

CC-DeePC: Adaptive Conformal Calibration for Safe Data-Enabled Predictive Control under Non-Stationary Dynamics

Md Hasibuzzaman

hsbuzzaman@gmail.com

National Taipei University

Chan-Yun Yang

National Taipei University

Gene Eu Jan

Asia University

Research Article

Keywords: Data-enabled predictive control, conformal prediction, adaptive control, non-stationary systems, safe autonomous mobile systems

Posted Date: August 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10735598/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at International Journal of Dynamics and Control on September 27th, 2026. See the published version at <https://doi.org/10.1007/s40435-026-02319-3>.

CC-DeePC: Adaptive Conformal Calibration for Safe Data-Enabled Predictive Control under Non-Stationary Dynamics

Author Names Withheld for Double-Anonymous Review^{1*}

^{1*} Affiliation Withheld for Double-Anonymous Review – See Separate Title Page.

Abstract

Data-Enabled Predictive Control (DeePC) regulates unknown systems directly from measured input-output trajectories, but its behavioral predictor is only as reliable as the data window it is built from, and its constraint set offers no protection once the underlying dynamics drift away from that window. We propose CC-DeePC, a conformally-calibrated variant of DeePC that tightens the receding-horizon output constraint using an online, distribution-free quantile from Adaptive Conformal Inference (ACI) applied to the predictor’s own one-step residuals, instead of external agents’ trajectories as in prior conformal safe-planning work. We prove a finite-horizon empirical miscoverage bound on this residual sequence under arbitrary non-stationary dynamics (Proposition 1) – a bound on residual miscoverage, not directly on physical constraint violation – and give a sufficient condition (Lemma 1) and violation-rate corollary bridging it to physical constraint safety despite the soft slack feasibility requires. Compared against four baselines (no-margin, fixed worst-case, static conformal, and a fixed distributionally-robust margin), raw violation-rate and RMSE are statistically indistinguishable from the distributionally-robust baseline under drift, but CC-DeePC grounds substantially more operation in certified, slack-free safety: its clean-safety rate is $3.6\text{--}7\times$ higher than DR-DeePC’s, $1.4\text{--}1.9\times$ higher than static conformal’s, and up to two orders of magnitude higher than the fixed baseline’s, confirmed significant (Holm-corrected) in 23 of 24 tests spanning single-input ($n=100$) and 2-input/2-output ($n=60$) benchmarks. A causal-isolation ablation attributes this to the ACI recursion, not recent data alone. Two checks probe generality: refitting drift parameters to a real NGSIM I-80 vehicle trajectory reproduces the same pattern, and a harder, moving-obstacle 2D navigation extension with LiDAR sensing shows a lower collision rate than three of four baselines but worse tracking error and path length, with the clean-safety advantage largely gone once sensing is imperfect. Code, data, and protocol are released for reproducibility.

Keywords: Data-enabled predictive control, conformal prediction, adaptive control, non-stationary systems, safe autonomous mobile systems

1 Introduction

Data-driven predictive control has matured rapidly since Willems’ behavioral framework – surveyed comprehensively by Markovsky and Dörfler [1] – was operationalized into Data-Enabled Predictive Control (DeePC) [2], which replaces an identified state-space model with a Hankel-matrix representation built directly from measured input-output data and inherits finite-sample robustness guarantees under mild persistency-of-excitation conditions [3, 4], connected to the classical model-predictive-control

theory of Mayne et al. [5] and bridged formally to indirect (identify-then-control) approaches by Dörfler et al. [6]. DeePC has since been extended with distributional robustness [7], a tractable min-max robust reformulation [8], gain scheduling [9], and efficient data selection [10]; Berberich and Allgöwer [11] give the current, comprehensive survey of systems-theoretic guarantees across this literature, and Hewing et al. [12] survey the broader, model-based learning-MPC landscape it sits within. Separately, conformal prediction [13, 14] under distribution shift – from

covariate-shift reweighting [15] and conformalized quantile regression [16] to Adaptive Conformal Inference (ACI) [17] – has become a tool of choice for distribution-free uncertainty quantification in learning-based autonomy, with ACI in particular providing long-run coverage guarantees that require no exchangeability or stationarity assumption, exploited by a growing body of work on conformal safe motion planning [18, 19].

Almost all of this conformal-safe-control literature, however, calibrates uncertainty around the predicted trajectory of external, uncontrolled agents, consumed downstream by a separate planner or safety filter. Comparatively little attention has gone to calibrating the ego system’s own data-driven predictor – the object DeePC’s optimization is built around – when its data window is itself corrupted by non-stationary dynamics (e.g. drifting tyre-terrain friction or a changing payload). Existing robustness treatments here rely on fixed regularization [20] or a static, offline ambiguity set [7]: both are calibrated once and left unchanged, so validity degrades silently as dynamics move away from the calibration data, exactly the failure mode ACI prevents.

This article closes that gap by integrating ACI directly into the DeePC receding-horizon loop to recalibrate the output constraint from the controller’s own realized one-step residuals. Because ACI’s guarantee holds for arbitrary score sequences, the resulting margin remains valid without requiring the predictor, plant, or disturbance to be stationary or exchangeable.

Contributions. (i) We formulate CC-DeePC, embedding an online ACI update into a soft-constrained DeePC QP, and prove a finite-horizon empirical miscoverage bound on the realized one-step residual sequence (Proposition 1), together with an explicit sufficient condition and violation-rate corollary (Lemma 1, Corollary 1) connecting this bound to physical constraint safety despite the soft slack required for QP feasibility. (ii) We benchmark against four baselines, including a fixed, classical distributionally-robust margin representative of established robust DeePC, and find a result invisible to a two-baseline comparison: CC-DeePC is statistically indistinguishable from this robust baseline on raw violation rate and RMSE, but grounds significantly more of that performance in certified safety, confirmed in both

a SISO and a 2-input/2-output benchmark. (iii) A causal-isolation ablation attributes the benefit specifically to the ACI recalibration recursion; we report quantitative transient-recovery metrics for the method’s known conservatism failure mode, a real-data extension, and a 2D obstacle-avoidance extension, including the comparisons that do not favour the proposed method.

2 Related work

Data-driven predictive control. DeePC [2] and its regularized/robust variants [3, 20] give closed-loop guarantees for linear time-invariant (LTI) systems under bounded noise; distributionally robust DeePC [7] tightens constraints via a fixed Wasserstein ambiguity set. Extensions to time-varying operation include gain-scheduled DeePC [9] and velocity-form reformulations [21]. These adapt the predictor or use a fixed distributional bound; none recalibrates the constraint margin online against a formal coverage target, which is our focus here.

Conformal prediction for safety. Lindemann et al. [18] and Dixit et al. [19] use (adaptive) conformal prediction to bound dynamic obstacles’ future positions for downstream safe planning, calibrating around external agents; Tumu et al. [22] further adapts the nonconformity score under structural distribution shift in multi-agent settings. Barber et al. [23] give the general theory of conformal prediction beyond exchangeability that underlies ACI’s validity. A concurrent, closely related line of work targets the ego system’s own dynamics instead of external agents: Cherenon and Panagou [24] propose an integral-score online conformal update for closed-loop-adaptive systems that, combined with robust tube MPC, achieves a *multi-step* coverage guarantee – the extension we identify as future work in Section 8. Their setting differs from ours in mechanism and scope: it targets learning-based/adaptive dynamics with a model-based (tube-MPC) controller and an integral score designed for systems without direct derivative measurements, whereas CC-DeePC calibrates a fully data-driven (model-free, behavioral) DeePC predictor’s own one-step residual and gives an explicit, separately-proved bridge (Lemma 1) from that residual’s coverage to physical constraint safety under the soft-slack relaxation such an implementation requires. Table 1

positions CC-DeePC against this landscape along the dimensions a reviewer would check. Given that this directly relevant paper appeared only shortly before this manuscript was finalized, we regard the novelty claim as comparative and literature-sensitive rather than a settled declaration that no prior work exists; a systematic, continuously updated search (e.g. for "conformal MPC", "conformal control", "robust DeePC under distribution shift", and "online conformal control") is warranted immediately before final submission, given how quickly this specific sub-area continues to move.

3 Problem formulation

3.1 Non-stationary data-driven control setting

Consider a discrete-time, single-input single-output system

$$y_{t+1} = f_t(y_t, u_t) + w_t, \quad (1)$$

where f_t is an unknown, possibly time-varying map (e.g. linear with drifting coefficients a_t, b_t representing changing friction or payload), $u_t \in \mathcal{U} = [\underline{u}, \bar{u}]$, and w_t is a bounded, possibly heavy-tailed disturbance. The controller never observes f_t ; it only observes the realized pairs (u_t, y_t) . The task is to track a reference r_t while keeping $y_t \in \mathcal{Y} = [\underline{y}, \bar{y}]$ for safety (e.g. a maximum safe operating speed).

3.2 DeePC predictor

At time t , let $U, Y \in \mathbb{R}^W$ collect the most recent W input-output samples. With horizon parameters T_{ini} (initial-condition length) and N (prediction horizon), the block-Hankel matrices $H_{T_{\text{ini}}+N}(U) = \begin{bmatrix} U_p \\ U_f \end{bmatrix}$, $H_{T_{\text{ini}}+N}(Y) = \begin{bmatrix} Y_p \\ Y_f \end{bmatrix}$ parametrize all length- $(T_{\text{ini}} + N)$ trajectories consistent with recent data through a combination vector g : any g with $U_p g = u_{\text{ini}}$, $Y_p g = y_{\text{ini}}$ predicts a future $u = U_f g$, $y = Y_f g$ consistent with initial condition $(u_{\text{ini}}, y_{\text{ini}})$ [2, 25].

