Conformal prediction for multi-dimensional time series by ellipsoidal sets

Chen Xu [1] Hanyang Jiang [1] Yao Xie [1]

[1]Georgia Institute of Technology

This work is supported by CMMI-2112533 and NSF DMS-2134037.

Motivation and Problem Setup

Conformal prediction (CP) has become a popular distribution-free technique to perform uncertainty quantification for complex machine learning algorithms. Recent works have especially developed sequential CP methods for time-series (e.g., `SPCI` [4]). However, most use cases focused on univariate time-series, where our primary interest in this work is to build prediction regions for multivariate time-series in the form of ellipsoids.

We assume a sequence of observations (X_t, Y_t), $t = 1, 2, \ldots$, where $Y_t \in \mathbb{R}^p$ are continuous p-dimensional outputs 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

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

We differ from past works in the following aspects. First, compared to copula-based methods [3] which build hyper-rectangular prediction regions, our construction of ellipsoids is more direct and simple, as we avoid having to optimize the copula choice and design. Second, compared to probabilistic forecasting approaches [1], our CP-based methods have theoretical guarantees and are model-agonostic. We will demonstrate empirical benefits over both approaches.

Table 1. A 2 × 2 taxonomy of conformal prediction approaches (not an exhaustive list), categorized based on the dimension of the response variable Y (rows) and data assumptions (columns).

	Exchangeable	Non-exchangeable
Univariate Y	(Volkhonskiy et al., 2017) (Barber et al., 2021; Kim et al., 2020)	(Zaffran et al., 2022; Xu & Xie, 2023a) (Xu & Xie, 2023b; Barber et al., 2023)
Multivariate Y	(Messoudi et al., 2021; Diquigiovanni et al., 2022) (Johnstone & Ndiaye, 2022; Feldman et al., 2023)	Ours (Stankeviciute et al., 2021; Sun Yu, 2024)

Our approach

The main novelty of `MultiDimSPCI` is the design of non-conformity scores that explicitly take into account the entry-level dependency in Y_t, with subsequent construction of ellipsoidal prediction regions using the scores.

More precisely, let $\hat{\varepsilon}_t = Y_t - \hat{f}(X_t)$ be the continuous *prediction* residual in $\mathbb{R}^p$ and let $\widehat{\Sigma} \in \mathbb{R}^{p \times p}$ be the corresponding covariance estimator over the prediction residuals. Note that when p is large, $\widehat{\Sigma}$ may not be invertible. Hence, given $\rho > 0$, we consider a *low-rank* approximation $\widehat{\Sigma}_\rho$ of $\widehat{\Sigma}$ by truncating singular values of $\widehat{\Sigma}$ that are smaller than ρ.

Algorithm 1 Multi-dimensional `SPCI` (`MultiDimSPCI`)

Require: Training data $\{(X_t, Y_t)\}_{t=1}^T$, prediction algorithm $\mathcal{A}$, significance level α, quantile regression algorithm $\mathcal{Q}$, positive threshold $\rho > 0$.

Ensure: Prediction intervals $\widehat{C}_{t-1}(X_t, \alpha), t > T$

Obtain $\hat{f}$ and residuals $\{\hat{\varepsilon}_t\}_{t=1}^T \subset \mathbb{R}^p$ (computed on the holdout set) with $\mathcal{A}$ and $\{(X_t, Y_t)\}_{t=1}^T$
Compute non-conformity scores $\mathcal{E}_T$ from $\{\hat{\varepsilon}_t\}_{t=1}^T$ and $\widehat{\Sigma}_\rho$ using (2)
for $t > T$ **do**
  Use quantile regression to obtain $\widehat{Q}_t \leftarrow \mathcal{Q}(\mathcal{E}_T)$
  Obtain uncertainty set $\widehat{C}_{t-1}(X_t, \alpha)$ as in (3).
  Obtain new residual $\hat{\varepsilon}_t$
  Update residual set $\{\hat{\varepsilon}_t\}_{t=1}^T$ by adding $\hat{\varepsilon}_t$ and removing the oldest one and update $\mathcal{E}_T$
end for

