vehicles

MDPI

Article

A Calculation Method for Automated Driving Test Scenario Complexity

Jie Zeng [1,2], Ling Zheng [1,*] and Haozhong Wang [2]

[1] College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China; cjzengjie@cmhk.com
[2] China Merchants Vehicle Testing & Technology Research Institute Co., Ltd., Chongqing 401329, China
* Correspondence: zling@cqu.edu.cn; Tel.: +86-138-8397-3824

Abstract

Scenario-based validation requires efficient screening of large test libraries, but existing complexity metrics often combine traffic hazard and automated-driving-system workload into a single construct. This study proposes a two-dimensional framework that separates scenario hazardousness from system processing difficulty. Hazardousness combines normalized acceleration and time-to-collision measures. Processing difficulty aggregates perception, decision-making, and execution demands, with weights obtained from a triangular fuzzy analytic hierarchy process. The framework was compared with four baseline metrics using 2000 Latin-hypercube-sampled cut-in scenarios and was explored in 22 closed-field scenarios tested with three production Level 2 vehicles. The proposed metric produced a broad and less concentrated distribution of cases across the normalized score range. However, total complexity was not significantly correlated with any vehicle score in the 22-case dataset; therefore, the field experiment supports only preliminary discriminant analysis, not predictive validity. A 10% hierarchical-weight perturbation analysis showed that the three largest global weights remained the same set in 82.3% of 10,000 simulations. The model is currently limited by empirically assigned perception constants, a small field sample, few ultra-high-complexity cases, and omission of road curvature, crosswinds, tire condition, and actuator delay. The additive aggregation assumes that interactions among criteria can be neglected for the intended comparison; this assumption has not been validated. The framework is best viewed as an interpretable scenario-screening tool requiring further calibration and validation.

Keywords: automated driving; scenario complexity; safety validation

Check for updates

Academic Editors: Ralf Stetter and Mohammed Chadli

Received: 5 July 2026
Revised: 11 September 2026
Accepted: 24 September 2026
Published: 25 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Safety validation of automated driving systems has gradually shifted from mileage-based testing to scenario-based testing [1]. Test scenarios provide standardized conditions for eliciting the functional and performance limits of automated driving systems and for comparing safety performance across platforms [2]. As operational design domains expand, the number, range, and spatiotemporal coupling of scenario elements increase rapidly. Simple scenario counts or single risk indicators therefore cannot efficiently identify scenarios with high testing value [3]. An interpretable, computable, and verifiable framework for quantifying scenario complexity is needed to support scenario grading, rational allocation of testing resources, and prioritization of high-risk test cases.

Previous studies define traffic-scenario complexity from several perspectives. Sussman described complex transportation systems as interconnected systems whose aggregate behavior can be difficult to predict [3]. Ma et al. related complexity to uncertainty in perception, cognition, and prediction [4]. Zhang et al. used observable drivable areas [5], Li et al. considered the number and influence of scenario elements [6], and Wang et al. combined information entropy with a gravitational model [7]. Zhu et al. linked scenario elements to perception, decision-making, and execution modules [8], whereas Sun et al. incorporated operational-design conditions, connected functions, driver perception, and vehicle capability [9].

International research has increasingly organized automated-driving validation around structured scenarios. Menzel et al. distinguished functional, logical, and concrete scenario levels [10]. Riedmaier et al. reviewed scenario-based safety assessment, including coverage, data sources, and evaluation metrics [11]. Neurohr et al. developed a criticality-analysis process for identifying causal factors in open-traffic contexts [12]. The European HEADSTART project connected simulation, X-in-the-loop, proving ground, and field testing within a harmonized methodology [13]. Nalic et al. used stress testing to increase the occurrence of safety-critical virtual scenarios [14], and Ponn et al. identified challenging highway scenarios from real-driving data [15]. These contributions strengthen scenario generation and coverage analysis, but they generally do not provide a single interpretable score that explicitly separates inherent traffic hazard from perception, planning, and control burden. This distinction motivates the present framework.

Although these studies provide a theoretical foundation for complexity assessment, substantial differences remain in their evaluation objects, indicator boundaries, and validation strategies. Two principal limitations can be identified:

- Some studies equate scenario risk with scenario complexity or evaluate complexity only in terms of system processing load. Consequently, they do not provide a unified framework that distinguishes inherent scenario hazards from the processing demands imposed on an automated driving system.
- Existing methods provide insufficient characterization of the spatiotemporal coupling among static environments, dynamic traffic participants, and sudden traffic events. Static aggregation or single-time-step indicators cannot capture how complexity evolves from hazard onset to the critical collision-avoidance phase.

Accordingly, this study defines scenario complexity as the combined challenge imposed on an automated driving system within a specified operational design domain and under a stated reference-system model. The study makes the following contributions:

- A two-dimensional framework separates inherent traffic hazardousness from the perception, decision-making, and execution demands imposed on a reference automated-driving system.
- The proposed metric is compared with four baseline metrics on 2000 Latin-hypercube-sampled cut-in scenarios, and its hierarchical weights are examined through a 10% perturbation analysis.
- An exploratory closed-field study uses 22 scenarios and three production Level 2 vehicles. The study reports the full-sample associations without deleting discordant cases and identifies the limits of the current evidence.

2. Theoretical Framework

The total scenario-complexity score C is a dimensionless scalar in [0, 1]. It combines the scalar hazardousness score Hs and scalar system-processing-difficulty score Ds. Hs measures ego-vehicle kinematic and interaction risk. Ds measures the demands placed on

the perception, decision-making, and execution modules of a defined reference automated-driving system.

The score is conditional rather than vehicle independent: it is calculated for fixed reference sensing ranges, proxy driving strategies, normalization thresholds, and indicator weights. Comparisons are therefore valid only when these assumptions are held constant.

Equations (1) and (2) define a fixed-weight additive approximation under the stated reference-system assumptions. Cross-criterion feedback, changes in relative importance with context, and interaction terms are not represented. The separation of hazardousness and processing difficulty is a conceptual organization of the indicators and does not demonstrate their independence. The resulting score is a candidate index for exploratory screening; its adequacy as a measure of overall complexity requires validation of the aggregation structure.

The contribution is not a new acceleration or time-to-collision equation. Those components are established risk measures. The contribution is their organization into two explicitly defined dimensions, the operationalization of perception, decision-making, and execution burden under a reference-system model, and their evaluation under a common normalization in simulation and closed-field data.

$$C = \omega_1 H_s + \omega_2 D_s \tag{1}$$

$$D_s = \omega_{21} D_{Per} + \omega_{22} D_{Pla} + \omega_{23} D_{Con} \tag{2}$$

The top-level weights sum to one. The three local weights under Ds also sum to one. All component indicators are normalized before aggregation, so Equations (1) and (2) map scalar inputs to a scalar output rather than performing matrix operations.

For example, if Hs = 0.70 and Ds = 0.50, with top-level weights 0.59 and 0.41, Equation (1) gives C = 0.618. If the perception, decision-making, and execution scores are 0.60, 0.40, and 0.30 and their local weights are 0.32, 0.46, and 0.22, Equation (2) gives Ds = 0.442.

2.1. Scenario Hazardousness

Scenario hazardousness quantifies potential safety threats at kinematic levels, consisting of ego-vehicle motion risk H_{self} and interactive risk with surrounding traffic participants H_{inter}. Normalized acceleration metrics and time-to-collision indicators are adopted for quantification separately. Nonlinear mapping functions are introduced to convert acceleration and time-to-collision values into dimensionless indicators within comparable ranges, given the nonlinear correlation between kinematic parameters and collision risks.

