Prospects for Applied Mathematics and Data Analysis

ISSN: 2836-4449 Vol: 05, No. 02, 2025 PP. 15–21

Article DOI: https://doi.org/10.54216/PAMDA.050203

FULL LENGTH ARTICLE

A Mathematical Framework for Adaptive Rolling Conformal Quantile Boosting under Temporal Distribution Shift: Application to Hour-Ahead $PM_{2.5}$ Forecast Intervals

Aiyared Iampan[1,*] **Said Broumi**[2]

[1] School of Science, University of Phayao, 19, Moo 2, Tambon Mae Ka, Amphur Mueang, Phayao 56000, Thailand

[2] Laboratory of Information Processing, Faculty of Science Ben MSik, University of Hassan II, Casablanca, Morocco

Emails: aiyared.ia@up.ac.th; broumisaid78@gmail.com

Received: May 10, 2025 Revised: August 11, 2025 Accepted: October 16, 2025 ⋆ Corresponding author

ABSTRACT

Prediction intervals for temporally dependent data require both conditional quantile estimation and a calibration mechanism capable of responding to distribution shift. An adaptive rolling conformal quantile boosting (ARCQB) formulation is developed in which boosted quantile functions provide a nonlinear base interval and a sequential state variable controls the empirical conformal quantile. For target miscoverage α, the calibration state follows a projected stochastic recurrence, $\alpha_{t+1} = \Pi_{\mathcal{A}}\{\alpha_t + \gamma(\alpha - e_t)\}$, where e_t is the realized miss indicator. A telescoping identity links the time-averaged miss frequency to the state displacement and projection residuals; in the unprojected bounded case, the calibration error is $O(T^{-1})$. The interval width admits the exact decomposition $w_t = w_t^{(0)} + 2q_t$, separating predictive sharpness from conformal inflation. Numerical evaluation uses a strictly chronological one-hour-ahead design on hourly Beijing air-quality measurements. For nominal 90% coverage, raw boosted quantiles attain 83.18%, static conformal calibration 87.40%, and rolling conformal calibration 89.87%. ARCQB attains 90.05% with mean width 47.14 μg m^{-3} and the lowest interval score, 72.86. Its maximum seasonal coverage deviation is 0.38 percentage points, compared with 7.99 points for the uncalibrated interval. The numerical behavior is therefore consistent with the feedback relation predicted by the calibration dynamics, while high-pollution regimes remain the principal source of conditional under-coverage.

Keywords: Conformal prediction ■ Adaptive calibration ■ Quantile regression ■ Stochastic recurrence ■ Distribution shift ■ Time-series uncertainty

1. INTRODUCTION

For a temporally ordered process, a point predictor and an uncertainty set solve different mathematical problems. The first estimates a conditional location functional; the second must control the frequency with which future observations fall outside a data-dependent set. In air-quality forecasting these objectives diverge sharply because pollutant concentrations are persistent, heteroscedastic, seasonal, and subject to abrupt regime changes. Recent forecasting models have improved deterministic accuracy using recurrent, decomposition-based, spatial–temporal, and transfer-learning architectures [5, 7, 9–11, 20, 22]. Their predictive intervals, when available, do not automatically retain nominal coverage after the error distribution changes.

Conformal prediction supplies a model-agnostic calibration layer based on empirical score ranks. Exact finite-sample arguments are strongest under exchangeability [2], while recent results extend conformal reasoning to non-exchangeable and dependent observations [3, 17]. Online formulations replace a fixed calibration level by a state that reacts to realized coverage errors [6]. This feedback interpretation is especially natural for environmental series: forecast errors arrive sequentially, and the uncertainty layer can adapt even if the underlying nonlinear predictor remains fixed.

Recent reviews describe the rapid expansion of machine-learning air-quality forecasting [8, 12], including spatially informed deep predictors [16], alternative input and horizon structures [18], and recurrent hybrid schemes [23]. Applied mathematical formulations based on data-driven differential equations provide a complementary perspective [15]. For predictive inference, training-conditional behavior and stability remain central limitations of distribution-free procedures [4, 13]; chronological validation is therefore preferred to random time-series partitioning [14].

The mathematical object considered here is a one-step interval sequence $\{I_t\}$ generated by two coupled components. The first is a family of boosted conditional quantile estimators $\widehat{q}_\tau(\mathbf{x}_t)$. The second is a rolling empirical score distribution with state-dependent tail probability. The resulting ARCQB recursion is analyzed through rank calibration, a projection-corrected telescoping identity, monotonicity of empirical quantiles, and exact width decomposition. These properties separate what is guaranteed by conformal rank arguments from what is induced by sequential feedback under temporal shift.

The numerical analysis is deliberately secondary to the formulation. Hourly observations are split chronologically into model-fitting, calibration, and future test periods; no future interpolation or random repartitioning is used. The comparison isolates calibration because the raw quantile, static conformal, rolling conformal, and adaptive intervals share the same fitted lower and upper quantile functions. The empirical question is therefore narrow: whether the mathematical calibration dynamics correct the under-coverage of a fixed nonlinear interval forecast.

