



TRACE-CRC: Trajectory-Adaptive Conformal Risk Control for Multi-Step Channel State Information Prediction

Downloaded from: <https://research.chalmers.se>, 2026-09-07 12:52 UTC

Citation for the original published paper (version of record):

Rezaei, K., Sattari, M., Aliakbari, J. et al (2026). TRACE-CRC: Trajectory-Adaptive Conformal Risk Control for Multi-Step Channel State Information Prediction. *Proceedings of Machine Learning Research*, 329: 984-1009

N.B. When citing this work, cite the original published paper.

TRACE-CRC: Trajectory-Adaptive Conformal Risk Control for Multi-Step Channel State Information Prediction

Kiarash Rezaei

Mehdi Sattari

Javad Aliakbari

Tommy Svensson

Paolo Monti

Carlos Natalino

KIARASHR@CHALMERS.SE

MEHDI.SATTARI@CHALMERS.SE

JAVADA@CHALMERS.SE

TOMMY.SVENSSON@CHALMERS.SE

MPAULO@CHALMERS.SE

CARLOS.NATALINO@CHALMERS.SE

Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

Editor: Ernst Ahlberg, Ulf Johansson, Henrik Boström, Alberto Carlevaro, Johan Hallberg Szabadváry and Lars Carlsson

Abstract

Reliable prediction of time-varying channel state information (CSI) is essential for efficient wireless communication. Each CSI frame is a matrix-valued representation of the wireless channel response, and a sequence of CSI frames forms a temporal channel trajectory. Modern deep learning-based CSI predictors, however, often provide only point predictions and lack calibrated uncertainty estimates. This limitation is particularly problematic in multi-step CSI prediction, where the target is a sequence of future CSI matrices, and downstream decisions such as beamforming or scheduling may fail if any part of the predicted trajectory is unreliable. We propose trajectory-adaptive calibration and error profiling with conformal risk control (TRACE-CRC), a method for trajectory-aware uncertainty quantification in multi-step CSI prediction. TRACE-CRC constructs Frobenius-norm uncertainty balls around predicted CSI matrices and controls the risk that at least one future frame is uncovered. Instead of calibrating each future step independently, TRACE-CRC combines future-step-dependent error profiling, trajectory difficulty stratification, and learn-then-test (LTT) risk control. Empirically, TRACE-CRC achieves reliable trajectory-level coverage with substantially smaller uncertainty balls than conservative multi-step corrections, while avoiding the trajectory undercoverage of compact stepwise and adaptive conformal baselines.

Keywords: conformal prediction, uncertainty quantification, multi-step forecasting, wireless communications, channel state information

1. Introduction

Recent advances in wireless technologies, such as massive multiple-input multiple-output (MIMO), have significantly improved the efficiency of wireless communication systems. In a MIMO system, multiple antennas are used at the transmitter and receiver so that several signal paths can be exploited simultaneously to improve data rate and reliability. Fully realizing these gains requires accurate channel state information (CSI), a high-dimensional complex-valued representation of the wireless channel response. CSI characterizes the strength and phase of the propagation paths and supports transmission decisions such as beamforming, precoding, and resource allocation. The ability to fully exploit these technologies critically depends on the accurate acquisition of CSI, a high-dimensional complex-valued

representation of the wireless channel response. However, acquiring CSI remains challenging due to the high dimensionality, complex propagation environment, and multi-modality of wireless channels (Björnson et al., 2017).

The conventional approach for CSI acquisition relies on pilot transmission and channel estimation at the receiver. CSI prediction provides an alternative: future channel states are forecast from past observations, helping mitigate channel aging and reduce pilot overhead, which are key bottlenecks in next-generation wireless systems (Stenhammar et al., 2024; Kim et al., 2021). Traditional CSI prediction methods include Kalman filtering (Kim et al., 2021), autoregressive models (Baddour and Beaulieu, 2005), and linear extrapolation (Yin et al., 2020). While analytically tractable, these methods often fail to capture the complex nonlinear dynamics of wireless channels.

Recent deep learning methods, including recurrent neural networks (Jiang and Schotten, 2019), transformer-based architectures (Jin et al., 2025), and diffusion models (Sattari et al., 2025), have significantly improved CSI prediction accuracy. Nevertheless, deep learning-based CSI predictors typically produce point estimates without calibrated uncertainty information. This limits reliability in deployment, since downstream decisions made by the wireless control plane, such as beamforming, scheduling, and link adaptation, depend not only on the predicted channel but also on the confidence associated with that prediction. In the multi-step setting considered here, the predictor outputs several future CSI frames; we refer to each future prediction step as a forecast horizon, or simply a horizon, and to the full sequence as a predicted CSI trajectory.

Conformal prediction (CP) provides a model-agnostic framework for constructing calibrated prediction sets around black-box model outputs (Vovk et al., 2005; Angelopoulos and Bates, 2022). This makes CP an appropriate candidate for uncertainty quantification in CSI prediction: instead of replacing the predictor, CP can wrap an existing CSI prediction model with uncertainty regions. Yet, multi-step CSI prediction poses challenges that standard split CP does not fully address. In other words, the prediction target is a trajectory of high-dimensional complex CSI matrices; errors can grow across the forecast horizon, and different channel-evolution patterns can exhibit varying levels of predictability. Moreover, for a wireless deployment, reliability should be assessed at the trajectory level, where a predicted CSI trajectory may be unreliable if any future CSI frame falls outside its uncertainty region.

Existing CP methods for wireless communications have mainly focused on classification, robust optimization, context-shift calibration, or uncertainty for specific channel prediction pipelines (Cohen et al., 2023b). In parallel, the broader time-series CP literature has developed tools for sequential adaptation, horizon dependence, and multi-step prediction, while conformal risk control (CRC) provides mechanisms for calibrating predictive systems under user-specified losses (Angelopoulos et al., 2022b). However, existing methods do not directly target the combination of matrix-valued CSI prediction, multi-step trajectory-level reliability, and explicit risk control certification considered in this work.

This paper proposes *TRACE-CRC*, a trajectory-adaptive CRC framework for multi-step CSI prediction. It constructs Frobenius-norm uncertainty balls around predicted CSI matrices, adapts their radii using horizon-dependent error profiles and trajectory-level difficulty strata, and certifies trajectory-level risk using Learn-then-Test (LTT). The resulting framework aims to provide calibrated uncertainty estimates for the complete predicted CSI

trajectory rather than only for isolated forecast horizons, shifting the calibration target from marginal per-step coverage to control of the full-trajectory failure event. This trajectory-level perspective supports more informed downstream wireless decisions, where a single unreliable future CSI frame can affect beamforming, scheduling, precoding, or link adaptation across the predicted trajectory. The source code and scripts required to reproduce the experiments are publicly available at <https://github.com/kiarashRezaei/trace-crc>.

The main contributions of this work are summarized as follows:

- We formulate multi-step CSI reliability as a trajectory-level CRC problem, where a predicted CSI trajectory is considered unreliable if at least one future CSI matrix falls outside its uncertainty ball.
- We extend structured multi-step conformal prediction to matrix-valued CSI trajectories by constructing per-horizon Frobenius-norm uncertainty balls whose radii adapt to both horizon-dependent error growth and heterogeneous trajectory difficulty.
- We combine trajectory-adaptive conformal calibration with an LTT risk control layer that selects an uncertainty rule with a finite-sample trajectory-level risk control.
- We evaluate TRACE-CRC against horizon-wise, trajectory-level, online adaptive, weighted, joint multi-step, and conformal risk control baselines. TRACE-CRC achieves trajectory coverage 0.933 at target coverage 0.90, while substantially reducing the conservatism of simultaneous-correction baselines.

2. Related Work

2.1. Conformal Prediction for Marginal Coverage

Classical split conformal methods provide finite-sample marginal coverage under the exchangeability assumption by calibrating a nonconformity score on held-out data (Papadopoulos et al., 2002; Lei et al., 2018). Several extensions improve adaptivity or robustness under more complex data settings. Conformalized quantile regression (CQR) combines quantile regression with conformal calibration to obtain adaptive regression intervals (Romano et al., 2019). Weighted conformal methods address covariate shift and more general forms of non-exchangeability by reweighting calibration residuals (Tibshirani et al., 2019; Barber et al., 2023). These methods form the foundation for uncertainty calibration, but they primarily target marginal prediction-set validity rather than trajectory-level reliability over multiple forecast horizons.

2.2. Conformal Prediction for Time Series and Multi-Step Forecasting

Sequential and time-series prediction often violate the exchangeability assumptions underlying classical CP. This has motivated the development of conformal methods for temporally dependent, nonstationary, and distribution-shifting data. Ensemble batch prediction intervals (EnbPI) constructs bootstrap-based prediction intervals for dynamic time series and updates residuals sequentially during deployment (Xu and Xie, 2021, 2023a). Adaptive conformal inference (ACI) updates the miscoverage level online to maintain long-run coverage

under distribution shift (Gibbs and Candès, 2021), with related adaptive variants, aggregated adaptive conformal inference (AgACI), developed for time-series forecasting (Zaffran et al., 2022). Other approaches exploit temporal structure in residuals or nonconformity scores: sequential predictive conformal inference (SPCI) estimates future residual quantiles using sequential residual quantile regression (Xu and Xie, 2023b), while kernel-based optimally weighted conformal prediction intervals (KOWCPI) uses kernel-based optimally weighted residual quantile regression for non-exchangeable time series (Lee et al., 2025). Multi-step forecasting introduces an additional challenge because uncertainty must be calibrated across multiple forecast horizons. CopulaCPTS models dependence among horizon-wise conformity scores through an empirical copula to obtain valid multi-step uncertainty regions while avoiding overly conservative Bonferroni corrections (Sun and Yu, 2024); related multi-step conformal approaches include ConForME and online multi-step conformal forecasting methods (Galvão Lopes et al., 2024; Wang and Hyndman, 2024). Several of these methods are adapted as baselines in our experiments where adaptive time-series methods are evaluated through separate horizon-wise and trajectory-wise variants, while multi-step methods address dependence across horizons without CRC-style risk certification. TRACE-CRC combines horizon-dependent structure and trajectory-level CRC certification in one framework.

