Sequential Predictive Conformal Inference for Time Series

Chen Xu [1] Yao Xie [1]

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology

Motivation and Problem Setup

Conformal prediction (CP) has become a popular distribution-free technique to perform uncertainty quantification for complex machine learning algorithms. However, conformal prediction for time series has been a challenging case because such data do not satisfy the exchangeability assumption in conformal inference, and thus we need to adjust existing or even develop new sequential CP algorithms with theoretical guarantees.

More precisely, both the traditional conformal inference and the sequential conformal inference we consider are general-purpose wrappers that can be used around any predictive model for any data and proceed by defining "non-conformity scores". However, there are also significant differences: Traditional conformal prediction assumes exchangeable training and test data to obtain performance guarantees, which leads to exchangeable non-conformity scores, and do not receive feedback during prediction. In contrast, sequential CP observes non-exchangeable data sequences and leverages feedback during prediction to yield more adaptive and accurate intervals.

In this work, we assume a sequence of observations (X_t, Y_t), $t = 1, 2, \ldots$, where Y_t are continuous scalar variables and $X_t \in \mathbb{R}^d$ denote features, which may either be the history of Y_t or contain exogenous variables helpful in predicting the value of Y_t. Given T training data and a user-specified significance level $\alpha \in [0, 1]$, we want to create prediction intervals $\widehat{C}_{t-1}(X_t)$ sequentially (at level α) such that

$$\mathbb{P}(Y_t \in \widehat{C}_{t-1}(X_t)|X_t) \to 1 - \alpha \text{ as } T \to \infty. \tag{1}$$

Our approach

The main novelty of `SPCI` is the time-adaptive re-estimation of residual quantiles over time, upon leveraging the temporal dependency among residuals.

More precisely, let $\hat{\epsilon}_t$ be the continuous *prediction* residual (or any user-designed non-conformity score). Let $F(z|\mathcal{H}_t)$ be the time-invariant conditional distribution of $\hat{\epsilon}_t$ given the history $\mathcal{H}_t$. `SPCI` thus estimates the conditional quantile $Q_t(p)$ for $p \in [0, 1]$, where we use quantile random forest (QRF) as the estimator [4] to establish theoretical guarantees. Together with a point predictor $\hat{f}(X_t)$, the interval $\widehat{C}_{t-1}(X_t)$ in this work has the form

Algorithm 1 Sequential Predictive Conformal Inference (`SPCI`)

Require: Training data $\{(X_t, Y_t)\}_{t=1}^T$, prediction algorithm $\mathcal{A}$, significance level α, quantile regression algorithm $\mathcal{Q}$.
Output: Prediction intervals $\widehat{C}_{t-1}(X_t), t > T$

```
Obtain f̂ and prediction residuals ε̂ with A and {(X_t, Y_t)}_{t=1}^T
for t > T do
    Use quantile regression to obtain Q̂_t ← Q(ε̂)
    Obtain prediction interval Ĉ_{t-1}(X_t) as in (2)
    Obtain new residual ε̂_t
    Update residuals ε̂ by sliding one index forward
    (i.e., add ε̂_t and remove the oldest one)
end for
```