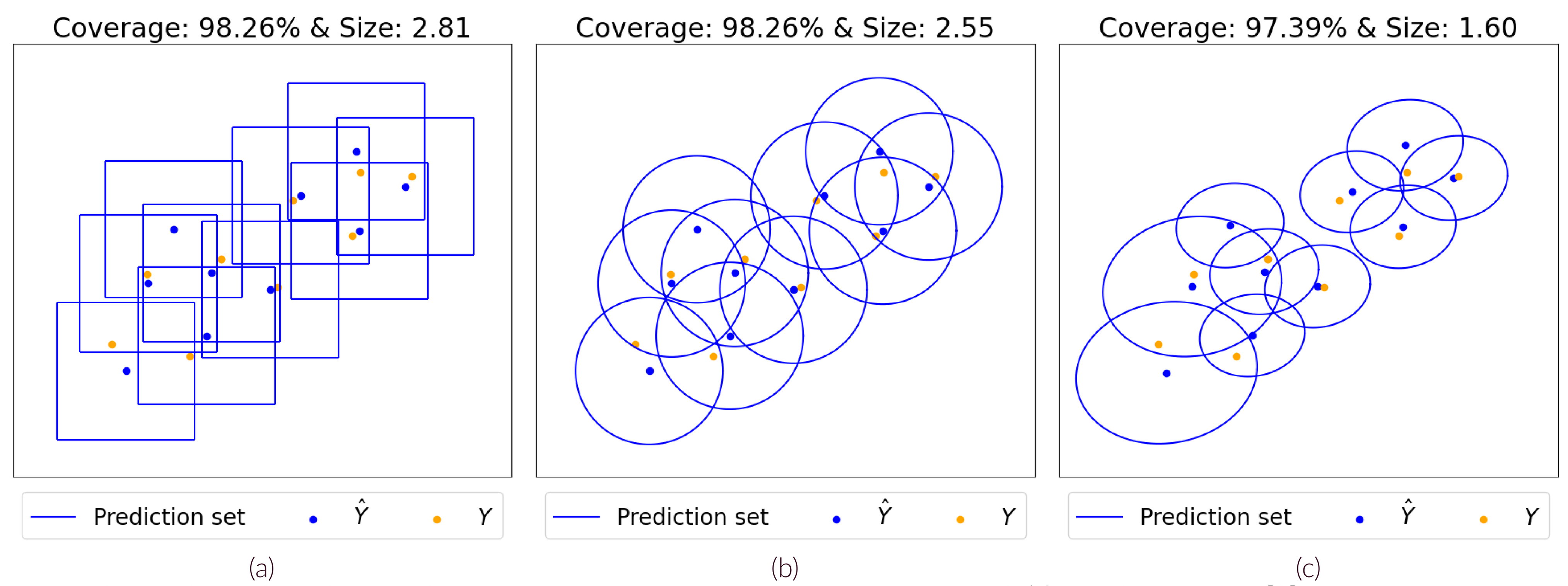

Figure 1. Comparison of multivariate CP method on real two-dimensional wind data. Left (a): Empirical copula [2] which constructs coordinate-wise prediction intervals. Middle (b): Spherical confidence set introduced in [3]. Right (c): our proposed ellipsoidal confidence set via `MultiDimSPCI`. While all methods yield coverage at least above the target 95% on test data, our method yields the smallest average size.