2.3. Conformal Risk Control

Beyond coverage guarantees, conformal risk control (CRC) calibrates prediction sets with respect to user-specified losses. Distribution-free risk-controlling prediction sets provide finite-sample control of expected losses for set-valued predictions (Bates et al., 2021). The LTT framework casts calibration as a multiple-testing problem and provides high-confidence risk control for general predictive algorithms (Angelopoulos et al., 2022a). CRC further specializes this perspective to monotone losses and generalizes split CP from coverage control to broader risk control (Angelopoulos et al., 2022b). Recent work has extended risk control ideas to non-exchangeable data (Farinhas et al., 2024), localized and online guarantees (Zecchin and Simeone, 2024), fairness-aware calibration (Zhang et al., 2024), and application-oriented risk control for model alignment and efficient inference (Overman et al., 2024; Jazbec et al., 2024). A closely related line of work applies CRC to sequential prediction and control, including wireless networking applications (Zecchin et al., 2024). TRACE-CRC instead applies CRC to matrix-valued multi-step CSI prediction with a trajectory-level failure loss.

2.4. Conformal Prediction in Wireless Communications

CP has recently been explored in several wireless communication settings. Standard split CP and cross-validation-based CP have been used for wireless classification tasks such as few-shot demodulation and modulation classification (Cohen et al., 2023a,b). Online CP has been studied for channel prediction in wireless systems (Cohen et al., 2023b). To address the training-deployment mismatch, (Yoo et al., 2026) proposes a meta-learned weighted CP method for context-dependent covariate shift. For decision-oriented wireless optimization, (Su et al., 2025) use split CP to construct channel uncertainty sets for robust beamforming. Most closely related to CSI prediction, (Kim et al., 2026) use CQR within a deep

conformal Bayes filter to obtain calibrated uncertainty for MIMO channel prediction. A broader overview of CP-based calibration for wireless AI is provided by [Simeone et al. \(2025\)](#). Compared with these works, TRACE-CRC differs in both the prediction object and the reliability target. It calibrates Frobenius-norm uncertainty balls for matrix-valued multi-step CSI trajectories and certifies trajectory-level failure risk, rather than focusing on classification, context-shift calibration, robust decision sets, or uncertainty for individual channel predictions.

3. Dataset and Problem Formulation

We study conformal uncertainty quantification for multi-step CSI prediction. In this setting, a pretrained CSI predictor observes past CSI frames and outputs a future CSI trajectory over multiple forecast horizons. Since downstream wireless decisions may depend on the consistency of the predicted channel evolution, our goal is not only to assess point-prediction accuracy, but also to construct Frobenius-norm uncertainty balls around predicted CSI matrices that provide trajectory-level reliability. We first describe the CSI dataset and prediction model, and then formalize the corresponding trajectory-level conformal prediction objective.

3.1. Dataset

We consider CSI trajectories generated from an underlying channel distribution. Informally, this distribution describes how the wireless channel evolves over time under random propagation conditions such as mobility, scattering, and multipath propagation effects. A CSI frame at time t is a complex-valued matrix

$$\mathbf{H}_t \in \mathbb{C}^{N_t \times N_c},$$

where N_t is the antenna dimension and N_c is the subcarrier or frequency dimension. Thus, each CSI frame can be viewed as a high-dimensional matrix-valued observation of the channel at one time instant.

The finite dataset consists of sampled realizations of this channel process

$$\mathcal{D} = \left\{ \left(\mathbf{H}_1^{(i)}, \mathbf{H}_2^{(i)}, \dots, \mathbf{H}_{T_i}^{(i)} \right) \right\}_{i=1}^M,$$

where i indexes trajectories, T_i is the length of trajectory i , t indexes time within a trajectory, and M is the number of trajectories. We use the CSI dataset generation setup from [Sattari et al. \(2025\)](#). The dataset consists of 1,000 independent CSI samples, each with $N_t = 16$ transmit antennas and $N_c = 16$ subcarriers, operating at a carrier frequency of 28 GHz. User velocities are uniformly sampled between 30 and 120 km/h, while the channel model is randomly selected from the 3GPP-compliant CDL models (CDL-A–CDL-E) ([3gp, 2022](#)). Each sample contains 100 time steps (frames), with consecutive CSI frames separated by approximately $33.3 \mu\text{s}$. Since the CSI is complex-valued, the real and imaginary components are stored as separate channels. Consequently, each sample is represented as a tensor of shape $100 \times 2 \times 16 \times 16$, where the first dimension corresponds to the 100 time steps, the second dimension represents the real and imaginary components of the CSI,

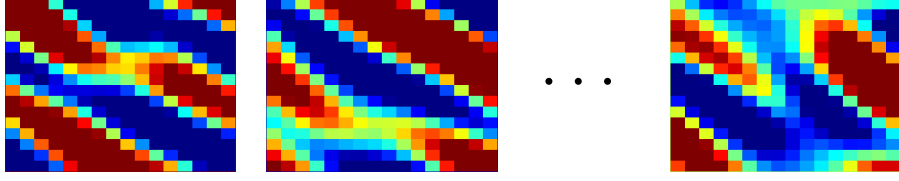


Figure 1: Snapshots of the real part of a CSI frame at different time steps.

and the remaining two dimensions correspond to the antenna and subcarrier dimensions, respectively.

The exchangeability assumption applies across complete CSI trajectories, not across frames within a trajectory. Hence, temporal dependence and channel evolution within each trajectory are allowed. In our simulation, trajectories are generated independently using the same channel-generation procedure and parameter distributions, making trajectory-level exchangeability reasonable in this setting. Figure 1 shows example snapshots from one trajectory by visualizing the real part of the CSI matrices at different time horizons. The variation across snapshots illustrates the temporal evolution of the wireless channel.

3.2. CSI Prediction

Time-varying CSI prediction can be viewed as a matrix-valued time-series forecasting problem. Let $\mathbf{H}_p$ denote the past CSI frames used as input, and let $\mathbf{H}_f$ denote the future CSI frames to be predicted

$$\mathbf{H}_p = (\mathbf{H}_{t-N_p+1}, \mathbf{H}_{t-N_p+2}, \dots, \mathbf{H}_t), \quad \mathbf{H}_f = (\mathbf{H}_{t+1}, \mathbf{H}_{t+2}, \dots, \mathbf{H}_{t+N_f}).$$

Here, $\mathbf{H}_p$ and $\mathbf{H}_f$ are random variables induced by the underlying channel distribution, while the dataset in Subsection 3.1 contains sampled realizations of these quantities.

Let $\mathcal{G}$ denote the class of admissible predictors mapping past CSI trajectories to future CSI trajectories. Under the mean squared error criterion, the population-optimal predictor $g^*(\mathbf{H}_p)$ is

$$g^*(\mathbf{H}_p) = \arg \min_{g \in \mathcal{G}} \mathbb{E} [\|\mathbf{H}_f - g(\mathbf{H}_p)\|^2]. \quad (1)$$

The optimal prediction rule is the conditional mean estimator

$$g^*(\mathbf{H}_p) = \mathbb{E} [\mathbf{H}_f | \mathbf{H}_p] = \int \mathbf{h}_f \mathbb{P}(\mathbf{h}_f | \mathbf{H}_p) d\mathbf{h}_f. \quad (2)$$

In practice, computing the minimum mean square error (MMSE) predictor is challenging because the conditional law of $\mathbf{H}_f$ given $\mathbf{H}_p$, denoted by $\mathbb{P}(\mathbf{H}_f | \mathbf{H}_p)$, is unknown and high-dimensional. A deep learning-based predictor g_θ trained with mean square error (MSE) can be viewed as a data-driven approximation to this conditional mean. However, deterministic predictors do not explicitly capture the stochastic and potentially multi-modal evolution of future CSI trajectories.

Recent diffusion-based CSI predictors address this limitation by modeling future CSI generation probabilistically. Following Sattari et al. (2025), we use a pretrained predictor

based on a temporal encoder and diffusion generator. The temporal encoder extracts representations from the past CSI frames, while the diffusion generator produces future CSI samples. At inference time, the model is used autoregressively: each predicted CSI frame is fed back as input for predicting the next frame.

For a sampled trajectory i , the past CSI input is $\mathbf{H}_p^{(i)} = (\mathbf{H}_{t-N_p+1}^{(i)}, \dots, \mathbf{H}_t^{(i)})$. Let g_θ denote the trained diffusion-based CSI predictor, where θ represents the learned parameters of the temporal encoder and diffusion generator. At inference time, g_θ generates a predicted future CSI trajectory from the past CSI window

$$\hat{\mathbf{H}}_f^{(i)} = g_\theta(\mathbf{H}_p^{(i)}) = (\hat{\mathbf{H}}_{t+1}^{(i)}, \dots, \hat{\mathbf{H}}_{t+N_f}^{(i)}).$$

While the autoregressive scheme provides a flexible prediction framework, it also induces horizon-dependent uncertainty. Prediction errors can propagate through the feedback loop and accumulate over future steps, so the reliability of the predicted CSI trajectory may deteriorate as the forecast horizon increases. This motivates the conformal uncertainty quantification problem considered next.

3.3. Conformal Problem Formulation

Building on the multi-step CSI prediction setup in Subsection 3.2, we now formulate trajectory-level conformal uncertainty quantification. Each future index $j \in \{1, \dots, N_f\}$ is a forecast horizon, and the collection of all N_f future CSI frames forms the predicted CSI trajectory.

For trajectory i and forecast horizon j , we define the Frobenius residual as

$$E_{i,j} = \left\| \mathbf{H}_{t+j}^{(i)} - \hat{\mathbf{H}}_{t+j}^{(i)} \right\|_F, \quad j = 1, \dots, N_f. \quad (3)$$

The Frobenius residual is well-suited for CSI uncertainty quantification because each CSI frame is a high-dimensional complex-valued matrix whose entries jointly describe the channel response across the spatial-frequency domain. Thus, a Frobenius-norm radius defines an uncertainty ball around the predicted CSI matrix in the full spatial-frequency channel space

$$\mathcal{B}_{i,j}(r_{i,j}) = \left\{ \mathbf{G} \in \mathbb{C}^{N_t \times N_c} : \left\| \mathbf{G} - \hat{\mathbf{H}}_{t+j}^{(i)} \right\|_F \leq r_{i,j} \right\}. \quad (4)$$

Here, $\mathbf{G}$ is a generic candidate CSI matrix in the same spatial-frequency space as $\mathbf{H}_{t+j}^{(i)}$. The ball is centered at the predicted CSI frame $\hat{\mathbf{H}}_{t+j}^{(i)}$. The radius $r_{i,j}$ may depend on the forecast horizon and on quantities computed from the past CSI input and the predicted future CSI trajectory, but not on the unobserved true future CSI frame.