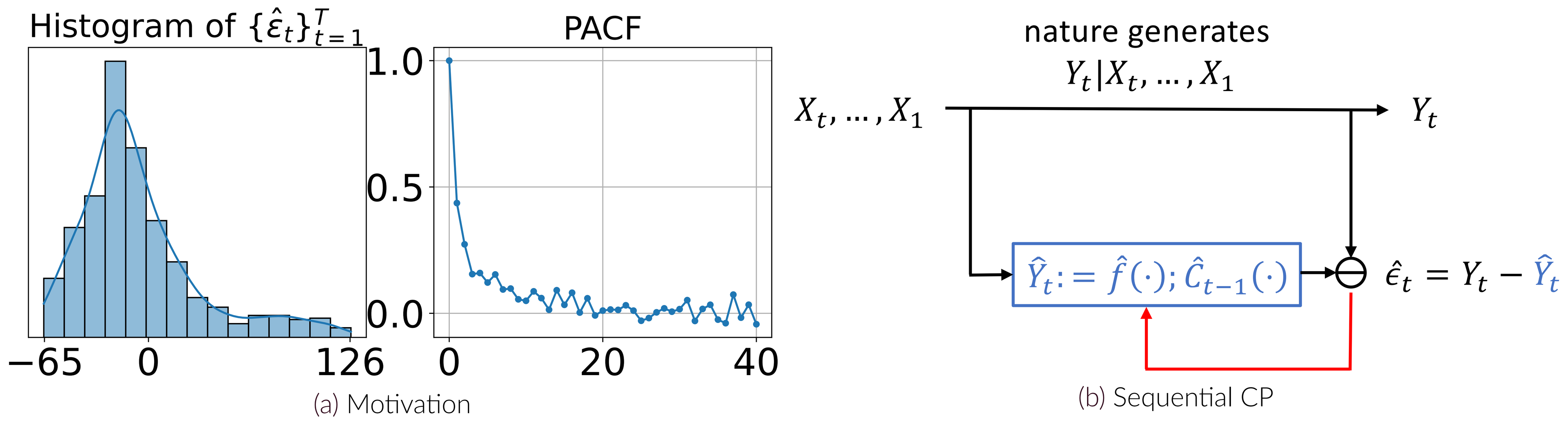


Figure 1. Motivation and the sequential CP paradigm (e.g., `SPCI`). In (a), we visualize the histogram and partial auto-correlation functions of residuals, which are asymmetrically distributed with non-negligible dependency. In (b), the sequential CP framework leverages new prediction residuals when constructing the time-adaptive prediction intervals.

$$\widehat{C}_{t-1}(X_t) := [\hat{f}(X_t) + \widehat{Q}_t(\hat{\beta}), \hat{f}(X_t) + \widehat{Q}_t(1 - \alpha + \hat{\beta})], \tag{2}$$

where $\hat{\beta}$ minimizes interval width empirically. In traditional CP, $\widehat{Q}_t(\cdot)$ is the empirical quantile function.

Table 1. Marginal coverage and width by all methods on three real time series. The target coverage is 0.9, and entries in the bracket indicate standard deviation over three runs. `SPCI` outperforms competitors with a much narrower interval width and does not lose coverage.

	Wind coverage	Wind width	Electric coverage	Electric width	Solar coverage	Solar width
SPCI	0.95 (1.50e-2)	2.65 (1.60e-2)	0.93 (4.79e-3)	0.22 (1.68e-3)	0.91 (1.12e-2)	47.61 (1.33e+0)
EnbPI	0.93 (6.20e-3)	6.38 (3.01e-2)	0.91 (6.84e-4)	0.32 (9.11e-4)	0.88 (4.25e-3)	48.95 (3.38e+0)
AdaptiveCI	0.95 (5.37e-3)	9.34 (3.56e-2)	0.95 (1.81e-3)	0.51 (7.25e-3)	0.96 (1.39e-2)	56.34 (1.15e+0)
NEX-CP	0.96 (8.21e-3)	6.68 (7.73e-2)	0.90 (2.05e-3)	0.45 (2.16e-3)	0.90 (7.73e-3)	102.80 (5.25e+0)
DeepAR	0.95 (5.32e-3)	6.86 (7.86e-3)	0.91 (3.45e-3)	0.62 (2.56e0-3)	0.92 (5.35e-3)	80.23 (4.94e+0)
TFT	0.92 (6.34e-2)	7.56 (5.34e-3)	0.95 (2.34e-2)	0.66 (2.34e-3)	0.93 (2.84e-3)	74.82 (4.23e+0)

We remark that `SPCI` intervals in (2) differ from conformal quantile regression [5], which directly fits conditional quantile functions on the *response variables* Y. `SPCI` also differs from probabilistic forecasting approaches [3, 6] as our method is compatible with any user-specified point prediction model, remains distribution-free, and provides coverage guarantees.

Theoretical analyses: We prove (1) in this work. Specifically, we extend the consistency analyses of QRF estimators for i.i.d. data to dependent sequences. To do so, we primarily impose constraints on the covariance function over residuals to avoid strong dependency. Details can be found in the proof of Theorem 2 in the work.

