**Delft University of Technology**

**Delft Center for Systems and Control**

Technical report 23-007

# PRISMA: A novel approach for deriving probabilistic surrogate safety measures for risk evaluation*

E. de Gelder, K. Adjenughwure, J. Manders, R. Snijders,
J.-P. Paardekooper, O. Op den Camp, A. Tejada, and B. De Schutter

*If you want to cite this report, please use the following reference instead:*

E. de Gelder, K. Adjenughwure, J. Manders, R. Snijders, J.-P. Paardekooper, O. Op den Camp, A. Tejada, and B. De Schutter, "PRISMA: A novel approach for deriving probabilistic surrogate safety measures for risk evaluation," *Accident Analysis & Prevention*, vol. 192, p. 107273, Nov. 2023. doi:10.1016/j.aap.2023.107273

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: https://www.dcsc.tudelft.nl

* This report can also be downloaded via https://pub.bartdeschutter.org/abs/23_007.html

# PRISMA: A Novel Approach for Deriving Probabilistic Surrogate Safety Measures for Risk Evaluation

Erwin de Gelder[a,b,*], Kingsley Adjenughwure[c], Jeroen Manders[a], Ron Snijders[d], Jan-Pieter Paardekooper[a,e], Olaf Op den Camp[a], Arturo Tejada[a,f], Bart De Schutter[b]

[a] *TNO, Integrated Vehicle Safety, Helmond, The Netherlands*
[b] *Delft University of Technology, Delft Center for Systems and Control, Delft, The Netherlands*
[c] *TNO, Sustainable Urban Mobility & Safety, The Hague, The Netherlands*
[d] *TNO, Monitoring & Control Services, Groningen, The Netherlands*
[e] *Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands*
[f] *Eindhoven University of Technology, Dynamics and Control Group, Eindhoven, The Netherlands*

**Abstract**

Surrogate Safety Measures (SSMs) are used to express road safety in terms of the safety risk in traffic conflicts. Typically, SSMs rely on assumptions regarding the future evolution of traffic participant trajectories to generate a measure of risk, restricting their applicability to scenarios where these assumptions are valid. In response to this limitation, we present the novel Probabilistic RISk Measure derivAtion (PRISMA) method. The objective of the PRISMA method is to derive SSMs that can be used to calculate in real time the probability of a specific event (e.g., a crash). The PRISMA method adopts a data-driven approach to predict the possible future traffic participant trajectories, thereby reducing the reliance on specific assumptions regarding these trajectories. Since the PRISMA is not bound to specific assumptions, the PRISMA method offers the ability to derive multiple SSMs for various scenarios. The occurrence probability of the specified event is based on simulations and combined with a regression model, this enables our derived SSMs to make real-time risk estimations.

To illustrate the PRISMA method, an SSM is derived for risk evaluation during longitudinal traffic interactions. Since there is no known method to objectively estimate risk from first principles, i.e., there is no known risk ground truth, it is very difficult, if not impossible, to objectively compare the relative merits of two SSMs. Instead, we provide a method for benchmarking our derived SSM with respect to expected risk trends. The application of the benchmarking illustrates that the SSM matches the expected risk trends.

Whereas the derived SSM shows the potential of the PRISMA method, future work involves applying the approach for other types of traffic conflicts, such as lateral traffic conflicts or interactions with vulnerable road users.

## 1. Introduction

Road safety is an important key performance indicator in transportation. In addition to the suffering of people as a consequence of crashes in traffic, these crashes cause enormous societal and economic losses. As a result, road safety research is an important research topic. For example, in 2018[1], there were over 6.7 million crashes in the U.S.A. (National Center for Statistics and Analysis, 2020), which is about 1.3 crashes per 1 million vehicle kilometers driven. These crashes in 2018 led to 2.7 million injured people and 37 thousand fatalities (National Center for Statistics and Analysis, 2020). Furthermore, apart from these societal losses, the economic costs of all crashes in the U.S.A. in 2018 was 242 billion dollars (National Center for Statistics

*Corresponding author
[1] At the time of writing, more recent results were not yet available.

and Analysis, 2020). Similarly, the European Commission (2020) reported over 22 thousand fatalities in 2019.

Road safety can be expressed in terms of injuries, fatalities, or crashes per kilometer of driving, but "that is a slow, reactive process" (Arun et al., 2021). Furthermore, "crashes are rare events and historical crash data does not capture near crashes that are also critical for improving safety" (Wang et al., 2021). An alternative for expressing road safety that does not rely on historical crash data is the use of safety indicators that directly measure the safety risk in traffic conflicts (Tarko, 2018b; Arun et al., 2021; Wang et al., 2021). Traffic conflicts are far more frequent than traffic crashes and the frequency of traffic conflicts can be used to predict the frequency of crashes (Davis et al., 2011; Tarko, 2018a). To define traffic conflicts, thresholds on so-called Surrogate Safety Measures (SSMs) are used, where SSMs characterize the risk of a crash or harm given an initial condition (Arun et al., 2021). SSMs vary from measures that estimate the remaining time until a crash, such as the well-known Time to Collision (TTC) (Hayward, 1972), to metrics that estimate the probability that a human driver cannot avoid a crash, see, e.g., (Wang and Stamatiadis, 2014).

SSMs typically rely on assumptions of what drivers or systems controlling the vehicles of interest are capable of doing and how their future trajectories — given an initial condition — will develop. For example, TTC (Hayward, 1972), the ratio of the distance toward and the speed difference with an approaching object, is computed by assuming a constant relative velocity. As a result of these assumptions, SSMs are only applicable in certain types of scenarios. For example, TTC is only applicable when approaching an object. More complex SSMs consider, e.g., a human model that can react to a risky situation by braking (Wang and Stamatiadis, 2014) or the uncertainty over the future ambient traffic state (Mullakkal-Babu et al., 2020). Regardless of the complexity of these models, however, these SSMs consider neither the specific capabilities of the driver or of the system controlling the vehicle, nor the local context for predicting the future of the vehicle's environment.

This paper presents the Probabilistic RISk Measure derivAtion (PRISMA) method, which is a data-driven approach for deriving SSMs that are not limited to certain types of scenarios. Because the method is not bound to certain predetermined assumptions about driver behavior, the derived SSMs can be adapted to the situations in which they are applied. In addition, to avoid relying on predetermined assumptions on how the ambient traffic evolves over time, the PRISMA method includes a data-driven approach for modeling the variations of the trajectories of the ambient traffic. Monte Carlo simulations are employed to predict the safety risk given these variations. To enable the real-time evaluation of the derived SSMs, we use the Nadaraya-Watson (NW) kernel estimator (Wasserman, 2006) for local regression. The PRISMA method has the following characteristics:

- The derived SSMs give a probability that a specified event, e.g., a crash or a near miss will happen in the near future, e.g., within the next 10 seconds, given an initial state and the foreseen evolutions of traffic participant trajectories. Since a traffic conflict can be defined as the probability of an unsuccessful evasion in a traffic interaction, according to Davis et al. (2011), a probability is easier to interpret than, e.g., a value ranging from 0 to infinity such as the TTC.

- Next to deriving new SSMs, i.e., new ways to estimate the probability of an event such as a crash, it is possible to reproduce already existing measures that provide a probability. Therefore, the PRISMA method can be seen as a generalization for deriving such existing SSMs.

- A driver behavior model can be used. It is also possible to use a model of an Automated Driving System (ADS), such that the derived SSM estimates the safety risk if this ADS controls the vehicle.

- Because a data-driven approach is adopted, the derived SSM adapts to the recorded data. In this way, it is possible to adapt the SSM to, e.g., the local traffic behavior provided that this local traffic behavior is captured by the recorded data.

- The PRISMA method is not limited to one type of scenario.

We illustrate the PRISMA method and its benefits by means of a case study. The case study demonstrates that when using the PRISMA method with the assumptions of the SSM of Wang and Stamatiadis (2014), both the SSM derived by the PRISMA method and the latter yield the same result. The case study continues with evaluating the crash risk of three longitudinal traffic conflicts which are a priori known to be, respectively, safe (i.e., no crash possible), moderately safe, and unsafe (i.e., a crash occurs), based on vehicle kinematics. We evaluate the risk of each of the scenarios using the SSM by Wang and Stamatiadis (2014) and an SSM derived by the PRISMA method, based on data from the Next Generation SIMulation (NGSIM) (Alexiadis et al., 2004). Moreover, since a comparison between these measures is not directly possible in general scenarios, a method to benchmark SSMs using expected risk trends is introduced in the case study.

This article is organized as follows. Section 2 provides an overview of SSMs described in the literature. The proposed PRISMA method is presented in Section 3. In Section 4, we illustrate the method in a case study. Some implications of this work are discussed in Section 5. The article is concluded in Section 6.

## 2. Literature review

Risk in the context of traffic safety is often defined as the probability of a crash occurring (Hakkert et al., 2002). Most SSMs are derived under specific assumptions of the expected behavior of the driving participants under a specific driving scenario. Several SSMs have been developed under such assumptions with the goal of quantifying the risk involved in driving in traffic on the road (Minderhoud and Bovy, 2001; Ozbay et al., 2008; Cunto and Saccomanno, 2009; Laureshyn et al., 2010). In general, the risk is quantified in terms of the proximity between two traffic agents in time and/or space, the ability to perform evasive actions like braking or swerving, or the magnitude of such actions (Shi et al., 2018; Zheng et al., 2020). In a potential crash situation, the proximity indicator is close to zero while the magnitude of evasive action is close to the limits of the driver and the vehicle (Zheng et al., 2020). The above clustering of SSMs in terms of time, space, and evasive action is common in the literature; so our literature review follows this pattern of clustering SSMs. We focus on the most commonly used measures in each cluster and their underlying assumptions. Additionally, we discuss some well-known SSM-based metrics, i.e., metrics that are derived from other SSMs.

The most common SSMs are time-based. A popular time-based proximity indicator is the TTC, which is an estimate of the remaining time until two vehicles collide and is defined as the time remaining until two vehicles collide if they would continue on the same course and speed (Hayward, 1972). The assumption for the TTC is that the relative speed and course will remain the same. In addition, the TTC is only relevant when two objects are approaching each other. These assumptions make it difficult to use it for various driving scenarios. Several other time-based SSMs have been derived from or based on the TTC. Notable among those are:

- the time-exposed TTC, which measures the amount of time the TTC is below a certain threshold (Minderhoud and Bovy, 2001);
- the Time Integrated TTC (TIT), which calculates the total area in a TTC versus time diagram where the TTC is below a certain threshold (Minderhoud and Bovy, 2001);
- the Modified TTC (MTTC), which is able to calculate the TTC for cases where vehicles do not keep a constant speed and the follower is slower than the leader (Ozbay et al., 2008); and
- the Time to Collision Disturbance (TTCD), which calculates the TTC in case the leader is decelerating with a constant deceleration (Xie et al., 2019).

For the MTTC and TTCD, the relative speed is not assumed to be constant, but new assumptions on the acceleration and speed of the objects are introduced.

Other time-based proximity indicators include Post-Encroachment Time (PET), which measures the "time between the moment that a vehicle leaves the area of potential collision, i.e., the area in which the paths of the two vehicles intersect, and the other vehicle arrives in the same area" (Mahmud et al., 2017) and

Time Headway (THW). PET can only be calculated when the collision area of the two participants is known. This assumption makes it mostly useful for scenarios with obvious crossing conflicts like intersections.

For distance-based proximity indicators, the Potential Index for Collision with Urgent Deceleration (PICUD) measures the remaining distance between vehicles during an emergency stop (Iida et al., 2001; Uno et al., 2003) and the Proportion of Stopping Distance (PSD) measures the remaining distance to the potential point of collision divided by the minimum acceptable stopping distance (Allen et al., 1978; Guido et al., 2011; Mahmud et al., 2017). These two measures assume that the vehicles will apply the maximum deceleration during emergency situations. This makes them suitable for emergency situations for which these assumption will most likely hold. For non-critical situations, however, the deceleration that the drivers will apply, may vary. More recently, a distance-based measure that assumes "correct" driving behavior has been proposed (Shalev-Shwartz et al., 2017). This measure calculates the minimum safety distance between a follower and its leader, such that no crash occurs if the leader vehicle brakes with a specified deceleration and the follower brakes after a specified reaction time with another specified deceleration. Based on the definition of this measure, it is not suitable for driving situations where the driver does not follow the description of "correct" driving given above.

