sustainability

MDPI

Article

Towards Safe and Sustainable Autonomous Vehicles Using Environmentally-Friendly Criticality Metrics

Sorin Liviu Jurj *,†, Tino Werner †, Dominik Grundt †, Willem Hagemann † and Eike Möhlmann †

Institute of Systems Engineering for Future Mobility, German Aerospace Center e.V (DLR), Escherweg 2, 26121 Oldenburg, Germany; tino.werner@dlr.de (T.W.); dominik.grundt@dlr.de (D.G.); willem.hagemann@dlr.de (W.H.); eike.moehlmann@dlr.de (E.M.)

* Correspondence: sorin.jurj@dlr.de; Tel.: +49-441-770507-251

† These authors contributed equally to this work.

Abstract: This paper presents a mathematical analysis of several criticality metrics used for evaluating the safety of autonomous vehicles (AVs) and also proposes novel environmentally-friendly metrics with the scope of facilitating their selection by future researchers who want to evaluate both safety and the environmental impact of AVs. Regarding this, first, we investigate whether the criticality metrics which are used to quantify the severeness of critical situations in autonomous driving are well-defined and work as intended. In some cases, the well-definedness or the intendedness of the metrics will be apparent, but in other cases, we will present mathematical demonstrations of these properties as well as alternative novel formulas. Additionally, we also present details regarding optimality. Secondly, we propose several novel environmentally-friendly metrics as well as a novel environmentally-friendly criticality metric that combines the safety and environmental impact in a car-following scenario. Third, we discuss the possibility of applying these criticality metrics in artificial intelligence (AI) training such as reinforcement learning (RL) where they can be used as penalty terms such as negative reward components. Finally, we propose a way to apply some of the metrics in a simple car-following scenario and show in our simulation that AVs powered by petrol emitted the most carbon emissions (54.92 g of CO_2), being followed closely by diesel-powered AVs (54.67 g of CO_2) and then by grid-electricity-powered AVs (31.16 g of CO_2). Meanwhile, the AVs powered by electricity from a green source, such as solar energy, had 0 g of CO_2 emissions, encouraging future researchers and the industry to develop more actively sustainable methods and metrics for powering and evaluating the safety and environmental impact of AVs using green energy.

Keywords: autonomous vehicles; criticality metrics; safety; sustainability

Citation: Jurj, S.L.; Werner, T.; Grundt, D.; Hagemann, W.; Möhlmann, E. Towards Safe and Sustainable Autonomous Vehicles Using Environmentally-Friendly Criticality Metrics. *Sustainability* **2022**, *14*, 6988. https://doi.org/10.3390/su14126988

Academic Editors: Rosolino Vaiana and Vincenzo Gallelli

Received: 11 May 2022
Accepted: 4 June 2022
Published: 7 June 2022

1. Introduction

The research interest in the domain of AVs, especially regarding their safety, has grown exponentially in the last few years. This is mainly due to recent advancements in the field of AI, especially regarding deep RL algorithms, which are showing promising results when implemented in AI components found in AVs, especially when combined with prior knowledge [1].

Concerning traffic scenarios, the safety of all traffic participants is considered to be the most important aspect on which the researchers should focus, this being especially reflected by projects such as VVM - Verification and Validation Methods for Automated Vehicles Level 4 and 5 [2], SET Level - Simulation-based Development and Testing of Automated Driving [3], as well as KI Wissen - Automotive AI powered by Knowledge [4], all three projects being funded by the German Federal Ministry for Economic Affairs and Climate Action. In addition to these, many other projects of the VDA Leitinitiative autonomous and connected driving [5] bring together various research partners from the industry and academia to solve challenging and contemporary research problems related to the AV domain, emphasizing the relevance of criticality and safety in traffic.

With regards to the meaning of criticality, despite the existent ambiguity regarding its definition in both industry and academia, for an easier understanding of its meaning in the context of this paper, we follow the definition given by the work in [6], namely: the combined risk of the involved actors when the traffic situation is continued.

Regarding this, to assess how critical a traffic situation is, the existent research works found in the literature focus on the use of so-called criticality metrics for automated driving [7,8]. However, because AVs are operating in a complex traffic environment where a high number of actors are present, such as AVs, non-AVs, and pedestrians, to name only a few, it is imperative to not only identify the suitable criticality metrics that can mitigate dangerous situations as it is currently done in the literature [7,8] but also to implement and evaluate them efficiently regarding their environmental impact as well.

This is of high importance, especially when the transportation sector is known to be a key contributor to climate change, accounting for more than 35% of carbon dioxide emissions in the United States alone [9]. It is therefore imperative that existent and future researchers do not only use existent metrics that can evaluate critical situations in traffic, but also make efforts in proposing novel environmentally friendly criticality metrics that can be used to evaluate the AVs impact on the environment and economy as well. A recent effort in this direction is made by a new global initiative that tries to catalyze impactful research work at the intersection of climate change and machine learning such as the Climate Change AI [10] organization as well as in recent works that try to encourage researchers to power and evaluate their deep learning-based systems using green energy [11,12].

Therefore, in this paper, we present a mathematical analysis of 43 criticality metrics [7,8] to determine if they are well-defined as well as if they are working as intended within their scope to easily facilitate their selection for criticality assessment in the context of AV safety evaluation. Furthermore, due to recent emergent paradigms, such as Green AI [13], which encourage researchers to move towards more sustainable methods that are environmentally friendly and inclusive, we also propose several green metrics that are used to create a novel green-based criticality metric, which is suitable for evaluating a critical scenario not only regarding safety but also regarding the environmental impact in a car-following scenario.

Furthermore, from the perspective of AI training, we assess whether existing criticality metrics are suitable to serve as a component of the objective function, (which was already mentioned in [7] (Section 3.1.1)), for example, of the reward function in RL. To this end, the used criticality metrics must work as intended so that actions and, therefore, policies themselves that represent the desired behavior are flagged as optimal.

Our main contributions are: (i) analysis of existing criticality metrics in terms of well-definedness, intendedness, and optimality; (ii) integration of existing loss functions in RL and of emission estimations into the criticality metrics framework; (iii) investigation of the suitability of the criticality metrics for AI training; (iv) illustrative simulations of the metrics applied in a car-following scenario.

The paper is organized as follows. In Section 2, we present the related work. Section 3 details the mathematical analysis of 43 criticality metrics as well as the proposed novel formulas regarding making some of them work as intended. Section 4 presents the proposed green-based criticality metrics. Section 5 presents our contribution regarding the usage of criticality metrics for AI training. In Section 6 we present the application of the metrics. Finally, in Section 7, we present the conclusions, limitations and future work of this paper.

2. Related Work

An extensive overview of criticality metrics in autonomous driving has been given by the authors in [7,8].

The usage of criticality metrics is not restricted to the evaluation of traffic scenarios but can be extended to the training of autonomous driving agents by integrating suitable metrics into the reward function. This technique is called reward shaping and allows for prior knowledge to be included in the training, as seen in [14].

Three of these criticality metrics, namely Headway (HW), Time Headway (THW), and Deceleration to Safety Time (DST), were implemented and tested in an Adaptive Cruise Control (ACC) use case, as detailed by the authors in [1]. In their work, the authors have shown that different RL models can be evaluated for the ACC use case using these metrics, however, the DST metric at the very least does not coincide with the supposed objective of this function.

The ecological impact of autonomous driving has been discussed in many works such as the ones in [15–19]. These works do not only consider fuel consumption or emissions but also analyze the socio-ecological aspects like a higher driving demand if AVs are available or indirect implications like reduced land use due to optimized parking. Moreover, the work in [20] proposes a model for estimating the emissions and evaluating it in different scenarios w.r.t., for example, the relative part of AVs in the traffic.

The cited references generally consider the fuel consumption and the emissions for evaluation. These measures can be seen as green-based metrics, which have already been used for AI training. For example, the authors in [21] train a deep RL model that is encouraged to minimize emissions, and the authors in [22] propose a deep RL controller based on a partially observed Markov Decision Problem for connected vehicles so that eco-driving is encouraged where battery state-of-charge and safety aspects (e.g., speed limits or safety distances) are integrated into the model. Additionally, the work in [23] presents an extensive overview of eco-driving RL papers where the reward function is nearly always state-of-charge or fuel consumption. The authors in [24] propose a hybrid RL strategy where conflicting goals such as saving energy and accelerating are captured by a long-short term reward (LSTR). To not let energy-saving jeopardize safety, the acceleration energy is only penalized for accelerations, not for decelerations. The reward function also consists of a green-pass reward term, which essentially encourages reaching the stopping line of an intersection when the traffic light is green (i.e., driving forward-looking). Some of these references do not only focus on carbon dioxide emissions but also consider, for example, carbon monoxide, methane, or nitrogen oxides. Besides training AVs, ecological aspects are also taken into consideration regarding traffic system controls [25].

3. Mathematical Analysis of Criticality Metrics

In the following, we present the mathematical analysis of several criticality metrics. As for the notation in the subsequent parts, please see the Abbreviations and Nomenclature sections where the most frequent abbreviations and symbols used in this paper are presented.

The criticality metrics work presented in [7,8], serves as a collection of metrics regarding vehicle conduction. In this section of the paper, we evaluate, from a mathematical framework, two aspects of these criticality metrics.

Note that, as in the literature, we always refer to these criticality functions as criticality "metrics", although most of them are not metrics in the mathematical sense.

First, we determine if they are well-defined, meaning that there is no ambiguity in the interpretation of the function for all the input values that the metric covers. A good example of a not well-defined metric would be $f(x) = 1/x$ for $x \in [-1, 1]$, since at the value of $x = 0$ (which exists on the input range), it is not clear which value the function outputs.

Second, if the metric is well-defined, we evaluate if it works as intended. This requires a different analysis for each metric. An example of a metric that does not work as intended would be a metric that is supposed to give a distance between two vehicles positions p_1 and p_2, usually for $p_1, p_2 \in \mathbb{R}^2$, and is defined as follows:

$$DistanceBetweenVehicles(p_1, p_2) = p_1 + p_2$$

The metric is well defined, as any two vectors have a unique additional output, but a distance should decrease (or at least stay constant) in magnitude as $||p_1 - p_2||$ decreases, but that is not the case with the function $DistanceBetweenVehicles$.

Third, we discuss the requirements for the agent's behaviors that arise from the individual criticality metrics. The criticality metrics quantify the criticality of traffic scenes or traffic scenarios so that it would be a desirable property that they make different concrete scenarios comparable to determine which one (and with it, usually, which agent) behaved best.

In some cases, the well-definedness or the intendedness of the metric will be apparent, but in other cases, we present mathematical demonstrations of these properties. In some cases, we present alternative formulas for the metric. It is important to mention here also that for the criticality metrics that are not well-defined or do not work as intended, it is infeasible to replace them with new ones, and thus their intendedness cannot be analyzed.

Following, to make all of the criticality metrics analyzed in this paper more accessible, we organize them according to their scales such as time, distance, velocity, acceleration, jerk, index, probability, and potential. As for the target values of the individual metrics, we will borrow those collected in the supplementary web page [8] of the work in [7]. These target values are not necessarily desirable but are used for scenario classification.

Note that the position predicates $p(t)$ seen in the Nomenclature section of this paper implicitly refer to predictions via a dynamic motion model (DMM) if applied to future time points. As some metrics are defined by aggregating other metrics over time, it is also possible to apply such a DMM for each time point retrospectively. In principle, predictions can also be completed by the means of machine learning, as in quantifying the energy absorption of a bump shock absorber in the work presented in [26].