Let $A_{i,j}$ denote the event that trajectory i is covered at forecast horizon j

$$A_{i,j} = \left\{ \mathbf{H}_{t+j}^{(i)} \in \mathcal{B}_{i,j}(r_{i,j}) \right\} = \{E_{i,j} \leq r_{i,j}\}. \quad (5)$$

Thus, membership in the Frobenius ball is equivalent to the residual being no larger than the assigned radius.

Let $\alpha_h \in (0, 1)$ denote the target horizon-wise miscoverage level. A horizon-wise objective marginally controls these events.

$$\mathbb{P}(A_{i,j}) \geq 1 - \alpha_h, \quad j = 1, \dots, N_f. \quad (6)$$

However, marginal horizon-wise coverage does not generally imply that the entire predicted CSI trajectory is covered. We therefore define the trajectory-level coverage event as $\mathcal{C}_i^{\text{traj}} = \bigcap_{j=1}^{N_f} A_{i,j}$. Its complement corresponds to at least one uncovered horizon

$$\left(\mathcal{C}_i^{\text{traj}}\right)^c = \{\exists j \in \{1, \dots, N_f\} : E_{i,j} > r_{i,j}\}.$$

The population trajectory-level risk is

$$\mathcal{R}_{\text{traj}} = \mathbb{P}\left(\left(\mathcal{C}_i^{\text{traj}}\right)^c\right) = \mathbb{P}(\exists j \in \{1, \dots, N_f\} : E_{i,j} > r_{i,j}). \quad (7)$$

Equivalently, the trajectory failure indicator is defined as

$$L_i^{\text{traj}} = \mathbf{1}\{\exists j \in \{1, \dots, N_f\} : E_{i,j} > r_{i,j}\}. \quad (8)$$

Since the population trajectory-level risk ($\mathcal{R}_{\text{traj}}$) is unknown, we formulate the goal as a CRC problem. The calibration procedure returns a data-dependent uncertainty rule $\hat{\mathcal{B}}$, which assigns radii $\hat{r}_{i,j}$ using only the past CSI input and the predicted future CSI trajectory. Under trajectory-level exchangeability between the data used for risk control selection and future deployment trajectories, we seek a procedure such that

$$\mathbb{P}\left(\mathcal{R}_{\text{traj}}(\hat{\mathcal{B}}) \leq \alpha\right) \geq 1 - \delta. \quad (9)$$

Here, α is the target trajectory failure level, δ is the allowed probability of certifying an uncertainty rule whose trajectory risk exceeds α , and the outer probability is over the data used by the CRC procedure. The guarantee requires exchangeability across complete CSI trajectories, as discussed in Section 3.1.

4. TRACE-CRC: Trajectory-Adaptive Conformal Risk Control

4.1. Method Overview

The trajectory-level objective in (9) requires a rule that the corresponding uncertainty ball cover the full future CSI trajectory, rather than isolated forecast horizons. A single-radius trajectory-level conformal method can be inefficient because it applies the same radius across all horizons, while copula-based approaches explicitly model dependence across horizons, as in CopulaCPTS (Sun and Yu, 2024). TRACE-CRC adopts a different approach: it avoids explicit dependence modeling and instead adapts the radius to both horizon profile and predicted trajectory difficulty.

For trajectory i assigned to difficulty group $g_i \in \{0, 1\}$, TRACE-CRC sets the radius at forecast horizon j as

$$r_{i,j}^* = r_{g_i,j}^* = \lambda^* q_{g_i} w_j, \quad j = 1, \dots, N_f, \quad (10)$$

where w_j is the horizon difficulty profile, $\{q_0, q_1\}$ are the two group-wise conformal quantiles, q_{g_i} selects the quantile for the assigned group, and λ^* is the global multiplier selected by LTT risk control.

The components w_j , $\{q_g\}_{g \in \{0,1\}}$, and λ^* are estimated using disjoint calibration splits. TRACE-CRC then returns the Frobenius uncertainty ball $\mathcal{B}_{i,j}^*$ in (4) with radius $r_{i,j}^*$. A future CSI trajectory is covered if every true future frame lies inside its corresponding horizon-wise uncertainty ball.

4.2. Three-Way Calibration Split

After splitting the available trajectory samples into calibration and test sets, TRACE-CRC further partitions the calibration CSI trajectories into three disjoint index subsets,

$$\mathcal{D}_{\text{cal}} = \mathcal{D}_{\text{prof}} \dot{\cup} \mathcal{D}_{\text{cp}} \dot{\cup} \mathcal{D}_{\text{val}}.$$

Thus, $i \in \mathcal{D}_{\text{prof}}$, $i \in \mathcal{D}_{\text{cp}}$, or $i \in \mathcal{D}_{\text{val}}$ indicates that trajectory i is assigned to the corresponding stage. For each calibration trajectory i , we obtain the multi-step prediction from the predictor described in Subsection 3.2 and compute the Frobenius residuals according to (3). These residuals provide the calibration scores used by TRACE-CRC.

The profile subset $\mathcal{D}_{\text{prof}}$ is used to estimate the horizon difficulty profile w_j and to fit the trajectory difficulty model that assigns each trajectory i to a group $g_i \in \{0,1\}$. The conformal calibration subset $\mathcal{D}_{\text{cp}}$ is used to compute the group-wise conformal quantiles q_g . Finally, the validation subset $\mathcal{D}_{\text{val}}$ is reserved for LTT risk control selection of the global multiplier λ^* . Because the three subsets are disjoint, the horizon profile, grouping rule, and group-wise quantiles are fixed before the LTT hypothesis tests are applied on $\mathcal{D}_{\text{val}}$. This separation keeps the profiling, conformal calibration, and LTT testing stages statistically distinct, simplifying the validity argument. However, it may reduce sample efficiency when calibration trajectories are scarce. Cross-fitted or cross-conformal variants could improve data reuse, although retaining the same LTT risk-control guarantee would require additional analysis.

4.3. Horizon Difficulty Profile

Prediction errors in multi-step CSI forecasting are horizon dependent, often increasing or changing shape across future horizons. TRACE-CRC therefore estimates a relative horizon difficulty profile $w_1, \dots, w_{N_f}$ using the profile split $\mathcal{D}_{\text{prof}}$.

For each forecast horizon j , we first compute a raw horizon difficulty estimate as the empirical upper-quantile residual

$$w_j^{\text{raw}} = Q_{1-\alpha_h}(\{E_{i,j} : i \in \mathcal{D}_{\text{prof}}\}).$$

where $Q_{1-\alpha_h}$ denotes the empirical $(1 - \alpha_h)$ -quantile. The parameter α_h is used only to estimate the relative horizon difficulty profile and is distinct from the final trajectory-level risk target α . Since horizon-wise empirical quantiles can fluctuate under finite calibration data, we smooth the raw profile by a local moving average. Specifically,

$$\tilde{w}_j = \frac{1}{|\mathcal{N}_K(j)|} \sum_{k \in \mathcal{N}_K(j)} w_k^{\text{raw}},$$

where $\mathcal{N}_K(j) = \{k \in \{1, \dots, N_f\} : |k - j| \leq \lfloor K/2 \rfloor\}$. In experiments, $K = 3$, so each horizon is smoothed using itself and its immediate neighbors, with boundary horizons averaged over the available terms. We then stabilize the profile by imposing the lower floor $c_{\min} = \rho \cdot \text{median}(\tilde{w}_1, \dots, \tilde{w}_{N_f})$, where $\rho > 0$ is a stabilization parameter. The final profile is normalized to unit mean

$$w_j = \frac{\max(\tilde{w}_j, c_{\min})}{N_f^{-1} \sum_{\ell=1}^{N_f} \max(\tilde{w}_\ell, c_{\min})}. \quad (11)$$

The normalized profile satisfies $N_f^{-1} \sum_{j=1}^{N_f} w_j = 1$. Values $w_j > 1$ correspond to harder-than-average forecast horizons, while $w_j < 1$ correspond to easier horizons.

4.4. Trajectory Difficulty Stratification

TRACE-CRC stratifies predicted CSI trajectories according to estimated prediction difficulty. The goal is to learn a simple mapping from a trajectory-level feature vector $\mathbf{x}_i$, computed from the predicted trajectory, to a predicted difficulty score $\hat{d}_i$. The resulting scores are used to form lower- and higher-difficulty groups.

For each profile index $i \in \mathcal{D}_{\text{prof}}$, we define

$$z_j^{(i)} = \left\| \hat{\mathbf{H}}_{t+j}^{(i)} \right\|_F, \quad j = 1, \dots, N_f,$$

and construct the trajectory-level feature vector

$$\mathbf{x}_i = \left[\mu_i, \sigma_i, z_{\max}^{(i)}, z_{\min}^{(i)}, \text{range}_i, \text{slope}_i, \text{TV}_i, \text{curvature}_i, z_1^{(i)}, z_{N_f}^{(i)} \right]. \quad (12)$$

Here, μ_i , σ_i , $z_{\max}^{(i)}$, and $z_{\min}^{(i)}$ are the mean, standard deviation, maximum, and minimum of the sequence $\{z_j^{(i)}\}_{j=1}^{N_f}$, with $\text{range}_i = z_{\max}^{(i)} - z_{\min}^{(i)}$ and $\text{slope}_i = z_{N_f}^{(i)} - z_1^{(i)}$. The term TV_i denotes trajectory variation, defined as the mean absolute first-order difference of the predicted norm sequence, while curvature_i denotes the mean absolute second-order difference:

$$\text{TV}_i = \frac{1}{N_f - 1} \sum_{j=1}^{N_f-1} \left| z_{j+1}^{(i)} - z_j^{(i)} \right|, \quad \text{curvature}_i = \frac{1}{N_f - 2} \sum_{j=2}^{N_f-1} \left| z_{j+1}^{(i)} - 2z_j^{(i)} + z_{j-1}^{(i)} \right|.$$

These features summarize the magnitude, variability, trend, and local smoothness of the predicted CSI trajectory.

For the same profile index $i \in \mathcal{D}_{\text{prof}}$, where the true future CSI is available for calibration, we define the residual-based trajectory difficulty target as

$$d_i = \max_{1 \leq j \leq N_f} \frac{E_{i,j}}{w_j}. \quad (13)$$

This score measures how difficult trajectory i is after accounting for the typical error scale at each forecast horizon. A large value of d_i indicates that the trajectory contains at least one unusually large residual relative to the expected horizon difficulty.