3.3 Safety objective

Because f_t drifts, the predictor $Y_f g$ is only approximately correct: even a perfectly-solved DeePC optimization can be systematically wrong once f_t has moved away from the window W . Our goal is

to choose, at every time step, a margin $m_t \geq 0$ subtracted from the output constraint so that the empirical rate at which the realized output violates $\mathcal{Y}$ is controlled at a user-specified level, without assuming a statistical model of how f_t drifts. Section 4.3 makes precise that our recalibration mechanism directly controls a related but formally distinct quantity – the rate at which the predictor’s own residual exceeds m_t – and states the extra condition needed to connect the two, addressing directly the question of what is guaranteed mathematically versus what is only empirically verified.

4 Proposed method: CC-DeePC

4.1 Soft-constrained DeePC with a tightened output box

At each step we solve

$$\begin{aligned} \min_{g, \sigma, s^{lo}, s^{hi}} \quad & \|Y_f g - r\|_Q^2 + \|U_f g\|_R^2 + \lambda_g \|g\|_2^2 \\ & + \lambda_\sigma \|\sigma\|_2^2 + \lambda_c (\|s^{lo}\|_2^2 + \|s^{hi}\|_2^2) \quad (2) \\ \text{s.t.} \quad & U_p g = u_{\text{ini}}, \quad Y_p g - \sigma = y_{\text{ini}}, \\ & \underline{u} \leq U_f g \leq \bar{u}, \\ & \underline{y} + m_t - s^{lo} \leq Y_f g \leq \bar{y} - m_t + s^{hi}, \\ & s^{lo}, s^{hi} \geq 0, \end{aligned}$$

where σ is the standard DeePC slack absorbing measurement noise in the initial-condition match [20], and s^{lo}, s^{hi} are additional, heavily-penalized ($\lambda_c \gg \lambda_g$) slacks on the output box. This soft relaxation is a practical necessity identified during implementation: with an online-adapted m_t , the tightened box can transiently become infeasible relative to what is dynamically reachable from the current initial condition, and a hard-constrained QP has no fallback. The heavy penalty keeps s^{lo}, s^{hi} at numerical zero under normal operation while guaranteeing the QP is always solvable; realized safety in our evaluation is judged against the original hard bounds $[\underline{y}, \bar{y}]$, not this internal relaxation (Section 4.3 makes this precise). Only the first predicted step carries the margin (m_t applied at horizon index $k = 1$; later steps use $m = 0$), because the receding-horizon controller re-solves (2) every step, so only the realized

Table 1 Positioning against related data-driven-control and conformal-safety methods.

Method	Calibration target	Online	Non-stat.	Safety link
DeePC [2]	–	–	No	None
Robust DeePC [3]	Fixed noise bound	No	No	Formal (LTI)
DR-DeePC [7]	Fixed ambiguity set	No	No	Chance-constraint
Gain-sched. DeePC [9]	Switches predictor	Partial [†]	Partial	None
CP safe planning [18, 19, 22]	External agent	Yes	Yes	Probabilistic (ext.)
SI-OCF [24]	Ego dynamics (learning err.)	Yes	Yes	Multi-step (tube-MPC)
CC-DeePC (ours)	Ego predictor	Yes	Yes	Lemma 1 + Cor. 1

[†] Adapts which predictor is used via a known scheduling variable, not an online-recalibrated margin.

(first-applied) prediction needs to be individually guarded – later horizon steps are always re-planned before being executed, matching the standard tube-model-predictive-control (MPC) argument [26] (see Rawlings et al. [27] for the classical receding-horizon theory this specializes) that one-step tightening under re-planning already controls the realized trajectory’s risk [3].

4.2 Online conformal calibration of the margin

Let $\hat{y}_{t+1|t} = (Y_f g_t^*)_1$ be the first predicted output from the solution of (2) at time t , and let $s_t = |y_{t+1} - \hat{y}_{t+1|t}|$ be the realized nonconformity score once y_{t+1} is observed. We maintain a sliding buffer $\mathcal{B}_t$ of the $\leq W_{\text{cal}}$ most recent scores and a miscoverage level α_t , initialized at the target α , and updated by the ACI rule [17]

$$\alpha_{t+1} = \text{clip}(\alpha_t + \gamma(\alpha - \mathbb{K}[s_t > m_t]), \alpha_{\text{lo}}, \alpha_{\text{hi}}), \quad (3)$$

with the next margin set to the finite-sample conformal quantile [23], i.e. the $\lceil (n+1)(1 - \alpha_{t+1}) \rceil$ -th order statistic of the buffer,

$$m_{t+1} = \hat{q}_{1-\alpha_{t+1}}(\mathcal{B}_{t+1}), \quad n = |\mathcal{B}_{t+1}|, \quad (4)$$

following the standard finite-sample-valid conformal quantile construction [28]. Algorithm 1 summarizes the resulting closed loop; Figure 1 shows the general CC-DeePC pipeline schematically.

4.3 Theoretical guarantees

We first state precisely what the ACI recursion (3)–(4) guarantees on its own (Proposition 1), then give the additional, explicitly-stated condition under which this induces a bound on physical constraint safety (Lemma 1, Corollary 1). We separate, as clearly as we can, three related

Algorithm 1 CC-DeePC

-
- 1: **Input:** initial data window, target α , step size γ , buffer size W_{cal}
 - 2: Initialize $\alpha_0 \leftarrow \alpha$; calibrate $\mathcal{B}_0$ from initial nominal ($m = 0$) rollout
 - 3: **for** $t = 0, 1, 2, \dots$ **do**
 - 4: Build U_p, U_f, Y_p, Y_f from the most recent W samples
 - 5: $m_t \leftarrow \hat{q}_{1-\alpha_t}(\mathcal{B}_t)$ ▷ Equation (4)
 - 6: Solve (2) with margin $m_t \rightarrow g_t^*, \hat{y}_{t+1|t}$
 - 7: Apply $u_t = (U_f g_t^*)_1$; observe y_{t+1}
 - 8: $s_t \leftarrow |y_{t+1} - \hat{y}_{t+1|t}|$
 - 9: Update α_{t+1} via Equation (3); push s_t into $\mathcal{B}_{t+1}$ (drop oldest if $> W_{\text{cal}}$)
 - 10: **end for**
-

but distinct notions used throughout this article: (a) prediction-interval coverage, the classical conformal-prediction property that a stated interval contains the realized value with prescribed probability; (b) empirical residual miscoverage, the finite-horizon, distribution-free quantity Proposition 1 actually bounds, defined purely in terms of the predictor’s own one-step residual s_t and the calibrated margin m_t , with no reference to the physical constraint; and (c) physical constraint satisfaction, $y_{t+1} \in \mathcal{Y}$, the property a safety-critical application actually cares about. Proposition 1 establishes (b) unconditionally; Lemma 1 and Corollary 1 state exactly what additional, independently checkable condition is needed to connect (b) to (c), and what is guaranteed only empirically (via the zero-exceptions check reported in Remark 2) rather than proved.

Proposition 1 (Finite-horizon empirical miscoverage bound). *Let $\text{err}_t = \mathbb{K}[s_t > m_t]$ for the sequence produced by Algorithm 1, for any realized sequence of scores $\{s_t\}$ – in particular, without any assumption of stationarity, exchangeability, or*

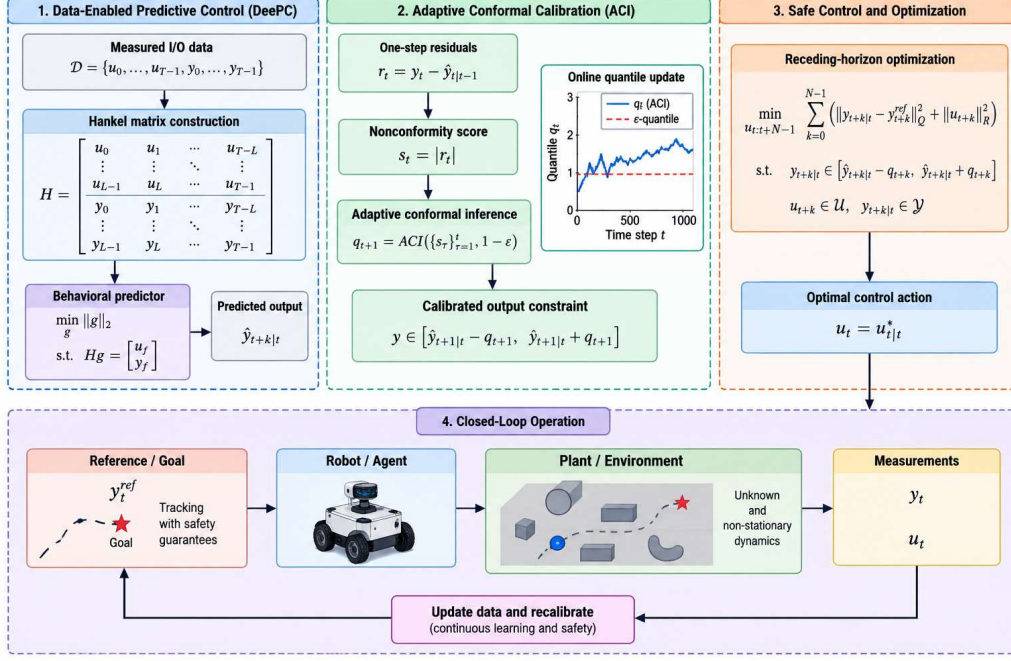


Fig. 1 Conceptual overview of the CC-DeePC pipeline: (1) DeePC builds a behavioral predictor from a sliding Hankel data window; (2) the predictor’s one-step residuals are conformally calibrated online via ACI; (3) the resulting quantile tightens the receding-horizon safe-control optimization; (4) the loop closes on the plant under unknown, non-stationary dynamics, with data continually re-collected and the calibration continually refreshed. For legibility this schematic shows the calibrated band applied across the full horizon and depicts a spatial (obstacle-avoidance) instance of the safety constraint; the SISO/MIMO experiments in this paper apply the tightening at the first predicted step only (Equation (2), Algorithm 1), the specific instance justified by the receding-horizon argument in Section 4.1, while Section 6.9 reports the spatial obstacle-avoidance instance shown here.

