technologies

MDPI

Article

Sensor-Health- and Belief-Aware Risk-Adaptive High-Order Control Barrier Function Safety Filtering for Dynamic Obstacle Avoidance

Yongsheng Ma, Guobao Zhang and Yongming Huang *

School of Automation, Southeast University, Nanjing 210018, China; mayon322470@seu.edu.cn (Y.M.)
* Correspondence: eamon2527@gmail.com

Abstract

Control-barrier-function-based safety filters are promising for autonomous driving, but most existing formulations treat obstacle perception as deterministic or account only for bounded ego state-estimation errors. This becomes limiting when obstacle existence, position, motion, and sensing quality vary online. We present a sensor-health- and belief-aware risk-adaptive high-order control barrier function (HOCBF) safety filter for dynamic obstacle avoidance. The method uses obstacle belief from a perception/tracking module, inflates residual obstacle uncertainty according to an object-wise sensor-health score, and converts upper-tail risk into adaptive HOCBF tightening through conditional value-at-risk (CVaR). Sensor health enters the controller through both covariance inflation and online CVaR confidence scheduling. The resulting quadratic program combines deterministic ego-error robustness with probabilistic perception uncertainty while minimally modifying the nominal control input. The zero-slack solution guarantees forward invariance of the risk-tightened safe set under the stated assumptions, whereas the slack-activated mode provides a quantified least-violation fallback rather than a strict safety guarantee. Simulations on a nonlinear 3-DOF bicycle model evaluate critical cut-in, sudden perception degradation, merge-bottleneck, fixed-CVaR, sensitivity, runtime-scaling, heterogeneous multi-obstacle, and heavy-tailed uncertainty cases.

Keywords: autonomous driving; high-order control barrier function; safety filter; sensor health; state estimation; risk-aware control

1. Introduction

Autonomous driving requires safety-critical control laws that remain reliable not only under nominal traffic evolution but also under moving obstacles, prediction ambiguity, and sensing degradation. In realistic traffic scenes, obstacle existence, obstacle motion, and interaction geometry are rarely known exactly, while the downstream controller is still required to react at the time scale of the vehicle control loop. A safety filter built on purely deterministic perception outputs is therefore difficult to trust in deployment. Conversely, overly conservative worst-case margins can degrade maneuverability and may become unnecessarily restrictive in routine traffic. This tension has made control-barrier-function-based safety filtering attractive, because it offers a principled way to modify a nominal controller only when needed while preserving safety constraints.

For vehicle safety problems, one of the first technical difficulties is the relative degree of the constraint itself. Obstacle-avoidance and lane-boundary requirements are naturally

Check for updates

Academic Editor: Tamás Haidegger

Received: 23 April 2026
Revised: 11 May 2026
Accepted: 16 May 2026
Published: 20 May 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.

described in position coordinates, whereas the control input acts through longitudinal acceleration and steering dynamics. High-order control barrier functions (HOCBFs) provide a systematic way to handle such constraints and have become a standard tool in safety-critical control of nonlinear systems [1]. More recent developments have extended this framework to discrete-time high-order settings and augmented-dynamics constructions for constrained robotic systems [2,3]. These advances clarify how high-relative-degree safety constraints can be enforced once the relevant state and environment information is available. What they do not resolve, however, is how uncertainty in obstacle existence, obstacle motion, or sensing quality should enter the barrier condition itself.

Once uncertainty is taken seriously, the robustness of the safety filter becomes the next issue. Considerable effort has therefore been devoted to CBF-based safety filtering under imperfect state information and disturbances. Representative examples include measurement-robust CBFs, high-order measurement-robust CBFs, tunable input-to-state safe formulations, disturbance-observer-assisted safety filtering, parameterized robustification, and observer–controller co-design [4–10]. These studies substantially improve the applicability of CBFs when the controller has access only to erroneous state estimates or bounded disturbance descriptions. Even so, the uncertainty treatment in this literature remains predominantly deterministic: uncertainty is usually represented by an error envelope, a disturbance bound, or a robustness margin. This is well suited to state-estimation mismatch, but it does not explicitly represent probabilistic obstacle existence, predicted obstacle covariance, or changes in sensing quality that should alter the confidence placed in a perception output.

Related questions have also been studied from a stochastic or belief-space perspective. Stochastic CBFs extend safety guarantees to stochastic systems [11,12], while belief control barrier functions use the probabilistic output of state estimators to synthesize risk-aware control laws in belief space [13]. In parallel, CVaR-based and related risk-aware barrier formulations have shown that tail-risk criteria can be incorporated directly into safety-critical control design [14,15]. Chance-constrained motion-planning and obstacle-avoidance methods further demonstrate the value of probabilistic reasoning when vehicles interact with moving obstacles [16,17]. Despite these advances, most of these formulations are either developed for general stochastic systems or remain closer to the planning layer. They do not directly answer the vehicle-level question considered here: how to combine relative-degree-two obstacle-avoidance constraints, deterministic state-estimation robustness, probabilistic obstacle belief, and online sensing degradation in a single HOCBF-based safety filter.

Recent sensor-based and incremental CBF formulations reduce dependence on exact model knowledge by using sensor-driven or reduced-order safety models [18]. This direction is complementary to the present work. The focus here is not only model uncertainty but also perception-belief uncertainty and online sensor-health degradation in dynamic-obstacle HOCBF constraints.

This gap becomes even more evident when one looks at autonomous-driving applications and perception-safety studies together. CBF-based techniques have been used in adaptive cruise control, autonomous lane change, roundabout crossing, and on-ramp merging, demonstrating the practical value of barrier-based safety enforcement in vehicle systems [19–22]. At the same time, studies on learned perception safety, obstacle-detection faults, and perception fault diagnosis have made it increasingly clear that sensing reliability cannot be treated as a secondary issue: once the perception layer becomes unreliable, the downstream safety mechanism must react in a structured and verifiable way [23–25]. Yet the connection between these two strands is still weak. In most vehicle-oriented CBF safety filters, obstacle information is still treated as deterministic or absorbed into fixed uncertainty margins. In many perception-safety studies, by contrast, the discussion stops at

fault detection, fault accommodation, or supervisory switching, without translating online sensing health into barrier-tightening terms inside a vehicle-level safety filter.

Taken together, these observations point to a specific missing piece. Existing HOCBF formulations explain how high-relative-degree safety constraints can be enforced. Robust CBF variants explain how bounded estimation mismatch can be absorbed. Belief- and risk-aware barrier formulations explain how probabilistic uncertainty may enter safety analysis. What remains insufficiently addressed is a unified control-oriented interface that maps online sensor health to obstacle-belief uncertainty, propagates that uncertainty through short-horizon obstacle prediction, and then converts the resulting risk into tightening terms that can be enforced directly in a dynamic-obstacle HOCBF quadratic program for autonomous driving.

To fill this gap, a sensor-health- and belief-aware risk-adaptive HOCBF safety filter is developed for dynamic obstacle avoidance. The key idea is to use an online sensor-health indicator to calibrate obstacle-belief uncertainty, propagate this uncertainty through short-horizon obstacle prediction, and transform the resulting tail risk into adaptive HOCBF tightening via a conditional CVaR criterion. In this way, deterministic estimation-error robustness and probabilistic perception uncertainty are incorporated within a unified vehicle-level safety-filtering framework. Compared with a fixed-margin robust formulation, the resulting design adjusts conservatism according to the current sensing condition. Compared with a purely belief-aware formulation, the uncertainty update is explicitly modulated by sensor health. Compared with a planner-only risk treatment, the final tightening term enters directly into the HOCBF constraint enforced by the safety filter.

The main contributions of this paper are summarized as follows:

- A perception-belief interface is introduced for dynamic-obstacle safety filtering. Each obstacle is represented by an existence probability, a posterior mean, a posterior covariance, and an object-wise sensor-health score.
- A health-aware CVaR-HOCBF formulation is developed. The sensor-health score inflates the residual obstacle uncertainty and schedules the CVaR confidence level online, thereby increasing conservativeness under degraded perception.
- A real-time QP-based safety filter is derived for control-affine vehicle dynamics with state-estimation error bounds. The formulation explicitly separates strict zero-slack safety guarantees from slack-activated least-violation behavior.
- The method is evaluated against deterministic robust HOCBF, chance-constrained HOCBF, belief-aware health-unaware HOCBF, and fixed-CVaR HOCBF baselines, with additional sensitivity, runtime, multi-obstacle, and heavy-tailed uncertainty experiments.

Figure 1 summarizes the overall workflow of the proposed method. Online sensor health, deterministic state-estimation uncertainty, and probabilistic obstacle belief are integrated through a unified risk-adaptive pipeline, where health-aware obstacle prediction and barrier tightening are embedded into the final HOCBF-QP safety filter.

The remainder of this paper is organized as follows. Section 2 formulates the vehicle model, sensing model, obstacle belief, and safety objective. Section 3 develops the health-aware probabilistic obstacle prediction, the CVaR-based risk tightening, and the HOCBF-QP safety filter. Section 4 presents simulation studies and discussion. Section 5 concludes the paper.

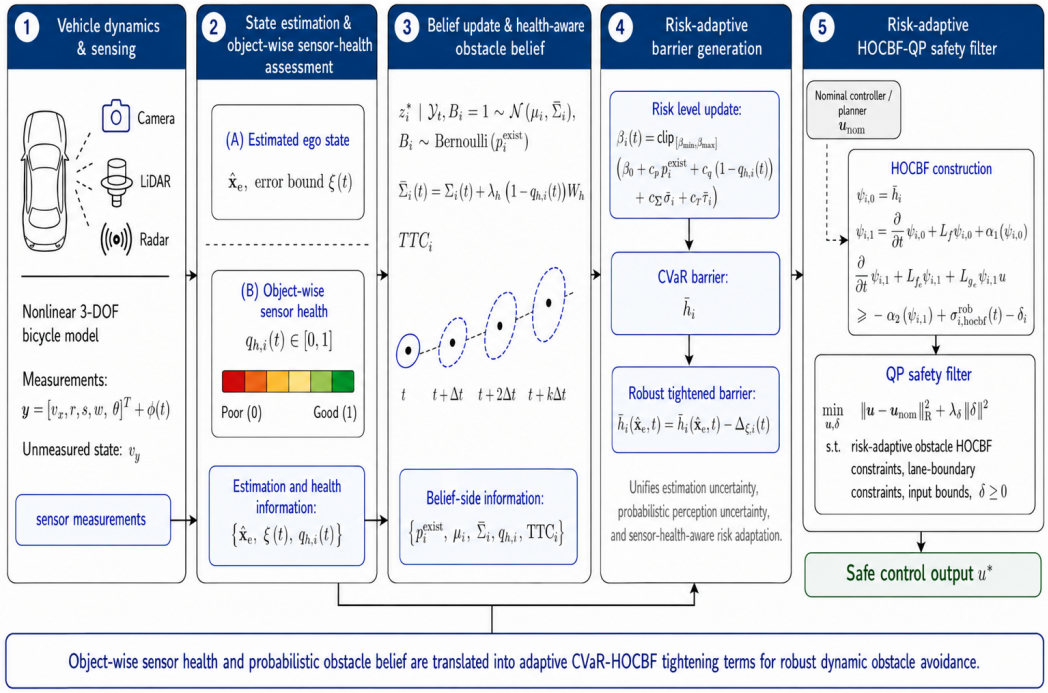


Figure 1. Overall framework of the proposed sensor-health- and belief-aware risk-adaptive HOCBF safety filter. Online sensor health and probabilistic obstacle belief are translated into adaptive HOCBF tightening terms, which are then enforced in the final HOCBF-QP safety filter for robust dynamic obstacle avoidance.

2. Problem Formulation

Throughout the paper, the true obstacle state is denoted by $z_i^\star$, whereas z_i denotes the random belief variable produced from perception/tracking information. The controller uses object-wise health $q_{h,i}(t) \in [0,1]$. When a single diagnostic health value is useful for plotting, it is only the conservative aggregation $q_h(t) = \min_{i \in \mathcal{I}_a(t)} q_{h,i}(t)$.

2.1. Ego-Vehicle Dynamics and Sensing Model

The ego vehicle is modeled as

$$\dot{x}_e = f_e(x_e) + g_e(x_e)u + d_e(t), \tag{1}$$

with state and input

$$x_e = \begin{bmatrix} v_x & v_y & r & s & w & \theta \end{bmatrix}^\top, \qquad u = \begin{bmatrix} a & \delta_f \end{bmatrix}^\top.$$

The measurement model is

$$y = Cx_e + \phi(t), \tag{2}$$

and, in the vehicle setting considered here,

$$y = \begin{bmatrix} v_x & r & s & w & \theta \end{bmatrix}^\top + \phi(t),$$

so that v_y is reconstructed by state estimation.

