Accident Analysis and Prevention 162 (2021) 106403

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Performance evaluation of surrogate measures of safety with naturalistic driving data

Chang Lu [a,b,1], Xiaolin He [c,1], Hans van Lint [d], Huizhao Tu [a,b], Riender Happee [c,d], Meng Wang [d,*]

[a] College of Transportation Engineering, Tongji University, 4800 Cao'an Road, Shanghai 201804, China
[b] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China
[c] Department of Cognitive Robotics, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands
[d] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands

ARTICLE INFO

Keywords:
Surrogate measure of safety
Performance evaluation
Naturalistic driving data
Empirical analysis

ABSTRACT

Surrogate measures of safety (SMoS) play an important role in detecting traffic conflicts and in traffic safety assessment. However, the underlying assumptions of SMoS are different and a certain SMoS may be adequate/inadequate for different applications. A comprehensive approach to evaluate the validity and applicability of SMoS is lacking in the literature. This study proposes such a framework that supports evaluating SMoS in multiple dimensions. We apply the framework to gain insights into the characteristics of six widely-used SMoS for longitudinal maneuvers, i.e., Time to Collision (TTC), single-step Probabilistic Driving Risk Field (S-PDRF), Deceleration Rate to Avoid a Crash (DRAC), Potential Index for Collision with Urgent Deceleration (PICUD), Proactive Fuzzy Surrogate Safety Metric (PFS), and the Critical Fuzzy Surrogate Safety Metric (CFS). To ensure comparability, all measures are calibrated with the same risk detection criterion. Four performance indicators, i.e., Prediction Accuracy, Timeliness, Robustness, and Efficiency are computed for all six SMoS and validated using naturalistic driving data. The strengths and weaknesses of all six measures are compared and analyzed elaborately. A key result is that not a single SMoS performs well in all performance dimensions. S-PDRF performs best in terms of Robustness but consumes the most time for computation. TTC is the most efficient but performs poorly in terms of Timeliness and Robustness. The proposed evaluation approach and the derived insights can support SMoS selection in active vehicle safety system design and traffic safety assessment.

1. Introduction

Road traffic injury is listed in the top ten major causes of mortality and morbidity worldwide (Blas and Kurup, 2010). Road traffic crashes result in approximately 1.35 million deaths and between 20 and 50 million non-fatal injuries each year (WHO, 2020).

Traffic accidents are mainly caused by human misjudgments (Nadimi et al., 2016). With the considerable growth of communication technology, computer capabilities, and data collection technology, active vehicle safety systems have been developed to reduce collision risk.

Surrogate measures of safety (SMoS) play an essential role in detecting traffic conflicts and evaluating traffic safety (Gao et al., 2020; Shi et al., 2021). They are handy for developing real-time collision warning systems (Bao et al., 2012) and decision-making for automated driving (Mullakkal-Babu et al., 2020). The selection of proper SMoS is difficult but essential for risk detection and safety evaluation. Generally, SMoS are chosen on the basis of their application scope and usefulness (Mullakkal-Babu et al., 2017).

From the perspective of application scope, SMoS have been classified by interaction styles, i.e., longitudinal interaction, lateral interaction, and two-dimensional interaction. Common longitudinal SMoS are time to collision (TTC) (Hayward, 1972; Mahmud et al., 2017; Mullakkal-Babu et al., 2017), deceleration required to avoid collision (DRAC) (Cooper and Ferguson, 1976), and potential indicator of collision with urgent deceleration (PICUD) (Uno et al., 2002; Fazekas et al., 2017). Lateral SMoS like post encroachment time (PET) (Songchitruksa and

* Corresponding author.
E-mail addresses: 1610756@tongji.edu.cn (C. Lu), x.he-2@tudelft.nl (X. He), J.W.C.vanLint@tudelft.nl (H. van Lint), huizhaotu@tongji.edu.cn (H. Tu), r.happee@tudelft.nl (R. Happee), m.wang@tudelft.nl (M. Wang).
[1] The authors contribute equally.

https://doi.org/10.1016/j.aap.2021.106403
Received 23 May 2021; Received in revised form 6 August 2021; Accepted 13 September 2021

Tarko, 2006), have been used as a risk measure in lane change controllers, safety assessment of intersections, and lateral vehicle maneuvers. In addition, some SMoS can be used in both longitudinal and lateral driving, e.g., single-step Probabilistic Driving Risk Field (S-PDRF) (Mullakkal-Babu et al., 2020).

If we limit the scope to just *risk detection in the longitudinal dimension*, there are still quite a number of SMoS with different features. From the perspective of operational attributes, i.e., time, speed, acceleration, and deceleration. SMoS can be classified into four categories, i.e., time-based (e.g., TTC, time headway), distance-based (e.g., PICUD), deceleration-based (e.g., DRAC), and others (Fazekas et al., 2017; Mullakkal-Babu et al., 2017). Considering the time argument used for quantifying risk, SMoS can be divided into two categories, i.e., imminent risk and potential risk (Nadimi et al., 2016; Songchitruksa and Tarko, 2006). SMoS which detect imminent risk evaluate this risk based on the current state (e.g., acceleration, velocity, and position) of vehicles only. SMoS which detect *potential risk* use prediction (extrapolations) of that state. For example, TTC is a well-known SMoS for imminent risk detection, whereas PICUD is used to detect potential risk. Considering the uncertainty of parameters, SMoS can in turn be divided into two categories, that is, SMoS with *deterministic parameters* (e.g., TTC and PICUD) and ones with *probabilistic parameters* (e.g., S-PDRF). The main difference between these two categories is whether behaviors of interacting vehicles are assumed deterministic or stochastic.

To evaluate the usefulness of SMoS within an application scope, SMoS have been extensively reviewed (Li et al., 2020a; Li et al., 2020b; Li et al., 2020c). Guido et al. compared SMoS (including DRAC, TTC, proportion of stopping distance, time integrated time-to-collision, and crash potential index) with respect to traffic conditions and variations in roundabout geometry (Guido et al., 2011). Tak et al. estimated the collision risk of TTC, Stopping Headway Distance (SHD), and deceleration-based SMoS in both deceleration and acceleration phases (Tak et al., 2018). The results show that TTC and SHD obviously overestimate the collision risk during deceleration. Vogel compared TTC and time headway for vehicles in a car following situation. The results show that the percentage of small time headways is relatively constant across different locations, while the percentage of small TTC values varies between different locations (Vogel, 2003).

Recent comparisons among SMoS focus on Accuracy and distributions of different risk levels under various scenarios, which consist of different locations, relative acceleration rates, relative speeds, traffic flow levels, etc. (Mullakkal-Babu, 2020). We argue that there are more dimensions of interest and relevance when applying SMoS. For instance, Timeliness (from the first moment to detect risk to the riskiest moment), Robustness (with respect to uncertainty in both traffic and ambient conditions), and computation time are highly relevant criteria for real-time applications. Limited literature exists on evaluating SMoS on these criteria.

Another problem is that the evaluation criteria for SMoS are typically inconsistent in the literature. There is no generally accepted consensus to classify acceptable and non-acceptable risk levels produced by different SMoS. In other words, those comparisons look at SMoS with specific thresholds to distinguish risky interactions from safe interactions, with possibly a few parameters representing the reaction and maneuver capabilities of drivers/vehicles. The consequence of this praxis is that there is a lack of comparability in these studies, which is amongst other things due to the lack of a proper calibration method for SMoS.

To overcome the aforementioned problems, this paper aims to develop an approach to evaluate the performance of SMoS. Considering the variety and degree of maturity of SMoS under different application scopes, we choose SMoS tailored to detect risk in the longitudinal dimension only (the risk of rear-end crash without lateral movements) in this study. To calibrate the thresholds and key parameters of SMoS, we use naturalistic driving data, which is widely used to study the assessment of SMoS because of the degree of realism in behaviors and interactions under risky events (Mattas et al., 2020; Mullakkal-Babu et al., 2020; Stapel et al., 2017). Using naturalistic driving data can provide fidelity and realism, which makes the transferability of the results easier.

