Sequential conformal prediction for time series

Chen Xu, Yao Xie

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

References

The talk is based on

- (EnbPI) Xu, C. and Xie, Y. Conformal prediction for time-series
  1. Conference version: *International Conference on Machine Learning, pp. 11559–11569. PMLR, 2021 (ICML 2021 Oral).*
  2. Journal version: *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(10):11575–11587, 2023 (IEEE TPAMI).*
- (SPCI) Xu, C. and Xie, Y. Sequential predictive conformal inference for time series, *International Conference on Machine Learning, pp. 38707–38727. PMLR, 2023 (ICML 2023).*
- (MultiDimSPCI) Xu, C., Jiang, H., and Xie, Y. Conformal prediction for multi-dimensional time series by ellipsoidal sets, *International Conference on Machine Learning, 2024 (ICML 2024 Spotlight).*

Outline

- Motivation and Goal
- Warm-up example
- Conformal Prediction: exchangeable (i.i.d.) data
- Conformal Prediction: time-series
- Theoretical guarantee
- Numerical results
- Conclusion

Motivation

- Forecasting is everywhere.
  Ex. My black-box NN model predicts the financial return to be 5% tomorrow.

Motivation

- Forecasting is everywhere.
  Ex. My black-box NN model predicts the financial return to be 5% tomorrow.
- Yet, uncertainty naturally exists.
  Ex. Low signal-to-noise ratio, high variation, randomness in model fitting...

Motivation

- Forecasting is everywhere.
  Ex. My black-box NN model predicts the financial return to be 5% tomorrow.
- Yet, uncertainty naturally exists.
  Ex. Low signal-to-noise ratio, high variation, randomness in model fitting...
- How can we quantify such uncertainty?
  Ex. Can we say with high probability, what the true return would be?

Goal

- Uncertainty quantification (UQ) for forecasting problem.
- Data: $(X_i, Y_i), X_i \in \mathbb{R}^d, Y_i \in \mathbb{R}$.
  X_i is feature vector and Y_i is uni-variate response.

Goal

- Uncertainty quantification (UQ) for forecasting problem.
- Data: $(X_i, Y_i), X_i \in \mathbb{R}^d, Y_i \in \mathbb{R}$.
  X_i is feature vector and Y_i is uni-variate response.
- Given a significance level $\alpha \in [0, 1]$ and new feature X_j, produce a prediction interval $C(X_j, \alpha)$:

$$\mathbb{P}(Y_j \in C(X_j, \alpha)) \geq 1 - \alpha. \tag{1}$$

Goal

- Uncertainty quantification (UQ) for forecasting problem.
- Data: $(X_i, Y_i), X_i \in \mathbb{R}^d, Y_i \in \mathbb{R}$.
  X_i is feature vector and Y_i is uni-variate response.
- Given a significance level $\alpha \in [0, 1]$ and new feature X_j, produce a prediction interval $C(X_j, \alpha)$:

$$\mathbb{P}(Y_j \in C(X_j, \alpha)) \geq 1 - \alpha. \quad (1)$$

- Example: with 95% probability (i.e., $\alpha = 0.05$), true stock price (i.e., Y_j) is within the constructed interval (i.e., $C(X_j, \alpha)$).

Goal

- Uncertainty quantification (UQ) for forecasting problem.
- Data: $(X_i, Y_i), X_i \in \mathbb{R}^d, Y_i \in \mathbb{R}$.
  X_i is feature vector and Y_i is uni-variate response.
- Given a significance level $\alpha \in [0, 1]$ and new feature X_j, produce a prediction interval $C(X_j, \alpha)$:

$$\mathbb{P}(Y_j \in C(X_j, \alpha)) \geq 1 - \alpha. \tag{1}$$

- Example: with 95% probability (i.e., $\alpha = 0.05$), true stock price (i.e., Y_j) is within the constructed interval (i.e., $C(X_j, \alpha)$).
- Can be generalized to confidence region for multivariate outputs.