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The lateral velocity v_y is not assumed to be directly measured. It is reconstructed by the state estimator using vehicle dynamics, inertial measurements, steering input, and longitudinal velocity. In implementation, all ego states can be produced by a fusion estimator; the simplified measurement matrix only highlights that v_y is the least directly observable state. Table 1 summarizes the main notation used throughout the paper.

Table 1. Notation used in the paper.

Symbol	Meaning
$x_e = [v_x, v_y, r, s, w, \theta]^\top$	Ego vehicle state
$u = [a, \delta_f]^\top$	Longitudinal acceleration and front steering input
y	Measured ego signal
$\hat{x}_e$	Estimated ego state
$\tilde{x}_e = x_e - \hat{x}_e$	Ego state-estimation error
$\xi(t)$	Component-wise deterministic bound on $\tilde{x}_e$
$z_i^\star$	True state of obstacle i
z_i	Random obstacle belief variable
B_i	Binary existence indicator of obstacle i
p_i^{exist}	Existence probability of obstacle i
μ_i	Posterior mean of obstacle i
Σ_i	Posterior covariance of obstacle i before health inflation
$\bar{\Sigma}_i$	Health-inflated posterior covariance
$q_{h,i}$	Object-wise sensor-health score
β_i	CVaR confidence level for obstacle i
L_i	Safety loss associated with obstacle i
h_i^{mean}	Posterior-mean barrier surrogate
$\tilde{h}_i$	CVaR-tightened safety function
$\bar{h}_i$	Ego-error-robust CVaR-tightened safety function
$\psi_{i,0}, \psi_{i,1}$	HOCBF auxiliary functions
$\mathcal{J}(t)$	Slackened QP constraint-row set
δ_j	Slack variable for constraint row $j \in \mathcal{J}(t)$
η	Shared slack-envelope variable in the implementation QP
ρ_i	Slack penalty weight
$k_{F,i}, k_{\alpha,i}$	Non-negative component-wise robustness-sensitivity vectors
$\mathcal{I}_a(t)$	Active obstacle set

Table 2 lists practical sensing sources for the ego-state quantities in the measurement model.

Table 2. Practical ego-state sensing sources.

Ego Quantity	Practical Source
v_x	wheel speed, GNSS/INS odometry
r	yaw-rate gyro in IMU
s, w, θ	localization from GNSS/INS, HD map, and lane geometry
v_y	observer or fusion estimator using vehicle dynamics and inertial signals

Assumption 1. *The functions $f_e(\cdot)$ and $g_e(\cdot)$ are locally Lipschitz on the operating domain, and there exist constants $d_{\max} > 0$ and $\phi_{\max} > 0$ such that*

$$\|d_e(t)\| \le d_{\max}, \qquad \|\phi(t)\| \le \phi_{\max}, \qquad t \ge 0.$$

Let $\hat{x}_e$ denote the estimated state and $\tilde{x}_e := x_e - \hat{x}_e$ the estimation error. The ego estimator is assumed to provide a known component-wise error envelope $\xi(t) \succeq 0$ such that

$$|x_e(t) - \hat{x}_e(t)| \preceq \xi(t), \qquad \forall t \geq 0. \tag{3}$$

The envelope can be generated by residual monitoring, innovation consistency checking, set-membership bounds, or covariance-to-bound calibration. The theoretical safety guarantee holds for trajectories for which this envelope is valid. This does not require perfect state estimation; if the envelope is violated by an unmodeled estimator failure, the formal HOCBF safety guarantee is not claimed.

2.2. Dynamic Obstacles and Risk Objective

For each active obstacle $i \in \mathcal{I}_a(t)$, define the true road-aligned obstacle state and the corresponding random belief variable as

$$z_i^\star = \begin{bmatrix} s_i^\star & w_i^\star & v_{s,i}^\star & v_{w,i}^\star \end{bmatrix}^\top, \qquad z_i = \begin{bmatrix} s_i & w_i & v_{s,i} & v_{w,i} \end{bmatrix}^\top.$$

The perception/tracking module provides z_i through a posterior belief rather than the controller directly observing $z_i^\star$. Each obstacle is associated with existence probability

$$p_i^{\text{exist}} \in [0, 1]. \tag{4}$$

Let $B_i \in \{0, 1\}$ denote the corresponding existence indicator, and let $q_{h,i}(t) \in [0, 1]$ be the object-wise sensor-health indicator.

The obstacle-avoidance barrier is extended from a fixed ellipse to a speed-dependent safety region. Define

$$[r]_+^\epsilon = \frac{1}{2}\left(r + \sqrt{r^2 + \epsilon^2}\right), \qquad |r|_\epsilon = \sqrt{r^2 + \epsilon^2}.$$

Let $P_i^+(t)$ denote the current or predicted position covariance used by the safety filter, with diagonal entries $P_{i,ss}^+(t)$ and $P_{i,ww}^+(t)$. The longitudinal and lateral safety scales are

$$a_i(t) = \text{clip}_{[a_{\min}, a_{\max}]}\left(a_0 + c_{a,v}[v_x(t) - \mu_{v_s,i}(t)]_+^\epsilon + c_{a,\Sigma}\sqrt{P_{i,ss}^+(t)}\right), \tag{5}$$

$$b_i(t) = \text{clip}_{[b_{\min}, b_{\max}]}\left(b_0 + c_{b,v}|\dot{w}(t) - \mu_{v_w,i}(t)|_\epsilon + c_{b,\Sigma}\sqrt{P_{i,ww}^+(t)} + c_b\right). \tag{6}$$

The obstacle-avoidance barrier is

$$h_i(x_e, z_i^\star, t) := \frac{(s - s_i^\star)^2}{a_i^2(t)} + \frac{(w - w_i^\star)^2}{b_i^2(t)} - 1, \tag{7}$$

For evaluation against the true obstacle geometry, the covariance terms are set to zero, while the posterior-mean velocity terms are replaced only in offline metric computation by the corresponding true obstacle velocities. The constant-size ellipse special case is recovered by setting the speed- and uncertainty-expansion coefficients to zero. The HOCBF derivation retains the time derivatives of $a_i(t)$ and $b_i(t)$; for example, with $d_s = s - s_i^\star$ and $d_w = w - w_i^\star$,

$$\dot{h}_i = \frac{2d_s(\dot{s} - \dot{s}_i^\star)}{a_i^2} - \frac{2d_s^2\dot{a}_i}{a_i^3} + \frac{2d_w(\dot{w} - \dot{w}_i^\star)}{b_i^2} - \frac{2d_w^2\dot{b}_i}{b_i^3}.$$

For a random loss L, the conditional value-at-risk is formally defined as

$$\mathrm{CVaR}_{\beta}(L) = \inf_{\eta \in \mathbb{R}} \left[\eta + \frac{1}{1-\beta} \mathbb{E}[(L-\eta)_+] \right], \qquad \beta \in (0,1). \tag{8}$$

For obstacle i,

$$\mathrm{CVaR}_{\beta_i}(L_i \mid B_i = 1) = \inf_{\eta \in \mathbb{R}} \left[\eta + \frac{1}{1-\beta_i} \mathbb{E}[(L_i-\eta)_+ \mid B_i = 1] \right]. \tag{9}$$

Larger β_i corresponds to a more conservative upper-tail risk evaluation. Define the loss variable $L_i(x_e, z_i, t) := -h_i(x_e, z_i, t)$. Following risk-aware barrier formulations based on CVaR [14,15], the main safety requirement is the conditional CVaR constraint

$$\mathrm{CVaR}_{\beta_i(t)}\big(L_i(x_e, z_i, t) \mid B_i = 1\big) \le 0, \qquad i \in \mathcal{I}_a(t). \tag{10}$$

The associated CVaR-tightened safety function is $\tilde{h}_i(x_e, t) = -\mathrm{CVaR}_{\beta_i}(L_i(x_e, z_i, t) \mid B_i = 1)$. A chance-constrained counterpart may also be considered [16]. In the controller design below, safety is enforced with respect to a short-horizon predicted obstacle position rather than the instantaneous obstacle state. This choice reflects that lane-change safety depends on the anticipated interaction over the next short time interval as well as the current geometry. Accordingly, the belief over z_i is propagated to the predicted position p_i^+, and the operational CVaR constraint is evaluated on the resulting prediction-based barrier.

2.3. Problem Statement

Given (1)–(3), the belief information, and (10), we seek a sensor-health- and belief-aware risk-adaptive HOCBF safety filter that minimally modifies the nominal input $u_{\mathrm{nom}}(t)$ while enforcing prediction-based dynamic-obstacle safety under bounded disturbances, estimation error, and sensing degradation. Since the operational barrier depends on position states while the vehicle input enters through second-order kinematics, the resulting obstacle-avoidance constraint has relative degree two, and a HOCBF formulation is required [1].

3. Method

The proposed controller combines health-aware obstacle prediction, CVaR-based barrier tightening, a chance-constrained corollary, and a HOCBF-QP safety filter. The construction builds on belief-space safety filtering and risk-aware barrier formulations while specializing them to sensor-health-aware dynamic obstacle avoidance [13–15].

3.1. Health-Aware Probabilistic Obstacle Prediction

The perception/tracking interface for obstacle i is

$$\mathcal{B}_i(t) = \left\{ p_i^{\mathrm{exist}}(t), \mu_i(t), \Sigma_i(t), q_{h,i}(t) \right\}.$$

The existence indicator satisfies $B_i(t) \sim \mathrm{Bernoulli}(p_i^{\mathrm{exist}}(t))$, and the obstacle posterior is

$$z_i(t) \mid \mathcal{Y}_t,\ B_i(t) = 1 \sim \mathcal{N}(\mu_i(t), \bar{\Sigma}_i(t)),$$

where $\mathcal{Y}_t$ is the perception/tracking history. The belief $\mathcal{B}_i(t)$ is not assumed to be produced by the controller; it is the interface between the perception/tracking module and the safety filter.

Here c_i is the detector confidence, e_i and S_i are the innovation and innovation covariance, n_i^{miss} is the missed-detection count, and d_i is the cross-sensor disagreement vector.

The proposed controller does not require a specific detector or tracker architecture. It requires only standard outputs commonly available in perception and tracking systems: confidence, innovation, covariance, missed-detection count, and cross-sensor consistency. The weights in Algorithm 1 are calibrated on validation logs by minimizing the mismatch between the health score and normalized tracking-reliability labels, subject to $\omega_c + \omega_\nu + \omega_m + \omega_a + \omega_\Sigma = 1$ and non-negative weights. The same calibrated weights and smoothing coefficient are then fixed for all reported scenarios. The resulting object-wise $q_{h,i}(t)$ enters both the health-inflated covariance and the adaptive CVaR confidence schedule.

Algorithm 1: Online object-wise sensor-health estimation.

Input: $c_i, e_i, S_i, n_i^{\text{miss}}, d_i, \Sigma_i, q_{h,i}(t_{k-1})$.
Output: $q_{h,i}(t_k)$.

1. *Compute* $\nu_i = e_i^\top S_i^{-1} e_i$.
2. *Compute*

$$q_i^{\text{conf}} = \text{clip}_{[0,1]}(c_i), \quad q_i^{\text{nis}} = \exp\left(-\frac{\max(0, \nu_i - \nu_0)}{\gamma_\nu}\right),$$

$$q_i^{\text{miss}} = \exp\left(-\frac{n_i^{\text{miss}}}{n_0}\right), \quad q_i^{\text{agree}} = \exp\left(-\frac{d_i^\top R_d^{-1} d_i}{\gamma_d}\right),$$

$$q_i^{\text{cov}} = \exp\left(-\frac{\text{tr}(W_\Sigma^{-1}\Sigma_i)}{\gamma_\Sigma}\right).$$

3. *Compute*

$$\hat{q}_{h,i} = \text{clip}_{[0,1]}\left(\omega_c q_i^{\text{conf}} + \omega_\nu q_i^{\text{nis}} + \omega_m q_i^{\text{miss}} + \omega_a q_i^{\text{agree}} + \omega_\Sigma q_i^{\text{cov}}\right),$$

where $\omega_c + \omega_\nu + \omega_m + \omega_a + \omega_\Sigma = 1$ *and all weights are non-negative.*

4. *Apply temporal smoothing:*

$$q_{h,i}(t_k) = (1 - \alpha_q) q_{h,i}(t_{k-1}) + \alpha_q \hat{q}_{h,i}(t_k), \qquad \alpha_q \in (0, 1].$$

5. *Return* $q_{h,i}(t_k)$.