The proposed framework involves three steps. The first step is to extract rear-end conflict events and to classify them into different risk levels. The second step is to select appropriate SMoS which can recognize different driving risks based on the naturalistic driving data. It is noted that calibrating the thresholds of different SMoS is of utmost importance in order to ensure comparability. In this study, based on the selected events, the thresholds of different SMoS will be calibrated. The calibration criterion is that SMoS with the thresholds can just recognize all events with high risk. The third step is to select indicators to comprehensively assess the performance of different SMoS based on the extracted and clustered driving event data, and derive insights into the performance of selected SMoS.

Compared to previous studies, our contributions are: 1) a systematic framework to evaluate the performance of SMoS. A key component here is the calibration method which ensures the comparability among SMoS; 2) new insights into the strengths and weaknesses of six typical SMoS for risk in the longitudinal dimension. The comparative study can assist the design of active vehicle safety systems and safety-oriented Intelligent Transportation Systems.

The remainder of the paper is structured as follows. In Section 2, the data selection and analyzing method are described. In Section 3, the indicators used for evaluating the performance of SMoS are proposed. The main results with naturalistic driving data are summarized and discussed in Section 4, and the findings and further research suggestions are presented in Section 5.

2. Methodology

In order to comprehensively evaluate the performance of different SMoS, a framework for performance evaluation of SMoS, including three necessary components, i.e., car-following events, SMoS, and performance indicators, is proposed. The sketch of the framework is shown in Fig. 1. Firstly, SMoS should be selected to cover different categories, and their performance indicators should be determined in advance. To conform the risk detection criteria of each SMoS, car-following events were extracted from naturalistic driving study to calibrate those. Besides that, those car-following events were also used as real cases to evaluate the performance of SMoS. Details of the naturalistic driving data source, events extraction, and SMoS calibration will be introduced in the remaining parts of this section.

2.1. Data source

In this paper, the naturalistic driving data from the 100-car Naturalistic Driving Study (NDS) was used, which includes driving data series of 60 crashes events and 768 near-miss events (Virginia Tech Transportation Institute, 2014). Common kinematic information of the subject vehicle and surrounding vehicle information was recorded in this database, such as vehicle's longitudinal velocity, acceleration, gap to neighbor vehicles, relative velocity to them, etc. The sampling rate is 0.1 s and 30 s before the events and 10 s after the events were covered. There are two reasons why this database was selected. On the one hand, the duration of driving recordings was suitable and sufficient to provide necessary details for safety assessment. On the other hand, compared to other databases, this NDS has clear labels for different critical events, which include the conflict type and the conflict reasons. Such labels can provide basic criteria to classify different critical events and can establish a basis for risk classification more accurately as well.

We selected events from the database based on four criteria: (1) Only the events of rear-end conflicts were selected, as the most common SMoS mainly focus on longitudinal conflicts; (2) Since the current study only focuses on rear-end conflicts, the events which were handled by steering

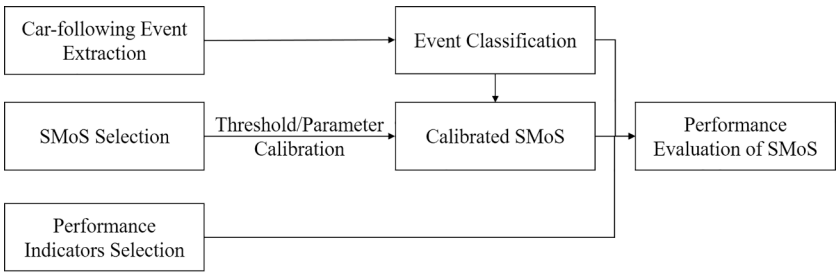

Fig. 1. The framework for performance evaluation of SMoS.

or the combination of steering and braking were excluded because they are also related to risk in the lateral dimension; (3) The events should have a complete data recording, which included the all the details during the whole event; (4) Since the driver's reaction is an important factor to cluster the event into different risk levels, the events without clear driver reaction (e.g., with no clear "elbow" point in the deceleration curve) were excluded. More details for preliminary data selection can be found in a previous publication (Xiong et al., 2019).

According to the criteria above, 97 near-miss events were selected (Xiong et al., 2019). However, in this paper, we paid special attention to the relative movement of the following vehicle and the leading vehicles. Thus, after further check, 13 near-miss events with incomplete radar data regarding the leading vehicle were excluded from further analysis. Apart from the near-miss events, we also checked the events where the subject vehicle was hit by the rear vehicle to focus on the risk assessment of the rear vehicle based on the data from the rear radar. After the check, one extra crash event was added to the dataset. Finally, 85 events of rear-end conflict were selected for our analysis. The conflict scenarios include freeway driving, urban driving, garage driving, etc., and a narrative of the conflict within each event was also included in the dataset. The velocity over all selected events ranged from 0 to 114 km/h.

2.2. Scenario classification

These 85 events were clustered into different risk levels to assess the performance of SMoS. Thus, we can identify whether different SMoS can recognize different risk levels with the naturalistic driving data. Spectral clustering was used for the clustering analysis. Spectral clustering based on graph theory treats all data as points in space and assumes all of the data points are connected by edges assigned with weights which are called a connected graph (Von Luxburg, 2007). The more significant the similarity between different points is, the higher the edge weight is. Spectral clustering creates groups where the weighted sum of edges between different groups is as small as possible but is high within a group. For clustering, braking profiles were extracted for these events. In usual practice, statistical values such as mean and maximum are computed as feature variables for clustering. However, the statistical computation may ignore some effective information of time series. Therefore, not only the average deceleration change rate over the deceleration profiles but also the range distance metric among different deceleration profiles computed with Dynamic Time Warping were used for clustering (Xiong et al., 2019). Therefore, the clustering algorithm could still work, although different events have different durations. Corresponding to the discussion at the beginning of Section 2.2 categories were selected for the identification of risk levels. Considering the limited sample size, 2 clusters were selected for risk level identification. In one cluster, higher maximum and mean deceleration level as well as faster average deceleration rate was featured compared to the other cluster. By using all the methods mentioned above, 28 events in the first cluster with higher average maximum deceleration and average mean deceleration were marked as high-level risk and 57 events in the second cluster were clustered into medium/low risk levels. Note that the events with medium/low risk level only mean they are relatively safe rather than absolutely safe.

2.3. SMoS calibration

There are various metrics for safety measurement and in all metrics, the selected thresholds or scales are vital parameters influencing the final performance. However, in most of the previous studies, when different metrics were compared, the specific thresholds were determined only by subjective experience, while the value range was obtained by statistical results of actual crash data (Virginia Tech Transportation Institute, 2014). Consequently, when these metrics were used to recognize driving risk during driving, the results were not convincing because of the experience-based threshold. As mentioned in Section 2.2, the biggest difference between events with high risk and those with low risk is that there is no unsafe moment in events with low risk. In this study, thresholds of each metric will be calibrated to maximize detection (correctly detect all events with highly unsafe moments) and to minimize false detections (incorrectly label events with low risk as dangerous events). Furthermore, some complex measures, for example, S-PDRF, have many other internal parameters influencing performance. These parameters should be calibrated as well.

Significantly, the calibrated values may be different from widely-used values in literature because of the calibration target. For example, the threshold of TTC may be a little higher than 4 s in this paper instead of 1.5–3 s, which is commonly used (Papazikou et al., 2019).

The calibration procedure is introduced as follows. Firstly, we should find out which key parameters that influence the performance are to be calibrated. For simple measures like TTC, there are no internal parameters to be calibrated and the only parameter which can be calibrated is the warning threshold value. Secondly, the extracted conflict events were used for different SMoS to compute the risk levels for every time step. Thirdly, possible values of parameters, including thresholds that can influence the final performance, should be traversed based on the driving data. The final step is to check which set of parameters has the best performance regarding individual SMoS. In this study, the basic criterion is that the safety measure can identify all events at high risk. Those SMoS that cannot meet this criterion should be excluded from further analysis.

3. Performance indicators for Surrogate measures of safety

In this section, to cover SMoS with different features, six SMoS for comparison were selected. Additionally, four indicators to assess the performance of SMoS were introduced.