Warm-up example

- Suppose
  (1) Data (X_i, Y_i) are i.i.d, $i = 1, \ldots, n$.
  (2) Linear model $Y_i = X_i^T \beta + \epsilon_i, \epsilon_i \sim N(0, 1)$.

Warm-up example

- Suppose
  (1) Data (X_i, Y_i) are i.i.d, $i = 1, \ldots, n$.
  (2) Linear model $Y_i = X_i^T \beta + \epsilon_i, \epsilon_i \sim N(0, 1)$.
- Then we know that
  (1) $\hat{Y}_j = X_j^T \hat{\beta}, \hat{\beta} = (X^T X)^{-1} X^T Y$.
  (2) $C(X_j, \alpha) = \hat{Y}_j \pm t(\alpha, n) \times \mathrm{S.E.}(\hat{\mathrm{Y}})$.

Warm-up example

- Suppose
  (1) Data (X_i, Y_i) are i.i.d, $i = 1, \ldots, n$.
  (2) Linear model $Y_i = X_i^T \beta + \epsilon_i, \epsilon_i \sim N(0, 1)$.
- Then we know that
  (1) $\hat{Y}_j = X_j^T \hat{\beta}, \hat{\beta} = (X^T X)^{-1} X^T Y$.
  (2) $C(X_j, \alpha) = \hat{Y}_j \pm t(\alpha, n) \times \mathrm{S.E.}(\hat{\mathrm{Y}})$.
- Main limitation:
  (1) Linear $X \to Y$ with known distribution of Y.
  (2) No model mis-specification for forecasting.
  (3) Data are independent.

Warm-up example

- Suppose
  (1) Data (X_i, Y_i) are i.i.d, $i = 1, \ldots, n$.
  (2) Linear model $Y_i = X_i^T \beta + \epsilon_i, \epsilon_i \sim N(0, 1)$.
- Then we know that
  (1) $\hat{Y}_j = X_j^T \hat{\beta}, \hat{\beta} = (X^T X)^{-1} X^T Y$.
  (2) $C(X_j, \alpha) = \hat{Y}_j \pm t(\alpha, n) \times \text{S.E.}(\hat{\text{Y}})$.
- Main limitation:
  (1) Linear $X \to Y$ with known distribution of Y.
  (2) No model mis-specification for forecasting.
  (3) Data are independent.
- Can we have something that circumvent the limitations above?
  We will see that *conformal prediction* can address (1) and (2) with theoretical guarantees, where we further address (3).

Solution: conformal prediction (CP)

- In a nutshell, conformal prediction first defines "conformity scores" $s(X, Y, \hat{f})$ (e.g., $s = Y - \hat{f}(X)$). Then,

$$C(X_j, \alpha) = \{y \in \mathbb{R} : s(X_j, y, \hat{f}) \leq Q_{1-\alpha}(\mathcal{S})\}. \tag{1}$$

$\mathcal{S} = \{s(X_i, Y_i, \hat{f})\}$ computed on a "hold-out" set and $Q_{1-\alpha}$ returns the empirical quantile.

Solution: conformal prediction (CP)

- In a nutshell, conformal prediction first defines "conformity scores" $s(X, Y, \hat{f})$ (e.g., $s = Y - \hat{f}(X)$). Then,

$$C(X_j, \alpha) = \{y \in \mathbb{R} : s(X_j, y, \hat{f}) \leq Q_{1-\alpha}(\mathcal{S})\}. \quad (1)$$

  $\mathcal{S} = \{s(X_i, Y_i, \hat{f})\}$ computed on a "hold-out" set and $Q_{1-\alpha}$ returns the empirical quantile.
- Thus, (1) includes all y with scores "conforming" to the past.
- The idea mainly originated in (Papadopoulos et al., 2007), with quite extensive development so far. More comprehensive review can be found in (Angelopoulos & Bates, 2023).

