This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Grasping Causality for the Explanation of Criticality for Automated Driving

TJARK KOOPMANN[1], LINA PUTZE[1], LUKAS WESTHOFEN[1], ROMAN GANSCH[2], AHMAD ADEE[3], CHRISTIAN NEUROHR[1]

[1]German Aerospace Center (DLR) e.V., Institute of Systems Engineering for Future Mobility, Oldenburg, Germany
[2]Robert Bosch GmbH, Renningen, Germany
[3]BMW AG, Munich, Germany

Corresponding author: Christian Neurohr (e-mail: christian.neurohr@dlr.de).

The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Climate Action within the project 'VVM – Verification & Validation Methods for Automated Vehicles Level 4 and 5'.

ABSTRACT Safeguarding automated driving systems at SAE levels 4 and 5 is a multi faceted challenge, for which classical distance-based approaches become infeasible. To alleviate this, contemporary scenario-based approaches suggest a decomposition into scenario classes combined with the statistical analysis of these classes regarding their criticality. Unfortunately, relying solely on associative statistics may fail to recognize the causalities leading to critical scenarios. These scenarios are prerequisite for the scenario-based development of safe automated driving systems. As to incorporate causal knowledge within the development process, this work introduces a formalization of causal queries. Answering these facilitates a causal understanding of safety-relevant influencing factors. This formalized causal knowledge can be used to specify and implement safety principles that provably reduce their associated criticality. Based on Judea Pearl's causal theory, we define a causal relation as a causal structure together with a context, both related to a suitable domain ontology. The focus lies on modeling the effect of such influencing factors on criticality as measured by appropriate criticality metrics. Our main example is a causal relation for the influencing factor 'reduced coefficient of friction' and its effect on the Brake-Threat-Number. As availability and quality of data are important to answer the causal queries, we also discuss requirements on real-world and synthetic data acquisition. Overall, this work contributes to establish formal causal considerations within the safety process for automated driving systems.

INDEX TERMS safety, automated driving, criticality, causal inference, statistical analysis

I. INTRODUCTION

Traffic safety research is inherently tied to the investigation of causal questions such as 'Which chain of events led to the accident?' or 'Are strict speed limits effective at increasing general safety in traffic?'. To prevent accidents, the theoretical part of teaching humans to drive safely usually includes numerous potential dangers, e.g., a wet road surface. This is combined with a causal explanation as to why an instance of this danger could lead to a critical situation, e.g., 'less available tire-to-road friction implies decreased maneuverability'. Moreover, this abstract danger is accompanied by some safety principle to mitigate the potential risk, e.g., 'if the road surface is wet, drive at reduced speed'. The human brain excels at identifying and mitigating instances of such dangers in the traffic world. This is due to a causal understanding derived from prior experiences, e.g., knowledge obtained from driving lessons or parents, leading to solid predictions even for unseen circumstances.

Automated driving systems (ADSs) are complex systems that are expected to operate fully automatically within their limited operational design domain (ODD, at SAE Level 4) or without any limitations (SAE Level 5). Here, ODD is defined as the 'operating conditions under which a given driving automation system [..] is specifically designed to function' [1]. Even more importantly, these systems should navigate such ODDs *safely* through all instances of danger. As to adapt the human learning process for automated driving, a rigorous process of identification and formalization of the knowledge to be learned is necessary. Neurohr et al. established a methodical *criticality analysis* for ADSs, aiming

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

to systematically identify critical influencing factors called *criticality phenomena* (CP) and analyze the underlying causal relations [2]. The work at hand extends the criticality analysis by leveraging *causal theory* [3] as a novel approach to model and analyze causal relations with formal rigor. Explanations of how CP increase criticality that are based on causal models with formal semantics enable the derivation and implementation of *safety principles* (SPs) – which are changes to the ADS that either reduce the occurrence of CP or mitigate their effect on criticality. SPs derived from causal knowledge provably reduce criticality and can therefore serve as quantitative evidence in a safety argumentation.

As we will see in subsection II-C, formal methods which tackle causal question are sparse in automotive safety. Hence, our distinct contributions to this field are as follows:

(i) a formalization of causal queries within the criticality analysis,
(ii) the application of causal theory for the formal modeling and analysis of causal relations,
(iii) the introduction of quantities to evaluate the modeling quality of such causal relations, and
(iv) as an example, a causal analysis of the CP reduced coefficient of friction.

These contributions differ from contemporary approaches for understanding causality in the context of vehicle development, e.g. fault tree analysis. Our approach allows to cover the modeling complexity required for safe automated driving, while providing a formal basis grounded in the theories of random variables and directed acyclic graphs.

The following section II introduces the methodical foundations of the criticality analysis, as introduced in [2], the required concepts from causal theory [3] and discusses how causal effects can be estimated from data as well as other relevant related work. In section III we explore how causal theory can be applied to causal relations of CP by representing them as causal structures. We discuss requirements and modeling principles for causal structures that lead to a rigorous definition of a causal relation and its validation. As to assess the modeling quality of causal relations, we introduce various indicator quantities and evaluate them on a small example. Section IV provides a detailed example of the proposed modeling and analysis approach for the criticality phenomenon reduced coefficient of friction. Further, we discuss in section V how causal relations constructed in this way relate to requirement elicitation for data acquisition, for real world data and in simulation environments, followed by ideas for future work in section VI.

II. PRELIMINARIES

In this section, we lay the foundations for this work by introducing prior work on the criticality analysis for ADSs in subsection II-A and the basics of causal theory in subsection II-B. Regarding related work, discussed in subsection II-C, we focus on the intersection of research on automotive safety, causality and intelligent systems.

A. CRITICALITY ANALYSIS FOR AUTOMATED DRIVING SYSTEMS

Any complex system that operates within highly complex, open and unpredictable contexts needs to undergo a rigorous safety assessment before deployment [4]. In particular for ADS-equipped vehicles (at SAE Level $\geq$ 4) operating in urban areas, all relevant traffic situations and influencing factors should be considered during the design- and testing phase [2]. It is widely recognized that such claims cannot be substantiated using statistical, distance-based approaches to testing [5], [6]. To alleviate this, scientific and industrial advances as e.g. driven by the VVM and SET Level projects[1] suggest an iterative, scenario-based approach [7]. For this, a suitable first step can be a *criticality analysis*, where the operational domain is analyzed and decomposed w.r.t. potentially safety-relevant influencing factors [2].

In the context of a criticality analysis, criticality is defined as *the combined risk of all actors within a traffic situation or scenario*, cf. [2, Definition 1, Remark 1]. Its main goal is to identify so-called *criticality phenomena* (CP) which are *concrete influencing factors in a scenario which are associated with increased criticality*, cf. [2, Definition 2].

For details on the identification and formalization of CP (*step 1*,Figure 1), we refer to Babisch et al. [8]. Therein, the authors formalize a catalog of CP in urban areas in the query language of a large German accident database. Based on this, the CP could be identified in a large subset of urban traffic accidents in Germany. To assess the associative relevance of CP (*step 2*,Figure 1), their associated risk can be estimated [8, Section 4].

The relevant CP are then analyzed regarding their underlying causalities to derive SPs which prevent their occurrence or, at least, mitigate their effects. Such SPs aim to reduce the causal effect of a CP on the measurable aspects of criticality[2], enabling ADS developers to increase product safety. The main outputs of a criticality analysis are a) criticality phenomena, b) their causal relations, and c) safety principles.

As running example, we consider the CP *reduced coefficient of friction* between the road and the tires of an ADS-operated vehicle. The criticality analysis collects a finite and manageable list of such CP in the operational domain. Their explanations can then become the basis for generic SPs, e.g., 'drive carefully during rain and freezing temperatures'.

When such mechanisms are taught to humans in driving school, it rests upon the causal knowledge of the driving teacher, that, for example, the combination of rain and freezing temperatures can lead to an icy road. This condition causes a *reduced coefficient of friction*, which increases the probability of unstable driving dynamics.

SPs that are based on causal knowledge are understood during the design phaseof an ADS and implemented in the final product, drastically increase its safety. Purely associative

[1] https://vvm-projekt.de/en, https://setlevel.de/en

[2] ISO 26262 defines 'safety mechanisms', a related albeit narrower concept for fault corrections of functional components [9].

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

implementations based on machine learning algorithms are prone to inefficient and unsafe behavior. In contrast, through implementation of causal knowledge the operational domain becomes predictable and manageable.

This work examines the role of causal analysis of CP within complex operational domains. Initially, the authors defined a *causal relation* rather syntactically as a *directed, acyclic graph* $CR = (P, E)$ *where* P *is the set of nodes – variables described by propositions on traffic scenarios – and* $E \subseteq P \times P$ *is the set of edges – causal links between the variables* [2, Definition 7]. Here, we sharpen this definition using Pearl's causal theory. More generally, we extend the criticality analysis methodically by incorporating SPs, leading to the process sketched in Figure 1. Therein, we begin the associative analysis with a CP, ensure its observability, and establish its associative relevance. Following this, the causal analysis starts by determining the novelty of the CP. A core part is the modeling of the CP's causal relation, which reflects the analyst's understanding of the CP's emergence and its effect on criticality. The causal relation enables predictions and serves as a backbone for subsequent analyses and queries.

Initially, the causal relation is checked for its explanatory potential: does it explain the emergence of the CP? does it explain how the CP increases criticality? If the explanations are not satisfactory, we re-enter the modeling step. Otherwise, the causal effect of the CP on criticality is analyzed. If the causal effect is negligible (e.g., due to a previously identified spurious association), we discard the CP and exit the analysis. Otherwise, we are presented with a CP that is both warranted for mitigation or prevention and understood causally. Hence, we analyze the causal relation w.r.t. the variables that are suitable candidates for mitigation or prevention, i.e. variables with sufficiently strong causal effects and feasible controllability at design- or runtime. Once candidate variables are identified, we derive SPs whose effectiveness is evaluated using interventions. If the SP is valueable. it reduces the CP's causal effect on criticality significantly. Otherwise, the candidate variables need to be refined.

Note that this process intentionally relies on causal language. We condense the arising causal queries to the following four aspects and indicate which steps of Figure 1 require their evaluation:

- **Q1** The explanation of a CP by a set of predecessors in a causal relation (*steps 3, 5*)
- **Q2** The causal effect of a CP on criticality as measured by a suitable metric (*steps 7, 8*)
- **Q3** The explanation of measured criticality by a CP (*step 6*)
- **Q4** The causal effect of SPs on reducing measured criticality (*step 11*)

A central aspect of this work, which is presented in section III, is the formalization of these causal queries using the do-calculus, as introduced in subsection II-B. To motivate the formalization of these causal queries, let us consider an example. Recall that we are investigating the CP *reduced coefficient of friction*, where we formalize the coefficient of friction μ being *reduced* as $\mu \leq 0.5$. Suspecting that a *reduced coefficient of friction* negatively impacts a vehicle's longitudinal maneuverability, we measure criticality using the Brake Threat Number (BTN) [10, Section 5.2].

Assuming the CP's novelty (**Q1**) the analyst formalizes his understanding of the CP by modeling the causal assumptions on the emergence of a reduced road-tire friction and its impact on the BTN. When the CP is already explained by the causal relation of a related CP, for example *degree of wetness*, it is not considered novel.

We can now use this model to answer causal queries. As a first step, it is imperative to validate the causal assumptions: does the model explain the presence of $\mu \leq 0.5$ within the context (**Q1**) by incorporating the influences of *weather*, *road surface*, *vehicle dynamics*, and *vehicle tires* appropriately? If so, we ask whether $\mu \leq 0.5$ explains the increase of BTN to a satisfactory extent (**Q3**). Moreover, it is essential to quantify the effect of the CP formalized as $\mu \leq 0.5$ on the criticality metric BTN (**Q2**). The amplitude of the causal effect can then be used to evaluate whether the CP should be investigated further.

Assuming the CP's causal relevance, SPs that reduce the impact of reduced road-tire friction can be derived from this causal model. Causal theory provides the formal tools to guide this process. By examining the causal effects of the CP's predecessors on the CP and of the CP's successor variables on the measured criticality in the causal model (**Q2**), we are able to identify suitable candidates. In our example, we may suspect that the mitigation/prevention variables $\{$*road surface, degree of wetness*$\}$ have a large effect on μ. Moreover, they are controllable in the sense that their values can be excluded from the ADS's ODD, i.e., the ADS shall not operate under those conditions. A derived SP then defines unsafe value ranges for the identified variables. For example, the ADS shall not operate on *cobblestone* roads or whenever the *precipitation* is ≥ 10 mm/h.