We then fit a ridge regression model $\hat{d}(\mathbf{x}) = \beta_0 + \beta^\top \mathbf{x}$ on $\{(\mathbf{x}_i, d_i) : i \in \mathcal{D}_{\text{prof}}\}$, using regularization parameter η . We use ridge regression to keep the stratification model low-variance and stable under limited calibration data. Afterwards, we apply it to the trajectories indexed by $\mathcal{D}_{\text{cp}}$. For each $\ell \in \mathcal{D}_{\text{cp}}$, this gives a predicted difficulty score $\hat{d}_\ell = \hat{d}(\mathbf{x}_\ell)$. For two strata, we set $\tau = \text{median}\{\hat{d}_\ell : \ell \in \mathcal{D}_{\text{cp}}\}$ and assign each conformal calibration trajectory by $g_\ell = \mathbf{1}\{\hat{d}_\ell > \tau\}$. Thus, $g_\ell = 0$ denotes the lower-difficulty group and $g_\ell = 1$ denotes the higher-difficulty group. The same fitted model and threshold are used to assign any validation or test trajectory i to a group,

$$g_i = \mathbf{1}\{\hat{d}(\mathbf{x}_i) > \tau\}. \quad (14)$$

4.5. Group-wise Conformal Calibration

Given the horizon profile in (11) and the group assignment rule in (14), TRACE-CRC calibrates a conformal scale within each difficulty stratum. For each conformal calibration index $\ell \in \mathcal{D}_{\text{cp}}$, where the true future CSI is available, we compute the same trajectory difficulty score defined in (13),

$$d_\ell = \max_{1 \leq j \leq N_f} \frac{E_{\ell,j}}{w_j}.$$

On $\mathcal{D}_{\text{prof}}$, this score was used to fit the trajectory difficulty model; on $\mathcal{D}_{\text{cp}}$, it is used as the conformal calibration score. For each group $g \in \{0, 1\}$, let

$$\mathcal{S}_g = \{d_\ell : \ell \in \mathcal{D}_{\text{cp}}, g_\ell = g\}, \quad n_g = |\mathcal{S}_g|.$$

The group-wise conformal quantile q_g is

$$q_g = Q_{1-\alpha_{\text{cp}}}^+(\mathcal{S}_g),$$

where $Q_{1-\alpha_{\text{cp}}}^+$ denotes the split-conformal empirical quantile at level $\lceil (n_g + 1)(1 - \alpha_{\text{cp}}) \rceil / n_g$. If $\lceil (n_g + 1)(1 - \alpha_{\text{cp}}) \rceil > n_g$, we set $q_g = +\infty$. Here, α_{cp} is the group-wise conformal calibration level. Thus, q_g is computed from the residual-based calibration scores of trajectories assigned to group g .

For a generic group $g \in \{0, 1\}$, the preliminary horizon-dependent radius is

$$r_{g,j} = q_g w_j, \quad j = 1, \dots, N_f. \quad (15)$$

Thus a trajectory i assigned to group g_i uses $r_{g_i,j} = q_{g_i} w_j$. Since $d_\ell \leq q_g$ implies $E_{\ell,j} \leq q_g w_j$ for every horizon j , q_g controls the trajectory-level scale within each difficulty group, while w_j distributes that scale across forecast horizons.

4.6. Learn-then-Test Risk Control

The group-wise conformal calibration step produces preliminary radii $q_g w_j$. TRACE-CRC then applies LTT risk control to select a global multiplier λ for these radii. Let $\Lambda = \{\lambda_1, \dots, \lambda_m\}$ be a finite grid of candidate multipliers. For a validation index $i \in \mathcal{D}_{\text{val}}$, let $g_i \in \{0, 1\}$ denote its assigned difficulty group. For each $\lambda \in \Lambda$, the candidate radius at horizon j is

$$r_{i,j}(\lambda) = r_{g_i,j}(\lambda) = \lambda q_{g_i} w_j.$$

On the validation subset, we evaluate each candidate multiplier using the trajectory-level failure loss $L_i(\lambda) = \mathbf{1} \{ \exists j \in \{1, \dots, N_f\} : E_{i,j} > \lambda q_{g_i} w_j \}$. Thus, $L_i(\lambda) = 1$ if at least one future CSI frame in trajectory i falls outside its corresponding Frobenius uncertainty ball. The empirical validation risk is

$$\widehat{\mathcal{R}}_{\text{traj}}(\lambda) = \frac{1}{n_{\text{val}}} \sum_{i \in \mathcal{D}_{\text{val}}} L_i(\lambda), \quad n_{\text{val}} = |\mathcal{D}_{\text{val}}|,$$

and the corresponding population trajectory risk for a fresh CSI trajectory sample is $\mathcal{R}_{\text{traj}}(\lambda) = \mathbb{P}(L_i(\lambda) = 1)$. For each candidate multiplier, TRACE-CRC tests

$$H_0(\lambda) : \mathcal{R}_{\text{traj}}(\lambda) > \alpha \quad \text{against} \quad H_1(\lambda) : \mathcal{R}_{\text{traj}}(\lambda) \leq \alpha,$$

where α is the target trajectory failure level. We test this hypothesis using a one-sided Hoeffding–Bentkus (HB) p -value

$$p(\lambda) = p_{\text{HB}} \left(\widehat{\mathcal{R}}_{\text{traj}}(\lambda), n_{\text{val}}, \alpha \right)$$

(Bentkus, 2004). Let $\mathcal{A} \subseteq \Lambda$ be the set of accepted multipliers after correction. TRACE-CRC selects $\lambda^* = \min_{\lambda \in \mathcal{A}} \lambda$, which gives the smallest certified multiplier among the accepted candidates. If $\mathcal{A} = \emptyset$, no multiplier in the candidate grid is certified; in this case, the grid must be enlarged, or the method returns no certified radius on the grid.

For trajectory i , the final radius at horizon j is

$$r_{i,j}^* = r_{g_i,j}^* = \lambda^* q_{g_i} w_j. \quad (16)$$

4.7. Prediction Bands and Risk Control Guarantee

For a new CSI context, TRACE-CRC first computes the trajectory-level feature vector $\mathbf{x}_i$ from the predicted future CSI trajectory, as defined in (12). It then assigns the trajectory to a difficulty group g_i using (14) and applies the final radius $r_{i,j}^*$ from (16) at each forecast horizon. The resulting trajectory-level prediction band is the collection of horizon-wise Frobenius balls

$$\widehat{\mathcal{B}}_i^* = \{ \mathcal{B}_{i,j}(r_{i,j}^*) \}_{j=1}^{N_f},$$

where each ball is centered at the corresponding predicted CSI frame $\widehat{\mathbf{H}}_{t+j}^{(i)}$, as defined in (4).

Under trajectory-level exchangeability between the validation losses used by LTT and future deployment losses, and with FWER controlled at level δ , the selected rule $\widehat{\mathcal{B}}^*$ satisfies the CRC guarantee in (9), certifying trajectory-level risk at level α . As discussed in Section 3.1, this assumption concerns exchangeability across complete trajectories and does not require independence among frames within a trajectory.

5. Experimental Setup

We evaluate TRACE-CRC on the multi-step CSI prediction task using the pretrained diffusion-based CSI predictor of Sattari et al. (2025); see Sections 3.1 and 3.2 for details. The predictor was trained on 10,000 trajectory-level samples following the training procedure

described in Sattari et al. (2025). We use an inference SNR of 20 dB and 20 reverse-diffusion sampling steps. TRACE-CRC is applied post hoc to the pretrained CSI predictor. Its calibration is performed offline, while deployment requires only evaluation of the calibrated scaling rule and uncertainty radii. The online computational cost is therefore expected to be small relative to diffusion inference. For the complex-valued CSI matrices, all residuals and uncertainty radii are computed using the Frobenius norm. From each 100-frame CSI sample described in Section 3.1, we extract a 30-frame trajectory for prediction. The predictor uses $N_p = 10$ past frames as input and predicts $N_f = 20$ future frames. Thus, conformal calibration and evaluation are performed over a forecast horizon of $N_f = 20$.

At inference time, the 1,000 prediction trajectories are partitioned using a fixed random holdout. We assign 30% of the trajectories to calibration and retain 70% for testing, yielding $n_{\text{cal}} = 300$ and $n_{\text{test}} = 700$. This split is shared by all conformal methods and ensures that no CSI frames from the same trajectory are assigned to both calibration and test sets. We use a relatively small calibration set to evaluate TRACE-CRC in a limited-calibration regime while retaining a large independent test set for stable estimation of trajectory-level performance. Since diffusion inference is stochastic, the complete evaluation is repeated over five random inference seeds while keeping the same calibration and test split, and results are aggregated across runs. For online adaptive methods, test trajectories are processed in a fixed dataset-index order for reproducibility; this order does not represent temporal progression.

The calibration trajectories are further divided into three disjoint subsets required by TRACE-CRC: the profile subset $\mathcal{D}_{\text{prof}}$, the conformal calibration subset $\mathcal{D}_{\text{cp}}$, and the validation subset $\mathcal{D}_{\text{val}}$. We allocate 10% of the calibration trajectories to profile learning, 15% of the remaining calibration trajectories to group-wise conformal calibration, and the rest to LTT validation, resulting in $n_{\text{prof}} = 30$, $n_{\text{cp}} = 40$, and $n_{\text{val}} = 230$. We allocate the largest calibration split to LTT validation because the risk-control certificate depends directly on validation evidence. We set $\alpha = 0.10$, corresponding to a target trajectory coverage of 0.90, and use $\alpha_{\text{cp}} = \alpha_h = 0.10$ for group-wise conformal calibration and horizon-profile estimation. For LTT, we set $\delta = 0.10$, corresponding to a 90% confidence level for the risk-control certificate. These values define a representative operating point and are kept fixed across methods to enable a controlled comparison of their calibration strategies. The LTT candidate family consists of 27 prespecified multipliers, including 10 uniformly spaced on $[0.70, 1.30]$ and additional candidates on $[1.40, 3.00]$. The λ -grid allows both mild deflation and inflation of the preliminary conformal radii. Additional method-specific parameters are reported in Appendix B.