3.1. Time-Scale Criticality Metrics

3.1.1. Encroachment Time (ET)

The ET metric was proposed in [27]. It is supposed to measure the time that an actor A_1 takes to encroach a designated conflict area CA, and is defined as the difference between the time step t_{exit} where the vehicle leaves CA and the time step t_{entry} where it enters it:

$$ET(A_1, CA) = t_{\text{exit}}(A_1, CA) - t_{\text{entry}}(A_1, CA)\,. \tag{1}$$

The ET metric has no run-time capability.

Well Definedness

Since the function consists of a simple substraction, the ET metric is well-defined as long as t_{exit} and t_{entry} are well-defined, meaning that A_1 enters CA at some time $t_{\text{entry}}(A_1, CA)$ and exits CA at some time $t_{\text{exit}}(A_1, CA)$, which is reasonable to expect.

Intendedness

The metric works as intended: If A_1 enters CA at time $t_{\text{entry}}(A_1, CA)$ and exits CA at time $t_{\text{exit}}(A_1, CA)$, then the time spent in CA is $t_{\text{exit}}(A_1, CA) - t_{\text{entry}}(A_1, CA)$, as defined.

Optimality

There are no target values for ET. Generally, the ET and therefore the time in the critical area should be as short as possible but it can be misleading if the concrete situation does not allow for it, e.g., if the conflict area is an occluded intersection, which A_1 evidently should not pass with high speed.

3.1.2. Post Encroachment Time (PET)

The PET metric [27] intends to calculate the time gap between one actor (A_1) leaving and another actor (A_2) entering a designated conflict area. This assumes A_1 completely passes CA before A_2 enters it, and the metric is defined as follows:

$$PET(A_1, A_2, CA) = t_{\text{entry}}(A_2, CA) - t_{\text{exit}}(A_1, CA)\,. \tag{2}$$

The PET metric has no run-time capability.

Well Definedness

Since this function consist of a simple substraction, the PET metric is well-defined as long as t_{exit} and t_{entry} are well-defined.

Intendedness

The PET metric works as intended: If A_1 enters CA at time $t_{\text{entry}}(A_1, CA)$ and exits CA at time $t_{\text{exit}}(A_1, CA)$, and A_2 enters CA at time $t_{\text{entry}}(A_2, CA)$ and exits CA at time $t_{\text{exit}}(A_2, CA)$, then the time gap between A_1 exiting and A_2 entering is $t_{\text{entry}}(A_2, CA) - t_{\text{exit}}(A_1, CA)$, as defined.

Optimality

In the ideal situation where the conflict area is known and that A_2 can observe the end of the conflict zone and A_1, A_2 can modify its velocity and acceleration so that it enters the conflict area at the appropriate time step, at least approximately by using predictions for the future trajectories of A_1 until A_1 exits the conflict area.

3.1.3. Predictive Encroachment Time (PrET)

The work in [7] does not give an explanation for the metric, only the formula, so we try to understand its purpose based on:

$$PrET(A_1, A_2, t) = \min(\{|\tilde{t}_1 - \tilde{t}_2| \mid p_1(t + \tilde{t}_1) = p_2(t + \tilde{t}_2),\ \tilde{t}_1, \tilde{t}_2 \geq 0\} \cup \{\infty\})\,. \quad (3)$$

The PrET metric has the run-time capability.

Well Definedness

The formula for the metric is well defined since it's defined as the minimum of a non-empty set. It should be noted that this formula assumes knowledge of the positions of both vehicles at all times greater than t (or having a model to predict them).

Intendedness

Based on the formula, the intention of the metric seems to be to find the minimum time difference between both vehicles passing through the same point, for any point. If that's the case, the metric works as intended.

Optimality

To attain an optimal PrET, which we denote by $PrET^*$, the following agent has to modify its velocity and acceleration at each time step according to the leading agent. Although PrET is a scene level criticality metric (as seen in Figure 5 of the work presented in [7]), $PrET^*$ can be at least approximately attained for the whole scenario. Note that in the special case of a constant velocity v of the leading agent, the rear agent just has to maintain the optimal distance corresponding to $PrET^*$ while driving with the same velocity v.

3.1.4. Time Headway (THW)

The THW metric intends to calculate the time until actor A_1 reaches the position of a lead vehicle A_2. It is defined by the following formula:

$$THW(A_1, A_2, t) = \min\{\Delta t \geq 0 \mid p_1(t + \Delta t) = p_2(t)\}\,. \quad (4)$$

The THW metric has the run-time capability.

Well Definedness

The formula for the metrics is not well defined, since it would require knowledge of the position of A_1 at all times greater than t, or a method of predicting it that is not mentioned.

Intendedness

The metric would work as intended if the position of A_1 was known at all times greater than t. Since it would return the first value of the set of times when A_1 reaches the position of A_2.

Proposed Solution

We will propose two new formulas for the metric, one that predicts future positions of A_1 by assuming constant velocity (4a) and one assuming constant acceleration (4b):

$$THW(A_1, A_2, t) = \min\{\Delta t \geq 0 \mid p_1(t) + v_1(t)\Delta t = p_2(t)\} = (p_2(t) - p_1(t))/v_1(t)\,. \quad (4a)$$

$$THW(A_1, A_2, t) = \min\{\Delta t \geq 0 \mid p_1(t) + v_1(t)\Delta t + a_1\Delta t^2/2 = p_2(t)\}\,. \quad (4b)$$

where $v_1(t)$ is the speed of A_1 at time t, a_1 is the acceleration of A_1 at time t.

Optimality

Having a prediction model for both the agent's trajectories, the rear agent can modify its movements so that the target value for THW is attained.

3.1.5. Time To Collision (TTC)

The TTC metric intends to return the minimal time until vehicle A_1 and vehicle A_2 collide using an underlying one-track prediction model for both actors where d is a distance metric, w.l.o.g. the Euclidean distance.

$$TTC(A_1, A_2, t) = \min\left(\{\Delta t \geq 0 \mid d(p_1(t + \Delta t), p_2(t + \Delta t)) = 0\} \cup \{\infty\}\right) \quad (5)$$

The TTC metric is run-time capable.

Well Definedness

The metric is well-defined since it's defined as the minimum of a non-empty set, and we are provided a method to predict the positions of both actors.

Intendedness

The metric works as intended:
If $TTC = \infty$, then $\{\Delta t \geq 0 \mid d(p_1(t + \Delta t), p_2(t + \Delta t)) = 0\} = \emptyset$.
Therefore, at no point in time greater than time t will the two vehicles collide.
If $TTC = \hat{t}$, then $\hat{t} \in \{\Delta t \geq 0 \mid d(p_1(t + \Delta t), p_2(t + \Delta t)) = 0\}$.
Therefore, the vehicles collide at time $\hat{t}$. If there was another time t' such that $t' \geq t$ and $t' \leq \hat{t}$, then the vehicles would not collide at time t', since $\hat{t}$ is the minimum of such collision times.
Therefore, $\hat{t}$ is the first time at which the two vehicles collide.

Optimality

The TTC metric is rather conflictive with other criticality metrics as it does not guide the following agent. From the perspective of criticality metrics like THW, it would be desirable to keep an appropriate velocity-dependent distance from the leading agent. Of course, if the leading agent brakes, the TTC becomes finite due to the reaction time of the following agent. Although one can compare different braking maneuvers, the TTC values depend mostly on the braking behavior of the leading agent. From the pure TTC perspective, however, a high TTC value would be desirable although there are different target values for this metric. The implication to the rear agent would be to keep a sufficiently large distance from the leading agent. In principle, if the reaction time can be estimated, the optimal THW would be the sum of the target value of the TTC and the reaction time, at least in the ideal situation where the agent can perform the same braking maneuver as the leading agent, which would require the same or at least a similar vehicle type. Note that

some of the target values of TTC exceed those of THW, which would signify that the rear agent may not make sense as the rear agent clearly would violate the TTC requirements if it could not brake even more efficiently than the leading agent.

3.1.6. Time Exposed TTC (TET)

The TET metric [28,29] intends to measure the amount of time for which the TTC is below a given target value τ during a fixed time interval $[t_0, t_e]$, and is defined by the following formula:

$$TET(A_1, A_2, \tau) = \int_{t_0}^{t_e} \mathbf{1}_{TTC(A_1,A_2,t)\leq\tau} \mathrm{dt}\,. \tag{6}$$

The TET metric has a retrospective run-time capability.

Well Definedness

Even though not all indicator functions are integrable, we can assume that the sets defined by $TTC(A_1, A_2, t) \leq \tau$ are discretizable and therefore measurable, so with that assumption, the integral is well defined.

Intendedness

A Riemann integral of an indicator function is equal to the measure of the indicator set, so the metric works as intended.

Optimality

If a TTC of τ or less is critical, it would be optimal to have a TET of zero. Evidently, TET shares the disadvantages of TTC.

3.1.7. Time Integrated TTC (TIT)

TIT is supposed to aggregate the difference between the TTC and a target value τ in a given time interval $[t_0, t_e]$.

$$TIT(A_1, A_2, \tau) = \int_{t_0}^{t_e} \mathbf{1}_{TTC(A_1,A_2,t)\leq\tau}(\tau - TTC(A_1, A_2, t))\mathrm{dt}. \tag{7}$$

The TIT metric has a retrospective run-time capability.

Well Definedness

The formula for the metric is well defined assuming as in TET that the sets defined by $TTC(A_1, A_2, t) \leq \tau$ are discretizable and therefore measurable, and assuming that TTC is continuous on t since it would be an integral of the product of two integrable functions (therefore integrable).

Intendedness

The metric works as intended since by definition the integral is an aggregation function, and the integrand is a measure of the difference between the TTC and τ.

Optimality

TIT simply scales the TET, so our analysis for TET remains valid for TIT.

3.1.8. Potential Time To Collision (PTTC)

The PTTC metric has been proposed by [30]. However, it is important to mention that, in this paper, we focus on analyzing the PTTC version of the metric found in [7]. If A_1 travels at a constant speed v_1 and the leading agent A_2 decelerates at a constant rate a_2 (with starting speed v_2), then the PTTC metric is defined as follows:

$$PTTC(A_1, A_2, t) = \frac{1}{a_2}\left(-\dot{d} \pm \sqrt{\dot{d}^2 + 2d}\right) \tag{8}$$

for $d = p_2(t) - p_1(t)$ and $\dot{d} = v_2(t) - v_1$.

The PTTC metric is run-time capable.

Well Definedness

The metric is not well defined, since the "$\pm$" on the equation is never disambiguated.

Intendedness

The metric intends to return the value of the simple physics problem of finding the time of collision given the initial values at time t. By doing some arithmetic steps we will find the solution to this problem and compare it to the value of the metric.

If we name Δt as the amount of time elapsed from time t (time of measurements), we want to find for which value of Δt the projected distance between vehicles at time $t + \Delta t$ is 0:

$$p(t + \Delta t) = 0 \iff p_2(t) - p_1(t) + v_2(t)\Delta t - v_1\Delta t - \frac{a_2}{2}\Delta t^2 = 0 \tag{8a}$$

Applying the quadratic equation to the above equation, we get:

$$\Delta t = \frac{-(v_2(t) - v_1) \pm \sqrt{(v_2(t) - v_1)^2 - 4(-\frac{a_2}{2})(p_2(T) - p_1(t))}}{-2\frac{a_2}{2}} = \frac{-\dot{d} \pm \sqrt{\dot{d}^2 + 2a_2 d}}{-a_2} \tag{8b}$$

which gives different values than the proposed formula, so even if the $\pm$ was desambiguated the metric would not be working as intended.