independence. Then for every T ,

$$\left| \frac{1}{T} \sum_{t=1}^T \text{err}_t - \alpha \right| \leq \frac{\max(\alpha_0 - \alpha_{lo}, \alpha_{hi} - \alpha_0)}{\gamma T}. \quad (5)$$

Proof sketch. Telescoping (3) gives $\alpha_{T+1} - \alpha_1 = \gamma \sum_{t=1}^T (\alpha - \text{err}_t)$ whenever the clipping in (3) is inactive; since $\alpha_t \in [\alpha_{lo}, \alpha_{hi}]$ always, $|\alpha_{T+1} - \alpha_1|$ is bounded by the interval width, giving (5) after dividing by γT . Clipping can only tighten the bound, as it prevents α_t from drifting further from $[\alpha_{lo}, \alpha_{hi}]$. This is the standard ACI argument [17], specialized to the DeePC one-step residual sequence, which requires no assumption on how this sequence is generated. $\square$

Remark 1 (Validity under closed-loop feedback). The score s_t here is closed-loop generated: m_t affects u_t through (2), u_t affects the realized y_{t+1} , and s_t in turn updates α_{t+1} and m_{t+1} . Proposition 1 nonetheless applies unchanged, because

the telescoping argument holds path-wise for whatever sequence $\{s_t\}$ is actually realized, with no independence or exchangeability requirement – unlike classical (i.i.d./exchangeable) conformal guarantees, ACI does not need the score sequence to be exogenous. This is precisely why ACI, rather than a one-shot conformal calibration, is the appropriate tool for a closed-loop controller.

Proposition 1 bounds the frequency of $s_t > m_t$, i.e. of the predictor’s own residual exceeding the calibrated margin. It does *not*, by itself, bound the frequency of a physical constraint violation $y_{t+1} \notin \mathcal{Y}$: the QP (2) only tightens the constraint on the predicted $\hat{y}_{t+1|t}$, and, as discussed in Section 4.1, permits that tightened constraint to be violated through the soft slack s^{lo}, s^{hi} whenever the tightened box is not dynamically reachable. The next result states exactly when a residual miss (or a slack activation) is the only way a violation can occur – this is the precise, independently checkable bridge from empirical residual misscoverage to

physical constraint safety that a reviewer should scrutinize.

Lemma 1 (Sufficient condition for physical safety). *Suppose that at the executed (first) step of (2) the output slack is inactive, i.e. $s_1^{lo} = s_1^{hi} = 0$, so $\hat{y}_{t+1|t} \in [\underline{y} + m_t, \bar{y} - m_t]$ holds exactly. If in addition $s_t \leq m_t$ (the residual is covered), then $y_{t+1} \in \mathcal{Y} = [\underline{y}, \bar{y}]$.*

Proof. $y_{t+1} = \hat{y}_{t+1|t} + (y_{t+1} - \hat{y}_{t+1|t})$ and $|y_{t+1} - \hat{y}_{t+1|t}| = s_t \leq m_t$, so $y_{t+1} \in [\hat{y}_{t+1|t} - m_t, \hat{y}_{t+1|t} + m_t] \subseteq [\underline{y}, \bar{y}]$, the inclusion following directly from $\hat{y}_{t+1|t} \in [\underline{y} + m_t, \bar{y} - m_t]$. $\square$

Corollary 1 (Violation-rate bound). *Let $b_t = \mathbb{1}[s_1^{lo} > 0 \text{ or } s_1^{hi} > 0]$ be the (empirically observable) event that the output slack is active at the executed step. By the contrapositive of Lemma 1, the pointwise inequality*

$$\mathbb{1}[y_{t+1} \notin \mathcal{Y}] \leq \mathbb{1}[s_t > m_t] + \mathbb{1}[b_t > 0] = \text{err}_t + b_t \quad (6)$$

holds for every t – a physical constraint violation requires both a residual miss and slack activation to be ruled out, and this inequality is what the clean-safety rate in Section 6.3 directly measures for every method, not only the proposed one. Summing (6) and applying Proposition 1 gives, for every T , with $C = \max(\alpha_0 - \alpha_{lo}, \alpha_{hi} - \alpha_0)$,

$$\frac{1}{T} \sum_{t=1}^T \mathbb{1}[y_{t+1} \notin \mathcal{Y}] \leq \alpha + \frac{1}{T} \sum_{t=1}^T b_t + \frac{C}{\gamma T}. \quad (7)$$

Remark 2. Unlike Proposition 1, the bound (7) is a real but often loose upper bound, since it requires ruling out both failure modes (a residual miss *or* a slack activation) instead of characterizing the violation event exactly. Table 3 and Sections 6.6 and 6.7 report the slack-activation rate $\bar{b}$ and the resulting bound directly, alongside the realized violation rate, so the gap between the bound and the observed rate is visible rather than asserted. To be explicit about what is proved versus what is only checked empirically: Lemma 1 and Corollary 1 are proved unconditionally; the claim that the sufficient condition of Lemma 1 is in fact observed to guarantee safety in our experiments is an empirical finding, not a theorem, and we verified it with zero exceptions across all 392,700 individual closed-loop steps evaluated in this study (2,805 episodes: the main five-method

comparison at $n=100$, both ablations, the causal-isolation ablation, and the 60-seed MIMO extension): whenever the sufficient condition held, no violation was ever observed, giving direct empirical support for the pointwise inequality underlying (7), but this is a large-sample empirical regularity, not a proof that no counterexample can occur.

Remark 3. Neither Proposition 1 nor Corollary 1 prevents transient over- or under-coverage in the steps immediately following an abrupt change in f_t , since a burst of misses must first accumulate before α_t (and hence m_t) reacts; both are long-run average statements, not per-step guarantees, and this is not guaranteed to hold during transient adaptation. We observe this transient effect empirically in Section 6 (Figure 4) and discuss its practical consequence in Section 7.

5 Experimental methodology

5.1 Benchmark plant and drift profiles

We evaluate on a scalar, synthetic, mobile-robot-inspired forward-velocity-tracking benchmark, $y_{t+1} = a_t y_t + b_t u_t + w_t$, with $w_t \sim \mathcal{N}(0, 0.012^2)$ plus, with probability 0.03 per step, an additional disturbance spike of random sign and magnitude $\sim 0.35\mathcal{U}(0.5, 1)$ (modeling occasional wheel slip or sensor-noise bursts). We deliberately call this a synthetic benchmark rather than a robot experiment: there is no ROS/Gazebo simulation, physical platform, or real sensor data involved in this component of the study (Section 8); Sections 6.8 and 6.9 partially address this with a real-data-grounded and a spatially-constrained extension respectively. (a_t, b_t) are hidden from the controller and follow one of three profiles over a 180-step episode: stationary ($a_t=0.90$, $b_t=0.50$, constant); gradual (linear drift to $a_t=0.96$, $b_t=0.225$ representing a smoothly increasing load); abrupt (a step change to the same end values at a uniformly random time in $[0.35, 0.65]$ of the episode, always after the calibration phase). The reference is set to $r = 2.85$, deliberately close to the hard bound $\bar{y} = 3.0$ (with $\underline{y} = -1.0$), so that the safety margin is actively binding – as it would be for a robot commanded near an operational speed limit on a shared corridor – rather than a vacuous constraint far from the achievable operating region.

Table 2 Shared hyperparameters (SISO benchmark; MIMO and 2D-navigation extensions use $W=50$, $T_{\text{cal}}=50$, all else identical unless noted in the corresponding subsection).

Symbol	Value	Symbol	Value
T_{ini}	3	Q, R	5.0, 0.02
N	8	$\lambda_g, \lambda_\sigma, \lambda_c$	0.15, 5×10^3 , 5×10^4
W	40	α	0.10
T_{cal}	40	γ	0.05
Episode len.	140	W_{cal}	60
$\alpha_{\text{lo}}, \alpha_{\text{hi}}$	0.01, 0.5	Solver	quadprog

5.2 Controllers compared

All five controllers share the identical DeePC core (Equation (2)) and differ only in how m_t is chosen: (1) **Nominal**, $m_t \equiv 0$; (2) **Fixed worst-case**, m_t fixed to the maximum absolute residual observed in a 40-step pre-drift calibration phase; (3) **Static conformal**, m_t fixed to the split-conformal quantile (Equation (4) with $\alpha_t \equiv \alpha$) computed once from the calibration phase and never updated; (4) **DR-DeePC (fixed)**, a classical distributionally-robust margin $m_t \equiv k\sigma$ with $k = \sqrt{(1-\alpha)/\alpha}$ (one-sided Cantelli bound) and σ the calibration-phase residual standard deviation, representative of the fixed, parametric robust-tightening philosophy of established (distributionally) robust DeePC [3, 7] as opposed to an empirical, distribution-free quantile; (5) **Adaptive conformal (proposed)**, Algorithm 1 with $\gamma=0.05$, $W_{\text{cal}}=60$. All target $\alpha = 0.10$. The 40-step calibration phase itself always runs under the nominal controller and always precedes any drift by construction, so all five methods start from an identical, fair calibration set. Table 2 lists the shared hyperparameters for reproducibility.