In terms of indicators relating to performing evasive actions, the Deceleration Rate to Avoid Collision (DRAC) is the most widely used. The DRAC is calculated as the ratio of the difference in speed between a following vehicle and a leading vehicle and their closing time (Almqvist et al., 1991; Mahmud et al., 2017). Another indicator is the Crash Potential Index (CPI), which calculates the probability that a vehicle's DRAC will exceed its Maximum Available Deceleration Rate (MADR) in a given time interval (Cunto and Saccomanno, 2009). The DRAC is not a risk measure on its own, if it is not compared with the braking capacity. This is a limitation and this is why the CPI measure has been developed. Both DRAC and CPI are mostly suitable for a car-following situation and are not suitable for lateral movements (Mahmud et al., 2017).

Wang and Stamatiadis (2014) and Xie et al. (2019) propose an SSM-based metric, i.e., they derive a probability using the TTC and TTCD, respectively, and some assumptions. In particular, Wang and Stamatiadis (2014) assume distributions of the vehicle braking capability and the driver's reaction time, while Xie et al. (2019) assume a distribution of the deceleration rate of the leader. Although these probabilities are suitable for various car-following situations, lane-change conflicts, and crossing conflicts, they are limited because they use the TTC and TTCD, respectively, in their calculations; so these probabilities are undefined when the TTC and TTCD, respectively, are undefined.

Saunier and Sayed (2008) derive a probability based on the TTCs that are estimated using hypothetical trajectories of the different traffic participants. Similar to the approach we present in Section 3, the hypothetical trajectories of the traffic participants are based on a data-driven model (Saunier et al., 2007). In (Saunier and Sayed, 2008), also the probabilities of the different hypothetical trajectories are considered. In order to keep the computation of the metric feasible in real time, the number of hypothetical trajectories to be considered is limited. Altendorfer and Wilkmann (2021) also calculate the probability of a collision. They provide a general framework that can consider any arbitrary prediction model of the trajectories of the traffic participants. To ensure real-time computations, their example considers Gaussian distributions and simplified geometries of the traffic participants, such that they can use numerical integration instead of Monte Carlo simulations.

Shi et al. (2018) use indicators like TIT, CPI, and PSD to measure the effectiveness of risk indicators for predicting crashes. The idea is to use a combination of indicators and thresholds on the indicators to predict whether an interaction may become a crash. This results in new indicators, but they inherit the union of the assumptions of the other indicators. Mullakkal-Babu et al. (2020) propose a probabilistic driving risk field. The method derives the risk a vehicle is exposed to using a kinematic approach with the inclusion of uncertainty in the vehicle's future state. Mullakkal-Babu et al. (2020) define this for an encounter between the ego vehicle and a road obstacle, such as other vehicles or objects. This research shares similar ideas with our proposed method of risk estimation, but Mullakkal-Babu et al. (2020) do not use a data-driven approach to derive the SSM. Furthermore, the future state of the vehicle is estimated with a fixed distribution (i.e., a normal distribution). This limits the application in scenarios where the data may have an entirely different distribution.

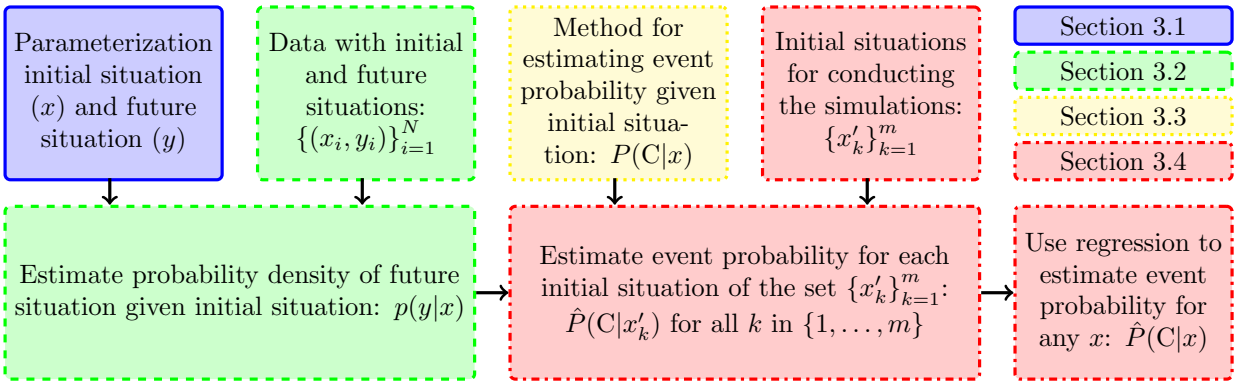



Figure 1: Schematic overview of the PRISMA method and the organization of Section 3. The mathematical symbols are further explained in Sections 3.1 to 3.4.

To estimate crash probabilities based on existing SSMs, the probabilistic approach using the Extreme Value Theory (EVT) has been applied successfully (Songchitruksa and Tarko, 2006; Tarko, 2012; Wang et al., 2021). For example, based on a specific TTC value, EVT can be used to predict the probability of a crash. Using EVT, the crash probability is estimated by assuming the generalized extreme value distribution and fitting the parameters of the distribution using either the "block maxima" approach or the "peak over" approach (Wang et al., 2021). It is also possible to combine multiple SSMs using the EVT. The advantage of EVT is that EVT provides probabilities that are directly linked to historical data and that these probabilities have been used successfully to predict the frequency of crashes (Songchitruksa and Tarko, 2006; Åsljung et al., 2017). Disadvantages of EVT are that it might inherit the assumptions of the SSMs that it uses to estimate the crash probability and that it assumes a fixed distribution of the extreme events, which is only justified if a lot of data is used. Furthermore, as the estimated crash probability is solely based on the fitted distribution, it does not consider potential changes to the driver's behavior (model).

## 3. Probabilistic RISk Measure derivAtion

In this section, we propose the PRISMA method which is a method for deriving a measure that quantifies the risk of a certain event, such as a crash, in a particular situation in which a vehicle - hereafter, the *ego vehicle* - is in and that is applicable for real-time use. The PRISMA method is schematically shown in Figure 1 and consists of four steps:

1. The parameterization of the "initial situation" and the possible "future situations" (Section 3.1);
2. Based on the initial situation, the estimation of the probability (density) for the possible future situations (Section 3.2);
3. The estimation of the probability of the specified event based on the initial and the future situations (Section 3.3); and
4. Local regression in order to speed up the calculations and to make it possible to use the SSM in real time (Section 3.4).

In this article, the following notation is used. To denote a probability function, $P(\cdot)$ is used. A probability density function (PDF) is denoted by $p(\cdot)$. The probability of $a$ given $b$ is denoted by $P(a|b)$. Similarly, a conditional PDF is denoted by $p(\cdot|\cdot)$. To denote the estimation of any of the aforementioned functions, a circumflex is used, e.g, $\hat{P}(a)$ denotes the estimated probability of $a$.

### *3.1. Parameterize initial and future situations*

The first step is to parameterize the initial situation the ego vehicle is in. In other words, the initial situation needs to be described using $n_x$ numbers that are stacked into one vector $x \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$. This

vector contains relevant aspects for determining the risk. As an example, $x$ could contain the speed of the ego vehicle and the distance toward its preceding vehicle. In Section 4, we will consider more examples.

Next to describing the initial situation, the future situation is described using $n_y$ numbers stacked into one vector $y \in \mathcal{Y} \subseteq \mathbb{R}^{n_y}$. Together with $x$, $y$ contains enough information to describe how the relevant future, e.g., the next 5 seconds, around the ego vehicle develops over time. As an example, $y$ could contain the speed for the next 5 seconds of the leading vehicle (if any) that is in front of the ego vehicle. In Section 4, we will consider more examples.

Let C denote an event, e.g., a crash or a near miss, such that the probability of this event is $P(\mathrm{C})$. The goal of our SSM is to estimate the probability of the event C given a particular situation $x$, i.e., $P(\mathrm{C}|x)$. We do this by considering all future situations, $\mathcal{Y}$, and calculating the probability of the event C given each possible value of $y$. Using integration, we obtain $P(\mathrm{C}|x)$:

$$P(\mathrm{C}|x) = \int_{\mathcal{Y}} P(\mathrm{C}|x, y)p(y|x)\,\mathrm{d}y. \tag{1}$$

### *3.2. Estimate $p(y|x)$*

In this section, we propose a method to estimate $p(y|x)$, i.e., the PDF of $y$ given $x$. Using the product rule for probability, we can write:

$$p(y|x) = \frac{p(x, y)}{p(x)} = \frac{p(x, y)}{\int_{\mathcal{Y}} p(x, y)\,\mathrm{d}y}. \tag{2}$$

Thus, it suffices to estimate $p(x, y)$.

Our proposal is to estimate $p(x, y)$ in a data-driven manner. A data-driven approach brings several benefits. First, the estimate automatically adapts to local driving styles and behaviors, which can change from region to region, provided that the data are obtained from the same local traffic. Second, assumptions such as a constant speed of other vehicles, are not needed. For our data-driven approach, let us assume that we have obtained $N$ situations from data. For the $i$-th situation, we denote the initial situation and the future situation by $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$, respectively. The remainder of this subsection describes how we estimate $p(x, y)$ using $\{(x_i, y_i)\}_{i=1}^{N}$.

#### *3.2.1. Kernel density estimation*

We first explain how to estimate $p(x, y)$ if we assume that all $n_x + n_y$ parameters depend on each other. If the shape of the PDF is known, a particular functional form can be fitted to the data, e.g., by estimating the parameters of a distribution by maximizing the likelihood. For example, if it is known that the data $\{(x_i, y_i)\}_{i=1}^{N}$ come from a multivariate normal distribution, it suffices to estimate the mean and the covariance. If, however, the shape is unknown, fitting a particular parametric distribution may lead to very inaccurate results (Chen, 2017). Furthermore, the shape of the estimated PDF might change as more data are acquired. Assuming a functional form of the PDF and fitting the parameters of the PDF to the data may therefore lead to inaccurate fits unless extensive manual tuning is applied.

In the remainder of this work, we assume that the shape of the PDF $p(x, y)$ is unknown a priori. Therefore, we employ a non-parametric approach using Kernel Density Estimation (KDE) (Rosenblatt, 1956; Parzen, 1962) because the shape of the PDF is then automatically computed and KDE is highly flexible regarding the shape of the PDF. Note, however, that the PRISMA method can also work with other non-parametric methods for estimating a PDF (cf. (Durkan et al., 2019; Peerlings et al., 2022)). Using KDE, the estimated PDF becomes:

$$\hat{p}(x, y) = \frac{1}{N}\sum_{i=1}^{N} K_H\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_i \\ y_i \end{bmatrix}\right), \tag{3}$$

where $K_H(\cdot)$ is an appropriate kernel function with an $(n_x + n_y)$-by-$(n_x + n_y)$ symmetric positive definite *bandwidth* or *smoothing* matrix $H$. The choice of the kernel $K_H(\cdot)$ is not as important as the choice of the

bandwidth matrix $H$ (Turlach, 1993). We use the often-used Gaussian kernel (Duong, 2007):

$$K_H(u) = \frac{1}{(2\pi)^{(n_x+n_y)/2}|H|^{1/2}} \exp\left\{-\frac{1}{2}u^\mathsf{T}H^{-1}u\right\}. \tag{4}$$

The bandwidth matrix $H$ controls the width of the kernel, or, in other words, the influence of each data point (i.e., $\begin{bmatrix}x_i^\mathsf{T} & y_i^\mathsf{T}\end{bmatrix}^\mathsf{T}$) on nearby regions (see (Wand and Jones, 1994) for a more extensive explanation of the bandwidth matrix). There are many different ways of estimating the bandwidth matrix, ranging from simple reference rules like, e.g., Silverman's rule of thumb (Silverman, 1986) to more elaborate methods; see (Turlach, 1993; Chiu, 1996; Jones et al., 1996; Bashtannyk and Hyndman, 2001; Zambom and Dias, 2013) for reviews of different bandwidth selection methods.

To estimate $P(\mathrm{C}|x)$ of (1), we need to draw samples from $\hat{p}(y|x)$. Drawing samples from the estimated PDF in (3) is straightforward: two random numbers are drawn, one to choose a random generator kernel out of the $N$ kernels that are used to construct the KDE, and one random number from that kernel. Sampling from $\hat{p}(y|x)$ works similarly, but instead of using an equal probability for each random generator kernel to be selected, different probabilities are used based on $x$. For more information on sampling from a conditional PDF obtained using KDE, see (Holmes et al., 2007; de Gelder et al., 2021).