Solution: conformal prediction (CP)

- In a nutshell, conformal prediction first defines "conformity scores" $s(X, Y, \hat{f})$ (e.g., $s = Y - \hat{f}(X)$). Then,

$$C(X_j, \alpha) = \{y \in \mathbb{R} : s(X_j, y, \hat{f}) \leq Q_{1-\alpha}(\mathcal{S})\}. \tag{1}$$

  $\mathcal{S} = \{s(X_i, Y_i, \hat{f})\}$ computed on a "hold-out" set and $Q_{1-\alpha}$ returns the empirical quantile.

- Thus, (1) includes all y with scores "conforming" to the past.
- The idea mainly originated in (Papadopoulos et al., 2007), with quite extensive development so far. More comprehensive review can be found in (Angelopoulos & Bates, 2023).
- Intervals using (1) are
  (a) distribution-free in Y .
  (b) model-free in $\hat{f}$.
  (c) theoretically valid with $1 - \alpha$ coverage.

Limitation of CP

- Main limitation is the *dependency* assumption on data (X_i, Y_i).

Limitation of CP

- Main limitation is the *dependency* assumption on data (X_i, Y_i).
- CP requires (X_i, Y_i) are *exchangeable* (e.g., i.i.d.), which means for a sequence of r.v.s $Z_i, i \geq 1$ and a permutation σ:

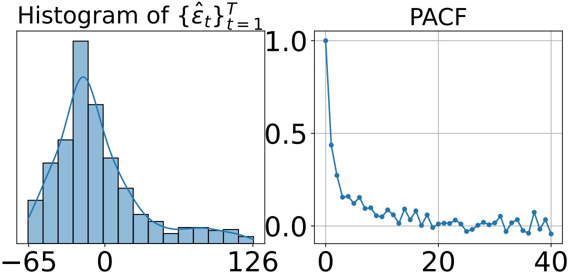
$$(Z_1, Z_2, Z_3, \ldots) \stackrel{d}{=} (Z_{\sigma(1)}, Z_{\sigma(2)}, Z_{\sigma(3)}, \ldots)$$

Limitation of CP

- Main limitation is the *dependency* assumption on data (X_i, Y_i).
- CP requires (X_i, Y_i) are *exchangeable* (e.g., i.i.d.), which means for a sequence of r.v.s $Z_i, i \geq 1$ and a permutation σ:

$$(Z_1, Z_2, Z_3, \ldots) \stackrel{d}{=} (Z_{\sigma(1)}, Z_{\sigma(2)}, Z_{\sigma(3)}, \ldots)$$

- However, exchangeability is RARE for time-series, due to strong serial correlation.



Extend CP beyond i.i.d.

- How to extend CP in this more general setting is a key and relatively new question.

Extend CP beyond i.i.d.

- How to extend CP in this more general setting is a key and relatively new question.
- Some concurrent works:
  (1) Adaptively adjust α_t depending on $\mathbb{1}(Y_j \in C(X_j, \alpha_t))$
  – Related works: (Gibbs & Candes, 2021; Zaffran et al., 2022).
  (2) Re-weighting in $Q_{1-\alpha}(\mathcal{S})$ based on "difference" between training and testing data.
  – Related works: (Tibshirani et al., 2019; Barber et al., 2023).

Extend CP beyond i.i.d.

- How to extend CP in this more general setting is a key and relatively new question.
- Some concurrent works:
  (1) Adaptively adjust α_t depending on $\mathbb{1}(Y_j \in C(X_j, \alpha_t))$
  – Related works: (Gibbs & Candes, 2021; Zaffran et al., 2022).
  (2) Re-weighting in $Q_{1-\alpha}(\mathcal{S})$ based on "difference" between training and testing data.
  – Related works: (Tibshirani et al., 2019; Barber et al., 2023).
- Our work is one of the first to understand this problem, which is not in conflict with these works (can be combined).

Extend CP beyond i.i.d.