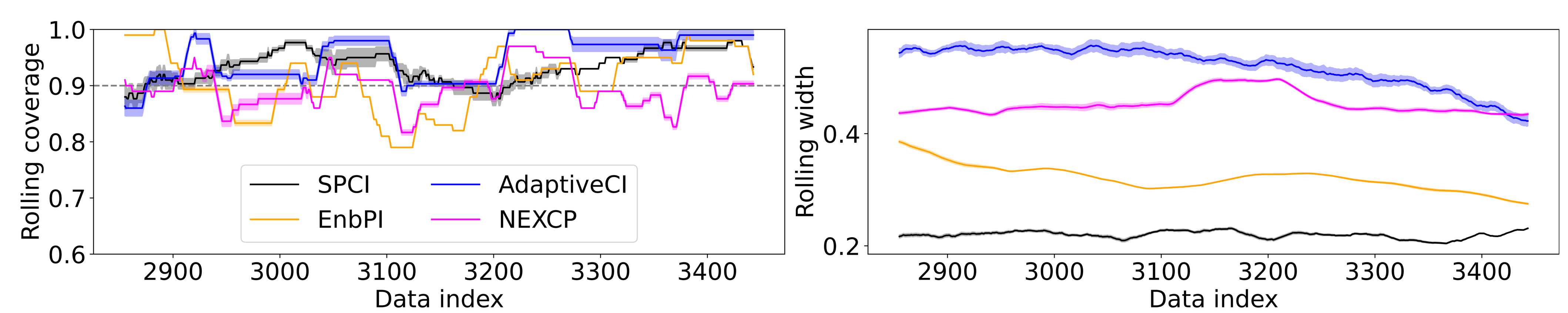


Figure 2. Rolling coverage and interval width over the electric time-series by different methods. `SPCI` in black not only yields valid rolling coverage but also consistently yields the narrowest prediction intervals. Furthermore, the variance of `SPCI` results over trials is also small, as shown by the shaded regions over coverage and width results.