*3.2.2. Assuming independence*

Due to the curse of dimensionality (Scott, 2015), estimating $p(x, y)$ with one KDE according to (3) becomes inaccurate if $n_x + n_y$ becomes large. One option to avoid this curse of dimensionality is to assume that one or more parameters are independent of the other parameters. E.g., suppose that $y^\mathsf{T} = \begin{bmatrix}\bar{y}^\mathsf{T} & \tilde{y}^\mathsf{T}\end{bmatrix}$, such that $\tilde{y}$ is independent of $x$ and $\bar{y}$. Then we can write

$$p(x, y) = p(x, \bar{y}, \tilde{y}) = p(x, \bar{y})p(\tilde{y}). \tag{5}$$

In this case, we would need to estimate $p(x, \bar{y})$ and $p(\tilde{y})$, which can be done in a similar manner as presented in Section 3.2.1. Because these two PDFs have fewer variables than $p(x, y)$, the two estimated PDFs will suffer less from the curse of dimensionality (Scott, 2015).

Another option is to model $p(y|x)$ as a cascade of conditional probabilities. For example, using the partitioning $y^\mathsf{T} = \begin{bmatrix}\bar{y}^\mathsf{T} & \tilde{y}^\mathsf{T}\end{bmatrix}$, $p(x|y)$ can be approximated using two conditional densities:

$$p(y|x) = p(\bar{y}, \tilde{y}|x) = p(\bar{y}|\tilde{y}, x)p(\tilde{y}|x) \approx p(\bar{y}|\tilde{y})p(\tilde{y}|x). \tag{6}$$

This approximation is valid if $\bar{y}$ and $x$ are *conditionally independent given* $\tilde{y}$ (Nagler and Czado, 2016). The same partitioning can be applied to $p(\bar{y}|\tilde{y})$ and $p(\tilde{y}|x)$ until only two-dimensional PDFs need to be estimated. Although this will lead to larger approximation errors, the lower-dimensional PDFs can be estimated more accurately. For more information on this approach, we refer the reader to (Aas et al., 2009; Nagler and Czado, 2016). Note that when relying on the assumption of independence or the assumption of conditional independence, these assumptions should be justified, e.g., through the use of some statistical tests like the Pearson's chi-squared test.

*3.2.3. Reduce number of parameters using singular value decomposition*

Another way to avoid the curse of dimensionality is to use a Singular Value Decomposition (SVD) (Golub and Van Loan, 2013) to reduce the number of parameters. In the field of machine learning, Principal Component Analysis (PCA) is commonly used for dimensionality reduction (Abdi and Williams, 2010; Hasan and Abdulazeez, 2021) and PCA uses the SVD. With an SVD, the parameters $x$ and $y$ are transformed into a lower-dimensional vector of parameters in such a way that the reduced vector of parameters describes as much of the variation as possible. To do this, an SVD is made of the matrix that contains all $N$ observed situations:

$$\begin{bmatrix}x_1 - \mu_x & \cdots & x_N - \mu_x \\ y_1 - \mu_y & \cdots & y_N - \mu_y\end{bmatrix} = U\Sigma V^\mathsf{T}. \tag{7}$$

Here, $\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $\mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i$. The matrices $U \in \mathbb{R}^{(n_x+n_y)\times(n_x+n_y)}$ and $V \in \mathbb{R}^{N\times N}$ are orthonormal, i.e., $U^{-1} = U^{\mathsf{T}}$ and $V^{-1} = V^{\mathsf{T}}$. Moreover, $\Sigma \in \mathbb{R}^{(n_x+n_y)\times N}$ has only zeros except at the diagonal: the $(j,j)$-th element is $\sigma_j$, $j \in \{1,\ldots,\bar{N}\}$ with $\bar{N} = \min(n_x+n_y, N)$, such that

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\bar{N}} \geq 0. \tag{8}$$

Because these so-called singular values are in decreasing order, we can approximate $x$ and $y$ by setting $\sigma_j = 0$ for $j > d$ with $d$ chosen[2] such that $n_x < d < n_x + n_y$:

$$\begin{bmatrix} x_i - \mu_x \\ y_i - \mu_y \end{bmatrix} = \sum_{j=1}^{\bar{N}} \sigma_j v_{ij} u_j \approx \sum_{j=1}^{d} \sigma_j v_{ij} u_j, = \begin{bmatrix} \bar{U}_1 \\ \bar{U}_2 \end{bmatrix} \bar{\Sigma} \bar{v}_i, \tag{9}$$

where $v_{ij}$ is the $(i,j)$-th element of $V$ and $u_j$ is the $j$-th column of $U$. Moreover, $\bar{U}_1$ is the $n_x$-by-$d$ upper left submatrix of $U$, $\bar{U}_2$ is the $n_y$-by-$d$ lower left submatrix $U$, $\bar{\Sigma} \in \mathbb{R}^{d\times d}$ is the diagonal matrix with the first $d$ singular values on its diagonal, and $\bar{v}_i^{\mathsf{T}} = \begin{bmatrix} v_{i1} & \cdots & v_{id} \end{bmatrix}$. Thus, with $\mu_x$, $\mu_y$, $\bar{U}_1$, $\bar{U}_2$, and $\bar{\Sigma}$, the $(n_x+n_y)$-dimensional vector $\begin{bmatrix} x_i^{\mathsf{T}} & y_i^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$ is approximated using the $d$-dimensional vector $\bar{v}_i$.

Instead of estimating the PDF of $\begin{bmatrix} x_i^{\mathsf{T}} & y_i^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$, we now estimate the PDF of $\bar{v}_i$ using KDE as described in Section 3.2.1. Note that the choice of $d$ includes a trade-off. Choosing $d$ too small results in too much loss of detail, while choosing $d$ too large will give accuracy problems when estimating the PDF of the new parameters. For more information on choosing an appropriate value of $d$, we refer the reader to (de Gelder et al., 2022). To sample from $\hat{p}(y|x)$, we can sample from the estimated distribution of $\bar{v}_i$. Because (9) is a linear mapping, the sample $\bar{v}$ that is drawn from the estimated distribution of $\bar{v}_i$ is subject to a linear constraint:

$$\bar{U}_1 \bar{\Sigma} \bar{v} = x - \mu_x. \tag{10}$$

In (de Gelder et al., 2021), an algorithm is provided for sampling from a KDE with a Gaussian kernel of (4) such that the resulting samples are subject to a linear constraint such as (10).

### *3.3. Estimate $P(\mathrm{C}|x)$ using a Monte Carlo simulation*

Monte Carlo simulations are used to estimate $P(\mathrm{C}|x)$, i.e., the probability of an event C given the initial situation described by $x$. The details of the simulation depend on the actual application. For example, if the goal of our SSM is to evaluate the risk that a human-driven vehicle collides, the simulation should involve a human driving behavior model. On the other hand, if the goal is to evaluate the risk of a crash when an ADS is controlling the vehicle, the simulation should include the model of this ADS.

A straightforward way to compute $P(\mathrm{C}|x)$ is to repeat a certain number of simulations with the same $x$ and count the number of simulations that result in the event C. If $N_{\mathrm{sim}}$ denotes the number of simulations and $N_{\mathrm{C}}$ is the number of events C, then $P(\mathrm{C}|x)$ could be estimated using

$$\hat{P}(\mathrm{C}|x) = \frac{N_{\mathrm{C}}}{N_{\mathrm{sim}}}. \tag{11}$$

An important choice for estimating $P(\mathrm{C}|x)$ is the number of simulations, $N_{\mathrm{sim}}$. One approach is to keep increasing $N_{\mathrm{sim}}$ until there is enough confidence in the estimation of (11). For example, the Clopper-Pearson interval (Clopper and Pearson, 1934) or the Wilson score interval (Wilson, 1927) can be used to determine the confidence of the estimation of (11). A disadvantage of (11) is that only the fact whether the event C occurred or not is used, while the simulation provides more information, such as the minimum distance between two objects or the impact speed in case of a crash. Therefore, we provide an alternative approach to estimate $P(\mathrm{C}|x)$.

[2] We have $d < n_x + n_y$, such that the dimension is reduced (from $n_x + n_y$ to $d$) and we have $d > n_x$, such that the number of linear constraints in (10) ($n_x$) is smaller than the number of variables ($d$).

For the alternative approach, let us assume that one simulation run provides more information than just the fact that the event C occurred or not. Let $z \in \mathbb{R}^{n_z}$ be a continuous variable representing the result of a simulation run and let $\mathcal{Z}_\mathrm{C}$ denote the set of possible simulation results in which the event C occurred. Thus, $z \in \mathcal{Z}_\mathrm{C}$ if and only if the simulation results in the event C. We assume $\mathcal{Z}_\mathrm{C}$ is known; see, e.g., the example in Section 4.1.2. Therefore, we have

$$P(\mathrm{C}|x) = P(z \in \mathcal{Z}_\mathrm{C}|x) = \int_{\mathcal{Z}_\mathrm{C}} p(z|x)\,\mathrm{d}z. \tag{12}$$

Similar as with the estimation of $p(x, y)$ in Section 3.2, we employ KDE to estimate $p(z|x)$:

$$\hat{p}(z|x) = \frac{1}{N_\mathrm{sim}} \sum_{j=1}^{N_\mathrm{sim}} K_{H_z}(z_j - z), \tag{13}$$

where $z_j$ denotes the result of the $j$-th simulation and $H_z$ denotes an appropriate bandwidth matrix. The kernel function $K_{H_z}(\cdot)$ is similarly defined as (4). We can now estimate $P(\mathrm{C}|x)$ by substituting $\hat{p}(z|x)$ of (13) for $p(z|x)$:

$$\hat{P}(\mathrm{C}|x) = \hat{P}(z \in \mathcal{Z}_\mathrm{C}|x) = \int_{\mathcal{Z}_\mathrm{C}} \hat{p}(z|x)\,\mathrm{d}z = \frac{1}{N_\mathrm{sim}} \sum_{j=1}^{N_\mathrm{sim}} \int_{\mathcal{Z}_\mathrm{C}} K_{H_z}(z_j - z)\,\mathrm{d}z. \tag{14}$$

Similar as with (11), we need to choose the number of simulations $N_\mathrm{sim}$. Our proposal is to keep increasing $N_\mathrm{sim}$ until the uncertainty of the estimated probability $\hat{P}(z \in \mathcal{Z}_\mathrm{C}|x)$ is below a certain threshold. As a measure for the uncertainty of the estimated probability $\hat{P}(z \in \mathcal{Z}_\mathrm{C}|x)$, we use the variance of $\hat{P}(z \in \mathcal{Z}_\mathrm{C}|x)$. Thus, we keep increasing $N_\mathrm{sim}$ until the variance of $\hat{P}(z \in \mathcal{Z}_\mathrm{C}|x)$ is below a threshold $\epsilon > 0$. The variance follows from (Nadaraya, 1964):

$$\mathrm{Var}\Big[\hat{P}(z \in \mathcal{Z}_\mathrm{C}|x)\Big] = \frac{P(z \in \mathcal{Z}_\mathrm{C}|x)(1 - P(z \in \mathcal{Z}_\mathrm{C}|x))}{N_\mathrm{sim}}. \tag{15}$$

Because $P(z \in \mathcal{Z}_\mathrm{C}|x)$ is unknown, we cannot directly use (15). Instead, we substitute the estimated counterpart of (14) for $P(z \in \mathcal{Z}_\mathrm{C}|x)$. Thus, $N_\mathrm{sim}$ is increased until the following condition is met:

$$\frac{\hat{P}(z \in \mathcal{Z}_\mathrm{C}|x)\Big(1 - \hat{P}(z \in \mathcal{Z}_\mathrm{C}|x)\Big)}{N_\mathrm{sim}} < \epsilon. \tag{16}$$

*3.4. Regression for real-time estimation of $P(\mathrm{C}|x)$*

To evaluate the risk measure during real-time operation of the ego vehicle, the expression of (14) is problematic, because it would require $N_\mathrm{sim}$ simulations. Even if the calculation is accelerated using a technique such as importance sampling, it might take too long. Therefore, we propose to evaluate (14) only for some fixed $\{x'_k\}_{k=1}^m$. Next, regression is used to estimate (14). One option is to choose a parametric model, e.g., a logistic model, and estimate the parameters of the model using $\Big\{\Big(x'_k, \hat{P}(\mathrm{C}|x'_k)\Big)\Big\}_{k=1}^m$. Up to our knowledge, however, there is no good reason to assume a particular parametric model, so we use a non-parametric regression technique to estimate (14). More specifically, we use the Nadaraya-Watson (NW) kernel estimator (Wasserman, 2006), because it automatically smooths the data (as is demonstrated in Section 4.1.3) and the approximation is guaranteed to give a number between 0 and 1, also when extrapolating the data. The NW kernel estimator is given by:

$$\hat{P}(\mathrm{C}|x) \approx \frac{\sum_{k=1}^m K_{H_\mathrm{NW}}(x - x'_k)\hat{P}(\mathrm{C}|x'_k)}{\sum_{k=1}^m K_{H_\mathrm{NW}}(x - x'_k)}. \tag{17}$$

Here, $\hat{P}(\mathrm{C}|x'_k)$ is based on (14) and $K_{H_{\mathrm{NW}}}(\cdot)$ represents the Gaussian kernel given by (4). Two important choices have to be made: The choice of $\{x'_k\}_{k=1}^m$ for which to evaluate (14) and the choice of the bandwidth matrix $H_{\mathrm{NW}}$. We suggest to base the design points $\{x'_k\}_{k=1}^m$ on the data that is used to estimate $p(y|x)$ in Section 3.2, i.e., $\{x_i\}_{i=1}^N$, such that all $x_i$ have at least one design point $x'_k$ nearby. In other words, $\{x'_k\}_{k=1}^m$ is chosen such that

$$\min_k (x_i - x'_k)^{\mathsf{T}} W (x_i - x'_k) \leq 1, \quad \forall i \in \{1, \ldots, N\}, \tag{18}$$

where $W$ denotes a weighting matrix. Note that if $W$ is the identity matrix, then (18) calculates the minimum squared Euclidean distance. In general, $W$ is a diagonal matrix. Choosing the diagonal elements of $W$ is a trade-off; if the elements are too large, then too many details are lost in the approximation of (17); if the elements are too small, it takes too long to evaluate (14) $m$ times, as $m$ increases for lower diagonal elements of $W$. The bandwidth matrix $H_{\mathrm{NW}}$ might be based on $W$, e.g., $H_{\mathrm{NW}} = W^{-1}$. Alternatively, $H_{\mathrm{NW}}$ might be based on the measurement uncertainty of $x$ if this measurement uncertainty is significant, where a larger $H_{\mathrm{NW}}$ applies in case of a larger measurement uncertainty of $x$. Note that if $H_{\mathrm{NW}}$ is a diagonal matrix with positive values on the diagonal that are close to zero, then the NW kernel estimation of (17) acts like nearest-neighbor interpolation.