2. MATHEMATICAL FRAMEWORK

2.1 Information set and one-step state map

Let $(\Omega, \mathcal{F}, \mathbb{P})$ support a stochastic process $\{(Y_t, \mathbf{Z}_t)\}_{t\ge 1}$, where $Y_t \in \mathbb{R}_+$ is the $PM_{2.5}$ concentration and $\mathbf{Z}_t \in \mathbb{R}^d$ contains contemporaneously observed pollutants and meteorological variables. The information available at forecast origin t is

$$\mathcal{F}_t = \sigma\{Y_s, \mathbf{Z}_s : s \le t\}. \tag{1}$$

A valid one-hour-ahead predictor must be $\mathcal{F}_t$-measurable.

Let $\mathcal{L} = \{1, 2, 3, 6, 12, 24\}$ denote lag indices and $\mathcal{R} = \{6, 12, 24\}$ rolling-window lengths. Define

$$\mu_t^{(r)} = \frac{1}{r}\sum_{j=0}^{r-1} Y_{t-j}, \tag{2}$$

$$(\sigma_t^{(r)})^2 = \frac{1}{r-1}\sum_{j=0}^{r-1}(Y_{t-j} - \mu_t^{(r)})^2. \tag{3}$$

The state vector is

$$\mathbf{X}_t = \Phi(\mathcal{F}_t) = \left[\mathbf{Z}_t^\top, \mathbf{c}_t^\top, \{Y_{t-\ell}\}_{\ell\in\mathcal{L}}, \{\mu_t^{(r)}, \sigma_t^{(r)}\}_{r\in\mathcal{R}}\right]^\top, \tag{4}$$

where $\mathbf{c}_t$ contains sine–cosine encodings of wind direction, hour, and day of year. Equation (4) makes the causality restriction explicit: no component depends on Y_{t+1}.

The target representation is

$$Y_{t+1} = m(\mathbf{X}_t) + \varepsilon_{t+1}, \qquad \mathbb{P}(\varepsilon_{t+1} \le u \mid \mathcal{F}_t) = F_t(u), \tag{5}$$

with no stationarity assumption on F_t.

2.2 Conditional quantiles and boosted risk minimization

For $\tau \in (0, 1)$, the conditional quantile is

$$q_\tau(\mathbf{x}) = \inf\{y : \mathbb{P}(Y_{t+1} \le y \mid \mathbf{X}_t = \mathbf{x}) \ge \tau\}. \tag{6}$$

It minimizes conditional pinball risk. With residual $u = y - f(\mathbf{x})$,

$$\rho_\tau(u) = u\,(\tau - \mathbb{I}\{u < 0\}). \tag{7}$$

For training index set $\mathcal{T}_{\mathrm{tr}}$, the empirical problem is

$$\widehat{f}_\tau = \arg\min_{f\in\mathcal{H}_M} \sum_{t\in\mathcal{T}_{\mathrm{tr}}} \rho_\tau(Y_{t+1} - f(\mathbf{X}_t)), \tag{8}$$

where $\mathcal{H}_M$ is an additive tree class. A functional boosting representation is

$$F_{m,\tau}(\mathbf{x}) = F_{m-1,\tau}(\mathbf{x}) + \nu h_{m,\tau}(\mathbf{x}), \quad m = 1, \ldots, M, \tag{9}$$

with learning rate ν and tree $h_{m,\tau}$ fitted to the pinball-loss pseudo-residual

$$r_{i,m}^{(\tau)} \in \tau - \mathbb{I}\{Y_{i+1} < F_{m-1,\tau}(\mathbf{X}_i)\}. \tag{10}$$

Three fitted functions are used: $\widehat{f}_{.05}$, $\widehat{f}_{.50}$, and $\widehat{f}_{.95}$. To prevent crossing at evaluation time, define

$$L_t^{(0)} = \min\{\widehat{f}_{.05}(\mathbf{X}_t), \widehat{f}_{.95}(\mathbf{X}_t)\}, \tag{11}$$

$$U_t^{(0)} = \max\{\widehat{f}_{.05}(\mathbf{X}_t), \widehat{f}_{.95}(\mathbf{X}_t)\}. \tag{12}$$

The raw interval is $I_t^{(0)} = [L_t^{(0)}, U_t^{(0)}]$ and its width is $w_t^{(0)} = U_t^{(0)} - L_t^{(0)}$.

2.3 Finite-sample conformal correction

For a calibration pair $(\mathbf{X}_i, Y_{i+1})$, use the nonnegative interval score

$$S_i = \max\{L_i^{(0)} - Y_{i+1},\ Y_{i+1} - U_i^{(0)},\ 0\}. \tag{13}$$

For m scores $S_1, \ldots, S_m$ and target miscoverage α, define

$$k_m(\alpha) = \min\{m, \lceil (m+1)(1-\alpha)\rceil\} \tag{14}$$

and let $S_{(k_m)}$ denote the corresponding order statistic. The static conformal interval is

$$I_t^{(S)} = [L_t^{(0)} - S_{(k_m)},\ U_t^{(0)} + S_{(k_m)}]. \tag{15}$$

Theorem 1 (Split-conformal rank coverage)**.** *If $(S_1, \ldots, S_m, S_{m+1})$ is exchangeable and independent of the model-fitting sample, then*

$$\mathbb{P}\{Y_{m+1} \in I_{m+1}^{(S)}\} \ge 1 - \alpha. \tag{16}$$

Proof. Under exchangeability, the rank of S_{m+1} among $m+1$ scores is uniform over admissible ranks, modulo ties. Failure requires $S_{m+1} > S_{(k_m)}$. By Eq. (14), the number of failing ranks is at most $\alpha(m+1)$; hence the failure probability is at most α. □

The theorem is a reference result, not a stationarity claim for atmospheric data. Temporal dependence motivates replacing the full score sample by a moving empirical distribution.