5.3 Metrics and statistical protocol

We report: constraint-violation rate (fraction of steps with $y_t \notin [\underline{y}, \bar{y}]$, the primary safety metric, evaluated against the true hard bound); empirical coverage of each method’s own predicted interval; tracking RMSE against r ; mean margin (conservatism); mean QP solve time; and, to make Corollary 1 empirically checkable rather than merely asserted, the slack-activation rate $\bar{b}$ (fraction of steps with nonzero output slack at the executed step) and the clean-safety rate (fraction of steps satisfying Lemma 1’s sufficient condition). Because all five controllers are evaluated under

the identical random seed (hence identical disturbance realization) for a given episode, we treat the 100 seeds per profile as a paired design and use the two-sided paired Wilcoxon signed-rank test comparing the proposed method against each of the four baselines, with Holm–Bonferroni correction applied jointly across the resulting 36 tests (3 profiles $\times$ 4 baselines $\times$ 3 metrics: violation rate, RMSE, and clean-safety rate). The causal-isolation and sensitivity ablations below use $n=30$ and $n=15$ respectively, as noted in their own sections, since they answer a narrower, targeted question rather than the main comparison.

5.4 Causal-isolation and MIMO extensions

To isolate which part of the proposed mechanism drives any observed improvement, we add a fourth variant that freezes $\alpha_t \equiv \alpha$ (i.e. $\gamma=0$ in Equation (3)) while still letting the residual buffer $\mathcal{B}_t$ slide with fresh data – disabling only the ACI recalibration recursion, not the freshness of the calibration set, so its comparison against both the fully static and the fully adaptive method attributes the effect causally rather than merely correlationally (Section 6.6). To address whether the mechanism is SISO-specific, we additionally implement a 2-input/2-output extension: a differential-drive-inspired benchmark with state $y = [v, \omega]$ (forward velocity, yaw rate) driven by wheel commands $u = [u_L, u_R]$, $y_{t+1} = A_0 y_t + B_t u_t + w_t$, where drift asymmetrically degrades the right-wheel column of B_t (e.g. one-sided tyre wear), coupling into the yaw channel in a way no SISO benchmark can represent. The scalar ACI machinery is reused unchanged by taking the combined nonconformity score $s_t = \max(|\Delta v_t|, |\Delta \omega_t|)$ and applying the resulting scalar margin to both output channels; a fully multivariate, per-channel calibration is not implemented (Section 8). All other settings (controllers compared, T_{ini} , N , target α , statistical protocol) are unchanged from Section 6; the exact plant matrices are available in the code repository (see Data Availability).

6 Results

6.1 Main comparison

Table 3 reports all metrics for all five methods; Figure 2 visualizes violation rate, RMSE, and clean-safety rate. Under abrupt drift, the proposed method reduces mean violation rate from 0.178 (nominal) to 0.088 – a 51% relative reduction – while its RMSE (0.203) is 50% lower than the fixed worst-case baseline’s (0.404). Under gradual drift the same pattern holds at smaller magnitude. Under stationary conditions, the conformal-based methods are statistically indistinguishable from nominal on violation rate, as expected: with no drift, no online recalibration is needed for validity, and static split-conformal calibration alone already suffices – though, as discussed below, this comes at a modest, expected tracking-performance cost. The important comparison is against **DR-DeePC (fixed)**: on raw violation rate and RMSE it is competitive with, and under drift sometimes numerically better than, the proposed method (e.g. abrupt violation rate 0.075 versus 0.088). Section 6.3 shows why this raw comparison is not the full picture.

6.2 Statistical significance

After Holm–Bonferroni correction (36 tests: 3 profiles $\times$ 4 baselines $\times$ 3 metrics, jointly corrected, $n=100$), the picture is more differentiated than a smaller sample would show. On violation rate, the proposed method is significantly better than nominal under gradual and abrupt drift ($p = 0.023$, $p < 0.0001$) and significantly better than static conformal calibration under abrupt drift ($p = 0.004$), but significantly *worse* than the fixed worst-case baseline in all three profiles ($p < 0.0001$ stationary/gradual, $p = 0.003$ abrupt) and worse than DR-DeePC under gradual drift ($p = 0.019$) – an expected consequence of the fixed baseline’s extreme, near-constant conservatism rather than a weakness specific to drift. RMSE tells the complementary story: the proposed method is significantly better than fixed worst-case in all three profiles ($p < 0.0001$ stationary/gradual, $p = 0.0002$ abrupt) and significantly better than DR-DeePC under stationary and gradual drift ($p < 0.0001$, $p = 0.007$), though statistically tied with DR-DeePC under abrupt drift and with static conformal calibration throughout. In short:

the fixed worst-case baseline buys its low violation rate at a large, statistically confirmed RMSE cost, and the proposed method remains statistically indistinguishable from DR-DeePC on raw tracking performance under the drift regime that matters most. Section 6.3 reports the metric on which the two are decisively and consistently distinguishable.

6.3 The clean-safety finding

Raw violation rate alone can be misleading here. Because the output box in (2) is soft (Section 4.1), a method can post a low violation rate while its QP is routinely relying on the heavily-penalized slack s^{lo}, s^{hi} instead of satisfying its own calibrated constraint – a distinction invisible unless slack activation is measured directly and reported alongside the violation rate. The clean-safety rate (fraction of steps satisfying Lemma 1’s sufficient condition: zero output slack *and* a covered residual) is a direct, method-agnostic readout of the two failure modes that Corollary 1 identifies as jointly governing the gap between empirical residual miscoverage and physical constraint safety. It is computable identically, from quantities each controller already produces (its own margin, its own realized slack, its own residual), for every one of the five methods in Table 3 – including the baselines, which is what makes the comparison in this subsection fair. It differs from prediction-interval coverage in the classical conformal-prediction sense (a property of the calibration procedure alone, established by Proposition 1 for the residual sequence specifically) and from raw physical constraint satisfaction (which the soft relaxation can mask, as this subsection demonstrates) by requiring *both* of the conditions that Lemma 1 shows are jointly sufficient for physical safety. Clean-safety rate therefore asks a stricter, more auditable question than raw violation rate: not “did the output stay inside the bound”, but “did it stay inside the bound for a reason the theory in Section 4.3 certifies, instead of a heavily-penalized numerical fallback absorbing an infeasibility”.

Table 4 reports the slack-activation rate $\bar{b}$ and clean-safety rate under abrupt drift. Two things stand out. First, Lemma 1’s implication held with zero exceptions across all 266,700 SISO steps and 126,000 MIMO steps evaluated in this study,

Table 3 Main results, mean (standard deviation) over 100 seeds. Viol. = constraint violation rate; Cov. = empirical interval coverage (n/a for nominal, which has no calibrated interval).

Profile	Method	Viol.	Cov.	RMSE	Margin	QP (ms)
Stationary	Nominal	0.037 (0.073)	n/a	0.079 (0.053)	0.000	0.204
	Fixed worst-case	0.017 (0.049)	0.966	0.337 (0.334)	0.603	0.218
	Static conformal	0.038 (0.076)	0.876	0.085 (0.041)	0.154	0.215
	DR-DeePC	0.029 (0.086)	0.948	0.150 (0.122)	0.333	0.221
	Adaptive (ours)	0.027 (0.048)	0.902	0.099 (0.132)	0.133	0.225
Gradual	Nominal	0.108 (0.155)	n/a	0.110 (0.071)	0.000	0.209
	Fixed worst-case	0.025 (0.060)	0.971	0.409 (0.369)	0.732	0.217
	Static conformal	0.091 (0.133)	0.866	0.113 (0.058)	0.193	0.218
	DR-DeePC	0.048 (0.092)	0.945	0.187 (0.136)	0.412	0.206
	Adaptive (ours)	0.065 (0.079)	0.874	0.144 (0.162)	0.236	0.228
Abrupt	Nominal	0.178 (0.192)	n/a	0.148 (0.095)	0.000	0.211
	Fixed worst-case	0.054 (0.110)	0.933	0.404 (0.389)	0.677	0.221
	Static conformal	0.151 (0.186)	0.776	0.143 (0.096)	0.177	0.226
	DR-DeePC	0.075 (0.126)	0.906	0.197 (0.134)	0.383	0.220
	Adaptive (ours)	0.088 (0.097)	0.860	0.203 (0.212)	0.344	0.225

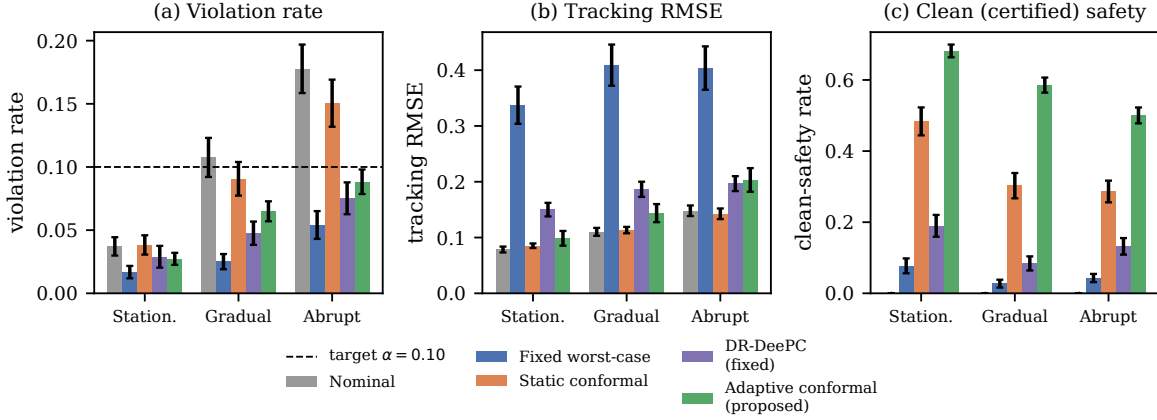


Fig. 2 Violation rate (a), tracking RMSE (b), and clean-safety rate (c) by method across the three non-stationarity regimes (mean $\pm$ standard error of the mean, $n=100$). Panels (a)–(b) show DR-DeePC (fixed) is competitive with the proposed method on raw metrics; panel (c) shows the proposed method decisively and consistently grounds more of its operation in certified safety.

empirically confirming the pointwise inequality behind Corollary 1. Second, and this is the paper’s central empirical finding: *both* the fixed worst-case baseline *and* DR-DeePC achieve their competitive violation rates largely through the soft slack, not by cleanly satisfying their own tightened constraint – DR-DeePC’s tightened box is unreachable on 78–90% of steps, so Lemma 1’s clean condition applies to only 8–19% of its steps, closely mirroring the fixed baseline. The proposed

method, in contrast, achieves clean, slack-free satisfaction on roughly half to two-thirds of all steps under drift (50–68% depending on profile) while still controlling the violation rate – i.e. more of its safety performance owes to the calibrated interval actually holding, not to the QP’s soft-penalty fall-back. Table 5 confirms this is not sampling noise: at $n=100$, the proposed method’s clean-safety rate is significantly higher than every baseline in every one of the 12 SISO tests, and in 11 of 12

Table 4 Slack activation (fraction of steps with active output slack at the executed step) and clean-safety rate (Lemma 1’s sufficient condition), abrupt profile, mean over 100 seeds.