SPs target a CP either by reducing its probability of occurrence or by mitigating its effects. Note that this differs from defining SPs independently of CP. Consider the generic SP 'always hold a minimum distance of 5m to other traffic participants'. While applying this SP certainly reduces the BTN, it does not target the causality underlying a *reduced coefficient of friction*. This is problematic as

1) we do not know whether the SP can be generalized to similar situations (e.g., braking on icy roads without other traffic participants),
2) in a safety case, there is hence no possibility to argue that that a certain CP is mitigated by implementing an SP, and
3) implementing non-specific SPs leads to ADSs that perform inadequately in absence of that CP (e.g., holding unnecessarily large distance to parking vehicles).

This indicates that SPs ought to be checked for effectiveness (**Q4**). In our example, we can analyze whether an intervention on *road surface* $\neq$ *cobblestone* or *degree of wetness* < 10 mm/h *precipitation* reduces either

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

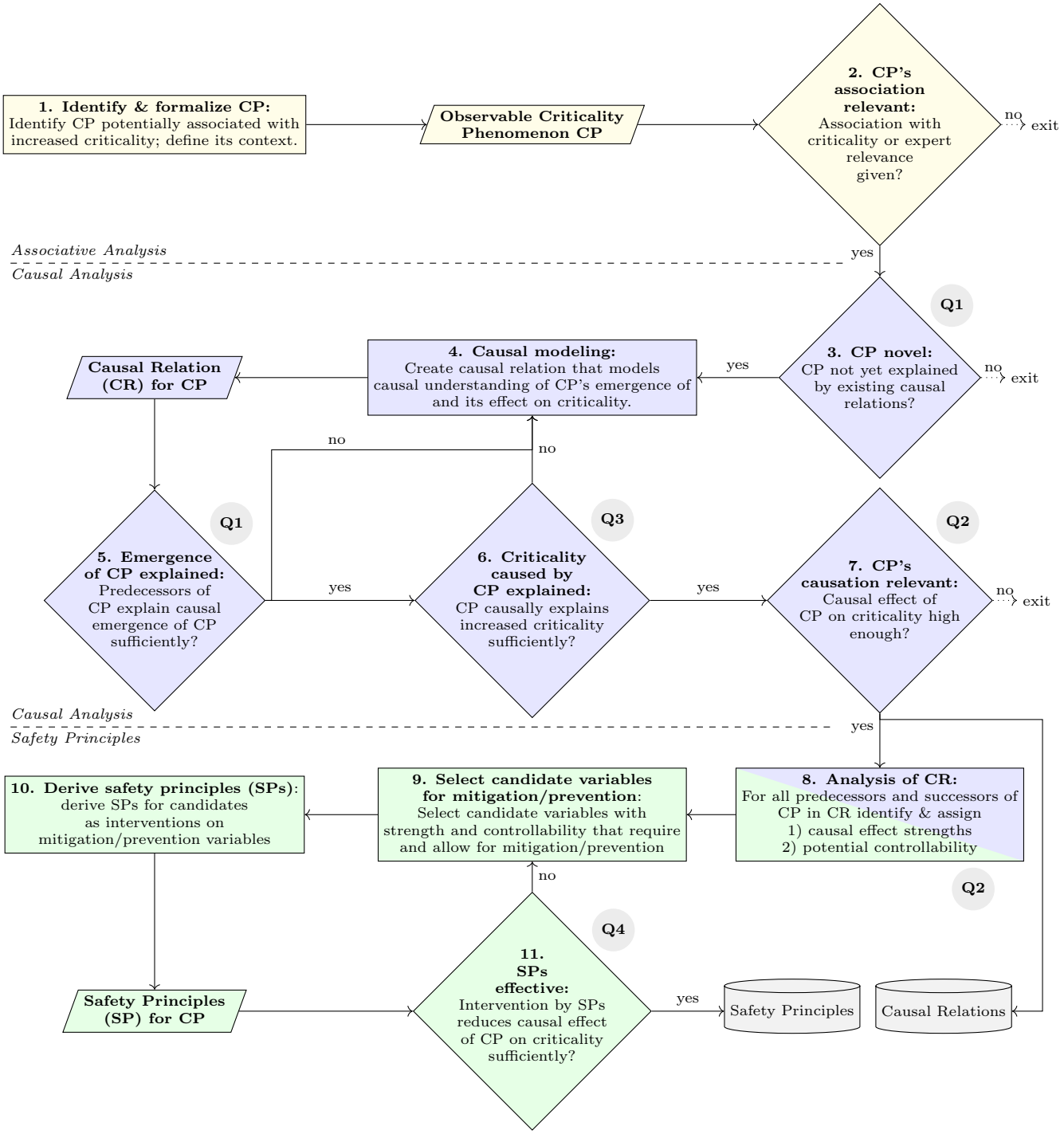


Figure 1: The procedure of the criticality analysis. The goal is to identify and formalize criticality phenomena, unveil the causalities behind then, and to derive effective safety principles. Steps that are relevant for causal analysis are annotated with their respective causal queries **Q1** to **Q4**.

- the probability of occurrence of a *reduced coefficient of friction*, or
- the its effect on criticality as measured by the BTN.

If this is the case, we found a valid SP that can be used for the design and implementation of an ADS.

So far we sketched how causal queries within a criticality analysis can contribute to the development of safe ADSs. As a foundation to formalize and answer such causal queries, we now revisit the basics of Pearl's causal theory.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

B. INTRODUCTION TO CAUSAL THEORY

Pearl's hierarchy of causation divides the topic of causality into three levels of cognitive ability: the level *association* is connected to observing ('If I see X, what does it tell me about Y?'), the level of *intervention* is concerned with doing ('If I do X, will it change Y?') and the level of *counterfactuals* covers the imagination ('If X had occurred, what would have been Y?) [11]. While the traditional concepts of probability theory can infer associations by estimating the probability of past and future events, they lack regarding interventional or counterfactual queries. This work is restricted to the first two levels of this hierarchy and omits counterfactuals.

To deal with interventional queries, experimental setups are commonly used. In many scientific fields like, psychology or medicine, the study design of randomized controlled trials (RCTs) is considered as a standard instrument to prove causal relationships. However, investigating causal effects in real-world traffic using experimental setups has some serious drawbacks. Conducting experiments in live traffic possibly endangers human lives and obtaining informed consent of the trial's participants is impossible. Moreover, experiments performed on proving grounds or in simulations are typically not transferable without further evidences, as the complexity of traffic is hardly reproducible in artificial environments.

The framework of causal theory introduced by Judea Pearl provides the means to investigate causalities based on non-experimental data [3], [12]. This includes a formal notation to express these causalities by combining graphical models with Bayesian statistics. The main idea is that, in nature, on a sufficient level of detail, there exists a collection of causal mechanisms that can be represented as deterministic functional relationships in so-called *structural causal models (SCMs)*:

Definition 1 (Structural Causal Model (SCM), [11]): A **structural causal model** $\mathcal{M}$ is a tuple $(\mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}))$, where

(i) $\mathbf{U}$ is a set of background variables, called *exogenous variables*, that are determined by factors outside the model;
(ii) $\mathbf{V}$ is a set $\{V_1, V_2, \ldots, V_n\}$ of variables, called *endogenous variables*, that are determined by other variables in the model – that is, variables in $\mathbf{U} \cup \mathbf{V}$;
(iii) $\mathcal{F}$ is a set of functions $\{f_1, f_2, ..., f_n\}$ such that each f_i, $i \in \{1, \ldots, n\}$, is a mapping from the domains of $U_i \cup PA_i$ to the domain of V_i, where $U_i \subseteq \mathbf{U}$ is a subset of exogenous variables and $PA_i \subseteq \mathbf{V} \setminus \{V_i\}$ is a subset of endogenous variables called *parents* of V_i. This means that f_i assigns values $v_i = f_i(pa_i, u_i)$ to V_i and the entire set $\mathcal{F}$ forms a mapping from the domain of $\mathbf{U}$ to the domain of $\mathbf{V}$;
(iv) $P(\mathbf{U})$ is the joint probability distribution of $\mathbf{U}$.

We remark that an SCM $\mathcal{M}$ induces a joint probability distribution $P_{\mathcal{M}}(\mathbf{V})$ on the set of endogenous variables $\mathbf{V}$ by combining $P(\mathbf{U})$ with the functions in $\mathcal{F}$.

The existence of functional relationships within an SCM can be represented by directed graphs. A *causal structure*, denoted by $\mathcal{S} = (\mathbf{V} \cup \mathbf{U}, \mathbf{E})$, consists of a set of variables $\mathbf{V} \cup \mathbf{U}$ and a set of directed edges $\mathbf{E}$ which represent possible direct functional relationships between the variables. An example of a causal structure is given in Figure 2. Correlations among exogenous variables are typically illustrated by dashed bidirectional edges. Unidirectional edges in a causal structure model the possibility of a direct causal influence. Therefore, it is the absence of such edges which carries the causal information in a causal structure. For example, the missing edge between V_1 and V_3 in Figure 2 models the assumption that V_1 has no direct influence on V_3, while the path $V_1 \to V_2 \to V_3$ illustrates the possibility of an indirect influence over V_2. Similarly, the absence of dashed bidirectional edges between the exogenous variables U_1 and U_2 models the assumption that they are independent. Note that any consecutive sequence of edges in a causal structure is called a *path*, regardless of their direction.

A causal structure satisfies the *Markov Condition* if the graph is acyclic and the exogenous variables are independent. This characteristic ensures that each variable is independent of its non-descendents given its parents. The Markov condition indicates whether a set of parents is considered as complete in the sense that all relevant immediate causes of a variable are modeled. Hence, it can serve as a criterion for the correct level of abstraction.

To enable a formal representation of interventions the so called *do-operator* is introduced. Let $\mathcal{M} = (\mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}))$ be an SCM, $X \subseteq \mathbf{V}$ a set of variables and x a concrete assignment of X. The operator $\mathrm{do}(X = x)$, abbreviated as $\mathrm{do}(x)$, simulates an *intervention* on the variables in X. This means the intervention deletes the functions defining X from the model and substitutes $X = x$ in every occurrence in the model. More formally, an intervention $\mathrm{do}(X = x)$ generates a submodel $\mathcal{M}_x = (\mathbf{U}, \mathbf{V}, \mathcal{F}_x, P(\mathbf{U}))$ with $\mathcal{F}_x = \{f_i \mid i = 1, \ldots n, V_i \notin X\} \cup \{X \leftarrow x\}$ [11, Definition 3]. The *post-interventional distribution* of a set of variables $Y \subseteq \mathbf{V}$ is then given by the distribution that is induced by the submodel $\mathcal{M}_x$:

$$P_{\mathcal{M}}(Y \,|\, \mathrm{do}(X = x)) := P_{\mathcal{M}_x}(Y)\,.$$

Graphically, such an intervention corresponds to the deletion of of all edges coming into X, cf. Figure 3.

Conventional statistical approaches implement interventions in experimental setups, such as RCTs, to investigate the post-interventional distribution. The introduced formalism enables the estimation of causal effects based on the pre-interventional distribution – without additional experimental data. This characteristic is called *identifiablity*:

Definition 2 (Identifiability, [12, Definition 2.3.4]): A causal effect is **identifiable** from a graph if it can be computed uniquely from any positive probability of the observed variables.

The so-called *do-calculus* consists of three basic rules to check identifiability [13]. A causal query is identifiable if and only if it can be transformed to an expression without the do-operator by repeated application of the do-calculus, cf. [13, Theorem 7]. There exist criteria to check identifiability directly on the graph. One of them is the *backdoor*

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

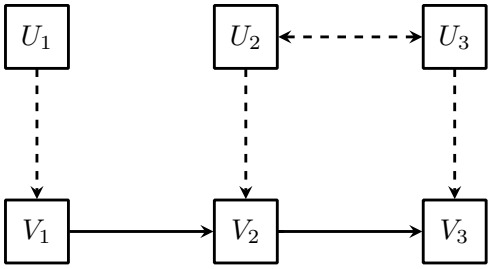


Figure 2: Example of a causal structure (pre-intervention) with dependent and independent exogenous variables.