Equations (3)–(5) define ego-vehicle motion risk. The longitudinal and lateral accelerations at time t are first normalized by their respective threshold values and then mapped to [0, 1].

$$\varphi_{x-a}(t) = \left(\frac{a_x(t)}{a_{x_max}}\right)^3, \quad 01, \quad a_x(t) \geq a_{x_max} \tag{3}$$

$$\varphi_{y-a}(t) = \left(\frac{a_y(t)}{a_{y_max}}\right)^3, \quad 01, \quad a_y(t) \geq a_{y_max} \tag{4}$$

The curly braces in Equations (3) and (4) denote piecewise functions, not sets. The first branch applies below the lower threshold, the middle branch applies between the lower and upper thresholds, and the last branch caps the normalized risk at 1 above the upper threshold.

$$H_{self} = max\{\varphi_{x-a}(t), \varphi_{y-a}(t)\} \tag{5}$$

Equations (6)–(10) define interaction risk. Longitudinal and lateral time to collision are calculated from relative position and relative velocity between the ego vehicle and each target. The maximum normalized risk over targets and time is retained.

$$\varphi_{x-TTC}(t) = \frac{|d_{x-relative}(t)|}{|VUT_{x-speed}(t) - OBJ_{x-speed}(t)|} \tag{6}$$

$$\varphi_{y-TTC}(t) = \frac{|d_{y-relative}(t)|}{|VUT_{y-speed}(t) - OBJ_{y-speed}(t)|} \tag{7}$$

$$\varphi_{x_TTC_nor}(t) = \left(\frac{TTC_{max} - \varphi_{x-TTC}(t)}{TTC_{max}}\right)^3, \quad 0 < \varphi_{x-TTC}(t)0, \quad \varphi_{x-TTC}(t) > TTC_{max} \tag{8}$$

$$\varphi_{y_TTC_nor}(t) = \left(\frac{TTC_{max} - \varphi_{y-TTC}(t)}{TTC_{max}}\right)^3, \quad 0 < \varphi_{y-TTC}(t)0, \quad \varphi_{y-TTC}(t) > TTC_{max} \tag{9}$$

$$H_{inter} = max\{\varphi_{x_TTC_nor}(t), \varphi_{y_TTC_nor}(t)\} \tag{10}$$

Total scenario hazardousness is calculated through Equation (11). ω_{11} and ω_{12} stand for weight coefficients of ego-vehicle motion risk and interactive risk, which are solved via the fuzzy analytic hierarchy process.

$$H_s = \omega_{11} H_{self} + \omega_{12} H_{inter} \tag{11}$$

2.2. Scenario Perception Difficulty

Scenario perception difficulty represents the recognition and tracking burden imposed on the reference system by static obstacles and dynamic participants. All indicators were computed offline from synchronized simulation or test logs at 0.1 s intervals. The value at each interval was calculated first, and the maximum over the scenario duration was retained. This peak statistic captures the most demanding instant but not the duration of high workload.

$$Per(t) = \omega_{st} Per_{st}(t) + \omega_{dy} Per_{dy}(t) \tag{12}$$

$$D_{Per} = max\{Per(t)\} \tag{13}$$

A common reference sensing region was used for all test cases: 200 m ahead, 100 m behind, and 5 m on each side of the ego vehicle. Thus, the longitudinal interval was [−100 m, 200 m] and the lateral interval was [−5 m, 5 m]. These fixed boundaries enable comparisons within this study. Engineering applications should adjust them to the actual sensor configuration, target class, and effective detection probability.

2.2.1. Static Obstacle Perception Difficulty

Static obstacle perception difficulty is solved via Equation (14). Basic perception pressure quantifies inherent recognition difficulty of different static obstacles, while perception noise terms comprehensively integrate environmental visibility, illumination, weather conditions, target surface texture and obstacle occlusion ratio.

$$Per_{st}(t) = \log_2\left(1 + \frac{1 - e^{-\sum_{i=1}^{n} SPD_{i_st_pre} \cdot SPD_{i_st_noise}}}{1 + e^{-\sum_{i=1}^{n} SPD_{i_st_pre} \cdot SPD_{i_st_noise}}}\right) \tag{14}$$

The baseline pressure represents target-class recognition difficulty before environmental and occlusion effects are added. Common static targets share a value of 1.0 because no class-specific calibration dataset was available; atypical shapes are assigned 1.2 as an engineering prior (Table 1). Scenario-to-scenario variation is produced primarily by visibility, illumination, weather, texture, and observable-area terms. These values are not empirical detection probabilities and require calibration against sensor-performance data.

Table 1. Basic perception pressure of static obstacles.

Serial Number	Obstacle Category	Obstacle Type	Pressure Value
1	Road Obstacle	Traffic barrier	1.0
2	Road Obstacle	Traffic cone	1.0
3	Road Obstacle	Carton	1.0
4	Road Obstacle	Warning triangle	1.0
5	Road Obstacle	Concrete isolation strip	1.0
6	Road Obstacle	Special-shaped obstacle	1.2

The static-obstacle perception noise is defined in Equation (15).

$$SPD_{i_st_noise} = light_i(t) \cdot visibility_i(t) \cdot weather_i(t) \cdot quality_i(t) \cdot Tar_i(t) \quad (15)$$

Illumination, visibility, weather and target texture quality quantify environmental and target attribute interference on perception performance; observable feature proportion describes obstacle occlusion severity. Parameter values of each interference factor are listed in Table 2.

$$Tar_i(t) = 2 - \frac{S_{OC}(t)}{S_{tr}} \quad (16)$$

where S_{tr} denotes full forward projection area of unoccluded obstacles; $S_{OC}(t)$ represents observable forward projection area of obstacles at time step t.

Table 2. Parameter values for static-obstacle perception noise.

Serial Number	Variable	Value Range	Coefficient
1	Visibility	≥500 m	1.0
2	Visibility	200–500 m	1.1
3	Visibility	100–200 m	1.2
4	Visibility	≤100 m	1.4
5	Illumination	Daytime without backlight	1.0
6	Illumination	Daytime backlight/strong glare	1.1
7	Illumination	Night with road lighting	1.1
8	Illumination	Night without road lighting	1.2
9	Illumination	Night with strong dazzling light	1.3
10	Weather Condition	Clear sky	1.0
11	Weather Condition	Light rain	1.1
12	Weather Condition	Moderate rain	1.2
13	Weather Condition	Heavy rain	1.4
14	Weather Condition	Dense fog	1.4
15	Target Texture Quality	Normal	1.0
16	Target Texture Quality	Surface contamination	1.1
17	Target Texture Quality	Partial feature loss	1.2
18	Target Texture Quality	Color similar to background	1.4

2.2.2. Dynamic Traffic Participant Perception Difficulty

Equation (17) defines dynamic-participant perception difficulty. The baseline term represents target-class recognition difficulty, and the noise term represents environmental interference, occlusion, and tracking load.

$$Per_{dy}(t) = \log_2\left(1 + \frac{1 - e^{-\sum_{i=1}^{n} SPD_{i_dy_pre} \cdot SPD_{i_dy_noise}}}{1 + e^{-\sum_{i=1}^{n} SPD_{i_dy_pre} \cdot SPD_{i_dy_noise}}}\right) \quad (17)$$

For dynamic targets, baseline pressure has the same interpretation. Common vehicles, riders, and pedestrians are assigned 1.0, whereas atypical vehicles and tricycles are assigned 1.2 (Table 3). The equal values avoid unsupported distinctions among common classes;

environmental interference, occlusion, and relative motion provide the differential load. These constants remain engineering priors rather than calibrated parameters.