- How to extend CP in this more general setting is a key and relatively new question.
- Some concurrent works:
  (1) Adaptively adjust α_t depending on $\mathbb{1}(Y_j \in C(X_j, \alpha_t))$
  – Related works: (Gibbs & Candes, 2021; Zaffran et al., 2022).
  (2) Re-weighting in $Q_{1-\alpha}(\mathcal{S})$ based on "difference" between training and testing data.
  – Related works: (Tibshirani et al., 2019; Barber et al., 2023).
- Our work is one of the first to understand this problem, which is not in conflict with these works (can be combined).
- In retrospect, our method can be viewed as a type of re-weighting scheme in (2).

Initial solution, EnbPI

- We consider *time-series* data, where X_i can be $(Y_{i-1}, \ldots, Y_{i-d})$ and/or contains other features.

Initial solution, EnbPI

- We consider *time-series* data, where X_i can be $(Y_{i-1}, \ldots, Y_{i-d})$ and/or contains other features.
- The method builds on (Kim et al., 2020) to build "ensemble" models that increase the accuracy of $\hat{f}$ *without* hold-out set.

Initial solution, EnbPI

- We consider *time-series* data, where X_i can be $(Y_{i-1}, \ldots, Y_{i-d})$ and/or contains other features.
- The method builds on (Kim et al., 2020) to build "ensemble" models that increase the accuracy of $\hat{f}$ *without* hold-out set.
- Suppose conformity score $s(X, Y, \hat{f}) = \hat{\epsilon}_t = Y - \hat{f}(X)$,

$$C(X_t, \alpha) = [\hat{f}(X_t) + Q_{\beta_t^*}(\mathcal{S}_t), \hat{f}(X_t) + Q_{1-\alpha+\beta_t^*}(\mathcal{S}_t)]. \quad (1)$$

where β_t^* tries to minimized interval width and $\mathcal{S}_t$ incorporates latest $\hat{\epsilon}_t$ in a rolling fashion.

Initial solution, EnbPI

- We consider *time-series* data, where X_i can be $(Y_{i-1}, \dots, Y_{i-d})$ and/or contains other features.
- The method builds on (Kim et al., 2020) to build "ensemble" models that increase the accuracy of $\hat{f}$ *without* hold-out set.
- Suppose conformity score $s(X, Y, \hat{f}) = \hat{\epsilon}_t = Y - \hat{f}(X)$,

$$C(X_t, \alpha) = [\hat{f}(X_t) + Q_{\beta_t^*}(\mathcal{S}_t), \hat{f}(X_t) + Q_{1-\alpha+\beta_t^*}(\mathcal{S}_t)]. \quad (1)$$

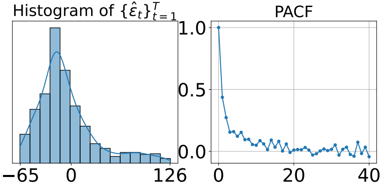
  where β_t^* tries to minimized interval width and $\mathcal{S}_t$ incorporates latest $\hat{\epsilon}_t$ in a rolling fashion.
- We can show when asymptotic coverage of $C(X_t, \alpha)$ under consistent training of f with $Y = f(X) + \epsilon$.
- Details can be found in works below (Xu & Xie, 2021, 2023a).

(Xu & Xie, 2021) Conformal prediction interval for dynamic time-series, in ICML 2021 (long presentation).

(Xu & Xie, 2023) Conformal prediction for time-series, in IEEE TPAMI.

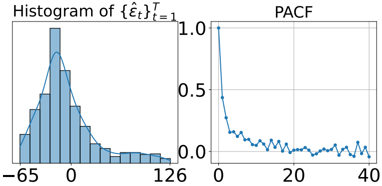
Latest solution, SPCI

- We could do better, by leveraging the dependency among residuals (conformity scores) $\hat{\epsilon}_t$:



Latest solution, SPCI

- We could do better, by leveraging the dependency among residuals (conformity scores) $\hat{\epsilon}_t$:



- In particular, we would adaptively re-estimate the *quantile* of $\hat{\epsilon}_t$ via a quantile regression model fitted on residuals :

$$C(X_t, \alpha) = [\hat{f}(X_t) + \hat{\mathcal{Q}}_{\beta_t^*}(\mathcal{S}_t), \hat{f}(X_t) + \hat{\mathcal{Q}}_{1-\alpha+\beta_t^*}(\mathcal{S}_t)]. \quad (1)$$

- Details can be found in (Xu & Xie, 2023b).