Proposed Solution

We will propose a new formula for the metric, by disambiguating the equation:

$$\frac{-\dot{d} \pm \sqrt{\dot{d}^2 + 2a_2 d}}{-a_2} \tag{8c}$$

In order to desambiguate this equation, we will need to impose the condition that $\Delta t \geq 0$, therefore:

$$\frac{-\dot{d} \pm \sqrt{\dot{d}^2 + 2a_2 d}}{-a_2} \geq 0 \tag{8d}$$

Since $a_2 > 0$:

$$\iff -\dot{d} \pm \sqrt{\dot{d}^2 + 2a_2 d} \leq 0 \tag{8e}$$

$$\iff \pm\sqrt{\dot{d}^2 + 2a_2 d} \leq \dot{d} \tag{8f}$$

Now we have two distinct cases to analyze: $\dot{d} \geq 0$ and $\dot{d} < 0$. If $\dot{d} \geq 0$, then the equation is equal to:

$$\pm\sqrt{1 + \frac{2a_2 d}{\dot{d}^2}} \leq 1 \tag{8g}$$

Since $\sqrt{1 + \frac{4a_2 d}{\dot{d}^2}} > 1$ the "$\pm$" desambiguates as "$-$" and the solution is:

$$\Delta t = \frac{-\dot{d} - \sqrt{\dot{d}^2 + 2a_2 d}}{-2a_2} \tag{8h}$$

Now if $\dot{d} < 0$, then the equation is equal to:

$$-\pm\sqrt{1+\frac{2a_2d}{\dot{d}^2}} \geq 1, \tag{8i}$$

and the "$\pm$" desambiguates as "$-$" as well. Therefore the new formula for the metric is:

$$PTTC(A_1, A_2, t) = \frac{\dot{d} + \sqrt{\dot{d}^2 + 2a_2d}}{a_2}. \tag{8j}$$

Optimality

As PTTC is just a special case of TTC, our analysis for TTC remains valid for PTTC.

3.1.9. Worst Time To Collision (WTTC)

The WTTC metric intends to extend the usual TTC by considering multiple traces of actors as predicted by an over-approximating DMM. For the sets $Tr_1(t)$ and $Tr_2(t)$ of traces of actor 1 resp. actor 2 at time t, is defined as follows:

$$WTTC(A_1, A_2, t) = \min_{p_1 \in Tr_1(t), p_2 \in Tr_2(t)} (\{\Delta t \geq 0 \mid d(p_1(t+\Delta t), p_2(t+\Delta t)) = 0\} \cup \{\infty\}),. \tag{9}$$

The WTTC metric has the run-time capability.

Well Definedness

It is not clear how to obtain the traces of the actors.

3.1.10. Time To Maneuver (TTM)

The definition of the TTM metric in the original source found in [31] is different than the definition presented in [7]. It is important to mention that, in this paper, we use the definition of TTM found in [7].

The TTM metric is supposed to return the latest possible time in the interval $[0, TTC]$ such that an actor A_1 performing the considered avoidance maneuver m would lead to collision avoidance (or $-\infty$ is a collision is inevitable).

$$\begin{aligned} TTM(A_1, A_2, t, m) = \max\ (\{\tilde{t} \in [0, TTC(A_1, A_2, t)]\ \mid \\ d(p_{1,m}(t+s), p_2(t+s)) > 0\ \forall s \geq \tilde{t}\} \cup \{-\infty\}) \end{aligned} \tag{10}$$

where $p_{1,m}(t')$ denotes the position of actor 1 at some time $t' > t$ if maneuver m has been executed. The TTM metric has run-time capability.

Well Definedness

The metric is well-defined since it is defined as the maximum of a non-empty set, assuming we are provided a method to predict the positions of both actors.

Intendedness

We will prove by examplethat the metric is not working as intended. Let us assume that:

$$\hat{t} = TTM(A_1, A_2, t, m) < TTC(A_1, A_2, t)$$

and that $\hat{t} \neq -\infty$. Then:

$$\hat{t} \in \{\tilde{t} \in [0, TTC(A_1, A_2, t)]\ \mid\ d(p_{1,m}(t+s), p_2(t+s)) > 0\ \forall s \geq \tilde{t}\}.$$

Then:

$$d(p_{1,m}(t+s), p_2(t+s)) > 0\ \forall s \geq \hat{t}.$$

If $\hat{t} < t_1 < TTC(A_1, A_2, t)$ then $d(p_{1,m}(t+s), p_2(t+s)) > 0\ \forall s \geq t_1$ so,

$$t_1 \in \{\tilde{t} \in [0, TTC(A_1, A_2, t)] \mid d(p_{1,m}(t+s), p_2(t+s)) > 0 \; \forall s \geq \tilde{t}\}.$$

Since $t_1 > \hat{t}$ and also exists on the set, then $\hat{t}$ is not the maximum of the set, which is absurd. Therefore, $TTM(A_1, A_2, t, m) = TTC(A_1, A_2, t)$ or $TTM(A_1, A_2, t, m) = -\infty$, which proves the metric is not working as intended.

Proposed Solution

$$\begin{aligned} TTM(A_1, A_2, t, m) = \max\, (\{\tilde{t} \in [0, TTC(A_1, A_2, t)] \mid \\ d(\hat{p}_1(\tilde{t}, t+s), p_2(t+s)) > 0 \; \forall s \geq \tilde{t}\} \cup \{-\infty\}), \end{aligned} \tag{10a}$$

where $\hat{p}_1(\tilde{t}, t+s)$ is the predicted position of A_1 at time $t+s$ if A_1 started performing the maneuver at time $\tilde{t}$. The proposed solution is well-defined and works as intended.

Optimality

At the level of TTM, the agent cannot be reasonably guided as the evasion maneuver and should be executed as quickly as possible.

3.1.11. Time To Brake (TTB)

This section on the criticality metrics paper [7] is a reference to the TTM metric.

3.1.12. Time To Kickdown (TTK)

This section on the criticality metrics paper presented in [7] is a reference to the TTM metric.

3.1.13. Time To Steer (TTS)

This section on the criticality metrics paper [7] is a reference to the TTM metric.

3.1.14. Time To React (TTR)

The TTR metric aims to approximate the latest time until a reaction over a predefined set of maneuvers M is required. It is defined as follows:

$$TTR(A_1, A_2, t) = \max_{m \in M} TTM(A_1, A_2, t, m)\,. \tag{11}$$

The TTR metric has the run-time capability.

Well Definedness

The metric's properties hinge mostly on TTM's properties. If we assume TTM is well-defined, then TTR is well defined, since TTR is the maximum of a finite set of values.

Intendedness

Assuming TTM works as intended, then TTR also works as intended.

Optimality

See the TTM metric.

3.1.15. Time To Zebra (TTZ)

The TTZ metric intends to measure the time until actor A_1 reaches a zebra crossing CA, and is defined as follows:

$$TTZ(A_1, CA, t) = \min\left(\{\Delta t \geq 0 \mid d(p_1(t+\Delta t), p_{Zebra}(t+\Delta t)) = 0\} \cup \{\infty\}\right). \tag{12}$$

There is a small detail to correct in the definition of TTZ, which is that the zebra crossing position does not change with time, so $p_{Zebra}(t+\Delta t)$ should be p_{Zebra}, and the formula should be:

$$TTZ(A_1, CA, t) = \min\left(\{\Delta t \geq 0 \mid d(p_1(t+\Delta t), p_{Zebra}) = 0\} \cup \{\infty\}\right). \tag{12a}$$

The TTZ metric has the run-time capability.

Well Definedness

The definition does not offer a way to obtain or predict $p_1(t+\Delta t)$. Other than that it is well-defined.

Intendedness

Assuming $p_1(t+\Delta t)$ can be computed, the TTZ metric should work as intended.

Optimality

The TTZ metric solely measures the time needed until the zebra crossing is reached and is therefore useless as the agent has to attain and even cross it eventually.

3.1.16. Time To Closest Encounter (TTCE)

The TTCE is supposed to measure the time $\Delta t > 0$ for which the distance d to other actors in a scenario becomes minimal.

$$TTCE(A_1, A_2, t) = \arg\min_{\Delta t \geq 0} d(p_1(t+\Delta t), p_2(t+\Delta t))\,. \tag{13}$$

The TTCE metric has the run-time capability.

Well Definedness

The metric is not well-defined since there are eventually multiple times that the distance to the other actor becomes minimal, so it is ambiguous, which is the TTCE.

Proposed Solution

$$DCE(A_1, A_2, t) = \min_{\Delta t \geq 0} d(p_1(t+\Delta t), p_2(t+\Delta t))\,, \tag{13a}$$

$$TTCE(A_1, A_2, t) = \min\left(\{\Delta t \geq 0 \mid d(p_1(t+\Delta t), p_2(t+\Delta t)) = DCE(A_1, A_2, t)\}\right) \tag{13b}$$

where $DCE(A_1, A_2, t)$ is the closest distance the two actors achieve for times greater that t. DCE is well-defined since it is a minimum of $d(p_1(t+\Delta t), p_2(t+\Delta t))$, which is a lower-bounded function (the lower bound is 0). $TTCE$ is well-defined since, by definition of DCE, there is a time $\tilde{t}$ such that $DCE(A_1, A_2, t) = d(p_1(t+\Delta t), p_2(t+\Delta t))$, and therefore the set of times $\{\Delta t \geq 0 \mid d(p_1(t+\Delta t), p_2(t+\Delta t)) = DCE(A_1, A_2, t)\}$ is non-empty.

Intendedness

The proposed solution works as intended.

Optimality

Not applicable, as the distance itself defines the criticality and not the time step unless the DCE is already critical and there would be a concrete scenario where one can argue why attaining the DCE at the given time step TTCE is even more critical than attaining it at some other time step.

3.2. Distance-Scale Criticality Metrics

3.2.1. Headway (HW)

The HW metric is supposed to measure the distance to a lead vehicle, and is defined as follows:

$$HW(A_1, A_2, t) = d(p_1(t), p_2(t)) \,. \tag{14}$$

Given the simplicity of the concept and the formula, it is evident that it is well-defined, and works as intended. Note that d is usually the Euclidean distance, so HW is indeed a metric in the mathematical sense. The HW metric has the run-time capability.

3.2.2. Accepted Gap Size (AGS)

The AGS metric intends to quantify the gap or the actual space between actors desired or required for others to make a positive action decision.

It is defined as follows:

$$AGS(A_1, t) = \min\{s \geq 0 \,|\, action(A_1, t, s) = 1\}, \tag{15}$$

where $action(A_1, t, s)$ is a (complex) model predicting on a binary scale, based on the circumstances at time t, whether A_1 will come to a positive action decision for the gap size s. The run-time capability of AGS depends on the used model and inputs.

Well Definedness

The metric does not consider the cases where the set $\{s \geq 0 \,|\, action(A_1, t, s) = 1\}$ is empty, that is, the cases where the actor will never make a positive action decision at time t.

As such, its well-definedness hinges heavily on the properties of the action model.

Optimality

It completely depends on the states and the action space, so there are no implications for the agent.

3.2.3. Distance of Closest Encounter (DCE)

The section on the criticality metrics paper [7] is just a reference to the TTCE metric (which we already covered in this paper).

3.2.4. Proportion of Stopping Distance (PSD)

The PSD metric is defined as the distance to a conflict area CA divided by the minimum stopping distance (MSD):

$$\begin{aligned} PSD(A_1, CA, t) &= \frac{d(p_1(t), p_{CA}(t))}{\mathrm{MSD}(A_1, t)}, \text{ with} \\ \mathrm{MSD}(A_1, t) &= \frac{\|v_1(t)\|_2^2}{2|a_{1,long,min}(t)|} \,. \end{aligned} \tag{16}$$

Here, $a_{1,long,min}$ is the smallest acceleration (negative) available for actor A_1. It should be noted that no intention is provided for the metric, just its mathematical definition, so it will not be possible to decide if it works as intended. The PSD metric has the run-time capability.