Table 3. Basic perception pressure of dynamic traffic participants.

Serial Number	Target Category	Target Type	Pressure Value
1	Motor Vehicle	Passenger car	1.0
2	Motor Vehicle	Truck	1.0
3	Motor Vehicle	Bus	1.0
4	Motor Vehicle	Special-shaped vehicle	1.2
5	Two/Three-Wheeler	Bicycle	1.0
6	Two/Three-Wheeler	Electric bicycle	1.0
7	Two/Three-Wheeler	Motorcycle	1.0
8	Two/Three-Wheeler	Tricycle	1.2
9	Pedestrian	Adult pedestrian	1.0
10	Pedestrian	Child pedestrian	1.0

Dynamic target perception noise coefficient is defined in Formula (18).

$$SPD_{i_dy_noise} = light_i(t) \cdot visibility_i(t) \cdot weather_i(t) \cdot quality_i(t) \cdot Tar_i(t) \cdot motivation_i(t) \quad (18)$$

Equations (19)–(21) use relative velocity to represent target-tracking load. This choice is intentionally simple. Acceleration, yaw rate, and trajectory uncertainty should be added when suitable calibration data are available.

$$\Delta v = |v_1 - v_2| \quad (19)$$

$$motivation_i = 1.5 - 0.5e^{-\Delta v} \quad (20)$$

$$motivation_i \in [1, 1.5) \quad (21)$$

2.3. Scenario Decision-Making Difficulty

Scenario decision-making difficulty contains three components: traffic-rule constraints, available drivable space, and sudden actions by other road users. Equations (22) and (23) aggregate these components and retain the maximum value over time.

$$Pla(t) = \omega_{rul} \cdot Rule(t) + \omega_{reg} \cdot Region(t) + \omega_{acc} \cdot Accident(t) \quad (22)$$

$$D_{Pla} = max\{Pla(t)\} \quad (23)$$

2.3.1. Traffic Regulation Complexity

Traffic-rule complexity represents the combined planning burden of simultaneously active rules and potential right-of-way conflicts. The initial coefficients in Table 4 use traffic-penalty severity only as an engineering proxy; penalty severity is not equivalent to algorithmic difficulty. Applications in another jurisdiction should recalibrate the coefficients for local rules, right-of-way conflicts, and the number of feasible planning branches.

$$Rule(t) = \log_2\left(1 + \frac{1 - e^{-\sum_{i=1}^{n} R_i}}{1 + e^{-\sum_{i=1}^{n} R_i}}\right) \quad (24)$$

Table 4. Initial weight coefficients of traffic regulation items.

Serial Number	Target Category	Target Type	Penalty Points	Pressure Value
1	Lane Change Violation	Lane crossing over solid lines/guide lines	1 demerit point	1.0
2	Speeding	Driving beyond upper/lower speed limits	10–20% over speed limit, 3 demerit points	1.1
3	Improper Lane Usage	Continuous lane line crossing	3 demerit points	1.2
4	Right-of-Way Violation	Failure to yield to pedestrians/priority vehicles at intersections	3 demerit points	1.3
5	Signal Light Violation	Incorrect response to traffic signal lights	6 demerit points	1.5

2.3.2. Drivable Space Area

$Region(t)$ represents the lateral space available to the ego vehicle during a traffic conflict. It is zero when no conflicting obstacle occupies the lane. Otherwise, Equations (25) and (26) combine the real-time drivable width and width-variation rate $\varepsilon(t)$.

$$\varepsilon(t) = \frac{w(t)-w(t-0.1)}{0.1} \tag{25}$$

$$Region(t) = \log_2\left(1 + \frac{1-e^{-[1-\frac{w(t)}{L}][1+\varepsilon(t)]}}{1+e^{-[1-\frac{w(t)}{L}][1+\varepsilon(t)]}}\right) \tag{26}$$

where $w(t)$ denotes real-time drivable lateral width at time step t; L represents standard lane width.

2.3.3. Sudden Traffic Incidents

Sudden traffic incidents are abrupt changes in the states of surrounding traffic participants that disrupt the normal motion of the ego vehicle. Incident difficulty is calculated using Equation (27), and the corresponding parameter values are listed in Table 5.

$$Accident(t) = \log_2\left(1 + \frac{1-e^{-\sum_{i=1}^{n} N_i}}{1+e^{-\sum_{i=1}^{n} N_i}}\right) \tag{27}$$

Table 5. Parameter Values of Sudden Traffic Incidents.

Serial Number	Target Category	Incident Type	Weight Coefficient
1	Pedestrian	Reverse turning back	1.00
2	Pedestrian	Sudden road crossing	1.30
3	Pedestrian	Collapse on road	1.34
4	Two/Three-Wheeler	Sudden rollover	1.24
5	Two/Three-Wheeler	Sudden stop	1.28
6	Two/Three-Wheeler	Serpentine driving	1.29
7	Two/Three-Wheeler	Sudden lane cut-in	1.58
8	Motor Vehicle	Emergency braking	1.18
9	Motor Vehicle	Emergency lane cut-in	1.64
10	Motor Vehicle	Cargo scattering	1.71

2.4. Scenario Execution Difficulty

Scenario execution difficulty represents road-induced demands on motion control and actuator response. The current implementation includes road gradient, pavement adhesion, and pavement roughness, as defined in Equations (28) and (29).

$$Con(t) = \omega_\alpha \cdot \alpha(t) + \omega_\beta \cdot \beta(t) + \omega_\gamma \cdot \gamma(t) \tag{28}$$

$$D_{Con} = max\{Con(t)\} \tag{29}$$

Road gradient, pavement adhesion coefficient and pavement roughness quantify longitudinal load variation, tire-road maximum available adhesion and road disturbance, respectively. The current indicator system excludes road curvature, crosswind interference, tire condition and actuator performance boundaries, with applicable scope limitations elaborated in the Discussion section.

$$\alpha(x) = \begin{cases} 0, & Slope < 0.03 \\ 0.1, & 0.03 \leq Slope < 0.06 \\ 0.3, & 0.06 \leq Slope < 0.09 \\ 0.5, & 0.09 \leq Slope < 0.12 \\ 0.6, & 0.12 \leq Slope < 0.15 \\ 1, & Slope \geq 0.15 \end{cases} \tag{30}$$

$$\beta(x) = \begin{cases} 0, & Friction \geq 0.7 \\ 0.3, & 0.5 \leq Friction < 0.7 \\ 0.6, & 0.35 \leq Friction < 0.5 \\ 0.8, & 0.2 \leq Friction < 0.35 \\ 1, & Friction < 0.2 \end{cases} \tag{31}$$

$$\gamma(x) = \begin{cases} 0, & IRI < 2.0 \\ 0.4, & 2.0 \leq IRI < 3.5 \\ 0.6, & 3.5 \leq IRI < 5.0 \\ 1, & IRI \geq 5.0 \end{cases} \tag{32}$$

The current execution-difficulty index includes only road gradient, pavement adhesion, and pavement roughness. Road curvature, crosswinds, tire condition, and actuator response delay are not included, which limits the scope of the model. The framework is therefore most applicable to straight or gently curved roads under conventional weather conditions. Additional indicators are required for strong crosswinds, sharp curves, and other demanding conditions.

3. Weight Calculation Based on FAHP

The fuzzy analytic hierarchy process (FAHP) extends the analytic hierarchy process [16] by representing an expert's pairwise judgment with a triangular fuzzy number [17]. Each number contains a lower, most likely, and upper value. The judgments are aggregated, converted to crisp values, and normalized to obtain local weights.

