Time-series Uncertainty Quantification using Conformal Prediction

Github Repository: https://github.com/hamrel-cxu/EnbPI

Chen Xu, Yao Xie
Industrial and Systems Engineering (ISyE)
Georgia Institute of Technology
cxu310@gatech.edu, yao.xie@isye.gatech.edu

Georgia Tech | ISyE H. Milton Stewart School of Industrial and Systems Engineering

1. Introduction and Problem Setup

Goal and Challenges: We adopt ideas in conformal prediction to construct prediction intervals for time-series. There are typically four challenges:

- *Time-series are dynamic* (e.g. highly stochastic and non-stationary).
- *Spatio-temporal Correlation* exists within data from different location.
- *Ensemble ML models* are computationally expensive if one retrains the ensembles to build intervals.
- *Distribution-free Intervals* are desirable, given unknown data distribution.

Contribution: Conformal prediction methods are distribution-free but likely have no coverage guarantee beyond exchangeable data. In contrast, we

- Proposes *Ensemble Batch Prediction Interval* (**EnbPI**), an efficient wrapper around any *ensemble model that has been trained.*
- Theoretically, prediction intervals enjoy *approximate marginal coverage*, which is exact asymptotically.
- Empirically, **EnbPI** reaches *exact marginal coverage* and even conditional coverage on many tasks. Modifications can handle problems with missing data and perform anomaly detection

Setup: Assume for any $t \geq 1$, $Y_t = f(X_t) + \epsilon_t$, where $Y_t \in \mathbb{R}, X_t \in \mathbb{R}^d, \epsilon_t \sim F$ and training data $\{(X_t, Y_t)\}_{t=1}^T$ are available.

For each $t > T$, we build a $(1-\alpha)$-level prediction interval $C^\alpha_{T,t}$ so that

$$P\left(Y_t \in C^\alpha_{T,t}\right) \geq 1 - \alpha, \qquad (1)$$

where $C^\alpha_{T,t}$ indicates previous T data are used to construct the interval at level $1-\alpha$. We call (1) as marginal coverage.

2. Algorithm

Fix a regression model $\mathscr{A} : \{\mathbb{R}^d \times \mathbb{R}\}^N \to (\mathbb{R}^d \to \mathbb{R})$ that outputs predictors $\hat{f}$. We construct the intervals for $t > T$ as

$$C^\alpha_{T,t} := \hat{f}_{-t}(X_t) \pm (1-\alpha) \text{ quantile of } \{\hat{\epsilon}_i\}_{i=t-1}^{t-T}, \qquad (2)$$

where residuals $\hat{\epsilon}_i := |Y_i - \hat{f}_{-i}(X_i)|$. To avoid overfitting, we require that datum (X_j, Y_j) is never used to train $\hat{f}_{-j}$ for any $j \geq 1$. Then, **EnbPI** consists of three steps:

1. (Training) Fit B bootstrap models $\hat{f}^b := \mathscr{A}(\{(X_i, Y_i) \mid i \in S_b\})$, where S_b is an index set sampled with replacement T indices from $[T]$.
2. (Training) Initialize $\boldsymbol{\epsilon} = \{\}$. For $i \in [T]$. Compute the "leave-one-out" ensemble model $\hat{f}^\phi_{-i}(X_i) := \phi\left(\left\{\hat{f}^b(X_i) \mid i \notin S_b\right\}\right)$ via the aggregation function ϕ (e.g. mean or median). Compute and save residuals $\hat{\epsilon}_i$ in $\boldsymbol{\epsilon}$.
3. (Prediction) For $t > T$, let $\hat{f}^\phi_{-t}(X_t) := (1-\alpha)$ quantile of $\left\{\hat{f}^\phi_{-i}(X_t)\right\}_{i=1}^T$.
   Compute $C^\alpha_{T,t}$ as in (2) using $\hat{f}^\phi_{-t}(X_t)$ and $\boldsymbol{\epsilon} = \{\hat{\epsilon}_i\}_{i=t-1}^{t-T}$.
   Whenever new observations Y_t are available as feedback, compute $\hat{\epsilon}_t := |Y_t - \hat{f}^\phi_{-t}(X_t)|$, let $\boldsymbol{\epsilon} = \left(\boldsymbol{\epsilon} - \{\hat{\epsilon}_1\}\right) \cup \{\hat{\epsilon}_t\}$, and reset indices of $\boldsymbol{\epsilon}$.

3. Theoretical Guarantee

Without loss of generality, we analyze the coverage at $t = T+1$, but the Theorem holds at any $t > T$.