Table 3 reports the sensor-health estimator hyperparameters used in the simulation configuration.

Table 3. Sensor-health estimator hyperparameters used in the simulation configuration. The range $[q_{\min}, q_{\max}]$ is applied object-wise to each $q_{h,i}$ before it enters covariance inflation and risk scheduling.

Parameter	Value	Role
ν_0	3.0	NIS threshold before health degradation is applied
γ_ν	0.45	NIS decay scale in q_i^{nis}
n_0	4 frames	missed-detection decay scale
γ_d	0.65	cross-sensor disagreement decay scale
R_d	$\text{diag}(0.45^2, 0.22^2, 0.60^2, 0.35^2)$	disagreement normalization for position and velocity components
W_Σ	$\text{diag}(0.35^2, 0.18^2, 0.40^2, 0.22^2)$	covariance-score normalization in q_i^{cov}
W_h	$\text{diag}(0.35^2, 0.18^2, 0.40^2, 0.22^2)$	covariance-inflation shape in $\bar{\Sigma}_i = \Sigma_i + \lambda_h(1 - q_{h,i})W_h$
γ_Σ	5.5	covariance health decay scale
$\omega_c, \omega_\nu, \omega_m, \omega_a, \omega_\Sigma$	0.25, 0.25, 0.20, 0.15, 0.15	weights for confidence, innovation, missed-detection, agreement, and covariance scores
α_q	0.12	temporal smoothing coefficient
$q_{\min}, q_{\max}$	0.03, 0.99	clipped object-wise health range
Update frequency	20 Hz	one health update per 0.05 s control step
Dropout rule	$n_i^{\text{miss}} \geq 4$ frames	sustained missed detections drive the health score toward the lower range

For each active obstacle $i \in \mathcal{I}_a(t)$, let the Gaussian obstacle belief be represented as in belief-space safety filtering [13]. Before health inflation, the tracker covariance is Σ_i; the inflated belief used by the safety filter is

$$z_i \mid B_i = 1 \sim \mathcal{N}(\mu_i, \bar{\Sigma}_i), \qquad \bar{\Sigma}_i(t) = \Sigma_i(t) + \lambda_h\big(1 - q_{h,i}(t)\big)W_h,$$

where $\lambda_h > 0$ and $W_h = \mathrm{diag}(\sigma_s^2, \sigma_w^2, \sigma_{v_s}^2, \sigma_{v_w}^2)$ prevents dimensionally inconsistent inflation by scaling position and velocity components according to their own physical units. The coefficient λ_h is a dimensionless health-inflation gain because W_h carries the physical covariance units of the obstacle state. When $q_{h,i} = 1$, no health-induced covariance is added. When $q_{h,i} = 0$, the added covariance is $\lambda_h W_h$, representing the maximum residual perception uncertainty assigned to a completely degraded object track.

A component-wise validation-residual rule can be used to calibrate this gain. Let

$$r_{j,k} = z^{\star}_{j,k} - \mu_{j,k}, \qquad \bar{\sigma}^2_{j,k} = \sigma^2_{j,k} + \lambda_h(1 - q_{h,k})w_j,$$

where j indexes a state component, k indexes a validation sample, and w_j is the corresponding diagonal entry of W_h. To cover degraded-perception residuals at confidence level $1 - \epsilon$, choose

$$\lambda_h \geq \max_{j,k} \left[\frac{r^2_{j,k}/z^2_{1-\epsilon/2} - \sigma^2_{j,k}}{(1 - q_{h,k} + \epsilon_q)w_j} \right]_+, \qquad z_{1-\epsilon/2} = \Phi^{-1}(1 - \epsilon/2),$$

where $\epsilon_q > 0$ avoids numerical blow-up when $q_{h,k} \approx 1$. This selects the smallest health-inflation gain that covers the selected validation residual envelope; in practice, a safety factor can be applied to the resulting value.

Remark 1. *The health-dependent covariance inflation does not assume that the physical obstacle mean remains correct under arbitrary sensor faults. The posterior mean μ_i is the output of the perception/tracking module after data association and bias compensation. The health score $q_{h,i}$ represents the residual reliability of this posterior and inflates the uncertainty around it. Persistent sensor bias must either be absorbed into the tracker mean or represented by an additional bias envelope:*

$$z_i^{\star} = \mu_i + \hat{b}_i + b_i + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \bar{\Sigma}_i), \quad \|b_i\|_{W_b^{-1}} \leq r_{b,i}, \quad r_{b,i} = b_0(1 - q_{h,i}).$$

When no explicit bias estimator is available, $\hat{b}_i = 0$. For a local linearization $L_i(z_i) \approx \ell_i^{\top} z_i + \ell_{0,i}$,

$$\mathrm{CVaR}_{\beta_i}(L_i) = \ell_i^{\top}(\mu_i + \hat{b}_i) + \ell_{0,i} + \kappa(\beta_i)\sqrt{\ell_i^{\top}\bar{\Sigma}_i\ell_i} + r_{b,i}\|\ell_i\|_{W_b}, \quad \kappa(\beta) = \frac{\varphi(\Phi^{-1}(\beta))}{1 - \beta}.$$

The last term $r_{b,i}\|\ell_i\|_{W_b}$ is the worst-case contribution of an undetected bounded bias.

A short-horizon predicted obstacle position is defined by

$$p_i^{+}(t) = C_{\tau_i} z_i(t), \qquad C_{\tau_i} = \begin{bmatrix} 1 & 0 & \tau_i & 0 \\ 0 & 1 & 0 & \tau_i \end{bmatrix}, \qquad p_i^{+}(t) \sim \mathcal{N}\big(\bar{p}_i(t), P_i(t)\big),$$

with

$$\bar{p}_i(t) = C_{\tau_i}\mu_i(t), \qquad P_i(t) = C_{\tau_i}\bar{\Sigma}_i(t)C_{\tau_i}^{\top}.$$

The obstacle-wise CVaR confidence level is updated by

$$\beta_i(t_k) = \mathrm{clip}_{[\beta_{\min}, \beta_{\max}]}\Big(\beta_0 + c_p p_i^{\mathrm{exist}}(t_k) + c_q(1 - q_{h,i}(t_k)) + c_{\Sigma}\bar{\sigma}_i(t_k) + c_{\tau}\bar{\tau}_i(t_k)\Big),$$

and is held constant over $[t_k, t_{k+1})$. The symbol $\bar{p}_i$ is reserved only for the predicted obstacle-position mean; the existence channel is denoted by p_i^{exist}. The normalized uncertainty term is

$$\bar{\sigma}_i(t) = \text{clip}_{[0,1]}\left(\frac{\text{tr}(W_\beta^{-1}\bar{\Sigma}_i(t))}{\sigma_{\text{ref}}}\right),$$

and

$$\text{TTC}_i(t) = \frac{\bar{p}_{i,s}(t) - s(t)}{\max(v_x(t) - \mu_{v_s,i}(t), \epsilon_v)}, \quad \bar{\tau}_i(t) = \exp\left(-\frac{\text{TTC}_i(t)}{\tau_0}\right),$$

with $\text{TTC}_i(t) = +\infty$ if the posterior-mean obstacle is not ahead or the ego vehicle is not closing in. This update law increases β_i when the obstacle is more likely to exist, the sensor health is lower, the posterior uncertainty is larger, or the time-to-collision is shorter. The affine monotone law is used as a calibrated surrogate rather than an arbitrary heuristic. Its monotonicity constraints encode the required safety ordering, and the affine form keeps the mapping interpretable, avoids discontinuous confidence jumps, and preserves the convex QP structure because $\beta_i(t)$ is evaluated before solving the control QP.

The scheduling gains can be fitted offline by

$$\min_{c_p, c_q, c_\Sigma, c_\tau} \sum_k \left(\beta_i(t_k) - \beta_i^{\text{tar}}(t_k)\right)^2 + \lambda_c \|c\|_2^2 \quad \text{s.t.} \quad c_p, c_q, c_\Sigma, c_\tau \geq 0, \quad \beta_i(t_k) \in [\beta_{\min}, \beta_{\max}],$$

where $c = [c_p, c_q, c_\Sigma, c_\tau]^\top$. The target β_i^{tar} can be selected from validation logs by matching empirical upper-tail loss coverage or by enforcing a desired violation-rate envelope. Online control only evaluates the closed-form schedule.

3.2. CVaR-Based Risk Tightening of the Dynamic Obstacle Barrier

Define

$$p_e = S x_e, \qquad S = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \qquad D_i(t) = \text{diag}\left(\frac{1}{a_i^2(t)}, \frac{1}{b_i^2(t)}\right) \succ 0,$$

where $a_i = a_i(t)$ and $b_i = b_i(t)$ are the speed-dependent scales in (5)–(6); the fixed-ellipse implementation is the special case with constant a_i, b_i. The true barrier is

$$h_i(x_e, p_i^+, t) = \left(p_e - p_i^+\right)^\top D_i(t)\left(p_e - p_i^+\right) - 1.$$

In the CVaR tightening below, the speed-dependent scales are locally frozen at each sampling instant: $a_i(t_k)$ and $b_i(t_k)$ are evaluated from the posterior mean obstacle velocity and the estimated ego velocity, and are treated as known over $[t_k, t_{k+1})$. Thus D_i is deterministic within the one-step Gaussian surrogate used for the closed-form CVaR bound. Higher-order uncertainty in these scales is not claimed to be covered by the closed-form proof; it is handled conservatively through covariance inflation in $\bar{\Sigma}_i$ or, for non-Gaussian samples, through the empirical CVaR extension in Supplementary Section S8. A posterior-mean barrier surrogate and the associated linearized barrier uncertainty scale are

$$h_i^{\text{mean}}(x_e, t) = \left(p_e - \bar{p}_i\right)^\top D_i(t)\left(p_e - \bar{p}_i\right) - 1,$$
$$\sigma_{h,i}(x_e, t) = 2\sqrt{\left(p_e - \bar{p}_i\right)^\top D_i(t) P_i D_i(t)\left(p_e - \bar{p}_i\right)}.$$

Let

$$\kappa(\beta_i) = \frac{\varphi(\Phi^{-1}(\beta_i))}{1 - \beta_i}, \qquad \chi_{b,i}(t) = r_{b,i}\|\ell_i\|_{W_b},$$

where φ and Φ denote the standard normal probability density function and cumulative distribution function, respectively. The residual-bias term $\chi_{b,i}$ is set to zero when no explicit bias envelope is used. The CVaR-tightened barrier is

$$\tilde{h}_i(x_e, t) = h_i^{\text{mean}}(x_e, t) - \kappa(\beta_i)\sigma_{h,i}(x_e, t) - \chi_{b,i}(t).$$

Theorem 1 (CVaR-tightened barrier under belief uncertainty and state-estimation error)**.** *Fix an obstacle $i \in \mathcal{I}_a(t)$ and a sampling interval $[t_k, t_{k+1})$ on which $\beta_i \in (0,1)$ is held constant and $D_i(t)$ is locally frozen from the posterior mean velocities. Assume that $\tilde{h}_i$ is differentiable on Ω_x and that the component-wise estimation bound satisfies $|x_e - \hat{x}_e| \le \xi(t)$. Define*

$$\begin{aligned} \Delta_{\xi,i}(t) &:= \sup_{|e| \preceq \xi(t)} \left|\nabla_{x_e}\tilde{h}_i(\hat{x}_e(t) + e, t)\right|^\top \xi(t), \\ \bar{h}_i(\hat{x}_e, t) &:= \tilde{h}_i(\hat{x}_e, t) - \Delta_{\xi,i}(t) \\ &= h_i^{\text{mean}}(\hat{x}_e, t) - \kappa(\beta_i)\sigma_{h,i}(\hat{x}_e, t) - \chi_{b,i}(t) - \Delta_{\xi,i}(t). \end{aligned} \tag{11}$$

If $\bar{h}_i(\hat{x}_e, t) \ge 0$, then

$$\text{CVaR}_{\beta_i}\left(-h_i(x_e, p_i^+, t) \mid B_i = 1\right) \le 0.$$

Proof. The proof is given in Appendix A; it combines a conservative Gaussian surrogate, standard CVaR identities [26,27], and the component-wise estimated-state correction. □

Remark 2. *One admissible implementation evaluates the supremum by interval or vertex bounds over the error box $|\tilde{x}_e| \preceq \xi$. A simpler fallback is the Lipschitz bound $\Delta_{\xi,i}(t) = L^x_{\tilde{h},i}\|\xi(t)\|$, with $L^x_{\tilde{h},i} \le L^x_{h^{\text{mean}},i} + \kappa(\beta_i)L^x_{\sigma,i}$. If P_i is fixed over a sampling interval, then $L^x_{\sigma,i} \le 2\|(D_i P_i D_i)^{1/2} S\|$.*

Equivalently, without an explicit residual-bias envelope, the estimated-state CVaR barrier can be written as

$$\tilde{h}_i(\hat{x}_e, t) = -\text{CVaR}_{\beta_i}(L_i(\hat{x}_e, z_i, t) \mid B_i = 1).$$

The component-wise ego-error bound then produces the final robust barrier

$$\Delta_{\xi,i}(t) = \sup_{|e| \preceq \xi(t)} \left|\nabla_{x_e}\tilde{h}_i(\hat{x}_e(t) + e, t)\right|^\top \xi(t), \qquad \bar{h}_i(\hat{x}_e, t) = \tilde{h}_i(\hat{x}_e, t) - \Delta_{\xi,i}(t).$$

Thus the construction follows the chain $h_i^{\text{mean}} \to \tilde{h}_i \to \bar{h}_i \to \psi_{i,0} \to \psi_{i,1} \to$ HOCBF-QP.