3.1. Surrogate measures of safety for comparison

3.1.1. Time to Collision (TTC) (Hayward, 1972)

TTC can be defined as the required collision time if the two subjects in conflicts maintain their motion state as shown in Eq. (1).

$$TTC = \begin{cases} \dfrac{gap}{v_f - v_l}, v_f > v_l \\ \infty, v_f \le v_l \end{cases} \quad (1)$$

where v_f, v_l represent the velocity of the following vehicle and leading vehicle respectively. *gap* is the distance headway minus vehicle length of the leading vehicle. In this most common definition of TTC, acceleration is considered to be zero, which is a practical choice also given the noisy character of acceleration estimates (Happee et al., 2017).

From the perspective of TTC, when $v_f \le v_l$, it means that the vehicle is safe. If $v_f > v_l$ and TTC $\le$ threshold, it is defined that the vehicle is in an unsafe state. To be sure, as discussed in the study (Kuang et al., 2015), any scenario in which the follower's speed is lower than the leader's is regarded as a safe situation could be unreasonable for saturated situations where vehicles are traveling at similar speeds and with small headway.

3.1.2. Single-step Probabilistic driving risk field (S-PDRF) (Mullakkal-Babu et al., 2020)

The artificial potential field is a prominent paradigm that the vehicle can use the field gradient at its location to control the motion while avoiding obstacles. Wang et al. used this artificial field theory to model driving risk considering the influence of driver, vehicle, and other road characteristics (Wang et al., 2016). Li et al. used this theory to developed a warning strategy to prevent traffic accidents (Li et al., 2020a; Li et al., 2020b; Li et al., 2020c). The field theory has also been used to model traffic flow (Ni, 2013). S-PDRF is an approach to assess driving risk, which employs a probabilistic motion prediction scheme, within the framework of artificial potential field theory (Mullakkal-Babu et al., 2020). This metric can estimate the driving risk continuously considering two important aspects of the driving risk: crash probability and crash severity. The computation of S-PDRF can be modeled as Eq. (2).

$$R_{n,s} = 0.5 M_s \beta^2 |\Delta v_{s,n}|^2 \bullet \mathrm{p(n,s)} \quad (2)$$

where $R_{n,s}$ is the collision risk between subject vehicle and a neighbor vehicle in Joules computed by S-PDRF. $|\Delta v_{s,n}|$ denotes the relative velocity between the subject vehicle and a neighbor vehicle. $\beta = \frac{M_n}{M_s + M_n}$ denotes the mass ratio, with $M_s$denoting the mass of subject vehicle and M_n denoting the mass of neighbor vehicle. $p(n,s)$ is the collision probability between subject vehicle and neighbor vehicle ranging from [0,1].

The S-PDRF estimates the probability of collision at a single future time instant; consequently, the collision probability is only related to the overlap in space. To estimate the collision probability at a future time, the predicted position of subject vehicle according to the motion state at the current time and the range of predicted positions of a neighbor are used. The risk considering the probable behavior of the neighbor vehicle is estimated by using a stochastic approach. The probability functions of acceleration variability can be estimated by treating acceleration signals as a random variable (Wagner et al., 2016). The acceleration variability is assumed to follow a Gaussian distribution (Ko et al., 2010). Hence, we can obtain the probability density function of collision as Eq. (3).

$$p(n,s|\tau) = N\left(\frac{\Delta X - \Delta V_X \tau}{0.5\tau^2} | \mu_X, \sigma_X\right) \bullet \mathrm{N}\left(\frac{\Delta \mathrm{Y} - \Delta \mathrm{V_Y} \tau}{0.5\tau^2} | \mu_\mathrm{Y}, \sigma_\mathrm{Y}\right) \quad (3)$$

where N is the collision probability density function. Parameters μ and σ denote the mean and standard deviation of the distribution in acceleration both longitudinally and laterally; $\Delta X$and ΔYdenote the relative spacing in X and Y direction separately between the subject vehicle and the neighbor vehicle; ΔV_X and ΔV_Y denotes the longitudinal relative velocity and lateral relative velocity. τ is the prediction time horizon.

In this study, only collision probability was considered when S-PDRF was computed. When the mean and the standard deviation of the acceleration distribution are 0 and 1 separately, the collision probability can be visualized as Fig. 2. When S-PDRF is greater than the predefined threshold, it means that the vehicle is unsafe. Another thing that should be noticed is that the dataset we used is only related to rear-end conflicts so that focusing on the longitudinal computation is sufficient in this study. Therefore, we omitted the lateral component of PDRF during computation to save the computing resources without losing any performance. Besides, S-PDRF that was used in this study was S-PDRF, of which the disadvantage is the potential risk within a prediction horizon may be lost. However, this drawback can be overcome with a shorter prediction horizon or by multi-step Probabilistic Driving Risk Field, which was introduced in (Mullakkal-Babu et al., 2020).

3.1.3. Deceleration rate to avoid a crash (DRAC) (Cooper and Ferguson, 1976)

DRAC is the squared differential speed between a following vehicle and its corresponding leading vehicle, divided by their closing gap. The leading vehicle is responsible for the initial action (i.e., braking for a traffic light or stop sign and changing lanes and/or accepting a gap), and the following is the vehicle immediately affected by the leading vehicle action and must respond to avoid dangerous interactions. The definition can be shown as Eq (4).

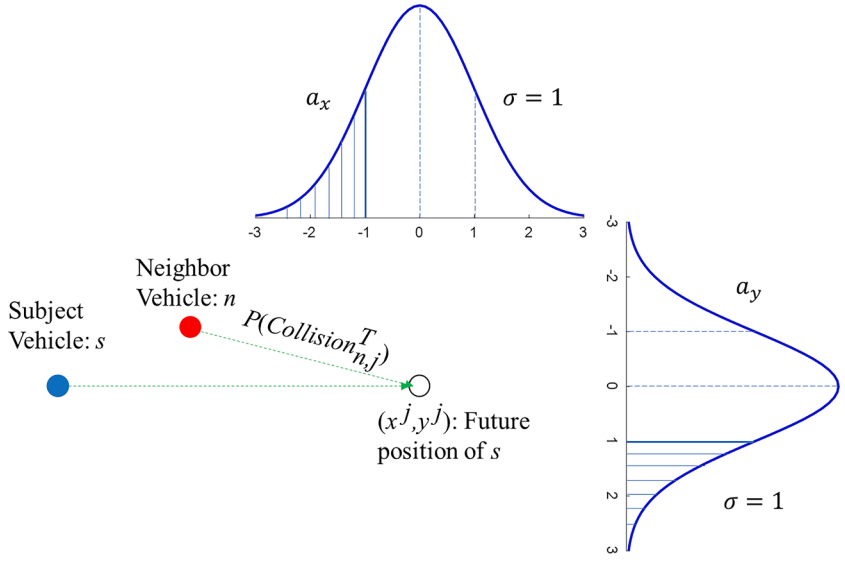

Fig. 2. Collision probability computed by PDRF.

$$DRAC = \begin{cases} \frac{(v_f - v_l)^2}{gap}, v_f > v_l \\ 0, v_f \le v_l \end{cases} \quad (4)$$

When DRAC is greater than the threshold, it is defined that the vehicle is unsafe. When DRAC exceeds the maximum braking deceleration, a collision is unavoidable by braking only.

3.1.4. Potential index for collision with urgent deceleration (PICUD) (Uno et al., 2002)

PICUD is defined as the distance between the two considered vehicles when they completely stop as shown in Eq. (5). Estimation of PICUD requires two predetermined parameters, namely: reaction time and deceleration rate.

$$PICUD = \frac{v_l^2 - v_f^2}{2\alpha} + gap - v_f \Delta t \quad (5)$$

where Δt is the reaction time of the rear vehicle, *gap* is the distance between the leading vehicle and the following vehicle. When PICUD is less than zero, it is defined that the vehicle is unsafe.