Assumptions:

1. *Data Regularity:* The error process $\{\epsilon_t\}_{t\geq 1}$ is stationary, strongly mixing, with bounded sum of mixing coefficients. The common CDF F is Lipschitz with $L > 0$.
2. *Estimation Quality:* There exists a real sequence $\{\delta_T\}_{T\geq 1}$ that converges to zero such that $\sum_{t=1}^T (\hat{\epsilon}_t - \epsilon_t)^2 / T \leq \delta_T^2$.

Theorem (Informally): *For any $\alpha \in (0,1)$, the prediction interval $C^\alpha_{T,T+1}(X_t)$ from EnbPI satisfies*

$$\left| P\left(Y_{T+1} \notin C^\alpha_{T,T+1}\right) - \alpha \right| \leq C\left((\log T/T)^{1/3} + \delta_T^{2/3}\right). \qquad (3)$$

C can be identified under Assumption 1 and 2.

Implications:

1. *RHS in (3):* It is a worst-case analysis; empirical coverage is almost always $1-\alpha$ over all test points. Meanwhile, different error assumptions (e.g., independent, stationary, etc.) yield factors different from $(\log T/T)^{1/3}$ (Xu et al., 2021a, Corollary 1—3).
2. *Assumptions:* 1 allows arbitrary data dependency and 2 holds true for many classes of f and algorithms $\mathscr{A}$ (Xu et al., 2021a, Section 4.2).
3. *Theorem Applicability:* It holds for split/inductive conformal prediction (ICP) (Papadopoulos et al., 2007). However, we advise using **EnbPI** as
   1. T in the Theorem becomes the size of *calibration data*.
   2. LOO ensemble predictor $\hat{f}_{-i}$ are in general better predictors.

4. Experiments

We show prediction intervals of **EnbPI** in various settings. We focus on solar radiation level prediction. In particular,

- Figure 1 compares **EnbPI** with traditional time-series methods.
- Figure 2 compares **EnbPI** with existing conformal prediction methods.
- Figure 3 shows the conditional coverage performance of **EnbPI.**

It is consistent that **EnbPI** always maintain valid coverage for time-series data. This is true at any significant level α for any regression model.

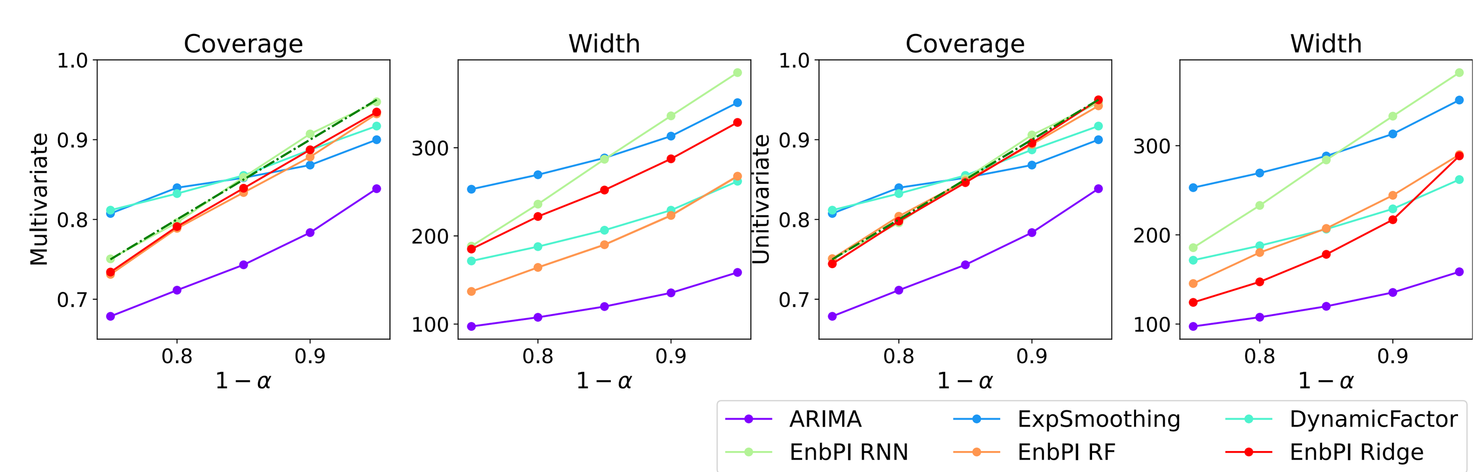

Figure 1: Marginal coverage against traditional time-series methods. We see that **EnbPI** yields consistent coverage when competing methods may fail or yield more conservative intervals