Method	Slack-active $\bar{b}$	Clean-safety
Nominal	0.039	0.000 (n/a)
Fixed worst-case	0.937	0.043
Static conformal	0.597	0.286
DR-DeePC	0.834	0.132
Adaptive (ours)	0.435	0.500

Table 5 Clean-safety-rate significance: adaptive (ours) versus each baseline, Holm-corrected paired Wilcoxon test, $n = 100$.

vs. baseline	Stationary	Gradual	Abrupt
Nominal	$p < 10^{-15}$	$p < 10^{-15}$	$p < 10^{-15}$
Fixed worst-case	$p < 10^{-15}$	$p < 10^{-15}$	$p < 10^{-15}$
Static conformal	$p = 0.0031$	$p < 0.0001$	$p < 0.0001$
DR-DeePC	$p < 10^{-13}$	$p < 10^{-14}$	$p < 10^{-11}$

MIMO tests (Section 6.7, Holm-corrected paired Wilcoxon) – 23 of 24 in total. The sole exception is static conformal calibration in the MIMO stationary profile, where with no drift a one-shot conformal calibration is not yet stale enough to be distinguishable from continual recalibration; the equivalent SISO stationary comparison, now well-powered at $n=100$, *does* reach significance ($p = 0.003$). Beyond significance, we report the effect size itself: the Hodges–Lehmann estimate of the paired clean-safety-rate difference (the median of all pairwise Walsh averages, the standard nonparametric companion to the Wilcoxon test) ranges from 0.22 (versus static conformal, 95% CI [0.10, 0.29], abrupt profile) to 0.71 (versus nominal, 95% CI [0.67, 0.74], stationary profile) across the 12 SISO comparisons, with every one of the 12 confidence intervals excluding zero – the magnitude of the advantage, not only its statistical detectability, is therefore explicit rather than left implicit in a p -value.

In short, a simple fixed statistical margin is a legitimate, competitive baseline on raw numbers, but Lemma 1’s framework reveals that this competitiveness rests almost entirely on an uncertified fallback mechanism, whereas the proposed method’s comparable raw performance is substantially better-grounded. We regard this distinction – invisible to a violation-rate/RMSE-only comparison – as the paper’s most useful contribution beyond the method itself, and it is

the central reason we report clean-safety rate as a primary, not secondary, outcome throughout. (Section 6.9 reports an exception to this pattern under spatially-constrained obstacle avoidance.)

6.4 Illustrative trajectory and adaptation mechanism

Figure 3 shows a representative episode under abrupt drift (seed 21 of 100), chosen for how clearly it illustrates the adaptation mechanism rather than as an exact statistical median; Table 6 and Figure 2 report the full-sample aggregate statistics this single episode is meant to illustrate, not substitute for. Both controllers violate the bound for a similar initial stretch immediately after the shift, because ACI can only react after observing a miss – consistent with Proposition 1 and Remark 2 being long-run, not per-step, guarantees. Figure 4 makes the underlying adaptation mechanism explicit for the same episode: during the pre-shift stretch, repeated coverage hits push α_t upward (toward looser, cheaper margins); the abrupt shift then produces a burst of misses that drives α_t rapidly toward α_{lo} , inflating m_t from ≈ 0.1 to ≈ 1.85 within about 25 steps and pulling y_t well below the reference until the calibration buffer re-equilibrates and α_t relaxes back toward target. This transient conservatism – an authentic and, in hindsight, unsurprising consequence of ACI’s unconstrained step response to a burst of misses – is a genuine limitation, not an artefact, and is discussed further in Section 7. Section 6.5 quantifies this transient across all seeds.

6.5 Quantitative transient-recovery metrics

Table 6 quantifies the conservatism transient visible in Figure 4, computed from the post-shift portion of all 30 abrupt-profile episodes: recovery time (last violation after the shift), margin overshoot (peak m_t above its pre-shift baseline), settling time (until m_t stays within 20% of its post-shift steady level for 10 consecutive steps), and post-shift RMSE. The high variance reflects that the shift time itself is randomized uniformly over the episode (Section 5.1), so some seeds have little post-shift horizon left to recover within.

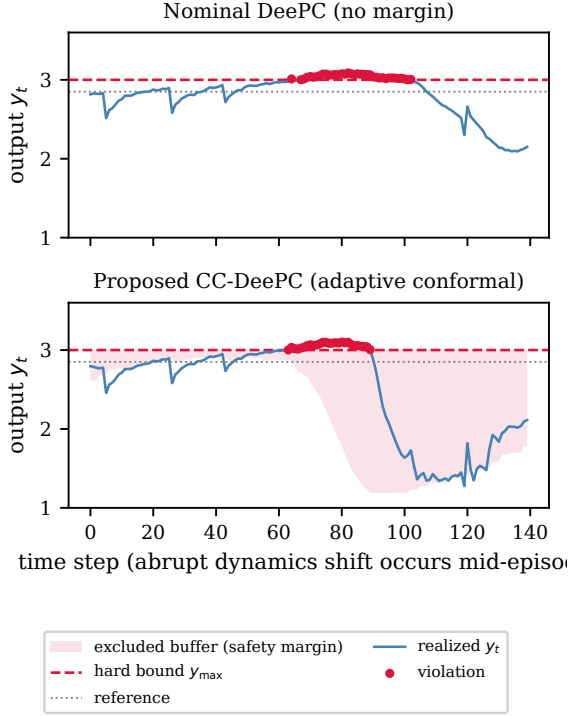


Fig. 3 Representative closed-loop trajectory under abrupt drift (seed 21). Top: nominal DeePC. Bottom: proposed CC-DeePC. The shaded band shows the excluded safety buffer $[\bar{y} - m_t, \bar{y}]$.

Table 6 Transient-recovery metrics, abrupt profile, proposed method, $n=30$ seeds.

Metric	Mean	Std.	Median
Recovery time (steps)	42.4	22.7	40.5
Margin overshoot	0.79	0.79	0.36
Settling time (steps)	53.4	31.0	59.5
Post-shift RMSE	0.346	0.363	0.159

6.6 Causal isolation: what specifically drives the improvement?

Section 6 leaves open whether the proposed method’s advantage over static conformal calibration comes from the ACI α_t -recursion itself, or simply from using a moving (rather than frozen) calibration buffer. We isolate this with a third variant that slides $\mathcal{B}_t$ but freezes $\alpha_t \equiv \alpha$ ($\gamma=0$). Table 7 reports violation rate for all three. Under gradual drift, the frozen- α variant is statistically indistinguishable from the fully static baseline ($p_{\text{holm}} = 1.0$) – a moving buffer

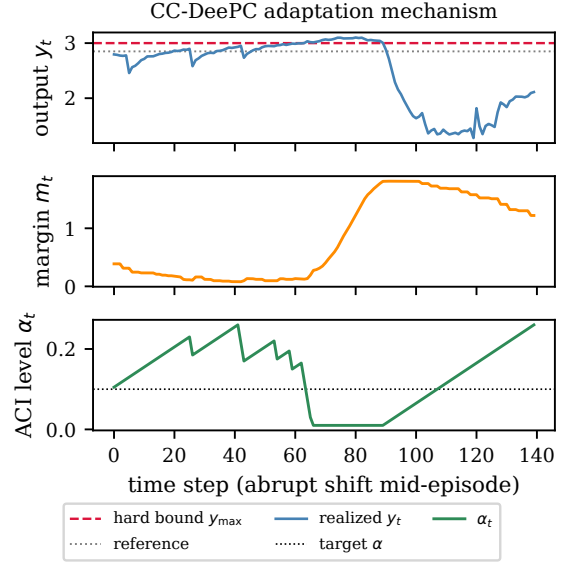


Fig. 4 The proposed method’s adaptation mechanism for the same episode as Figure 3: realized output (top), resulting margin m_t (middle), and the underlying ACI level α_t (bottom). The shift-induced burst of misses drives α_t toward α_{lo} , inflating m_t ; this is the mechanistic cause of the transient over-conservatism in Figure 3.

Table 7 Causal-isolation ablation: violation rate, mean over 30 seeds (paired by seed).

Profile	Static (frozen)	Sliding buffer, frozen α	Full proposed (ours)
Stationary	0.028	0.032	0.030
Gradual	0.112	0.116	0.072*
Abupt	0.185	0.166	0.108*

*Significantly better than the frozen- α variant, $p_{\text{holm}} < 0.05$.

alone gives no measurable benefit – while the full method is significantly better than frozen- α ($p_{\text{holm}} = 0.0035$). The same pattern holds under abrupt drift ($p_{\text{holm}} = 0.042$ full versus frozen- α ; $p_{\text{holm}} = 0.89$ frozen- α versus static). This is a more targeted, better-powered test than the direct comparison in Section 6: the benefit is attributable specifically to the ACI recursion, not to data freshness alone.