## 4. Case study

In the first part of the case study, we illustrate that the PRISMA method generalizes the SSM proposed by Wang and Stamatiadis (2014). Here, we also demonstrate the effect of $\epsilon$ on the accuracy of the SSM derived by the PRISMA method and we show the difference between $\hat{P}(\mathrm{C}|x)$ of (14) and the approximation of $\hat{P}(\mathrm{C}|x)$ using the NW kernel estimator of (17). In Section 4.2, we demonstrate how the PRISMA method can be used to create a new SSM that calculates the risk of a crash in a longitudinal interaction between two vehicles. The SSM derived in Section 4.2 is qualitatively analyzed in Section 4.3. To also quantitatively analyze SSMs, Mullakkal-Babu et al. (2017) proposed a benchmarking method which we apply in Section 4.4.

### *4.1. Comparison with Wang and Stamatiadis' measure*

Wang and Stamatiadis (2014) provide an SSM, which we denote by WS, that calculates the probability of a crash under certain assumptions. We first explain how WS is calculated. Next, Section 4.1.2 shows how to estimate this SSM using our method. In Section 4.1.3, we illustrate the results of both.

#### *4.1.1. Measure of Wang and Stamatiadis*

The SSM WS calculates the probability of a crash of the ego vehicle and the leading vehicle, where the ego vehicle is following an initially slower driving leading vehicle. The SSM WS is based on the following assumptions (Wang and Stamatiadis, 2014):

- the leading vehicle keeps a constant speed;
- the (driver of the) ego vehicle starts to brake after its reaction time, denoted by $t_{\mathrm{r}}$;
- based on (Green, 2000), the reaction time $t_{\mathrm{r}}$ is distributed according to a log-normal distribution, such that the mean is 0.92 s and the standard deviation is 0.28 s;
- when the ego vehicle reacts, it brakes with its MADR, denoted by $a_{\mathrm{MADR}}$; and
- $a_{\mathrm{MADR}}$ is distributed according to a truncated normal distribution with a mean of $9.7\,\mathrm{m/s^2}$, a standard deviation of $1.3\,\mathrm{m/s^2}$, a lower bound of $L = 4.2\,\mathrm{m/s^2}$ (Cunto, 2008), and an upper bound of $U = 12.7\,\mathrm{m/s^2}$ (Cunto, 2008).

To calculate WS at a given time $t$, the speed difference between the ego vehicle and the leading vehicle, $\Delta v(t)$, and the TTC, $t_{\mathrm{TTC}}(t)$, are used. Note that $t_{\mathrm{TTC}}(t)$ is the ratio of the gap, $g(t)$, between the ego vehicle and the leading vehicle and $\Delta v(t)$. If $\Delta v(t) \leq 0$, then the ego vehicle drives slower and there is no

risk of a future crash according to Wang and Stamatiadis (2014), so $\mathrm{WS}(t) = 0$. Given $a_{\mathrm{MADR}}$, the driver of the ego vehicle needs to react within

$$t_{\max}(t) = t_{\mathrm{TTC}}(t) - \frac{\Delta v(t)}{2a_{\mathrm{MADR}}} \tag{19}$$

in order to avoid a crash. Using the distributions of $a_{\mathrm{MADR}}$ and $t_{\mathrm{r}}$, we can calculate the probability that this is the case, resulting in:

$$\mathrm{WS}(t) = \begin{cases} 0 & \text{if} \quad \Delta v(t) \leq 0 \\ \int_{\hat{L}}^{U} \int_{0}^{t_{\max}(t)} p(t_{\mathrm{r}}) p(a_{\mathrm{MADR}}) \, \mathrm{d}t_{\mathrm{r}} \, \mathrm{d}a_{\mathrm{MADR}} & \text{if} \quad \Delta v(t) > 0 \wedge \frac{\Delta v(t)}{2t_{\mathrm{TTC}}(t)} < U \,, \\ 1 & \text{otherwise} \end{cases} \tag{20}$$

with $\hat{L} = \max\left(L, \frac{\Delta v(t)}{2t_{\mathrm{TTC}}(t)}\right)$, $p(t_{\mathrm{r}})$ is the log-normal probability density of $t_{\mathrm{r}}$, and $p(a_{\mathrm{MADR}})$ is the truncated normal probability density of $a_{\mathrm{MADR}}$.

*4.1.2. Replicating Wang and Stamatiadis' measure*

Because WS is based on $\Delta v(t)$ and $t_{\mathrm{TTC}}(t)$, these two variables are also used by the PRISMA method to describe the initial situation:

$$x^{\mathsf{T}}(t) = \begin{bmatrix} \Delta v(t) & t_{\mathrm{TTC}}(t) \end{bmatrix}. \tag{21}$$

The leading vehicle is assumed to have a constant speed, so $x(t)$ of (21) already describes the future situation of the leading vehicle. Therefore, there is no need to estimate $p(y|x)$. At the start of each simulation run, the driver of the ego vehicle is not braking. After the reaction time $t_{\mathrm{r}}$, the driver starts braking with $a_{\mathrm{MADR}}$. The random parameters $t_{\mathrm{r}}$ and $a_{\mathrm{MADR}}$ are similarly distributed as described in Section 4.1.1.

Since we are interested in the probability of a crash, the event C denotes a crash. A simulation run ends if either the ego vehicle and the leading vehicle are colliding or if the gap between the ego vehicle and the leading vehicle is not decreasing. Depending on the reason for a simulation run to end, we consider the following result:

- If the ego vehicle and the leading vehicle are colliding, we are interested in the "severity" of the crash. This is expressed using the speed difference: $v_{\mathrm{l}}(t_{\mathrm{end}}) - v_{\mathrm{e}}(t_{\mathrm{end}})$, where $t_{\mathrm{end}}$ denotes the final time of the simulation run.
- If there is no crash, we are interested in how close the two vehicles came. Therefore, the minimum gap is used, which is $g(t_{\mathrm{end}})$.

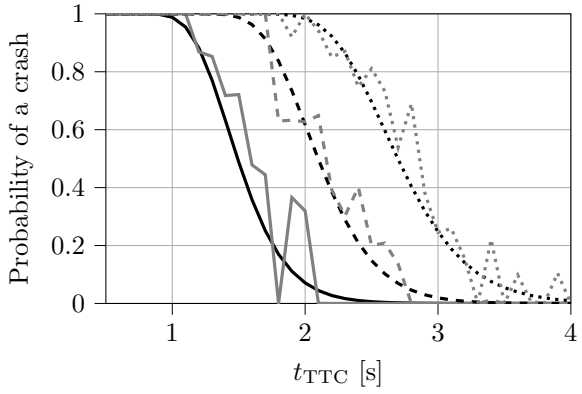
Thus, we have:

$$z = \begin{cases} v_{\mathrm{l}}(t_{\mathrm{end}}) - v_{\mathrm{e}}(t_{\mathrm{end}}) & \text{if crash} \\ g(t_{\mathrm{end}}) & \text{otherwise} \end{cases}. \tag{22}$$

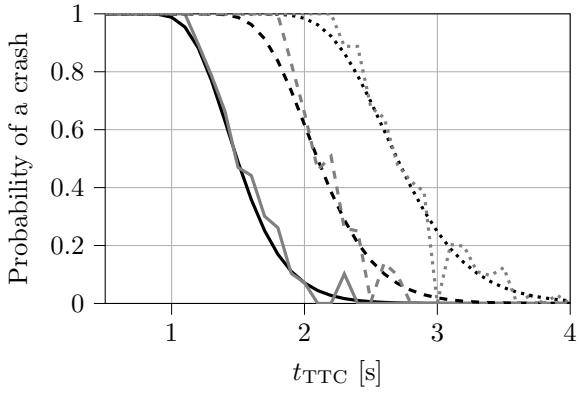
Clearly, $z \leq 0$ indicates a crash, so $\mathcal{Z}_{\mathrm{C}} = (-\infty, 0]$. The minimum number of simulations to estimate $P(\mathrm{C}|x)$ is set to 10. The number of simulations is further increased until the condition in (16) with $\epsilon = 0.2$ or $\epsilon = 0.02$ is met. For the design points $\{x'_k\}_{k=1}^{m}$, we use a rectangular grid with $\Delta v$ ranging from 0 m/s till 40 m/s with steps of 2 m/s and $t_{\mathrm{TTC}}$ ranging from 0.5 s till 4 s in steps of 0.1 s. Thus, $m = 21 \cdot 36 = 756$. For $H_{\mathrm{NW}}$, a diagonal matrix is chosen with the diagonal elements corresponding to the square of the step size of the grid, i.e., $4\,\mathrm{m}^2/\mathrm{s}^2$ and $0.01\,\mathrm{s}^2$.

*4.1.3. Comparison*

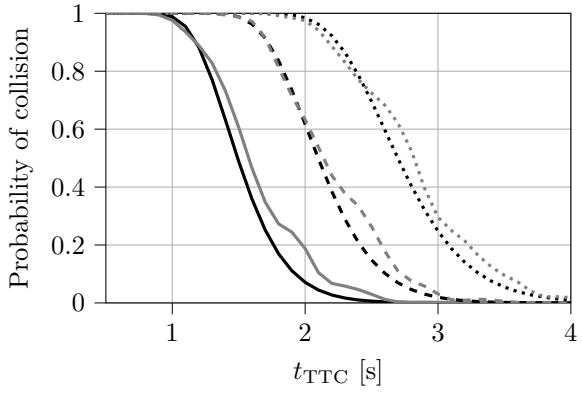
Figure 2 shows the results of the comparison between the measure of Wang and Stamatiadis (2014) and the measure derived using the PRISMA method described in Section 3. The black lines in Figure 2 denote WS of (20). These lines show that for lower values of $t_{\mathrm{TTC}}$, WS increases. Also, for increasing values of $\Delta v$ (solid, dashed, and dotted lines), the risk measure WS increases. Both these observations match the intuition that a lower TTC and a higher speed difference are less safe.



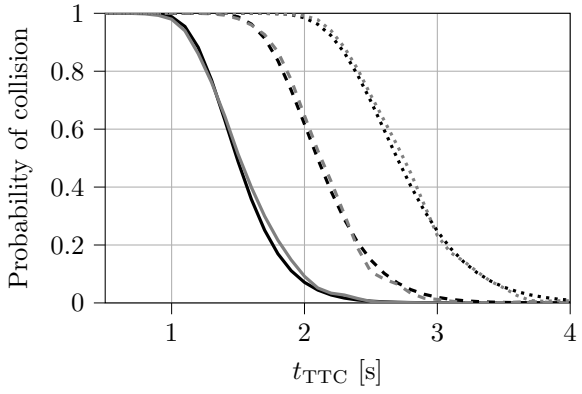

(a) Comparison of WS of (20) (black) and $\hat{P}(\mathrm{C}|x)$ of (14) (gray) with $\epsilon = 0.2$.