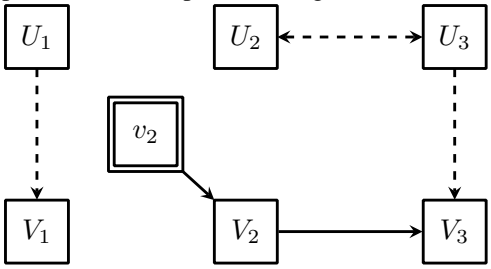


Figure 3: The causal structure of Figure 2 post-intervention for $\mathrm{do}(V_2 = v_2)$.

criterion. The main idea is to find choose a set of variables that blocks all spurious associations between X and Y and does not open any new spurious associations. In the language of causal models, a path p between two variables is called *blocked* (or *d-separated*) by a set of variables S if either p contains a chain $i \to m \to j$ or a fork $i \leftarrow m \to j$ such that the middle variable m is in S or if p contains an inverted fork $i \to m \leftarrow j$ such that neither the collider m, nor its descendants are in S. The back-door criterion states that a set S of measurable variables is *admissible for adjustment* (or *adjustment set*) if no element of S is a descendant of X and the elements of S block all paths from X to Y that contain an directed edge into X. For such S, the causal effect of X on Y is identifiable and given by [12, Theorem 3.3.2]:

$$P(Y = y \mid \mathrm{do}(X = x)) = \sum_s P(Y = y \mid X = x, S = s) P(S = s)\,, \tag{1}$$

where we sum over all possible values s that the variables in S can take. The criterion holds for graphs of any shape, even for *semi-Markovian models*, which are acyclic graphs whose exogenous variables may be dependent.

Let us remark that neither the do-calculus nor the backdoor criterion require any information about the specific functions $\mathcal{F}$ of an SCM or the joint probability distribution of the exogenous variables $P(\mathbf{U})$. Therefore, the framework of causal inference enables the estimation of causal effects based on the causal structure and suitable non-experimental data.

There are many libraries implementing the do-calculus to different degrees, such as pgmpy [14], pyAgrum [15], PyWhy/DoWhy [16] and causallib [17]. In this work, we rely on pgmpy for the example of section III and on pyAgrum for section IV.

C. RELATED WORK

1) Bayesian Networks in Safety Considerations

Bayesian Networks [18] are similar to causal models in that they are both probabilistic graphical models representing a set of random variables and their conditional dependencies via a directed acyclic graph. They have been considered for safety considerations, in particular for risk analysis and reliability topics, in many domains [19]. In his thesis, Rauschenbach provided the foundations for modeling errors in complex systems using Bayesian networks [20]. Rauschenbach and Nuffer applied Bayesian networks in this context for a probabilistic Failure Mode and Effect Analysis (probFMEA) and for evaluation of functional safety metrics [21]. Steps towards the application of Bayesian networks in the automotive domain have also been taken, e.g., to learn models for the probability distributions of initial scene configurations from real-world data [22]. Moreover, Adee et al. utilize Bayesian networks to model and assess triggering conditions and performance limitations regarding the safety of the intended functionality [23], and even for the discovery of novel triggering conditions [24]. In recent work, Maier and Mottok present the tool *BayesianSafety* that builds on Bayesian networks to combine environmental influences, machine learning and causal reasoning for automotive safety analyses [25]. As causal effects cannot be inferred from the conditional probability distributions (CPDs) alone, these Bayesian approaches do not enable the user to query about causality. In contrast, our work relies on the do-calculus which allows to answer causal queries necessary for designing *effective* SPs.

2) Causality in Automotive Safety

The well-known automotive safety standards ISO 26262 [9] and ISO 21448 [26] already consider the problem of causality. In particular, the identification of factors that cause hazardous behavior. For this, classical methods for hazard analysis and risk assessment are recommended, such as Fault Tree Analysis (FTA) [27] or System Theoretic Process Analysis (STPA) [28]. In [29], [30], the authors present a safety analysis for ADS at SAE Level $\geq$ 3 that relies on an extended FTA for causal analysis of hazards with a focus on triggering conditions in the sense of ISO 21448. However, FTA may not be applicable in case of dependent events and is generally restricted to tree-shaped models. The do-calculus extends the class of models from trees to acyclic graphs, which allows to model (acyclic) dependencies between events. Moreover, classical safety analyses such as FTA focus on identification of hazards for a concrete item, i.e., the specifics of the system are known. A criticality analysis is also applicable in case only an abstract use case description is given, as it focuses on CP and their underlying causal relations for a class of systems [2], [31].

For the validation of sensor models, Linnhoff et al. [32] introduce the *Perception Sensor Collaborative Effect and Cause Tree* (PerCollECT) where they model, as an informal graph, the effect of system-independent causes on the active perception of sensor models. While providing crucial infor-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

mation about the causal assumptions on the sensor model, the relevance estimation is expert-based and therefore inherently qualitative. In contrast to our approach, PerCollECT does not provide a formal basis for the definition of nodes and edges grounded in probability theory. In [33], Linnhoff et al. measure the influence of environmental conditions on automotive lidar sensors. The data measured in this approach could be advantageously combined with a causal structure to derive the CPDs necessary to obtain a causal model.

Closest to our work is a series of publications by Maier et al. First, they sketch how causal models could be used for automotive safety and scenario-based testing [34]. Second, the authors put forward a refinement of their initial approach [35]. Here, Maier et al. demonstrate on an example how a causal model can be built for an SAE level 1 system and use simulation data to estimate conditional probabilities as well as predictive and diagnostic causal queries. Similar to our work, they aim to systematically construct the causal model based on available knowledge. However, the focus lies on testing and integration into a model-based engineering workflow, whereas our approach considers the specifics of early design phases – assuming only few, abstract system properties. Most importantly, our approach is particularly focused on SAE levels ≥ 4, which has significant implications. For example, data collection efforts are considerably harder for these levels of automation. We show how data collection efforts can be informed by the results of our causal analysis (e.g., using adjustment sets). Another example is the much more complex context, which requires us to integrate ontologies into causal models to achieve a consistent underlying context formalization.

III. CAUSAL RELATIONS AS SEMI-MARKOVIAN CAUSAL MODELS

On the basis of Pearl's causal theory introduced in subsection II-B, we now propose an approach to achieve a causal understanding of CP and their influence on criticality. In particular, we seek for formalize and answer the causal queries **Q1**-**Q4** outlined in subsection II-A. As a main idea, we provide a formal definition of a CP's *causal relation* that enables a comprehensive analysis of causality for automated driving.

A. CRITICALITY PHENOMENA AND CRITICALITY METRICS IN TERMS OF CAUSAL THEORY

The language of causal theory requires an adequate definition of the terms *criticality phenonenon* and *criticality metric*. Recall, that a CP is defined as *a concrete influencing factor in a scenario associated with increased criticality* [2, Definition 2]. Expressing a CP in terms of a causal theory means that it has to be described using random variables. In particular, when analyzing the causal influence of a CP on measured criticality, the do-calculus prescribes the investigation of an intervention $\mathrm{do}(X = x)$ for a suitable random variable X and a value x. Therefore, the CP has to occur as value of a random variable – not as the random variable itself. Accordingly, we introduce a binary random variable X with value range $\mathrm{Im}(X) = \{cp, \neg cp\}$ where the value cp models the CP and $\neg cp$ is its negation.

As a simple example, consider the CP *heavy rain* ($= cp$), roughly understood a strong form of liquid precipitation ($= X$). One possible formalization of the random variable X could be based on the precipitation rate. By determining a threshold α for the precipitation rate, X becomes a binary random with values cp = *heavy rain*, if the precipitation rate is higher than α and $\neg cp$, if it at most α. A formalization of both the random variable X and the value cp representing the CP shall be based on careful design decisions. In particular, regarding its relevance in the criticality analysis and its measurability in terms of basic concepts (such as precipitation) [36]. In some cases, domain standards and regulations are available to guide these decisions, but even in such cases the semantics can vary. For *heavy rain* there exist several thresholds and different units characterizing it: The American Meteorological Association defines *heavy rain* as a precipitation in the form of liquid water drops with a maximum rate over 7.6mm in sixty minutes or more than 0.76mm in six minutes[3]. In contrast, Germany's national meteorological service DWD defines heavy rain as a precipitation of more than 10mm in sixty minutes or more than 1.7mm in ten minutes[4]. Even formalizations not using the basic concept of precipitation are conceivable, e.g., based on the expected probability of occurrence.

To investigate causal effects it is crucial that the value range of the random variable X associated with a CP is measurable in the real world and preferably also in simulation. Hence, the value range and its unit need to be chosen adequately. For a unified semantics, the chosen formalizations of CP should be denoted in a domain ontology. If appropriate data or measurement standards are already known, the value range and its unit should be specified accordingly. The actual formalization of the CP can be shifted to a later point in time when the causal structure will be combined with data. Its general measurabilitym, however, is mandatory. Clearly, the chosen formalization influences the causal analysis. Causal effects might even vanish, e.g., when choosing a very low precipitation rate to formalize heavy rain. Refinement of the formalizations can hence become necessary [2].

In order to measure criticality we employ so-called *criticality metrics*, which are functions that estimate certain aspects of criticality in a scene or scenario [10]. More precisely, we assume criticality metrics to be of the form $\varphi : Situ \rightarrow [0, \infty)$, mapping from the set of traffic situations $Situ$ to non-negative real values quantifying criticality – with larger values implying increased criticality. As each CP affects different aspects of criticality, the selection of the metric heavily impacts the causal analysis. Consequently, we need to evaluate possible metrics for every CP individually. It can be advantageous to investigate the causal effects of a CP on multiple metrics to get a broader picture of its influence on criticality. Westhofen

[3]https://glossary.ametsoc.org/wiki/Rain
[4]https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv2=101812&lv3=101906

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

et al. provide guidance on selecting suitable criticality metrics for the task at hand based on a state-of-the-art review [10]. In the example of *heavy rain*, the braking distance of the ADS-operated vehicle might be a simple candidate metric to measure its effect on criticality. The modeling process of a causal structure for cp = *heavy rain* $\in \mathrm{Im}(X)$ with X = *precipitation* and φ = *braking distance* is illustrated by Figure 4. There, in a first refinement step, two variables hypothesized to be relevant for the causal effect of X on φ are added to the graph as nodes together with the respective edges.

B. REQUIREMENTS ON AND PRINCIPLES FOR MODELING CAUSAL RELATIONS

In the previous subsection, we framed CP and criticality metrics in causal language and gave an idea on the modeling and refinement of causal structures, as illustrated by Figure 4. We now introduce the context of a causal structure to obtain a formal definition for the causal relation of a CP.

The overarching goal is to derive a causal structure that can be combined with data such that the causal effect of the investigated CP on criticality is identifiable in terms of the do-calculus. Essentially, the causal relations in the criticality analysis correspond to the concept of causal structures, cf. subsection II-B. However, there are additional requirements. Each CP defines an abstract class of scenarios – the set of all scenarios where the CP is present. Therefore, explicitly modeling all things that are potentially relevant to the CP's causal relation leads to arbitrarily complex causal structures. Due to this issue, it appears natural to partition the scenarios for a given CP into subsets which can be modeled more efficiently. Pearl describes a *context* as a collection of assumptions and constraints that are assumed to be true and that every probability distribution is implicitly conditioned on [12, pp.4-5]. Based on this idea, we define a context for causal structures that represents a subset of scenarios by defining constraints on the existence of objects (e.g., a certain vehicle). These objects are defined in terms of background knowledge (e.g., what entails a vehicle). Both the objects and the background knowledge are considered to be an *ontology*.

Definition 3 (Context): The *context of a causal structure* is a set of statements based on a suitable ontology about the existence of objects and constraints on these objects.

For the automotive domain, there exist multiple promising ontologies that might be used to define contexts of causal structures, e.g., A.U.T.O. [36] or ASAM OpenXOntology [37]. The 6-Layer Model (6LM) as presented in [38] may be used to structure the statements provided by the context. Note that an increased number of assumptions and constraints in the context leads to simpler causal structures at the cost of reducing its coverage. The context serves as an existential basis for building the causal structure for the CP, which includes modeling its emergence as well as its influence on a criticality metric. Combining causal structures with the requirements of a criticality analysis, we arrive at the following:

Definition 4 (Causal Relation): A *causal relation* for a CP cp is a tuple $(\mathcal{S}, \mathcal{C})$ with respect to a suitable traffic domain ontology $\mathcal{O}$, where $\mathcal{S} = (\mathbf{V} \cup \mathbf{U}, \mathbf{E})$ is a causal structure and $\mathcal{C}$ is a context such that

(i) there exists a binary random variable X with $\mathrm{Im}(X) = \{cp, \neg cp\}$ corresponding to a node in $\mathbf{V}$,
(ii) there exists a criticality metric $\varphi : \mathit{Situ} \rightarrow [0, \infty)$ corresponding to a sink of $\mathcal{S}$ (i.e., no outcoming edges),
(iii) the variables in $\mathbf{U}$ are the exogenous variables corresponding to error terms,
(iv) the variables in $\mathbf{V}$ are defined on properties of the objects in $\mathcal{O}$, and
(v) the context $\mathcal{C}$ is defined based on $\mathcal{O}$.