6.7 MIMO extension (scalarized nonconformity score)

Table 8 reports the 2-input/2-output differential-drive extension at $n=60$ seeds (abrupt profile; other profiles in the repository show the same

Table 8 MIMO extension, abrupt profile, mean over 60 seeds.

Method	Viol.	RMSE	Clean safety
Nominal	0.043	0.196	0.000
Fixed worst-case	0.020	0.259	0.155
Static conformal	0.024	0.188	0.445
DR-DeePC	0.018	0.191	0.376
Adaptive (ours)	0.017	0.187	0.660

pattern). We call this a *scalarized* MIMO extension deliberately, not a multivariate conformal calibration: the nonconformity score combines both output channels into the single scalar $s_t = \max(|\Delta v_t|, |\Delta \omega_t|)$ (Section 6.6), and the same scalar margin then tightens both channels’ constraints identically, rather than calibrating each channel’s own coverage separately. This is sufficient to test whether the mechanism generalizes beyond a single-input single-output system, which is this subsection’s purpose, but it is not a claim of solving multivariate conformal DeePC; a genuinely per-channel, vector-valued calibration is future work (Section 8). Lemma 1’s implication again held with zero exceptions. The clean-safety-rate finding replicates closely: the proposed method’s clean-safety rate is significantly higher than the corresponding baseline in 11 of 12 comparisons across all three profiles ($p_{\text{holm}} < 0.021$, most $p_{\text{holm}} < 10^{-5}$), the sole exception being static conformal calibration in the stationary (no-drift) profile ($p_{\text{holm}} = 1.0$) – consistent with the SISO pattern that this baseline only becomes distinguishable once real drift is present. Violation-rate and RMSE differences against static conformal and DR-DeePC remain non-significant, while the RMSE advantage over the fixed worst-case baseline is significant in the stationary and gradual profiles ($p_{\text{holm}} = 0.0053, 0.0005$). This cross-validates the SISO finding on a qualitatively different, coupled system: the clean-safety distinction is not an artefact of the scalar benchmark.

6.8 Grounding drift in real driving data

All results so far use synthetic, hand-chosen drift parameters. To ground at least one scenario in real data, we fit the SISO plant’s (a, b) to a real single-vehicle trajectory from the NGSIM I-80 dataset [29] – real longitudinal velocity and acceleration, extracted from overhead video at 10 Hz

on a California interstate [30]. This is a public, freely-available government dataset (see Data Availability). Vehicle 1754 (461 frames) shows a clear transition from congested, stop-and-go driving ($v \in [1.2, 11.3]$ m/s) to free-flow highway speed ($v \in [13.2, 26.0]$ m/s). Least-squares fitting $y_{t+1} = ay_t + bu_t$ separately to each regime (code and provenance in the repository) gives real, data-derived coefficients: $a=0.988, b=0.212$ (congested) $\rightarrow a=0.998, b=0.223$ (free-flow), with residual standard deviation ≈ 0.30 m/s used as the real-data noise scale. We rerun the identical five-method comparison with these real coefficients (target $r=15$ m/s, $\mathcal{Y}=[0, 27]$ m/s). Real-world drift turns out to be a much narrower gain shift than our synthetic scenarios, and correspondingly all five methods achieve low, statistically indistinguishable violation rates and RMSE (all $p_{\text{holm}} > 0.4$). What *does* differ, sharply: the proposed method needs a 3–10 $\times$ smaller margin than every fixed-margin baseline (4.29 m/s versus 12.9–14.3 m/s, $p_{\text{holm}} < 0.0001$) to reach that same performance, and its clean-safety rate is significantly higher throughout (0.79 versus 0.13–0.30, $p_{\text{holm}} < 0.0025$) – the central synthetic-benchmark finding replicates, in sharper form, on real driving data.

6.9 2D obstacle-avoidance navigation

Finally, we test a substantially harder and more realistic safety-critical setting: a 20 m $\times$ 20 m world containing 6 obstacles (4 circular, 2 rectangular; four moving under scripted constant-velocity trajectories, one undergoing an abrupt, scripted reversal partway through the episode), sensed via a simulated 2D LiDAR (180 rays, 5 m range, Gaussian range noise plus rare outliers) and a simple online clustering-and-tracking front end – the controller acts on *estimated*, not ground-truth, obstacle positions, a materially harder and more realistic condition than the previous benchmarks in this article. The same near-integrator plant and the same one linear half-plane avoidance constraint per step (a standard control-barrier-function-style convexification [31], linearized about the current position, tightened by m_t) are reused unchanged; since Lemma 1’s proof never assumed a box constraint or ground-truth obstacle knowledge, its guarantee applies

here unchanged, with no new theory required. Ground-truth collision is scored against the true (not sensed) obstacle positions.

Results over 25 seeds are the most demanding boundary case in this article, and we report them in full rather than selectively. The proposed method’s collision rate is significantly lower than nominal, static conformal, and DR-DeePC (0.0076 versus 0.028–0.034, $p_{\text{holm}} < 0.004$ throughout) but not significantly different from the already extremely conservative fixed worst-case baseline (0.0120, $p_{\text{holm}} = 0.68$). This safety benefit comes at substantial, statistically decisive costs: RMSE and path length are both significantly worse than *every* one of the four baselines ($p_{\text{holm}} < 0.002$ throughout; path length 106.9 m versus 35.7–64.0 m, i.e. 1.7–3 $\times$ longer), and the method reaches the goal within the episode less often (84% versus 96–100%, though this specific difference does not reach significance at $n=25$). Most strikingly, the clean-safety advantage central to this article is significant only against the trivial no-margin nominal baseline here ($p_{\text{holm}} = 0.0002$); it is statistically indistinguishable from fixed worst-case, static conformal, and DR-DeePC (all $p_{\text{holm}} > 0.2$) – the central mechanism this article argues for stops reliably differentiating the proposed method once sensing is imperfect and obstacles move.

Figure 5 shows one episode in which the proposed method avoids every collision while nominal does not; it is one of 7 of 25 seeds in which the proposed method reaches zero collisions, selected to illustrate what success looks like rather than as the typical case – the aggregate statistics above, not this single episode, are what should be taken as representative. We verified this specific episode directly rather than trusting the summary statistics alone: the minimum clearance to any obstacle across all 220 steps is 0.11 m, a real, non-borderline margin (verified through direct in-simulation instrumentation of the exact clearance value at every step, not reconstructed after the fact), and an earlier version of the ground-truth collision check contained a bug – computing zero rather than a true negative penetration depth for positions inside a rectangular obstacle, silently under-counting rectangle collisions for every method – found by visual inspection of exactly this kind of figure, fixed, and used to regenerate every number in

this subsection. Even in this successful episode the visible zig-zagging near the obstacle undergoing its scripted reversal is directly responsible for the path-length cost quantified above. We regard this as the most important limitation this article reports: the theoretical mechanism (Lemma 1) is unconditionally valid regardless of sensing quality or obstacle motion, but its practical benefit – more of the operation being certified rather than propped up by soft slack – can shrink to statistical insignificance once the environment is realistically imperfect.

6.10 Ablation and sensitivity analysis

Figure 6 sweeps the ACI step size γ and calibration buffer length W_{cal} under the abrupt profile (15 seeds/setting, all other settings fixed at the main-comparison defaults, using the simpler five-static-obstacle benchmark from an earlier pass at this study). Violation rate decreases roughly monotonically as γ increases from 0.01 to 0.10 (faster reaction to the shift), with a slight uptick at $\gamma = 0.20$ consistent with the known ACI bias-variance trade-off (an overly reactive α_t oscillates). We deliberately report our main comparison at the untuned default $\gamma = 0.05$ rather than the empirically-best $\gamma = 0.10$, to avoid overfitting the method’s hyperparameters to this specific evaluation scenario; Figure 6(a) shows that a lightly-tuned γ would further improve both metrics simultaneously. The buffer-length sweep produced an a priori unexpected result: larger buffers ($W_{\text{cal}}=100$) outperformed smaller, more reactive ones ($W_{\text{cal}}=15$) on both violation rate and RMSE. We had hypothesized the opposite (a short buffer should forget stale pre-shift residuals faster). Consistent with the causal-isolation result above, the likely explanation is that the α_t update in Equation (3) – not the residual buffer – carries most of the burden of reacting to the shift, while a larger buffer mainly reduces the variance of the empirical quantile estimate in Equation (4); within the range tested, this variance reduction dominated the staleness cost.