Well Definedness

The metric is not well-defined, since it is not clear what the output would be when MSD is zero (and MSD is zero when v_1 is zero, which is possible).

Optimality

Not applicable, as it is not evident if the conflict area should be avoided or if it is inevitable.

3.3. Velocity-Scale Criticality Metrics

As for optimality, it would be optimal not to have any collision. Since the metrics concentrate on quantifying the impact of a (potential) collision, any maneuvers which lead to zero collision risk are optimal, therefore, one cannot expect to find a unique optimal maneuver.

3.3.1. Conflict Severity (CS)

The CS metric intends to estimate the severity of a potential collision in a scenario. It is defined as follows where the time to accident (TTA) is defined as $TTA(A_1, A_2) = TTC(A_1, A_2, t_{evasive})$ for the starting point $t_{evasive}$ of an evasive maneuver and where M_1 and M_2 are the masses of the vehicles of actor 1 resp. actor 2:

$$CS(A_1, A_2) = \Delta v(A_1, A_2, t_{evasive}) - \left(TTA(A_1, A_2) \cdot \|a_1(t_{evasive})\|_2 \cdot \frac{M_2}{M_1 + M_2} \right). \tag{17}$$

The CS metric has no run-time capability, as TTA can only be computed once an evasive maneuver has been identified.

Well Definedness

The metric is not well-defined, since it is not clear how the acceleration $a_1(t_{evasive})$ of actor 1 at $t_{evasive}$ is measured—it could be referring to the acceleration the vehicle had at time $t_{evasive}$, or the acceleration it took at that point in time as part of its evasive maneuver.

3.3.2. Delta-v (Δv)

The Δv metric is defined as the change in speed over collision duration, and is defined as follows for the velocity of actor 1 for the time step $t_{aftercol}$ after the collision and $t_{beforecol}$ before the collision:

$$\Delta v(A_1) = \|v_1(t_{aftercol})\|_2 - \|v_1(t_{beforecol})\|_2 \tag{18}$$

$$\Delta v(A_1, A_2, t) = \frac{M_2}{M_1 + M_2} (\|v_2(t)\|_2 - \|v_1(t)\|_2) \tag{18a}$$

The Δv metric has the run-time capability.

Well Definedness

The metric is well-defined as it is just a vector norm subtraction.

Intendedness

The metric (as exemplified in equation 18) works as intended since its intention is imposed in its formula. The intention of the equation 18a is not given.

3.4. Acceleration-Scale Criticality Metrics

3.4.1. Deceleration to Safety Time (DST)

For an actor A_1 following another actor A_2, the DST metric intends to calculate the deceleration required by A_1 in order to maintain a safety time of $t_s \geq 0$ seconds under the assumption of constant velocity v_2 of actor A_2. Since DST requires it, it is defined as follows:

$$DST(A_1, A_2, t, t_s) = \frac{(v_1(t) - v_2)^2}{2(d(p_1(t), p_2(t)) - v_2 \cdot t_s)}. \tag{19}$$

The DST metric has the run-time capability.

Well Definedness

The metric is not well-defined since $d(p_1(t), p_2(t))$ could be equal to $v_2 \cdot t_s$ and it would not be clear how to divide by their difference (zero) in that case.

Intendedness

Even ignoring the edge case of $d(p_1(t), p_2(t)) = v_2 \cdot t_s$, in cases where $d(p_1(t), p_2(t)) < v_2 \cdot t_s$, then:

$$\frac{(v_1(t) - v_2)^2}{2(d(p_1(t), p_2(t)) - v_2 \cdot t_s)} <= 0,$$

even in cases where $v_1 >> v_2$ and would require A_1 to decelerate urgently in order to avoid a collision.

Optimality

As DST computes a required deceleration, it implicitly poses restrictions on the agent's behavior as it should ensure that the required acceleration is always comfortable for the person in the vehicle.

3.4.2. Required Longitudinal Acceleration ($a_{long,req}$)

For two actors A_1, A_2 at time t, $a_{long,req}$ is supposed to measure the average negative longitudinal acceleration required by actor A_1 to avoid a collision in the future. It is defined by the following formula:

$$a_{long,req}(A_1, A_2, t) = \max\{a_{1,long} \leq 0 \mid \forall\, \Delta t \geq 0 : d(p_1(t + \Delta t), p_2(t + \Delta t)) > 0\}. \tag{20}$$

The $a_{long,req}$ metric has a run-time capability.

Well Definedness

The metric is well-defined as long as a method for predicting the positions of the vehicles is provided.

Intendedness

The equation finds the minimum deceleration needed, not the average as intended.

Optimality

Similarly, as for DST, this required acceleration should be comfortable.

3.4.3. Required Lateral Acceleration ($a_{lat,req}$)

The metric is intended to provide the minimal absolute lateral acceleration in either direction that is required for a steering maneuver to evade collision. It is defined as follows:

$$a_{lat,req}(A_1, A_2, t) = \min\{|a_{1,lat,left}(A_1, A_2, t)|, |a_{1,lat,right}(A_1, A_2, t)|\} \tag{21}$$

where:

$$a_{1,lat,k}(A_1, A_2, t) = a_{2,lat,k} + \frac{2(v_{2,lat}(t) - v_{1,lat}(t))}{TTC(A_1, A_2, t)} + \frac{2}{TTC(A_1, A_2, t)^2} \cdot \left[sign\left(\frac{W_1 + W_2}{2}\right) + p_{2,lat}(t) - p_{1,lat}(t) \right]$$

where W_i denotes the width of A_i and where $k \in \{left, right\}$. Here, $p_{\cdot,lat}$ and $v_{\cdot,lat}$ denote the lateral position resp. velocity component of the respective actor. The $a_{lat,req}$ metric has run-time capability.

Well Definedness

The metric is not well-defined, since $a_{2,lat,k}$ is not defined, as it cannot use the definition of $a_{1,lat,k}$, or the definition would be circular, and thus again not well-defined.

Optimality

Similarly, as for DST, this required acceleration should be comfortable.

3.4.4. Required Acceleration (a_{req})

The metric is defined as follows:

$$a_{req}(A_1, A_2, t) = \sqrt{a_{long,req}(A_1, A_2, t)^2 + a_{lat,req}(A_1, A_2, t)^2}\,. \tag{22}$$

It should be noted that no intention is given for the definition of a_{req}, so it is not possible to analyze its intendedness. The a_{req} metric has a run-time capability.

Well Definedness

The metric is merely a norm, so assuming both accelerations are well-defined, it is well-defined.

Optimality

See the required lateral and longitudinal acceleration metric.

3.5. Jerk-Scale Criticality Metrics

As for optimality, the jerks clearly should be comfortable.

3.5.1. Lateral Jerk (LatJ)

This section refers to the longitudinal jerk metric, which will be evaluated in the next section.

3.5.2. Longitudinal Jerk (LongJ)

Jerk is the rate of change in acceleration, and is defined as follows:

$$LatJ(A_1, t) = j_{1,lat}(t), \quad LongJ(A_1, t) = j_{1,long}(t) \tag{23}$$

for the longitudinal resp. lateral jerk $j_{1,long}(t)$ resp. $j_{1,lat}(t)$ of actor 1 at time t. The metrics are just extracted from the input, so they are well-defined and work as intended by definition. The LongJ metric has a run-time capability.

3.6. Index-Scale Criticality Metrics

As for optimality, the following metrics ACI, AM, CI, and CPI depend on the ego agent's policy (not on concrete actions) and also on the transition model. Therefore, these metrics can be understood on a policy level so that an optimal policy would achieve very low values of these metrics here.

3.6.1. Accident Metric (AM)

The AM metric intends to evaluate whether an accident happened in a scenario Sc:

$$AM(Sc) = \begin{cases} 0 & \text{no accident happened during } Sc, \\ 1 & \text{otherwise.} \end{cases} \tag{24}$$

The ACI metric has no run-time capability.

Well Definedness

The function is well-defined though it does not provide a method of determining if there was an accident.

Intendedness

It is clear that by definition the metric works as intended.

Optimality

Having no accident is optimal.

3.6.2. Brake Threat Number (BTN)

For actor A_1, the BTN metric is defined as the required longitudinal acceleration imposed on actor A_1 by actor A_2 at time t, divided by the longitudinal acceleration that is at most available to A_1 in that scene, i.e.,

$$BTN(A_1, A_2, t) = \frac{a_{long,req}(A_1, A_2, t)}{a_{1,long,min}(t)}. \tag{25}$$

The BTN metric has the run-time capability.

Well Definedness

Assuming the minimum longitudinal acceleration is less than 0, which is reasonable, the metric is well-defined. However, the metric will not be defined for cases where the vehicle is broken down and not moving.

Intendedness

An intention is not given, just the definition of the metric, so it is not possible to analyze its intendedness.

Optimality

If the value of the metric is at least 1, it is not possible to avoid a collision by braking, so BTN has to be always smaller than 1.

3.6.3. Steer Threat Number (STN)

The STN metric is defined as the required lateral acceleration divided by the lateral acceleration that is at most available to actor A_1:

$$STN(A_1, A_2, t) = \frac{a_{lat,req}(A_1, A_2, t)}{a_{1,lat,min}(t)}. \tag{26}$$

The STN metric has the run-time capability.

Well Definedness

Assuming the minimum lateral acceleration is less than 0, which is reasonable, the metric is well-defined. However, the metric will not be defined for cases where the vehicle is broken and can not steer.

Intendedness

An intention is not given, just the definition of the metric, so it is not possible to analyze its intendedness.

Optimality

Similarly, as for BTN, a value of at least 1 indicates that the agent cannot avoid a collision by steering.

3.6.4. Conflict Index (CI)

The CI metric intends to enhance the PET metric with a collision probability estimation as well as a severity factor, and is defined as follows:

$$CI(A_1, A_2, CA, \alpha, \beta) = \frac{\alpha \Delta K_e}{e^{\beta PET(A_1, A_2, CA)}} \tag{27}$$

Here, β is a calibration factor that depends on the scenario properties. The parameter $\alpha \in [0, 1]$ is another calibration factor that represents the relative part of the energy that affects the passengers while ΔK_e is the absolute change in kinetic energy acting on the vehicle's body before and after the predicted collision. The CI has no run-time capability, as PET can only be determined a posteriori.

Well Definedness

The metric is not well-defined, since ΔK_e is not known, and a way to estimate it is not provided.

Intendedness

It is not clear the specific intention of the metric, more than to give a general idea of how likely a crash is weighted by the severity of the eventual crash, which would by definition work as intended assuming everything is well-defined.

3.6.5. Crash Potential Index (CPI)

The CPI metric intends to calculate the average probability that a vehicle can not avoid a collision by deceleration, and is defined as follows:

$$CPI(A_1, A_2) = \frac{1}{t_e - t_0} \int_{t_0}^{t_e} P(a_{long,req}(A_1, A_2, t) < a_{1,long,min}(t)) \mathrm{dt}\,. \tag{28}$$

The CPI metric has no run-time capability.

Well Definedness

Assuming the integral is well-defined, the metric is well-defined.

Intendedness

The metric works as intended by definition.