(b) Comparison of WS of (20) (black) and $\hat{P}(\mathrm{C}|x)$ of (14) (gray) with $\epsilon = 0.02$.




(c) Comparison of WS of (20) (black) and the approximation of $\hat{P}(\mathrm{C}|x)$ using (17) (gray) with $\epsilon = 0.2$.




(d) Comparison of WS of (20) (black) and the approximation of $\hat{P}(\mathrm{C}|x)$ using (17) (gray) with $\epsilon = 0.02$.

Figure 2: Comparison of WS of (20) (black lines) and $\hat{P}(\mathrm{C}|x)$ of (14) (a,b) or the approximation of $\hat{P}(\mathrm{C}|x)$ using (17) (c,d) (gray lines) for a speed difference of $\Delta v = 10\,\mathrm{m/s}$ (solid lines), $\Delta v = 20\,\mathrm{m/s}$ (dashed lines), and $\Delta v = 30\,\mathrm{m/s}$ (dotted lines). Here, $\hat{P}(\mathrm{C}|x)$ is based on the same underlying assumptions as WS, see Section 4.1.1. The influence of the parameter $\epsilon$, which determines the number of simulations to estimate $P(\mathrm{C}|x)$, is illustrated by using different values.

The gray lines in Figure 2 denote $\hat{P}(\mathrm{C}|x)$. In Figures 2a and 2b, $\hat{P}(\mathrm{C}|x)$ of (14) is used and the gray lines in Figures 2c and 2d represent the approximation of $\hat{P}(\mathrm{C}|x)$ using the NW kernel estimator of (17). Figure 2 illustrates that $\hat{P}(\mathrm{C}|x)$ follows the same trend as WS. Figure 2 also illustrates the effect of the choice of the threshold $\epsilon$. In general, for a lower value of $\epsilon$, the number of simulations $N_{\mathrm{sim}}$ used in (13) is higher. As a result, it can be expected that the estimation $\hat{P}(\mathrm{C}|x)$ is closer to $P(\mathrm{C}|x)$ (cf. (15)). A comparison of Figure 2a ($\epsilon = 0.2$) and Figure 2b ($\epsilon = 0.02$) demonstrates this effect. Figure 2 further illustrates the regression using the NW kernel estimator: the gray lines in Figures 2c and 2d can be seen as smoothed versions of the gray lines in Figures 2a and 2b, respectively.

### *4.2. Developing an SSM for longitudinal interactions*

To further illustrate the PRISMA method, we apply it to derive an SSM that calculates the risk of a crash in a longitudinal interaction between two vehicles. The SSM is based on the NGSIM data set (Alexiadis et al., 2004). The NGSIM data set contains vehicle trajectories obtained from video footage of cameras that were located at several motorways in the U.S.A. The derived SSM estimates the risk of a crash of the ego vehicle with its leading vehicle. To describe the initial situation at time $t$, $n_x = 4$ parameters are used:

- the speed of the leading vehicle ($v_{\mathrm{l}}(t)$);

- the acceleration of the leading vehicle ($a_\mathrm{l}(t)$);
- the speed of the ego vehicle ($v_\mathrm{e}(t)$); and
- the log of the gap between the leading vehicle and the ego vehicle[3] $\log g(t)$.

Thus, we have:

$$x^\mathsf{T}(t) = \begin{bmatrix} v_\mathrm{l}(t) & a_\mathrm{l}(t) & v_\mathrm{e}(t) & \log g(t) \end{bmatrix}. \tag{23}$$

The speed of the leading vehicle at $n_\mathrm{h} = 50$ instances, each $\Delta t = 0.1\,\mathrm{s}$ apart, describes the future situation:

$$y^\mathsf{T}(t) = \begin{bmatrix} v_\mathrm{l}(t+\Delta t) & \cdots & v_\mathrm{l}(t+n_\mathrm{h}\Delta t) \end{bmatrix}. \tag{24}$$

It is assumed that $y(t)$ depends on $v_\mathrm{l}(t)$ and $a_\mathrm{l}(t)$. To model this dependency with a single kernel density estimator would give us a PDF with $n_\mathrm{h}+2$ dimensions. To reduce the dimensionality, we use an SVD as described in Section 3.2.3 with[4] $d = 4$. In total, 18182 longitudinal interactions between two vehicles have been analyzed. Here, a longitudinal interaction between a leading and following vehicle within the same lane always begins when the THW is less than or equal to 2 s, or when $g$ is less than or equal to 20 m. The interaction ceases when both the THW exceeds 4 s and $g$ becomes greater than 40 m, or when both vehicles are no longer the closest pair in the same lane. This scheme is commonly referred to as a hysteresis loop. For each second of an interaction, we extract an "initial situation" $x_i$ according to (23) and a corresponding "future situation" $y_i$ according to (24). This leads to $N = 469453$ data points. Based on Silverman's rule of thumb (Silverman, 1986), we use a bandwidth matrix $H = h^2 I_4$ for the KDE with $h \approx 0.186$ and $I_4$ denoting the 4-by-4 identity matrix.

To demonstrate the sampling from the estimated density of the reduced parameter vector subject to a linear constraint such as (10), the plots in Figure 3 show 50 different future situations in the form of (24). Figure 3a assumes an initial situation with $v_\mathrm{l} = 15\,\mathrm{m/s}$ and $a_\mathrm{l} = 1\,\mathrm{m/s^2}$ and Figure 3b assumes an initial situation with $v_\mathrm{l} = 15\,\mathrm{m/s}$ and $a_\mathrm{l} = -1\,\mathrm{m/s^2}$. Note that the same PDF is used to produce the lines in Figure 3; the only difference between Figure 3a and Figure 3b is a different linear constraint (based on $v_\mathrm{l}$ and $a_\mathrm{l}$) on the generated samples. In case a simulation run is longer than 5 s, the speed of the leading vehicle is assumed to remain constant after these 5 s. Note that a simulation run is rarely longer than 5 s, so this assumption does not have a significant effect on the results.

To estimate $P(\mathrm{C}|x)$ (Section 3.3), we use the Intelligent Driver Model Plus (IDM+) (Schakel et al., 2010) for modeling the ego vehicle driver behavior and response. In addition to IDM+, we assume that the driver has a reaction time that is similarly distributed as $t_\mathrm{r}$ in Section 4.1.1 and that the MADR is similarly distributed as $a_\mathrm{MADR}$ in Section 4.1.1. The simulation result $z$ is defined according to (22). The minimum number of simulations to estimate $P(\mathrm{C}|x)$ is set to 10 and this number is further increased until the condition in (16) with $\epsilon = 0.1$ is met.

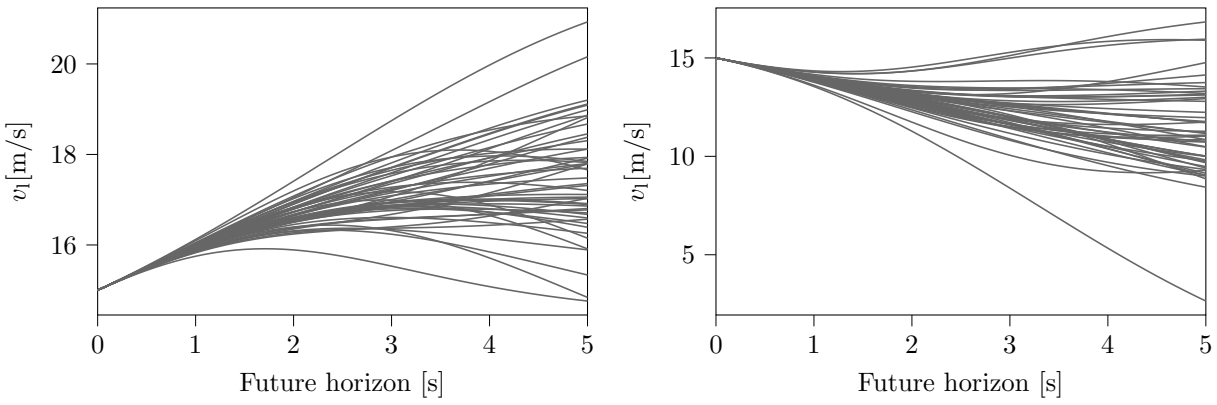
To calculate $P(\mathrm{C}|x)$ using (17), we create a grid of points $\{x'_k\}_{k=1}^m$ using the method explained in Section 3.4. For $W$, we use a diagonal matrix with diagonal elements: 0.25, 4, 0.25, and 0.25, which is a trade-off between keeping many points such that the estimation in (17) is accurate while also keeping the total number of points for which $P(\mathrm{C}|x)$ is estimated manageable. With this choice of $W$, we have $m = 10129$. For the regression of (17), we use $H_\mathrm{NW} = W^{-1}$.

### *4.3. Analyzing the SSMs for longitudinal interactions*

The heat maps in Figure 4 show how the developed SSM depends on the input variables $v_\mathrm{l}$ and $g$. The other two parameters, $v_\mathrm{e}$ and $a_\mathrm{l}$, are fixed for each heat map. The heat maps show that the estimated crash probability is practically 0 if both $v_\mathrm{l}$ and $g$ are large. This seems reasonable, because in that case, the ego vehicle is at a safe distance from its leading vehicle while the approaching speed is small. In addition, for a

[3] Note that the log is used, such that there are, relatively speaking, more simulations performed with a small initial gap, cf. (18).

[4] Note that because we assume that $y(t)$ depends on 2 parameters of $x(t)$, i.e., $v_\mathrm{l}(t)$ and $a_\mathrm{l}(t)$, we need to choose $d$ such that $2 < d < n_\mathrm{h} + 2$.




(a) Initial situation: $v_l = 15\,\mathrm{m/s}$ and $a_l = 1\,\mathrm{m/s^2}$.

(b) Initial situation: $v_l = 15\,\mathrm{m/s}$ and $a_l = -1\,\mathrm{m/s^2}$.

Figure 3: 50 potential future situations samples from the KDE that is constructed using data from the NGSIM data set.

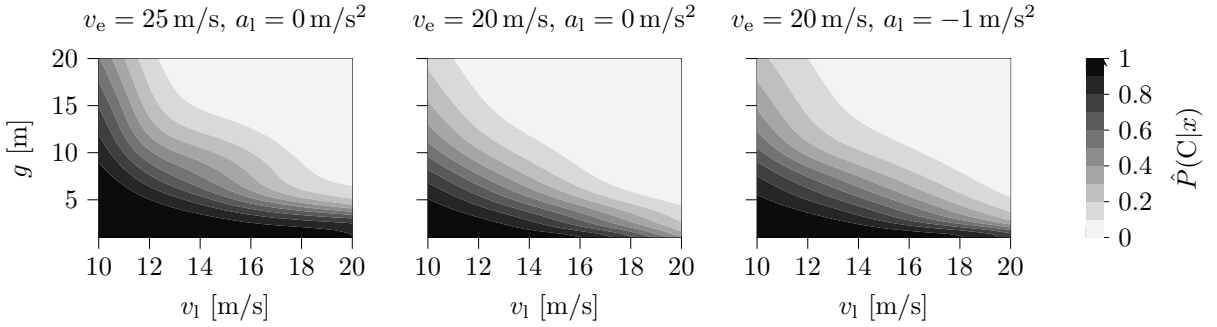



Figure 4: Heat maps of the SSM described in Section 4.2 as a function of the speed of the leading vehicle ($v_l$) and the gap between the ego vehicle and the leading vehicle ($g$). For each heat map, the other two input parameters are fixed at $v_e = 25\,\mathrm{m/s}$ (left) or $v_e = 20\,\mathrm{m/s}$ (center and right) and $a_l = 0\,\mathrm{m/s^2}$ (left and center) or $a_l = -1\,\mathrm{m/s^2}$ (right). The estimated crash probability ranges from 0 (white) to 1 (black).

fixed $v_e$, we see that the crash risk increases as the difference in speed increases, as is expected. The same applies for a decreasing distance between the two vehicles. For small values of $v_l$ and $g$, the estimated crash probability is practically 1. The left and center heat maps of Figure 4 show that for a higher speed of the ego vehicle, the crash probability is estimated to be higher. Similarly, the right and center heat maps of Figure 4 show that for a lower initial acceleration of the leading vehicle, the crash probability is estimated to be higher.

In Figure 5, the evaluations of the measure described in Section 4.2 are shown for 3 different scenarios. Each of the 3 scenarios considers an ego vehicle and a leading vehicle driving in front of the ego vehicle. Both vehicles are driving in the same direction and in the same lane. For comparison, the right plots also include the evaluations of WS of (20).

