Uncertainty-Aware Control Barrier Functions with Self-Adaptive Online Conformal Prediction for Safe Reinforcement Learning

[OUSSAMA AKIR]Higher School of Communications of Tunis (Sup'Com)>Higher School of Communication
University of Carthage, Tunis, Tunisia>University of Carthage, Tunis, Tunisia
Tunis, Tunisia>Tunis, Tunisia

[oussama.akir@supcom.tn]

October 21, 2025

Abstract

Safe reinforcement learning (RL) is essential for deploying autonomous systems in safety-critical applications such as energy systems, autonomous vehicles, and medical devices. While RL has demonstrated remarkable success, ensuring safety during both training and deployment remains challenging. Existing approaches either provide safety guarantees that ignore model uncertainty or quantify uncertainty without formal safety certificates.

This paper introduces **Uncertainty-Aware Control Barrier Functions (U-CBF)**, a novel framework integrating formal safety guarantees with adaptive uncertainty quantification for safe RL. Our approach combines ensemble-based uncertainty quantification, Self-Adaptive Online Conformal Prediction (SAOCP) providing distribution-free uncertainty bounds with guaranteed coverage, and uncertainty-aware CBF constraints for provably safe action selection.

We validate U-CBF on a microgrid energy management benchmark, demonstrating superior safety (99.2% constraint violation rate vs. 95.8% for state-of-the-art baselines) while maintaining competitive performance. Crucially, we deploy U-CBF on real microgrid hardware, achieving 98.7% safety over 6-hour experiments with zero emergency stops, verified IEEE 1547 compliance, and quantified economic benefits ($840/year savings, 2.3-year payback period).

The main contributions include: (1) theoretical analysis proving U-CBF achieves forward invariance with probability $(1 - \alpha)$ under SAOCP-calibrated uncertainty, (2) SAOCP algorithm for adaptive online conformal prediction maintaining coverage under distribution shift, (3) comprehensive empirical evaluation demonstrating 80% reduction in safety violations, and (4) real-world hardware validation with regulatory compliance verification. This work establishes a principled framework for uncertainty-aware safety in RL, bridging formal verification and practical deployability.

Keywords: Safe Reinforcement Learning, Control Barrier Functions, Conformal Prediction, Uncertainty Quantification, Energy Systems, Hardware Validation

ArXiv Categories:

- **Primary**: cs.LG (Machine Learning)
- **Secondary**: cs.SY (Systems and Control), cs.RO (Robotics), math.OC (Optimization and Control)

1 Introduction

Safe reinforcement learning (RL) is critical for deploying autonomous systems in safety-critical domains such as energy management [18], autonomous vehicles [12], and medical devices [27]. While RL has achieved remarkable success in complex decision-making tasks [15, 21], ensuring safety during both training and deployment remains a fundamental challenge. Existing approaches either provide safety guarantees that ignore model uncertainty [4] or quantify uncertainty without formal safety certificates [8].

1.1 Motivation

Control Barrier Functions (CBFs) [4] offer provable safety guarantees by maintaining system states within safe sets through real-time constraint enforcement. However, standard CBF formulations assume perfect knowledge of system dynamics—an assumption violated in most real-world applications. When model errors exceed expected bounds, CBF constraints may fail, leading to safety violations despite theoretical guarantees.

Conversely, uncertainty quantification methods such as Bayesian neural networks [9] and ensemble models [13] capture epistemic uncertainty but lack mechanisms to translate these estimates into rigorous safety guarantees. Recent work on conformal prediction [20, 26] provides distribution-free uncertainty bounds with finite-sample coverage guarantees. However, fixed confidence levels degrade under distribution shift—a common occurrence in RL due to policy updates and environmental non-stationarity.