3.3. Chance-Constrained Corollary

If the overall violation probability satisfies $\mathbb{P}(B_i = 1,\, h_i < 0) \le \alpha_i$, then for $p_i^{\text{exist}} > \alpha_i$ it is sufficient to impose [16]

$$\tilde{h}_i^{\text{cc}}(\hat{x}_e, t) := h_i^{\text{mean}}(\hat{x}_e, t) - \Phi^{-1}\left(1 - \frac{\alpha_i}{p_i^{\text{exist}}}\right)\sigma_{h,i}(\hat{x}_e) - \Delta^{\text{cc}}_{\xi,i}(t) \ge 0, \tag{12}$$

where $\Delta^{\text{cc}}_{\xi,i}$ is computed from the same component-wise ego-error box as in Theorem 1, using the quantile-tightened surrogate instead of the CVaR-tightened surrogate.

3.4. Risk-Adaptive HOCBF Construction

Because $\bar{h}_i(\hat{x}_e, t)$ is time-varying and the obstacle-avoidance constraint has relative degree two, define, following the HOCBF framework in [1],

$$\psi_{i,0}(\hat{x}_e, t) := \bar{h}_i(\hat{x}_e, t), \qquad \psi_{i,1}(\hat{x}_e, t) := \frac{\partial \psi_{i,0}}{\partial t} + L_{f_e}\psi_{i,0} + \alpha_1\big(\psi_{i,0}\big),$$

with class-$\mathcal{K}$ functions α_1 and α_2. The HOCBF constraint is

$$\frac{\partial \psi_{i,1}}{\partial t} + L_{f_e}\psi_{i,1} + L_{g_e}\psi_{i,1}u \geq -\alpha_2\big(\psi_{i,1}\big) + \sigma^{\text{rob}}_{i,\text{hocbf}}(t) - \delta_i, \tag{13}$$

where $\delta_i \geq 0$ and a sufficient robustness margin is obtained from the component-wise bounds

$$\big|F_i(x_e, u, t) - \hat{F}_i(\hat{x}_e, u, t)\big| \leq k_{F,i}^{\top}\xi(t) + B_{\psi_1,i}d_{\max},$$
$$\big|\alpha_2(\psi_{i,1}(x_e, t)) - \alpha_2(\hat{\psi}_{i,1}(t))\big| \leq k_{\alpha,i}^{\top}\xi(t),$$

where $k_{F,i} \succeq 0$ and $k_{\alpha,i} \succeq 0$. Hence

$$\sigma^{\text{rob}}_{i,\text{hocbf}}(t) = (k_{F,i} + k_{\alpha,i})^{\top}\xi(t) + B_{\psi_1,i}d_{\max}. \tag{14}$$

Within each sampling interval, β_i is constant, so the dominant time variation enters through $\bar{p}_i(t)$, $P_i(t)$, $\hat{x}_e(t)$, and $\xi(t)$. Related robustness ideas for state uncertainty and measurement uncertainty can also be found in [4,5].

3.5. Quadratic-Program Safety Filter

The safety filter computes the applied control by

$$\min_{u,\delta} \ \|u - u_{\text{nom}}\|_R^2 + \sum_{i \in \mathcal{I}_a} \rho_i \delta_i^2$$

subject to

$$\frac{\partial \psi_{i,1}}{\partial t} + L_{f_e}\psi_{i,1} + L_{g_e}\psi_{i,1}u \geq -\alpha_2\big(\psi_{i,1}\big) + \sigma^{\text{rob}}_{i,\text{hocbf}}(t) - \delta_i, \qquad \forall i \in \mathcal{I}_a(t),$$

$$u_{\min} \leq u \leq u_{\max}, \qquad \delta_i \geq 0, \qquad \forall i \in \mathcal{I}_a(t).$$

The compact notation above emphasizes the predicted-obstacle HOCBF rows. The implementation used for the reported experiments solves the generalized row-wise QP

$$\mathcal{J}(t) = \mathcal{J}_{\text{cur}}(t) \cup \mathcal{J}_{\text{pred}}(t) \cup \mathcal{J}_{\text{lane}}(t),$$

where $\mathcal{J}_{\text{cur}}$, $\mathcal{J}_{\text{pred}}$, and $\mathcal{J}_{\text{lane}}$ denote current-obstacle, predicted-obstacle, and upper/lower lane HOCBF rows, respectively. The implemented decision vector has $u \in \mathbb{R}^2$, one slack δ_j for each row $j \in \mathcal{J}(t)$, and a shared slack-envelope variable η:

$$\min_{u,\{\delta_j\},\eta} \ \|u - u_{\text{nom}}\|_R^2 + \sum_{j \in \mathcal{J}(t)} \rho_j \delta_j^2 + \rho_\eta \eta^2$$

subject to

$$\hat{F}_j(\hat{x}_e, u, t) + \alpha_{2,j}\big(\hat{\psi}_{j,1}(t)\big) \geq \sigma_j^{\text{rob}}(t) - \delta_j, \qquad \forall j \in \mathcal{J}(t),$$
$$0 \leq \delta_j \leq \eta, \qquad \eta \geq 0, \qquad u_{\min} \leq u \leq u_{\max}, \qquad |u - u_{\text{prev}}| \preceq \dot{u}_{\max}\Delta t.$$

Thus if $C = |\mathcal{J}(t)|$, the implementation QP has

$$n_{\text{var}} = 2 + C + 1$$

decision variables. The final $+1$ is the shared slack-envelope variable η, which explains the variable counts used in the runtime scaling study. Real-time feasibility with increasing active-obstacle count is evaluated in Supplementary Section S7. In the obstacle-only scaling benchmark, each of N active obstacles contributes one slackened row, so $C = N$ and the reported 95th-percentile and maximum solve times remain below the 50 ms control period for $N \leq 30$. The generalized slack penalty is

$$\rho_j = \rho_0\Big(1 + r_p p_i^{\text{exist}} + r_q(1 - q_{h,i}) + r_\Sigma \bar{\sigma}_i + r_\tau \bar{\tau}_i\Big).$$

For lane rows, the obstacle-dependent terms are omitted. For the reported simulations, $r_p = r_q = r_\Sigma = r_\tau = 0$, so all compared methods use the same fixed slack penalty while retaining the same generalized QP structure. The main formulation uses conditional CVaR; obstacle existence influences conservatism through β_i. The chance-constrained form in (12) is included only for comparison. When $\delta_j \equiv 0$ for all $j \in \mathcal{J}(t)$, the QP enforces the risk-adaptive HOCBF exactly; otherwise it should be interpreted as operating in a least-violation mode.

4. Simulation Experiments and Discussion

This section evaluates whether the proposed safety filter exposes the three quantities that matter for a safety-critical lane-changing task: the current-obstacle margin h_{cur}, the predicted-obstacle margin h_{pred}, and the road-boundary margin h_{lane}. We also report the control intervention magnitude

$$J_u(t) = \|u(t) - u_{\text{nom}}(t)\|_2,$$

which quantifies how strongly the safety filter departs from the nominal controller. All reported minimum margins are computed from the raw simulation logs, not from visual readings of the figures.

4.1. Simulation Setup and Compared Methods

The ego vehicle is simulated with the nonlinear road-aligned bicycle surrogate in (1). The state is $x_e = [v_x, v_y, r, s, w, \theta]^\top$ and the control is $u = [a, \delta_f]^\top$. The sampling period is 0.05 s and each rollout lasts 14 s. The actuator bounds are $a \in [-2.5, 2.0]$ m/s^2 and $\delta_f \in [-0.34, 0.34]$ rad. The nominal controller tracks $v_x = 23$ m/s and attempts a lane change from the lower lane center $w = 0$ m to the upper lane center $w = 3.5$ m. Road boundaries are encoded as lane-barrier HOCBF constraints after subtracting the ego half-width and a 0.15 m buffer. The speed-dependent obstacle ellipse uses base scales $a_0 = 11.0$ m and $b_0 = 2.8$ m; the effective scales are updated online from posterior mean relative velocity and current/predicted position covariance, with the implementation coefficients summarized in the QP settings table below. The short-horizon prediction time is $\tau_i = 0.55$ s.

Five filters are compared. The proposed method uses the health-aware CVaR tightening in (11). The fixed-CVaR baseline uses the same CVaR form but keeps $\beta_i = 0.90$. The chance-constrained baseline uses the Gaussian quantile tightening in (12). The belief-aware, health-unaware baseline uses the belief covariance but suppresses the sensor-health degradation channel by setting $q_{h,i} = 1$ in the belief update. In the plot legends, this baseline is abbreviated as "belief no health" for space economy. The deterministic HOCBF baseline ignores probabilistic obstacle uncertainty and retains only the deterministic estimated-state

robustness correction. All methods share the same nominal controller, actuator bounds, road-boundary constraints, and slack penalty.

Table 4 summarizes the simulation scenarios and their placement in the manuscript or Supplementary Material.

Table 4. Scenario design used in the simulation study. Scenario A is treated as a warm-up case and is reported in Supplementary Section S1; the main text focuses on Scenarios B–D.

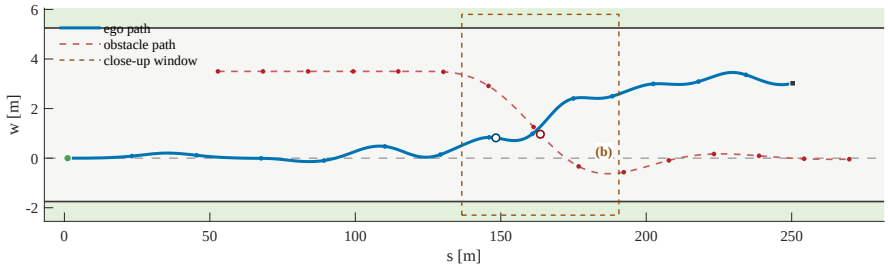
Scenario	Purpose	Main Stressor	Placement
A	Baseline cut-in	Feasible single-obstacle cut-in under nominal sensing degradation	Supp. S1
B	Critical cut-in	More critical single-obstacle encounter that reveals local geometry and control cost	Main text
C	Sudden perception degradation	Sensor health drops over the interaction window, increasing belief uncertainty and risk tightening	Main text
D	Merge bottleneck	Moving obstacle and narrowing road boundary jointly compress the feasible corridor	Main text

4.2. Scenario B: Critical Single-Obstacle Cut-In

Scenario B tests a critical single-obstacle cut-in under nominal road geometry. Figure 2 shows the global trajectory and the local critical encounter, including vehicle footprints and speed-dependent current/predicted obstacle contours.

Figure 3 shows that all methods remain safe, but with different margin-effort trade-offs. The deterministic HOCBF has the smallest current-obstacle margin, $h_{\text{cur,min}} = 0.420$, whereas the proposed filter keeps $h_{\text{cur,min}} = 0.931$, $h_{\text{pred,min}} = 3.673$, and $h_{\text{lane,min}} = 0.561$ with mean intervention 2.910.