The first scenario in Figure 5 (top row) shows a scenario in which the leading vehicle initially drives with a speed of 20 m/s. The leading vehicle starts to decelerate after 3 s toward a speed of 10 m/s with an average deceleration of $3\,\mathrm{m/s^2}$. The ego vehicle initially drives with a speed of 24 m/s at a distance of 40 m from the leading vehicle. The ego vehicle starts decelerating after 2 s toward a speed of 8 m/s within 4 s. It takes 4 s more to reach the speed of the leading vehicle. Because the ego vehicle always maintains a relatively large distance toward the leading vehicle, both SSMs do not qualify this scenario as risky, considering the estimated crash probability that stays below 0.1.

The second scenario in Figure 5 (center row) differs from the first scenario in that the ego vehicle starts

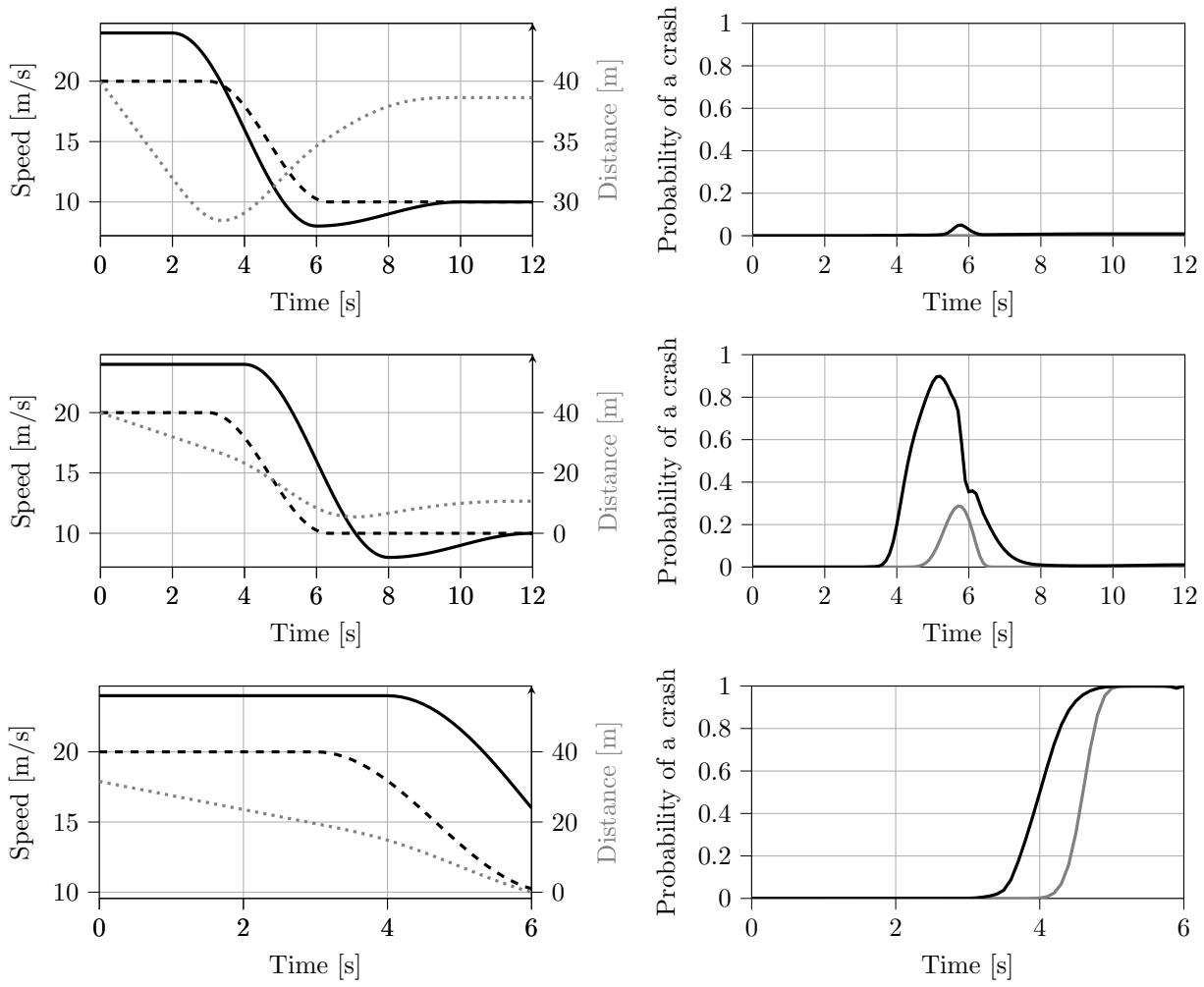



Figure 5: Demonstration of SSMs for 3 hypothetical scenarios. The left plots show the speeds of the ego vehicle (solid black line) and leading vehicle (dashed black line) and the distance between the ego vehicle and the leading vehicle (dotted gray line, scale on the right of the plot). The right plots show the estimated probability of a crash corresponding to the 3 scenarios according to the SSM explained in Section 4.2 (black lines) and the SSM of Wang and Stamatiadis (2014) explained in Section 4.1.1 (gray lines).

to decelerate 2 s later. As a result, the ego vehicle approaches the leading vehicle up to a distance of 5.4 m. According to $\hat{P}(\mathrm{C}|x)$ from Section 4.2 (black line in the right plot of Figure 5), the probability of a crash reaches almost 1, indicating that around that time, the risk of a crash is high. The local minimum of $\hat{P}(\mathrm{C}|x)$ at around 6 s illustrates the effect of the numerical approximation of $P(\mathrm{C}|x)$. Because we have used $\epsilon = 0.1 > 0$, the resulting estimation may have an error. When lowering the threshold $\epsilon$, the resulting $\hat{P}(\mathrm{C}|x)$ in the center right plot in Figure 5 will be smoother. This goes, however, at the cost of an increased number of simulations[5].

The third scenario in Figure 5 (bottom row) differs from the second scenario in that the initial distance between the ego vehicle and the leading vehicle is 31.5 m instead of 40 m. As a result, the ego vehicle collides with the leading vehicle after 6 s. As expected, the SSMs in Figure 5 indicate a crash probability of 1. The difference between $\hat{P}(\mathrm{C}|x)$ and WS is that $\hat{P}(\mathrm{C}|x)$ increases earlier. Note that $\hat{P}(\mathrm{C}|x)$ increasing sooner than WS does not necessarily mean that it is better: because there is no objective truth for an SSM, we cannot argue that one SSM is better than another SSM. Hence, in the next section, we will present a quantitative approach to benchmark an SSM.

Note that the choice of $d = 4$ can be considered to be somewhat arbitrary. Therefore, we have repeated this case study described with $d = 3$ and $d = 5$. With $d = 3$, the results are comparable to the results with $d = 4$ and there is hardly any difference noticeable. With $d = 5$, the derived SSM fluctuates a bit more, which resulted in a less smooth SSM. Most likely, this is the case because the KDE becomes less reliable for $d = 5$. We cannot argue objectively which choice of $d$ is better, but since a lower $d$ generally leads to more loss of information and the SSM with $d = 5$ is less smooth, we have opted for $d = 4$.

### *4.4. Benchmarking an SSM with expected risk trends*

In this section, we demonstrate an approach for benchmarking an SSM that is based on expected risk trends discussed in Mullakkal-Babu et al. (2017), who argue that the risk increases if the approaching speed of the ego vehicle toward the leading vehicle increases. Furthermore, the risk of the ego vehicle colliding with its leading vehicle increases with a higher ego vehicle speed (Aarts and Van Schagen, 2006). Note that as a result of the first risk trend, the risk of a collision with a vehicle behind the ego vehicle decreases with high ego vehicle speed, but since the current benchmarking only considers the risk of the ego vehicle colliding with its leading vehicle, the risk of a collision with a vehicle behind the ego vehicle is not further considered. The risk also increases with a higher driver reaction time (Klauer et al., 2006). On the other hand, the risk decreases with a higher road friction (Wallman and Åström, 2001) or a larger intervehicle spacing (Mullakkal-Babu et al., 2017).

To check whether the developed SSM follows these 5 expected risk trends[6], we evaluate the partial derivatives of the measure of (17). The intuition is as follows: If the expected risk trend for an input X (e.g., the ego vehicle speed) is that the risk increases as X increases, then we expect the partial derivative of our SSM with respect to X to be positive. Furthermore, if we evaluate the partial derivative at many points, we expect that at least the majority of these evaluated partial derivatives is positive. Similarly, if we expect that the risk measure decreases with increasing X, then we expect that at least the majority of the evaluated partial derivatives is negative. Note that because the proposed benchmarking method only considers the partial derivatives, no claim can be made regarding the actual accuracy of the SSM.

To illustrate the approach for benchmarking an SSM, we use the SSM of Section 4.2 with a few different assumptions. Because we have not described an expected trend regarding $a_\mathrm{l}$, we simply use $a_\mathrm{l} = 0$. Also, because the expected risk trend for the relative speed is defined, we use the relative speed, i.e., $\Delta v = v_\mathrm{e} - v_\mathrm{l}$, instead of $v_\mathrm{l}$. For the same reason, instead of assuming a random reaction time $t_\mathrm{r}$ and MADR $a_\mathrm{MADR}$, these

[5]Alternatively, the bandwidth matrix $H_\mathrm{NW}$ may be increased. On the one hand, this will lower the variance of the error, but, on the other hand, it will increase the bias of the result. We refer the interested reader to (Chen, 2017) for more details on the effect of $H_\mathrm{NW}$.

[6]In (Mullakkal-Babu et al., 2017), a sixth expected risk trend is mentioned based on (Evans, 1994), namely the vehicle mass. Our interpretation of (Evans, 1994), however, is that the ratio of masses of two colliding vehicles influences the safety risk and that one cannot argue that a higher mass of the ego vehicle necessarily increases the safety risk. Therefore, we exclude the ego vehicle mass from our analysis.

Table 1: Percentiles of the partial derivatives of the SSM and the corresponding expected risk trends.

| | $v_\mathrm{e} - v_\mathrm{l}$ | $v_\mathrm{e}$ | $t_\mathrm{r}$ | $g$ | $a_\mathrm{MADR}$ |
|---|---|---|---|---|---|
| Expected trend | Increase | Increase | Increase | Decrease | Decrease |
| Maximum | 0.1629 | 0.1555 | 1.3136 | 0.0010 | 0.0037 |
| 99th percentile | 0.1585 | 0.1162 | 0.8968 | 0.0002 | 0.0002 |
| 95th percentile | 0.1495 | 0.0524 | 0.6765 | 0.0000 | -0.0000 |
| 90th percentile | 0.1346 | 0.0151 | 0.5351 | -0.0000 | -0.0000 |
| 75th percentile | 0.0746 | 0.0012 | 0.2917 | -0.0002 | -0.0003 |
| 50th percentile | 0.0114 | 0.0001 | 0.0605 | -0.0070 | -0.0054 |
| 25th percentile | 0.0004 | 0.0000 | 0.0022 | -0.0320 | -0.0290 |
| 10th percentile | 0.0000 | -0.0000 | 0.0000 | -0.0545 | -0.0654 |
| 5th percentile | 0.0000 | -0.0002 | 0.0000 | -0.0645 | -0.0880 |
| 1st percentile | 0.0000 | -0.0007 | -0.0020 | -0.0781 | -0.1337 |
| Minimum | -0.0035 | -0.0030 | -0.0180 | -0.1076 | -0.2030 |

are now considered as input to our measure. Finally, instead of using the log of the gap between the ego vehicle and the leading vehicle, we use the gap as a direct input. Thus, we have:

$$x^\mathsf{T} = \begin{bmatrix} v_\mathrm{e} - v_\mathrm{l} & v_\mathrm{e} & t_\mathrm{r} & g & a_\mathrm{MADR} \end{bmatrix}. \tag{25}$$

We compute $\hat{P}(\mathrm{C}|x)$ using (17) where the points $\{x'_k\}_{k=1}^m$ are taken from a grid. For each input variable, 10 different values at equal distance are used, resulting in $m = 10^5$. Here, $v_\mathrm{e} - v_\mathrm{l}$ ranges from $0\,\mathrm{m/s}$ to $20\,\mathrm{m/s}$, $v_\mathrm{e}$ ranges from $10\,\mathrm{m/s}$ to $30\,\mathrm{m/s}$, $t_\mathrm{r}$ ranges from $0.5\,\mathrm{s}$ to $1.5\,\mathrm{s}$, $g$ ranges from $5\,\mathrm{m}$ to $30\,\mathrm{m}$, and $a_\mathrm{MADR}$ ranges from $4\,\mathrm{m/s^2}$ to $10\,\mathrm{m/s^2}$. A threshold $\epsilon = 0.02$ is used. For the bandwidth matrix $H_\mathrm{NW}$, we use a diagonal matrix with the $(i,i)$-th entry corresponding to the squared difference between two consecutive values of the $i$-th entry of $x$. For example, the first value is $(20\,\mathrm{m/s}/(10-1))^2 \approx 4.9\,\mathrm{m^2/s^2}$. The other values on the diagonal are: $4.9\,\mathrm{m^2/s^2}$, $0.012\,\mathrm{s^2}$, $7.7\,\mathrm{m^2}$, and $0.44\,\mathrm{m^2/s^4}$. For each input variable listed in (25), we evaluate the partial derivative of (17) at each $x'_k$, $k \in \{1, \ldots, m\}$.

