. In the case study, though, the limited amount of data appeared to contribute substantially more to the uncertainty of the risk than the limited number of simulation runs.

Future work involves applying the proposed method with more data and a state-of-the-art ADS. Because it is expected that the fatality rate of an ADS is on the order of 10^{-7} fatalities per hour of driving or less, the amount of data to obtain enough certainty is likely to be orders of magnitude larger than the data set used in the case study. The proposed method for quantifying the uncertainty of the estimated risk can be a measure for determining whether more data are required. Other future work involves incorporating the proposed method into the type approval framework for future ADSs of SAE levels 3, 4, and 5, such as proposed in the draft European Union implementing regulation for ADS (European Commission 2022).

Funding

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101006664. This work reflects only the authors' views, and neither CINEA nor the EC is responsible for any use that may be made of the information it contains.

ORCID

Erwin de Gelder http://orcid.org/0000-0003-4260-4294
Olaf Op den Camp http://orcid.org/0000-0002-6355-134X

References

Antona-Makoshi J, Uchida N, Yamazaki K, Ozawa K, Kitahara E, Taniguchi S. 2019. Development of a safety assurance process for autonomous vehicles in Japan. Paper presented at: 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV). p. 1–18, https://www-esv.nhtsa.dot.gov/Proceedings/26/26ESV-000286.pdf.

Bunge J, Fitzpatrick M. 1993. Estimating the number of species: a review. J Am Stat Assoc. 88(421):364–373. doi: 10.1080/01621459.1993.10594330.

Burns P. 2002. Robustness of the Ljung-Box test and its rank equivalent. Technical report. doi: 10.2139/ssrn.443560.

Cadre B. 2006. Kernel estimation of density level sets. J Multivariate Anal. 97(4):999–1023. doi: 10.1016/j.jmva.2005.05.004.

Chan C-Y. 2017. Advancements, prospects, and impacts of automated driving systems. Int J Transp Sci Technol. 6(3):208–216. doi: 10.1016/j.ijtst.2017.07.008.

Chen Y-C. 2017. A tutorial on kernel density estimation and recent advances. Biostat Epidemiol. 1(1):161–187. doi: 10.1080/24709360.2017.1396742.

de Gelder E, Adjenughwure K, Manders J, Snijders R, Paardekooper J-P, Op den Camp O, Tejada Ruiz A, De Schutter B. 2023. PRISMA: a novel approach for deriving probabilistic surrogate safety measures for risk evaluation. Under review. https://arxiv.org/abs/2303.07891.

de Gelder E, Elrofai H, Khabbaz Saberi A, Op den Camp O, Paardekooper J-P, De Schutter B. 2021. Risk quantification for automated driving systems in real-world driving scenarios. IEEE Access. 9:168953–168970. doi: 10.1109/ACCESS.2021.3136585.

de Gelder E, Manders J, Grappiolo C, Paardekooper J-P, Op den Camp O, Schutter D. 2020. B. Real-world scenario mining for the assessment of automated vehicles. Paper presented at: IEEE International Transportation Systems Conference (ITSC). p. 1073–1080. doi: 10.1109/ITSC45102.2020.9294652.

de Gelder E, Paardekooper J-P, Khabbaz Saberi A, Elrofai H, Op den Camp O, Kraines S, Ploeg J, De Schutter B. 2022. Towards an ontology for scenario definition for the assessment of automated vehicles: An object-oriented framework. IEEE Trans Intell Veh. 7(2):300–314. doi: 10.1109/TIV.2022.3144803.

de Gelder E, Paardekooper J-P, Op den Camp O, Schutter D. 2019. B. Safety assessment of automated vehicles: How to determine whether we have collected enough field data? Traffic Inj Prev. 20(S1):162–170. doi: 10.1080/15389588.2019.1602727.

Duong T. 2007. ks: Kernel density estimation and kernel discriminant analysis for data in R. Journal of Statistical Software. 21(7):1–16. doi: 10.18637/jss.v021.i07.

Efron B. 1979. Bootstrap methods: another look at the jackknife. Ann Stat. 7(1):1–26. doi: 10.1214/aos/1176344552.