Table 6 maps five plain-language judgments, from equal importance to extreme importance, to triangular fuzzy numbers. Experts compared only indicators that share the same parent node.

Ten experts completed the pairwise-comparison questionnaire independently. The panel combined five industry specialists and five academic specialists so that testing practice and methodological knowledge were both represented. The industry group had at least six years of experience in closed-field automated-driving tests, AI-SAP development, or ADAS safety assessment. The academic group held doctoral degrees and had at least eight years of research experience in driving scenarios or perception-decision algorithms. All experts received the same indicator definitions, scenario examples, and linguistic scale. They scored anonymously and independently. For each pairwise comparison, the ten triangular fuzzy judgments were averaged and then processed as described above. Tables 7–13 report the aggregated matrices and resulting weights.

Table 6. Fuzzy linguistic evaluation scales.

Linguistic Scale	Triangular Fuzzy Number $\tilde{a}_{ij}$	Semantic Interpretation
Equal Importance	$(1,1,1)$	Indicator i shares identical importance with indicator j
Slightly More Important	$(1,2,3)$	Indicator i is slightly more important than indicator j
Distinctly More Important	$(2,3,4)$	Indicator i is distinctly more important than indicator j
Strongly More Important	$(3,4,5)$	Indicator i is strongly more important than indicator j
Extremely More Important	$(4,5,6)$	Indicator i is extremely more important than indicator j
Reciprocal Judgment	$\left(\frac{1}{u_{ij}}, \frac{1}{m_{ij}}, \frac{1}{l_{ij}}\right)$	Relative importance of indicator j compared with indicator i
Equal Importance	$(1,1,1)$	Indicator i shares identical importance with indicator j

Table 7. Matrix 1.

Primary Layer Matrix 1	Scenario Hazardousness	System Processing Difficulty
Scenario Hazardousness	(1,1,1)	(1,1.5,2)
System Processing Difficulty	(1/2,1/1.5,1)	(1,1,1)

Table 8. Matrix 2.

Primary Layer Matrix 2	Time-to-Collision	Braking Deceleration
Time-to-Collision	(1,1,1)	(1,1.5,2)
Braking Deceleration	(1/2,1/1.5,1)	(1,1,1)

Table 9. Matrix 3.

Primary Layer Matrix 3	Perception Difficulty	Decision-Making Difficulty	Execution Difficulty
Perception Difficulty	(1,1,1)	(1/2,1/1.5,1)	(1,1.5,2)
Decision-Making Difficulty	(1,1.5,2)	(1,1,1)	(1.5,2,2.5)
Execution Difficulty	(1/2,1/1.5,1)	(1/2.5,1/2,1/1.5)	(1,1,1)

Table 10. Matrix 4.

Primary Layer Matrix 4	Static Obstacle Perception	Dynamic Target Perception
Static Obstacle Perception	(1,1,1)	(1/2,1/1.5,1)
Dynamic Target Perception	(1,1.5,2)	(1,1,1)

Table 11. Matrix 5.

Primary Layer Matrix 5	Regulation Complexity	Drivable Space	Sudden Incidents
Regulation Complexity	(1,1,1)	(1/2,1/1.5,1)	(1/2.5,1/2,1/1.5)
Drivable Space	(1,1.5,2)	(1,1,1)	(1/2,1/1.5,1)
Sudden Incidents	(1.5,2,2.5)	(1,1.5,2)	(1,1,1)

Table 12. Matrix 6.

Primary Layer Matrix 6	Road Gradient	Pavement Adhesion	Pavement Roughness
Road Gradient	(1,1,1)	(1/5,1/4,1/3)	(1,1,1)
Pavement Adhesion	(3,4,5)	(1,1,1)	(3,4,5)
Pavement Roughness	(1,1,1)	(1/5,1/4,1/3)	(1,1,1)

Table 13. Weight coefficients of scenario complexity evaluation indicators.

Target Layer	Criterion Layer	Criterion Layer Weight	Primary Indicator Layer	Relative Weight	Secondary Indicator Layer	Relative Weight
C_s	H_s	0.59	H_{inter}	—	—	0.34
			H_{self}	—	—	0.24
	D_s	0.41	D_{Per}	Static Obstacle Perception	0.41	0.05
				Dynamic Target Perception	0.59	0.07
			D_{Plan}	Regulation Complexity	0.23	0.04
				Drivable Space	0.32	0.06
				—	—	0.08
			D_{Con}	Road Gradient	0.16	0.02
				Pavement Adhesion	0.66	0.06
				Pavement Roughness	0.16	0.01

FAHP provides local preference weights for the specified hierarchy. Fuzzy judgments represent uncertainty in the elicited comparisons but do not remove the structural restrictions of hierarchical composition [17,18]. The subsequent additive model assumes that changes in one criterion can be evaluated without modeling changes in its relative contribution caused by other criteria. This assumption has not been independently verified. Shared variables and mechanisms can connect the indicators: relative velocity enters both collision-risk and tracking-load measures, while cut-in behavior can affect both time to collision and drivable space. Separate definitions therefore do not rule out double counting or interaction effects. Tables 7–13 report weights conditional on the specified hierarchy; they are not evidence of criterion independence.

Consistency ratios were interpreted only for matrices of order three or greater. For a reciprocal 2 × 2 matrix, the consistency ratio is zero by construction and is, therefore, not diagnostic. The three 3 × 3 matrices in Tables 9, 11 and 12 had consistency ratios of 0.0013, 0.0013, and 0.0000, respectively, all below the commonly used threshold of 0.10.

Weight robustness was assessed in 10,000 Monte Carlo simulations. Each published local weight was independently multiplied by a factor drawn from U(0.9, 1.1), and the weights were renormalized within each parent node. The 95% intervals for the three largest global weights were 0.317–0.381 for time-to-collision risk, 0.215–0.270 for ego-motion risk, and 0.073–0.097 for sudden incidents. The same three indicators remained the largest-weight set in 82.3% of simulations, and the largest absolute shift in any global weight was 0.052. These results indicate moderate, not complete, robustness. Because the archived field worksheet contains total scores but not all ten scenario-level subindicator values, field-scenario ranks could not be recomputed under every perturbation.

The consistency ratios assess the internal coherence of the specified pairwise judgments. The perturbation analysis assesses changes in weights while the hierarchy and additive equations remain fixed. Neither check tests the absence of dependence or estab-

lishes robustness to a different model structure. In particular, stability of the largest weights does not demonstrate stability of scenario rankings under interaction effects.

4. Experimental Design and Validation

The field experiment examined whether the calculated complexity scores were associated with collision-avoidance scores. It did not assume that a scenario itself 'performs'. The 22 scenarios were executed with each of three vehicles, and every vehicle-scenario combination was repeated three times. Because the sample was small and observational, the analysis was designed as exploratory validation rather than proof of predictive validity.

Pearson correlations were calculated for each vehicle and for the mean score across vehicles. All 22 cases were retained. Vehicle-specific results were emphasized because differences in sensing and control capability can modify the relationship between scenario characteristics and test outcomes.

4.1. Baseline Complexity Calculation Methods

To quantify the advantages of the proposed dual-dimensional integrated complexity index (M5), four representative single-dimensional methods from the literature were selected as baselines (M1–M4). All methods were recalculated using identical scenario inputs, normalization rules, and reference sensing ranges to ensure a fair comparison. Their principal mechanisms and limitations are summarized below.