(**a**)



(**b**)

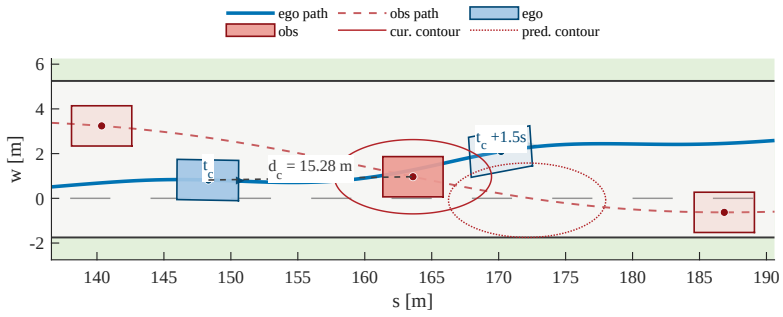


Figure 2. Scenario B: critical cut-in geometry. Panel (**a**) shows the macro trajectory and the dashed brown close-up window. Panel (**b**) shows the vehicle-shape snapshot at the critical encounter. The blue solid curve denotes the ego path, the red dashed curve denotes the obstacle path, the blue and red rectangles denote ego and obstacle footprints, respectively, and the solid and dotted red contours denote the current and predicted obstacle safety contours.

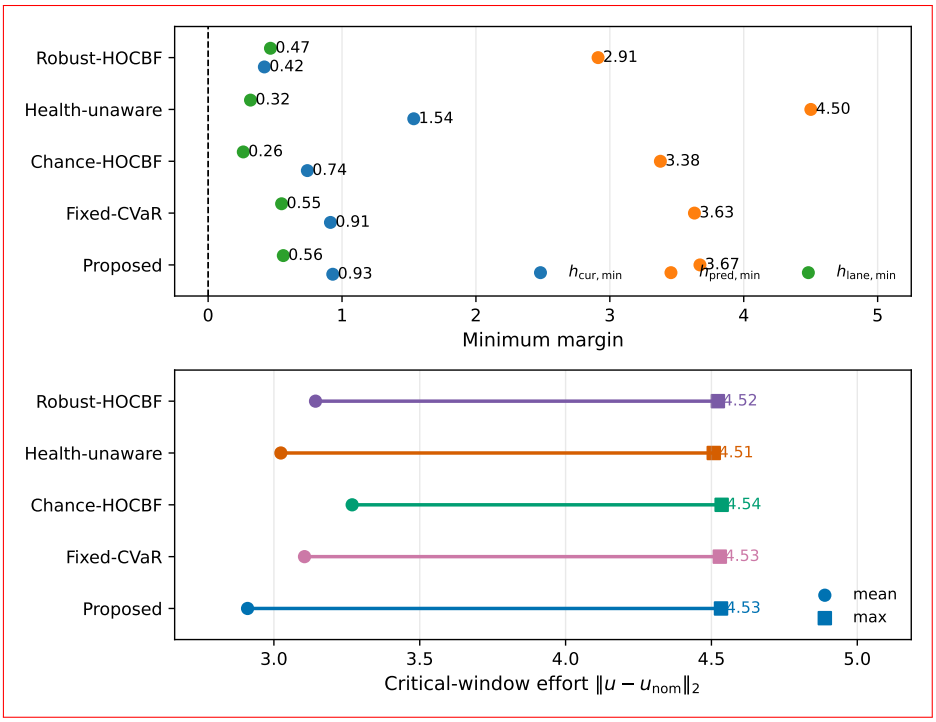


Figure 3. Scenario B safety and control-effort summaries. In the upper panel, blue, orange, and green markers denote minimum current-obstacle, predicted-obstacle, and lane margins, respectively; the vertical dashed black line marks the zero safety boundary. In the lower panel, colored horizontal segments compare the critical-window mean intervention (circle markers) and maximum intervention (square markers) for each method.

4.3. Scenario C: Sudden Perception Degradation

Scenario C isolates sudden perception degradation. The obstacle interaction is similar to Scenario B, but the object-wise health score drops during the critical interval. Figure 4 shows the resulting chain: lower $q_{h,i}$, larger predicted barrier uncertainty, increased adaptive CVaR confidence, and stronger prediction tightening.

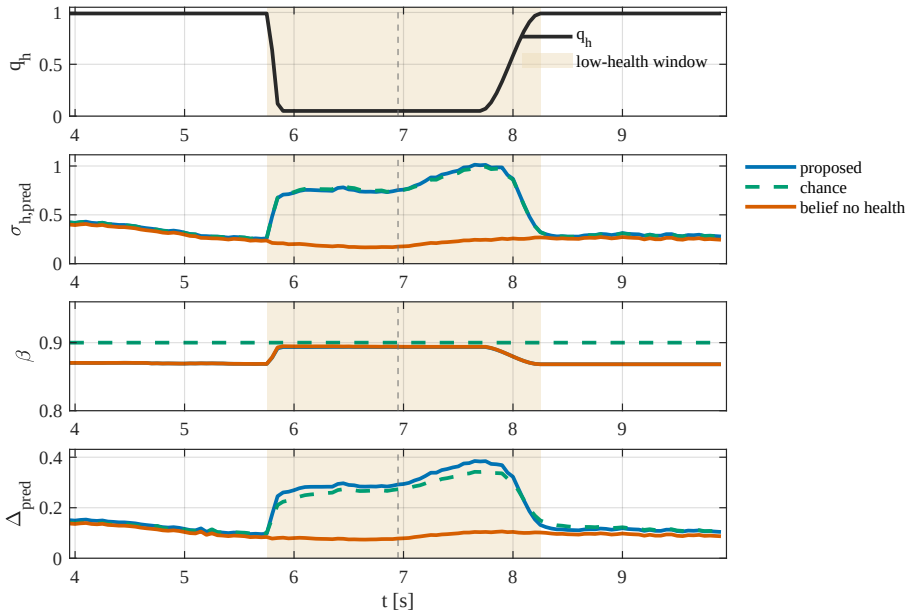


Figure 4. Scenario C risk-adaptation timeline. The low-health interval causes the predicted barrier uncertainty to grow; the proposed filter increases its risk confidence and effective HOCBF tightening accordingly.

All margins remain positive in this rollout. The proposed method obtains $h_{\text{cur,min}} = 1.460$, $h_{\text{pred,min}} = 4.517$, and $h_{\text{lane,min}} = 0.375$, with mean intervention 3.151, comparable to the chance-constrained value 3.037. The key result is the interpretable health-dependent adaptation, which deterministic and fixed-risk formulations do not provide.

4.4. Scenario D: Merge Bottleneck with Joint Lane and Obstacle Constraints

Scenario D adds a merge-like bottleneck, so the ego vehicle cannot avoid the cut-in obstacle by using a large lateral detour. Figure 5 shows the compressed feasible corridor where the obstacle barrier and road boundary jointly constrain the maneuver.

As shown in Figure 6, the deterministic HOCBF loses lane-boundary margin, with $h_{\text{lane,min}} = -0.021$. The proposed filter keeps all three margins positive, $h_{\text{cur,min}} = 1.114$, $h_{\text{pred,min}} = 3.825$, and $h_{\text{lane,min}} = 0.340$, at the cost of higher mean intervention (3.841 versus 3.212 for the deterministic baseline).

Table 5 reports the main quantitative results for Scenarios B–D.

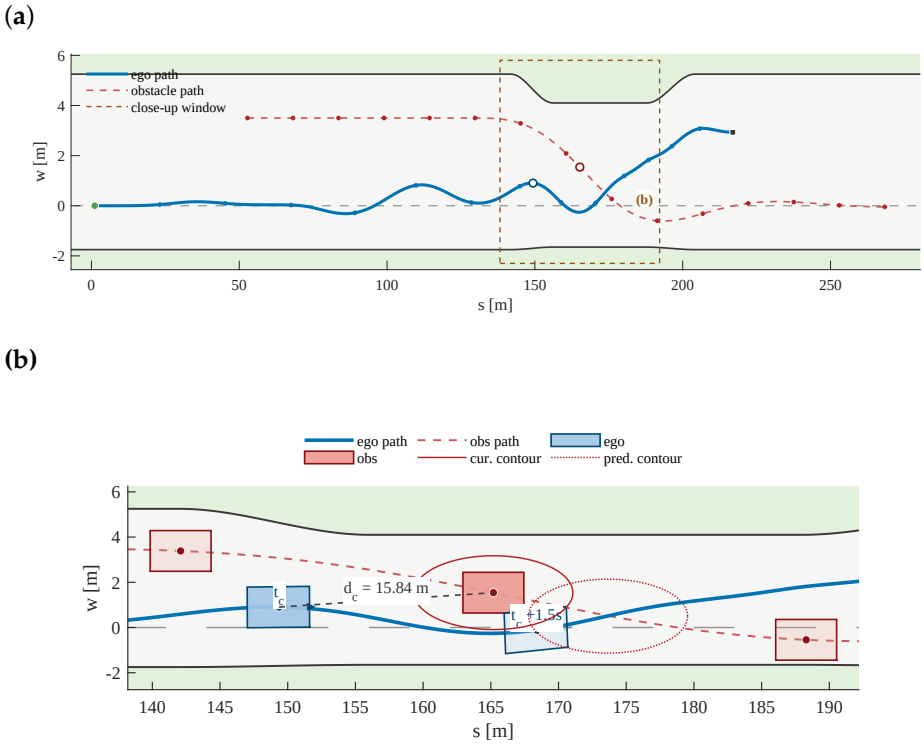


Figure 5. Scenario D: merge bottleneck geometry. Panel (**a**) shows the macro trajectory through the compressed road segment and the dashed brown close-up window. Panel (**b**) shows the local critical geometry with vehicle footprints, speed-dependent obstacle contours, and narrowed road boundaries. The blue solid curve denotes the ego path, the red dashed curve denotes the obstacle path, the blue and red rectangles denote ego and obstacle footprints, respectively, and the solid and dotted red contours denote the current and predicted obstacle safety contours.

Table 5. Main quantitative results for Scenarios B–D from the simulation logs. Positive margin values indicate that the corresponding safety boundary is respected. The intervention cost is reported over the critical time window centered at the proposed method's minimum current-obstacle margin.

Scenario	Method	$h_{\text{cur,min}}$	$h_{\text{pred,min}}$	$h_{\text{lane,min}}$	Mean J_u	Max J_u
B	proposed	0.931	3.673	0.561	2.910	4.533
B	fixed-CVaR	0.913	3.632	0.549	3.105	4.529
B	chance-constrained	0.741	3.378	0.262	3.268	4.535
B	belief-aware, health-unaware	1.536	4.501	0.316	3.024	4.508
B	deterministic	0.420	2.912	0.465	3.143	4.522

Table 5. *Cont.*

Scenario	Method	$h_{\text{cur,min}}$	$h_{\text{pred,min}}$	$h_{\text{lane,min}}$	Mean J_u	Max J_u
C	proposed	1.460	4.517	0.375	3.151	4.514
C	fixed-CVaR	1.244	4.347	0.384	2.937	4.503
C	chance-constrained	1.500	4.555	0.381	3.037	4.514
C	belief-aware, health-unaware	1.435	4.482	0.378	2.772	4.512
C	deterministic	1.273	4.414	0.416	2.927	4.502
D	proposed	1.114	3.825	0.340	3.841	4.541
D	fixed-CVaR	1.120	3.831	0.239	3.881	4.541
D	chance-constrained	1.142	3.941	0.380	3.623	4.541
D	belief-aware, health-unaware	1.630	4.626	0.365	3.730	4.536
D	deterministic	0.338	2.519	-0.021	3.212	4.533