3.1.5. Proactive Fuzzy Surrogate safety metric (PFS) (Mattas et al., 2020)

The formulation of PFS is reported as Eqs. (6)–(8).

$$d_{safe} = v_f \Delta t + \frac{v_f^2}{2b_{f,comf}} - \frac{v_l^2}{2b_{l,max}} \quad (6)$$

$$d_{unsafe} = v_f \Delta t + \frac{v_f^2}{2b_{f,max}} - \frac{v_l^2}{2b_{l,max}} \quad (7)$$

$$PFS = \begin{cases} 1, gap \le d_{unsafe} \\ 0, gap \ge d_{safe} \\ \frac{gap - d_{safe}}{d_{unsafe} - d_{safe}}, gap \in (d_{unsafe}, d_{safe}) \end{cases} \quad (8)$$

with v_f being the speed of the rear vehicle, v_l the speed of the front vehicle, Δt the reaction time of the rear vehicle, *gap* the distance between the leading vehicle and the following vehicle, $b_{f,comf}$ the comfortable deceleration of the rear vehicle, $b_{f,max}$ the maximum deceleration of the rear vehicle and $b_{l,max}$ the maximum deceleration of the leading vehicle. The front vehicle's maximum deceleration must be considered at least as hard as the maximum deceleration of the rear vehicle. The assumption is that when the leading vehicle starts decelerating with the maximum possible deceleration, the following vehicle continues driving with constant speed for time Δt and then starts to decelerate with its comfortable deceleration. Both vehicles decelerate until they come to a stop, as this is the worst-case scenario. On the one hand, if the distance is enough for the following vehicle to stop and avoid the collision, then the distance is certainly safe. On the other hand, if the following vehicle after reaction time decelerates as hard as possible and still does not avoid an impact, the distance is certainly unsafe.

When PFS is greater than the threshold, it is defined that the vehicle is unsafe.

3.1.6. Critical Fuzzy Surrogate safety metric (CFS) (Mattas et al., 2020)

The formulation of CFS is reported as Eqs. (9)–(15).

$$a_f^{'} = max(a_f, -b_{f,comf}) \quad (9)$$

$$v_f^{'} = v_f + a_f^{'} \Delta t \quad (10)$$

if $v_f^{'} \le v_l$:

$$d_{safe} = d_{unsafe} = \frac{(v_f - v_l)^2}{2a_f^{'}} \quad (11)$$

else if if$v_f^{'} > v_l$

$$d_{new} = \left(\frac{v_f + v_f^{'}}{2} - v_l\right)\Delta t \quad (12)$$

$$d_{safe} = d_{new} + \frac{\left(v_f + a_f^{'}\Delta t - v_l\right)^2}{2b_{f,comf}} \quad (13)$$

$$d_{unsafe} = d_{new} + \frac{\left(v_f + a_f^{'}\Delta t - v_l\right)^2}{2b_{f,max}} \quad (14)$$

$$CFS = \begin{cases} 1, & gap \le d_{unsafe} \\ 0, & gap \ge d_{safe} \\ \frac{gap - d_{safe}}{d_{unsafe} - d_{safe}}, & gap \in (d_{unsafe}, d_{safe}) \end{cases} \quad (15)$$

The assumption is that if the leading vehicle keeps constant speed, the following vehicle continues driving with constant acceleration for time Δt, and then starts to decelerate with its comfortable deceleration. On the one hand, if the distance is enough for the following vehicle to stop before crashing, the distance is certainly safe. On the other hand, if the following vehicle decelerates as hard as possible after reaction time and still does not avoid an impact, the distance is certainly unsafe. The constant acceleration assumption for the follower during the reaction time separates the cases of the follower already accelerating or decelerating, with the former being more dangerous than the latter. When CFS is greater than the threshold, it is defined that the vehicle is unsafe.

3.2. Summary of SMoS

The features of all SMoS are summarized in Table 1.

Based on the features and component parameters of each SMoS, key parameters needed to be calibrated of each SMoS are shown in Table 2. Specific settings of other parameters are introduced in Section 4.

Fig. 3 describes how values of SMoS change, in terms of only relative speed and gap. The relative speed is equal to the speed of the following vehicle minus the speed of the leading vehicle. The gap represents the

Table 1
Features of SMoS.

SMoS	Attribute	Assumption of the leading vehicle	Assumption of the following vehicle
TTC	Tim-based	Keep current velocity	Keep current velocity
S-PDRF	Distance-based	The distribution of acceleration rate is Gaussian distribution (Mullakkal-Babu et al., 2020)	Keep current velocity
DRAC	Deceleration-based	Keep current velocity	Keep current velocity
PICUD	Distance-based	Brake with a proper deceleration	Brake with a proper deceleration
PFS	Distance-based	Brake with the maximum possible deceleration	Drive with constant speed for time Δt and then decelerate with deceleration less than maximum possible deceleration.
CFS	Distance-based	Keep current velocity	Brake with the maximum deceleration.

Table 2
Key Parameters Needed to Calibrated of Each SMoS.

SMoS	Key parameters needed to calibrated
TTC	The threshold time (s)
S-PDRF	Prediction horizon (s)
	Mean of the acceleration distribution of traffic vehicle (m/s^2)
	Standard deviation of the acceleration distribution of other vehicles (m/s^2)
	The threshold value
DRAC	Threshold deceleration rate (m/s^2)
PICUD	Deceleration rate to stop (m/s^2)
PFS	The threshold value
CFS	The threshold value

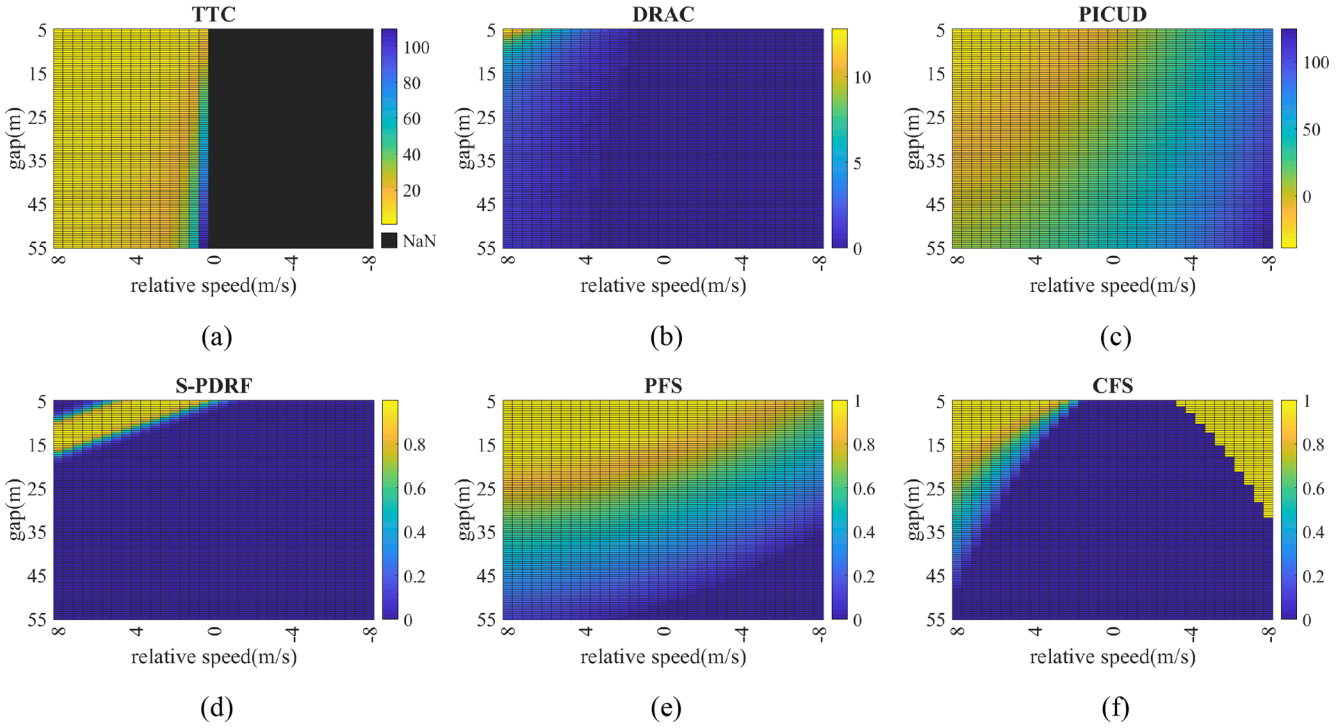

Fig. 3. Visualization of risk variation over different relative speeds and different gaps using different SMoS (yellow indicates higher risk). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

gap between the following vehicle and the leading vehicle. The speed of the following vehicle was set as 10 m/s (Mullakkal-Babu et al., 2017).