Remark 1 (Causal Relation):

- Regarding (ii): in principal, it is possible to model several criticality metrics in one causal structure.
- Regarding (ii): the chosen criticality metric φ has to be suitable for measuring the aspects of criticality associated with cp
- Regarding (ii): the causal structure's dependencies turn the criticality metric φ into a random variable.
- Regarding (iii): formally, there exists an exogenous variable for every endogenous variable, however, these errors terms are usually not modeled explicitly.

As to illustrate subsection 4 and subsection 1, consider the example of Figure 5, which depicts a simple, generic causal relation. The causal structure contains 10 nodes including the random variable X near the center of the graph and a criticality metric φ as a sink. The example assumes that the variables V_2, V_3 and V_4 causally influence X, while X influences φ through V_7 and V_8. These random variables encode properties of the objects a, b, and c defined in the context $\mathcal{C}$. Taken together, the causal structure and the context of Figure 5 define a causal relation.

a: Modeling of Causal Relations

In order to build the causal structure, the relevant 'propositions on traffic scenarios' related to the CP and the criticality metric need to be represented as random variables and structured as directed, acyclic graph. In this graph, each edge represents the possibility of a direct functional relationship between the connected variables, i.e. a causal assumption, without necessarily being explicit.

To enable an instantiation of the causal structure with data, the variables need to be expandable to random variables. Hence, every variable has to be investigated regarding the possibility to assign a value range including a unit. Typically, the inability to assign a value range is due to an inadequate definition of the variable or an inappropriate abstraction level. In that case, we have to decide whether it is sufficient to keep the variable as an unobserved background factor or if further concretization is required.

Note that the application of the do-calculus is defined on discrete variables. In fact, many variables in traffic are usually connected to continuous distributions such as posi-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

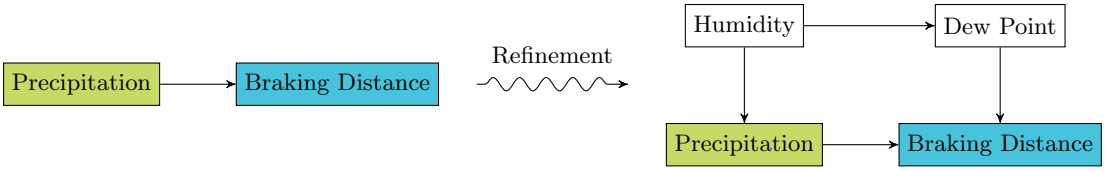


Figure 4: Example of modeling a causal structure, starting with a criticality phenomenon (green) and a criticality metric (blue). The indicated refinement step adds two variables which influence the criticality phenomenon and/or the criticality metric.

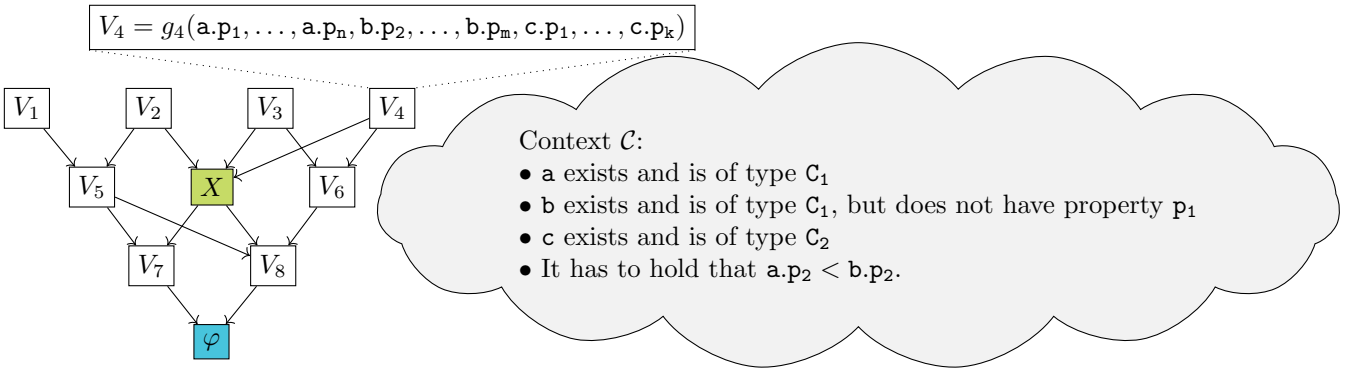


Figure 5: Example of generic causal relation with causal structure $\mathcal{S}$ on the left, error terms omitted, and context $\mathcal{C}$ on the right. X takes the values $\mathrm{Im}(X) = \{cp, \neg cp\}$ and φ is a criticality metric. $\mathtt{C_1}$ and $\mathtt{C_2}$ are classes and $\mathtt{a}$, $\mathtt{b}$ and $\mathtt{c}$ are objects in the domain ontology $\mathcal{O}$, cf. subsection 4. The inputs of the random variable V_4 are indicated by a function g_4 taking as inputs the properties of the objects in the ontology $\mathcal{O}$.

tions, velocities, or accelerations. Even if variables are not discrete in reality, discretization ought to happen during the modeling process. Here, discretization must be fine enough to capture the key features, yet coarse enough such the empirical distribution can be estimated from data. Also, consider that discretization always implies error terms. These error terms may only be neglected or even estimated for rather well-behaved variables [7].

b: Temporal Aspects

Let us remark that every causal assumption implicates a certain temporal dependency between the corresponding nodes, as the cause is assumed to precede the effect. However, modeling the causal structure without explicit consideration of the temporal aspects is a form of abstraction which can lead to cycles in the structure, as the dependencies of scenario properties might vary at different points of time [39, Chapter 10]. Therefore, the question if we need to model temporal dependencies explicitly is reduced to asking whether there exists an adequate level of abstraction. Here, we focus on modeling causal dependencies without temporal aspects and leave the investigation on time-dependent structures for future research.

c: Validation of Causal Relations

The quantitative assessment of causal relations requires the combination of the causal structure with data. As we can not expect that a data set covers every variable, we introduce the following definition allowing the usage of partial data:

Definition 5 (Instantiation of Causal Relations): Let $\mathrm{CR} = (\mathcal{S}, \mathcal{C})$ with $\mathcal{S} = (\mathbf{V} \cup \mathbf{U}, \mathbf{E})$ be a causal relation and $N \subseteq \mathbf{V}$ a subset of nodes with $X \in N$ and $\mathrm{Im}(X) = \{cp, \neg cp\}$. Given a data set D we call CR *partially instantiated* by D w.r.t. N if the CPDs of each variable contained in N are instantiated by D. If N additionally contains a node φ representing the criticality metric and at least one adjustment set for the causal effect of X on φ, we call a causal relation CR *instantiated* by D w.r.t. N.

In order to assure the validity of a CP's causal relation the criticality analysis evaluates whether the causal relation models the emergence of the CP and its influence on criticality sufficiently (*steps 5,6* in Figure 1). For this, we propose an expert-based approach based on statistical hypothesis testing. Here, we assess the modeling quality of a causal relation CR w.r.t. a data set by employing some function κ with values in $[0, \infty)$ for which values closer to zero imply that the model fits to the data. These judging functions can be used as test statistic to decide based on a data set D whether the hypothesis – CR does not explain the emergence of the CP or its influence on the criticality metric sufficiently – should be rejected or not. If for a selected family of data sets the hypothesis can be rejected with a significance level $\alpha > 0$ or if the experts decide that those datasets that do not support a rejection are insignificant, we call CR *validated* w.r.t. κ with significance level α. To achieve a comprehensive validation process, an investigation of different judging functions is necessary. subsection III-C introduces some candidates that may be used for this validation. The construction of corresponding

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

test statistics, however, is future work.

C. QUANTIFICATION OF MODELING QUALITY

Since we clarified a common understanding of the type of variables that represent the CP and the criticality metric, we can now define the main quantities suggested for the iteration cycle. The indicators proposed in this section shall provide an insight into the causal queries **Q1**, **Q2**, and **Q3** and thus will be called causality indicator functions. While **Q4** is not directly addressed in this section, it can be approached using the quantities defined in this section as well.

The following quantities measure the increase in criticality caused by a CP using a difference respectively and a ratio of expected values, thereby addressing **Q2**:

Definition 6 (Average/Relative Causal Effect): Let $\mathrm{CR} = (\mathcal{S}, \mathcal{C})$ with $\mathcal{S} = (\mathbf{V} \cup \mathbf{U}, \mathbf{E})$ be a causal relation for a CP cp and D a data set instantiating CR w.r.t. a suitable set of nodes $N \subseteq \mathbf{V}$. The *average* and *relative causal effect of* X on φ are given by

$$ACE(X, \varphi) \coloneqq \mathbb{E}(\varphi \mid \mathrm{do}(X = cp)) - \mathbb{E}(\varphi \mid \mathrm{do}(X = \neg cp)),$$
$$RCE(X, \varphi) \coloneqq \frac{\mathbb{E}(\varphi \mid \mathrm{do}(X = cp))}{\mathbb{E}(\varphi \mid \mathrm{do}(X = \neg cp))}.$$

The causality indicator functions ACE and RCE generalize the average and relative treatment effect, known from RCTs, to non-experimental data using the do-calculus.

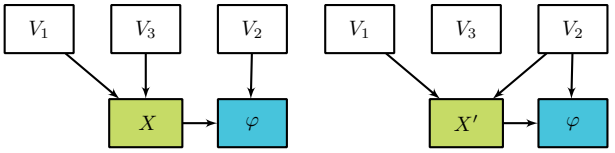
The causal query **Q3** evaluates whether the chosen criticality metric recognizes the effect of the CP and if so, whether the CP has a significant causal impact on the criticality. Therefore, we propose a quantity which incorporates the proportion of the measured criticality outside of the CP:

Definition 7 (Extend of Explanation of Criticality by CP): Let $\mathrm{CR} = (\mathcal{S}, \mathcal{C})$ with $\mathcal{S} = (\mathbf{V} \cup \mathbf{U}, \mathbf{E})$ be a causal relation for a CP cp and D a data set instantiating CR w.r.t. a suitable set of nodes $N \subseteq \mathbf{V}$. We further require the criticality metric φ in N to fulfill $\mathbb{E}(\varphi | \mathrm{do}(X = \neg cp)) \leq \mathbb{E}(\varphi | \mathrm{do}(X = cp))$. The *extent of the explanation* of φ by X is defined as

$$\sigma(X, \varphi) \coloneqq 1 - \frac{\mathbb{E}(\varphi | \mathrm{do}(X = \neg cp))}{\mathbb{E}(\varphi)}, \quad \text{denoted } cp \models_\sigma \varphi.$$

This causality indicator function formalizes the causal query **Q3** by evaluating the strength of the paths from X to φ within the causal relation. Note that, as $\mathbb{E}(\varphi)$ is purely stochastic, $\sigma(X, \varphi)$ includes a ratio of a causal quantity by an observational one.