1.2 Our Contribution

This paper introduces **Uncertainty-Aware Control Barrier Functions (U-CBF)**, a novel framework integrating formal safety guarantees with adaptive uncertainty quantification for safe RL. Our approach combines:

1. **Ensemble-based uncertainty quantification** capturing epistemic model uncertainty through neural network ensembles
2. **Self-Adaptive Online Conformal Prediction (SAOCP)** providing distribution-free uncertainty bounds with guaranteed coverage that automatically adapt to distribution shift
3. **Uncertainty-aware CBF constraints** integrating calibrated uncertainty into barrier function conditions for provably safe action selection

We validate U-CBF on a microgrid energy management benchmark, demonstrating superior safety (99.2% constraint violation rate vs. 95.8% for state-of-the-art baselines) while maintaining competitive performance. Crucially, we deploy U-CBF on real microgrid hardware, achieving 98.7% safety over 6-hour experiments with zero emergency stops and verified IEEE 1547 compliance [11].

Main contributions:

- Theoretical analysis proving U-CBF achieves forward invariance with probability $(1 - \alpha)$ under SAOCP-calibrated uncertainty
- SAOCP algorithm for adaptive online conformal prediction maintaining coverage guarantees under distribution shift
- Comprehensive empirical evaluation demonstrating 80% reduction in safety violations
- Real-world hardware validation with regulatory compliance and economic impact quantification
- Open-source implementation with reproducibility package

2 Related Work

2.1 Safe Reinforcement Learning

Safe RL methods can be categorized into constraint optimization [1, 23], shielding approaches [3], and worst-case robust policies [16]. Constrained Policy Optimization (CPO) [1] extends TRPO with safety constraints but lacks formal guarantees. Recovery RL [24] learns backup policies but requires extensive exploration. Our approach differs by providing formal safety certificates through CBFs while adapting to model uncertainty.

2.2 Control Barrier Functions

CBFs [4, 5] provide forward invariance guarantees for safe sets. Recent work integrates CBFs with learning-based control [7, 22]. However, these methods assume known dynamics or fixed uncertainty bounds. Robust CBFs [2] handle bounded disturbances but require conservative bounds that degrade performance. We address model uncertainty through adaptive conformal prediction.

2.3 Uncertainty Quantification in RL

Ensemble methods [13, 8] and Bayesian approaches [9] quantify epistemic uncertainty. Model-based RL leverages uncertainty for exploration [8] and safe policy optimization [6]. However, these methods lack finite-sample guarantees. Conformal prediction [20] provides distribution-free coverage but typically assumes i.i.d. data—violated in RL due to policy-induced distribution shift.

2.4 Conformal Prediction

Split conformal prediction [14] achieves exact coverage under exchangeability. Recent extensions handle distribution shift through weighted conformal prediction [25] and adaptive methods [10]. Our SAOCP algorithm extends these ideas to online settings with policy-induced non-stationarity, maintaining coverage through adaptive multiplier updates.

3 Problem Formulation

3.1 Markov Decision Process

We consider an infinite-horizon discounted MDP defined by tuple $(\mathcal{S}, \mathcal{A}, f, r, \gamma, s_0)$ where $\mathcal{S} \subset \mathbb{R}^{n_s}$ is the state space, $\mathcal{A} \subset \mathbb{R}^{n_a}$ is the action space, $f : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ is the (unknown) dynamics function, $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the reward function, $\gamma \in (0, 1)$ is the discount factor, and s_0 is the initial state distribution.

3.2 Safety Constraints

We define the safe set $\mathcal{C} \subset \mathcal{S}$ through barrier function $h : \mathcal{S} \to \mathbb{R}$:

$$\mathcal{C} = \{s \in \mathcal{S} \mid h(s) \geq 0\} \tag{1}$$

The goal is to learn policy $\pi : \mathcal{S} \to \mathcal{A}$ maximizing expected return while ensuring $s_t \in \mathcal{C}$ for all $t \geq 0$ with high probability.

3.3 Control Barrier Functions

A continuous function h is a Control Barrier Function (CBF) if there exists class-$\mathcal{K}$ function α such that for all $s \in \mathcal{C}$:

$$\sup_{a \in \mathcal{A}} [h(f(s,a)) - h(s) + \alpha(h(s))] \geq 0 \tag{2}$$

This condition ensures forward invariance: if $s_0 \in \mathcal{C}$, then $s_t \in \mathcal{C}$ for all $t \geq 0$.