(Xu & Xie, 2023b) Sequential predictive conformal inference for time series, in ICML 2023.

Latest solution, SPCI

- Prediction intervals in `SPCI`:

$$C(X_t, \alpha) = [\hat{f}(X_t) + \hat{\mathcal{Q}}_{\beta_t^*}(\mathcal{S}_t), \hat{f}(X_t) + \hat{\mathcal{Q}}_{1-\alpha+\beta_t^*}(\mathcal{S}_t)]. \quad (1)$$

- We used quantile random forest (QRF) (Meinshausen, 2006) as $\mathcal{Q}$ and train it auto-regressively on $\mathcal{S}_t$.

Latest solution, SPCI

- Prediction intervals in `SPCI`:

$$C(X_t, \alpha) = [\hat{f}(X_t) + \hat{\mathcal{Q}}_{\beta_t^*}(\mathcal{S}_t), \hat{f}(X_t) + \hat{\mathcal{Q}}_{1-\alpha+\beta_t^*}(\mathcal{S}_t)]. \quad (1)$$

- We used quantile random forest (QRF) (Meinshausen, 2006) as $\mathcal{Q}$ and train it auto-regressively on $\mathcal{S}_t$.
- We proved asymptotic conditional coverage guarantees of (1).

Latest solution, SPCI

- Prediction intervals in SPCI:

$$C(X_t, \alpha) = [\hat{f}(X_t) + \hat{\mathcal{Q}}_{\beta_t^*}(\mathcal{S}_t), \hat{f}(X_t) + \hat{\mathcal{Q}}_{1-\alpha+\beta_t^*}(\mathcal{S}_t)]. \quad (1)$$

- We used quantile random forest (QRF) (Meinshausen, 2006) as $\mathcal{Q}$ and train it auto-regressively on $\mathcal{S}_t$.
- We proved asymptotic conditional coverage guarantees of (1).
- Estimated residuals are weighted quantiles on $\mathcal{S}_t$ and have an interesting connection with (Barber et al., 2023).
- This is different from *conformalized quantile regression* (Romano et al., 2019) where $\mathcal{Q}$ is fitted on original (X_i, Y_i).

(Xu & Xie, 2023b) Sequential predictive conformal inference for time series, in ICML 2023.

SPCI for multi-dimensional time-series

- When $Y_t \in \mathbb{R}^p$ is a vector, we build prediction regions $C(X_t, \alpha) \subset \mathbb{R}^p$ subject to the same coverage guarantee.

SPCI for multi-dimensional time-series

- When $Y_t \in \mathbb{R}^p$ is a vector, we build prediction regions $C(X_t, \alpha) \subset \mathbb{R}^p$ subject to the same coverage guarantee.
- The key lies in the design of conformity score, where we use

$$s(X, Y, \hat{f}) = \hat{\epsilon}_t^T \hat{\Sigma}_\rho^{-1} \hat{\epsilon}_t, \tag{1}$$

where $\hat{\Sigma}_\rho$ is the low-rank approximation of sample covariance matrix $\hat{\Sigma}$ of residuals. The proposed method is called `MultiDimSPCI` (Xu et al., 2024).