3.6.6. Aggregated Crash Index (ACI)

The ACI metric intends to measure the collision risk for car-following scenarios and it is defined as follows:

$$ACI(S) = \sum_{j=1}^{n} CR_{L_j}(S), \tag{29}$$

where, given a collision tree derived from a probabilistic causal model, the concrete outcomes are represented by the tree's leaf nodes L_j. Every leaf node has a value C_{L_j} which is 0 in the case of no collision and 1 in the case of a collision. None-leaf nodes in the tree represent conditions that may occur during the scenario. Similar to CPI, the conditions are

defined based on other metrics, e.g., the current stopping time of the lead vehicle being smaller than a lognormally distributed reaction time. The collision risk $CR_{L_j}(S)$ of a leaf node L_j given a scene S is hence represented by $CR_{L_j}(S) = P(L_j) \cdot C_{L_j}$, where $P(L_j)$ is the probability of satisfying all conditions necessary to reach L_j in the collision tree when given the current conditions in the scene S. The ACI metric has no run-time capability.

Well Definedness

The well-definedness of the metric hinges completely on the well-definedness of CR.

Intendedness

Since each leaf represents an independent event, the addition of the probabilities adds up to the total probability of a crash, so it works as intended.

3.6.7. Pedestrian Risk Index (PRI)

The PRI metric intends to estimate the conflict probability and severity for pedestrian crossing scenarios, and is defined as follows:

$$PRI(A_1, CA) = \int_{t_{c_{start}}}^{t_{c_{stop}}} (s_{imp}(A_1, CA, t)^2 \cdot (t_s(A_1, t) - TTZ(A_1, CA, t)))\mathrm{dt}, \tag{30}$$

where:

$$\forall\, t \in [t_{c_{start}}, t_{c_{stop}}] : TTZ(P, CA, t) < TTZ(A_1, CA, t) < t_s(A_1, t)$$

and $t_s(A_1, t)$ is the time A_1 needs to come to a full stop at time t, including its reaction time.

The PRI metric theoretically has the run-time capability, but is primarily designed for a posteriori analysis.

Well Definedness

It is a reasonable assumption that the integrand is continuous, and therefore the integrand and in turn, the metrics are well-defined.

Intendedness

The PRI as defined does not estimate the conflict probability nor the severity of the conflict, but a combination of estimations for both.

Optimality

This metric makes little sense as it assumes that there exists a conflict period where the agent and the pedestrian collide. It would make more sense to inspect the time steps before to derive why the agent failed to avoid this situation.

3.6.8. Responsibility Sensitive Safety Dangerous Situation (RSS-DS)

The RSS-DS metric is intended for the identification of a dangerous situation S with a set of actors $\mathcal{A}$ and is defined as:

$$RSS\text{-}DS(A_1, \mathcal{A}) = \begin{cases} 1 & \exists A_i \in \mathcal{A} \setminus \{A_1\} : d^{lat}(A_1, A_i) < d^{lat}_{min} \wedge d^{long}(A_1, A_i) < d^{long}_{min} \\ 0 & \text{otherwise,} \end{cases} \tag{31}$$

where the safe lateral and longitudinal distances d^{lat}_{min} and d^{long}_{min} are formalized, depending on the current road geometry. The RSS-DS metric has the run-time capability.

Well Definedness

The metric is well-defined since the proposition conditioning the function is well-defined.

Intendedness

The metric is not working as intended. The paper never mentions a preferred actor A_1, which is the only actor that the metric makes sure is at a safe distance from others.

3.6.9. Solution

We can fix the issue by redefining the metric as:

$$RSS\text{-}DS(\mathcal{A}) = \begin{cases} 1 & \forall A \in \mathcal{A} : P_A, \\ 0 & \text{otherwise,} \end{cases} \tag{31a}$$

where:

$$P_A := \exists A_i \in \mathcal{A} \setminus \{A\} : d^{lat}(A, A_i) < d^{lat}_{min} \wedge d^{long}(A, A_i) < d^{long}_{min}$$

Optimality

It has to be understood here w.r.t. all actors so they have to jointly coordinate their maneuvers to achieve the optimal value 0 of RSS-DS.

3.6.10. Space Occupancy Index (SOI)

For a given scenario in the time interval $[t_0, t_e]$, the CI is defined as:

$$SOI(A_1, \mathcal{A}) = \sum_{t=t_0}^{t_e} C(A_1, \mathcal{A}, t). \tag{32}$$

where one defines a so-called personal space $Sp(A_i, t)$ for each actor and each time step and checks for a given time step t whether there is an overlap, i.e., a violation of any personal spaces of actor 1 and any other actor. Formally, the number $C(A_1, \mathcal{A}, t)$ of conflicts w.r.t. actor 1 in time step t is then defined by:

$$C(A_1, \mathcal{A}, t) = \sum_{A_j \in \mathcal{A} \setminus \{A_1\}} 1_{Sp(A_1,t) \cap Sp(A_j,t) \neq \emptyset}.$$

The SOI metric has a retrospective run-time capability.

Well Definedness

If we understand $Sp(A_i)$ as $Sp(A_i, t)$ and we assume that both t_0 and t_e are integers where $t_0 < t_e$ the metric is well-defined as set intersections are well-defined and so are finite sums.

Intendedness

The intention of the metric is not stated and not clear enough to verify if it works as intended.

Optimality

Same as mentioned before for the RSS-DS metric.

3.6.11. Trajectory Criticality Index (TCI)

The TCI metric intends to find a minimum difficulty value, i.e., how demanding even the easiest option for the vehicle will be under a set of physical and regulatory constraints. It is defined as follows:

$$TCI(A_1, S, t, t_H) = \min_{a_{1,long}, a_{1,lat}} \sum_{\tilde{t}=t}^{t+t_H} w_{long} R_{long}(\tilde{t}) + w_{lat} R^2_{lat}(\tilde{t}) + \frac{w_{long} a^2_{1,long}(\tilde{t}) + w_{lat} a^2_{1,lat}(\tilde{t})}{(\mu_{max} g)^2} \tag{33}$$

for some prediction horizon t_H and for the maximum coefficient of friction μ_{max} and the gravitational constant g. w_{long} and w_{lat} are w weights and R_{long} and R_{lat} margins for angle corrections, formally:

$$R_{long}(t) = \frac{\max(0, x(t) - r_{long}(t))}{d_{long}(t)}, \quad R^2_{lat}(t) = \frac{(y(t) - r_{lat}(t))^2 v(t - \Delta t)}{d^2_{lat}(t) v_{max}}$$

for the longitudinal and lateral positions $x(t)$ and $y(t)$ is the position, some step size Δt, the maximum velocity v_{max}, the longitudinal reference distance $r_{long}(t) = 2\text{s} \cdot v(t)$, the position $r_{lat}(t)$ with the highest lateral distance of all obstacles in the scene and for the maximum longitudinal and lateral deviations $d_{long}(t)$ and $d_{lat}(t)$ from r_{long}, r_{lat}.

The TCI metric theoretically has the run-time capability but is not designed for active trajectory control.

Well Definedness

The metric is not well-defined, since s is not defined and it is not clear how to compute $d_{long}(t)$, $d_{lat}(t)$ since the maximization set is not given.

Optimality

Of course, it is desirable to have low difficulty values where the concrete values depend on the concrete situation.

3.7. Probability-Scale Criticality Metrics

3.7.1. Collision Probability via Monte Carlo (P-MC)

The P-MC metric intends to produce a collision probability estimation based on future evolutions from a Monte Carlo path planning prediction and is defined as follows:

$$\text{P-MC}(A_1, S, t) = P(\mathcal{C}) = \int P(\mathcal{C} \mid \mathcal{U}) P(\mathcal{U}) d\mathcal{U} \tag{34}$$

where:

$$P(\mathcal{U}) := P(u_1, \ldots, u_k) := \prod_{j=1}^{k} P(u_j)^{\alpha_j},$$

$P(\mathcal{C} \mid \mathcal{U})$ is the collision probability of A_1 in S under inputs from some controller input set $\mathcal{U}$.

The P-MC metric has the run-time capability.

Well Definedness

The metric is not well-defined, since α_j is not defined and it is not clear how to compute $P(\mathcal{C} \mid \mathcal{U})$ or solve the integral.

3.7.2. Collision Probability via Scoring Multiple Hypotheses (P-SMH)

The P-SMH metric intends to assign probabilities to predicted trajectories and accumulate them into a collision probability. It is defined as follows:

$$\text{P-SMH}(A_1, \mathcal{A}, t) = \sum_{i=1}^{N} \sum_{j=1}^{M} \chi^i_j p_{A_1,i} p_{(\mathcal{A} \setminus A_1),j} \tag{35}$$

where χ^i_j equals one if and only if the i-th trajectory of A_1 and the j-th trajectory of the actors in $\mathcal{A} \setminus A_1$ lead to a collision, and $p_{A_1,i}$ resp. $p_{(\mathcal{A} \setminus A_1),j}$ are the probabilities of the trajectories being realized.

The P-SMH metric has the run-time capability, as has been demonstrated by evaluation in [32].

Well Definedness

The metric is not well-defined: a way to obtain trajectories is not given, and neither is a way of getting the probabilities of each trajectory.

3.7.3. Collision Probability via Stochastic Reachable Sets (P-SRS)

The P-SRS metric intends to estimate a collision probability using stochastic reachable sets and originates from [33]. Assuming a discretized controller input space and state space, let $p^h(t_k)$ denote the probability vector of the states reached in time step t_k for input partition h. These probability vectors are updated by a Markov chain model. The goal is to approximate the probability of a crash.

First, ref. [33] (Section V.B) shows how to compute the probability vectors w.r.t. time intervals $[t_k, t_{k+1}]$ given $p^h(t_k)$ for all input partitions h. By respecting vehicle dynamics, road information, speed limits, and interactions of the agents, they eventually compute the probability for a path segment e being attained in some interval $[t_k, t_{k+1}]$, denoted by $p_e^{path}([t_k, t_{k+1}])$. As the vehicles may not exactly follow the paths, the authors in [33] additionally model the lateral deviations from the paths, denoted by $p_f^{dev}([t_k, t_{k+1}])$, indicating the probability that the deviation from the path lands in some interval D_f where they assume that the probability is constant for intervals D_f in which the whole deviation range is discretized. Assuming that the path and deviation probabilities are independent, the actual position $p_{ef}^{pos} = p_e^{path} p_f^{dev}$ can be computed for each time interval and agent, enabling to compute the probability of crashes by summing up all the probabilities for cases where the vehicle bodies "overlap".

The P-SRS metric has the run-time capability but needs precomputations.

Well Definedness

The metric can only be computed if all of the required components, starting from the dynamics and the Markov chain model, are given, making it hard to apply in practice.

3.8. Potential-Scale Criticality Metrics

3.8.1. Potential Functions as Superposition of Scoring Functions (PF)

Thie PF metric does not have a unique definition, being open-ended. It just suggests specifying potential functions (open-ended as well) and applying a method of one's choosing to the combined potential function.

3.8.2. Safety Potential (SP)

The SP metric intends to measure how unsafe, with regards to collision avoidance, a situation is; it is defined as follows:

$$SP(A_1, A_2, t) = \rho_{1,2} = \|(t_{stop}(A_1) - t_{int}, t_{stop}(A_2) - t_{int})\|_k \tag{36}$$

where $k \in \mathbb{Z}_{>0} \cup \{\infty\}$ and where t_{int} requires a short-time prediction model of the trajectories and refers to the first time step of an intersection while $t_{stop}(A_i)$ denotes the time where actor i has achieved a full stop. The SP metric has the run-time capability.

Well Definedness

The metric is not well-defined since it is not clear how to compute $t_{stop}(A_i)$ or t_{int}.

4. Proposed Green-Based Criticality Metrics

Considering the importance of climate change and recent efforts in the literature to propose methods that can reduce the amount of CO_2 emissions, in this chapter, we propose several novel metrics that are then being used to propose a novel green-based criticality metric that combines not only the environmental impact but also the safety in a car-following scenario.