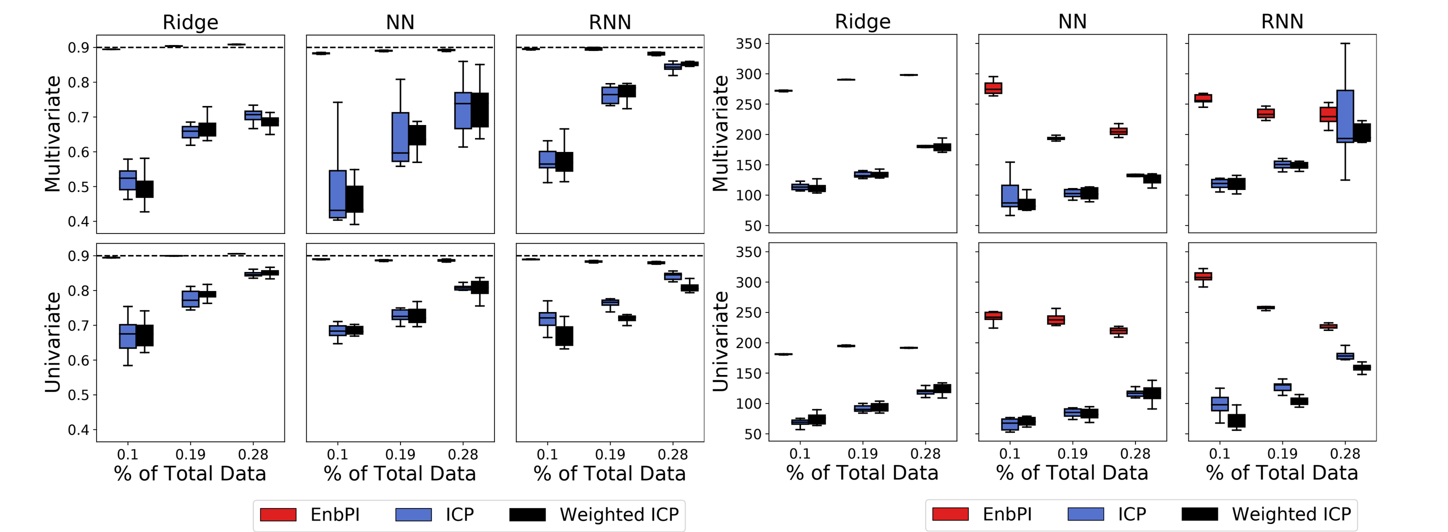

Figure 2: Marginal coverage against existing conformal prediction methods. Each box contains results from 10 independent trials. **EnbPI** intervals always satisfy valid coverage.