5.1. Baseline Methods

We compare TRACE-CRC with two groups of methods: conformal baselines from the literature and ablations of the proposed conformal risk control framework. Since the methods from the literature do not all target the same formal guarantee, we organize them into four categories: standard split-conformal baselines, time-series adaptive conformal baselines, structured multi-step baselines, and risk control ablations. Within each category, when applicable, we distinguish horizon-wise variants from trajectory-level variants according to whether calibration targets individual forecast steps or the full CSI trajectory. All meth-

ods are evaluated on the same held-out test trajectories using the reliability and efficiency metrics defined in Section 6.

Standard split-conformal baselines. We first include standard split-conformal regression baselines (Papadopoulos et al., 2002; Lei et al., 2018). Global Residual Conformal pools all calibration residuals $E_{i,j}$ across samples and horizons and uses a single radius $R_{i,j} = q$. Horizon-wise conformal instead computes a separate quantile q_j for each forecast horizon, yielding $R_{i,j} = q_j$. To represent trajectory-level split conformal calibration, we include Max-score split conformal, which uses the nonconformity score $S_i = \max_{1 \leq j \leq N_f} E_{i,j}$ and applies the resulting quantile $q_{\max}$ uniformly across all horizons. We also include a residual-quantile conformal baseline inspired by conformalized quantile regression (Romano et al., 2019). This baseline estimates a horizon-dependent residual quantile profile and then conformalizes it using the maximum residual excess over that profile.

Time-series adaptive conformal baselines. To compare against adaptive conformal methods designed for temporally ordered or nonstationary settings, we include exponentially weighted (EW) conformal variants. EW-Horizon conformal and EW-Trajectory conformal replace the unweighted empirical quantile with a recency-weighted quantile, assigning larger weights to more recent residuals. These baselines are motivated by conformal prediction under distribution shift and non-exchangeability (Tibshirani et al., 2019; Barber et al., 2023; Wang and Hyndman, 2024).

We also include EnbPI (Xu and Xie, 2021), ACI (Gibbs and Candès, 2021), and AgACI (Zaffran et al., 2022) as established adaptive online conformal time-series baselines. For each method, we report both a horizon and a trajectory variant. The horizon variant calibrates separately at each prediction step j , while the trajectory variant first reduces each predicted CSI trajectory to a scalar score, typically $\max_{1 \leq j \leq N_f} E_{i,j}$, and then calibrates this trajectory-level score. This distinction is important because horizon variants primarily target per-horizon reliability, whereas trajectory variants are more directly aligned with the trajectory-level objective considered in this work.

Structured multi-step baselines. We include two CRC baselines, Bonferroni-CRC and Sidak-CRC, which use classical multi-step corrections. Both apply simultaneous inference corrections across forecast horizons, using the Bonferroni (Bonferroni, 1936) and Sidak (Šidák, 1967) adjustments, respectively. Such corrections are commonly used to convert marginal intervals into simultaneous multi-step prediction intervals (Ravishanker et al., 1991). We also compare against CopulaCPTS, a joint multi-step conformal forecasting method that accounts for dependence across forecast horizons through an empirical copula construction (Sun and Yu, 2024).

risk control ablations. Finally, we include ablations of the proposed risk control framework to isolate the contribution of each component. Table 1 summarizes which methods use horizon profiling w_j , trajectory group-wise conformal quantiles q_g , and LTT risk control calibration (Bates et al., 2021; Angelopoulos et al., 2022b). Global-CRC uses only the LTT multiplier λ^* . Horizon-Profile CRC adds w_j , Trajectory-Stratified CRC adds q_g , and TRACE-CRC combines all three components. These ablations test whether LTT risk control alone is sufficient, or whether efficiency improves when risk control is combined with horizon profiling and trajectory stratification.

Table 1: Ablation structure for TRACE-CRC components.

Method	w_j	q_g	λ^*
Global-CRC	×	×	✓
Horizon-Profile CRC	✓	×	✓
Trajectory-Stratified CRC	×	✓	✓
TRACE-CRC	✓	✓	✓

6. Results and Discussion

We evaluate all methods on the same held-out CSI test trajectories. Our primary reliability metric is trajectory coverage (TC), and our efficiency metric is the horizon-wise average Frobenius radius AFR_j :

$$\text{TC} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \mathbf{1}\{\forall j \in \{1, \dots, N_f\} : E_{i,j} \leq r_{i,j}\}, \quad \text{AFR}_j = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} r_{i,j}.$$

We also report horizon-wise coverage HC_j and summarize it using mean horizon coverage (MHC) and worst-horizon coverage (WHC):

$$\text{HC}_j = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \mathbf{1}\{E_{i,j} \leq r_{i,j}\}, \quad \text{MHC} = \frac{1}{N_f} \sum_{j=1}^{N_f} \text{HC}_j, \quad \text{WHC} = \min_{1 \leq j \leq N_f} \text{HC}_j.$$

Figures show the horizon-wise coverage HC_j and average Frobenius radius AFR_j , averaged over five random diffusion inference seeds; shaded regions denote one standard deviation across seeds. For each seed, MHC is the mean of HC_j across horizons, WHC is its minimum, and AFR is the mean of AFR_j . Tables report the mean $\pm$ standard deviation of these per-seed metrics. MHC, WHC, and TC measure reliability relative to the target coverage $1 - \alpha = 0.90$, while AFR measures efficiency. Table 2 presents representative methods from each comparison category; complete results and tuning parameters are reported in Appendices A and B.

Horizon-wise methods produce compact uncertainty regions but insufficient trajectory coverage, with TC values of 0.710 for Horizon-wise conformal and 0.744 for AgACI-Horizon. Methods that explicitly target full-trajectory reliability generally improve TC but require larger radii; for example, EnbPI-Trajectory and Bonferroni-CRC attain TC values of 0.919 and 0.968, with AFR values of 21.80 and 20.00, respectively. TRACE-CRC achieves $\text{TC} = 0.933 \pm 0.011$ and $\text{AFR} = 13.64 \pm 0.41$, exceeding the 0.90 target while remaining more efficient than the trajectory-level and simultaneous representatives. It also uses smaller radii than Residual Quantile and CopulaCPTS, while achieving higher trajectory coverage.

Standard split-conformal baselines. Figure 2 compares the standard split-conformal baselines with TRACE-CRC. Global Residual Conformal provides high coverage at early horizons but degrades sharply over the forecast path, resulting in low trajectory coverage. Horizon-wise Conformal produces the most compact uncertainty balls, but its horizon-wise coverage remains mostly below the target and its TC is only 0.710. Residual Quantile Conformal better captures horizon-dependent error growth and reaches the trajectory target,

Table 2: Representative conformal methods for CSI uncertainty under the fixed random trajectory-level split. The target trajectory coverage is 0.90. Results are reported as mean $\pm$ standard deviation across five diffusion inference seeds. For each seed, MHC is the mean coverage across forecast horizons, WHC is the worst-horizon coverage, TC is full-trajectory coverage, and AFR is the mean Frobenius radius across forecast horizons and test trajectories.

Category	Method	MHC	WHC	TC	AFR
Standard split-conformal	Horizon-wise	0.872 ± 0.002	0.855 ± 0.006	0.710 ± 0.009	10.73 ± 0.06
	Residual Quantile	0.965 ± 0.002	0.939 ± 0.003	0.903 ± 0.007	14.30 ± 0.21
Time-series adaptive	AgACI-Horizon	0.896 ± 0.001	0.891 ± 0.001	0.744 ± 0.006	11.49 ± 0.06
	EnbPI-Trajectory	0.984 ± 0.001	0.937 ± 0.007	0.919 ± 0.008	21.80 ± 0.24
Structured multi-step	Bonferroni-CRC	0.993 ± 0.001	0.985 ± 0.003	0.968 ± 0.004	20.00 ± 0.29
	CopulaCPTS	0.969 ± 0.004	0.945 ± 0.004	0.892 ± 0.012	14.87 ± 0.29
Risk control / ablation	Global-CRC	0.992 ± 0.001	0.967 ± 0.005	0.957 ± 0.007	24.37 ± 0.54
	TRACE-CRC	0.978 ± 0.004	0.962 ± 0.006	0.933 ± 0.011	13.64 ± 0.41

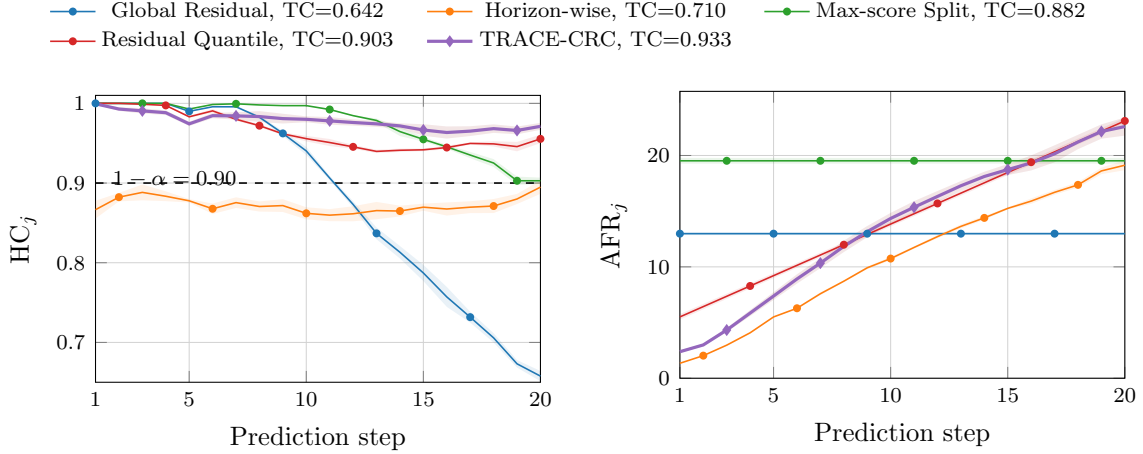


Figure 2: Horizon-wise coverage HC_j (left) and average Frobenius radius AFR_j (right) for standard split-conformal baselines and TRACE-CRC. Curves show means over five diffusion inference seeds, and shaded regions indicate ± 1 standard deviation across seeds. The dashed line marks the target coverage $1 - \alpha = 0.90$; legend entries report trajectory coverage (TC).

with $TC = 0.903$, although at a larger average radius than TRACE-CRC. Max-score Split Conformal directly calibrates a trajectory-level score, but produces substantially wider balls while still falling below the target, with $TC = 0.882$. TRACE-CRC achieves the strongest reliability–efficiency trade-off in this group, attaining $TC = 0.933$ with smaller radii than both trajectory-oriented split-conformal baselines.