M1 [8] uses only longitudinal and lateral time-to-collision values and therefore treats collision risk as the sole measure of scenario complexity. It does not account for static environmental interference, perception load, or vehicle execution constraints. Consequently, non-collision scenarios with high processing demands, such as a heavily occluded pedestrian in fog, are poorly represented, and the resulting scores are concentrated in the medium-complexity range.

M2 [19], proposed by Dong et al., calculates complexity from the information entropy of static elements, including the road, weather, and illumination. It excludes dynamic interactions and vehicle kinematics. The method therefore has limited sensitivity to cut-ins, emergency braking, and other interactive hazards, causing most samples to fall within low-complexity intervals.

M3 [20], developed by Wang et al., uses information entropy for static complexity and an improved gravitational model based on relative speed and distance for dynamic complexity. However, it does not distinguish inherent scenario hazards from the perception, decision-making, and execution demands imposed on the automated driving system. It also omits environmental occlusion, illumination interference, and pavement disturbance.

M4 [9], proposed by Sun et al., incorporates road, environmental, and traffic-participant indicators and determines their weights using a conventional analytic hierarchy process based on expert judgment. However, it neither separates inherent hazard from system processing difficulty nor represents the temporal evolution of complexity from hazard onset to collision avoidance. Perception noise and road-surface disturbances are also not quantified.

All four baseline methods and the proposed M5 integrated index adopt the same 0–1 unified dimensionless output range. The identical LHS-generated cut-in scenario dataset and fixed calculation platform are applied for all comparative measurements to eliminate interference from inconsistent preprocessing standards.

4.2. Test Scenario Configuration

Two independent datasets were designed to evaluate the proposed dual-dimensional complexity index. First, 2000 simulated cut-in scenarios generated by Latin hypercube

sampling were used to compare the score distributions and discrimination capability of M1–M5. Second, 22 hazardous cases from the AI-SAP specification were implemented in closed-field tests with three production Level 2 vehicles to examine the relationship between scenario complexity and collision-avoidance performance.

4.2.1. Simulated Scenarios Generated via LHS for Comparative Analysis

Latin hypercube sampling (LHS) [21] was used to generate the scenarios for comparative analysis. Unlike full factorial sampling, whose sample size increases exponentially with dimensionality, LHS provides broad coverage of a multidimensional parameter space with a manageable number of samples. The resulting scenario library was used to compare the five complexity metrics.

The LHS dataset was used only to compare the distributional coverage of the five metrics under identical inputs. Broader coverage is useful for screening, but it does not by itself demonstrate predictive accuracy.

The parameters listed in Table 14 were used to construct the cut-in scenarios. In total, 2000 scenarios were generated by LHS; the distributions of the principal sampled parameters are shown in Figure 1.

Table 14. Key parameters of the cut-in scenarios.

Parameter Category	Parameter Name	Symbol	Value Range
Ego vehicle state	Ego vehicle speed	v_{sv}	[30–130]
	Target vehicle speed	v_{tv}	[10–120]
Traffic participants	Lateral distance at the onset of cut-in	$Rs-lat$	[0.875–1.75]
	Time-to-collision at cut-in instant	φ_{x-TTC}	[1.0–5.0]
	Cut-in duration	T_{dur}	[1.5–8]
Road	Number of lanes	L_n	3
	Friction coefficient	β_f	[0.4–1]
Environment	Weather condition	$E_{weather}$	See Table 2
	Illumination condition	E_{light}	See Table 2

4.2.2. AI-SAP Standard Test Scenarios for Real-Vehicle Verification

The Intelligent Vehicle Safety Assessment Procedure (AI-SAP) is a unified test specification jointly developed by Tsinghua University and China Merchants Vehicle Testing & Technology Research Institute Co., Ltd. It organizes automated-driving safety assessment into four structural layers and seven evaluation dimensions and includes more than 50 quantitative indicators [22]. The specification covers weather, environmental interference, and diverse traffic-participant behaviors, enabling systematic assessment of representative hazardous driving conditions.

Twenty-two representative hazardous cases were selected from the AI-SAP scenario library for closed-field vehicle testing. They included construction-zone obstacles, stationary vehicles under low visibility, lead-vehicle emergency braking, nighttime pedestrian crossings, rainy lane cut-ins, and occlusion-induced motorcycle intrusions. Configuration parameters for the AI-SAP scenarios are available from the official release website [23].

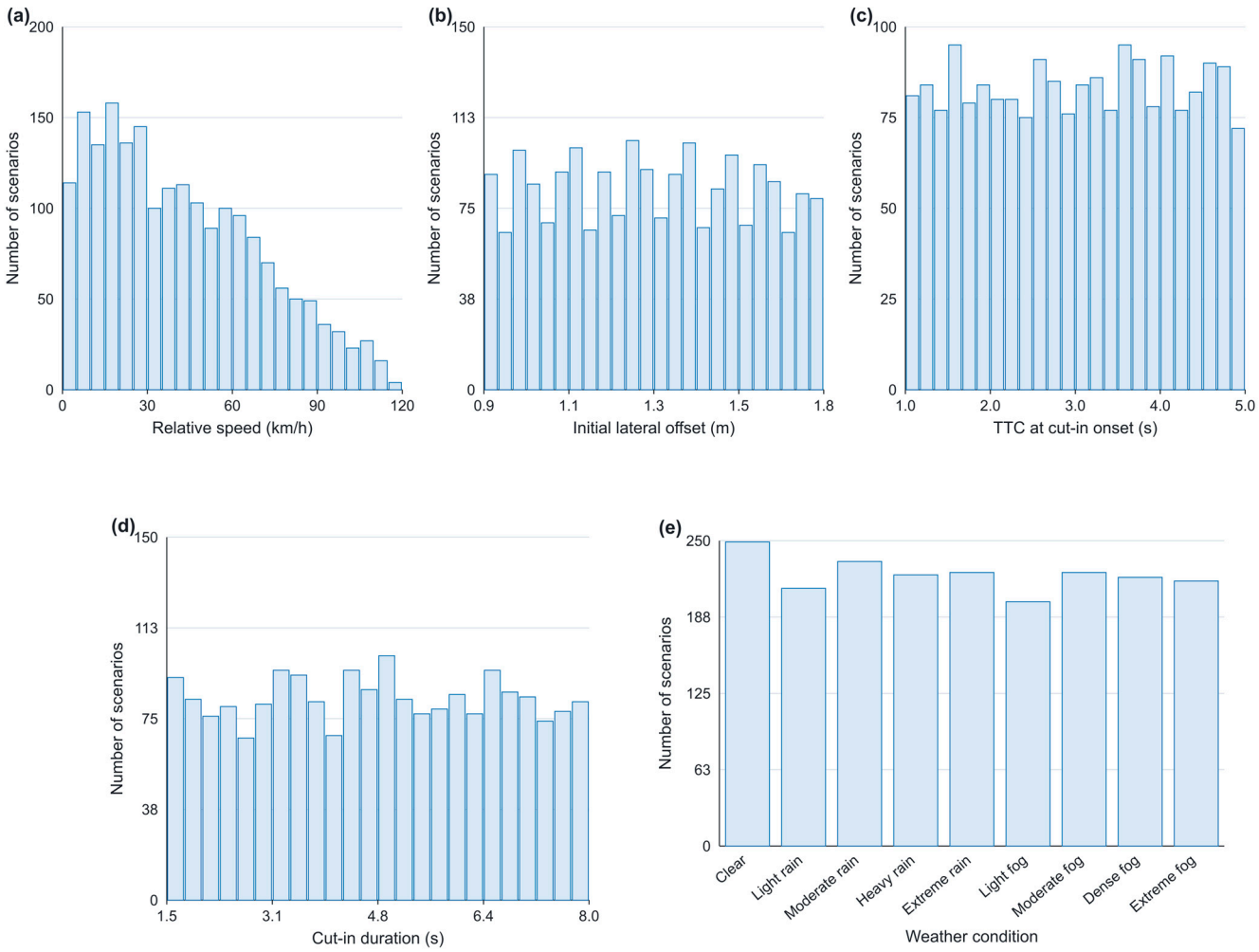


Figure 1. Distributions of the principal parameters in the 2000 scenarios generated by Latin hypercube sampling: (**a**) relative speed between the ego and target vehicles; (**b**) lateral offset at cut-in onset; (**c**) time to collision at cut-in onset; (**d**) cut-in duration; and (**e**) weather condition.