Table 1 shows the result of the benchmarking. It shows that the SSM follows the expected risk trends mostly. E.g., in more than 99 % of the cases, the partial derivative of the relative speed $(v_\mathrm{e} - v_\mathrm{l})$ is positive. For the remaining 1 %, the partial derivative is negative, albeit only slightly. One explanation is that this remaining 1 % is caused by the inaccuracies introduced by the numerical approximation of (14).

## 5. Discussion

Typically, SSMs rely on assumptions regarding the behavior of traffic participants. An advantage of the presented PRISMA method for deriving SSMs is that the PRISMA method is not bound to certain predetermined assumptions. We want to stress, however, that when using the PRISMA method for deriving an SSM, a set of assumptions is still needed. In fact, multiple SSMs can be derived by using the PRISMA method with different sets of assumptions. As a result, the PRISMA method can be used to derive multiple SSMs that are applicable in various types of scenarios, e.g., ranging from vehicle-following scenarios to scenarios at intersections. Note that although the PRISMA method is applicable in various types of scenarios, the current case study focuses on longitudinal traffic conflicts. In a future work, we will present the application of the PRISMA method for deriving SSMs for lateral traffic conflicts.

The PRISMA method uses data to adapt the SSMs to, e.g., the local traffic behavior. More specifically, the data are used to predict the possible future situations ($y$) given an initial situation ($x$). This can be an advantage because the data can be used to rely less on assumptions as to how the future develops given an initial situation. To fully benefit from this approach, the data should satisfy a few conditions. First, the recorded data need to represent the actual traffic behavior in which the SSMs are applied. Second, we

need enough data to estimate $p(y|x)$. In (de Gelder et al., 2019), a metric is presented that can be used to determine whether enough data have been collected to estimate $p(y|x)$ accurately.

The PRISMA method can still be applied in case no data are available. The first alternative is to use existing knowledge to determine an estimate of $p(y|x)$ instead of estimating $p(y|x)$ on the basis of data. For example, statistics or literature on driving behavior of traffic participants may be used. The second alternative is to use assumptions on how the future develops given an initial situation $x$. For example, when assuming that the speed of the leading vehicle in Section 4.1.2 remains constant, it is not needed to estimate $p(y|x)$. Note that a combination is also possible. For example, estimate $p(y|x)$ based on data in case $x$ is well represented in the data, but define $p(y|x)$ on the basis of existing knowledge and/or assumptions for the cases where $x$ is underrepresented in the data. A third alternative is to use other methods for predicting the trajectories of the other traffic participants, e.g., using hidden Markov models (Laugier et al., 2011), sequence similarity methods (Saunier et al., 2007), Gaussian mixture models (Wiest et al., 2012), or long short-term memory networks (Deo and Trivedi, 2018). For an overview of trajectory prediction models for vehicles and pedestrians, see (Lefèvre et al., 2014) and (Rudenko et al., 2020), respectively.

Note that the PRISMA method is used to derive SSMs that predict the probability of a specific event, such as a crash, i.e., the derived SSMs can be used as a measure of proximity of the specified event. However, the PRISMA method is not used to measure the severity of an interaction, i.e., the extent of harm in case the interaction leads to a crash. For measuring the severity of an interaction, typically energy-based SSMs are used (Wang et al., 2021). So, if there is a need to also have an indicator of the severity of an interaction, an energy-based SSM, e.g., see (Ozbay et al., 2008; Alhajyaseen, 2015; Laureshyn et al., 2017; Mullakkal-Babu et al., 2020), may be considered alongside an SSM derived using the PRISMA method.

We have illustrated the PRISMA method through different derived SSMs in the case study. The derived SSMs estimate the probability of a crash with a leading vehicle under different assumptions. Because of the focus on crashes, the resulting SSMs may still be low given an initial situation that is generally considered to be unsafe. For example, the SSM described in Section 4.2 gives a crash probability of approximately 14 % when approaching a leading vehicle that is driving at a constant speed of $v_\mathrm{l} = 12\,\mathrm{m/s}$ ($a_\mathrm{l} = 0\,\mathrm{m/s^2}$) with a speed of $v_\mathrm{e} = 25\,\mathrm{m/s}$ and a gap of $g = 20\,\mathrm{m}$ (see left heat map in Figure 4). In this initial situation, the THW is only $g/v_\mathrm{e} = 0.8\,\mathrm{s}$ and the TTC is only $g/(v_\mathrm{e} - v_\mathrm{l}) = 1.5\,\mathrm{s}$, whereas a THW of less than 1 s or a TTC of less than 1.5 s is considered unsafe (Vogel, 2003). In order to put more emphasis on such unsafe situations, different events — instead of crashes — can be considered. For example, we can derive an SSM that estimates the probability that the TTC is below 1 s within the next 5 seconds. More research is needed to investigate whether such SSMs can be of practical use, e.g., for evaluating whether a driver is actively pursuing large safety margins.

A few choices have to be made when using the PRISMA method for deriving SSMs. One such a choice is the set of initial situations $\{x_1, \ldots, x_m\}$ for which the probability $P(\mathrm{C}|x)$ is estimated. Generally speaking, for larger $m$, the approximation of $P(\mathrm{C}|x)$ in (17) improves. One disadvantage, however, is that more simulation runs are required when $m$ is larger, but because these simulation runs are performed offline, this problem might be solved by, e.g., parallel computing resources. Another disadvantage is that the computational cost of the approximation in (17) scales linearly with $m$. Especially when using this approximation for real-time evaluation of the SSM, this can be a bottleneck. One solution to this is to not use all $m$ initial situations for evaluating (17). The intuition is as follows: since (17) uses local regression, an initial situation $x_k$ can be removed from the set $\{x_1, \ldots, x_m\}$ if all neighboring data points give (approximately) the same probability of the event C, i.e., $|\hat{P}(\mathrm{C}|x_i) - \hat{P}(\mathrm{C}|x_k)|$ is below a threshold for all $x_i$, $i \neq k$ for which $\|x_i - x_k\|_2$ is below another threshold (assuming that $\hat{P}(\mathrm{C}|x)$ is sufficiently smooth). For example, the SSM that is shown in Figure 4, only a few initial situations are required in the upper right region of the heat maps, since the estimated probability is always lower than 0.1.

Another choice is the threshold $\epsilon$ that controls the number of simulation runs ($N_\mathrm{sim}$) that are used to estimate $P(\mathrm{C}|x)$. According to (16), $N_\mathrm{sim}$ is increased until the variance of the estimation error is below $\epsilon$, i.e., $\mathrm{Var}\Big[P(\mathrm{C}|x) - \hat{P}(\mathrm{C}|x)\Big] < \epsilon$. Therefore, a lower $\epsilon$ generally results in more accurate estimations of the probability, as illustrated in Figure 2. The downside, however, is that for a lower $\epsilon$, more offline simulation runs are required. Although a good choice of $\epsilon$ remains a topic of research, based on experience, we advice

to use a maximum threshold of $\epsilon = 0.1$ and lower values if the computational resources allow for this.

In the examples presented in Section 4, we have considered the leading vehicle as the only traffic participant other than the ego vehicle. The PRISMA method can be applied in scenarios with multiple traffic participants other than the ego vehicle. However, the number of parameters ($n_x$, i.e., the size of $x$) then becomes larger. As a result, two problems may arise. First, as $m$ grows exponentially with $n_x$, so does the number of simulation runs. Second, even if these simulation runs can be performed, the regression using (17) becomes slow due to the large $m$. To overcome these problems, an SSM can be computed for each traffic participant independently. For example, let $N_{\text{tp}}$ denote the number of traffic participants other than the ego vehicle. With $i \in \{1, \ldots, N_{\text{tp}}\}$, let $\mathrm{C}_i$ denote the event of colliding with the $i$-th traffic participant and let $x^i$ denote the initial situation considering the $i$-th traffic participant. Under the assumption that $P\big(\mathrm{C}_i|x^i\big)$ is independent of $x^j$ for all $i \neq j$, we can calculate the probability of colliding with one or more traffic participants using

$$1 - \prod_{i=1}^{N_{\text{tp}}} \big(1 - P\big(\mathrm{C}_j|x^j\big)\big). \tag{26}$$

For example, consider a scenario with multiple crossing pedestrians. Using the PRISMA method, we can derive an SSM that estimates the probability of colliding with a pedestrian. Then, after evaluating this SSM for each pedestrian, the probability of colliding with one or more pedestrians can be calculated using (26) without the need for an SSM that considers multiple pedestrians.

In the case study, we have shown how to analyze an SSM both qualitatively, using heat maps and testing the SSM in different scenarios, and quantitatively by benchmarking the SSM with expected risk trends (Mullakkal-Babu et al., 2017). Since the SSMs derived using the PRISMA method provide a probability, it is also possible to verify the estimated probability by comparing it with real data. This requires, however, an extensive data set that would allow for estimating the probability of the event C, e.g., a crash, in the near future given a certain situation a vehicle is in. It remains a topic for future work to use such a data set to verify the SSMs derived using the PRISMA method.

The PRISMA method is a novel approach for deriving probabilistic SSMs for risk evaluation. Some limitations, however, may hamper its use for real-world applications. First, as described earlier, if the set of initial situations $\{x_1, \ldots, x_m\}$ is too large, real-time calculation of the SSM may be difficult. As a consequence, the dimension of the vector describing the initial situation, $x$, cannot be too large, meaning that the initial situation needs to be encoded into a limited set of numbers. Second, KDE does not work well for large dimensions. Note, however, that we have provided some options for reducing the dimensionality, and, if reducing the dimensionality further is not an good option, the PRISMA method can also be applied when other methods are used for the probability density estimation. Third, many simulations may be required. It helps that the simulations can be conducted offline, but it may still be challenging to conduct the simulations in a reasonable time window. Fourth, although we claim that the PRISMA method can be used to derive multiple SSMs for different type of scenarios, we cannot claim that the derived SSMs are more valid than others SSMs, nor can we derive an SSM that is valid for all types of scenarios. Lastly, for the simulations, the response of the ego vehicle must be assumed. This may not coincide with the ego vehicle response in reality. In a future work, in case of a human driver, this may be tackled by considering the state of the human driver, such as whether the eyes are on the road and/or towards a conflicting traffic participant, as part of the state vector that is used to describe the initial situation.

## 6. Conclusions

Road safety is an important research topic. To quantify the safety at a vehicle level, Surrogate Safety Measures (SSMs) are often used to characterize the risk of a crash. We have proposed a novel approach called the Probabilistic RISk Measure derivAtion (PRISMA) method for deriving SSMs that calculate the probability that a certain event, e.g., a crash, will happen in the near future given an initial situation. Whereas traditional SSMs are generally only applicable in certain types of scenarios, the PRISMA method can be applied to various types of scenarios. Furthermore, because the PRISMA method is data-driven, the

derived SSMs can be adapted to the local traffic behavior that is captured by the data. Also, no assumptions on the driver behavior are made. Therefore, the PRISMA method has the potential for deriving multiple SSMs for quantifying the safety of a — possibly automated — vehicle.

We have illustrated that the PRISMA method can be used to reproduce known probabilistic SSMs. In an example, we have derived a new SSM based on the Next Generation SIMulation (NGSIM) data set that calculates the risk of a crash in a longitudinal interaction between two vehicles. Through several explanatory scenarios, it has been shown that the derived SSM correctly provides a quantification of the crash risk. We have also presented how the evaluation of the partial derivatives of the SSM can be used to benchmark an SSM using expected risk trends.

The SSMs derived using the presented PRISMA method can be used to warn drivers for unsafe situations and ensuring that proper attention is being paid to the road situation. Furthermore, the derived measures can prospectively estimate the impact of newly introduced systems on traffic safety. A limitation of the current study is that the presented approach is only applied to longitudinal traffic interactions. Future work involves applying the PRISMA method for the derivation of SSMs that measure the risk of lateral traffic interactions, interactions with vulnerable road users, and interactions with multiple (different types of) traffic participants. Furthermore, more research is needed to investigate whether the SSMs derived by the PRISMA method can be used to evaluate whether a driver is actively pursuing (large) safety margins.