Challenge: CBF condition assumes exact knowledge of dynamics f. With learned model $\hat{f}$, prediction errors can violate safety guarantees.

4 Methodology

4.1 Ensemble Dynamics Model

We learn ensemble of M neural networks $\{\hat{f}_i\}_{i=1}^{M}$ modeling dynamics. For state-action pair (s, a), ensemble predictions yield:

$$\hat{s}' = \frac{1}{M} \sum_{i=1}^{M} \hat{f}_i(s,a) \tag{3}$$

Empirical variance captures epistemic uncertainty:

$$\sigma^2(s,a) = \frac{1}{M} \sum_{i=1}^{M} \|\hat{f}_i(s,a) - \hat{s}'\|^2 \tag{4}$$

4.2 Self-Adaptive Online Conformal Prediction

Standard conformal prediction uses fixed quantile of calibration errors. Under distribution shift, coverage degrades. We propose SAOCP maintaining target coverage $(1-\alpha)$ through adaptive multiplier λ_t.

Calibration Set: Maintain buffer $\mathcal{D}_{\text{cal}} = \{(s_i, a_i, s_i')\}_{i=1}^{n_{\text{cal}}}$ of recent transitions.

Nonconformity Scores: For each calibration sample:

$$e_i = \frac{\|s_i' - \hat{s}_i'\|}{\sigma(s_i, a_i)} \tag{5}$$

Adaptive Quantile: Compute $(1-\alpha)$-quantile with adaptive multiplier:

$$q_t = \text{Quantile}(\{e_i\}_{i=1}^{n_{\text{cal}}}, 1-\alpha) \cdot \lambda_t \tag{6}$$

Multiplier Update: Track empirical coverage and update:

$$\lambda_{t+1} = \lambda_t \cdot (1 + \eta(\text{coverage}_t - (1-\alpha))) \tag{7}$$

where $\eta > 0$ is learning rate and coverage_t is fraction of recent predictions within bounds.

Uncertainty Bounds: For new prediction at (s, a):

$$\|\hat{f}(s,a) - f(s,a)\| \leq q_t \cdot \sigma(s,a) \quad \text{w.p. } \geq 1-\alpha \tag{8}$$

4.3 Uncertainty-Aware CBF

Integrate calibrated uncertainty into CBF constraint. For proposed action a:

$$h(\hat{f}(s,a) - q_t \sigma(s,a)\mathbf{1}) - h(s) + \alpha(h(s)) \geq 0 \tag{9}$$

where $\mathbf{1}$ is worst-case direction (gradient of h).

Safe Action Projection: If policy action a_π violates U-CBF, project to safe action via quadratic program:

$$a^* = \arg\min_{a \in \mathcal{A}} \quad \|a - a_\pi\|^2 \tag{10}$$

$$\text{s.t.} \quad h(\hat{f}(s,a) - q_t \sigma(s,a) \nabla h) - h(s) + \alpha(h(s)) \geq 0 \tag{11}$$

This ensures safety under calibrated uncertainty bounds.

4.4 Theoretical Guarantee

Theorem 1 (Safety Guarantee)**.** *Under SAOCP-calibrated uncertainty with coverage $\geq 1-\alpha$, U-CBF constraint ensures forward invariance of safe set $\mathcal{C}$ with probability $\geq 1-\alpha$.*

Sketch. SAOCP guarantees prediction error within $q_t \sigma(s,a)$ with probability $\geq 1-\alpha$. U-CBF constraint accounts for this uncertainty in worst-case direction. By CBF theory, forward invariance holds under exact dynamics. Since true dynamics lie within bounds w.p. $\geq 1-\alpha$, forward invariance holds w.p. $\geq 1-\alpha$. Full proof in supplementary material. □

5 Experimental Setup

5.1 Microgrid Energy Management

We evaluate on microgrid control problem with renewable energy and hydrogen storage:

State ($n_s = 6$): Battery SoC, H_2 tank level, water reservoir, heat storage, grid voltage, temperature

Actions ($n_a = 2$): Electrolyzer power (0-5kW), fuel cell power (0-3kW)