SPCI for multi-dimensional time-series

- When $Y_t \in \mathbb{R}^p$ is a vector, we build prediction regions $C(X_t, \alpha) \subset \mathbb{R}^p$ subject to the same coverage guarantee.
- The key lies in the design of conformity score, where we use

$$s(X, Y, \hat{f}) = \hat{\epsilon}_t^T \hat{\Sigma}_\rho^{-1} \hat{\epsilon}_t, \tag{1}$$

  where $\hat{\Sigma}_\rho$ is the low-rank approximation of sample covariance matrix $\hat{\Sigma}$ of residuals. The proposed method is called `MultiDimSPCI` (Xu et al., 2024).
- Compared to recent approaches (e.g., CopulaCPTS (Sun & Yu, 2024)), our $C(X_t, \alpha)$ is simpler and empirically performs better.

(Xu et al., 2024) Conformal prediction for multi-dimensional time series by ellipsoidal sets, in ICML 2024 (spotlight).

Theoretical guarantee: EnbPI

$$\hat{\epsilon}_t = Y_t - \hat{Y}_t = \underbrace{f_t(X_t) - \hat{f}_t(X_t)}_{\text{prediction error}} + \underbrace{\epsilon_t}_{\text{“nature”}}$$

Consider $f_t(X_t) = f(X_t)$:

- Analyze $t = T + 1$; can extend to $t > T + 1$
- Assumption 1 (Data regularity): Error process $\epsilon_1, \epsilon_2, \ldots$
  - stationary and strongly mixing
  - sum of mixing coefficients bounded by M
  - true CDF F is Lipschitz with constant $L > 0$
- Assumption 2 (Estimation quality)

$$\sum_{t=1}^{T} (\hat{f}_t(X_t) - f(X_t))^2 / T \leq \delta_T^2,$$

Theoretical guarantee (cont.)

- Given a training size T and $\alpha \in (0,1)$,

$$|\mathbb{P}(Y_{T+1} \notin \widehat{C}^{\alpha}_{T+1}) - \alpha| \leq C((\log T/T)^{1/3} + \delta_T^{2/3})$$

Implications

- Factor $(\log T/T)^{1/3}$ comes from assuming α-mixing errors, different error assumptions (e.g., independent, stationary, etc.) yield different rates
- Coverage gap dependent on T and accuracy of algorithm

Assumption 1 can be extended

- Independent $\{\epsilon_t\}_{t\geq 1}$

$$\text{Rate} = (\log(16T)/T)^{1/2}.$$

- Stationary linear processes $\epsilon_t = \sum_{j=1}^{\infty} \delta_j z_{t-j}$.

$$\text{Rate} = \log T/\sqrt{T}$$

  Faster than strongly mixing errors, slower than independent errors.

- Joint density of $\{\epsilon_t\}_{t=1}^{T+1}$ satisfies a logarithmic Sobolev inequality

$$\text{Rate} = (\log(cT)/T)^{1/3}$$

Assumption 2: “Good” predictive algorithm

- Assumption 2 holds true for many classes of algorithms
- No-free-lunch theorem:
  assumption on f is necessary in order for us to approximate it well.
- Examples
  - if f is sufficiently smooth,
    $$\delta_T = o(T^{-1/4})$$
    for neural networks sieve estimators (Chen and White, 1999).
  - If f is a sparse high-dimensional linear model,
    $$\delta_T = o(T^{-1/2})$$
    for Lasso and Dantzig selector (Bickel et al. 2009).

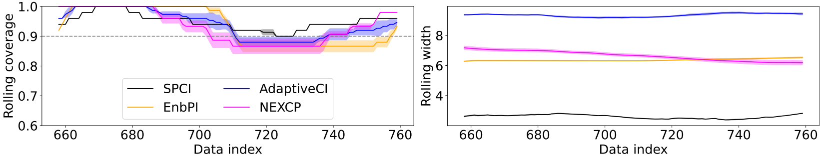
Empirical results: marginal coverage

- **Takeaway:** The conditional estimation of quantile of (unknown) residuals in `SPCI` significantly improves performance.