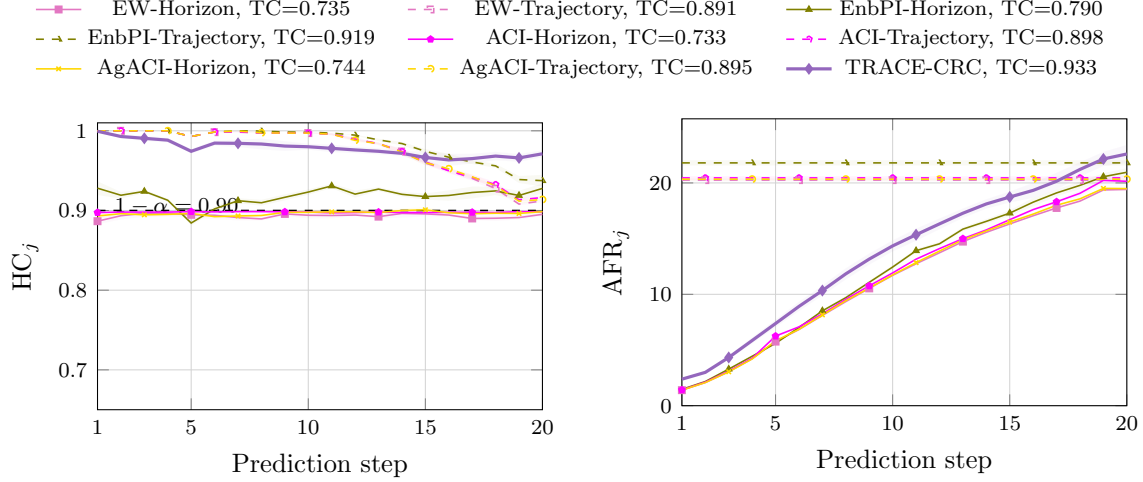


Figure 3: Horizon-wise coverage and average Frobenius radius for time-series adaptive conformal baselines and TRACE-CRC. Dashed curves denote trajectory-level variants, solid curves denote horizon-wise variants, and the diamond-marked curve denotes TRACE-CRC.

Time-series adaptive conformal baselines. Figure 3 compares TRACE-CRC with adaptive time-series conformal baselines. Horizon-level adaptive methods remain relatively efficient, with AFR values between 11.40 and 12.13, but their trajectory coverage is well below the target, with TC ranging from 0.733 to 0.790. Trajectory-level variants improve full-path reliability, reaching TC values between 0.891 and 0.919, but their AFRs increase substantially to 20.25–21.80. Their horizon-wise coverage is high at early prediction steps but generally decreases toward later horizons, reflecting the increasing difficulty of long-range prediction. TRACE-CRC provides higher trajectory coverage than all adaptive baselines while remaining substantially more efficient than their trajectory-level variants. These results indicate that temporal adaptation alone is insufficient for efficient trajectory-level reliability.

Structured multi-step baselines. Figure 4 compares TRACE-CRC with structured multi-step baselines. Bonferroni-CRC and Sidak-CRC achieve high trajectory coverage, with $TC = 0.968$, but their simultaneous corrections are highly conservative, yielding an average AFR of 20.00 and sharp radius increases at several horizons. Their curves coincide because both corrections select the same split-conformal order statistic in this finite-sample setting. CopulaCPTS is less conservative, with $AFR = 14.87$, but remains below the trajectory target at $TC = 0.892$; its horizon-wise coverage also exhibits greater variability across diffusion inference seeds at several prediction steps. TRACE-CRC attains $TC = 0.933$ with the smallest average radius in this group, $AFR = 13.64$, providing a more favorable reliability–efficiency trade-off than the structured baselines.

Risk control ablations. Figure 5 examines how horizon profiling and trajectory stratification affect the reliability–efficiency trade-off within the risk-control framework. Global-

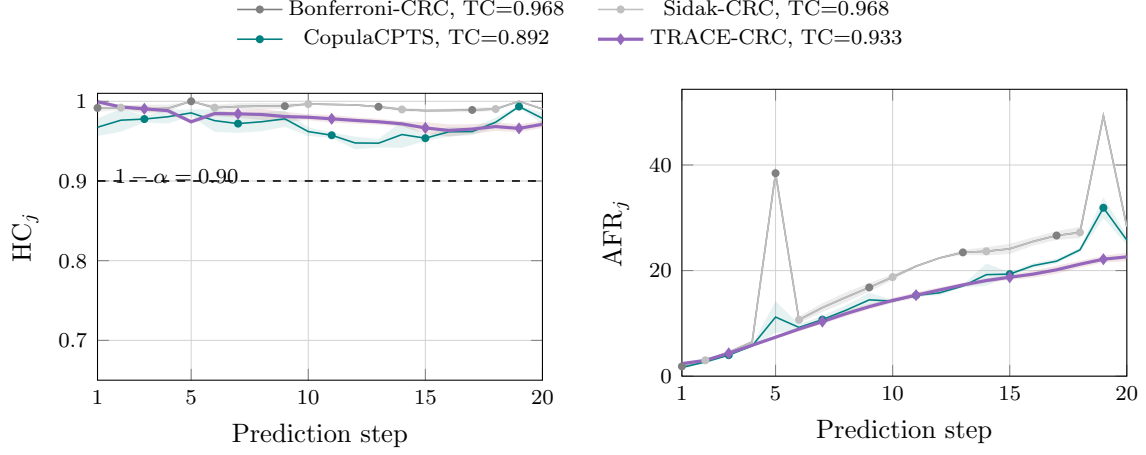


Figure 4: Horizon-wise coverage and average Frobenius radius for structured multi-step baselines and TRACE-CRC.

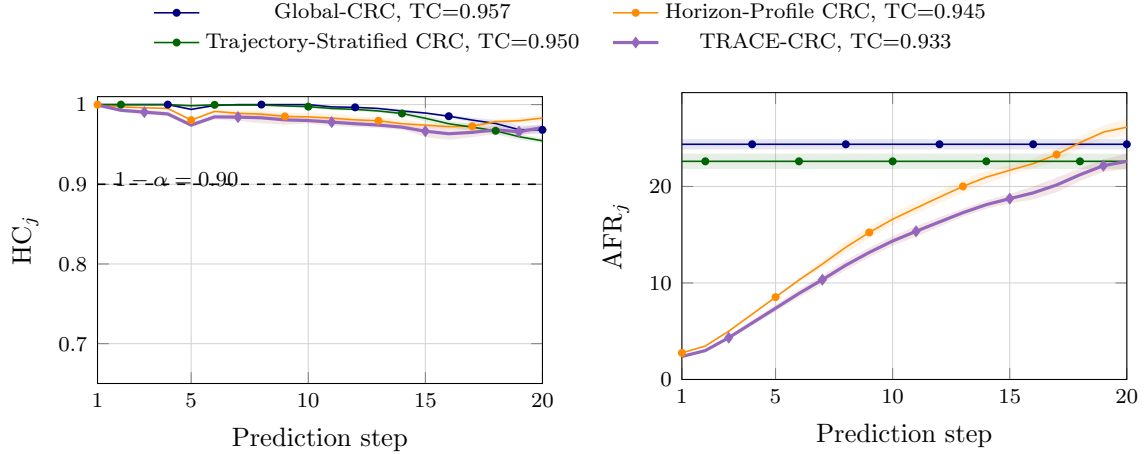


Figure 5: Horizon-wise coverage and average Frobenius radius for TRACE-CRC and its risk-control ablations.

CRC is the most conservative variant, attaining $TC = 0.957$ with a constant $AFR = 24.37$ across horizons. Trajectory-Stratified CRC achieves similar trajectory coverage, $TC = 0.950$, while reducing the constant radius to 22.60. Horizon-Profile CRC yields horizon-dependent radii and lowers the average AFR to 15.78, with $TC = 0.945$. TRACE-CRC combines both adaptations and achieves $TC = 0.933$ with the smallest average radius, $AFR = 13.64$. These results show that horizon profiling contributes most of the efficiency gain, while trajectory stratification provides an additional reduction in uncertainty radii when combined with horizon-aware calibration in TRACE-CRC.

Robustness to calibration partition. We evaluated TRACE-CRC over 50 runs using a fixed outer 300/700 calibration/test split while varying the internal allocation of the 300 calibration trajectories into $\mathcal{D}_{\text{prof}}$, $\mathcal{D}_{\text{cp}}$, and $\mathcal{D}_{\text{val}}$, together with five diffusion inference seeds. TRACE-CRC selected a formally certified multiplier in every run and achieved an average trajectory coverage of 0.940 ± 0.015 , exceeding the target level in every run. Variability was small and was driven primarily by the internal calibration partition rather than diffusion inference (Appendix C), indicating robustness to the particular calibration allocation.

Overall reliability–efficiency trade-off. Overall, MHC alone is not sufficient for evaluating multi-step CSI uncertainty: several methods achieve high MHC while their TC remains below the nominal target. Methods that explicitly target trajectory-level or simultaneous coverage improve TC but often increase AFR. TRACE-CRC provides a favorable trade-off by maintaining TC above the nominal target while reducing AFR relative to conservative simultaneous-correction baselines.

7. Conclusion

This work introduced TRACE-CRC, a trajectory-adaptive conformal risk control (CRC) framework for uncertainty quantification in multi-step CSI prediction. The central motivation is that standard marginal or horizon-wise coverage objectives are not sufficient when downstream decisions depend on the full predicted channel trajectory. TRACE-CRC constructs per-horizon Frobenius-norm uncertainty balls and calibrates them under a trajectory-level failure criterion, while incorporating horizon profiling and trajectory stratification to account for structured variation in CSI prediction errors.

Empirically, TRACE-CRC achieved a favorable reliability–efficiency trade-off. Horizon-wise and adaptive conformal baselines often produced compact uncertainty balls but failed to maintain reliable trajectory coverage, whereas conservative multi-step methods such as Bonferroni and Global-CRC improved trajectory coverage at the cost of substantially larger radii. The ablation study showed that combining horizon profiling with trajectory stratification reduces conservatism while maintaining trajectory-level reliability. More broadly, TRACE-CRC could serve as a foundation for trajectory-level uncertainty quantification in other complex time-series use cases in telecommunications. It may also have applications in areas that require reliability over an entire predicted sequence, including financial transaction monitoring, industrial sensing, and healthcare forecasting.