4.3. Ground-Test Application

Closed-field tests were conducted at the Chongqing Motor Vehicle Compulsory Inspection Proving Ground, Jinfeng Base, China Merchants Vehicle Testing & Technology Research Institute Co., Ltd., Chongqing, China. The site provided controlled road sections, artificial-rainfall and fog facilities, soft vehicle and two-wheeler targets, and synchronized onboard data acquisition. Three production Level 2 passenger vehicles were tested (see Figure 2). All used a forward monocular camera and front millimeter-wave radar; Vehicles 1 and 2 also had corner radars (see Table 15). Each vehicle-scenario combination was repeated three times.

Table 15. Sensor configurations of the three test vehicles.

Sensor Name	Vehicle 1	Vehicle 2	Vehicle 3
Front radar	✓	✓	✓
Front lidar	×	×	×
Front camera	✓	✓	✓
Corner radar	✓	✓	×

✓: means available; ×: means not available.



Figure 2. Test vehicles, equipment, and controlled environments used in the closed-field experiments: (**a**) three production Level 2 passenger vehicles with different sensor configurations; (**b**) a two-wheeled motorized dummy, a soft crash-target vehicle, and an onboard data-acquisition unit; and (**c**) the artificial-rainfall and fog chamber used to reproduce low-visibility and stationary-target scenarios.

The scenarios were adopted from AI-SAP, and the scoring rules followed its evaluation specification. A score of 1 was assigned when the vehicle avoided a collision through deceleration, stopping, or a lane change. A score of 0 was assigned when the collision speed exceeded 40 km/h; for lower collision speeds, a fractional score was calculated using Equation (33).

$$\alpha = \frac{v_0 - v}{v_0} \tag{33}$$

v_0 denotes target driving speed defined by test scenarios; v represents collision velocity between ego vehicles and surrounding obstacles.

5. Results

5.1. Comparison Results of Five Complexity Calculation Methods

Figure 3 compares the distributions obtained using the five complexity metrics. The results reveal substantial classification bias in the single-dimensional metrics M1–M4 and a more balanced distribution for the integrated metric M5.

Figure 3a shows that M1 is concentrated mainly in the 0.3–0.5 range, whereas M2 is concentrated at lower values. M3 and M4 occupy wider but still uneven ranges. M5 places observations throughout the occupied 0–0.8 range without concentrating most cases in one interval. This pattern can support stratified screening.

Figure 3b reports the mean score within each occupied interval. The comparison confirms that the five implementations use the same normalized scale, while their occupied intervals differ. M5 includes observations in high-complexity intervals that are empty or sparsely populated for several baselines. This is evidence of distributional discrimination, not evidence that M5 predicts vehicle failure.

Overall, M5 provides broad score coverage and a less concentrated distribution than most baselines for this simulated cut-in dataset. The result supports its use as a scenario-stratification measure, subject to external calibration and outcome-based validation.

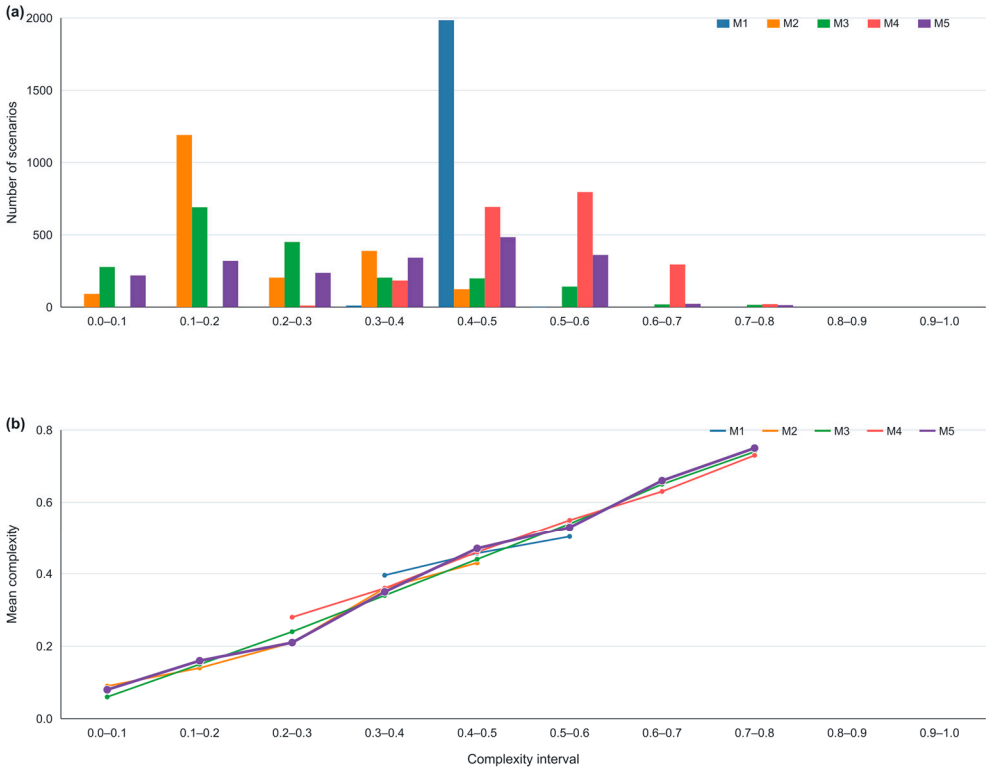


Figure 3. Comparison of five complexity metrics for 2000 simulated scenarios: (**a**) number of scenarios in each equal-width complexity interval and (**b**) mean complexity within each interval.

5.2. Real-Vehicle Test Application

Table 16 summarizes the scores of the three vehicles in the 22 scenarios. Pearson correlations were calculated among total complexity, ego-vehicle speed, and the vehicle scores. All cases were retained. Figure 4 reports two-sided tests with * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Table 16. Collision-avoidance scores of the three vehicles in the 22 test scenarios.