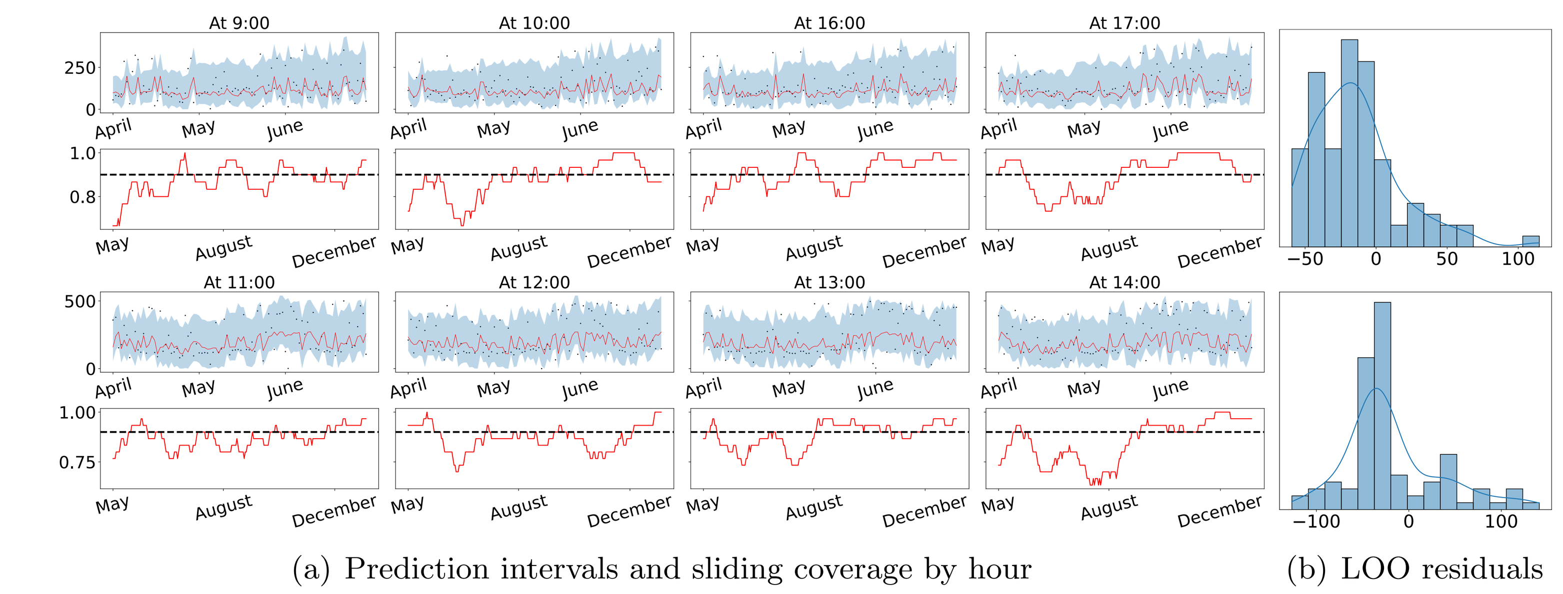

Figure 3: Conditional coverage of **EnbPI** on solar radiation level prediction. The coverage is evaluated at each hour, when the training data contain data from all hours.

5. Extensions

We are actively exploring applying EnbPI and its variants:

- Figure 4 considers multi-step ahead prediction.
- Time-series anomaly detection for traffic flows (Xu et al., 2021b).
- Time-series prediction set for classification (Xu et al., 2022).
- The possibility for change-point detection (cf. our journal version).

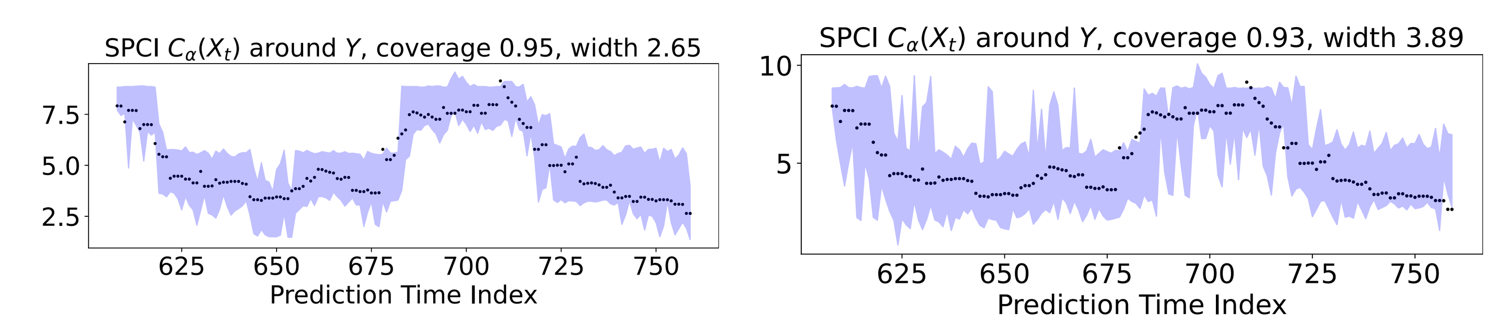

Figure 4: Multi-step ahead prediction based on an improved version of **EnbPI**, which we abbreviated SPCI. Paper to appear on arxiv soon.

References

Xu, Chen and Yao Xie (2021a). "Conformal prediction interval for dynamic time-series" oral paper/long talk in the Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021 (ICML 2021). Journal version under revision.

Xu, Chen and Yao Xie (2021b). "Conformal Anomaly Detection on Spatio-Temporal Observations with Missing Data" In: ICML 2021 Workshop—Distribution-free Uncertainty Quantification.

Xu, Chen and Yao Xie (2022). "Conformal prediction set for time-series" In: ICML 2022 Workshop—Distribution-free Uncertainty Quantification.

Tibshirani, R. J., Barber, R. F., Candes, E., and Ramdas, A. (2019). "Conformal prediction under covariate shift" In: Advances in Neural Information Processing Systems 32 (NeurIPS 2019), pages 2530–2540 .

Papadopoulos, H., Vovk, V., and Gammerman, A. (2007) "Conformal prediction with neural networks" In: 19th IEEE International Conference on Tools with Artificial Intelligence(ICTAI 2007), volume 2, pages 388–395.