Moving on to **Q1**, we discuss and define several indicators that allow insight into the emergence of the CP. As a first simple guess, one might think that a pairwise comparison of distributions of a subset of nodes between models is a feasible approach. In the special case of comparing the distribution of X in the model with the distribution of X in reality, we denote this with $\rho_1 = KL(P^{X_M} | P^{X_R})$, where KL denotes the Kullback-Leibler divergence [40]. However, it is very likely that even though one got the distribution of the occurrence



(a) Assumed Reality (b) Model of assumed reality

$V_1 =$ Summer	0.5	$V_2 =$ Slow	0.6	$V_3 =$ Oceanic	0.6
$V_1 =$ Winter	0.5	$V_2 =$ Fast	0.4	$V_3 =$ Continental	0.4

V_1	$V_1 =$ Summer		$V_1 =$ Winter	
V_3	$V_3 =$ Oceanic	$V_3 =$ Cont.	$V_3 =$ Oceanic	$V_3 =$ Cont.
$X = \neg cp$	0.4	0.2	0.3	0.4
$X = cp$	0.6	0.8	0.7	0.6

V_1	$V_1 =$ Summer		$V_1 =$ Winter	
V_2	$V_2 =$ Slow	$V_2 =$ Fast	$V_2 =$ Slow	$V_2 =$ Fast
$X' = \neg cp$	0.4	0.2	0.3	0.4
$X' = cp$	0.6	0.8	0.7	0.6

X, X'	$X, X' = \neg cp$		$X, X' = cp$	
V_2	$V_2 =$ Slow	$V_2 =$ Fast	$V_2 =$ Slow	$V_2 =$ Fast
$\varphi =$ Short	0.6	0.1	0.8	0.3
$\varphi =$ Long	0.4	0.9	0.2	0.7

Figure 6: Example of two causal models that share the same variables and conditional probability distributions, but differ by one edge in the graph. We use this constructed example to demonstrate the utility of the causality indicator functions from subsection III-C. The variables are all binary-valued and given as $\varphi =$ Braking Distance $\in$ {Short, Long}, $X = X' =$ Precipitation $\in$ {Light Rain $= \neg cp$, Heavy Rain $= cp$}, $V_1 =$ Season $\in$ {Summer, Winter}, $V_2 =$ Ego Forward Velocity $\in$ {Slow, Fast}, and $V_3 =$ Climate Region $\in$ {Oceanic, Continental}.

of CP correct, the dependencies leading to the emergence of CP are incorrect. To demonstrate this, we present an example of two causal models of the CP heavy rain, where one of the models is declared as reality (Figure 6a) while the other is a model that an expert came up with (Figure 6b), trying to model said reality. According to subsection III-A we model the CP as a binary variable X, where *CP* is equal to heavy rain and define φ as the braking distance with the possible values $\{Short, Long\} \hat{=} \{0, 1\}$. Since the parents of X in the two models are not equal, the index on X solely clarifies which CPDs are used. Further, with the other variables we denote V_1 as the season, V_2 as the ego forward velocity and V_3 is the climate region that is equally distributed to V_2 for this example. Computing the aforementioned quantities, we obtain $ACE = 0.2$, $RCE = 1.5$, $\sigma = 0.1416$ and $\rho_1 = 0$. Clearly, $\rho_1 = 0$ does not identify the insufficient modeling quality here. As a first improvement regarding **Q1** we suggest an indicator that incorporates stochastic dependencies among the variables:

Definition 8 (Extend of Explanation of Emergence of CP (Observational)): Let $\mathrm{CR} = (\mathcal{S}, \mathcal{C})$ with $\mathcal{S} = (\mathbf{V} \cup \mathbf{U}, \mathbf{E})$ be a causal relation for a CP cp, $N \subseteq \mathbf{V}$ a set of nodes, D_1 and D_2 data sets containing real world data and data from the model, respectively, such that $\mathcal{S}$ is partially instantiated w.r.t. N by D_1 and by D_2. Further, we denote with P and Q joint probability distributions for the variables in N, obtained by

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

estimating the distribution from D_1 and D_2, respectively. We then define the (observational) *extent of explanation of the emergence of cp* w.r.t. N as the Kullback-Leibler divergence of the distributions P and Q, i.e., $\rho_2^N = KL(P(N) \mid Q(N))$.

As this causality indicator rewards us with a combined causality indicator function for the emergence of the CP, that also includes the strengths of dependencies in both models, we obtain $\rho_2 = 0.0141 > 0$ for the example of Figure 6. However, due to the associative character of the metric, it can only infer what Peters et al. termed *observational equivalence* [39, Definition 6.47] and is thus not able to implicate equivalence of interventions. To tackle this issue, we introduce a quantity using the concept of *causal influence* as defined by Janzing et al. [41, Def. 2].

Definition 9 (Extend of Explanation of Emergence of CP (Interventional)): Let $\mathrm{CR} = (\mathcal{S}, \mathcal{C})$ with $\mathcal{S} = (\mathbf{V} \cup \mathbf{U}, \mathbf{E})$ be a causal relation for a CP cp, $N \subseteq \mathbf{V}$ a set of nodes, D_1 and D_2 data sets containing real world data and data from the model, respectively, such that CR is partially instantiated w.r.t. N by D_1 and by D_2. We denote for $n \in \mathbf{V}$ by $\mathrm{Out}(n) \subseteq \mathbf{E}$ the set of outgoing edges of the node n and denote the causal influence [41, Def. 2] of n computed using D_1 by $\mathcal{I}^R_{\mathrm{Out}(n)}$ and computed from D_2 by $\mathcal{I}^M_{\mathrm{Out}(n)}$. Finally, we obtain the vector representation of ρ_3^N as

$$\rho^N_{3,n} = \mathcal{I}^R_{\mathrm{Out}(n)} - \mathcal{I}^M_{\mathrm{Out}(n)} \text{ for all } n \in N \setminus X$$

and take the norm to define the (interventional) *extent of explanation of the emergence of cp* w.r.t. N as the scalar value

$$\rho_3^N = \|(\rho^N_{3,n})_{n \in N \setminus X}\|_2 \,.$$

For the example of Figure 6 and with $N = \{V_1, V_2, X\}$, we obtain $\rho_3^N = 0.0181 > 0$.

Note that we define ρ_1, ρ_2 and ρ_3 as causality indicator functions which inform experts about the quality of the causal relation. They should not be construed as decision metrics.

As already mentioned, ρ_2 is purely associative and can therefore only identify dependencies, not causalities. However, correct dependencies are necessary for modeling correct causalities. So computing ρ_2 contributes to the causal relation's iterative validation. Following Peters et al. [39], observational equivalence is a prerequisite for interventional equivalence. Furthermore, in practice it is beneficial to be able to compare only subsets of the vertices of a graph, especially during the design phase, which the presented causality indicator functions allow for. Note that the expressive power of the presented causality indicator functions is based on data quality in the comparison data set, since errors due to insufficient data are included. This underlines the notion of the indicator functions not being decision metrics but rather guidance for the expert.

We conjecture that there is room for improvement regarding distance measures for probability distributions in this context. Of particular interest are candidates that provide constructive feedback, e.g., in the form of witness functions, or that do not require to estimate the probability distributions such as the maximum mean discrepancy [42], [43].

IV. EXAMPLE CAUSAL ANALYSIS

In this section, we apply the modeling, instantiation and quantification steps of section III to the CP *reduced coefficient of friction*. Due to lack of real world data, we rely on synthetic data obtained from computer simulations for this example.

Layer	Property
(L1) Road Network and Traffic Guidance Objects	The road network shall consist of a T-junction.
(L2) Roadside Structures	Roadside structures may exist and are not further constrained.
(L3) Temporary Modifications of (L1) and (L2)	There shall be no temporary modifications to layer 1 and 2.
(L4) Dynamic Objects	An ego vehicle and a bicyclist with intersecting planned paths shall exist on different arms of the junction. Further, vehicles may be parked at the side of the road.
(L5) Environmental Conditions	Environmental conditions shall exist and remain unconstrained.
(L6) Digital Information	Digital information might exist, but shall remain unconstrained.

Table 1: Example context for the causal relation for the CP *reduced coefficient of friction*, based on the 6-Layer Model.

A. MODELING OF THE CAUSAL RELATION

We start by modeling the causal relation for the CP reduced coefficient of friction, according to the principles outlined in sections III-A and III-B in a specific context. More precisely, we limit the context of the causal relation to an occlusion scenario with a bicyclist at an urban junction. Table 1 describes this context, structured along along the 6-Layer-Model [38].

To find an adequate criticality metric φ, we conducted a suitability analysis, as proposed by Westhofen et al. [10]. The requirements on the criticality metrics in our example are listed in Table 2. This led to the candidate metrics Time-To-Brake, Time-To-React, Time-To-Steer, Required Lateral Acceleration, Required Longitudinal Acceleration, Required Acceleration, Brake-Threat-Number, Steer-Threat Number, Crash Potential Index, Responsibility Sensitive Safety, Trajectory Criticality Index, Collision Probability via Monte Carlo, Collision Probability via Scoring Multiple Hypotheses, Collision Probability via Stochastic Reachable Sets, and Safety Potential, cf. [10].

As the criticality in the example context depends on a potential braking maneuver, we chose the Brake-Threat-Number (BTN) as criticality metric φ. Based on the BTN's

Property	Requirement
Scenario type	Intersections
Validity	High
Sensitivity	High
Inputs	Available longitudinal control
Specificity	Medium
Reliability	High
Output scale	Ordinal
Subject type	Automation
Run-time capability	None
Target values	None
Prediction model	No prediction necessary

Table 2: Requirements on the criticality metric.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

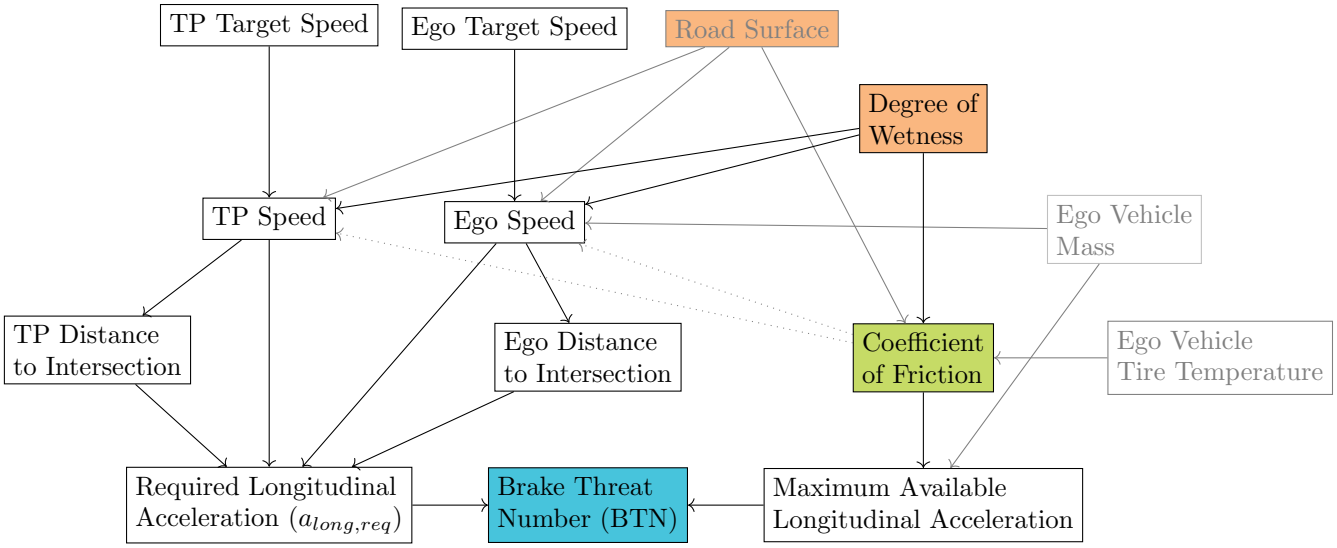


Figure 7: Causal structure for the CP *reduced coefficient of friction* given another traffic participant (TP). The green node models $X \in \{cp, \neg cp\}$, the blue node models the criticality metric $\varphi = \text{BTN}$, and the orange nodes are the confounding variables. Nodes that are greyed out could not be realized in the experiment due to the limitations of the CARLA simulator. The dotted edges indicate a possible iterative modeling step based on the experimental results of subsection IV-C.

definition, its relation to the Required Longitudinal Acceleration ($a_{long,req}$), and iterative refinement, our modeling efforts result in the causal structure depicted in Figure 7. This causal structure strikes a balance between complexity and practical feasibility, as the instantiation of causal relations becomes increasingly difficult with each additional variable due to the amount of required data.

Additionally, we provide a highly detailed causal structure for the *reduced coefficient of friction* in the appendix, cf. Figure 9 together with Table 4. While this detailed graph approximates the reality more closely, it is essentially impossible to instantiate due to its size.

B. INSTANTIATION OF THE CAUSAL RELATION

Combining both the context of Table 1 and the causal structure of Figure 7, we obtain a causal relation for the *reduced coefficient of friction* in the sense of Definition 4. Next, we determine suitable parameter ranges for a logical scenario to generate synthetic data from. These data are used to instantiate the causal relation according to Definition 5. For this, the parameter space of the scenario is defined to contain the possible values of the exogenous variables in the graph. Thereby, the road network may take any shape that is admissible by the context. As to meaningfully apply the method, all variables in the causal structure need to be measurable. However, aiming for instantiation to calculate only single causal effects, we can reduce the nodes for which data are required and, therefore, need to be measured, according to equation (1).