Elfring J, Appeldoorn R, van den Dries S, Kwakkernaat M. 2016. Effective world modeling: Multisensor data fusion methodology for automated driving. Sensors. 16(10):1–27. doi: 10.3390/s16101668.

Elrofai H, Paardekooper J-P, de Gelder E, Kalisvaart S, Op den Camp O. 2018. Scenario-based safety validation of connected and automated driving. Technical report. Helmond (the Netherlands): Netherlands Organization for Applied Scientific Research, TNO. http://publications.tno.nl/publication/34626550/AyT8Zc/TNO-2018-streetwise.pdf.

European Commission. 2022. Commision implementing act AD v4.1. Draft implementing regulation. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=PI_COM:Ares(2022)2667391.

Goodman LA. 1960. On the exact variance of products. J Am Stat Assoc. 55(292):708–713. doi: 10.1080/01621459.1960.10483369.

Gyllenhammar M, Johansson R, Warg F, Chen D, Heyn H-M, Sanfridson M, Söderberg J, Thorsén A, Ursing S. 2020. Towards an operational design domain that supports the safety argumentation of an automated driving system. Paper presented at: 10th European Congress on Embedded Real Time Systems (ERTS). https://www.diva-portal.org/smash/get/diva2:1390550/FULLTEXT01.pdf.

Hauer F, Schmidt T, Holzmüller B, Pretschner A. 2019. Did we test all scenarios for automated and autonomous driving systems? Paper presented at: IEEE Intelligent Transportation Systems Conference (ITSC). p. 2950–2955. doi: 10.1109/itsc.2019.8917326.

Hayward JC. 1972. Near miss determination through use of a scale of danger. Technical Report TTSC-7115. State College: Pennsylvania State University. https://onlinepubs.trb.org/Onlinepubs/hrr/1972/384/384-004.pdf.

ISO 21448. 2022. Road vehicles – safety of the intended functionality. Geneva (Switzerland): Standard, International Organization for Standardization. https://www.iso.org/standard/77490.html.

ISO 26262. 2018. Road vehicles – functional safety. Geneva (Switzerland): Standard, International Organization for Standardization. https://www.iso.org/standard/68383.html.

Kaur P, Taghavi S, Tian Z, Shi W. 2021. A survey on simulators for testing self-driving cars. Paper presented at: Fourth International Conference on Connected and Autonomous Driving (MetroCAD). p. 62–70. doi: 10.1109/MetroCAD51599.2021.00018.

Ljung GM, Box GE. 1978. On a measure of lack of fit in time series models. Biometrika. 65(2):297–303. doi: 10.1093/biomet/65.2.297.

Luetge C. 2017. The German ethics code for automated and connected driving. Philos Technol. 30(4):547–558. doi: 10.1007/s13347-017-0284-0.

Menzel T, Bagschik G, Maurer M. 2018. Scenarios for development, test and validation of automated vehicles. Paper presented at: IEEE Intelligent Vehicles Symposium (IV), p. 1821–1827. doi: 10.1109/ivs.2018.8500406

Milanés V, Shladover SE. 2014. Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data. Transp Res Part C: Emerging Technol. 48:285–300. doi: 10.1016/j.trc.2014.09.001.

Mullakkal-Babu FA, Wang M, Farah H, van Arem B, Happee R. 2017. Comparative assessment of safety indicators for vehicle trajectories on highways. Transp Res Rec. 2659(1):127–136. doi: 10.3141/2659-14.

Najm WG, Smith JD, Yanagisawa M. 2007. Pre-crash scenario typology for crash avoidance research. Technical Report DOT HS 810 767, U.S. Cambridge (MA): Department of Transportation Research and Innovative Technology Administration. https://rosap.ntl.bts.gov/view/dot/6281/dot_6281_DS1.pdf.

Owen AB. 2013. Monte Carlo theory, methods and examples. https://statweb.stanford.edu/~owen/mc/.

Paardekooper J-P, Montfort S, Manders J, Goos J, de Gelder E, Op den Camp O, Bracquemond A, Thiolon G. 2019. Automatic identification of critical scenarios in a public dataset of 6000 km of public-road driving. Paper presented at: 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV). https://www-esv.nhtsa.dot.gov/Proceedings/26/26ESV-000255.pdf.