Risk distributions of all six SMoS are different. TTC is undefined when the following vehicle is slower than the leading vehicle, which results in fluctuations in risk during dynamic driving. PICUD and PFS have a smoother transition from "safe blue" to "unsafe yellow" than others. As shown in Fig. 3 (d), it is difficult for S-PDRF to detect risk with low gap and high relative speed at the same time. This is because, in the S-PDRF, the predicted subject vehicle position is ahead of the leading vehicle after a certain prediction step but still behind the leading vehicle in the previous step, which should not be counted. This problem can be resolved by shortening the prediction horizon of S-PDRF or using the multi-step Probabilistic Driving Risk Field. CFS may misjudge the risk status when the speed of the leading vehicle is much higher than the speed of the following vehicle.

3.3. Prediction Accuracy

As mentioned before, we classified car-following events into two categories. In a binary classification problem, there are four possible outcomes: true positive (*TP*), false positive (*FP*), true negative (*TN*), and false negative (*FN*). In this context, *TP* is the number of unsafe samples predicted as the unsafe class, *FP* is the number of safe samples predicted as the unsafe class, *FN* is the number of unsafe class samples predicted as the safe class, and *TN* is the number of safe samples predicted as the safe class.

In this paper, we used four widely-used classification evaluation indicators (Li et al., 2020a; Li et al., 2020b; Li et al., 2020c), including Precision, Recall, Accuracy and F1-score, to carry out the evaluation of the SMoS performance, which were defined as Eqs. (16)–(19) (Abou Elassad et al., 2020):

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{18}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{19}$$

Precision manifests how well the model predicts (i.e., a measure of exactness) and Recall manifests how well the model does not miss the target (i.e., a measure of completeness). F1-score is the weighted harmonic mean of the two and represents a realistic measure of SMoS

performance.

3.4. Timeliness

In this perspective, SMoS shall detect unsafe scenarios as early as possible.

For each car following event, the Timeliness indicator in this paper is defined as the time duration from the first unsafe timing to the end timing of each event. For each event, the Timeliness indicator is defined as (20):

$$Timeliness = t_{end} - t_{unsafe,first} \tag{20}$$

For crash events, the crash moment is the end timing of the event. For near-miss events, there is not a consensus definition about the most critical moment. The last moment in the recording is used as the end timing of the near-miss event. The mean Timeliness value of all events of each SMoS is used to represent the Timeliness performance. Therefore, the result of Timeliness can reflect the relative performance among SMoS.

3.5. Robustness

In real life, there may be a discrepancy between the real velocity of a surrounding vehicle and the value detected by sensors (Varghese and Boone, 2015). The error of the range rate may result in a misjudgment about the risk level of events.

Robustness of SMoS can be defined as the capability of maintaining the prediction accuracy while it is influenced by errors of the range rate. From the perspective of Robustness, a SMoS with good Robustness performance can identify the true risk level of each event affected by the error. In order to reduce the impact of the single random error on risk identification, a series of errors following the same distribution is generated. The mean of absolute values of differences between F1-scores under different range rate errors and the F1-score unaffected by errors are selected to represent the Robustness of SMoS. The computation method is as shown in Eq. (21).

$$Robustness = \frac{\sum_1^n |F1_i - F1_0|}{n} \tag{21}$$

where $F1_i$ and $F1_0$ represent F1-scores under different range rate errors and the F1-score unaffected by errors, respectively. A smaller Robustness represents a smaller influence caused by errors on risk detection results. Details of error generation will be explained in Section 4.4.

3.6. Efficiency

SMoS can be used in real-time collision warning systems, where the computation performance of SMoS is critical. Complex SMoS tend to take longer to compute. The Efficiency in this paper is defined as the computation time for each SMoS of all time steps in all events. The computation time is affected by the capability of the computer. Therefore, to ensure comparability, the computation should be done on the same computer. And values of Efficiency can only reflect relative relationships among SMoS. The smaller Efficiency represents a shorter computing time to compute SMoS.

4. Results and discussion

4.1. Threshold/Parameters calibration for SMoS

We used the selected 85 events to test the performance of the 6 different SMoS. Two samples out of the 85 events shown in Fig. 4 are selected to show the values of these 6 SMoS intuitively. In this figure, the first subplot shows the gap between the leading vehicle and the following vehicle. The other subplots show the six SMoS along with the time where the green marks represent the moments in low risk and the red marks represent the moments in high risk. It is defined that specific SMoS can just recognize a high-risk event if at least one moment in high risk is detected. The Timeliness of each event is the duration from the first moment marked with a red point to the end moment. After calibration, we found all 6 SMoS can identify all events in high risk. Therefore, all SMoS were kept for further analysis. The calibrated parameters for different SMoS are shown in Table 3.

As we can see in Table 3, the threshold for DRAC is very small (0.35 m/s^2). In this study, thresholds of each metric will be calibrated to maximize detection (correctly detect all events with highly unsafe moments) and to minimize false detections (incorrectly label events with low risk as dangerous events). Because of the calibration target, calibrated values are influenced by the data used for calibration. The calibrated values may be different from widely-used values in literature. There is one event with high risk in our dataset that can only be detected when the threshold for DRAC is very small. That is the reason why the calibrated threshold for DRAC is very small. Details of that event are as shown in Fig. 4 (a). The speed difference between the leading vehicle and the following vehicle is very small, but these two vehicles' speeds are fast. This scenario is always unsafe in reality. Due to the shortcoming of SMoS, which only considering imminent risk, these SMoS cannot easily detect this type of event. Similarly, the curve of TTC shows the same trend, and the threshold of TTC is bigger than widely-used values.

We did not calibrate the threshold value of PICUD. Because as described in Section 3, the threshold of PICUD is fixed, which means only when its value is smaller than 0 can it give a warning. Therefore, we calibrated another key parameter of PICUD instead. Additionally, the calibration results were based on the dataset we mentioned in Section 2.2 and the parameters of SMoS can be different regarding different datasets.

Additionally, the response time was set as 1 s (Green, 2000), the comfortable deceleration was set as 1 m/s^2 and the maximum deceleration was set as 6.8 m/s^2 which is given by the North Florida Transportation Planning Organization.

4.2. Prediction Accuracy

As the discussion above, we tested all 85 events with the 6 SMoS and the confusion matrix is shown in Table 4. The values of Precision, Recall, Accuracy, and F1-score were also computed according to Eqs. (16)–(19). The results are shown in Table 5.

It can be seen from Table 5 that the values of Recall for all 6 SMoS are 1. Because the calibration target is to ensure that the referred SMoS can recognize all unsafe events, which means that *FN* case should not appear. According to Eq. (17), the value of Recall should be 1 if *FN* is 0.

The Precision was only around 0.35, indicating that *FP* rate is relatively high. In other words, most of the events in low risk are identified as unsafe class. It should be noted that in order to achieve the calibration purpose, a high Recall rate has to be firstly guaranteed. Therefore, *FP* is not intendedly controlled during calibration when *FN* is strictly ensured to be 0. Another point is that the absolute value of Precision is not the most important comparability but meeting the same calibration criteria for all 6 SMoS should be guaranteed during calibration.

Similarly, according to Eqs. (16) and (18), if *TN* is bigger than 0, Accuracy is always slightly bigger than Precision but at the same level.

4.3. Timeliness

According to the discussion in Section 3.4 and Eq. (11), the results are shown in Table 6.