As available real world data for this specific application is scarce, we instantiated the causal relation using the open source automotive simulator CARLA [44]. To model a rather unbiased perception system, we employed a YOLOv4 object

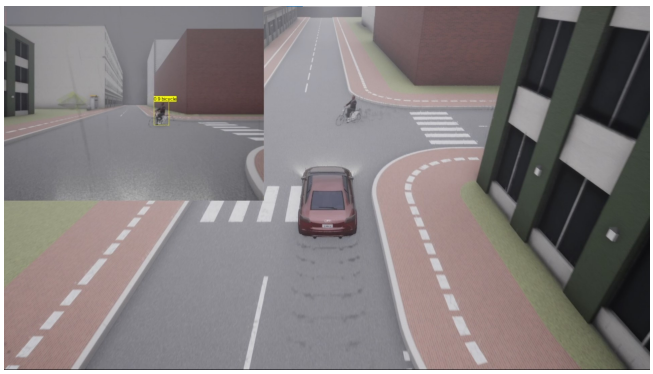


Figure 8: Screenshot of the employed CARLA simulation showing the detection of the bicyclist under *hard rain* weather condition within the context of Table 1.

detection, that receives its input from an RGB camera at the top of the wind shield of the ego vehicle. Depending on this perception, the ego vehicle may perform a braking maneuver to avoid a collision with the bicyclist. Based on the relation between weather and friction [45], we discretely mapped the weather presets in CARLA to fixed coefficients of friction:

$$\begin{aligned} \textit{clear weather} &\mapsto \mu = 0.9, \\ \textit{wet condition} &\mapsto \mu = 0.6, \\ \textit{hard rain} &\mapsto \mu = 0.4\,. \end{aligned}$$

Moreover, we realized the friction coefficient's variation in the CARLA simulation using so-called *friction triggers* spanning the entire map. For each weather preset, 600 simulation runs were executed and logged, totaling 1800 simulation traces. For the technical instantiation, the empirical

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

distribution of each variable in the graph was estimated by applying a suitable binning strategy to the data and then deriving the required conditional probability tables. The type of binning depends on the properties of the variable and its interpretation. For example, this work assumes X to be binary, so the variable coefficient of friction must be discretized using some threshold $\tau \in [0, 1]$. For other variables, such as $\varphi = \mathrm{BTN}$, we want their value to (approximately) coincide with the binning label, to render them interpretable. Lastly, we would also like to obtain interpretability of variables such as positions, distances and speeds. The problem here is that the number of bins propagates down to the variables' children in the causal graph. Therefore, data acquisition might become infeasible due do the required amount of data. As for now, we alleviate this using equal frequency binning This, in turn, implicitly introduces assumptions about the real world probability distributions of the exogenous variables. As of now, we can not solve this problem, because determining representative probability distributions is a very hard problem in practice [46].

C. EVALUATION OF CAUSAL EFFECTS

In subsection III-C, we defined several causality indicator functions to evaluate the strength of causal effects, the extent of explanation of φ by X, and the extent of explanation of the emergence of cp. As we require the random variable $X \in \{cp, \neg cp\}$ to be binary, we choose a threshold τ that formalizes the *reduced coefficient of friction* as $\mu \leq \tau$. For the demonstration, we evaluated causal effects using $\tau \in \{0.5, 0.7\}$ as thresholds, cf. Table 3. Recall, that the BTN has value range $[0, \infty)$ and that a $\mathrm{BTN} > 1$ indicates an accident scenario. For both values of τ, we obtained an ACE of ≈ 0.14 which, together with RCE's of ≈ 1.23 and ≈ 1.26 respectively, indicates a quite substantial effect in the given context. The main difference between $\tau = 0.5$ and $\tau = 0.7$ is the explainable share of criticality caused by cp, doubling from $\sigma = 7\%$ to $\sigma = 14\%$. The increase in σ is due to relatively more critical scenarios being identified when formalizing the *reduced coefficient of friction* with $\tau = 0.7$ rather than $\tau = 0.5$.

However, to compute these causal effects, we used synthetic data from an unvalidated simulation. Therefore, we cannot transfer these findings to the real world. We can merely state that the CP *reduced coefficient of friction* is actually implemented in the CARLA simulator. In order to use results from such a causal analysis in a safety case, causal effects need to be estimated from real world data.

One advantage of using synthetic data is that we could fully instantiate the causal relation of Figure 7 in the sense of Definition 5. Hence it is possible to calculate causal effects between all the variables in the causal structure. To demonstrate this we additionally calculated the the causal effect of the *degree of wetness* on the *ego speed*, cf. Table 3. Here, we note a very slight absolute and relative decrease in the ego's speed for an increased degree of wetness. As we did not implement any change in the ego's behavior based on precipitation, this must have a different cause. Analyzing the vehicle dynamics model employed in CARLA, it turns out that a *reduced coefficient of friction* leads to decreased acceleration and, thus, slightly lower speeds of all traffic participants. Therefore, we discovered a potentially missing edge in the causal structure, triggering another iteration in the modeling cycle (see the dotted edges in Figure 7).

Table 3: Computation of the causal quantities defined in subsection III-C based on the causal structure of Figure 7 and synthetic data obtained using the CARLA simulator in the context of Table 1. The parameter τ models different thresholds for the coefficient of friction to be *reduced*.

Quantity	X		φ	Value
	Coefficient of Friction (cp / $\neg cp$)			
ACE	$\mu \leq 0.5$	$\mu > 0.5$	BTN	0.1415
RCE	$\mu \leq 0.5$	$\mu > 0.5$	BTN	1.2316
σ	$\mu \leq 0.5$	$\mu > 0.5$	BTN	0.0737
ACE	$\mu \leq 0.7$	$\mu > 0.7$	BTN	0.1445
RCE	$\mu \leq 0.7$	$\mu > 0.7$	BTN	1.2569
σ	$\mu \leq 0.7$	$\mu > 0.7$	BTN	0.1478
	Degree of Wetness (cp / $\neg cp$)			
ACE	*wet*	*clear*	Ego Speed	-0.0018
RCE	*wet*	*clear*	Ego Speed	0.9659
ACE	*hard rain*	*wet*	Ego Speed	-0.0012
RCE	*hard rain*	*wet*	Ego Speed	0.975

Evaluating the quantities ρ_1, ρ_2 and ρ_3 is omitted here, as this would require real world data or, as a first iteration, a second, more refined model.

V. REQUIREMENT ELICITATION FOR DATA ACQUISITION

During the validation of a causal relation for the derivation and assessment of SPs, the causality indicator functions from subsection III-C have to be repeatedly estimated from data. This section analyzes the requirements on the data acquisition process – in the real world and in simulation environments.

Note that the iterative modeling of causal relations (*steps 4.–7.* in Figure 1) require real world data, as ensuring and demonstrating the validity of simulation environments is an unsolved problem. While there exists research regarding the validation of simulation environments and simulation models in the context of automated driving [43], [47]–[50], there exist no well defined and widely applicable validity metrics and thresholds. However, once a CP's causal relation is established and validly implemented, a simulation can be used to determine the effectiveness of SPs (*steps 8.-11.* in Figure 1).

A. REAL WORLD DATA

Once an initial expert-based causal relation has been established, it is necessary to collect data for its iterative validation (*steps 4.-6.* in Figure 1). In particular, we need data for the causal structure's relevant variables to evaluate the causality indicator functions. Here, the relevance of variables regarding the data collection process can be defined in terms of adjustment sets.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

In the example of Figure 7, the set $S = \{$*road surface*, *degree of wetness*$\}$ (orange nodes) is an adjustment set for the causal effect of the CP *reduced coefficient of friction* on $\varphi = \mathrm{BTN}$. Hence, we can compute the ρ_i and σ if we collect data for the CP, the criticality metric φ, and the variables in S. This requires these variables to measurable in the real world. Being able to actively control these variables helps to limit costs and efforts for data collection. For example, if we need more data driving on *cobblestone*, we want to collect data on such road surfaces specifically.

Moreover, data collection is influenced by a causal relation's context in terms of location, presence of other traffic participants, and the general setting. For accurate results it is crucial to tightly estimate the CPDs for the relevant variables, as we calculate the causal effect according to equation (1). The more data points we collect for these variables, the better. Consider an example where we obtain bird-view traffic data from drones equipped with cameras hovering above a fixed location. Due to the type of data collection there will be few data points including rain, heavy wind, or gusts. Relying on such data for estimating causal effects will be problematic when weather-related variables are involved.

B. SIMULATION ENVIRONMENTS

During modeling and iterative refinement of a causal relation, some variables may not be available in the simulation environment, e.g., black ice. Even if the variable is implemented in the simulation, its effect on other variables in the causal relation might not, e.g., the effect of black ice on the road-tire friction.

In these cases, implementing a causal relation in a simulation leads to requirements on the simulator itself. The causal structure provides important information about the interfaces that every valid simulation model should necessarily have. Once a causal relation has been validated and shall be used inside a simulation to demonstrate the effectiveness of a given SP, the validity of the simulation environment regarding the context of the causal relation becomes essential.

Causal theory provides us with a formula for the causal effect that is independent of the choice of an adjustment set. As the same must be true for the causal effect in the simulation, we obtain an additional requirement regarding the simulation's validity. If this independence is not given, validity issues are likely to exist within the simulation environment.

Once the simulation's validity is assured in the given context, we require that every variable in the adjustment set can be recorded and controlled. Of course, the CP and the criticality metric must be recorded as well.

VI. CONCLUSION AND FUTURE WORK

In this work, we sketched how causal theory applied in the context of a criticality analysis can contribute to the safety of ADSs. In particular, we introduced a formalization of causal queries to improve our understanding of CP. For this, we defined a causal relation as a causal structure that models the effect of a CP on criticality combined with a context. Moreover, we took steps towards the assessment of a causal relation's modeling quality by introducing various comparative quantities. Our main example is a causal relation for the CP reduced coefficient of friction and its effect on the ADS-operated vehicle's required acceleration via the BTN.

There are limitations to this work: (i) As this work focuses on providing a theoretical basis, the approaches practical relevance has not been established. However, increasing this relevance is a hard problem as it requires large amounts of real world data, which also have to satisfy the requirements described in subsection V-A. (ii) So far, we have considered causal relations only on scene-level, i.e., independently of temporal developments. Contemporary approaches to ADS safety are mostly scenario-based. Hence, extending our approach to include dynamic developments seems necessary. Although there exist prior work on dynamic causal models [51], [52], the inclusion of time-dependent variables is a weakness of causal theory.

This article leaves open many possible extensions of the presented framework and its application. Let us briefly consider some of these possibilities for future work.

Temporal Causal Dependencies: The mutual interactions between traffic participants may introduce cycles in the graph when modeled explicitly on scenario level. For this, we propose the usage of dynamic causal models, where each node represents a stochastic process and edges are modeled with consideration of time, allowing for mutual interactions without introducing cycles.

Modularity of Causal Models (regarding Clustered Components): Due to the combinatorial character of contexts when including multiple traffic participants, it is beneficial to be able to easily add or remove entities from causal models. Further, causal models may need to be updated to changes in traffic, e.g., the introduction of personal mobility devices. As causal models and fault trees share some key features, e.g. being graph-based, prior work on the modularization of fault trees should be considered. For Bayesian networks. there exists a modularization approach by Koller and Pfeffer [53].

Modularity of Causal Models (regarding Criticality Phenomena): So far we have only considered causal models for single criticality phenomena. However, synergies between criticality phenomena must be examined to obtain a complete understanding of the phenomenon in each context. Since there is a large variety of criticality phenomena, we propose the examination of interlocking of causal models for different criticality phenomena as future work. One could imagine that the causal models might be stored in a connected database, thus sharing knowledge of the possibly relevant synergies, similar to the CHIELD database for causal hypotheses in evolutionary linguistics [54].

Validation of Causal Models: The derivation of valid SPs requires good estimates of the CPDs for the adjustment set variables, the CP, and the criticality metric. Therefore, the quantitative results of the proposed causal analysis hinge on data quality. Future work should define guidance when the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

quantitative results can be considered significant.

Calculus for Stochastic Interventions: For now, we have only considered interventions that set variables to a specific value. However, stochastic interventions provide a theory to adjust the distributions of the variables of a causal model [55]. Extending our approach in this direction could be a powerful tool to gather useful information about the model or to model SPs as stochastic intervention on variables.

Overall Safety Argumentation: The report of the German ethics commission on automated and connected driving requires a demonstration of a reduction of harm to be causal in nature: 'The licensing of automated systems is not justifiable unless it promises to produce at least a diminution in harm compared with human driving' [56, p. 10]. For this, the causal models presented in this work may be applied in a more abstract overarching safety argumentation. Particularly, the fundamental question whether the introduction of an ADS results in less harm than had it not been introduced points to a causative understanding of a positive risk balance [57].

References

[1] SAE International, "J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 2021.