Parzen E. 1962. On estimation of a probability density function and mode. Ann Math Stat. 33(3):1065–1076. doi: 10.1214/aoms/1177704472.

Ploeg J, de Gelder E, Slavík M, Querner E, Webster T, de Boer N. 2018. Scenario-based safety assessment framework for automated vehicles. Paper presented at: 16th ITS Asia-Pacific Forum. p. 713–726. https://arxiv.org/abs/2112.09366.

Pütz A, Zlocki A, Bock J, Eckstein L. 2017. System validation of highly automated vehicles with a database of relevant traffic scenarios. Paper presented at: 12th ITS European Congress. p. 1–8, https://www.pegasusprojekt.de/files/tmpl/pdf/12th%20ITS%20European%20Congress_Folien.pdf.

Riedmaier S, Ponn T, Ludwig D, Schick B, Diermeyer F. 2020. Survey on scenario-based safety assessment of automated vehicles. IEEE Access. 8:87456–87477. doi: 10.1109/ACCESS.2020.2993730.

Roesener C, Sauerbier J, Zlocki A, Fahrenkrog F, Wang L, Várhelyi A, de Gelder E, Dufils J, Breunig S, Mejuto P, et al. 2017. A comprehensive evaluation approach for highly automated driving. Paper presented at: 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV). https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000259.pdf.

Rosenblatt M. 1956. Remarks on some nonparametric estimates of a density function. Ann Math Stat. 27(3):832–837. doi: 10.1214/aoms/1177728190.

Rosique F, Navarro PJ, Fernández C, Padilla A. 2019. A systematic review of perception system and simulators for autonomous vehicles research. Sensors. 19(3):648. doi: 10.3390/s19030648.

SAE J3016. 2021. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Technical report. Warrendale (PA): SAE International.

Scholtes M, Westhofen L, Turner LR, Lotto K, Schuldes M, Weber H, Wagener N, Neurohr C, Bollmann MH, Körtke F, et al. 2021. 6-layer model for a structured description and categorization of urban traffic and environment. IEEE Access. 9:59131–59147. doi: 10.1109/ACCESS.2021.3072739.

Thorn E, Kimmel S, Chaka M. 2018. A framework for automated driving system testable cases and scenarios. Technical Report DOT HS 812623. Washington (DC): National Highway Traffic Safety Administration. https://rosap.ntl.bts.gov/view/dot/38824.

Turlach BA. 1993. Bandwidth selection in kernel density estimation: a review. Technical report, Berlin: Institut für Statistik und Ökonometrie, Humboldt-Universität zu. https://www.researchgate.net/publication/2316108_Bandwidth_Selection_in_Kernel_Density_Estimation_A_Review.

Wang C, Xie Y, Huang H, Liu P. 2021. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. Accid Anal Prev. 157:106157. doi: 10.1016/j.aap.2021.106157.

Wang W, Liu C, Zhao D. 2017. How much data are enough? A statistical approach with case study on longitudinal driving behavior. IEEE Trans Intell Veh. 2(2):85–98. doi: 10.1109/tiv.2017.2720459.

Wasserman L. 2006. All of nonparamatric statistics. New York (NY): Springer.

Westhofen L, Neurohr C, Koopmann T, Butz M, Schütt B, Utesch F, Neurohr B, Gutenkunst C, Böde E. 2022. Criticality metrics for automated driving: A review and suitability analysis of the state of the art. Arch Comput Methods Eng. 30:1–35. doi: 10.1007/s11831-022-09788-7.

World Forum for Harmonization of Vehicle Regulations. 2021. E/ECE/TRANS/505/Rev.3/Add.156 Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems. Geneva (Switzerland): United Nations. https://unece.org/sites/default/files/2021-03/R157e.pdf

Xiao L, Wang M, van Arem B. 2017. Realistic car-following models for microscopic simulation of adaptive and cooperative adaptive cruise control vehicles. Transp Res Rec. 2623(1):1–9. doi: 10.3141/2623-01.

Zhang P. 1996. Nonparametric importance sampling. J Am Stat Assoc. 91(435):1245–1253. doi: 10.1080/01621459.1996.10476994.