Timeliness reflects the warning time margin before the risk occurs. Generally speaking, the larger the value is, the earlier the risk can be recognized. The results show that the Timeliness values for all SMoS ranged from 5.24 s to 14.62 s, and PFS, PICUD, and CFS have better performances regarding warning time margin. However, it is noted that

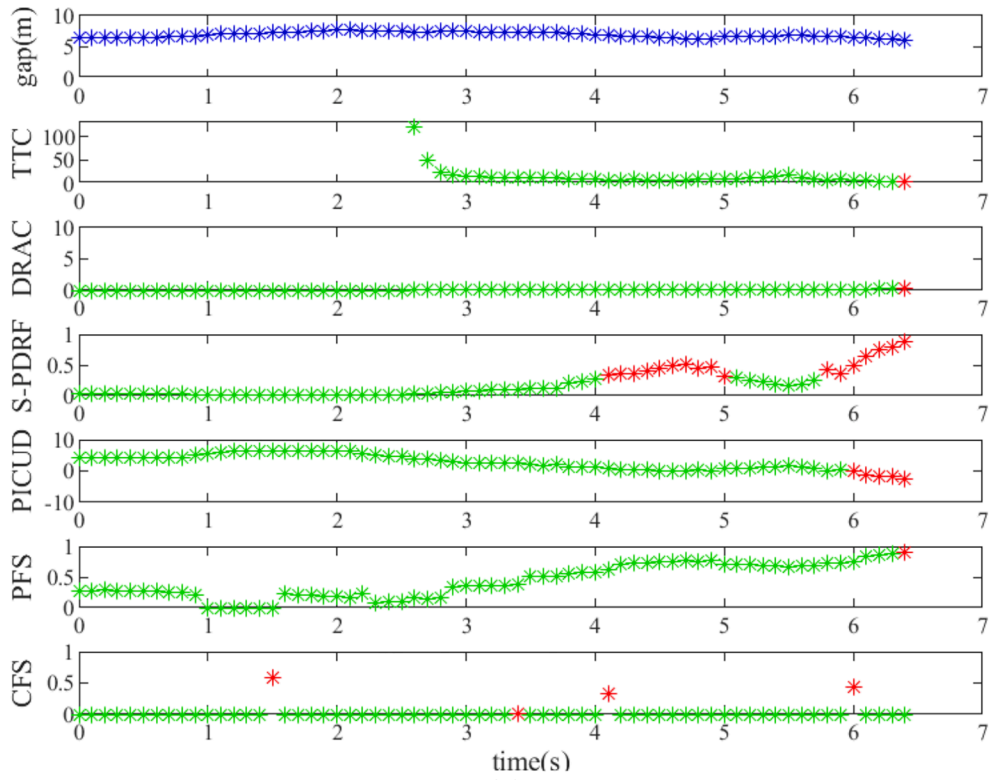

Fig. 4. The values of 6 SMoS in two sample events.

Table 3
Calibrated Parameters for Different SMoS.

SMoS	Parameters for calibration	Calibrated value
TTC	Threshold time(s)	4.5
S-PDRF	Prediction horizon(s)	1.5
	Mean of the acceleration distribution of traffic vehicle (m/s^2)	1
	Standard deviation of the acceleration distribution of traffic vehicle (m/s^2)	1
	Threshold value	0.3
DRAC	Threshold value of deceleration rate (m/s^2)	0.35
PICUD	deceleration rate to stop (m/s^2)	3.4
PFS	Threshold value	0.9
CFS	Threshold value	0.01

Table 4
Confusion Matrix.

SMoS	TP	FP	TN	FN
TTC	28	50	7	0
S-PDRF	28	50	7	0
DRAC	28	52	5	0
PICUD	28	55	2	0
PFS	28	54	3	0
CFS	28	53	4	0

Table 5
Performance of Prediction Accuracy.

SMoS	Precision	Recall	Accuracy	F1-score
TTC	0.3590	1	0.4118	0.5283
S-PDRF	0.3590	1	0.4118	0.5283
DRAC	0.3500	1	0.3882	0.5185
PICUD	0.3373	1	0.3529	0.5045
PFS	0.3415	1	0.3647	0.5091
CFS	0.3457	1	0.3765	0.5138

Table 6
Performance of Timeliness.

SMoS	Timeliness(s)	Standard deviation(s)
TTC	5.24	6.45
S-PDRF	6.68	7.66
DRAC	8.76	8.88
PICUD	13.95	11.66
PFS	14.62	12.48
CFS	10.93	10.68

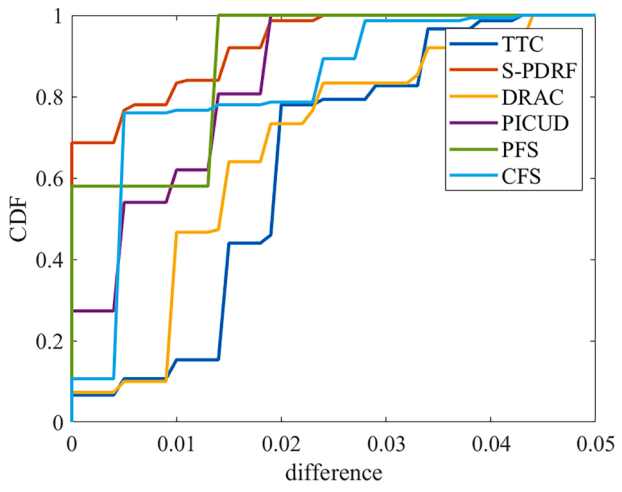

Fig. 5. Absolute values of differences between F1-scores under different range rate errors and the F1-score unaffected by errors.

results were obtained based on the calibrated parameters regarding the specific dataset so that the results can only reflect relative relationships among these SMoS. Additionally, all SMoS can have longer Timeliness if we tune their parameters at the cost of more *FP* events and a lower Precision, which will be discussed in Section 4.6.

4.4. Robustness

In this paper, we added the assumed error (like white noise) on range rate data collected by radar, which means the relative velocity between the following vehicle and the leading vehicle. The value distribution of the error obeys the normal distribution. The mean value ranges from −1m/s to 1 m/s and the interval is 0.1 m/s. The standard value ranges from 0 m/s to 1 m/s and the interval is 0.1 m/s. In order to minimize the effect of random factors, ten error samples were generated under each pair of the mean value and the standard value. Totally, we generated 2310 ($21 \times 11 \times 10 = 2310$) error samples. We identified the unsafe status for different events with errors again with these six SMoS. For each SMoS, 2310 F1-scores were obtained. The Cumulative Distribution Function (CDF) of absolute values of differences between F1-scores under different range rate errors and the F1-score unaffected by errors for all SMoS is shown in Fig. 5.

According to the discussion in Section 3.5 and Equation (18), we can compute the standard deviation of F1-score as the indicator to assess Robustness of different SMoS. The results are shown in Table 7.

The value of Robustness in this paper is essentially a standard deviation of risk detection rates influenced by random noise. Theoretically, if the value of Robustness is smaller, the risk detection rates of the referred SMoS are less affected by noise. Additionally, according to the error type selected in this study, SMoS that depend less on "relative velocity" directly will have a better performance in Robustness. The results in Table 7 correspond to the discussion above. DRAC, TTC, and CFS performed worse obviously because relative velocity is directly used in the computation of them. For other SMoS, the relative velocity is used indirectly, although they are related to relative velocity to other road users more or less.

4.5. Efficiency

The simulation for different SMoS should be conducted in the same device for the reason that the Efficiency was defined as the computation time for each time step. The experiments are conducted on a workstation with AMD Ryzen 9 5950X 16-Core Processor CPU and 128 GB RAM. The

Table 7
Performance of Robustness.

SMoS	Robustness
TTC	0.0186
S-PDRF	0.0037
DRAC	0.0163
PICUD	0.0080
PFS	0.0057
CFS	0.0090

Table 8
Performance of Efficiency.

SMoS	Computation time (ms)
TTC	0.23
S-PDRF	3226.24
DRAC	0.60
PICUD	24.54
PFS	296.63
CFS	82.82

Table 9
Ranking of SMoS by Performance Indicators.