Safety Constraints: SoC $\in [20\%, 80\%]$, voltage $\in [200V, 250V]$, temperature $\in [10C, 80C]$

Objective: Minimize energy costs while maintaining safety and responding to variable renewable generation

5.2 Baselines

We compare against 5 methods:

- **PPO** [19]: Vanilla RL without safety
- **CPO** [1]: Constrained policy optimization
- **RCPO** [23]: Reward-constrained policy optimization
- **Safe RL** [17]: Safety layer with fixed bounds
- **CBF-RL** [7]: CBF with fixed uncertainty

5.3 Implementation Details

Ensemble: $M = 5$ networks, 2 hidden layers (256 units), ReLU activation

SAOCP: $\alpha = 0.05$, $\eta = 0.01$, $n_{\text{cal}} = 1000$, $\lambda_0 = 1.0$

CBF: $\alpha(h) = h$ (linear class-$\mathcal{K}$)

RL: PPO with learning rate 3e-4, batch size 256, 1M timesteps

All experiments run with 10 random seeds. Hardware experiments use real microgrid testbed (10kWh battery, 5kW electrolyzer, 3kW fuel cell).

6 Results

6.1 Simulation Performance

Table 1 shows performance comparison. U-CBF achieves 99.2% CVR compared to 95.8% for best baseline (CBF-RL), representing 80% reduction in violations. Return remains competitive at 412.3 vs. 425.1 for unconstrained PPO.

6.2 SAOCP Calibration Quality

Figure **??** shows SAOCP maintains 95% coverage throughout training while fixed confidence degrades to 78% due to distribution shift. Adaptive multiplier λ_t adjusts automatically, increasing when coverage drops and decreasing when over-conservative.

Table 1: Performance comparison (mean $\pm$ std over 10 seeds)

Method	CVR (%) ↑	Return ↑
PPO	87.3 ± 3.2	**425.1** $\pm$ 12.4
CPO	91.5 ± 2.8	398.7 ± 15.3
RCPO	92.8 ± 2.1	405.2 ± 11.8
Safe RL	94.1 ± 2.5	387.5 ± 14.7
CBF-RL	95.8 ± 1.9	401.3 ± 13.2
U-CBF (Ours)	**99.2** $\pm$ 0.4	412.3 ± 9.8

6.3 Hardware Validation

Real hardware experiments over 6 hours (21,600 timesteps at 1 Hz) demonstrate:

- **Safety**: 98.7% CVR, 0 emergency stops
- **Reliability**: MTBF > 6 hours vs. 2.1 hours for baseline
- **Compliance**: IEEE 1547 PASS (voltage regulation within 5%)
- **Economics**: \$840/year savings, 2.3-year payback, \$4,250 10-year NPV

Hardware performance closely matches simulation (98.7% vs. 99.2%), validating sim-to-real transfer.

6.4 Ablation Studies

Component ablation demonstrates:

- Removing SAOCP (fixed confidence): CVR drops to 96.8%
- Removing ensemble (single model): CVR drops to 94.5%
- Removing CBF (soft penalties): CVR drops to 92.1%

All components contribute to final performance.

7 Discussion

7.1 Key Insights

Adaptive calibration is essential: Fixed confidence intervals degrade under distribution shift common in RL. SAOCP's adaptive multiplier maintains coverage automatically.

Hardware validation confirms practicality: Close sim-to-real performance (99.2% $\rightarrow$ 98.7%) and zero

emergency stops over 6 hours demonstrate real-world viability.

Economic justification: $840/year savings with 2.3-year payback provides concrete business case beyond academic metrics.

7.2 Limitations

Computational cost: Ensemble model and SAOCP add 15-20% overhead vs. single model. Acceptable for 1-10 Hz control but may challenge faster systems.

Hyperparameter sensitivity: SAOCP learning rate η requires tuning. Too high causes oscillation, too low prevents adaptation.

Multi-constraint extension: Current formulation handles single barrier function. Multiple constraints require careful composition.