[2] C. Neurohr, L. Westhofen, M. Butz, M. H. Bollmann, U. Eberle, and R. Galbas, "Criticality Analysis for the Verification and Validation of Automated Vehicles," *IEEE Access*, vol. 9, pp. 18016–18041, 2021.

[3] J. Pearl *et al.*, "Causal inference in statistics: An overview," *Statistics surveys*, vol. 3, pp. 96–146, 2009.

[4] A. Poddey, T. Brade, J. E. Stellet, and W. Branz, "On the Validation of Complex Systems Operating in Open Contexts," *arXiv preprint arXiv:1902.10517*, 2019.

[5] W. Wachenfeld and H. Winner, "The Release of Autonomous Vehicles," in *Autonomous Driving: Technical, Legal and Social Aspects* (M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, eds.), pp. 425–449, Berlin, Heidelberg: Springer Berlin Heidelberg, 2016.

[6] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016.

[7] C. Neurohr, L. Westhofen, T. Henning, T. de Graaff, E. Möhlmann, and E. Böde, "Fundamental Considerations around Scenario-Based Testing for Automated Driving," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 121–127, 2020.

[8] S. Babisch, C. Neurohr, L. Westhofen, S. Schoenawa, and H. Liers, "Leveraging the GIDAS Database for the Criticality Analysis of Automated Driving Systems," *Journal of Advanced Transportation*, vol. 2023, no. 1, p. 1349269, 2023.

[9] International Organization for Standardization, "ISO 26262: Road vehicles – Functional safety," 2018.

[10] L. Westhofen, C. Neurohr, T. Koopmann, M. Butz, B. Schütt, F. Utesch, B. Neurohr, C. Gutenkunst, and E. Böde, "Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art," *Archives of Computational Methods in Engineering*, 2022.

[11] E. Bareinboim, J. D. Correa, D. Ibeling, and T. Icard, *On Pearl's Hierarchy and the Foundations of Causal Inference*, p. 507–556. New York, NY, USA: Association for Computing Machinery, 1 ed., 2022.

[12] J. Pearl, *Causality*. Cambridge University Press, 2 ed., 2009.

[13] I. Shpitser and J. Pearl, "Identification of joint interventional distributions in recursive semi-markovian causal models," in *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, AAAI'06, p. 1219–1226, AAAI Press, 2006.

[14] A. Ankan and A. Panda, "pgmpy: Probabilistic Graphical Models using Python," in *Proceedings of the 14th Python in Science Conference* (K. Huff and J. Bergstra, eds.), pp. 6 – 11, 2015.

[15] G. Ducamp, C. Gonzales, and P.-H. Wuillemin, "aGrUM/pyAgrum : a toolbox to build models and algorithms for Probabilistic Graphical Models in Python," in *10th International Conference on Probabilistic Graphical Models*, vol. 138 of *Proceedings of Machine Learning Research*, (Skørping, Denmark), pp. 609–612, Sept. 2020.

[16] A. Sharma and E. Kiciman, "Dowhy: An end-to-end library for causal inference," *arXiv preprint arXiv:2011.04216*, 2020.

[17] Y. Shimoni, E. Karavani, S. Ravid, P. Bak, T. H. Ng, S. H. Alford, D. Meade, and Y. Goldschmidt, "An Evaluation Toolkit to Guide Model Selection and Cohort Definition in Causal Inference," *arXiv preprint arXiv:1906.00442*, 2019.

[18] B. G. Marcot and T. D. Penman, "Advances in Bayesian network modelling: Integration of modelling technologies," *Environmental modelling & software*, vol. 111, pp. 386–393, 2019.

[19] P. Weber, G. Medina-Oliva, C. Simon, and B. Iung, "Overview on bayesian networks applications for dependability, risk analysis and maintenance areas," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 4, pp. 671–682, 2012.

[20] M. Rauschenbach, *Probabilistische Grundlage zur Darstellung integraler Mehrzustands-Fehlermodelle komplexer technischer Systeme*. PhD thesis, 2017.

[21] M. Rauschenbach and J. Nuffer, "Quantitative FMEA and Functional Safety Metrics Evaluation in Bayesian Networks," pp. 2475–2482, 2019.

[22] T. A. Wheeler, M. J. Kochenderfer, and P. Robbel, "Initial scene configurations for highway traffic propagation," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 279–284, 2015.

[23] A. Adee, R. Gansch, and P. Liggesmeyer, "Systematic Modeling Approach for Environmental Perception Limitations in Automated Driving," in *2021 17th European Dependable Computing Conference (EDCC)*, pp. 103–110, 2021.

[24] A. Adee, R. Gansch, P. Liggesmeyer, C. Glaeser, and F. Drews, "Discovery of perception performance limiting triggering conditions in automated driving," in *2021 5th International Conference on System Reliability and Safety (ICSRS)*, pp. 248–257, 2021.

[25] R. Maier and J. Mottok, "BayesianSafety - An Open-Source Package for Causality-Guided, Multi-model Safety Analysis," in *Computer Safety, Reliability, and Security* (M. Trapp, F. Saglietti, M. Spisländer, and F. Bitsch, eds.), (Cham), pp. 17–30, Springer International Publishing, 2022.

[26] International Organization for Standardization, "ISO 21448: Road vehicles – Safety of the intended functionality," 2022.

[27] W. E. Vesely, F. F. Goldberg, N. H. Roberts, and D. F. Haasl, "Fault Tree Handbook," tech. rep., Nuclear Regulatory Commission Washington DC, 1981.

[28] T. Ishimatsu, N. G. Leveson, J. Thomas, M. Katahira, Y. Miyamoto, and H. Nakao, "Modeling and Hazard Analysis using STPA," 2010.

[29] B. Kramer, C. Neurohr, M. Büker, E. Böde, M. Fränzle, and W. Damm, "Identification and Quantification of Hazardous Scenarios for Automated Driving," in *Model-Based Safety and Assessment* (M. Zeller and K. Höfig, eds.), (Cham), pp. 163–178, Springer International Publishing, 2020.

[30] E. Böde, M. Büker, W. Damm, M. Fränzle, B. Kramer, C. Neurohr, and S. Vander Maelen, "Identifikation und Quantifizierung von Automationsrisiken für hochautomatisierte Fahrfunktionen," tech. rep., OFFIS e.V., 2019.

[31] C. Neurohr, L. Westhofen, M. Butz, M. H. Bollmann, L. Putze, T. Koopmann, R. Gansch, M. Knoop, A. Rasch, B. Cojocaru, and J. Daube, "Advances on the Criticality Analysis for Automated Driving Systems," Mar. 2024.

[32] C. Linnhoff, P. Rosenberger, S. Schmidt, L. Elster, R. Stark, and H. Winner, "Towards Serious Perception Sensor Simulation for Safety Validation of Automated Driving - A Collaborative Method to Specify Sensor Models," in *24th International Conference on Intelligent Transportation Systems (ITSC)*, (Darmstadt), IEEE, 2021. Veranstaltungstitel: 24th International Conference on Intelligent Transportation Systems (ITSC).

[33] C. Linnhoff, K. Hofrichter, L. Elster, P. Rosenberger, and H. Winner, "Measuring the influence of environmental conditions on automotive lidar sensors," *Sensors*, vol. 22, no. 14, 2022.

[34] R. Maier, L. Grabinger, D. Urlhart, and J. Mottok, "Towards Causal Model-Based Engineering in Automotive System Safety," in *Model-Based Safety and Assessment* (C. Seguin, M. Zeller, and T. Prosvirnova, eds.), (Cham), pp. 116–129, Springer International Publishing, 2022.

[35] R. Maier, L. Grabinger, D. Urlhart, and J. Mottok, "Causal models to support scenario-based testing of adas," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[36] L. Westhofen, C. Neurohr, M. Butz, M. Scholtes, and M. Schuldes, "Using Ontologies for the Formalization and Recognition of Criticality for Automated Driving," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 519–538, 2022.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

[37] Association for Standardization of Automation and Measuring Systems (ASAM), "ASAM OpenXOntology," 2022.
[38] M. Scholtes, L. Westhofen, L. R. Turner, K. Lotto, M. Schuldes, H. Weber, N. Wagener, C. Neurohr, M. H. Bollmann, F. Körtke, J. Hiller, M. Hoss, J. Bock, and L. Eckstein, "6-Layer Model for a Structured Description and Categorization of Urban Traffic and Environment," *IEEE Access*, vol. 9, pp. 59131–59147, 2021.
[39] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, 2017.
[40] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
[41] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schölkopf, "Quantifying causal influences," *The Annals of Statistics*, vol. 41, no. 5, pp. 2324 – 2358, 2013.
[42] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, 2012.
[43] B. Neurohr, T. Koopmann, E. Möhlmann, and M. Fränzle, "Determining the Validity of Simulation Models for the Verification of Automated Driving Systems," *IEEE Access*, 2023.
[44] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Conference on robot learning*, pp. 1–16, PMLR, 2017.
[45] C. Lex, *Maximum Tire-Road Friction Coefficient Estimation*. PhD thesis, 2015.
[46] M. M. Zefreh and A. Török, "Distribution of traffic speed in different traffic conditions: An empirical study in budapest," *Transport*, vol. 35, no. 1, pp. 68–86, 2020.
[47] S. Riedmaier, D. Schneider, D. Watzenig, F. Diermeyer, and B. Schick, "Model Validation and Scenario Selection for Virtual-Based Homologation of Automated Vehicles," *Applied Sciences*, vol. 11, 2020.
[48] P. Rosenberger, J. T. Wendler, M. F. Holder, C. Linnhoff, M. Berghöfer, H. Winner, and M. Maurer, "Towards a Generally Accepted Validation Methodology for Sensor Models - Challenges, Metrics, and First Results," 2019.
[49] P. Rosenberger, M. F. Holder, N. Cianciaruso, P. Aust, J. F. Tamm-Morschel, C. Linnhoff, and H. Winner, "Sequential lidar sensor system simulation: a modular approach for simulation-based safety validation of automated driving," *Automotive and Engine Technology*, vol. 5, no. 3, pp. 187–197, 2020.
[50] B. Neurohr, T. de Graaff, A. Eggers, T. Bienmüller, and E. Möhlmann, *Providing Evidence for the Validity of the Virtual Verification of Automated Driving Systems*, pp. 5–13. 03 2024.
[51] G. Blondel, M. Arias, and R. Gavaldà, "Identifiability and Transportability in Dynamic Causal Networks," *International Journal of Data Science and Analytics*, vol. 3, 2017.
[52] R. Srinivasan, J. J. R. Lee, R. Bhattacharya, and I. Shpitser, "Path dependent structural equation models," in *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence* (C. de Campos and M. H. Maathuis, eds.), vol. 161 of *Proceedings of Machine Learning Research*, pp. 161–171, PMLR, 2021.
[53] D. Koller and A. Pfeffer, "Object-Oriented Bayesian Networks," 2013.
[54] S. G. Roberts, A. Killin, A. Deb, C. Sheard, S. J. Greenhill, K. Sinnemäki, J. Segovia-Martín, J. Nölle, A. Berdicevskis, A. Humphreys-Balkwill, H. Little, C. Opie, G. Jacques, L. Bromham, P. Tinits, R. M. Ross, S. Lee, E. Gasser, J. Calladine, M. Spike, S. F. Mann, O. Shcherbakova, R. Singer, S. Zhang, A. Benítez-Burraco, C. Kliesch, E. Thomas-Colquhoun, H. Skirgård, M. Tamariz, S. Passmore, T. Pellard, and F. Jordan, "CHIELD: the causal hypotheses in evolutionary linguistics database," *Journal of Language Evolution*, 2020.
[55] J. Correa and E. Bareinboim, "A Calculus for Stochastic Interventions: Causal Effect Identification and Surrogate Experiments," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 06, pp. 10093–10100, 2020.
[56] U. Di Fabio, M. Broy, R. Brüngger, U. Eichhorn, A. Grunwald, D. Heckmann, E. Hilgendorf, H. Kagermann, A. Losinger, M. Lutz-Bachmann, *et al.*, "Ethics Commission: Automated and Connected Driving," tech. rep., German Federal Ministry of Transport and Digital Infrastructure, 2017.
[57] T. Brade, B. Kramer, and C. Neurohr, "Paradigms in Scenario-Based Testing for Automated Driving," in *2021 International Symposium on Electrical, Electronics and Information Engineering*, pp. 108–114, 2021.
[58] S. Khaleghian, A. Emami, and S. Taheri, "A technical survey on tire-road friction estimation," *Friction*, vol. 5, pp. 123–146, 2017.
[59] B. N. J. Persson, "Rubber friction: role of the flash temperature.," *Journal of physics. Condensed matter : an Institute of Physics journal*, vol. 18 32, pp. 7789–823, 2006.
[60] R. Brach and M. Brach, "The Tire-Force Ellipse (Friction Ellipse) and Tire Characteristics," *SAE 2011 World Congress and Exhibition*, vol. 2296, 2011.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

TJARK KOOPMANN received the B.Sc. and M.Sc. degrees in mathematics from the University of Oldenburg, Germany, in 2018 and 2020, respectively. Since then, he has been working at the transportation division of the research institute OFFIS e.V., that became the institute for systems engineering for future mobility of the German Aerospace Center (DLR e.V.) at the start of 2022. His main research area is in the scenario-based safeguarding of automated driving systems, especially in the domain of causal analysis of criticality increasing factors. Further, he is also concerned with research regarding methods for the validation of simulation models and simulation environments.