SMoS	Prediction Accuracy	Timeliness	Robustness	Efficiency
TTC	1	6	6	1
S-PDRF	1	5	1	6
DRAC	3	4	5	2
PICUD	6	2	3	3
PFS	5	1	2	5
CFS	4	3	4	4

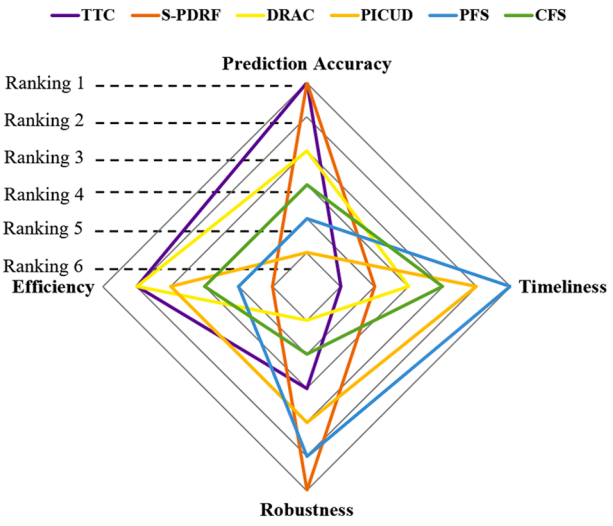

Fig. 6. Ranking of SMoS by performance indicators.

computation time of different SMoS is shown in Table 8.

SMoS considering more possible situations through a stochastic approach took much more time for computation. S-PDRF is the most complex as it requires spatial overlap computations and probability integration. Naturally, it is the most time-consuming one among all 6 SMoS. Conversely, TTC took the least time source as there is only one step to compute the value, and it is the simplest out of the 6 SMoS.

4.6. HYPERLINK "SPS:id::Sec9" Discussion

In order to have an overview of the performance of different SMoS, we ranked the performance of all mentioned SMoS regarding different performance indicators. The ranking results are shown in Table 9. We also visualize the results in Fig. 6.

As can be seen in Table 9, TTC has the best performance of Prediction Accuracy and Efficiency, which is the reason why TTC is popular and widely used in practice and research. However, TTC does not score that great on the performance indicators Robustness and Timeliness. S-PDRF has the best performance of Prediction Accuracy and Robustness. If the computation capability permits, using S-PDRF to predict the accidents or guide the system design is a good choice in our view. More in general, there are application-specific trade-offs to be considered. Computational Efficiency is clearly more relevant for real-time applications and is less an issue for off-line safety assessment. For example, in a real-time situation, a computation time of a second (as S-PDRF in Table 8) is unacceptable.

Arguably, Prediction Accuracy is one of the most important performance indicators. According to Eqs. (17) and (18), in case of Recall being with the value of 1, the biggest influencers to the value of Prediction Accuracy are false-positives (FP). PICUD and PFS performed poorly in terms of Prediction Accuracy. Because these two SMoS were considering the most critical braking condition, this might result in the overestimation of the potential risk. Therefore, some safe cases can be estimated as risky cases and there will be a bigger FP value, which can be found in Table 4. As discussed above, S-PDRF also considers potential risk, but the potential risk computation is more realistic. Therefore, S-PDRF can get a much better performance than PICUD and PFS. The other SMoS do not consider the potential risk in the future. Therefore, their performances are moderate.

The Prediction Accuracy of each SMoS was influenced by the threshold or key parameters. As for the 28 events with high risk in this study, the SMoS show different sensitivities with different values of the threshold and key parameters. S-PDRF and CFS can recognize all 28 high-risk events with an arbitrary threshold value from 0 to 1. PFS and PICUD can also detect all high-risk events with the proper threshold and key parameter range. However, TTC and DRAC are very sensitive to the threshold value and their performances are determined almost exclusively by the threshold value. Another factor that should be mentioned regarding S-PDRF is the sensitivity of its key parameters. There are more parameters within it compared to other SMoS. During the simulation, we found that the mean and especially the standard deviation of the normal distribution of neighbor vehicle's acceleration influenced the prediction performance substantially. Furthermore, the prediction performance was sensitive to the size of the prediction time horizon when the gap between the following vehicle and the leading vehicle is small. In such a case, a larger prediction horizon would always result in S-PDRF missing the collision, since a collision within a single prediction horizon would be ignored by S-PDRF. The upper left corner in Fig. 3 (d) shows the abovementioned situation. Therefore, a shorter prediction horizon or S-PDRF with multi-step prediction can resolve the problem, as discussed in Section 3.1.2 and (Mullakkal-Babu et al., 2020).

For Robustness, SMoS should not be too sensitive to the variance of inputs and it is better to give a relatively stable output corresponding to the change of the inputs. S-PDRF and PFS already consider uncertainty and prediction within their computation. Therefore, encountering the input noise, they have better performances of eliminating the influence of noise. Although CFS has a very similar form as PFS, uncertainty is not considered in the computation. Thus, CFS cannot perform as well as PFS regarding Robustness. This is the same reason that DRAC, PICUD and TTC also cannot perform well when noise is added to radar rate signal.

Timeliness reflects the warning time margin before the reference moment in each event. In Fig. 4, Timeliness is the time duration between the first red mark and the end of the event. Under this premise, Timeliness is effective to compare the warning time margin provided by SMoS. TTC and DRAC only consider the current status of the following vehicle and the leading vehicle. The change of the future status is not taken into account. In some cases, they cannot catch the risk decrease and gave a shorter warning margin accordingly, meaning that they overestimate the risk in the future. As for S-PDRF, in fact, it has a more 'realistic' Timeliness because the uncertainty was considered according to the actual traffic situation. For example, in Table 6, S-PDRF had a 6.68 s time margin which was relatively short among all SMoS. However, the fact is 6.68 s is a more reasonable value beyond which the values of Timeliness were somehow overestimated. Therefore, based on our definition of Timeliness, the performance of S-PDRF is not so outstanding. For the other SMoS, the prediction was included, and thus they perform better regarding Timeliness.

5. Conclusions

This paper develops an approach to evaluate the performance of SMoS. This method enables performance evaluation of SMoS along multiple dimensions, including Prediction Accuracy, Timeliness, Robustness, and Efficiency. To make a fair comparison among SMoS, a calibration method using naturalistic data with realistic car-following events is applied. We demonstrate how our method can be used to rationally compare 6 SMoS in terms of relative advantages and disadvantages. Connections between the performance of each SMoS and their features (definitions, assumptions, and their mathematical properties) are elaborately analyzed.

On the basis of the comparative results, we derived several findings: (1) Probabilistic driving risk field approach (S-PDRF) performs well in terms of Prediction Accuracy and Robustness at the expense of higher computational costs than all other SMoS; (2) SMoS that consider the potential risk (based on predictions), such as S-PDRF, PFS, and PICUD, have in principle a better performance in Accuracy, but how good the performance depends on the quality of their prediction about potential risk; (3) Performances of SMoS which considered the relative velocity can be significantly influenced by sensors noise. (4) Evaluating the performance of SMoS with a single indicator may create biases from the real findings. The proposed framework that supports evaluating SMoS along several dimensions is necessary.

The analysis clearly illustrates the inconsistency among all six SMoS in terms of 4 relevant and commonly used performance indicators. Although there are no quantitative criteria to give a comprehensive ranking for each SMoS, this study provides empirically underpinned arguments to prefer specific measures in cases where different performance indicators may be considered more important. The proposed framework for evaluating SMoS along these several dimensions provides a powerful tool to support those choices.

Apparently, SMoS winning on Timeliness (PFS, PICUD, and CFS) perform not great on Prediction Accuracy (TTC or S-PDRF). As both Timeliness and Accuracy are essential in real-time application, a combination of such SMoS can be explored in the future.

In addition, this analysis could be further improved by verifying the findings with more empirical accident data. Future research should also focus on developing an ensemble SMoS that considers the advantages of existing SMoS found in this study.

CRediT authorship contribution statement

Chang Lu: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Xiaolin He:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Hans van Lint:** Supervision, Writing – review & editing. **Huizhao Tu:** Supervision, Writing – review & editing. **Riender Happee:** Writing – review & editing. **Meng Wang:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the National Key R&D Program of China (2019YFE0108300) and Key Research Project from Shanxi Transportation Holdings Group (No. 20-JKKJ-1).