7.3 Future Work

Theoretical extensions: Tighter bounds on probability of safety, analysis of adaptive convergence rates

Multi-agent safety: Extend U-CBF to multi-agent systems with coupled constraints

Other domains: Autonomous driving, robotics, chemical process control

8 Conclusion

We introduced Uncertainty-Aware Control Barrier Functions (U-CBF), integrating formal safety guarantees with adaptive uncertainty quantification for safe reinforcement learning. Our Self-Adaptive Online Conformal Prediction (SAOCP) algorithm maintains calibrated uncertainty bounds under distribution shift, enabling U-CBF to achieve 99.2% safety in simulation and 98.7% on real hardware with zero emergency stops.

Key contributions include theoretical safety guarantees, 80% reduction in violations vs. baselines, real-world hardware validation with regulatory compliance, and economic justification ($840/year savings, 2.3-year payback). This work establishes a principled framework for uncertainty-aware safety in RL, bridging formal verification and practical deployability.

Reproducibility: Code, data, and pretrained models available at `https://github.com/[username]/ucbf-safe-rl`

References

[1] Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. Constrained policy optimization. In *International conference on machine learning*, pages 22–31. PMLR, 2017.

[2] Anayo K Alan, Andrew J Taylor, Chaozhe R He, Aaron D Ames, and Gábor Orosz. Control barrier functions for stochastic systems. *Automatica*, 130:109688, 2021.

[3] Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[4] Aaron D Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control barrier functions: Theory and applications. *2019 18th European Control Conference (ECC)*, pages 3420–3431, 2019.

[5] Aaron D Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control barrier functions: Theory and applications. In *2019 18th European control conference (ECC)*, pages 3420–3431. IEEE, 2019.

[6] Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. In *Advances in neural information processing systems*, pages 908–918, 2017.

[7] Richard Cheng, Gábor Orosz, Richard M Murray, and Joel W Burdick. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3387–3395, 2019.

[8] Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In *Advances in neural information processing systems*, pages 4754–4765, 2018.

[9] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *international conference on machine learning*, pages 1050–1059, 2016.

[10] Isaac Gibbs and Emmanuel Candès. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

[11] IEEE Standards Association. Ieee standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces, 2018. IEEE Std 1547-2018.

[12] B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):4909–4926, 2021.

[13] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in neural information processing systems*, pages 6402–6413, 2017.

[14] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

[15] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[16] Lerrel Pinto, James Davidson, Rahul Sukthankar, and Abhinav Gupta. Robust adversarial reinforcement learning. In *International Conference on Machine Learning*, pages 2817–2826. PMLR, 2017.

[17] Alex Ray, Joshua Achiam, and Dario Amodei. Benchmarking safe exploration in deep reinforcement learning. *arXiv preprint arXiv:1910.01708*, 2019.

[18] Hao Ren, Dongdong Xu, and Chen Chen. A comprehensive survey of safe reinforcement learning for energy management. *IEEE Transactions on Smart Grid*, 13(4):2829–2840, 2022.

[19] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[20] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

[21] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

[22] Andrew J Taylor, Victor D Dorobantu, Meera Krishnamoorthy, Hoang M Le, Yisong Yue, and Aaron D Ames. Learning for safety-critical control with control barrier functions. In *Learning for Dynamics and Control*, pages 708–717. PMLR, 2020.

[23] Chen Tessler, Daniel J Mankowitz, and Shie Mannor. Reward constrained policy optimization. In *International Conference on Learning Representations*, 2018.

[24] Brijen Thananjeyan, Ashwin Balakrishna, Suraj Nair, Michael Luo, Krishnan Srinivasan, Minho Hwang, Joseph E Gonzalez, Julian Ibarz, Chelsea Finn, and Ken Goldberg. Recovery rl: Safe reinforcement learning with learned recovery zones. In *IEEE Robotics and Automation Letters*, volume 6, pages 4915–4922. IEEE, 2021.

[25] Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in Neural Information Processing Systems*, 32, 2019.

[26] Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer Science & Business Media, 2005.

[27] Chao Yu, Jiming Liu, and Shamim Nemati. Reinforcement learning in healthcare: A survey. *ACM Computing Surveys*, 55(1):1–36, 2021.