2.4 Rolling empirical quantile under temporal shift

Let

$$\mathcal{H}_t(W) = \{S_{t-j} : j = 1, \ldots, \min(W, t-1)\} \tag{17}$$

contain only scores revealed before the forecast at t. For an effective miscoverage state $a_t \in (0, 1)$, write

$$Q_t(a_t; W) = Q_{1-a_t}(\mathcal{H}_t(W)), \tag{18}$$

where Q uses the finite-sample order-statistic convention in Eq. (14). The rolling interval is

$$I_t(a_t, W) = [L_t^{(0)} - Q_t(a_t; W),\ U_t^{(0)} + Q_t(a_t; W)]. \quad (19)$$

The non-adaptive rolling conformal method fixes $a_t \equiv \alpha$.

Proposition 1 (Width decomposition)**.** *For every t for which Q_t is defined,*

$$w_t(a_t, W) = w_t^{(0)} + 2Q_t(a_t; W). \quad (20)$$

Proof. Subtract the lower endpoint in Eq. (19) from the upper endpoint. □

Equation (20) separates the width produced by the quantile regression from the additional width required by empirical calibration.

2.5 Adaptive calibration dynamics

Let

$$e_t = \mathbb{I}\{Y_{t+1} \notin I_t(a_t, W)\} \in \{0, 1\}. \quad (21)$$

ARCQB updates the effective miscoverage state by

$$a_{t+1} = \Pi_{[a_-, a_+]} \left[a_t + \gamma(\alpha - e_t)\right], \quad (22)$$

where $0 < a_- < \alpha < a_+ < 1$, $\gamma > 0$, and Π is Euclidean projection. In the numerical implementation $(a_-, a_+) = (0.005, 0.30)$.

To expose the effect of projection, define

$$r_t = a_{t+1} - a_t - \gamma(\alpha - e_t). \quad (23)$$

Thus $r_t = 0$ whenever the unconstrained update remains inside $[a_-, a_+]$.

Proposition 2 (Exact calibration identity)**.** *For any $T \geq 1$, the projected recurrence satisfies*

$$\frac{1}{T}\sum_{t=1}^{T} e_t = \alpha - \frac{a_{T+1} - a_1}{\gamma T} + \frac{1}{\gamma T}\sum_{t=1}^{T} r_t. \quad (24)$$

Proof. From Eq. (23), $\gamma e_t = \gamma\alpha - (a_{t+1} - a_t) + r_t$. Summation over $t = 1, \ldots, T$ telescopes the state increments and division by γT gives Eq. (24). □

Corollary 1 (Bounded unprojected calibration error)**.** *If projection is inactive, $r_t = 0$, and the state remains in an interval of diameter D, then*

$$\left|\frac{1}{T}\sum_{t=1}^{T} e_t - \alpha\right| \leq \frac{D}{\gamma T}. \quad (25)$$

Hence the time-averaged miss frequency converges to α at rate $O(T^{-1})$ whenever the state is bounded.

Proposition 3 (Negative feedback)**.** *For a fixed score multiset $\mathcal{H}_t(W)$, $Q_t(a; W)$ is nonincreasing in a. Consequently, if $e_t = 1$, then the unconstrained update gives $a_{t+1} < a_t$ and*

$$Q_t(a_{t+1}; W) \geq Q_t(a_t; W), \quad (26)$$

so the next conformal correction is weakly enlarged, subject to the change in score history.

Proof. The empirical quantile level is $1 - a$. Decreasing a increases this level, and empirical quantiles are monotone in their probability argument. For $e_t = 1$, Eq. (22) changes the unconstrained state by $-\gamma(1 - \alpha) < 0$. □

The recurrence therefore has a direct control interpretation: misses create interval expansion pressure, while hits create contraction pressure. It does not imply exact conditional coverage for every $\mathbf{X}_t$.

2.6 Calibration objective and diagnostics

For an interval $[L, U]$, define the interval score

$$\begin{aligned} \mathcal{S}_\alpha(L, U; y) =& (U - L) + \frac{2}{\alpha}(L - y)\mathbb{I}\{y < L\} \\ &+ \frac{2}{\alpha}(y - U)\mathbb{I}\{y > U\}. \end{aligned} \quad (27)$$

The rolling window and gain are selected on the calibration period only:

$$(W^*, \gamma^*) = \underset{W\in\mathcal{W}, \gamma\in\Gamma}{\arg\min} \frac{1}{M}\sum_{t=1}^{M} \mathcal{S}_{.10}(L_t, U_t; Y_{t+1}), \quad (28)$$

with

$$\mathcal{W} = \{336, 720, 1440, 2160\}, \quad (29)$$

$$\Gamma = \{0, .001, .003, .005, .01, .02\}. \quad (30)$$

The optimum is $(W^*, \gamma^*) = (2160, 0.02)$. The case $\gamma = 0$ is the rolling conformal baseline.