4.1. Average Car CO_2 Emissions Per KM

According to the European Environment Agency [34], the average CO_2 emissions per km of a diesel-powered car is 127.0 g $\frac{CO_2}{km}$ while being 127.6 g $\frac{CO_2}{km}$ for a petrol-powered car. Furthermore, according to [35], the average CO_2 emissions per km of an electric car was 43% lower than the average CO_2 emissions per km of a diesel-powered car, therefore, 72.4 g $\frac{CO_2}{km}$.

4.1.1. Car CO_2 Emissions (CCO2E)

The CCO2E metric measures the amount of grams of CO_2 emitted by an average car on a given drive. We define:

$$\text{CCO2E}(d) := \begin{cases} 0 & \text{if the vehicle is powered by green energy} \\ d \cdot 127 \text{ g } CO_2 & \text{if the vehicle is diesel-powered} \\ d \cdot 127.6 \text{ g } CO_2 & \text{if the vehicle is petrol-powered} \\ d \cdot 72.4 \text{ g } CO_2 & \text{if the vehicle is powered by electricity from the grid} \end{cases} \tag{37}$$

where d is the distance travelled by the vehicle in the drive in kilometers.

4.1.2. Green Energy CO_2 Emissions Saved (GECO2ES)

The GECO2ES metric measures how much energy is saved in an electric vehicle by using green energy:

$$\text{GECO2ES}(d) \begin{cases} d \cdot 72.4 \text{ g } CO_2 & \text{if the vehicle is powered by green energy} \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

where d is the distance travelled by the vehicle in the drive in kilometers.

4.1.3. CO_2 Emissions Weighted Safety Distance (CO2EWSD)

With the previous metrics defined in this chapter, we now create a novel green-based criticality metric called CO2EWSD.

The CO2EWSD metric combines safety and environmental impact in a car-following scenario by measuring how safely the vehicle behind is driven (by measuring the percentage of time being at a distance greater than the given safety distance), and weighting it by the CO_2 emissions of the drive (by the vehicle behind).

It is defined as follows:

$$\text{CO2EWSD}(V_1, V_2, sd) = \frac{1}{1 + \text{CCO2E}(\int_{t_0}^{t_e} speed(V_1, t)dt)} \frac{1}{t_e - t_0} \int_{t_0}^{t_e} 1_{d(V_1,V_2,t)>sd} dt \tag{39}$$

where V_1 and V_2 are the vehicles in the scenario, V_1 being the agent vehicle behind V_2, sd is the given safety distance, t_0 and t_e are the start and end times of the scenario, and $d(V_1, V_2, t)$ is the distance between the vehicles at time t. $speed(V_1, t)$ if the absolute speed of V_1 at time t and therefore $\int_{t_0}^{t_e} speed(V_1, t)dt$ the total distance travelled by V_1 in the drive.

Thus,

$$\text{CCO2E}(\int_{t_0}^{t_e} speed(V_1, t)dt)$$

is the CO_2 emissions of the drive.

$$\frac{1}{1 + \text{CCO2E}(\int_{t_0}^{t_e} speed(V_1, t)dt)} \in (0, 1]$$

increases when the CO_2 emissions of the drive are lower and $\int_{t_0}^{t_e} 1_{d(V_1,V_2,t)>sd} dt$ is the amount of time the vehicle is driving at a distance greater than sd to the leading vehicle, so $\frac{1}{t_e-t_0} \int_{t_0}^{t_e} 1_{d(V_1,V_2,t)>sd} dt$ is the ratio of time spent driving at a safe distance of the total time.

5. Usage of Criticality Metrics for AI Training

In AI training, criticality metrics can be used as penalty terms, for example, as negative reward components in RL. The agent therefore successively learns to select appropriate actions, resulting in maneuvers that are not critical or in which criticality is sufficiently low, evaluating the selected metrics. As an action usually only considers acceleration/deceleration and changing the heading angle, parameters that cannot be influenced by the agent like payloads, the length of the vehicle, or generally its structure could only implicitly be considered when computing the rewards, e.g., higher payloads can be integrated into the computation of the braking distance. These parameters often correspond to passive safety and optimizing them is part of the manufacturing process [26], but does not correspond to the scope of this work.

Based on the analysis in the previous sections, we can conclude that many metrics are not useful for AI training or at most in a very limited way. These include PTTC, AGS, CS, Δv, TCI, PF, PRI, TTZ, TTCE, PSD, and WTTC (as the latter considers the worst trajectory while an RL agent selects the best one in the rollout during training) and, as for the ones with limited applicability, TTM/TTR and their relatives (one could quantify the difference between the latest time step computed by the metric and the time step where the actual maneuver/reaction happens) as well as ET (if one had information about the longest time an agent should spend in a conflict area so that one would penalize a longer stay).

Probability-scale metrics consider the whole policies and therefore may be used for safe RL training [36], provided that the required probability models are available. The same holds for ACI and CI.

The most suitable criticality metrics that can be turned into a reward component are THW and HW, which can be used in combination so that HW defines a minimum distance that could potentially be violated when using THW in the context of very low velocities. Note that rollouts are essentially nothing but samples of future trajectories so that it would suffice to be able to roll the other agent's trajectory out here, even without a prediction model for the future trajectories of the non-ego agent. The lateral and longitudinal jerk can be compared with comfortable values for these parameters so that maneuvers that would lead to a too strong jerk would be penalized accordingly. TET and TIT can be adapted to finite-horizon rollouts by replacing the integral with a sum. These metrics can be conflictive with HW or THW, therefore, the reward shaping must be conducted carefully.

Metrics that can also be used and work more implicitly include the required (lateral/longitudinal) acceleration. In dangerous traffic situations, some of the rollouts may include collisions while other ones are collision-free. However, one could inspect the deceleration that was necessary to prevent the collision and penalize these trajectories as well (of course, with a way smaller magnitude than those which led to a collision) so that the smoothest trajectories with still acceptable decelerations without collisions would be executed. This would similarly work for BTN and STN while BTN is more intuitive as STN would be valid only in situations where a simple deceleration would not work. AM corresponds to a simple (constant) collision penalty.

Metrics that could, in principle, be used (but at least with caution) include TTC (due to conflict with other metrics), PrET (conflictive with and less informative than THW as it does not take the dynamics at the closest time difference into account), PET (for example, if only one vehicle should be in the conflict area so that the ego vehicle must learn to decelerate or even stop before it if another vehicle is still located there, i.e., the trajectories that would lead to the ego vehicle entering if it is not allowed would be penalized) and DST (if computable, check whether the required deceleration would be comfortable).

The criticality metrics RSS-DS and SOI are not applicable if one trains a single agent as they depend on the movements of the other vehicles but can enter joint training of multiple

agents where one would penalize dangerous situations when at least two vehicles are too close.

Note that by the duality of reward terms and corresponding criticality metrics, one can also interpret reward terms like the Yaw loss [37], which penalizes non-optimal headings or the Off-road loss [38], which penalizes if the agent drives in the non-drivable area as criticality metrics.

5.1. Off-Road Loss

The off-road loss considers a whole trajectory $((x_1, y_1), \ldots, (x_H, y_H))$ for a given actor and computes the smallest Euclidean distance to the drivable area. Denoting $(u(x, y), v(x, y))$ as the nearest point in the drivable area, the off-road loss is given by:

$$\frac{1}{H} \sum_{h=1}^{H} ||(x_h, y_h) - (u(x_h, y_h), v(x_h, y_h))||_2 \tag{40}$$

which, provided that the nearest points can directly be identified, has run-time capability.

5.1.1. Well Definedness

Provided that the nearest points exist, which they do if there is a drivable area, the off-road loss is well-defined.

5.1.2. Intendedness

The intention is to penalize off-road trajectories, which is accomplished by this metric.

5.1.3. Optimality

An optimal trajectory that is entirely part of the drivable area receives an off-road loss of zero, therefore, such trajectories (which form an uncountably large set) are optimal.

5.2. Yaw Loss

The yaw loss again considers a whole trajectory $((x_1, y_1), \ldots, (x_H, y_H))$ for a given actor and quantifies deviations from the angle to the angle of the nearest lane. The angle corresponding to two consecutive waypoints (x_i, y_i) and (x_{i+1}, y_{i+1}) is given by $\theta_i = \theta(x_i, x_{i+1}, y_i, y_{i+1}) = \arctan((x_2 - x_1)/(y_2 - y_1))$. Denoting the angle of the nearest lane in time step i by θ_i^*, the yaw loss is the accumulated difference between θ_i^* and θ_i. Note that [37] (Equation (6)) implies that the difference is non-zero, which contradicts their definition in [37] (Equation (3)). We suggest to use:

$$\sum_{h=1}^{H-1} (\theta_i - \theta_i^*)^2 \tag{41}$$

as off-road loss for the whole trajectory. Note that the work in [37] also considers the yaw loss for intersections and for lane change where a pre-defined interval of heading differences is allowed so that the yaw loss is zero if the heading during lane change is contained in this interval.

Well Definedness

Provided that the nearest lane can be detected, the reference heading θ_i^* can be computed in run-time.

Intendedness

The intention is to penalize deviations from a pre-scribed heading, which is accomplished by this metric.

Optimality

An optimal trajectory is achieved if the heading always coincides with the desired heading resp. if the heading during a lane change and turn maneuvers is contained in a suitable interval. Note that again uncountably many optimal trajectories exist.

Note that one can assume some kind of transitivity of the metrics in the sense that if the behavior of the agent was good in one metric, it is very likely to also be good in some other metric, which facilitates the training as it can be regarded as a pre-selection of metrics and, therefore, of corresponding reward components. We can identify the following transitivity relations:

- Avoiding collisions, i.e., achieving a good AM, implies a good CS and Δv.
- Learning comfortable maneuvers, i.e., achieving a good LatJ and LongJ, implies that its behavior will also be good when evaluated in metrics like STN, BTN, $a_{long,req}$, and $a_{lat,req}$.
- Training according to AM combined with at least LatJ and LongJ, may be enhanced with distance-keeping metrics like THW, HW, or ACC, the agent is expected to drive forward-looking and therefore smoothly, so it should also achieve a good ACI, PSD, and DST as well as TTM and its variants.

Concerning the relation of classical criticality metrics and green-based metrics, it is important to note that the fuel/energy consumption and, therefore, the emissions, are depending on the driving behavior, for example, due to air resistance increasing quadratically with the velocity. The classical criticality metrics encourage the agent to drive forward-looking, avoiding large and unnecessary accelerations, and therefore saving energy. The green-based metrics are undeniably important for evaluation, but it would be hard to integrate them into training itself for numerous reasons like, by the argumentation above, reducing the air resistance would just correspond to an upper-velocity limit, that the agent cannot control the fuel type for a given vehicle or because it would be very difficult to compute the actual energy consumption, which would amount to knowing the friction between the wheels and the road, the weight, the shape of the car, i.e., how streamlined it is. Hence, we suggest using the green-based metrics mainly for evaluation at this point, while achieving ecological goals with a clever selection of criticality metrics from Section 3 for training.

Of course, performing AI training with reward terms corresponding to only a sparse selection of criticality metrics does neither exclude nor hinder an evaluation of the trained agent in terms of all criticality metrics.

6. Application of the Metrics

In this section, we will propose a way to apply the metrics in a simple scenario. To give the research community a chance to easily test and use the metrics considered in this section, we also implemented them in an HTML page using the newly released PyScript [39], which is a JavaScript module that allows the writing of Python code directly in HTML. The source code for the implementation is available at the following link: To be completed with a link later. The advantage of using PyScript over plain JavaScript is the ability to use existing Python code, including its dependencies, directly in the web browser.

6.1. Scenario

The scenario we will use to apply the metrics consists of a car-following scenario in which an agent has to follow a leading vehicle in a straight path.