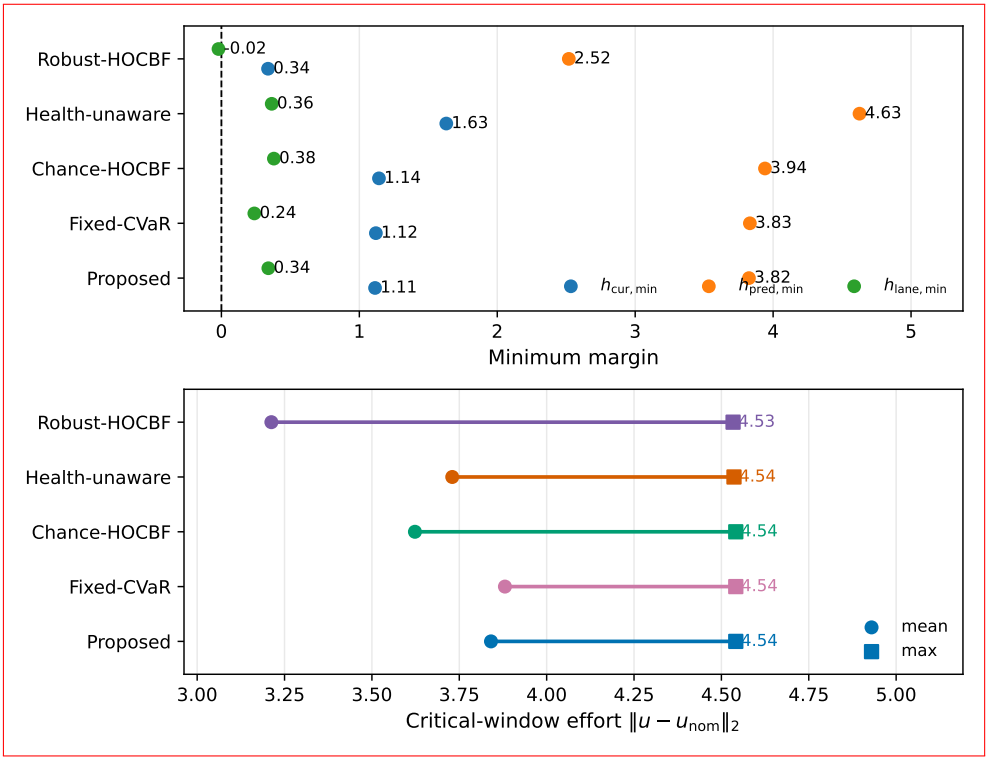


Figure 6. Scenario D safety and control-effort summaries from the same simulation logs as Table 5. In the upper panel, blue, orange, and green markers denote minimum current-obstacle, predicted-obstacle, and lane margins, respectively; the vertical dashed black line marks the zero safety boundary. In the lower panel, colored horizontal segments compare the critical-window mean intervention (circle markers) and maximum intervention (square markers) for each method. The deterministic baseline crosses the lane-boundary safety boundary, while the proposed method preserves positive current, predicted, and lane margins under bottleneck compression.

4.5. Additional Baselines, Solver Settings, and Comparison Protocol

The nominal controller tracks the reference lane-change trajectory and desired longitudinal speed. The desired acceleration and steering command are first computed by the nominal controller and then passed to the HOCBF-QP safety filter. With $u_{\text{nom}} = [a_{\text{nom}}, \delta_{f,\text{nom}}]^\top$,

$$a_{\text{nom}} = k_v(v_{\text{ref}} - v_x), \qquad e_w = w - w_{\text{ref}}, \quad e_\theta = \theta - \theta_{\text{ref}}, \qquad \delta_{f,\text{nom}} = -k_w e_w - k_\theta e_\theta - k_r r.$$

All methods use the same nominal controller, sampling period, input bounds, road-boundary constraints, and random seeds.

Table 6 summarizes the nominal controller, speed-dependent safety-region, and QP implementation settings used for all compared methods.

Table 7 positions the proposed method relative to deterministic robust CBF/HOCBF, stochastic CBF, chance-constrained CBF/MPC, distributionally robust or risk-aware CBF, and learning-based safety filters. The numerical baselines in Tables 5 and 8 are limited to CBF-QP variants that share the same vehicle dynamics, obstacle logs, nominal controller, and input bounds. We therefore do not claim numerical superiority over learning-based safety filters that were not reproduced in this simulation environment. Scenario E adds a heterogeneous multi-obstacle stress replay in Supplementary Section S5, Scenario F evaluates heavy-tailed perception uncertainty in Supplementary Section S8, and Supplementary Sections S10–S12 add component ablations, false-detection replays, and a curved-road dropout-latency replay.

Table 6. Nominal controller, speed-dependent safety-region, and QP implementation settings used for all compared methods.

Item	Value
Sampling time/horizon	$\Delta t = 0.05$ s, rollout length 14 s
Nominal speed and lane reference	$v_{\mathrm{ref}} = 23$ m/s; lane reference uses cubic smoothstep from $w = 0$ m to $w = 3.5$ m over $t = 3.0$–5.8 s
Nominal gains	$k_v = 0.32$, $k_w = 0.070$, $k_\theta = 1.75$, $k_r = 0.25$, $k_{v_y} = 0.08$
Nominal command limits	$\lvert\theta_{\mathrm{ref}}\rvert \le 0.18$ rad, $\lvert\delta_{f,\mathrm{nom}}\rvert \le 0.27$ rad
Input bounds	$a \in [-2.5, 2.0]$ m/s^2, $\delta_f \in [-0.34, 0.34]$ rad
Input-rate bounds	$\lvert\Delta a\rvert \le 4.0\Delta t$ m/s^2, $\lvert\Delta\delta_f\rvert \le 1.0\Delta t$ rad per step
Safety ellipse base scales	$a_0 = 11.0$ m, $b_0 = 2.8$ m
Speed/uncertainty scale coefficients	$c_{a,v} = 0.08$, $c_{b,v} = 0.10$, $c_{a,\Sigma} = 0.40$, $c_{b,\Sigma} = 0.25$, $c_b = 0.10$
Safety ellipse clipping bounds	$a_i \in [10.0, 17.0]$ m, $b_i \in [2.5, 4.1]$ m
QP solver	MATLAB R2020a `quadprog`, interior-point-convex with active-set fallback
Solver tolerance/iterations	constraint tolerance 10^{-8}, optimality tolerance 10^{-9}, step tolerance 10^{-12}, maximum 200 iterations
Warm start	disabled for the explicit `quadprog` initial point; previous input used in the objective and rate bounds
Slack penalty base ρ_0	5.0×10^4
Shared envelope penalty ρ_η	$50\rho_0$
Hardware	AMD Ryzen 7 8845H CPU with Radeon 780M integrated GPU

Table 7. Taxonomy of related safety-filter classes and how they relate to the implemented baselines. A check mark means the class directly addresses the listed feature in its standard form or in the cited representative work. Numerical comparisons in this paper are restricted to methods that can be implemented under the same vehicle model, obstacle logs, and HOCBF-QP architecture.

Class	Representative References	Stochastic Belief	Heavy-Tail/ Shift	Sensor Health	CBF/HOCBF Certificate	Role in This Paper
Deterministic robust CBF/HOCBF	[4,5]	–	–	–	✓	Implemented deterministic robust baseline with bounded ego-error correction.
Stochastic CBF	[11,12]	✓	partial	–	✓	Compared conceptually; motivates probabilistic safety constraints.
Chance-constrained CBF/MPC	[16]	✓	–	–	partial	Implemented Gaussian chance-constrained HOCBF baseline.
Belief-space safety filters	[13]	✓	–	–	✓	Implemented belief-aware, health-unaware baseline by fixing $q_{h,i} = 1$.
Risk-aware/ distributionally robust CBF	[14,15]	✓	✓	–	✓	Implemented fixed-CVaR and empirical-CVaR variants.
Learning-based safety filters	[23,24]	model dependent	model dependent	possible	possible	Discussed as related perception-control interfaces; not claimed numerically dominated.
Proposed health-aware CVaR-HOCBF	this work	✓	empirical extension	✓	✓	Object-wise health, covariance inflation, adaptive CVaR, and robust ego tightening.

Table 8 provides additional baseline metrics for Scenarios B–D.

Table 9 lists the risk-scheduling, covariance-inflation, and slack-penalty parameters used in the reported simulations.

Table 8. Additional baseline metrics for Scenarios B–D from the simulation logs. $\psi_{0,\mathrm{pred,min}}$ is the minimum risk-tightened predicted-obstacle HOCBF margin. Effort is $J_u(t) = \|u(t) - u_{\mathrm{nom}}(t)\|_2$. Slack duty is the fraction of samples with positive obstacle or lane slack, integrated slack is $\int \max_j \delta_j(t)\, dt$, $h_{\mathrm{slack,min}}$ is the minimum current-obstacle margin during slack-active samples, and t_{95} is the 95th-percentile QP solve time.

Scen.	Method	$\psi_{0,\mathrm{pred}}$	Mean J_u	Max J_u	Duty	Max δ	$\int \delta dt$	$h_{\delta,\mathrm{min}}$	t_{95}
B	Proposed	3.200	1.760	4.533	52.3%	9.194	9.609	0.931	6.53
B	Fixed-CVaR	3.153	1.710	4.529	49.5%	8.021	8.705	0.913	6.47
B	Chance-HOCBF	2.910	1.663	4.535	49.1%	6.204	9.610	0.742	5.96
B	Health-unaware	4.090	1.598	4.508	49.1%	8.853	10.803	1.536	6.24
B	Robust-HOCBF	2.590	1.618	4.522	50.5%	7.581	9.315	0.422	6.76
C	Proposed	3.317	1.667	4.514	46.6%	16.552	10.443	1.463	5.68
C	Fixed-CVaR	2.987	1.766	4.503	49.1%	16.651	10.673	1.244	5.78
C	Chance-HOCBF	3.372	1.660	4.514	47.3%	16.087	10.681	1.500	5.68
C	Health-unaware	4.049	1.781	4.512	50.2%	9.648	10.400	1.435	5.56
C	Robust-HOCBF	3.343	1.755	4.502	43.8%	17.233	9.522	1.273	5.94
D	Proposed	3.344	2.087	4.541	50.2%	6.590	9.837	1.114	5.58
D	Fixed-CVaR	3.347	2.097	4.541	49.8%	6.568	9.996	1.120	5.42
D	Chance-HOCBF	3.417	1.980	4.541	49.5%	5.389	8.842	1.142	5.50
D	Health-unaware	4.198	2.032	4.536	48.8%	7.032	9.610	1.630	5.57
D	Robust-HOCBF	2.174	1.701	4.533	50.9%	6.660	9.712	0.338	5.81

Table 9. Risk-scheduling, covariance-inflation, and slack-penalty parameters used in the reported simulations. Scenario-specific entries are shown where the closed-loop study intentionally changes the risk envelope.

Parameter	Value	Role
λ_h	5.5	Health-dependent covariance-inflation gain
$\beta_{\min}$	A/B/C/D: 0.88/0.84/0.84/0.86	Lower bound for adaptive CVaR confidence
$\beta_{\max}$	A/B/C/D: 0.95/0.92/0.94/0.94	Upper bound for adaptive CVaR confidence
β_0	A/B/C/D: 0.90/0.86/0.86/0.88	Baseline CVaR confidence before scheduling terms
c_p	0.008	Existence-probability scheduling gain
c_q	0.015; Scenario C: 0.028	Health-loss scheduling gain
c_Σ	0.0015; Scenario C: 0.0022	Covariance scheduling gain
c_τ	0.008	TTC scheduling gain
σ_{ref}	20.0	Normalization scale for $\bar{\sigma}_i$
W_β	$\mathrm{diag}(0.85^2, 0.50^2, 0.85^2, 0.50^2)$	Covariance normalization in the risk schedule
τ_0	5.0 s	TTC decay scale in $\bar{\tau}_i$
ϵ_v	0.10 m/s	Closing-speed floor for posterior-mean TTC
$r_p, r_q, r_\Sigma, r_\tau$	0, 0, 0, 0	Reported experiments use the fixed slack-penalty special case

4.6. Discussion

Across Scenarios B–D, the proposed method is most valuable when probabilistic obstacle uncertainty and geometric constraint compression are both active. Scenario B shows that the method preserves positive current, predicted, and lane margins in a critical cut-in without becoming the most aggressive controller in the critical window. Scenario C makes the internal risk-adaptation pathway visible: sensing degradation enlarges prediction uncertainty, raises the scheduled risk level, and strengthens the effective HOCBF tightening. Scenario D shows why this additional information matters in a compressed feasible corridor: the deterministic HOCBF incurs a small but explicit lane-boundary violation, whereas the proposed method remains within the joint safe set.

The results also clarify the role of the baselines. The chance-constrained filter can perform competitively when a fixed confidence level is already conservative enough, but it does not adapt its tightening level to sudden health degradation. The belief-aware,

health-unaware variant can preserve large margins in some cases, but it lacks a direct mechanism that links sensor-health loss to the belief update and the resulting risk schedule. The deterministic filter is computationally and conceptually simpler, but it is also the first to lose lane safety in the merge bottleneck. The proposed method is not the best on every effort metric; instead, it trades additional intervention in Scenario D for positive obstacle and lane margins. Taken together, these observations support the proposed health-aware CVaR-HOCBF construction as a practical bridge from perception degradation to safety-critical control action, particularly in scenarios where online sensing quality materially changes the required safety margin.