For a partition $\mathcal{G}$ of the test period into regimes g, define regime coverage

$$\widehat{C}_g = \frac{1}{n_g}\sum_{t\in g} \mathbb{I}\{Y_{t+1} \in I_t\} \quad (31)$$

and the maximum calibration discrepancy

$$D_\infty(\mathcal{G}) = \max_{g\in\mathcal{G}} |\widehat{C}_g - (1 - \alpha)|. \quad (32)$$

This diagnostic is used below for the four meteorological seasons.

3. NUMERICAL DESIGN

3.1 Data, causality, and temporal partition

The numerical series is the Aotizhongxin station from the Beijing Multi-Site Air Quality collection. The station file contains 35,064 hourly records from March 2013 through February 2017. The variables used at the forecast origin are $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, temperature, pressure, dew point, precipitation, wind speed, wind direction, and the derived temporal and lag features in Eq. (4). Missing predictor histories are removed; no future-value interpolation is used.

After construction of lagged and rolling states, 27,929 complete forecast origins remain. The chronological sets are

$$|\mathcal{T}_{tr}| = 13{,}989, \quad (33)$$

$$|\mathcal{T}_{cal}| = 6{,}902, \quad (34)$$

$$|\mathcal{T}_{te}| = 7{,}038. \quad (35)$$

Training runs from 2 March 2013 to 28 February 2015, calibration from 1 March 2015 to 29 February 2016, and testing from 1 March 2016 to 28 February 2017. The target means are 86.43, 76.49, and 81.64 $\mu g\,m^{-3}$, respectively; the corresponding 95th percentiles are 245, 252, and 259 $\mu g\,m^{-3}$. Figure 1 displays this strictly chronological design.

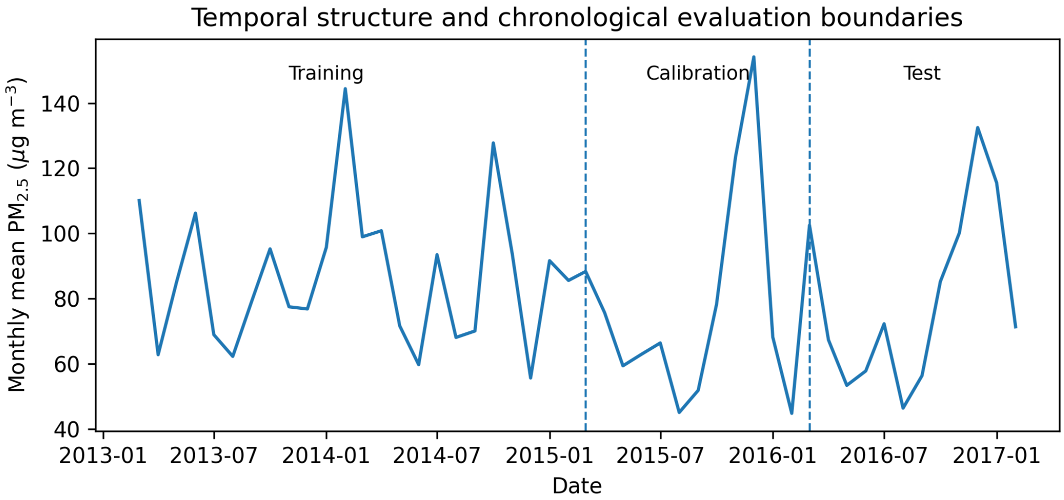

Figure 1: Chronological partition used in the numerical analysis. Model parameters are fitted before March 2015; calibration parameters are selected before March 2016; all final metrics are computed on the later test year.

3.2 Competing estimators

Persistence, $\widehat{Y}_{t+1} = Y_t$, is the short-horizon baseline. Ridge regression is fitted after selecting its penalty from $\{0.01, 0.1, 1, 10, 100\}$ by expanding-window validation inside $\mathcal{T}_{\mathrm{tr}}$; the selected penalty is 1.0. Histogram gradient boosting uses $M = 300$, $\nu = 0.05$, at most 31 terminal leaves, minimum leaf size 30, and ℓ_2 regularization 0.1. Identical tree-capacity parameters are used for the squared-error, median, 0.05-quantile, and 0.95-quantile fits.

The four interval estimators share the same pair $(L_t^{(0)}, U_t^{(0)})$: raw boosted quantiles; static conformal correction; rolling conformal correction with $W = 2160$; and ARCQB with $(W, \gamma) = (2160, 0.02)$. Calibration effects are therefore identified without refitting the underlying quantile functions.