A limitation of the current study is that the risk-control certificate and the tightness of the resulting Frobenius-norm uncertainty balls depend on the available calibration and validation data, the forecast horizon length, and the assumption of trajectory-level exchangeability between validation and deployment. This assumption may fail when the distribution of complete trajectories changes, for example because SNR, mobility, or blockage regimes shift after calibration, in which case validation losses may no longer represent deployment losses. Future work should examine sensitivity to calibration-set size, validation-set size, prediction horizon length, and other distribution shifts inherent in wireless communications, in addition to shift-aware recalibration strategies.

Acknowledgment

This work was supported in part by the EUREKA CELTIC-NEXT SUSTAINET-Advance project, funded by Vinnova (Sweden’s Innovation Agency) under Grant 2025-02987, in part by the Swedish Research Council (VR) through the 6G-NTN-E Research Environment under Grant 2024-06645.

References

- Study on channel model for frequencies from 0.5 to 100 GHz. Technical Report TR 38.901, 3GPP, Mar. 2022. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/.
- Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification, 2022. URL <https://arxiv.org/abs/2107.07511>.
- Anastasios N. Angelopoulos, Stephen Bates, Emmanuel J. Candès, Michael I. Jordan, and Lihua Lei. Learn then test: Calibrating predictive algorithms to achieve risk control, 2022a. URL <https://arxiv.org/abs/2110.01052>.
- Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control, 2022b. URL <https://arxiv.org/abs/2208.02814>.
- K. E. Baddour and N. C. Beaulieu. Autoregressive modeling for fading channel simulation. *IEEE Trans. Wireless Commun.*, 4(4):1650–1662, 2005. doi: 10.1109/TWC.2005.850327.
- Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023. doi: 10.1214/23-AOS2276. URL <https://projecteuclid.org/journals/annals-of-statistics/volume-51/issue-2/Conformal-prediction-beyond-exchangeability/10.1214/23-AOS2276.full>.
- Stephen Bates, Anastasios Angelopoulos, Lihua Lei, Jitendra Malik, and Michael I. Jordan. Distribution-free, risk-controlling prediction sets. *Journal of the ACM*, 68(6):43:1–43:34, 2021. doi: 10.1145/3478535. URL <https://doi.org/10.1145/3478535>.
- Vidmantas Bentkus. On hoeffding’s inequalities. *The Annals of Probability*, 32(2), April 2004. ISSN 0091-1798. doi: 10.1214/009117904000000360. URL <http://dx.doi.org/10.1214/009117904000000360>.
- Emil Björnson, Jakob Hoydis, and Luca Sanguinetti. Massive MIMO networks: Spectral, energy, and hardware efficiency. *Foundations and Trends® in Signal Processing*, 11(3-4): 154–655, 2017. ISSN 1932-8346. doi: 10.1561/20000000093. URL <http://dx.doi.org/10.1561/20000000093>.
- Carlo Emilio Bonferroni. *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936.

- Kfir M. Cohen, Sangwoo Park, Osvaldo Simeone, and Shlomo Shamai Shitz. Calibrating ai models for few-shot demodulation via conformal prediction. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023a. doi: 10.1109/ICASSP49357.2023.10096780.
- Kfir M. Cohen, Sangwoo Park, Osvaldo Simeone, and Shlomo Shamai Shitz. Calibrating ai models for wireless communications via conformal prediction. *IEEE Transactions on Machine Learning in Communications and Networking*, 1:296–312, 2023b. doi: 10.1109/TMLCN.2023.3319282.
- António F. Farinhas, Dennis Ulmer, Chrysoula Zerva, and André F. T. Martins. Non-exchangeable conformal risk control. In *International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=j511LaqEeP>.
- Aloysio Galvão Lopes, Eric Goubault, Sylvie Putot, and Laurent Pautet. ConForME: Multi-horizon conditional conformal time series forecasting. In *Proceedings of the Thirteenth Symposium on Conformal and Probabilistic Prediction with Applications*, volume 230 of *Proceedings of Machine Learning Research*, pages 345–365. PMLR, 2024. URL <https://proceedings.mlr.press/v230/galvao-lopes24a.html>.
- Isaac Gibbs and Emmanuel Candès. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems*, volume 34, pages 1660–1672, 2021. URL <https://proceedings.neurips.cc/paper/2021/hash/0d441de75945e5acbc865406fc9a2559-Abstract.html>.
- Metod Jazbec, Alexander Timans, Tin Hadži Veljković, Kaspar Sakmann, Dan Zhang, Christian A. Naesseth, and Eric Nalisnick. Fast yet safe: Early-exiting with risk control. In *Advances in Neural Information Processing Systems*, volume 37, pages 129825–129854, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/ea5a63f7ddb82e58623693fd1f4933f7-Abstract-Conference.html.
- W. Jiang and H. D. Schotten. Recurrent neural network-based frequency-domain channel prediction for wideband communications. In *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, pages 1–6, 2019. doi: 10.1109/VTCSpring.2019.8746352.
- Y. Jin, Y. Wu, Y. Gao, S. Zhang, S. Xu, and C.-X. Wang. Linformer: A linear-based lightweight transformer architecture for time-aware MIMO channel prediction. *IEEE Trans. Wireless Commun.*, pages 1–1, 2025. doi: 10.1109/TWC.2025.3558950.
- Dongwon Kim, Jinu Gong, and Joonhyuk Kang. Mimo channel prediction via deep learning-based conformal bayes filter, 2026. URL <https://arxiv.org/abs/2603.04764>.
- H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi. Massive MIMO channel prediction: Kalman filtering vs. machine learning. *IEEE Trans. Commun.*, 69(1):518–528, 2021. doi: 10.1109/TCOMM.2020.3027882.
- Jonghyeok Lee, Chen Xu, and Yao Xie. Kernel-based optimally weighted conformal time-series prediction. In *International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=oP7arLOWix>.

- 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. doi: 10.1080/01621459.2017.1307116. URL <https://www.tandfonline.com/doi/abs/10.1080/01621459.2017.1307116>.
- William Overman, Jacqueline Jil Vallon, and Mohsen Bayati. Aligning model properties via conformal risk control. In *Advances in Neural Information Processing Systems*, volume 37, pages 110702–110722, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/c79625091a4f8b5d3abe29f3b14fa43a-Abstract-Conference.html.
- Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *Machine Learning: ECML 2002*, pages 345–356. Springer, 2002. doi: 10.1007/3-540-36755-1_29.
- Nalini Ravishanker, Dipak K. Dey, and George C. Tiao. Multiple prediction intervals for time series: Comparison of simultaneous and marginal intervals. *International Journal of Forecasting*, 7(4):445–463, 1991.
- Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://papers.neurips.cc/paper_files/paper/2019/hash/5103c3584b063c431bd1268e9b5e76fb-Abstract.html.
- Mehdi Sattari, Javad Aliakbari, Alexandre Graell i Amat, and Tommy Svensson. Csi prediction using diffusion models, 2025. URL <https://arxiv.org/abs/2510.11214>.
- Zbyněk Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.
- Osvaldo Simeone, Sangwoo Park, and Matteo Zecchin. Conformal calibration: Ensuring the reliability of black-box ai in wireless systems, 2025. URL <https://arxiv.org/abs/2504.09310>.
- O. Stenhammar, G. Fodor, and C. Fischione. A comparison of neural networks for wireless channel prediction. *IEEE Wireless Commun.*, 31(3):235–241, 2024. doi: 10.1109/MWC.006.2300140.
- Xin Su, Qiushuo Hou, Ruisi He, and Osvaldo Simeone. Conformal robust beamforming via generative channel models. In *2025 IEEE 26th International Workshop on Signal Processing and Artificial Intelligence for Wireless Communications (SPAWC)*, pages 1–5, 2025. doi: 10.1109/SPAWC66079.2025.11143438.
- Sophia Huiwen Sun and Rose Yu. Copula conformal prediction for multi-step time series forecasting. In *International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=ojIJZDNIBj>.
- Ryan J. Tibshirani, Rina Foygel Barber, Emmanuel J. Candès, and Aaditya Ramdas. Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL <https://arxiv.org/abs/1904.06019>.

- Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer-Verlag, Berlin, Heidelberg, 2005. ISBN 0387001522.
- Xiaoqian Wang and Rob J. Hyndman. Online conformal inference for multi-step time series forecasting, 2024. URL <https://arxiv.org/abs/2410.13115>.
- Chen Xu and Yao Xie. Conformal prediction interval for dynamic time-series. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11559–11569. PMLR, 2021. URL <https://proceedings.mlr.press/v139/xu21h.html>.
- Chen Xu and Yao Xie. Conformal prediction for time series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):11575–11587, 2023a. doi: 10.1109/TPAMI.2023.3272339. URL <https://ieeexplore.ieee.org/document/10121511>.
- Chen Xu and Yao Xie. Sequential predictive conformal inference for time series. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 38707–38727. PMLR, 2023b. URL <https://proceedings.mlr.press/v202/xu23r.html>.
- H. Yin, H. Wang, Y. Liu, and D. Gesbert. Addressing the curse of mobility in massive MIMO with prony-based angular-delay domain channel predictions. *IEEE J. Sel. Areas Commun.*, 38(12):2903–2917, 2020. doi: 10.1109/JSAC.2020.3005473.
- Seonghoon Yoo, Sangwoo Park, Petar Popovski, Joonhyuk Kang, and Osvaldo Simeone. Calibrating wireless ai via meta-learned context-dependent conformal prediction. *IEEE Transactions on Signal Processing*, 74:423–438, 2026. doi: 10.1109/TSP.2026.3650912.
- Margaux Zaffran, Aymeric Dieuleveut, Olivier Féron, Yannig Goude, and Julie Josse. Adaptive conformal predictions for time series. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 25834–25866. PMLR, 2022. URL <https://proceedings.mlr.press/v162/zaffran22a.html>.
- Matteo Zecchin and Osvaldo Simeone. Localized adaptive risk control. In *Advances in Neural Information Processing Systems*, volume 37, pages 8165–8192, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/0f93c3e9b557980d93016671acd94bd2-Abstract-Conference.html.
- Matteo Zecchin, Sangwoo Park, and Osvaldo Simeone. Forking uncertainties: Reliable prediction and model predictive control with sequence models via conformal risk control, 2024. URL <https://arxiv.org/abs/2310.10299>.
- Lujing Zhang, Aaron Roth, and Linjun Zhang. Fair risk control: A generalized framework for calibrating multi-group fairness risks. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. URL <https://arxiv.org/abs/2405.02225>.