The formal zero-slack safety guarantee depends on three conditions: a valid ego estimation error envelope, calibrated obstacle belief, and QP feasibility without slack. When slack is activated, the method provides a least-violation action with an explicit slack-dependent violation bound, not strict forward invariance. The Gaussian closed-form CVaR tightening is exact only under the local Gaussian belief surrogate; for heavy-tailed or multimodal perception errors, empirical CVaR or distributionally robust CVaR should be used. Finally, the sensor-health score is not a replacement for perception fault diagnosis; it is a compact reliability interface between perception and control.

4.7. Supplementary Material Overview

Additional validation results are provided in the Supplementary Material, including the warm-up cut-in scenario, complete numerical summaries, constraint-time diagnostics, heterogeneous multi-obstacle stress replay, sensitivity analysis, runtime scaling, heavy-tailed and multimodal uncertainty tests, slack-bound visualization, component ablations, false-detection stress replays, and a curved-road dropout-latency replay. Table 10 summarizes the organization.

Table 10. Organization of the Supplementary Material.

Supplementary Section	Content	Purpose
S1	Warm-up cut-in	Basic sanity check
S2	Complete numerical summary	Full metric record
S3	Constraint-time diagnostics	Barrier and slack timeline
S4	Scenario C/D supplementary plots	Additional closed-loop evidence
S5	Multi-obstacle stress replay	Scalability and heterogeneous interaction
S6	Sensitivity analysis	Risk-scheduling robustness
S7	Runtime scaling	Real-time feasibility
S8	Heavy-tailed uncertainty	Non-Gaussian robustness
S9	Slack-bound illustration	Least-violation interpretation
S10	Component ablation	Isolated contribution of health scheduling, covariance inflation, CVaR, and ego-error compensation
S11	False-detection stress replays	False-positive and dropout behavior under maintained beliefs
S12	Curved-road dropout-latency replay	Realistic occlusion, dropout, and latency stress case

5. Conclusions

This paper presented a sensor-health- and belief-aware risk-adaptive HOCBF safety filter for dynamic obstacle avoidance. The method maps online sensor health and probabilistic obstacle belief to CVaR-based barrier tightening terms and embeds them in a relative-degree-two HOCBF-QP. The theoretical analysis establishes sufficient forward-safety conditions under zero-slack operation and clarifies the interpretation of slack-

activated least-violation behavior. Simulation results show that the proposed filter preserves current-obstacle, predicted-obstacle, and lane-boundary margins in critical cut-in, sudden perception-degradation, and merge-bottleneck scenarios, while keeping the required control intervention comparable to competing baselines and making the health-dependent risk adaptation explicit.

The slack-activated mode is also related to unavoidable-collision mitigation, where the objective changes from strict safety preservation to selecting the least harmful feasible maneuver. Connecting the proposed least-violation HOCBF-QP with crash-severity-aware decision making [28] is left as a future extension.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/technologies14050310/s1, Supplementary Sections S1–S12: warm-up cut-in validation, complete numerical summary, constraint-time diagnostics, Scenario C/D supplementary plots, multi-obstacle stress replay, sensitivity analysis, runtime scaling, heavy-tailed uncertainty evaluation, slack-bound illustration, component ablation, false-detection stress replays, and curved-road dropout-latency replay.

Author Contributions: Conceptualization, Y.M. and Y.H.; methodology, Y.M. and Y.H.; software, Y.M.; validation, Y.M., and G.Z.; formal analysis, Y.M.; investigation, Y.M.; resources, G.Z. and Y.H.; data curation, Y.M.; writing—original draft preparation, Y.M.; writing—review and editing, Y.M., Y.H., and G.Z.; visualization, Y.M.; supervision, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and code supporting the findings of this study are available from the corresponding author upon reasonable request.

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

Appendix A. Proof Details and Slack-Bound Summary

This appendix contains the proof details needed for the main manuscript. It gives the proof of Theorem 1, the robust HOCBF argument used by the zero-slack certificate, the slack-dependent least-violation bound, and a practical estimation procedure for the robustness constants.

Appendix A.1. Proof of Theorem 1

Proof of Theorem 1. For obstacle i, let $p_i^+ = \bar{p}_i + \eta_i$ with $\eta_i \mid (B_i = 1) \sim \mathcal{N}(0, P_i)$. Over one sampling interval, the speed-dependent matrix $D_i = \mathrm{diag}(1/a_i^2, 1/b_i^2)$ is locally frozen from posterior mean velocities and is deterministic for the Gaussian surrogate. With $e_i = p_e - \bar{p}_i$,

$$h_i(x_e, p_i^+, t) = (e_i - \eta_i)^\top D_i (e_i - \eta_i) - 1 \geq h_i^{\mathrm{mean}}(x_e, t) - 2e_i^\top D_i \eta_i,$$

because $D_i \succeq 0$. Thus the true loss $L_i = -h_i$ is upper bounded almost surely by the Gaussian linear surrogate

$$L_i^{\mathrm{lin}} = -h_i^{\mathrm{mean}}(x_e, t) + 2e_i^\top D_i \eta_i, \qquad \sigma_{h,i} = 2\sqrt{e_i^\top D_i P_i D_i e_i}.$$

The Gaussian CVaR identity gives

$$\mathrm{CVaR}_{\beta_i}(L_i^{\mathrm{lin}} \mid B_i = 1) = -h_i^{\mathrm{mean}}(x_e, t) + \kappa(\beta_i)\sigma_{h,i}(x_e, t),$$

where $\kappa(\beta_i) = \varphi(\Phi^{-1}(\beta_i))/(1-\beta_i)$. By the monotonicity of CVaR, $L_i \le L_i^{\mathrm{lin}}$ implies that the CVaR of the true loss is no larger than this surrogate value. Including the optional residual-bias envelope $\chi_{b,i}$ yields the tightened barrier $\tilde{h}_i = h_i^{\mathrm{mean}} - \kappa(\beta_i)\sigma_{h,i} - \chi_{b,i}$.

The component-wise ego-state error envelope then gives

$$\tilde{h}_i(x_e, t) \ge \tilde{h}_i(\hat{x}_e, t) - \sup_{|e| \preceq \xi(t)} \left|\nabla_{x_e}\tilde{h}_i(\hat{x}_e + e, t)\right|^\top \xi(t) = \bar{h}_i(\hat{x}_e, t).$$

Therefore $\bar{h}_i(\hat{x}_e, t) \ge 0$ is sufficient for $\mathrm{CVaR}_{\beta_i}(-h_i(x_e, p_i^+, t) \mid B_i = 1) \le 0$. □

Appendix A.2. Robust HOCBF Proof

Recall the risk-tightened and ego-error-robust HOCBF variable

$$\psi_{i,0}(x_e, t) = \bar{h}_i(x_e, t), \qquad \hat{\psi}_{i,0}(t) = \psi_{i,0}(\hat{x}_e, t),$$

and the first-order HOCBF auxiliary function

$$\psi_{i,1}(x_e, t) = \partial_t\psi_{i,0}(x_e, t) + \nabla_{x_e}\psi_{i,0}(x_e, t)^\top f_e(x_e) + \alpha_1(\psi_{i,0}(x_e, t)).$$

Let $\hat{\psi}_{i,1}(t)$ be defined by replacing x_e with $\hat{x}_e$. Define the true and estimated second-order HOCBF left-hand sides as

$$F_i(x_e, u, t) := \partial_t\psi_{i,1} + \nabla_{x_e}\psi_{i,1}^\top\big(f_e(x_e) + g_e(x_e)u + d_e(t)\big),$$

$$\hat{F}_i(\hat{x}_e, u, t) := \partial_t\psi_{i,1} + \nabla_{x_e}\psi_{i,1}^\top\big(f_e(\hat{x}_e) + g_e(\hat{x}_e)u\big),$$

where the derivatives in the second expression are evaluated at $\hat{x}_e$.

Assumption A1. *The functions f_e, g_e, $\psi_{i,0}$, and $\psi_{i,1}$ are continuously differentiable with respect to x_e on a compact operating domain Ω_x, uniformly over the considered sampling interval.*

Assumption A2. *There exist non-negative sensitivity vectors $k_{F,i}$ and $k_{\alpha,i}$, and a finite constant $B_{\psi_1,i}$, such that for all $x_e, \hat{x}_e \in \Omega_x$ satisfying $|x_e - \hat{x}_e| \preceq \xi(t)$ and all admissible $u \in \mathcal{U}$,*

$$|F_i(x_e, u, t) - \hat{F}_i(\hat{x}_e, u, t)| \le k_{F,i}^\top\xi(t) + B_{\psi_1,i}d_{\max},$$

$$|\alpha_2(\psi_{i,1}(x_e, t)) - \alpha_2(\hat{\psi}_{i,1}(t))| \le k_{\alpha,i}^\top\xi(t).$$

Lemma A1. *Under Assumptions A1 and A2,*

$$F_i(x_e, u, t) + \alpha_2(\psi_{i,1}(x_e, t)) \ge \hat{F}_i(\hat{x}_e, u, t) + \alpha_2(\hat{\psi}_{i,1}(t)) - (k_{F,i} + k_{\alpha,i})^\top\xi(t) - B_{\psi_1,i}d_{\max}.$$

Proof. Adding the two bounds in Assumption A2 gives the stated inequality. The additive disturbance term is separated through

$$\left|\nabla_{x_e}\psi_{i,1}(x_e, t)^\top d_e(t)\right| \le B_{\psi_1,i}d_{\max}.$$

The remaining state-mismatch terms are bounded on Ω_x by the mean-value theorem and the sampled or analytic sensitivity envelopes summarized in Appendix A.4. □

Theorem A1. *Fix obstacle $i \in \mathcal{I}_a(t)$ and suppose Assumptions A1 and A2 hold. If the QP enforces the zero-slack estimated-state condition*

$$\hat{F}_i(\hat{x}_e, u, t) + \alpha_2(\hat{\psi}_{i,1}(t)) \geq (k_{F,i} + k_{\alpha,i})^\top \xi(t) + B_{\psi_1,i} d_{\max},$$

then

$$F_i(x_e, u, t) + \alpha_2(\psi_{i,1}(x_e, t)) \geq 0.$$

Consequently, if $\psi_{i,0}(x_e, 0) \geq 0$ and $\psi_{i,1}(x_e, 0) \geq 0$, then the risk-tightened set remains forward invariant.

Proof. The first claim follows immediately from Lemma A1 and the zero-slack QP inequality. Since $F_i(x_e, u, t) = \dot{\psi}_{i,1}(x_e, t)$, the inequality is equivalent to

$$\dot{\psi}_{i,1}(x_e, t) \geq -\alpha_2(\psi_{i,1}(x_e, t)).$$

The forward-invariance statement follows by the standard HOCBF comparison argument: applying the comparison principle to $\psi_{i,1}$ preserves $\psi_{i,1} \geq 0$, and then applying it to $\dot{\psi}_{i,0} \geq -\alpha_1(\psi_{i,0})$ preserves $\psi_{i,0} \geq 0$. □

Corollary A1. *If the QP safety filter enforces the estimated-state HOCBF condition in Theorem A1 with $\delta_j(t) \equiv 0$ for every active obstacle and lane row, then the corresponding intersection of risk-tightened obstacle and lane safe sets is forward invariant under the stated assumptions.*

Proof. With zero slack, every enforced row satisfies the hypothesis of Theorem A1. Intersecting finitely many simultaneously enforced forward-invariant barrier conditions gives the joint certificate used by the QP. □

Remark A1. *If any slack variable is active, the strict certificate in Corollary A1 is no longer claimed; the QP is instead interpreted as a least-violation fallback.*

Appendix A.3. Slack-Dependent Violation Bound

For the scalar HOCBF chain with $\alpha_1(r) = k_1 r$, $\alpha_2(r) = k_2 r$, and $k_1, k_2 > 0$, the relaxed constraint gives

$$\dot{\psi}_{i,1} + k_2 \psi_{i,1} \geq -\delta_i(t).$$

Multiplying by $e^{k_2 t}$ and integrating yields

$$\psi_{i,1}(t) \geq e^{-k_2 t} \psi_{i,1}(0) - \int_0^t e^{-k_2(t-\tau)} \delta_i(\tau)\, d\tau.$$

Since $\dot{\psi}_{i,0} + k_1 \psi_{i,0} = \psi_{i,1}$,

$$\psi_{i,0}(t) \geq e^{-k_1 t} \psi_{i,0}(0) + \int_0^t e^{-k_1(t-\tau)} \left[e^{-k_2 \tau} \psi_{i,1}(0) - \int_0^\tau e^{-k_2(\tau - r)} \delta_i(r)\, dr \right] d\tau.$$

Accumulated slack therefore directly lowers the certified lower bound on the HOCBF margin. Zero slack recovers the forward-invariance certificate; positive slack gives a quantified least-violation interpretation rather than a strict safety guarantee:

$$A_{\delta,i}(t_0, t_1) = \int_{t_0}^{t_1} \delta_i(t)\, dt, \qquad T_{\delta,i}(t_0, t_1) = \int_{t_0}^{t_1} \mathbf{1}\{\delta_i(t) > 0\}\, dt,$$

$$\bar{\delta}_i = \frac{A_{\delta,i}(t_0, t_1)}{T_{\delta,i}(t_0, t_1) + \epsilon}, \qquad A^W_{\delta,i}(t) = \int_{t - T_W}^{t} \delta_i(\tau)\, d\tau.$$

The slack variable is therefore logged as a violation certificate, not as evidence of strict safety. If the sliding-window metric $A^{W}_{\delta,i}(t) > \eta_{\delta}$, the controller should raise a persistent least-violation flag and hand the situation to a higher-level fallback such as emergency braking, minimum-risk maneuver planning, or crash-severity-aware mitigation.