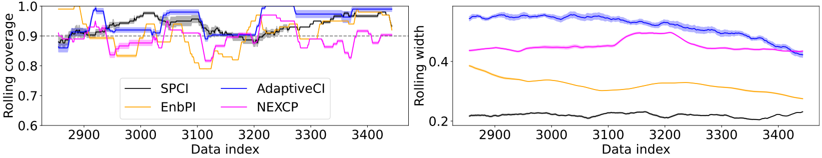
Table: Performance of `SPCI` against different methods on different datasets. Target coverage is 90%.

	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)

Empirical results: rolling coverage



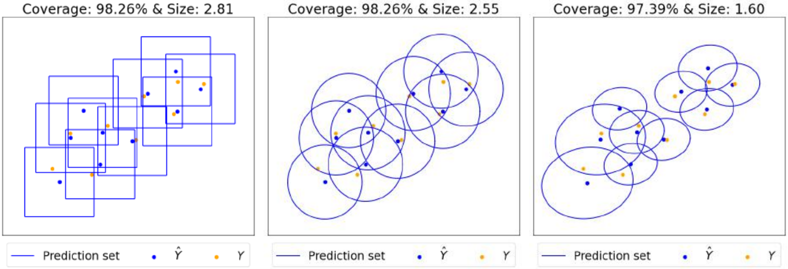
(a) Wind



(b) Electric

Figure: Comparison of rolling coverage of CP methods on two datasets. Window size is 50 or 100.

Multi-dimensional results in $\mathbb{R}^2$



(a) (b) (c)

Figure 1: Comparison of multivariate CP method on real two-dimensional wind data (see Section 5.2). Left (a): Empirical copula (Messoudi et al., 2021) which constructs coordinate-wise prediction intervals. Middle (b): Spherical confidence set introduced in (Sun & Yu, 2024). Right (c): our proposed ellipsoidal confidence set via `MultiDimSPCI`. While all methods yield coverage at least above the target 95% on test data, our method yields the smallest average size.

Multi-dimensional results beyond 2D.

Table 3: Real-data comparison of test coverage and average prediction set size by different methods. The target coverage is 0.95, and at each p, the smallest size of prediction sets is in **bold**. Our `MultiDimSPCI` yields the narrowest confidence sets without sacrificing coverage for two reasons. First, it explicitly captures dependency among coordinates of Y_t by forming ellipsoidal prediction sets. Second, it captures temporal dependency among non-conformity scores upon adaptive re-estimation of score quantiles.

(a) Wind data

Method	$p=2$ coverage	$p=2$ size	$p=4$ coverage	$p=4$ size	$p=8$ coverage	$p=8$ size
`MultiDimSPCI`	0.97	**1.60**	0.96	**7.02**	0.96	**72.10**
CopulaCPTS (Sun & Yu, 2024)	0.98	2.55	0.97	10.23	0.97	252.67
Local ellipsoid (Messoudi et al., 2022)	0.96	3.51	0.97	13.07	0.98	1.09e+3
Copula (Messoudi et al., 2021)	0.98	2.81	0.98	10.32	0.97	1.60e+3
TFT (Lim et al., 2021)	0.94	10.61	0.75	159.39	0.94	2.91e+4
DeepAR (Salinas et al., 2020)	0.96	7.07	0.76	67.97	0.96	1.79e+5

(b) Solar data

Method	$p=2$ coverage	$p=2$ size	$p=4$ coverage	$p=4$ size	$p=8$ coverage	$p=8$ size
`MultiDimSPCI`	0.96	1.68	0.96	**2.89**	0.97	**4.97**
CopulaCPTS (Sun & Yu, 2024)	0.99	4.36	0.99	37.56	0.99	3.28e+3
Local ellipsoid (Messoudi et al., 2022)	0.97	**1.32**	0.97	3.20	0.97	43.07
Copula (Messoudi et al., 2021)	0.99	4.11	0.99	27.73	0.99	1.42e+3
TFT (Lim et al., 2021)	0.99	13.68	0.99	71.72	0.93	1.19e+3
DeepAR (Salinas et al., 2020)	0.97	10.76	0.98	157.09	0.74	31.82