ID	Scenarios Description	C_s	Vehicle 1	Vehicle 2	Vehicle 3	v_{sv}
S1	Water-filled traffic barriers in construction zone	0.45	0%	0%	0%	120
S2	Traffic safety cones in construction zone	0.45	0%	0%	0%	120
S3	Tire obstacle	0.45	0%	0%	0%	120
S4	Small box obstacle	0.45	0%	0%	0%	120
S5	Triangular warning sign	0.45	0%	0%	0%	120
S6	Stationary vehicle ahead under foggy conditions	0.56	100%	100%	70%	80
S7	Stationary truck at tunnel entrance	0.38	100%	100%	0%	80
S8	Crash-damaged accident vehicle	0.62	100%	0%	0%	120
S9	Lead vehicle hard braking	0.46	100%	100%	100%	80
S10	Adjacent vehicle lane cut-in under rainy conditions	0.51	34.70%	100%	7%	60

Table 16. *Cont.*

ID	Scenarios Description	C_s	Vehicle 1	Vehicle 2	Vehicle 3	v_{sv}
S11	Adjacent vehicle lane cut-in under low-light night conditions	0.59	100%	100%	100%	60
S12	Three-wheeler lane cut-in	0.64	34.80%	11%	13.67%	60
S13	Stationary three-wheeler appearing after lead vehicle lane cut-out	0.72	100%	66.67%	38.62%	80
S14	Stationary vehicle in the adjacent lane suddenly accelerates and invades the ego lane	0.56	0%	0%	0%	80
S15	Conflict between ego vehicle and ramp-merging vehicle	0.47	100%	100%	100%	80
S16	Pedestrian traveling longitudinally along the road under foggy conditions	0.71	100%	100%	100%	60
S17	Umbrella-carrying pedestrian traveling longitudinally along road under rainy conditions	0.65	100%	100%	0%	60
S18	Dark-clothed pedestrian road crossing at night	0.38	34%	56.78%	0%	60
S19	Ghost pedestrian suddenly darting out with dynamic visual occlusion	0.82	100%	100%	100%	60
S20	Motorcycle lane cut-in with dynamic visual occlusion	0.71	4.70%	100%	26.98%	60
S21	Scooter rider in raincoat crossing roadway	0.58	100%	100%	0%	60
S22	Pedestrian pushing bicycle diagonally across road	0.45	100%	100%	100%	60

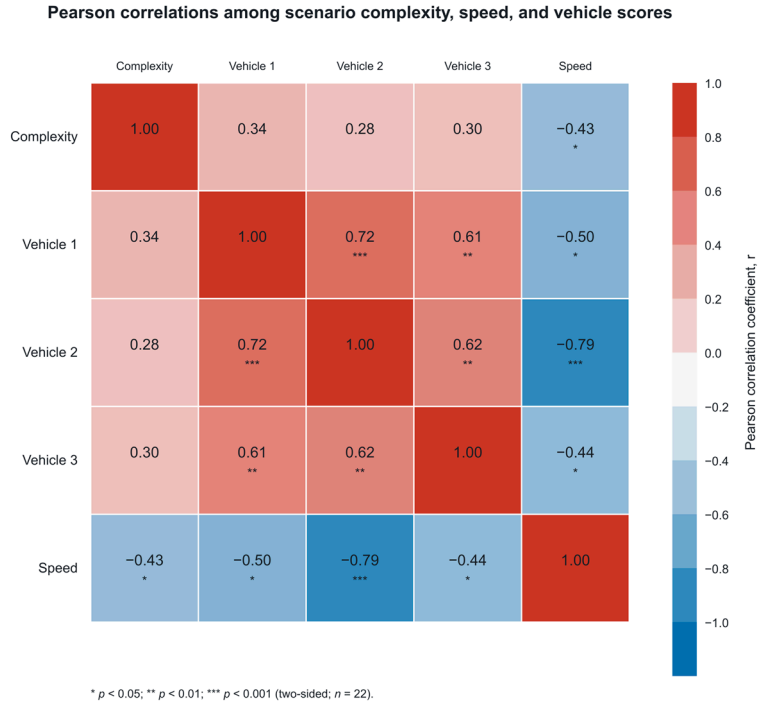


Figure 4. Pearson correlation matrix for total scenario complexity, ego-vehicle speed, and the score rates of the three test vehicles (n = 22). Asterisks indicate two-sided significance: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Vehicle scores were positively correlated with one another (Vehicle 1 versus Vehicle 2: r = 0.725, $p < 0.001$; Vehicle 2 versus Vehicle 3: r = 0.619, $p < 0.01$; Vehicle 1 versus Vehicle 3:

r = 0.614, $p < 0.01$). Total complexity was not significantly correlated with Vehicle 1 (r = 0.335, $p = 0.127$), Vehicle 2 (r = 0.282, $p = 0.204$), or Vehicle 3 (r = 0.298, $p = 0.178$). No cases were excluded.

For the mean score across vehicles, leave-one-scenario-out correlations ranged from 0.257 to 0.418. A 10,000-sample nonparametric bootstrap produced the wide 95% confidence interval reported above. Both checks show that the field estimate is uncertain and sensitive to the limited scenario sample, although it is not driven by a single case.

6. Discussion

6.1. Distributional Coverage of the Integrated Metric M5

The baseline metrics produce uneven score distributions because they represent restricted factor sets. M1 emphasizes time-to-collision, M2 emphasizes static elements, and M3–M4 cover different combinations of static and dynamic factors. Their different concentrations are therefore expected rather than necessarily erroneous.

M5 combines traffic hazardousness with perception, decision-making, and execution demands and consequently spreads this cut-in dataset across more score intervals. That property may reduce redundant sampling during scenario-library screening. It should not be interpreted as proof of superior prediction, because no accepted ground-truth complexity label was available.

The simulation set contains few cases above 0.8 and does not cover ramp and off-road edge cases. The framework also uses a peak statistic and therefore does not represent the duration or cumulative burden of a complex episode. These constraints limit generalization beyond the evaluated scenario families.

6.2. Field Associations and Validation Limits

The positive correlations among the three vehicles indicate that some scenarios were consistently easier or harder across platforms. They do not, however, validate the total-complexity score.

Ego-vehicle speed was negatively associated with collision-avoidance performance, most strongly for Vehicle 2 (r = −0.786, $p < 0.001$). Speed was also moderately negatively correlated with total complexity in this sample (r = −0.430, $p = 0.046$), illustrating that the field scenarios were not balanced across speed and complexity.

Total complexity was not significantly correlated with any individual-vehicle score. Its correlation with the mean vehicle score was also nonsignificant (r = 0.348, $p = 0.113$; bootstrap 95% confidence interval, −0.021 to 0.658). The previous post hoc analysis that removed discordant cases has been deleted. The field results therefore do not establish predictive validity.

With $n = 22$ and a two-sided alpha of 0.05, a Fisher-z approximation indicates 80% power only for correlations of approximately $|r| = 0.57$ or larger. The study is therefore underpowered for modest associations. Further limitations include empirical perception constants, few ultra-high-complexity cases, and omission of road curvature, crosswinds, tire condition, actuator delay, and possible dependence among criteria. Future studies should preregister exclusion rules, calibrate parameters against sensor data, expand the scenario set, and compare FAHP with covariance-aware or data-driven weighting methods.

Dependence among criteria is a limitation of model specification, in addition to the limitations of sample size and parameter calibration. Its effect on the scenario scores and rankings has not been quantified. A dependence-aware comparison requires explicit influence assumptions, suitable expert judgments or other justified estimates, and the complete scenario-level indicator matrix. The analytic network process (ANP) is one candidate for representing such influences [18]. The current results do not establish that the

additive implementation is adequate in scenarios where interaction or feedback materially changes criterion importance.

7. Conclusions

This study proposed a two-dimensional framework that separates traffic hazardousness from the perception, decision-making, and execution demands imposed on a reference automated-driving system. Established kinematic and time-to-collision indicators were organized within this framework, and FAHP was used to obtain hierarchical preference weights.

In 2000 simulated cut-in scenarios, M5 produced a broad, comparatively less concentrated score distribution, supporting its potential use for stratified scenario screening. A 10% weight-perturbation analysis showed moderate robustness of the dominant global weights. Distributional coverage, however, is not a measure of predictive accuracy.