Appendix A.4. Practical Estimation of Robustness Constants

The robust HOCBF proof assumes finite constants on a compact operating domain. In implementation, the ego state and input are restricted to

$$\Omega_x = [v_x^{\min}, v_x^{\max}] \times [v_y^{\min}, v_y^{\max}] \times [r^{\min}, r^{\max}] \times [s^{\min}, s^{\max}] \times [w^{\min}, w^{\max}] \times [\theta^{\min}, \theta^{\max}],$$

$$\mathcal{U} = [a^{\min}, a^{\max}] \times [\delta_f^{\min}, \delta_f^{\max}].$$

For obstacle row i, define the second-order HOCBF residual:

$$F_i(x_e, u, t) = \frac{\partial \psi_{i,1}}{\partial t} + L_{f_e}\psi_{i,1} + L_{g_e}\psi_{i,1}u.$$

The constants can then be bounded by

$$L^{x}_{\rho,i} = \sup_{x_e \in \Omega_x} \|\nabla_{x_e}\rho_i(x_e, t)\|, \qquad L_{\psi_1,i} = \sup_{x_e \in \Omega_x} \|\nabla_{x_e}\psi_{i,1}(x_e, t)\|,$$

$$B_{\psi_1,i} = \sup_{x_e \in \Omega_x} \|\nabla_{x_e}\psi_{i,1}(x_e, t)\|, \qquad K_i = \sup_{\substack{x_e \in \Omega_x,\ u \in \mathcal{U},\\ t \in [t_k, t_{k+1}]}} \|\nabla_{x_e}F_i(x_e, u, t)\|.$$

The same sampled gradient envelope is used for $L_{\psi_1,i}$ and $B_{\psi_1,i}$, with $L_{\psi_1,i}$ entering the α_2-composition bound and $B_{\psi_1,i}$ bounding the additive disturbance term. If $\alpha_2(r) = k_2 r$, then $L_{\alpha_2,i} = k_2$; for a nonlinear class-$\mathcal{K}$ function,

$$L_{\alpha_2,i} = \sup_{r \in [\psi_{1,i}^{\min}, \psi_{1,i}^{\max}]} |\alpha_2'(r)|.$$

In the reported implementation, these constants are computed on the compact simulation domain. The required gradients are evaluated by analytic differentiation or automatic differentiation on a tensor grid over Ω_x, $\mathcal{U}$, and the sampling interval. The maximum sampled value is multiplied by a safety factor $\gamma_{\text{Lip}} > 1$:

$$\widehat{L}^{x}_{\rho,i} = \gamma_{\text{Lip}} \max_{m} \left\| \nabla_{x_e}\rho_i(x_e^{(m)}, t) \right\|,$$

$$\widehat{K}_i = \gamma_{\text{Lip}} \max_{m,n,\ell} \left\| \nabla_{x_e}F_i(x_e^{(m)}, u^{(n)}, t^{(\ell)}) \right\|, \qquad \widehat{B}_{\psi_1,i} = \gamma_{\text{Lip}} \max_{m} \left\| \nabla_{x_e}\psi_{i,1}(x_e^{(m)}, t) \right\|.$$

All reported experiments use $\gamma_{\text{Lip}} = 1.2$. Table A1 summarizes the practical estimation of the robustness constants used in the robust HOCBF margin.

Table A1. Practical estimation of robustness constants used in the robust HOCBF margin.

Constant	Definition	Implementation Estimate
$L^{x}_{\rho,i}$	$\sup \|\nabla_{x_e}\rho_i\|$	Analytic/AD grid maximum on Ω_x times γ_{Lip}
$L_{\psi_1,i}$	$\sup \|\nabla_{x_e}\psi_{i,1}\|$	Sampled gradient envelope times γ_{Lip}
$B_{\psi_1,i}$	$\sup \|\nabla_{x_e}\psi_{i,1}\|$	Same envelope, used for additive disturbance robustness
K_i	$\sup \|\nabla_{x_e}F_i\|$ over $\Omega_x \times \mathcal{U} \times [t_k, t_{k+1}]$	Analytic/AD grid maximum times γ_{Lip}
$L_{\alpha_2,i}$	$\sup \|\alpha_2'(r)\|$	k_2 for the linear $\alpha_2(r) = k_2 r$ used in the simulations

References

1. Xiao, W.; Belta, C. High-Order Control Barrier Functions. *IEEE Trans. Autom. Control* **2022**, *67*, 3655–3662. [CrossRef]
2. Xiong, Y.; Zhai, D.H.; Tavakoli, M.; Xia, Y. Discrete-Time Control Barrier Function: High-Order Case and Adaptive Case. *IEEE Trans. Cybern.* **2023**, *53*, 3231–3239. [CrossRef] [PubMed]
3. Wang, H.; Peng, J.; Zhang, F.; Wang, Y. High-Order Control Barrier Function-Based Safety Control of Constrained Robotic Systems: An Augmented Dynamics Approach. *IEEE/CAA J. Autom. Sin.* **2024**, *11*, 2487–2496. [CrossRef]
4. Cosner, R.K.; Singletary, A.W.; Taylor, A.J.; Molnar, T.G.; Bouman, K.L.; Ames, A.D. Measurement-Robust Control Barrier Functions: Certainty in Safety with Uncertainty in State. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 6286–6291. [CrossRef]
5. Oruganti, P.S.; Naghizadeh, P.; Ahmed, Q. Safe Control Using High-Order Measurement Robust Control Barrier Functions. In Proceedings of the 2023 American Control Conference (ACC), San Diego, CA, USA, 31 May–2 June 2023; pp. 4148–4154. [CrossRef]
6. Alan, A.; Taylor, A.J.; He, C.R.; Orosz, G.; Ames, A.D. Safe Controller Synthesis with Tunable Input-to-State Safe Control Barrier Functions. *IEEE Control Syst. Lett.* **2022**, *6*, 908–913. [CrossRef]
7. Alan, A.; Taylor, A.J.; He, C.R.; Ames, A.D.; Orosz, G. Control Barrier Functions and Input-to-State Safety with Application to Automated Vehicles. *IEEE Trans. Control Syst. Technol.* **2023**, *31*, 2744–2759. [CrossRef]
8. Alan, A.; Molnar, T.G.; Das, E.; Ames, A.D.; Orosz, G. Disturbance Observers for Robust Safety-Critical Control with Control Barrier Functions. *IEEE Control Syst. Lett.* **2023**, *7*, 1123–1128. [CrossRef]
9. Alan, A.; Molnar, T.G.; Ames, A.D.; Orosz, G. Parameterized Barrier Functions to Guarantee Safety Under Uncertainty. *IEEE Control Syst. Lett.* **2023**, *7*, 2077–2082. [CrossRef]
10. Agrawal, D.R.; Panagou, D. Safe and Robust Observer-Controller Synthesis Using Control Barrier Functions. *IEEE Control Syst. Lett.* **2023**, *7*, 127–132. [CrossRef]
11. Clark, A. Control Barrier Functions for Stochastic Systems. *Automatica* **2021**, *130*, 109688. [CrossRef]
12. Nishimura, Y.; Hoshino, K. Control Barrier Functions for Stochastic Systems and Safety-Critical Control Designs. *IEEE Trans. Autom. Control* **2024**, *69*, 8088–8095. [CrossRef]
13. Vahs, M.; Pek, C.; Tumova, J. Belief Control Barrier Functions for Risk-Aware Control. *IEEE Robot. Autom. Lett.* **2023**, *8*, 8565–8572. [CrossRef]
14. Ahmadi, M.; Xiong, X.; Ames, A.D. Risk-Averse Control via CVaR Barrier Functions: Application to Bipedal Robot Locomotion. *IEEE Control Syst. Lett.* **2022**, *6*, 878–883. [CrossRef]
15. Kishida, M. Risk-Aware Control: Integrating Worst-Case Conditional Value-at-Risk with Control Barrier Function. *IET Control Theory Appl.* **2025**. [CrossRef]
16. Li, M.; Sun, Z.; Liao, Z.; Weiland, S. Moving Obstacle Collision Avoidance via Chance-Constrained MPC with CBF. *arXiv* **2023**, arXiv:2304.01639. [CrossRef]
17. Mustafa, K.A.; Ornia, D.J.; Kober, J.; Alonso-Mora, J. RACP: Risk-Aware Contingency Planning with Multi-Modal Predictions. *IEEE Trans. Intell. Veh.* **2025**, *10*, 228–243. [CrossRef]
18. Autenrieb, J.; Shin, H.S. Sensor-Based Safety-Critical Control Using an Incremental Control Barrier Function Formulation via Reduced-Order Approximate Models. In Proceedings of the 2025 IEEE American Control Conference (ACC), Denver, CO, USA, 8–10 July 2025; pp. 374–381.
19. Hu, C.; Wang, J. Trust-Based and Individualizable Adaptive Cruise Control Using Control Barrier Function Approach with Prescribed Performance. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 6974–6984. [CrossRef]
20. He, S.; Zeng, J.; Zhang, B.; Sreenath, K. Rule-Based Safety-Critical Control Design Using Control Barrier Functions with Application to Autonomous Lane Change. In Proceedings of the 2021 American Control Conference (ACC), Virtually, 25–28 May 2021; pp. 178–185. [CrossRef]
21. Abduljabbar, M.; Meskin, N.; Cassandras, C.G. Control Barrier Function-based Lateral Control of Autonomous Vehicle for Roundabout Crossing. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; pp. 859–864. [CrossRef]
22. Zhang, C.; Dai, L.; Zhang, H.; Wang, Z. Control Barrier Function-Guided Deep Reinforcement Learning for Decision-Making of Autonomous Vehicle at On-Ramp Merging. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 8919–8932. [CrossRef]
23. Dean, S.; Taylor, A.; Cosner, R.; Recht, B.; Ames, A. Guaranteeing Safety of Learned Perception Modules via Measurement-Robust Control Barrier Functions. In *Proceedings of the 2020 Conference on Robot Learning, Virtual, 16–18 November 2020*; PMLR: New York, NY, USA, 2021; Volume 155, pp. 654–670.
24. Bansal, A.; Kim, H.; Yu, S.; Li, B.; Hovakimyan, N.; Caccamo, M.; Sha, L. Perception Simplex: Verifiable Collision Avoidance in Autonomous Vehicles Amidst Obstacle Detection Faults. *Softw. Test. Verif. Reliab.* **2024**, *34*, e1879. [CrossRef]
25. Hou, W.; Li, W.; Li, P. Fault Diagnosis of the Autonomous Driving Perception System Based on Information Fusion. *Sensors* **2023**, *23*, 5110. [CrossRef]

26. Rockafellar, R.T.; Uryasev, S. Optimization of Conditional Value-at-Risk. *J. Risk* **2000**, *2*, 21–41. [CrossRef]
27. Norton, M.; Khokhlov, V.; Uryasev, S. Calculating CVaR and bPOE for Common Probability Distributions with Application to Portfolio Optimization and Density Estimation. *Ann. Oper. Res.* **2021**, *299*, 1281–1315. [CrossRef]
28. Wiseman, Y.; Grinberg, I. Circumspectly Crash of Autonomous Vehicles. In Proceedings of the 2016 IEEE International Conference on Electro Information Technology (EIT), Grand Forks, ND, USA, 19–21 May 2016; pp. 387–392.

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.