(c) Traffic data

Method	$p=2$ coverage	$p=2$ size	$p=4$ coverage	$p=4$ size	$p=8$ coverage	$p=8$ size
`MultiDimSPCI`	0.96	**1.31**	0.96	**1.93**	0.96	**2.98**
CopulaCPTS (Sun & Yu, 2024)	0.95	1.70	0.94	3.15	0.95	14.10
Local ellipsoid (Messoudi et al., 2022)	0.95	1.36	0.94	2.08	0.95	4.13
Copula (Messoudi et al., 2021)	0.95	1.44	0.95	3.90	0.94	40.60
TFT (Lim et al., 2021)	0.89	9.07	0.93	87.92	0.88	9.69e+2
DeepAR (Salinas et al., 2020)	0.87	13.53	0.88	57.20	0.82	9.89e+3

Additional collaborations

- Theoretical development of `SPCI` under kernel regression (Lee et al., 2024a), providing non-asymptotic marginal coverage gap.
- Transformer rather than quantile random forest inside SPCI (Lee et al., 2024b), capturing long-term dependency with improved empirical performances.

(Lee et al., 2024a) "Kernel-based optimally weighted conformal prediction intervals", *pre-print*.

(Lee et al., 2024b) "Transformer Conformal Prediction for Time Series", *pre-print, also in 2nd SPIGM @ ICML 2024*.

Summary

- Conformal prediction offers distribution-free and model-free uncertainty quantification with guarantees.
- We developed three approaches for CP in time-series:
  (1) EnbPI: rolling residual + ensemble predictor.
  (2) SPCI: adaptive re-estimation of residual quantiles.
  (3) MultiDimSPCI: multi-dimensional extension of SPCI.
- EnbPI has been implemented in **six** public packages: (1) MAPIE in scikit-learn. (2) Fortuna by Amazon AWS. (3) PUNCC by the Artificial and Natural Intelligence Toulouse Institute. (4) Functime for time-series machine learning at scale. (5) ConformalPrediction for Trustworthy AI in Julia. (6) skpro in sktime.

References I

Angelopoulos, A. N. and Bates, S. Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16 (4):494–591, 2023. ISSN 1935-8237. doi: 10.1561/2200000101. URL http://dx.doi.org/10.1561/2200000101.

Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani, R. J. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816 – 845, 2023. doi: 10.1214/23-AOS2276. URL https://doi.org/10.1214/23-AOS2276.

Gibbs, I. and Candes, E. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

Kim, B., Xu, C., and Barber, R. Predictive inference is free with the jackknife+-after-bootstrap. *Advances in Neural Information Processing Systems*, 33:4138–4149, 2020.

References II

Meinshausen, N. Quantile regression forests. *J. Mach. Learn. Res.*, 7:983–999, 2006.

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

Romano, Y., Patterson, E., and Candes, E. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, pp. 3543–3553, 2019.

Sun, S. H. and Yu, R. Copula conformal prediction for multi-step time series prediction. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=ojIJZDNIBj`.

Tibshirani, R. J., Barber, R. F., Candes, E., and Ramdas, A. Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems*, pp. 2530–2540, 2019.

References III

Xu, C. and Xie, Y. Conformal prediction interval for dynamic time-series. In *International Conference on Machine Learning*, pp. 11559–11569. PMLR, 2021.

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

Xu, C. and Xie, Y. Sequential predictive conformal inference for time series. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 38707–38727. PMLR, 23–29 Jul 2023b.

Xu, C., Jiang, H., and Xie, Y. Conformal prediction for multi-dimensional time series by ellipsoidal sets. In *Forty-first International Conference on Machine Learning*, 2024.

References IV

Zaffran, M., Féron, O., Goude, Y., Josse, J., and Dieuleveut, A. Adaptive conformal predictions for time series. In *International Conference on Machine Learning*, pp. 25834–25866. PMLR, 2022.