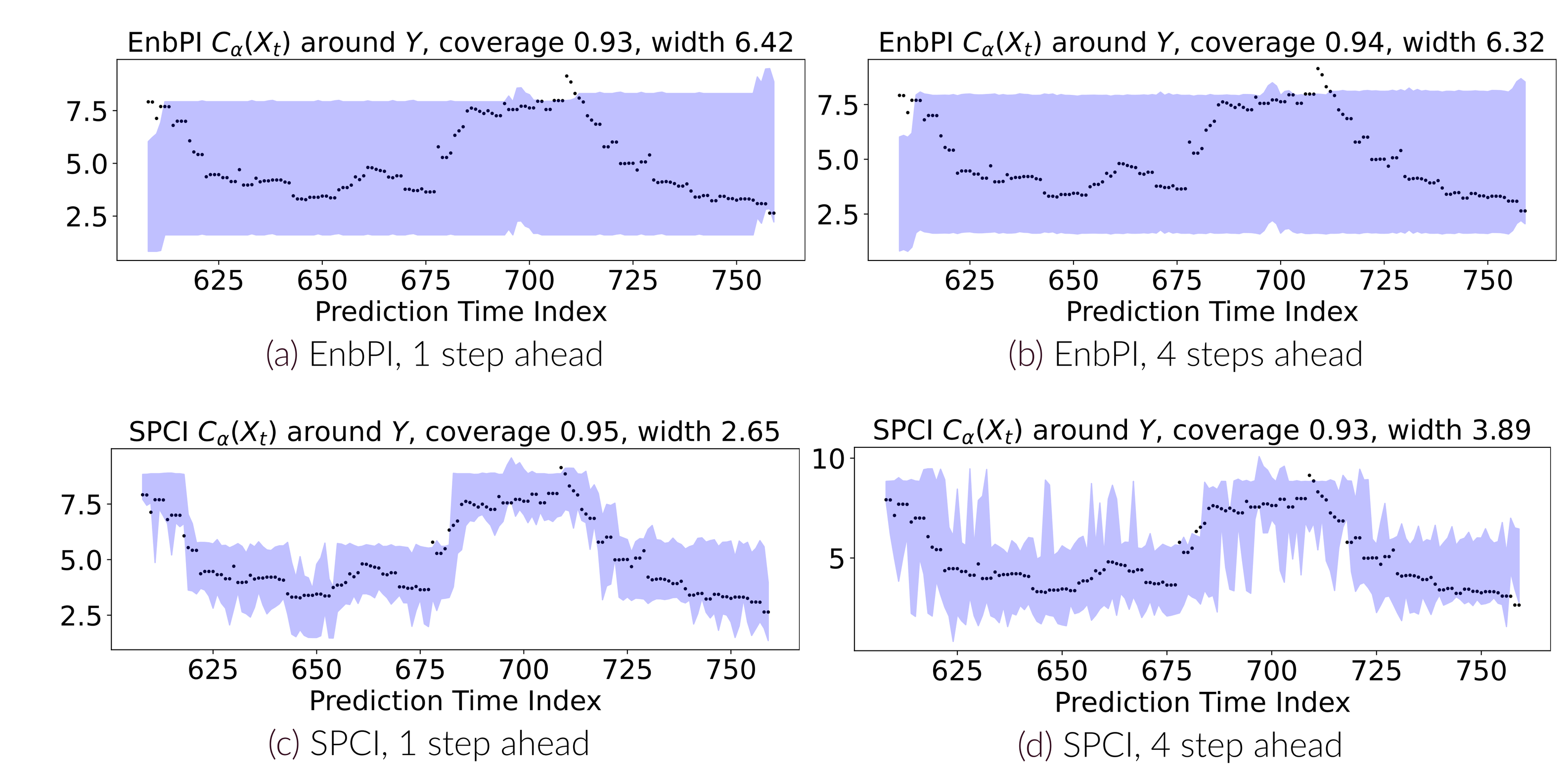
Experiments

We demonstrate the advantage of `SPCI` against CP methods and existing probabilistic forecasting approaches based on deep neural networks (NN), in terms of narrower intervals under valid empirical coverage. We also demonstrate the benefit of `SPCI` in multi-step ahead predictive inference (see Figure 3).

We specifically compare with three CP approaches: EnbPI [7, 8], AdaptiveCI [2], and NEX-CP [1],as well as two NN-based approaches: TFT [3] and DeepAR [6].

Table 2. Simulation on non-stationary time-series with $\alpha = 0.1$. We show that `SPCI` outperforms baselines in terms of interval width without sacrificing valid coverage.

SPCI		EnbPI		AdaptiveCI		NEX-CP	
Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width
0.92 (2.75e-3)	12.96 (2.56e-2)	0.90 (2.21e-3)	25.41 (4.79e-2)	0.90 (4.12e-3)	28.00 (5.81e-2)	0.93 (3.10e-3)	46.50 (6.29e-2)



(a) EnbPI, 1 step ahead (b) EnbPI, 4 steps ahead (c) SPCI, 1 step ahead (d) SPCI, 4 step ahead

Figure 3. Multi-step ahead prediction interval construction by `SPCI` and `EnbPI` on wind speed data. Compared to `EnbPI` results in subfigures (a) and (b), `SPCI` intervals are much narrower and more adaptive—`SPCI` intervals follow the trajectory of the time-series whereas `EnbPI` ones are overly conservative. In addition, `SPCI` interval increase in width as the predictive horizon increases, reflecting the existence of more uncertainty in long horizons.

Future directions

We aim to extend `SPCI` to multi-dimensional confidence region construction and refine the multi-step ahead predictive inference approach.

References

[1] Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Conformal prediction beyond exchangeability. *arXiv preprint arXiv:2202.13415*, 2022.

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

[3] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

[4] Nicolai Meinshausen. Quantile regression forests. *J. Mach. Learn. Res.*, 7:983–999, 2006.

[5] Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, pages 3543–3553, 2019.

[6] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

[7] Chen Xu and Yao Xie. Conformal prediction interval for dynamic time-series. In *International Conference on Machine Learning*, pages 11559–11569. PMLR, 2021.

[8] Chen Xu and Yao Xie. Conformal prediction for time series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45 (10):11575–11587, 2023. doi:10.1109/TPAMI.2023.3272339.