3.3 Losses and dependence-aware uncertainty

Point performance is measured by

$$\mathrm{MAE} = n^{-1} \sum_{i=1}^{n} |y_i - \widehat{y}_i|, \tag{36}$$

$$\mathrm{RMSE} = \left[n^{-1} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \right]^{1/2}, \tag{37}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \widehat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}. \tag{38}$$

For intervals, empirical coverage and mean width are

$$\widehat{C} = n^{-1} \sum_i \mathbb{I}\{L_i \le y_i \le U_i\}, \qquad \overline{w} = n^{-1} \sum_i (U_i - L_i). \tag{39}$$

Serial dependence in pairwise loss differences is handled with a 24-hour moving-block bootstrap using 2,000 replicates.

4. RESULTS

4.1 Point-risk comparison

Table 1 reports the one-hour-ahead losses. Median boosting minimizes MAE at 10.126 μg m^{-3}; ridge regression minimizes RMSE at 18.514 μg m^{-3} and maximizes R^2 at 0.9547. The loss ordering is expected because Eq. (7) at $\tau = .5$ targets the conditional median, whereas squared loss emphasizes large deviations. The moving-block bootstrap for the absolute-loss contrast between median boosting and persistence gives

$$\Delta_{\mathrm{MAE}} = -0.380\ \mu\mathrm{g\,m}^{-3}, \tag{40}$$

$$95\%\ \mathrm{CI} = [-0.642, -0.031]\ \mu\mathrm{g\,m}^{-3}. \tag{41}$$

Table 1. Test-year point forecast performance.

Model	MAE	RMSE	R^2
Persistence	10.506	19.663	0.9489
Ridge	10.492	**18.514**	**0.9547**
HGB mean	10.482	19.748	0.9484
HGB median	**10.126**	19.735	0.9485

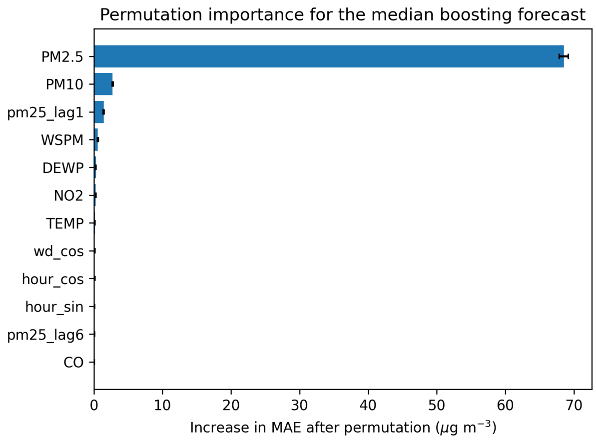

Figure 2: Permutation importance for the median boosted forecast. Bars report the increase in test MAE after feature permutation; error bars show the across-repeat standard deviation.

The permutation diagnostic in Figure 2 confirms that the current $PM_{2.5}$ measurement carries most of the one-hour-ahead signal. Permuting $PM_{2.5}$ increases MAE by 68.51 μg m^{-3}, compared with 2.74 μg m^{-3} for PM_{10} and 1.43 μg m^{-3} for the one-hour $PM_{2.5}$ lag. The remaining variables make substantially smaller marginal contributions once these persistent pollutant measurements are present.

Table 2. Test-year prediction interval performance; nominal coverage is 90%.

Method	Coverage (%)	Width	Interval score
Raw quantile boosting	83.18	**39.37**	74.74
Static conformal	87.40	41.98	73.54
Rolling conformal	89.87	44.67	73.28
ARCQB	**90.05**	47.14	**72.86**

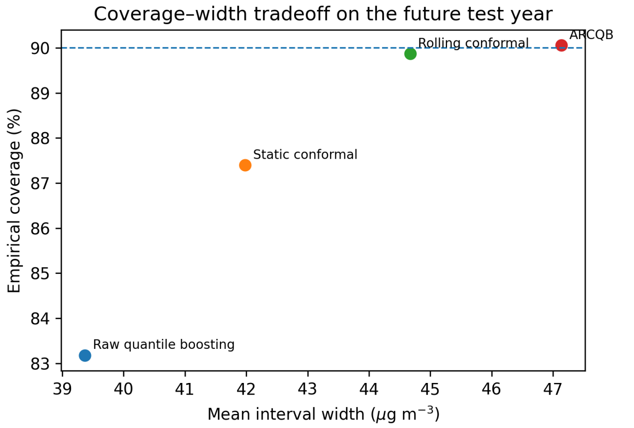

Figure 3: Coverage–width coordinates for the four interval constructions. The horizontal reference corresponds to nominal 90% coverage.

4.2 Coverage, width, and score

The nominal level is $1 - \alpha = 0.90$. Table 2 shows a clear calibration sequence. Raw quantile boosting under-covers by

$$0.9000 - 0.8318 = 0.0682, \tag{42}$$

or 6.82 percentage points. Static calibration reduces the gap to 2.60 points; rolling calibration reduces it to 0.13 points. ARCQB yields

$$\widehat{C}_A = 0.90054, \qquad |\widehat{C}_A - 0.90| = 5.40 \times 10^{-4}. \tag{43}$$

For ARCQB, the mean empirical correction is $\overline{Q} = 3.883\ \mu\text{g m}^{-3}$. The exact width relation in Eq. (20) gives

$$39.370 + 2(3.883) = 47.136\ \mu\text{g m}^{-3}, \tag{44}$$

which equals the observed mean width up to numerical precision. The increase in width is therefore attributable to the calibration term rather than a change in the base quantile predictor.