The lead vehicle will accelerate until reaching a speed of 15 m/s and then try to maintain this speed.

The agent will aim to stay at a safe distance from the leading vehicle. We will use a simple heuristic for the agent: if the leading vehicle is faster it will accelerate, otherwise, it will not.

6.2. Methodology

The way we approached this was the following: first, we coded using the Python programming language, for the scenario we described using the Carla simulator. Then, we collected the data from the simulation. Finally, we used the data to compute the value of the selected metrics.

6.3. Data

In the Table S1 file in the Supplement Data to this paper, we show the collected data from the simulation: positions and velocities at each step for both vehicles. Here, a data point was collected every 0.03 s (a step) for a total of 600 steps. The data in the table itself was too raw to draw any conclusions, but this data can be used to debug the pipeline, making sure the vehicles start where they should and progress accordingly.

However, in most cases, one would rather take a look at graphs of these values. In Figures 1 and 2, we observe these graphs.

Figure 1. Positions of agent (blue color) and the leading vehicle (red color) over each step from the simulation in meters.

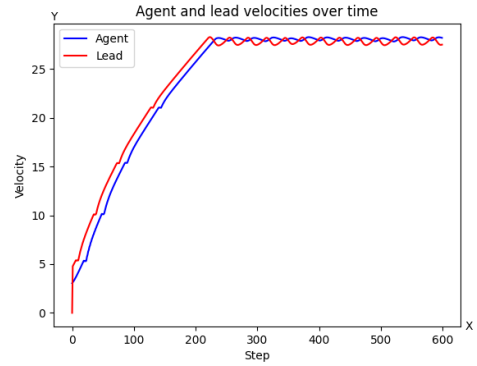

Figure 2. Velocities of agent (blue color) and the leading vehicle (red color) over each step from the simulation in m/s.

In Figure 1, we observe that the positions of the vehicles increase as they progress on the road, with the agent following from behind the leading vehicle. In Figure 2, we observe how at the beginning both agents accelerate, and at around step 200, the leading vehicle starts stabilizing at its target speed, and the agent vehicle follows suit.

6.4. Metrics

In the Table S2 file found in Supplement Data to this paper, we show a few step-by-step metrics obtained from the data, namely the HW, THW, TTC, and RSS-DS. Once again, with this data, we verify that the vehicles never collided as RSS-DS is always false. As is often the case, these values are more useful in the form of graphs.

Regarding this, in Figure 3, we see the HW increases as both agents accelerate, but then it starts decreasing, showing a possible flaw in the behavior of the agent since, if time continued, the value of the metric will go to 0.

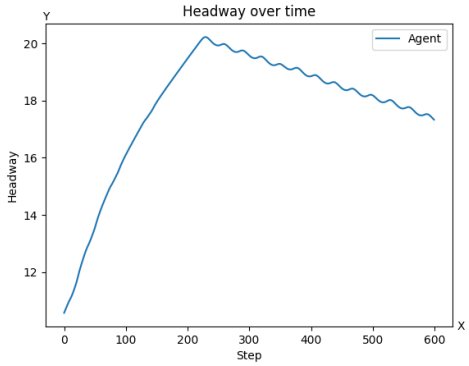

Figure 3. HW step by step from the simulation in meters.

In Figure 4, we can see the values of the THW metric that are less than infinity.

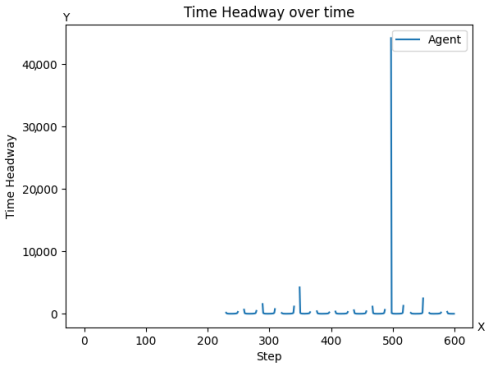

Figure 4. THW step by step from the simulation in seconds, assuming constant speeds. At times when the agent is not faster than the leading vehicle, the value is infinite (and therefore not shown).

Here, we observe intermittent gaps in the graph as the vehicle with more velocity is changing periodically in this scenario.

In Figure 5, we can observe the graph of the TTC metric.

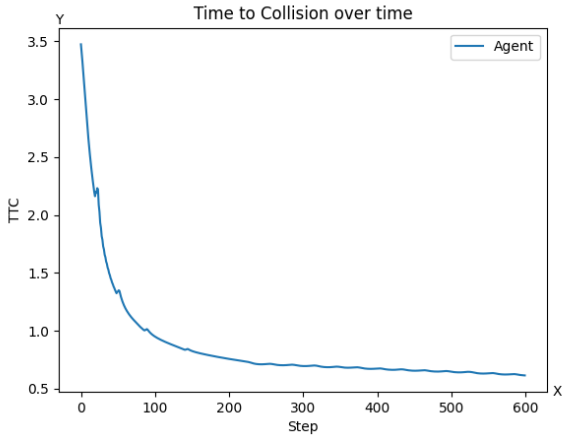

Figure 5. TTC over each step of the simulation in seconds, assuming a fixed position from the leading vehicle.

In this case, we consider TTC as the time the agent would take to crash against the leading vehicle if the agent maintained its current speed and the leading vehicle suddenly stopped moving. Here, we can see the values decreasing slowly as both vehicles accelerate and then decreasing slowly as the leading vehicle sustains its velocity.

In Figure 6, we show the values of the RSS-DS metric. Here, we can more easily see that it is 0 (false) at all times, confirming that there was no crash during the scenario.

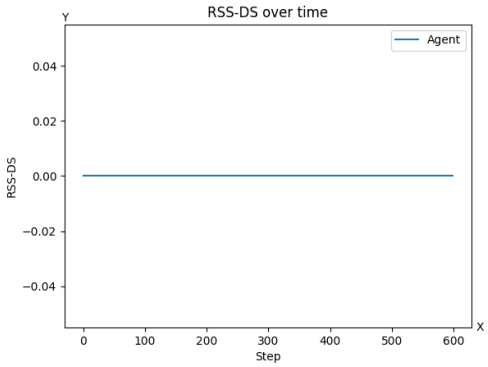

Figure 6. RSS-DS over each step of the simulation. Here, a value of 0 means that no collision was detected, and a value of 1 means that a collision was detected at the given step.

In Table 1, we show other aggregated metrics, namely the CO2EWSD, CCO2E, and GECO2ES.

Table 1. Aggregated metrics from the simulation.

Metric Name	Value (Grams of CO_2)
CO2EWSD	0.61
CCO2E (powered by grid-electricity)	31.16
CCO2E (diesel-powered)	54.67
CCO2E (petrol-powered)	54.92
GECO2ES	31.16

Here, we can observe that the vehicle would have emitted the most CO_2 if it was powered by petrol (54.92 g of CO_2), being followed closely by a diesel-powered vehicle (54.67 g of CO_2) and that an electric vehicle charged by electricity from the grid would have emitted 31.16 g of CO_2. Therefore, as expected, the savings of using a green-powered electric vehicle (GECO2ES) would be 31.16 g of CO_2.

7. Conclusions and Future Work

In this paper, we present a mathematical analysis of several criticality metrics used for evaluating the safety of AVs and also propose novel environmentally-friendly metrics with the scope to facilitate their selection by future researchers who want to evaluate both safety and the environmental impact of AVs. More exactly, we investigate if the existent criticality metrics found in the literature, which are used to quantify the severeness of critical situations in autonomous driving, are well-defined and work as intended. We found out that in some cases, the well-definedness or the intendedness of the metrics are apparent, but in other cases, we present mathematical demonstrations of these properties as well as propose alternative novel formulas for them. In addition, we also present details regarding optimality. Then, we propose several novel environmentally-friendly metrics as well as a novel environmentally-friendly criticality metric that combines safety and environmental impact. We also discuss the possibility of applying these criticality metrics in AI training such as RL algorithms and where these metrics can be used as penalty terms such as negative reward components. Here, we derived that it suffices to use a sophisticated selection for training as optimizing some metrics also optimizes other

ones, not requiring a redundant usage in training. Finally, regarding the application of the metrics, we propose a way to apply some of the metrics in a simple car-following scenario and show in a simulation that our proposed environmentally-friendly criticality metric called GECO2ES can be successfully used to evaluate AVs from the safety and environmental points of view. More exactly, we show that AVs powered by petrol emitted the most carbon emissions (54.92 g of CO_2), being followed closely by diesel-powered AVs (54.67 g of CO_2) and then by grid-electricity-powered AVs (31.16 g of CO_2) with the AVs powered by electricity coming from a green source such as solar energy, having no carbon emissions at all. Concluding, our work encourages future researchers and the industry to develop more actively sustainable methods and metrics that can be used to power AVs and also evaluate them regarding safety and environmental impact completely by using green energy. Regarding the limitations of this work, we are aware that safety and sustainability are just two facets of autonomous driving and that their acceptance also depends on other aspects such as performance-to-price value, travel time, or symbolic value, as seen in the work presented in [40]. As this work considers the training of an autonomous agent where safety, sustainability, and travel time can be optimized, the price or social values cannot be affected by AI training itself, therefore, this work is restricted to the former aspects. In future work, we plan to make use of these criticality metrics when training an AI in selected real use cases such as an overtaking scenario. Furthermore, we plan to make use of PyScript for AI, e.g., to share a DL model written entirely in Python through a website.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su14126988/s1.

Author Contributions: Conceptualization, S.L.J.; methodology, S.L.J., D.G., and T.W.; software, S.L.J. and D.G.; validation, S.L.J., D.G., T.W., and W.H.; formal analysis, S.L.J., D.G., and T.W.; investigation, S.L.J., D.G., T.W., and W.H.; resources, S.L.J., D.G., T.W., W.H., and E.M.; data curation, S.L.J., D.G., T.W., W.H., and E.M.; writing—original draft preparation, S.L.J., D.G., T.W., and W.H.; writing—review and editing, S.L.J., D.G., T.W., W.H., and E.M.; visualization, S.L.J., D.G., and T.W.; supervision, E.M.; project administration, D.G., T.W., and E.M.; funding acquisition, E.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry of Economic Affairs and Climate Action (BMWK) through the KI-Wissen project under grant agreement No. 19A20020M, and by the German Federal Ministry for Digital and Transport (BMDV) through the ViVre project under grant agreement No. 01MM19014E.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting reported results can be found in Supplementary Meterials .