LINA PUTZE received the B.Sc. and M.Sc. degrees in mathematics from the University of Münster in 2016 and 2019, specializing on the topics of stochastic processes, probability theory and its applications. She is currently working as a researcher at the German Aerospace Center (DLR) e.V. Institute of Systems Engineering for Future Mobility. The focus of her research is on methods to ensure safety of highly automated transport systems, including the identification and analysis of hazards and risk triggering scenario properties, causal analysis and risk assessment.

LUKAS WESTHOFEN received the degrees of B.Sc. and M.Sc. (with honors) in computer science by the RWTH Aachen University in 2015 and 2019, specializing on the topics of probabilistic programs and software verification. He currently works at the German Aerospace Center (DLR e.V.) Institute of Systems Engineering for Future Mobility, where his research focuses on developing methods to establish confidence in the safety of automated driving systems. As a PhD student at the TU Dortmund University, he investigates temporal, ontology-based querying as a tool for analyzing safety-critical scenarios.

ROMAN GANSCH received the B.Sc., M.Sc. and PhD degrees in electrical engineering with a focus on microelectronics from the Vienna University of Technology, Austria, in 2009, 2011, 2014 respectively. He has further received the M.Sc. degree in safety and systems engineering from the FH Campus Vienna, Austria, in 2019. From 2015 he worked as functional safety engineer for engine control units at Robert Bosch AG in Vienna, Austria, and switched to the central research department of Robert Bosch GmbH in Renningen, Germany in 2019. As research engineer he works on safety analysis methods for complex systems with a focus on probabilistic models and causal inference.

AHMAD ADEE received B.Sc. in mechatronics from the National University of Sciences and Technology, Islamabad Pakistan, and M.Sc. degree in robotics from Politechnika Warszawska, Poland and Ecole Centrale de Nantes, France in 2014 and 2018 respectively and PhD degree (Dr. Ing.) in computer sciences from RPTU Kaiserslautern-Landau, Germany in 2023. After a short period as a Safety Engineer with the TomTom N.V Netherlands, he started as SOTIF Expert at BMW AG. The focus of his research is on methods to ensure the safety of highly automated transport systems, including the identification and analysis of triggering conditions, causal and statistical analysis as well as defining verification and validation methods and strategies for such systems.

CHRISTIAN NEUROHR received the B.Sc. and M.Sc. degrees in mathematics in 2011 and 2013 respectively, both from Technische Universität Kaiserslautern, Germany and the Ph.D. degree (Dr. rer. nat.) from Carl von Ossietzky Universität Oldenburg, Germany in 2018. After a short period as a visiting researcher with the MAGMA group at the University of Sydney, he started his occupation as a postdoctoral researcher at the German Aerospace Center (DLR e.V.) Institute of Systems Engineering for Future Mobility where he is working in the area of scenario-based verification and validation of automated vehicles. Since 2023 he leads the 'Criticality Analysis' team within the division 'Theory and Design'.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

VII. APPENDIX

A. LIST OF VARIABLES FOR THE DETAILED CAUSAL STRUCTURE FOR THE REDUCED COEFFICIENT OF FRICTION.

Table 4: Variables in the detailed causal structure for the CP *reduced coefficient of friction*, cf. Figure 9, and their properties.

Variable	Interpretation	Range and Unit
Date and time of day	Time in a fixed time zone	$\{0,\dots,23\} \times \{0,\dots,59\}$
Global location	Area of the world where the scenario takes place	$\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [-\pi, \pi]$
Weather	Weather at the location where the scenario takes place	$\{\text{warm, cold, } \dots\} \times \{\text{clear sky, cloudy, } \dots\}$
Humidity	Average current relative humidity over the area of the scenario	$[0, \infty)$, %
Dew point	Average current dew point over the area of the scenario	$[0, \infty)$, K
Precipitation	Average precipitation in the general area of the scenario	$[0, \infty)$, $\mathrm{mm\,m^{-2}\,h^{-1}}$
Road list	List of relevant road networks	{road_1, …, road_n}
Relative position of ego	Position of ego relative to fixed origin	$(-\infty, \infty)^3$, m
Speed limit of ego	Minimal speed limit for ego for the next 200m	$[0, \infty)$, $\mathrm{m\,s^{-1}}$
Ego distance to next ISRL	Distance of ego to next intersection stopping requirement location, e.g. traffic light or stop sign	$[0, \infty)$, m
Target position of ego	Planned target location of ego relative to the origin	$(-\infty, \infty)^3$, m
Ego road curvature	Curvature of the road segment the ego is currently driving on	$[0, \infty)$, $\mathrm{m^{-1}}$
Global illumination	Average illuminance in a designated area around the ego	$[0, \infty)$, $\mathrm{cd\,m^{-2}}$
Air temperature	Average air temperature throughout the scenario state space	$[0, \infty)$, K
Degree of Wetness	Average thickness of water film on the road in area around ego	$[0, \infty)$, m
Winter slipperiness	Relative area of the road covered with ice in area around ego	$[0, 100]$, %
Road surface material	Material of the road surface in a designated area around ego	{asphalt, gravel, … }
Road surface contamination	Average contamination of the road surface, e.g. leaves or oil	$[0, \infty)^n$ for $n = \lvert \{\text{leaves, soil, oil, } \dots\} \rvert$
Road surface degradation	General condition of the road surface using the pavement condition index	$\{0, \dots, 100\}$
Road surface roughness	Aggregated distance of sample points to the mean road surface	$[0, \infty)$, m
Planned steering	Planned average steering of ego for the next x seconds	$\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$
Planned acceleration	Planned average acceleration of ego for the next x seconds	$[0, \infty)$, $\mathrm{m\,s^{-2}}$
Ego tire temperature	Temperature of the tires	$[0, \infty)$, K
Ego tire flash temp.	Aggregate of the temperature of the tires at the contact patch	$[0, \infty)$, K
Tire type	General classification of the tire type	{Touring, Winter, … }
Wet grip	EU tire label regarding wetgrip	{A,B,C,E,F}
Tire pressure	Aggregated tire pressure, e.g. minimum or average	$[0, \infty)$, $\mathrm{N\,m^{-2}}$
Maximal braking torque	Ability to brake at the brakes themselves	$[0, \infty)$, N m
Ego vehicle mass	Mass of ego including vehicle load	$(0, \infty)$, kg
Forward velocity of ego	Velocity of the ego vehicle relative to the road	$\mathbb{R}^3$, $\mathrm{m\,s^{-1}}$
Hub velocity of ego	Velocity of the ego measured by the system	$\mathbb{R}^3$, $\mathrm{m\,s^{-1}}$
Coefficient of friction	Average coefficient of friction between road surface and tires	$[0, \infty)$, 1
Max. avail. long. dec.	Maximal available longitudinal deceleration of the ego vehicle	$[0, \infty)$, $\mathrm{m\,s^{-2}}$
Max. avail. lat. dec.	Maximal available lateral deceleration of the ego vehicle	$[0, \infty)$, $\mathrm{m\,s^{-2}}$
Max. req. long. dec.	Maximal required longitudinal deceleration from the driving task	$[0, \infty)$, $\mathrm{m\,s^{-2}}$
Max. req. lat. dec.	Maximal required lateral deceleration from the driving task	$[0, \infty)$, $\mathrm{m\,s^{-2}}$
STN	Steer Thread Number	$[0, \infty)$, 1
BTN	Brake Thread Number	$[0, \infty)$, 1
agg(STN, BTN)	Aggregate of BTN and STN	$[0, \infty)$, 1

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3555177

B. DETAILED CAUSAL STRUCTURE FOR REDUCED COEFFICIENT OF FRICTION

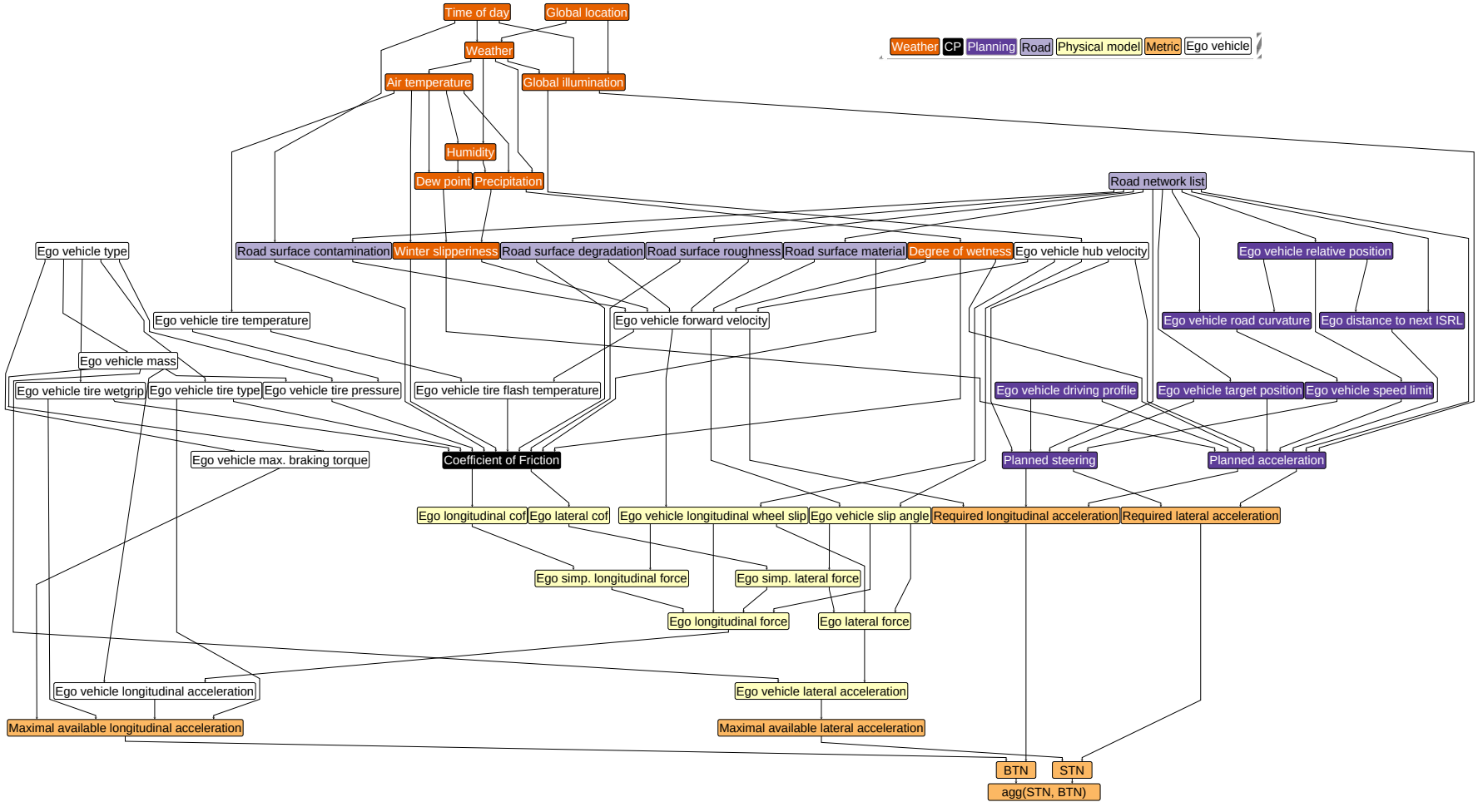


Figure 9: Highly detailed causal structure for the CP reduced coefficient of friction based on current research [58]–[60]. Due to its size, this causal structure is almost impossible to instantiate with limited resources. However, it describes the causalities underlying the coefficient of friction more accurately than the causal structure of Figure 7. Moreover, in this graph, variables are grouped according to areas of similarity, which can be considered a first step towards modularization of causal relations.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/