In the 22 closed-field cases, the scores of the three vehicles were positively correlated with one another, but total complexity was not significantly correlated with any vehicle score. No post hoc case deletion was used. The field experiment therefore provides exploratory evidence about scenario differentiation but does not establish predictive validity for vehicle performance.

The method is limited by empirically assigned perception constants, a small and statistically underpowered field sample, few scenarios above 0.8, and omission of road curvature, crosswinds, tire condition, and actuator delay. Future work should calibrate the indicators against sensor-performance data, add ultra-high-complexity and diverse road cases, model cumulative temporal burden, and evaluate covariance-aware or data-driven weighting before the framework is used for predictive assessment.

The present implementation should be interpreted as an exploratory additive approximation with an unverified dependence structure. Before it is used for general scenario ranking or predictive assessment, the indicator definitions and aggregation assumptions require evaluation against a model that represents material dependencies. Stable weights within the present hierarchy are insufficient to establish that adequacy.

Author Contributions: Conceptualization, J.Z.; Methodology, J.Z.; Software, J.Z.; Validation, J.Z. and H.W.; Formal analysis, H.W.; Investigation, L.Z.; Resources, L.Z.; Data curation, H.W.; Supervision, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Innovation Key R&D Program of Chongqing, no. CSTB2025TIAD-STX0037; Research on Evaluation Systems and Key Equipment for Scenario Safety of Intelligent Connected Vehicles (Contract No. TYSP202509234536), funded by China Merchants Investment Development Co., Ltd.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Author Haozhong Wang was employed by the company China Merchants Vehicle Testing & Technology Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI-SAP	Automotive Intelligent Safety Assessment Procedure
TTC	Time to Collision
FAHP	Fuzzy Analytic Hierarchy Process

References

1. Ma, Y.; Jiang, W.; Wu, J. Self-evolution scenarios for simulation tests of autonomous vehicles based on different models of driving styles. *China J. Highw. Transp.* **2023**, *36*, 216–228.
2. Xia, Q.; Duan, J.; Gao, F.; Chen, T.; Yang, C. Automatic generation method of test scenario for ADAS based on complexity. In Proceedings of the SAE 2017 Intelligent and Connected Vehicles Symposium, Kunshan, China, 26–27 September 2017.
3. Sussman, J.M. The new transportation faculty: The evolution to engineering systems. *Transp. Q.* **1999**, *53*, 15–26.
4. Ma, Y.; Pan, X.; Xiong, L.; Xing, X.; Bulut, S.; Chen, J. Definition and Quantification of the Complexity Experienced by Autonomous Vehicles in the Environment and Driving Task. In *Proceedings of the CICTP 2020*; American Society of Civil Engineers: Xi'an, China, 2020; pp. 1030–1042.
5. Zhang, L.; Ma, Y.; Xing, X.; Xiong, L.; Chen, J. Research on the complexity quantification method of driving scenarios based on information entropy. In *Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*; IEEE: New York, NY, USA, 2021; pp. 3476–3481.
6. Li, J.K.; Deng, W.W.; Ren, B.T. An evaluation method for test scenario complexity of intelligent vehicles. In Proceedings of the 2020 Annual Conference of China Society of Automotive Engineers, Shanghai, China, 17–29 October 2020; Volume 1.
7. Wang, R.; Sun, Y.F.; Song, J. Evaluation method and experimental verification of road test scenarios for automated vehicles. *Automot. Eng.* **2021**, *43*, 521–528.
8. Zhu, B.; Fan, T.X.; Zhao, W.B.; Li, W.N.; Zhang, P.X. Continuous test scenario complexity evaluation method for automated driving vehicles. *J. Jilin Univ. (Eng. Technol. Ed.)* **2025**, *55*, 456–467. (In Chinese) [CrossRef]
9. Sun, H.; Li, Y.R.; Zhang, L.L.; Zhai, Y.; Chen, Z.Y.; Chen, C. Scenario complexity calculation model of real road test based on operational design condition. *Automot. Eng.* **2024**, *46*, 1983–1992. (In Chinese) [CrossRef]
10. Menzel, T.; Bagschik, G.; Maurer, M. Scenarios for development, test and validation of automated vehicles. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1821–1827. [CrossRef]
11. Riedmaier, S.; Ponn, T.; Ludwig, D.; Schick, B.; Diermeyer, F. Survey on scenario-based safety assessment of automated vehicles. *IEEE Access* **2020**, *8*, 87456–87477. [CrossRef]
12. Neurohr, C.; Westhofen, L.; Butz, M.; Bollmann, M.H.; Eberle, U.; Galbas, R. Criticality analysis for the verification and validation of automated vehicles. *IEEE Access* **2021**, *9*, 18016–18041. [CrossRef]
13. European Commission. Harmonised European Solutions for Testing Automated Road Transport (HEADSTART). Project 824309, 2019–2022. Available online: https://cordis.europa.eu/project/id/824309 (accessed on 7 September 2026).
14. Nalic, D.; Li, H.; Eichberger, A.; Wellershaus, C.; Pandurevic, A.; Rogic, B. Stress testing method for scenario-based testing of automated driving systems. *IEEE Access* **2020**, *8*, 224974–224984. [CrossRef]
15. Ponn, T.; Breitfuss, M.; Yu, X.; Diermeyer, F. Identification of challenging highway-scenarios for the safety validation of automated vehicles based on real driving data. In Proceedings of the 15th International Conference on Ecological Vehicles and Renewable Energies (EVER), Monte-Carlo, Monaco, 10–12 September 2020. [CrossRef]
16. Saaty, T.L. Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* **2008**, *1*, 83–98. [CrossRef]
17. Buckley, J.J. Fuzzy hierarchical analysis. *Fuzzy Sets Syst.* **1985**, *17*, 233–247. [CrossRef]
18. Saaty, T.L.; Kulakowski, K. Axioms of the Analytic Hierarchy Process (AHP) and its Generalization to Dependence and Feedback: The Analytic Network Process (ANP). *arXiv* **2016**, arXiv:1605.05777v2.
19. Dong, X.; Zhang, D.; Mu, Y.; Zhang, T.; Tang, K. Complexity of driving scenarios based on traffic accident data. *Int. J. Automot. Technol.* **2024**, *25*, 23–36. [CrossRef]
20. Dong, H.; Shu, W.; Chen, C.; Sun, C.; You, C. Research on complexity evaluation method for dangerous driving conditions. *Automot. Eng.* **2020**, *42*, 808–814. (In Chinese) [CrossRef]
21. He, X.D.; Gou, W.X.; Liu, Y.S.; Gao, Z.Z. Structure and system nonprobabilistic reliability solution method based on enhanced optimal latin hypercube sampling. *Int. J. Struct. Stab. Dyn.* **2015**, *15*, 1450034. [CrossRef]
22. Two Evaluation Systems for Intelligent Vehicle Safety and Intelligent Cockpit Released in Chongqing, Tsinghua University. Available online: https://www.tsinghua.edu.cn/info/1182/114418.htm (accessed on 27 June 2026).
23. China Merchants Vehicle Test Technology Research Institute Co., Ltd. Release of Automotive Intelligent Safety Evaluation Specification (AI-SAP 2025 Version). Available online: https://mp.weixin.qq.com/s?__biz=MzIyNjU4MDUyMQ==&mid=2247517624&idx=1&sn=46d3a92e89fb0c72a9bd198a3cc5a61c&chksm=e91e9f6d753b7072b88c512e8c5fe22b8f238f422b8d7e20c78737f7c09603577fd4ced1a828#rd (accessed on 24 June 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.