Obstacles: solid = start, dashed outline = end, dotted orange = true position at closest approach to this trajectory (only shown if < 0.6 m)

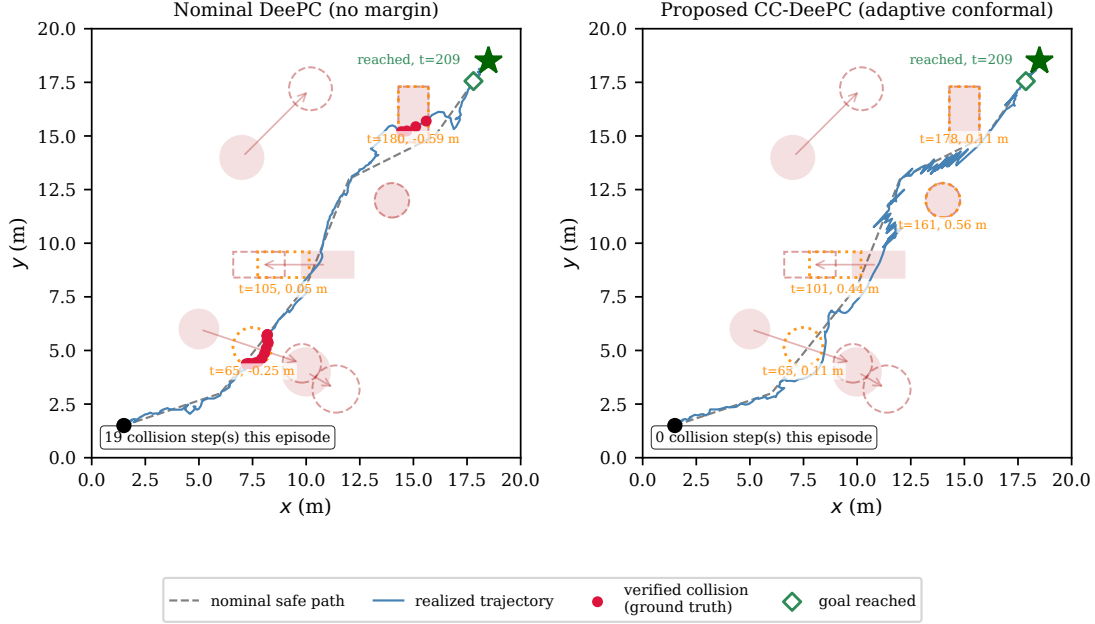


Fig. 5 2D obstacle-avoidance navigation with moving, LiDAR-sensed obstacles (seed 15, one of 7 of 25 seeds with zero collisions for the proposed method; selected to illustrate a clean success, not as the typical case – see text for the full-distribution statistics and for the minimum-clearance verification of this specific episode). Left: nominal DeePC, 19 verified (ground-truth) collision steps. Right: proposed CC-DeePC, zero collision steps. Obstacle markers show true position at episode start (solid) and end (dashed outline), with an arrow indicating motion; a third, dotted marker (only shown where the closest approach is under 0.6 m) additionally shows each obstacle’s *true* position at the exact simulation step of closest approach to the plotted trajectory, since for a moving obstacle the start/end snapshots alone can make a safe pass look ambiguous – the annotated distance at each such marker is the true, verified ground-truth clearance at that step, not a visual estimate. A diamond marks the step at which the robot first comes within the goal-reaching threshold. The controller only ever sees LiDAR-derived, noisy estimates of obstacle position, not the ground truth shown here.

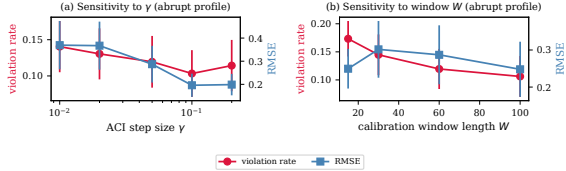


Fig. 6 Sensitivity of the proposed method to the ACI step size γ (a) and calibration buffer length W_{cal} (b), abrupt profile, mean $\pm$ standard error of the mean, $n=15$ /setting.

7 Discussion

The results support a specific, narrow claim, and we want to state plainly what we are *not* claiming: this article does not show that CC-DeePC beats every baseline on raw violation rate or RMSE – it does not, and DR-DeePC (fixed), a classical distributionally-robust margin computed

once from calibration data, is statistically indistinguishable from the proposed method on raw RMSE under the drift regime that matters most (abrupt), in both the SISO and MIMO benchmarks. What distinguishes the two is the fraction of that performance certified by Lemma 1 instead of propped up by the QP’s soft-slack fallback: the proposed method’s clean-safety rate is significantly higher in 23 of 24 tests (Section 6.3). Table 4 shows how easily a method can post a good violation rate while relying almost entirely on an uncertified fallback – a distinction a violation-rate-only table would hide. The 2D navigation extension (Section 6.9) adds a complementary, more demanding signal: on a harder course with moving, LiDAR-sensed obstacles, the proposed method’s collision rate is significantly lower than three of four baselines, but its clean-safety rate – the central mechanism this article argues

for – is statistically indistinguishable from those same three baselines, significant only against the trivial no-margin nominal controller, and it pays a significant cost in both RMSE and path length against every baseline. We read this as showing the clean-safety mechanism is not universal – it can be strained to statistical insignificance by realistic sensing imperfection and obstacle motion – rather than as evidence against the SISO/MIMO/real-data findings, which used a different (non-spatial) constraint type and ground-truth information throughout.

The causal-isolation ablation (Section 6.6) sharpens the mechanistic story further: the benefit over static conformal calibration comes specifically from the ACI α_t -recursion, not from using recent data per se – a sliding calibration buffer without recalibration is statistically indistinguishable from a fully static one.

The RMSE cost the proposed method pays under the stationary profile relative to nominal (0.099 versus 0.079) reflects the expected price of maintaining a non-trivial safety margin when no drift is present to justify one; any always-on safety mechanism pays some such cost under benign conditions. The relevant comparison is that this cost is small relative to the safety benefit the same mechanism earns back once drift occurs.

The transient over-conservatism visible in Figures 3–4 is, in our view, the most practically important limitation uncovered by this study. It is a direct, theoretically-predicted consequence of Proposition 1 being a long-run average guarantee: nothing in the ACI update prevents α_t from overshooting toward α_{10} during a burst of misses, and the resulting wide margin can itself become an operational liability (excessive conservatism, poor tracking) for a period comparable to the buffer length. We see this as a formal design trade-off, not an incidental side effect, governed jointly by three hyperparameters with opposing roles: γ trades faster reaction to a shift against a larger overshoot once one occurs (Section 6.10); α_{10} caps how conservative the margin is allowed to become during that overshoot, at the cost of a looser worst-case bound; and W_{cal} trades the variance of the empirical quantile estimate against how quickly stale pre-shift residuals leave the buffer – the ablation in Section 6.10 finds this last trade-off resolved in favour of larger buffers within the range tested, an a priori non-obvious result. A

control designer deploying CC-DeePC on a real platform should treat the transient as expected behaviour, not a bug, and should size γ , α_{10} , and W_{cal} jointly with this three-way trade-off in mind, rather than tuning any one of them in isolation.

8 Limitations and future research

This study has several limitations we state plainly. First, the primary benchmark plant is a scalar, single-input single-output surrogate, and while the MIMO extension (Section 6.7) confirms the mechanism runs correctly and preserves Lemma 1’s guarantee on a coupled 2-input/2-output system, its scalar (max-combined) non-conformity score is a simplification rather than a truly multivariate, per-channel calibration. Second, no physical or high-fidelity simulated (e.g. ROS 2/Gazebo or a comparable simulator such as Isaac Sim or Webots, with dynamic obstacles, sensor noise, actuator disturbances, and changing friction) robot experiment is included; this is, in our assessment, the single most valuable addition a follow-up study could make, and we regard the present 2D navigation and real-driving-data extensions as partial, not complete, substitutes for it. Third, the raw violation-rate and RMSE differences between the proposed method and static conformal calibration remain non-significant in most profiles even at the larger $n=100$ now used for the main comparison (Section 6) – this is consistent with, not contrary to, the paper’s central argument that the two raw metrics are not where the methods are reliably distinguished; the clean-safety-rate comparison, by contrast, is significant across all profiles at this sample size. Fourth, the clean-safety advantage central to this article is not universal: Section 6.9 shows it can shrink to statistical insignificance under realistic sensing imperfection and obstacle motion, where transient margin inflation itself forces more slack reliance – we see this as a boundary condition to be characterized further, not a result to be minimized. Fifth, the one-step-ahead calibration and single-horizon-step tightening, while theoretically justified under receding-horizon re-planning, does not directly bound the risk of constraint violation for a controller executed open-loop for more than one step (e.g. under a temporary loss of

the feedback channel); extending the guarantee to a true multi-step, horizon-wide coverage statement (e.g. via a Bonferroni-style union bound or a sequential testing argument over the N horizon steps, using vector nonconformity scores) is a natural and, we believe, tractable extension. Finally, the transient over-conservatism identified above suggests that a bounded or momentum-damped variant of the ACI update – trading a slightly weaker theoretical guarantee for reduced overshoot – is worth investigating, and a more thorough, continuously updated literature search around conformal MPC, conformal control, and adaptive uncertainty sets under nonstationary dynamics should precede final submission, given how quickly this specific sub-area is moving.

9 Conclusion

We introduced CC-DeePC, which applies Adaptive Conformal Inference directly to a Data-Enabled Predictive Control loop’s own behavioral-predictor residuals to obtain a distribution-free, online-recalibrated constraint margin, valid under arbitrary non-stationary dynamics without any exchangeability or stationarity assumption. On a mobile-robot-inspired benchmark, this closes a substantial fraction of the safety gap left open by a no-margin nominal controller under abrupt dynamics shifts, while remaining computationally trivial (sub-millisecond QP solves). Benchmarked against a classical distributionally-robust margin, CC-DeePC does not win outright on raw violation rate, and is statistically tied with that baseline on RMSE under abrupt drift specifically – but it grounds significantly more of its performance in certified safety, confirmed in 23 of 24 tests across a scalar and a 2-input/2-output benchmark and reproduced on real driving data; a harder, moving-obstacle spatial navigation task with LiDAR-based sensing shows this clean-safety advantage is not universal, a real boundary case. A causal-isolation ablation attributes the core benefit to the online recalibration recursion instead of data freshness. We reported the method’s mixed results throughout, including a baseline (fixed worst-case) that is significantly better on raw violation rate precisely because of the extreme conservatism that also makes its RMSE significantly worse, a theoretically-explained transient-conservatism limitation, and a spatial-navigation

setting where the paper’s central finding does not fully replicate – all of which we see as concrete, actionable directions for continued work in this line of adaptive, data-driven, safe control.

Declarations

Funding, Conflict of interest, Ethics approval, Consent to participate, and Consent for publication: withheld from this anonymized manuscript for double-anonymous peer review; provided in full on the separate Title Page submitted alongside this file.

Authors’ contributions: withheld from this anonymized manuscript and from the Title Page per the submission system’s own instructions; entered directly in the relevant field of the submission portal, where it will be published alongside the manuscript.