The interval-score contrast between ARCQB and static conformal calibration is -0.678, with a moving-block 95% interval $[-1.741, 0.239]$. The score improvement is therefore less decisive than the coverage correction; the principal gain is calibration accuracy. Figure 3 summarizes the associated coverage–width trade-off.

4.3 Observed adaptive-state dynamics

Across the test year, the adaptive state has mean 0.1126, median 0.1080, minimum 0.005, and maximum 0.276. The lower projection boundary is visited for 37 forecast states, so the practical recursion is not identical to the unprojected system in Corollary 1. Nevertheless, the exact projected identity in Eq. (24) remains valid.Using the final recursively implied state $a_{T+1} = 0.176$, $a_1 = 0.100$, $\gamma = 0.02$, $T = 7038$, and the realized projection

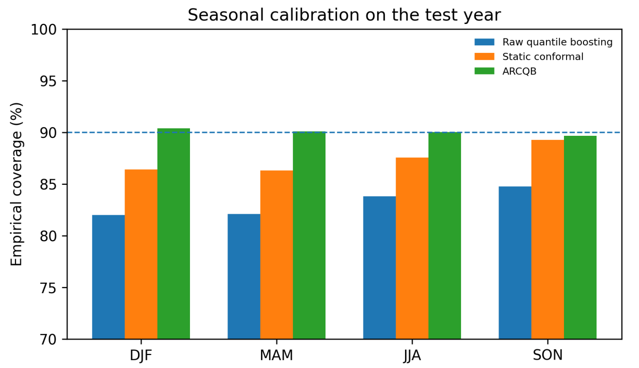

Figure 4: Seasonal empirical coverage for raw quantile boosting, static conformal calibration, and ARCQB. The dashed line marks nominal 90% coverage.

residuals, Eq. (24) evaluates to

$$\frac{1}{T}\sum_{t=1}^{T} e_t = 0.099460, \tag{45}$$

which is exactly the empirical miss fraction $1 - 0.900540$ up to rounding. This equality is algebraic rather than probabilistic; it verifies that the reported adaptive coverage is internally consistent with the implemented state recurrence.

The empirical conformal correction Q_t has mean 3.883 μg m $^{-3}$, median 2.137 μg m^{-3}, 95th percentile 12.722 μg m^{-3}, and maximum 87.072 μg m^{-3}. The large upper tail reflects episodes in which recent forecast errors require substantial interval inflation. Figure 5 shows how this correction behaves during a representative winter episode.

4.4 Seasonal calibration discrepancy

Table 3 evaluates Eq. (31) on the four seasons, while Figure 4 provides the corresponding visual comparison. ARCQB coverage ranges from 89.68% to 90.38%. Consequently,

$$D_\infty^{\text{ARCQB}} = 0.00383, \tag{46}$$

whereas

$$D_\infty^{\text{raw}} = 0.07994, \tag{47}$$

$$D_\infty^{\text{static}} = 0.03685, \tag{48}$$

$$D_\infty^{\text{rolling}} = 0.01237. \tag{49}$$

The adaptive recurrence reduces the worst seasonal calibration discrepancy by about 95% relative to the raw quantile interval.

4.5 Conditional stress at high pollution levels

The state available at the forecast origin also permits a pollution-regime diagnostic. Table 4 shows that for origin $PM_{2.5}$ below 35 $\mu\text{g m}^{-3}$, ARCQB coverage is 90.85%; for 35–75 $\mu\text{g m}^{-3}$ it is 91.80%; for 75–150 $\mu\text{g m}^{-3}$ it falls to 88.64%; and for at least 150 $\mu\text{g m}^{-3}$ it is 87.65%. Mean width rises monotonically from 21.17 to 108.92 $\mu\text{g m}^{-3}$ across these bands.

This failure mode is mathematically distinct from the seasonal result. The recurrence regulates the aggregate sequence of misses, but e_t in Eq. (21) is not conditioned on a pollution state. Therefore Eq. (24) controls a time average and does not imply $\mathbb{E}[e_t \mid \mathbf{X}_t \in A] = \alpha$ for every regime A. State-normalized conformity scores or stratified calibration are natural extensions, provided that the extreme-state sample is large enough for stable empirical quantiles.