## Acknowledgment

This research was supported in part by the SAFE-UP project (proactive SAFEty systems and tools for a constantly UPgrading road environment). SAFE-UP has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 861570.

## References

L. Aarts and I. Van Schagen. Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention*, 38(2): 215–224, 2006. doi: 10.1016/j.aap.2005.07.004.

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009. doi: 10.1016/j.insmatheco.2007.02.001.

H. Abdi and L. J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4): 433–459, 2010. doi: 10.1002/wics.101.

V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, and G. McHale. The next generation simulation program. *Institute of Transportation Engineers. ITE Journal*, 74(8):22–26, 2004.

W. K. M. Alhajyaseen. The integration of conflict probability and severity for the safety assessment of intersections. *Arabian Journal for Science and Engineering*, 40(2):421–430, 2015. doi: 10.1007/s13369-014-1553-1.

B. L. Allen, B. T. Shin, and P. J. Cooper. Analysis of traffic conflicts and collisions. *Transportation Research Board*, 667: 67–74, 1978.

S. Almqvist, C. Hydén, and R. Risser. Use of speed limiters in cars for increased safety and a better environment. *Transportation Research Record*, 1318, 1991.

R. Altendorfer and C. Wilkmann. A new approach to estimate the collision probability for automotive applications. *Automatica*, 127:109497, 2021. doi: 10.1016/j.automatica.2021.109497.

A. Arun, M. M. Haque, A. Bhaskar, S. Washington, and T. Sayed. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accident Analysis & Prevention*, 153(106016), 2021. doi: 10.1016/j.aap.2021.106016.

D. Åsljung, J. Nilsson, and J. Fredriksson. Using extreme value theory for vehicle level safety validation and implications for autonomous vehicles. *IEEE Transactions on Intelligent Vehicles*, 2(4):288–297, 2017. doi: 10.1109/TIV.2017.2768219.

D. M. Bashtannyk and R. J. Hyndman. Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 36(3):279–298, 2001. doi: 10.1016/S0167-9473(00)00046-3.

Y.-C. Chen. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1(1):161–187, 2017. doi: 10.1080/24709360.2017.1396742.

S.-T. Chiu. A comparative review of bandwidth selection for kernel density estimation. *Statistica Sinica*, 6:129–145, 1996.

C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. doi: 10.2307/2331986.

F. Cunto and F. F. Saccomanno. Simulated safety performance of rear-end and angled vehicle interactions at isolated intersections. *Canadian Journal of Civil Engineering*, 36(11):1794–1803, 2009. doi: 10.1139/L09-092.

F. J. C. Cunto. *Assessing Safety Performance of Transportation Systems Using Microscopic Simulation*. PhD thesis, University of Waterloo, 2008.

G. A. Davis, J. Hourdos, H. Xiong, and I. Chatterjee. Outline for a causal model of traffic conflicts and crashes. *Accident Analysis & Prevention*, 43(6):1907–1919, 2011. doi: 10.1016/j.aap.2011.05.001.

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

E. de Gelder, E. Cator, J.-P. Paardekooper, O. Op den Camp, and B. De Schutter. Constrained sampling from a kernel density estimator to generate scenarios for the assessment of automated vehicles. In *IEEE Intelligent Vehicles Symposium Workshops (IV Workshop)*, pages 203–208, 2021.

E. de Gelder, J. Hof, E. Cator, J.-P. Paardekooper, O. Op den Camp, J. Ploeg, and B. De Schutter. Scenario parameter generation method and scenario representativeness metric for scenario-based assessment of automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):18794–18807, 2022. doi: 10.1109/TITS.2022.3154774.

N. Deo and M. M. Trivedi. Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs. In *Intelligent Vehicles Symposium (IV)*, pages 1179–1184. IEEE, 2018. doi: 10.1109/IVS.2018.8500493.

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

C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. Neural spline flows. In *Advances in Neural Information Processing Systems*, 2019. URL `https://proceedings.neurips.cc/paper_files/paper/2019/file/7ac71d433f282034e088473244df8c02-Paper.pdf`.

European Commission. Road safety targets — monitoring report June 2020. Technical report, European Road Safety Observatory, 2020.

L. Evans. Driver injury and fatality risk in two-car crashes versus mass ratio inferred using Newtonian mechanics. *Accident Analysis & Prevention*, 26(5):609–616, 1994. doi: 10.1016/0001-4575(94)90022-1.

G. H. Golub and C. F. Van Loan. *Matrix Computations*, volume 3. John Hopkins University Press, 2013.

M. Green. "how long does it take to stop?" methodological analysis of driver perception-brake times. *Transportation Human Factors*, 2(3):195–216, 2000. doi: 10.1207/sthf0203_1.

G. Guido, F. Saccomanno, A. Vitale, V. Astarita, and D. Festa. Comparing safety performance measures obtained from video capture data. *Journal of Transportation Engineering*, 137(7):481–491, 2011. doi: 10.1061/(ASCE)TE.1943-5436.0000230.

A. S. Hakkert, L. Braimaister, and I. Van Schagen. The uses of exposure and risk in road safety studies. Technical Report R-2002-12, SWOV Institute for Road Safety, 2002.

B. M. S. Hasan and A. M. Abdulazeez. A review of principal component analysis algorithm for dimensionality reduction. *Journal of Soft Computing and Data Mining*, 2(1):20–30, 2021. URL `https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/8032`.

J. C. Hayward. Near miss determination through use of a scale of danger. Technical Report TTSC-7115, Pennsylvania State University, 1972.

M. P. Holmes, A. G. Gray, and C. L. Isbell. Fast nonparametric conditional density estimation. In *23rd Conference on Uncertainty in Artificial Intelligence*, pages 175–182, 2007.

Y. Iida, N. Uno, S. Itsubo, and M. Suganuma. Traffic conflict analysis and modeling of lane-changing behavior at weaving section. In *Infrastructure Planning*, pages 305–308, 2001.

M. C. Jones, J. S. Marron, and S. J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433):401–407, 1996. doi: 10.1080/01621459.1996.10476701.

S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical Report DOT HS 810 594, Virginia Tech Transportation Institute, 2006.

C. Laugier, I. E. Paromtchik, M. Perrollaz, M. Yong, J.-D. Yoder, C. Tay, K. Mekhnacha, and A. Nègre. Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety. *IEEE Intelligent Transportation Systems Magazine*, 3(4):4–19, 2011. doi: 10.1109/MITS.2011.942779.

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. doi: 10.1016/j.aap.2010.03.021.

A. Laureshyn, T. De Ceunynck, C. Karlsson, Å. Svensson, and S. Daniels. In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator. *Accident Analysis & Prevention*, 98:46–56, 2017. doi: 10.1016/j.aap.2016.09.026.

S. Lefèvre, D. Vasquez, and C. Laugier. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH Journal*, 1(1):1–14, 2014. doi: 10.1186/s40648-014-0001-z.

S. M. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli. Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs. *IATSS Research*, 41(4):153–163, 2017. doi: 10.1016/j.iatssr.2017.02.001.

M. M. Minderhoud and P. H. Bovy. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention*, 33(1):89–97, 2001. doi: 10.1016/S0001-4575(00)00019-1.

F. A. Mullakkal-Babu, M. Wang, H. Farah, B. van Arem, and R. Happee. Comparative assessment of safety indicators for vehicle trajectories on highways. *Transportation Research Record*, 2659(1):127–136, 2017. doi: 10.3141/2659-14.

F. A. Mullakkal-Babu, M. Wang, X. He, B. van Arem, and R. Happee. Probabilistic field approach for motorway driving risk assessment. *Transportation Research Part C: Emerging Technologies*, 118:102716, 2020. doi: 10.1016/j.trc.2020.102716.

E. A. Nadaraya. Some new estimates for distribution functions. *Theory of Probability & Its Applications*, 9(3):497–500, 1964. doi: 10.1137/1109069.

T. Nagler and C. Czado. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, 151(Supplement C):69–89, 2016. doi: 10.1016/j.jmva.2016.07.003.

National Center for Statistics and Analysis. Summary of motor vehicle crashes: 2018 data. Technical Report DOT HS 812 961, National Highway Traffic Safety Administration, 2020.

K. Ozbay, H. Yang, B. Bartin, and S. Mudigonda. Derivation and validation of new simulation-based surrogate safety measure. *Transportation Research Record*, 2083:105–113, 2008. doi: 10.3141/2083-12.

E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962. doi: 10.1214/aoms/1177704472.

D. E. Peerlings, J. A. van den Brakel, N. Baştürk, and M. J. Puts. Multivariate density estimation by neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, Early Access, 2022. doi: 10.1109/TNNLS.2022.3190220.

M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27 (3):832–837, 1956. doi: 10.1214/aoms/1177728190.

A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras. Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, 39(8):895–935, 2020. doi: 10.1177/0278364920917446.

N. Saunier and T. Sayed. Probabilistic framework for automated analysis of exposure to road collisions. *Transportation Research Record*, 2083(1):96–104, 2008. doi: 10.3141/2083-11.

N. Saunier, T. Sayed, and C. Lim. Probabilistic collision prediction for vision-based automated road safety analysis. In *IEEE Intelligent Transportation Systems Conference*, pages 872–878, 2007. doi: 10.1109/ITSC.2007.4357793.

W. J. Schakel, B. Van Arem, and B. D. Netten. Effects of cooperative adaptive cruise control on traffic flow stability. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 759–764, 2010. doi: 10.1109/itsc.2010.5625133.

D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, 2015.

S. Shalev-Shwartz, S. Shammah, and A. Shashua. On a formal model of safe and scalable self-driving cars. *arXiv preprint arXiv:1708.06374*, 2017.

X. Shi, Y. D. Wong, M. Z. F. Li, and C. Chai. Key risk indicators for accident assessment conditioned on pre-crash vehicle trajectory. *Accident Analysis & Prevention*, 117:346–356, 2018. doi: 10.1016/j.aap.2018.05.007.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. CRC press, 1986.

P. Songchitruksa and A. P. Tarko. The extreme value theory approach to safety estimation. *Accident Analysis & Prevention*, 38(4):811–822, 2006. doi: 10.1016/j.aap.2006.02.003.

A. P. Tarko. Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis & Prevention*, 45: 230–240, 2012. doi: 10.1016/j.aap.2011.07.008.

A. P. Tarko. Estimating the expected number of crashes with traffic conflicts and the Lomax distribution — a theoretical and numerical exploration. *Accident Analysis & Prevention*, 113:63–73, 2018a. doi: 10.1016/j.aap.2018.01.008.

A. P. Tarko. Surrogate measures of safety. In *Safe Mobility: Challenges, Methodology and Solutions*. Emerald Publishing Limited, 2018b. doi: 10.1108/S2044-994120180000011019.

B. A. Turlach. Bandwidth selection in kernel density estimation: A review. Technical report, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, 1993.

N. Uno, Y. Iida, S. Yasuhara, and M. Suganuma. Objective analysis of traffic conflict and modeling of vehicular speed adjustment at weaving section. *Infrastructure Planning Review*, 20:989–996, 2003. doi: 10.2208/journalip.20.989.

K. Vogel. A comparison of headway and time to collision as safety indicators. *Accident Analysis & Prevention*, 35(3):427–433, 2003. doi: 10.1016/S0001-4575(02)00022-2.

C.-G. Wallman and H. Åström. Friction measurement methods and the correlation between road friction and traffic safety: A literature review. Technical Report 911A, Swedish National Road and Transport Research Institute (VTI), 2001.

M. P. Wand and M. C. Jones. Multivariate plug-in bandwidth selection. *Computational Statistics*, 9(2):97–116, 1994.

C. Wang and N. Stamatiadis. Evaluation of a simulation-based surrogate safety metric. *Accident Analysis & Prevention*, 71: 82–92, 2014. doi: 10.1016/j.aap.2014.05.004.

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

L. Wasserman. *All of Nonparamatric Statistics*. Springer, 2006.

J. Wiest, M. Höffken, U. Kreßel, and K. Dietmayer. Probabilistic trajectory prediction with Gaussian mixture models. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 141–146, 2012. doi: 10.1109/IVS.2012.6232277.

E. B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927. doi: 10.1080/01621459.1927.10502953.

K. Xie, D. Yang, K. Ozbay, and H. Yang. Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. *Accident Analysis & Prevention*, 125:311–319, 2019. doi: 10.1016/j.aap.2018.07.002.

A. Z. Zambom and R. Dias. A review of kernel density estimation with applications to econometrics. *International Econometric Review (IER)*, 5(1):20–42, 2013.

L. Zheng, T. Sayed, and F. Mannering. Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. *Analytic Methods in Accident Research*, page 100142, 2020. doi: 10.1016/j.amar.2020.100142.