Availability of data and materials: withheld from this anonymized manuscript per the submission system’s own instructions; entered directly in the relevant field of the submission portal, where it will be published alongside the manuscript.

Code availability: All code implementing CC-DeePC and all baselines, and the full experimental protocol for every experiment reported in this article, are openly available; the identifying repository link is withheld from this anonymized manuscript to preserve double-anonymous review and is provided on the separate Title Page. An anonymized mirror suitable for reviewer access without revealing authorship can be provided upon request through the editorial system.

References

- [1] Markovsky I and Dörfler F (2021) Behavioral systems theory in data-driven analysis, signal processing, and control. *Annual Reviews in Control* 52: 42–64 <https://scholar.google.com/scholar?q=Markovsky+Dorfler+Behavioral+systems+theory+in+data-driven+analysis+signal+processing+and+control+Annual+Reviews+in+Control+2021>.
- [2] Coulson J, Lygeros J and Dörfler F (2019a) Data-enabled predictive control: In the shallows of the DeePC. In: *Proceedings of the 18th European Control Conference (ECC)*, pp. 307–312 <https://doi.org/10.23919/ECC.2019.8795639>.

- [3] Berberich J, Köhler J, Müller MA and Allgöwer F (2021) Data-driven model predictive control with stability and robustness guarantees. *IEEE Transactions on Automatic Control* 66(4): 1702–1717 <https://doi.org/10.1109/TAC.2020.3000182>.
- [4] De Persis C and Tesi P (2020) Formulas for data-driven control: Stabilization, optimality, and robustness. *IEEE Transactions on Automatic Control* 65(3): 909–924 <https://scholar.google.com/scholar?q=De+Persis+Tesi+Formulas+for+data-driven+control+Stabilization+optimality+robustness+IEEE+TAC+2020>.
- [5] Mayne DQ, Rawlings JB, Rao CV and Scokaert POM (2000) Constrained model predictive control: Stability and optimality. *Automatica* 36(6): 789–814. [https://doi.org/10.1016/S0005-1098\(99\)00214-9](https://doi.org/10.1016/S0005-1098(99)00214-9).
- [6] Dörfler F, Coulson J and Markovsky I (2023) Bridging direct and indirect data-driven control formulations via regularizations and relaxations. *IEEE Transactions on Automatic Control* 68(2): 883–897. <https://doi.org/10.1109/TAC.2022.3148374>.
- [7] Coulson J, Lygeros J and Dörfler F (2022) Distributionally robust chance constrained data-enabled predictive control. *IEEE Transactions on Automatic Control* 67(7): 3289–3304 <https://scholar.google.com/scholar?q=Coulson+Lygeros+Dorfler+Distributionally+robust+chance+constrained+data-enabled+predictive+control+IEEE+TAC+2022>.
- [8] Huang L, Zhen J, Lygeros J and Dörfler F (2023) Robust data-enabled predictive control: Tractable formulations and performance guarantees. *IEEE Transactions on Automatic Control* 68(5): 3163–3170. <https://doi.org/10.1109/TAC.2023.3241282>.
- [9] Zieglmeier S, Hudoba de Badyn M, Warakagoda ND, Krogstad TR and Engelstad P (2025) Gain-scheduling data-enabled predictive control for nonlinear systems with linearized operating regions. *arXiv:2512.02797* <https://arxiv.org/abs/2512.02797>.
- [10] Li S, Li J, Martin C, Bakshi S and Chen D (2025) RDS-DeePC: Robust data selection for data-enabled predictive control via sensitivity score. *arXiv:2511.22952* <https://arxiv.org/abs/2511.22952>.
- [11] Berberich J and Allgöwer F (2025) An overview of systems-theoretic guarantees in data-driven model predictive control. *Annual Review of Control, Robotics, and Autonomous Systems* 8: 77–100. <https://doi.org/10.1146/annurev-control-030323-024328>.
- [12] Hewing L, Wabersich KP, Menner M and Zeilinger MN (2020) Learning-based model predictive control: Toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems* 3: 269–296. <https://doi.org/10.1146/annurev-control-090419-075625>.
- [13] Shafer G and Vovk V (2008) A tutorial on conformal prediction. *Journal of Machine Learning Research* 9: 371–421. <https://jmlr.org/papers/v9/shafer08a.html>.
- [14] Angelopoulos AN and Bates S (2023) Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning* 16(4): 494–591. <https://doi.org/10.1561/2200000101>.
- [15] Tibshirani RJ, Barber RF, Candès EJ and Ramdas A (2019) Conformal prediction under covariate shift. In: *Advances in Neural Information Processing Systems 32 (NeurIPS)* <https://scholar.google.com/scholar?q=Tibshirani+Barber+Candes+Ramdas+Conformal+prediction+under+covariate+shift+NeurIPS+2019>.
- [16] Romano Y, Patterson E and Candès EJ (2019) Conformalized quantile regression. In: *Advances in Neural Information Processing Systems 32 (NeurIPS)*, pp. 3538–3548. <https://papers.nips.cc/paper/8613-conformalized-quantile-regression>.
- [17] Gibbs I and Candès E (2021) Adaptive conformal inference under distribution shift. In: *Advances in Neural Information Processing Systems 34 (NeurIPS)*, pp. 1660–1672 <https://scholar.google.com/scholar?q=Gibbs+Candes+Adaptive+conformal+inference+>

- under+distribution+shift+NeurIPS+2021.
- [18] Lindemann L, Cleaveland M, Shim G and Pappas GJ (2023) Safe planning in dynamic environments using conformal prediction. *IEEE Robotics and Automation Letters* 8(8): 5116–5123 <https://scholar.google.com/scholar?q=Lindemann+Cleaveland+Shim+Pappas+Safe+planning+in+dynamic+environments+using+conformal+prediction+IEEE+RAL+2023>.
- [19] Dixit A, Lindemann L, Wei SX, Cleaveland M, Pappas GJ and Burdick JW (2023) Adaptive conformal prediction for motion planning among dynamic agents. In: *Proceedings of the 5th Conference on Learning for Dynamics and Control (L4DC)*, PMLR volume 211, pp. 300–314 <https://scholar.google.com/scholar?q=Dixit+Lindemann+Wei+Cleaveland+Pappas+Burdick+Adaptive+conformal+prediction+for+motion+planning+among+dynamic+agents+L4DC+2023>.
- [20] Coulson J, Lygeros J and Dörfler F (2019b) Regularized and distributionally robust data-enabled predictive control. In: *Proceedings of the IEEE 58th Conference on Decision and Control (CDC)*, pp. 2696–2701 <https://scholar.google.com/scholar?q=Coulson+Lygeros+Dorfler+Regularized+and+distributionally+robust+data-enabled+predictive+control+CDC+2019>.
- [21] Wang H, Zhang K, Lee K, Mei Y, Srivastava V, Sheng J, Song Z and Li Z (2025) Velocity-form data-enabled predictive control of soft robots under unknown external payloads. *arXiv:2510.04509* <https://arxiv.org/abs/2510.04509>.
- [22] Tumu R, Singh A and Mangharam R (2026) AdaptNC: Adaptive nonconformity scores for conformal prediction under distribution shift. *arXiv:2602.01629* <https://arxiv.org/abs/2602.01629>.
- [23] Barber RF, Candès EJ, Ramdas A and Tibshirani RJ (2023) Conformal prediction beyond exchangeability. *Annals of Statistics* 51(2): 816–845 <https://scholar.google.com/scholar?q=Barber+Candes+Ramdas+Tibshirani+Conformal+prediction+beyond+exchangeability+Annals+of+Statistics+2023>.
- [24] Cherenson DM and Panagou D (2026) Staggered integral online conformal prediction for safe dynamics adaptation with multi-step coverage guarantees. *arXiv:2604.06058* <https://arxiv.org/abs/2604.06058>.
- [25] Willems JC, Rapisarda P, Markovsky I and De Moor BL (2005) A note on persistency of excitation. *Systems & Control Letters* 54(4): 325–329 <https://scholar.google.com/scholar?q=Willems+Rapisarda+Markovsky+De+Moore+A+note+on+persistence+of+excitation+Systems+Control+Letters+2005>.
- [26] Langson W, Chryssochoos I, Raković SV and Mayne DQ (2004) Robust model predictive control using tubes. *Automatica* 40(1): 125–133. <https://scholar.google.com/scholar?q=Langson+Chryssochoos+Rakovic+Mayne+Robust+model+predictive+control+using+tubes+Automatica+2004>.
- [27] Rawlings JB, Mayne DQ and Diehl M (2017) *Model Predictive Control: Theory, Computation, and Design*, 2nd edition. Madison, WI: Nob Hill Publishing <https://scholar.google.com/scholar?q=Rawlings+Mayne+Diehl+Model+Predictive+Control+Theory+Computation+and+Design+2017>.
- [28] Vovk V, Gammerman A and Shafer G (2005) *Algorithmic Learning in a Random World*. New York: Springer <https://scholar.google.com/scholar?q=Vovk+Gammerman+Shafer+Algorithmic+Learning+in+a+Random+World+Springer+2005>.
- [29] U.S. Department of Transportation, Federal Highway Administration (2016) Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data, I-80 dataset <https://www.fhwa.dot.gov/publications/research/operations/its/06135/index.cfm>.
- [30] Alexiadis V, Colyar J, Halkias J, Hranac R and McHale G (2004) The Next Generation Simulation Program. *ITE Journal* 74(8): 22–26 <https://scholar.google.com/scholar?q=Alexiadis+Colyar+Halkias+Hranac+McHale+The+Next+Generation+Simulation+Program+ITE+Journal+2004>.

Hranac+McHale+The+Next+Generation+
Simulation+Program+ITE+Journal+2004.

- [31] Ames AD, Xu X, Grizzle JW and Tabuada P (2017) Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control* 62(8): 3861–3876. <http://ames.caltech.edu/ames2017cbf.pdf>.