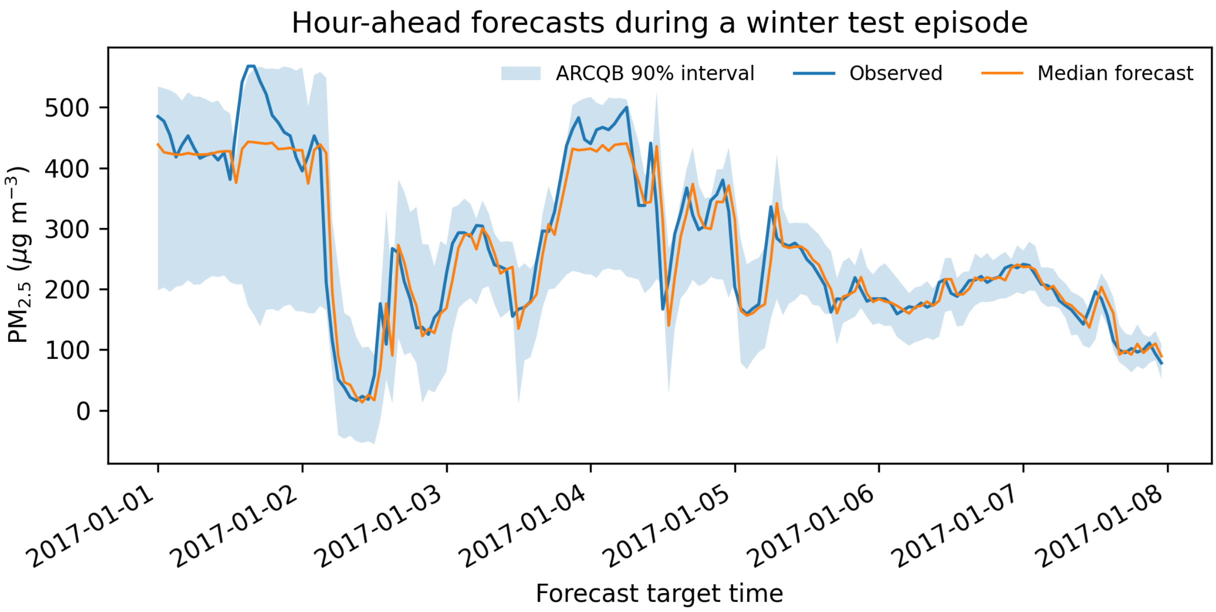

Figure 5: Observed concentration, median boosted forecast, and ARCQB 90% interval during 1–7 January 2017. Each interval is computed before its target hour is observed.

Table 3. Seasonal empirical coverage and mean width on the test year.

Season	Method	n	Coverage (%)	Mean width	Interval score
DJF	Raw quantile boosting	1934	82.01	52.47	109.91
DJF	Static conformal	1934	86.40	55.08	108.44
DJF	Rolling conformal	1934	89.14	58.36	108.09
DJF	ARCQB	1934	**90.38**	64.25	**106.41**
MAM	Raw quantile boosting	1593	82.11	36.78	71.37
MAM	Static conformal	1593	86.32	39.39	69.88
MAM	Rolling conformal	1593	88.76	42.36	**69.30**
MAM	ARCQB	1593	**90.08**	45.89	69.39
JJA	Raw quantile boosting	1737	83.82	30.20	53.91
JJA	Static conformal	1737	87.56	32.81	52.79
JJA	Rolling conformal	1737	91.13	36.07	52.61
JJA	ARCQB	1737	**90.04**	36.37	**52.60**
SON	Raw quantile boosting	1774	84.78	36.39	59.81
SON	Static conformal	1774	89.29	39.00	**59.10**
SON	Rolling conformal	1774	90.42	40.25	59.17
SON	ARCQB	1774	**89.68**	40.13	59.25

Table 4. ARCQB performance by origin PM$_{2.5}$ concentration band.

Band (μg m^{-3})	n	Coverage (%)	Width
< 35	2655	90.85	21.17
35–75	1647	91.80	39.56
75–150	1594	88.64	53.94
≥ 150	1142	87.65	108.92

5. DISCUSSION

The numerical results support three mathematical interpretations. First, conditional quantile estimation and coverage calibration are not equivalent. The same lower and upper boosted functions produce 83.18% raw coverage and 90.05% adaptive coverage; the difference arises entirely from the score process and state recurrence. Second, Eq. (20) makes the cost of calibration explicit. ARCQB adds an average 7.77 μg m^{-3} to the raw interval width, and that increase is accompanied by a 6.87-point improvement in coverage. Third, the seasonal discrepancy in Eq. (46) is much smaller than the pollution-stratum discrepancy, illustrating the distinction between long-run marginal control and statewise conditional calibration.

These findings complement recent air-quality work centered on increasingly complex deterministic architectures [1, 19, 21, 22]. A mathematically defined calibration layer can change uncertainty reliability without replacing the forecasting model. That separation is useful when prediction models are expensive to retrain or when the error law changes faster than the conditional location structure.

The exchangeable coverage theorem should not be read as a guarantee for the hourly atmospheric process. Dependence and drift violate the simplest rank assumptions, and current non-exchangeable conformal theory gives more nuanced conditions [3, 17]. The role of the adaptive recurrence is operational: it converts past coverage errors into a controlled change of the empirical quantile level. The exact identity in Eq. (24) describes that feedback algebraically; predictive validity under a particular stochastic dependence structure remains a separate theoretical question.

6. CONCLUSION

ARCQB expresses time-varying predictive uncertainty through a coupled quantile–calibration system. Boosted pinball-risk minimization defines the base interval; empirical score ranks define the conformal correction; and a projected feedback recurrence changes the effective tail probability after each realized forecast. The resulting formulation yields an exact width decomposition and an exact finite-horizon calibration identity, with an $O(T^{-1})$ miss-frequency bound in the bounded unprojected case.