Appendix A. Complete Results

Table 3 reports the full set of evaluated conformal methods. This table includes all standard split-conformal baselines, adaptive time-series baselines, structured multi-step baselines, and risk control ablations.

Table 3: Validity and efficiency of all conformal methods under the fixed 300/700 calibration–test split. Results are reported as mean $\pm$ standard deviation over five random diffusion inference seeds; the target trajectory coverage is 0.90.

Category	Method	MHC	WHC	TC	AFR
Standard split-conformal	Global Residual	0.880 ± 0.003	0.658 ± 0.006	0.642 ± 0.006	12.98 ± 0.13
	Horizon-wise	0.872 ± 0.002	0.855 ± 0.006	0.710 ± 0.009	10.73 ± 0.06
	Max-score Split	0.973 ± 0.001	0.901 ± 0.005	0.882 ± 0.005	19.52 ± 0.26
	Residual Quantile	0.965 ± 0.002	0.939 ± 0.003	0.903 ± 0.007	14.30 ± 0.21
Time-series adaptive	EW-Horizon	0.893 ± 0.001	0.885 ± 0.001	0.735 ± 0.005	11.40 ± 0.03
	EW-Trajectory	0.976 ± 0.001	0.908 ± 0.001	0.891 ± 0.002	20.25 ± 0.11
	EnbPI-Horizon	0.918 ± 0.003	0.884 ± 0.003	0.790 ± 0.010	12.13 ± 0.18
	EnbPI-Trajectory	0.984 ± 0.001	0.937 ± 0.007	0.919 ± 0.008	21.80 ± 0.24
	ACI-Horizon	0.898 ± 0.000	0.896 ± 0.001	0.733 ± 0.005	11.74 ± 0.03
	ACI-Trajectory	0.977 ± 0.001	0.912 ± 0.001	0.898 ± 0.001	20.46 ± 0.17
	AgACI-Horizon	0.896 ± 0.001	0.891 ± 0.001	0.744 ± 0.006	11.49 ± 0.06
	AgACI-Trajectory	0.977 ± 0.001	0.911 ± 0.003	0.895 ± 0.003	20.34 ± 0.15
Structured multi-step	Bonferroni-CRC	0.993 ± 0.001	0.985 ± 0.003	0.968 ± 0.004	20.00 ± 0.29
	Sidak-CRC	0.993 ± 0.001	0.985 ± 0.003	0.968 ± 0.004	20.00 ± 0.29
	CopulaCPTS	0.969 ± 0.004	0.945 ± 0.004	0.892 ± 0.012	14.87 ± 0.29
Risk control / ablation	Global-CRC	0.992 ± 0.001	0.967 ± 0.005	0.957 ± 0.007	24.37 ± 0.54
	Horizon-Profile CRC	0.984 ± 0.002	0.970 ± 0.004	0.945 ± 0.004	15.78 ± 0.38
	Trajectory-Stratified CRC	0.989 ± 0.002	0.954 ± 0.003	0.950 ± 0.005	22.60 ± 0.79
	TRACE-CRC	0.978 ± 0.004	0.962 ± 0.006	0.933 ± 0.011	13.64 ± 0.41

Appendix B. Method Parameters

Tables 4 and 5 report the selected baseline configurations and their validation performance, respectively. The search grids were fixed before evaluation. For each of the five random diffusion inference seeds used in the main experiment, the fixed 300-trajectory calibration set was divided into 210 tuning-training and 90 tuning-validation trajectories; the 700 test trajectories were not used for selection.

Horizon-wise and trajectory-level candidates were first required to satisfy $\text{MHC} \geq 0.90$ and $\text{TC} \geq 0.90$, respectively, for every inference seed. When no candidate met this all-seed criterion, selection minimized the worst-seed coverage shortfall, followed by the mean shortfall and then the mean AFR. Such cases are marked by $\dagger$. The exponentially weighted conformal baselines searched over β , EnbPI over ensemble size, window size, and centering rule, ACI and AgACI over their update-rate grids, and CopulaCPTS over calibration split and training epochs.

For the CRC variants, all 27 prespecified multipliers were included in a single Holm correction using one-sided Hoeffding–Bentkus p -values, with no empirical fallback when

Table 4: Parameters used for the standard, adaptive, and structured conformal baselines.

Method	Parameters used
Global Residual	Pooled split-conformal residual quantile, $\alpha = 0.10$.
Horizon-wise	Per-horizon split-conformal residual quantiles, $\alpha = 0.10$.
Max-score Split	Split conformal on $S_i = \max_{1 \leq j \leq T} E_{i,j}$, $\alpha = 0.10$.
Residual Quantile	50/50 training/calibration split; QuantileRegressor with quantile = 0.90, regularization = 10^{-4} , and solver = highs.
EW-Horizon [†]	$\beta = 0.02$; horizon-wise exponentially weighted quantiles with online residual-history updates.
EW-Trajectory [†]	$\beta = 0.02$; exponentially weighted trajectory-max quantiles with online residual-history updates.
EnbPI-Horizon	$B = 50$; full residual history; mean center; Ridge base model with penalty = 1; online update.
EnbPI-Trajectory	$B = 50$; window = 100; mean center; Ridge base model with penalty = 1; online update.
ACI-Horizon [†]	$\gamma = 0.05$; horizon-wise adaptive levels with online updates.
ACI-Trajectory [†]	$\gamma = 0.02$; trajectory-max adaptive level with online updates.
AgACI-Horizon [†]	$\Gamma = \{0, 0.005, 0.01, 0.02, 0.03, 0.05\}$; $\eta = 0.05$; horizon-wise online aggregation.
AgACI-Trajectory [†]	$\Gamma = \{0, 0.005, 0.01, 0.02, 0.03, 0.05\}$; $\eta = 0.20$; trajectory-level online aggregation.
Bonferroni-CRC	Per-horizon correction $\alpha_j = \alpha/T = 0.005$, with $T = 20$.
Sidak-CRC	Per-horizon correction $\alpha_j = 1 - (1 - \alpha)^{1/T} = 0.005254$, with $T = 20$.
CopulaCPTS	Quantile/copula calibration split = 0.50/0.50; epochs = 1200; algorithmic random seed = 123.

formal certification was unavailable. Table 6 reports the calibrated quantities and selected LTT multipliers used in the final evaluation.

Table 5: Validation performance of the selected baseline configurations, reported as mean $\pm$ standard deviation across five diffusion inference seeds.

Method	Val. MHC	Val. TC	Val. fail	Val. AFR
EW-Horizon [†]	0.8977 ± 0.0044	0.7511 ± 0.0151	0.2489 ± 0.0151	10.9397 ± 0.0901
EW-Trajectory [†]	0.9760 ± 0.0010	0.8956 ± 0.0054	0.1044 ± 0.0054	19.9327 ± 0.1515
EnbPI-Horizon	0.9099 ± 0.0071	0.8200 ± 0.0227	0.1800 ± 0.0227	11.0966 ± 0.0835
EnbPI-Trajectory	0.9778 ± 0.0009	0.9044 ± 0.0054	0.0956 ± 0.0054	20.2111 ± 0.1577
ACI-Horizon [†]	0.8996 ± 0.0014	0.7600 ± 0.0249	0.2400 ± 0.0249	11.0983 ± 0.2292
ACI-Trajectory [†]	0.9771 ± 0.0013	0.8956 ± 0.0054	0.1044 ± 0.0054	20.1203 ± 0.2402
AgACI-Horizon [†]	0.8981 ± 0.0042	0.7756 ± 0.0163	0.2244 ± 0.0163	10.8788 ± 0.1077
AgACI-Trajectory [†]	0.9766 ± 0.0015	0.8956 ± 0.0054	0.1044 ± 0.0054	19.9937 ± 0.2320
CopulaCPTS	0.9780 ± 0.0015	0.9222 ± 0.0141	0.0778 ± 0.0141	14.4977 ± 0.2848

Table 6: Method-specific parameters, calibrated quantities, and LTT outcomes for the conformal risk-control variants. Reported values are mean $\pm$ standard deviation across five diffusion inference seeds; brackets show the observed range of λ^* .

Method	Fixed parameters	Calibrated quantities	LTT outcome
Global-CRC	No method-specific hyperparameters	$q_{\text{global}} = 11.607 \pm 0.170$	$\lambda^* = 2.100 \pm 0.071$ [2.000, 2.200]
Horizon-Profile CRC	Smoothing window $K = 3$; stabilization ratio $\rho = 0.15$	$q_{\text{global}} = 11.607 \pm 0.170$	$\lambda^* = 1.360 \pm 0.055$ [1.300, 1.400]
Trajectory-Stratified CRC	Number of groups = 2; Ridge penalty $\eta = 1.0$	$q_0^{\text{raw}} = 13.877 \pm 2.521$; $q_1^{\text{raw}} = 23.850 \pm 1.907$	$\lambda^* = 1.253 \pm 0.099$ [1.167, 1.400]
TRACE-CRC	Number of groups = 2; smoothing window $K = 3$; stabilization ratio $\rho = 0.15$; Ridge penalty $\eta = 1.0$	$q_0 = 10.857 \pm 0.386$; $q_1 = 15.052 \pm 0.326$	$\lambda^* = 1.073 \pm 0.037$ [1.033, 1.100]

Appendix C. Robustness to the Internal Calibration Partition

To assess the sensitivity of TRACE-CRC to the internal calibration partition, we fixed the outer 300/700 trajectory-level calibration/test split and repeated the evaluation over 10 random reallocations of the 300 calibration trajectories into $\mathcal{D}_{\text{prof}}$, $\mathcal{D}_{\text{cp}}$, and $\mathcal{D}_{\text{val}}$. Each partition was evaluated using five diffusion inference seeds, yielding 50 runs in total. Table 7 summarizes the resulting variability. Coverage remained stable across runs, and the internal calibration partition had a larger effect than diffusion inference, although both sources of variability were modest.

Table 7: Variability of TRACE-CRC over 50 runs using 10 internal calibration partitions and five diffusion inference seeds. Internal-split and inference-seed SDs are marginal standard deviations across the corresponding averaged levels.

Metric	Overall SD	Internal split SD	Inference-seed SD
MHC	0.0062	0.0051	0.0021
WHC	0.0081	0.0063	0.0025
TC	0.0155	0.0109	0.0058
AFR	0.9753	0.6691	0.3114