References

Elamrani Abou Elassad, Z., Mousannif, H., Al Moatassime, H., 2020. A proactive decision support system for predicting traffic crash events: A critical analysis of imbalanced class distribution. Knowledge-Based Syst. 205, 106314. https://doi.org/10.1016/j.knosys.2020.106314.

Bao, S., Leblanc, D.J., Sayer, J.R., Flannagan, C., 2012. Heavy-truck drivers' following behavior with intervention of an integrated, in-vehicle crash warning system: A field evaluation. Hum. Factors 54 (5), 687–697, 10.1177%2F0018720812439412.

Blas, E., Kurup, A.S., 2010. Equity, social determinants and public health programmes. World Health Organization.

Cooper, D.F., Ferguson, N., 1976. Traffic studies at T-Junctions. 2. A conflict simulation Record. Traffic Eng. Control 17.

Fazekas, A., Hennecke, F., Kalló, E., Oeser, M., 2017. A Novel Surrogate Safety Indicator Based on Constant Initial Acceleration and Reaction Time Assumption. J. Adv. Transp. 2017, 1–9. https://doi.org/10.1155/2017/8376572.

Gao, K., Tu, H., Sun, L., Sze, N.N., Song, Z., Shi, H., 2020. Impacts of reduced visibility under hazy weather condition on collision risk and car-following behavior: Implications for traffic control and management. Int. J. Sustain. Transp. 14 (8), 635–642. https://doi.org/10.1080/15568318.2019.1597226.

Green, M., 2000. "How Long Does It Take to Stop?" Methodological Analysis of Driver Perception-Brake Times. Transp. Hum. Factors 2 (3), 195–216. https://doi.org/10.1207/STHF0203_1.

Guido, G., Saccomanno, F., Vitale, A., Astarita, V., Festa, D., 2011. Comparing safety performance measures obtained from video capture data. J. Transp. Eng. 137 (7), 481–491. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000230.

Happee, R., Gold, C., Radlmayr, J., Hergeth, S., Bengler, K., 2017. Take-over performance in evasive manoeuvres. Accid. Anal. Prev. 106, 211–222. https://doi.org/10.1016/j.aap.2017.04.017.

Hayward, J.C., 1972. Near-Miss Determination Through Use of a Scale of Danger. Highw. Res. Board 384, 24–34.

Ko, J., Guensler, R., Hunter, M., 2010. Analysis of effects of driver/vehicle characteristics on acceleration noise using GPS-equipped vehicles. Transp. Res. Part F: Traff. Psychol. Behav. 13 (1), 21–31. https://doi.org/10.1016/j.trf.2009.09.003.

Kuang, Y., Qu, X., Wang, S., 2015. A tree-structured crash surrogate measure for freeways. Accid. Anal. Prev. 77, 137–148. https://doi.org/10.1016/j.aap.2015.02.007.

Li, L., Gan, J., Yi, Z., Qu, X.u., Ran, B., 2020a. Risk perception and the warning strategy based on safety potential field theory. Accid. Anal. Prev. 148, 105805. https://doi.org/10.1016/j.aap.2020.105805.

Li, P., Abdel-Aty, M., Cai, Q., Islam, Z., 2020b. A Deep Learning Approach to Detect Real-Time Vehicle Maneuvers Based on Smartphone Sensors. IEEE Trans. Intell. Transp. Syst. 1–10 https://doi.org/10.1109/TITS.2020.3032055.

Li, Y., Zheng, Y., Morys, B., Pan, S., Wang, J., Li, K., 2020c. Threat Assessment Techniques in Intelligent Vehicles: A Comparative Survey. IEEE Intell. Transp. Syst. Mag. https://doi.org/10.1109/MITS.2019.2907633.

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

Mattas, K., Makridis, M., Botzoris, G., Kriston, A., Minarini, F., Papadopoulos, B., Re, F., Rognelund, G., Ciuffo, B., 2020. Fuzzy Surrogate Safety Metrics for real-time assessment of rear-end collision risk. A study based on empirical observations. Accid. Anal. Prev. 148, 105794. https://doi.org/10.1016/j.aap.2020.105794.

Mullakkal-Babu, F.A., 2020. Modelling Safety Impacts of Automated Driving Systems in Multi-Lane Traffic. TU Delft University. https://doi.org/10.4233/uuid:37c8ec51-59f0-4dec-a3fa-bdd3c69e09db.

Mullakkal-Babu, F.A., Wang, M., Farah, H., van Arem, B., Happee, R., 2017. Comparative Assessment of Safety Indicators for Vehicle Trajectories on Highways. Transp. Res. Rec. J. Transp. Res. Board 2659, 127–136, 10.3141%2F2659-14.

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

Nadimi, N., Behbahani, H., Shahbazi, H.R., 2016. Calibration and validation of a new time-based surrogate safety measure using fuzzy inference system. J. Traffic Transp. Eng. (English Ed. 3 (1), 51–58. https://doi.org/10.1016/j.jtte.2015.09.004.

Ni, D., 2013. A Unified Perspective on Traffic Flow Theory Part I : The Field Theory. Appl. Math. Sci. 7 (39), 1929–1946. https://doi.org/10.1061/41186(421)420.

Papazikou, E., Quddus, M., Thomas, P., Kidd, D., 2019. What came before the crash? An investigation through SHRP2 NDS data. Saf. Sci. 119, 150–161. https://doi.org/10.1016/j.ssci.2019.03.010.

Shi, X., Wang, Z., Li, X., Pei, M., 2021. The effect of ride experience on changing opinions toward autonomous vehicle safety. Commun. Transp. Res. 1, 100003. https://doi.org/10.1016/j.commtr.2021.100003.

Songchitruksa, P., Tarko, A.P., 2006. Practical method for estimating frequency of right-angle collisions at traffic signals. Transp. Res. Rec. 1953, 89–97, 10.1177%2F0361198106195300111.

Stapel, J., Mullakkal-Babu, F.A., Happee, R., 2017. Driver behavior and workload in an on-road automated vehicle. Proc. Road Saf. Simul. Int. Conf. 1–10. http://resolver.tudelft.nl/uuid:53b8c5eb-a1ad-43f4-b349-286e8dffd8f1.

Tak, S., Kim, S., Lee, D., Yeo, H., 2018. A comparison analysis of surrogate safety measures with car-following perspectives for advanced driver assistance system. J. Adv. Transp. 2018, 1–14. https://doi.org/10.1155/2018/8040815.

Uno, N., Iida, Y., Itsubo, S., Yasuhara, S., 2002. A microscopic analysis of traffic conflict caused by lane-changing vehicle at weaving section. Proc. 13th Mini-EURO Conf. Handl. Uncertain. Anal. Traffic Transp. Syst, 10–13.

Varghese, J.Z., Boone, R.G., 2015. Overview of Autonomous Vehicle Sensors and Systems. Int. Conf. Oper. Excell. Serv. Eng. 178–191.

Virginia Tech Transportation Institute, 2014. VTTI Data Warehouse.

Vogel, K., 2003. A comparison of headway and time to collision as safety indicators. Accid. Anal. Prev. 35 (3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2.

Von Luxburg, U., 2007. A tutorial on spectral clustering. Stat. Comput. 17 (4), 395–416. https://doi.org/10.1007/s11222-007-9033-z.

Wagner, P., Nippold, R., Gabloner, S., Margreiter, M., 2016. Analyzing human driving data an approach motivated by data science methods. Chaos, Solitons and Fractals 90, 37–45. https://doi.org/10.1016/j.chaos.2016.02.008.

Wang, J., Wu, J., Zheng, X., Ni, D., Li, K., 2016. Driving safety field theory modeling and its application in pre-collision warning system. Transp. Res. Part C Emerg. Technol. 72, 306–324. https://doi.org/10.1016/j.trc.2016.10.003.

WHO, 2020. Road traffic injuries [WWW Document]. URL https://www.who.int/health-topics/road-safety#tab=tab_1.

Xiong, X., Wang, M., Cai, Y., Chen, L., Farah, H., Hagenzieker, M., 2019. A forward collision avoidance algorithm based on driver braking behavior. Accid. Anal. Prev. 129, 30–43. https://doi.org/10.1016/j.aap.2019.05.004.