On the chronological $PM_{2.5}$ experiment, the raw 90% quantile interval covers 83.18% of future observations, whereas ARCQB covers 90.05%. The maximum seasonal deviation from nominal coverage falls from 7.99 to 0.38 percentage points. The remaining under-coverage in high-pollution states identifies the main unresolved mathematical p roblem: converting sequential marginal calibration into reliable state-conditional calibration under dependence and sparse extremes.

REFERENCES

[1] Aggarwal, A., & Toshniwal, D. (2021). A hybrid deep learning framework for urban air quality forecasting. *Journal of Cleaner Production, 329*, 129660.

[2] Barber, R. F., Candès, E. J., Ramdas, A., & Tibshirani, R. J. (2021). Predictive inference with the jackknife+. *The Annals of Statistics, 49*(1), 486–507.

[3] Barber, R. F., Candès, E. J., Ramdas, A., & Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. *The Annals of Statistics, 51*(2), 816–845.

[4] Bian, M., & Barber, R. F. (2023). Training-conditional coverage for distribution-free predictive inference. *Electronic Journal of Statistics, 17*(2), 2044–2066.

[5] Chang, Y.-S., Chiao, H.-T., Abimannan, S., Huang, Y.-P., Tsai, Y.-T., & Lin, K.-M. (2020). An LSTM-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research, 11*(8), 1451–1463.

[6] Gibbs, I., & Candès, E. J. (2021). Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems, 34*, 1660–1672.

[7] Jin, N., Zeng, Y., Yan, K., & Ji, Z. (2021). Multivariate air quality forecasting with nested long short term memory neural network. *IEEE Transactions on Industrial Informatics, 17*(12), 8514–8522.

[8] Liao, Q., Zhu, M., Wu, L., Pan, X., Tang, X., & Wang, Z. (2020). Deep learning for air quality forecasts: A review. *Current Pollution Reports, 6*(4), 399–409.

[9] Lu, X., Ye, C., Shan, M., Qin, B., Wang, Y., Xing, H., Xie, X., & Liu, Z. (2023). The prediction of $PM_{2.5}$ concentration using transfer learning based on ADGRU. *Water, Air, & Soil Pollution, 234*, 258.

[10] Ma, J., Ding, Y., Cheng, J. C. P., Jiang, F., Gan, V. J. L., & Xu, Z. (2020). A Lag-FLSTM deep learning network based on Bayesian optimization for multi-sequential-variant $PM_{2.5}$ prediction. *Sustainable Cities and Society, 60*, 102237.

[11] Mao, W., Wang, W., Jiao, L., Zhao, S., & Liu, A. (2021). Modeling air quality prediction using a deep learning approach: Method optimization and evaluation. *Sustainable Cities and Society, 65*, 102567.

[12] Masood, A., & Ahmad, K. (2021). A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production, 322*, 129072.

[13] Steinberger, L., & Leeb, H. (2023). Conditional predictive inference for stable algorithms. *The Annals of Statistics, 51*(1), 290–311.

[14] Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: A survey. *Big Data, 9*(1), 3–21.

[15] Wang, Y., Wang, H., & Zhang, S. (2020). Prediction of daily $PM_{2.5}$ concentration in China using data-driven ordinary differential equations. *Applied Mathematics and Computation, 375*, 125088.

[16] Xiao, F., Yang, M., Fan, H., Fan, G., & Al-Qaness, M. A. A. (2020). An improved deep learning model for predicting daily $PM_{2.5}$ concentration. *Scientific Reports, 10*, 20988.

[17] Xu, C., & Xie, Y. (2023). Sequential predictive conformal inference for time series. *Proceedings of the 40th International Conference on Machine Learning, PMLR 202*, 38707–38727.

[18] Yang, J., Yan, R., Nong, M., Liao, J., Li, F., & Sun, W. (2021). $PM_{2.5}$ concentrations forecasting in Beijing through deep learning with different inputs, model structures and forecast time. *Atmospheric Pollution Research, 12*(9), 101168.

[19] Zaini, N., Ean, L. W., Ahmed, A. N., Abdul Malek, M., & Chow, M. F. (2022). $PM_{2.5}$ forecasting for an urban area based on deep learning and decomposition method. *Scientific Reports, 12*, 17565.

[20] Zeng, Y., Chen, J., Jin, N., Jin, X., & Du, Y. (2022). Air quality forecasting with hybrid LSTM and extended stationary wavelet transform. *Building and Environment, 213*, 108822.

[21] Zhang, Z., Zeng, Y., & Yan, K. (2021). A hybrid deep learning technology for $PM_{2.5}$ air quality forecasting. *Environmental Science and Pollution Research, 28*, 39409–39422.

[22] Zhang, K., Yang, X., Cao, H., Thé, J., Tan, Z., & Yu, H. (2023). Multi-step forecast of $PM_{2.5}$ and PM_{10} concentrations using convolutional neural network integrated with spatial–temporal attention and residual learning. *Environment International, 171*, 107691.

[23] Zheng, G., Liu, H., Yu, C., Li, Y., & Cao, Z. (2022). A new $PM_{2.5}$ forecasting model based on data preprocessing, reinforcement learning and gated recurrent unit network. *Atmospheric Pollution Research, 13*(7), 101475.