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AVs	Autonomous Vehicles
AI	Artificial Intelligence
RL	Reinforcement Learning
ACC	Adaptive Cruise Control
AV	Automated Vehicle
CA	Conflict Area
DMM	Dynamic Motion Model
ET	Encroachment Time
PET	Post Encroachment Time
PrET	Predictive Encroachment Time
THW	Time Headway
TTC	Time To Collision
TET	Time Exposed TTC
TIT	Time Integrated TTC
PTTC	Potential Time To Collision
WTTC	Worst Time To Collision
TTM	Time to Maneuver
TTB	Time To Brake
TTK	Time To Kickdown
TTS	Time To Steer
TTR	Time To React
TTZ	Time To Zebra
TTCE	Time To Closest Encounter
HW	Headway
AGS	Accepted Gap Size
DCE	Distance of Closest Encounter
PSD	Proportion of Stopping Distance
CS	Conflict Severity
Δv	Delta-v
DST	Deceleration to Safety Time
$a_{long,req}$	Required Longitudinal Acceleration
$a_{lat,req}$	Required Lateral Acceleration
a_{req}	Required Acceleration
LatJ	Lateral Jerk
LongJ	Longitudinal Jerk
AM	Accident Metric
BTN	Brake Threat Number
STN	Steer Threat Number
CI	Conflict Index
CPI	Crash Potential Index
ACI	Aggregated Crash Index
PRI	Pedestrian Risk Index
RSS-DS	Responsibility Sensitive Safety Dangerous Situation
SOI	Space Occupancy Index
TCI	Trajectory Criticality Index
P-MC	Collision Probability via Monte Carlo
P-SMH	Collision Probability via Scoring Multiple Hypotheses
P-SRS	Collision Probability via Stochastic Reachable Sets
PF	Potential Functions as Superposition of Scoring Functions
SP	Safety Potential
CCO2E	Car CO_2 Emissions
GECO2ES	Green Energy CO_2 Emissions Saved
CO2EWSD	CO_2 Emissions Weighted Safety Distance

Nomenclature

The following symbols are used in this manuscript:

A_i	actor i
$\mathcal{A}$	set of all actors in a scene or scenario
t_0	starting time of a scenario
t_e	ending time of a scenario
t	a point in time
t_H	a time horizon
$p_O(t)$	position of object O at time t
$p_i(t)$	position of actor i at time t
$p_{i,m}(t)$	position of actor i at time t when conducting maneuver m
$d(p_1(t), p_2(t))$	euclidean distance of $p_1(t)$ and $p_2(t)$
$\dot{d}(p_1(t), p_2(t))$	derivative of euclidean distance d
$v_i(t)$	velocity of actor i at time t
$a_{i,min}(t)$	minimal available acceleration of actor i at time t
$a_{i,max}(t)$	maximal available acceleration of actor i at time t
$j_i(t)$	jerk of actor i at time t
v_{long}	longitudinal component of a vector v
v_{lat}	lateral component of a vector v
$u_i(t)$	control inputs of actor i at time t
$_i(t)$	sideslip angle of actor i at time t
$\psi_i(t)$	yaw angle of actor i at time t
$\omega_i(t)$	yaw rate of actor i at time t
F_{idxy}	tire forces of actor i with direction d for tire (x, y)
$c_{i\alpha f}$	front tire cornering stiffness of actor i
$c_{i\alpha r}$	rear tire cornering stiffness of actor i
l_{if}	distance from front axle to center of gravity of actor i
l_{ir}	distance from rear axle to center of gravity of actor i
L	distance from front to rear axle
m_i	mass of actor i
I_{iz}	moment of inertia of actor i
δ_{if}	front steering angle at the tires of actor i
τ	target value
$\|\cdot\|_2$	the Euclidean norm
$speed(V_1, t)$	absolute speed of vehicle 1 (V_1) and time t

References

1. Jurj, S.L.; Grundt, D.; Werner, T.; Borchers, P.; Rothemann, K.; Möhlmann, E. Increasing the Safety of Adaptive Cruise Control Using Physics-Guided Reinforcement Learning. *Energies* **2021**, *14*, 7572. https://doi.org/10.3390/en14227572.
2. VVM Consortium. VVM—Verification and Validation Methods for Automated Vehicles Level 4 and 5. Available online: https://www.vvm-projekt.de/en/ (accessed on 14 February 2022).
3. SET Level. SET Level—Simulation-Based Development and Testing of Automated Driving. Available online: https://setlevel.de/en (accessed on 14 February 2022).
4. KI Wissen Consortium. KI Wissen Project. Available online: https://www.kiwissen.de/ (accessed on 14 February 2022).
5. VDA. VDA Leitinitiative Autonomes und Vernetztes Fahren. Available online: https://en.vda.de/de/themen/innovation-und-technik/automatisiertes-fahren/vda-leitinitiative.html (accessed on 14 February 2022).
6. Neurohr, C.; Westhofen, L.; Butz, M.; Bollmann, M.H.; Eberle, U.; Galbas, R. Criticality Analysis for the Verification and Validation of Automated Vehicles. *IEEE Access* **2021**, *9*, 18016–18041. https://doi.org/10.1109/ACCESS.2021.3053159.
7. Westhofen, L.; Neurohr, C.; Koopmann, T.; Butz, M.; Schütt, B.; Utesch, F.; Kramer, B.; Gutenkunst, C.; Böde, E. Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art. *arXiv* **2021**, arXiv:2108.02403.
8. Westhofen, L.; Neurohr, C.; Koopmann, T.; Butz, M.; Schütt, B.U.; Utesch, F.; Kramer, B.; Gutenkunst, C.; Böde, E. Criticality Metrics for Automated Vehicles. Available online: https://criticality-metrics.readthedocs.io/en/latest/ (accessed on 2 May 2022).
9. United States Environmental Protection Agency. Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2019. Available online: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2019, (accessed on 16 February 2022).

10. Climate Change AI (CCAI). Climate Change AI (CCAI). Available online: https://www.climatechange.ai/ (accessed on 16 February 2022).
11. Jurj, S.L.; Rotar, R.; Opritoiu, F.; Vladutiu, M. Efficient Implementation of a Self-sufficient Solar-Powered Real-Time Deep Learning-Based System. In Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference, Halkidiki, Greece, 5–7 June 2022; Iliadis, L., Angelov, P.P., Jayne, C., Pimenidis, E., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 99–118.
12. Jurj, S.L.; Opritoiu, F.; Vladutiu, M. Environmentally-Friendly Metrics for Evaluating the Performance of Deep Learning Models and Systems. In Proceedings of the Neural Information Processing, Bangkok, Thailand, 18–22 November 2020; Yang, H., Pasupa, K., Leung, A.C.S., Kwok, J.T., Chan, J.H., King, I., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 232–244.
13. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. *arXiv* **2019**, arXiv:1907.10597.
14. Brys, T.; Harutyunyan, A.; Vrancx, P.; Taylor, M.E.; Kudenko, D.; Nowé, A. Multi-objectivization of reinforcement learning problems by reward shaping. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 2315–2322.
15. Martin, G.T. *Sustainability Prospects for Autonomous Vehicles: Environmental, Social, and Urban*; Routledge: Oxfordshire, UK, 2019.
16. Milakis, D. Long-term implications of automated vehicles: An introduction. *Transp. Rev.* **2019**, *39*, 1–8.
17. Taiebat, M.; Brown, A.L.; Safford, H.R.; Qu, S.; Xu, M. A review on energy, environmental, and sustainability implications of connected and automated vehicles. *Environ. Sci. Technol.* **2018**, *52*, 11449–11465.
18. Wadud, Z.; MacKenzie, D.; Leiby, P. Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. *Transp. Res. Part A Policy Pract.* **2016**, *86*, 1–18.
19. Fernández Llorca, D.; Gómez, E. *Trustworthy Autonomous Vehicles;* Technical Report; Joint Research Centre (Seville Site): Seville, Spain, 2021.
20. Mccarthy, J.F. Sustainability of Self-Driving Mobility: An Analysis of Carbon Emissions between Autonomous Vehicles and Conventional Modes of Transportation. Ph.D. Thesis, Harvard Extension School, Cambridge, MA, USA, 2017.
21. Xu, Z.; Cao, Y.; Kang, Y.; Zhao, Z. Vehicle emission control on road with temporal traffic information using deep reinforcement learning. *IFAC-PapersOnLine* **2020**, *53*, 14960–14965.
22. Zhu, Z.; Gupta, S.; Gupta, A.; Canova, M. A deep reinforcement learning framework for eco-driving in connected and automated hybrid electric vehicles. *arXiv* **2021**, arXiv:2101.05372.
23. Ganesh, A.H.; Xu, B. A review of reinforcement learning based energy management systems for electrified powertrains: Progress, challenge, and potential solution. *Renew. Sustain. Energy Rev.* **2022**, *154*, 111833.
24. Bai, Z.; Hao, P.; Shangguan, W.; Cai, B.; Barth, M. Hybrid Reinforcement Learning-Based Eco-Driving Strategy for Connected and Automated Vehicles at Signalized Intersections. *arXiv* **2022**, arXiv:2201.07833.
25. Kővári, B.; Szőke, L.; Bécsi, T.; Aradi, S.; Gáspár, P. Traffic Signal Control via Reinforcement Learning for Reducing Global Vehicle Emission. *Sustainability* **2021**, *13*, 11254.
26. Jimenez-Martinez, M. Artificial Neural Networks for Passive Safety Assessment. *Eng. Lett.* **2022**, *30*, 1–9.
27. Allen, B.L.; Shin, B.T.; Cooper, P.J. Analysis of traffic conflicts and collisions. *Transp. Res. Rec.* **1978**, *667*, 67–74.
28. Minderhoud, M.M.; Bovy, P.H. Extended time-to-collision measures for road traffic safety assessment. *Accid. Anal. Prev.* **2001**, *33*, 89–97.
29. Johnsson, C.; Laureshyn, A.; De Ceunynck, T. In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. *Transp. Rev.* **2018**, *38*, 765–785.
30. Wakabayashi, H.; Takahashi, Y.; Niimi, S.; Renge, K. Traffic conflict analysis using vehicle tracking system/digital vcr and proposal of a new conflict indicator. *Infrastruct. Plan. Rev.* **2003**, *20*, 949–956.
31. Sontges, S.; Koschi, M.; Althoff, M. Worst-case Analysis of the Time-To-React Using Reachable Sets. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1891–1897. https://doi.org/10.1109/IVS.2018.8500709.
32. Morales, E.S.; Membarth, R.; Gaull, A.; Slusallek, P.; Dirndorfer, T.; Kammenhuber, A.; Lauer, C.; Botsch, M. Parallel Multi-Hypothesis Algorithm for Criticality Estimation in Traffic and Collision Avoidance. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 2164–2171. https://doi.org/10.1109/IVS.2019.8814015.
33. Althoff, M.; Stursberg, O.; Buss, M. Model-Based Probabilistic Collision Detection in Autonomous Driving. *Intell. Transp. Syst. IEEE Trans.* **2009**, *10*, 299–310. https://doi.org/10.1109/TITS.2009.2018966.
34. European Environmental Agency. Average CO_2 Emissions from New Cars and New Vans Increased Again in 2019. Available online: https://www.eea.europa.eu/highlights/average-co2-emissions-from-new-cars-vans-2019#:~:text=On%20average%2C%20the%20CO2,the%20beginning%20of%20the%20monitoring (accessed on 29 April 2022).
35. Wietschel, M.; Kühnbach, M.; Rüdiger, D. *Die aktuelle Treibhausgas-emissionsbilanz von Elektrofahrzeugen in Deutschland*; Working Paper Sustainability and Innovation No. S02/2019; EconStor: Kiel, Germany, 2019
36. García, J.; Fern.; o Fernández. A Comprehensive Survey on Safe Reinforcement Learning. *J. Mach. Learn. Res.* **2015**, *16*, 1437–1480.
37. Greer, R.; Deo, N.; Trivedi, M. Trajectory Prediction in Autonomous Driving with a Lane Heading Auxiliary Loss. *IEEE Robot. Autom. Lett.* **2021**, *6*, 4907–4914.
38. Niedoba, M.; Cui, H.; Luo, K.; Hegde, D.; Chou, F.C.; Djuric, N. Improving movement prediction of traffic actors using off-road loss and bias mitigation. In Proceedings of the Workshop on 'Machine Learning for Autonomous Driving' at Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019.

39. Anaconda. PyScript: Python in the Browser. Available online: https://anaconda.cloud/pyscript-python-in-the-browser (accessed on 9 May 2022).
40. Jing, P.; Xu, G.; Chen, Y.; Shi, Y.; Zhan, F. The determinants behind the acceptance of autonomous vehicles: A systematic review. *Sustainability* **2020**, *12*, 1719.