Given a candidate value $Y \in \mathbb{R}^p$, let $\hat{\varepsilon} = Y - \hat{f}(X)$ be the new residual. Using the pseudo-inverse $\widehat{\Sigma}_\rho^{-1}$ of the low-rank approximation, we then define the scalar non-conformity score $e(Y)$ as

$$\hat{e}(Y) = (\hat{\varepsilon} - \bar{\varepsilon})^T \widehat{\Sigma}_\rho^{-1} (\hat{\varepsilon} - \bar{\varepsilon}), \tag{2}$$

where $\bar{\varepsilon}$ is the mean of prediction residual. Using $\widehat{\Sigma}_\rho^{-1}$, which is always well-defined, an ellipsoid with radius r can thus be written as. $\mathcal{B}(r, \bar{\varepsilon}, \widehat{\Sigma}_\rho) = \{x \in \mathbb{R}^p : (x - \bar{\varepsilon})^T \widehat{\Sigma}_\rho^{-1} (x - \bar{\varepsilon}) \le r\}$. Thus, the prediction region $\widehat{C}_{t-1}(X_t) \subset \mathbb{R}^p$ for a given confidence level α takes the form

$$\begin{aligned} \widehat{C}_{t-1}(X_t) &= \{Y : \widehat{Q}_t(\hat{\beta}) \le \hat{e}(Y) \le \widehat{Q}_t(1 - \alpha + \hat{\beta})\} \quad (3) \\ &= \hat{f}(X_t) + \mathcal{B}(\sqrt{\widehat{Q}_t(1-\alpha+\hat{\beta})}, \bar{\varepsilon}, \widehat{\Sigma}_\rho) \setminus \mathcal{B}(\sqrt{\widehat{Q}_t(\hat{\beta})}, \bar{\varepsilon}, \widehat{\Sigma}_\rho) \\ \hat{\beta} &= \underset{\beta \in [0, \alpha]}{\arg\min}\ V(\widehat{\Sigma}_\rho, \widehat{Q}_t(1-\alpha+\beta)) - V(\widehat{\Sigma}_\rho, \widehat{Q}_t(\beta)) \;\; (V \text{ is volume of } \mathcal{B}) \quad (4) \end{aligned}$$

In (3), $\widehat{Q}_t$ denotes a fitted quantile regressor on the non-conformity score, following `SPCI`.

Theoretical guarantee

Let $Y_t \in \mathbb{R}^p$ follow $Y_t = f(X_t) + \varepsilon_t$, where f is an unknown function and ε_t is the noise. We can obtain different bounds on the coverage gap under different dependency assumptions on $\{\varepsilon_t\}$ and on the eigenvalue behavior of $\Sigma = \text{Cov}(\varepsilon_t)$ and $\widehat{\Sigma} = \text{Cov}(\hat{\varepsilon}_t)$. In particular, $(L_T, C_\delta, \delta_T)$ in the coverage gaps converge to zero under additional assumptions on estimation quality of f and on tail behavior of eigenvalues of Σ and $\widehat{\Sigma}$, reaching asymptotic valid coverage.

Theorem (When $\{\varepsilon_t\}$ are i.i.d)

With probability $1 - \delta$, for any training size T and $\alpha \in (0, 1)$, we have

$$|\mathbb{P}(Y_{T+1} \in \widehat{C}_T(X_{T+1}) \mid X_{T+1} = x_{T+1}) - (1 - \alpha)| \le 12\sqrt{\frac{\log(16T)}{T}} + 4(L_T + 1)(C_\delta + \delta_T). \tag{5}$$

Theorem (When $\{\varepsilon_t\}$ are stationary and strongly mixing)

Assume the true covariance matrix Σ is known. For any training size T and $\alpha \in (0, 1)$, we have

$$|\mathbb{P}(Y_{T+1} \in \widehat{C}_T(X_{T+1}) \mid X_{T+1} = x_{T+1}) - (1 - \alpha)| \le 12\frac{(\frac{M}{2})^{1/3}(\log T)^{2/3}}{T^{1/3}} + 4(L_T + 1)\left(\frac{\delta_T}{\sqrt{\lambda}} + \delta_T\right). \tag{6}$$

Experiments

We demonstrate the advantage of `MultiDimSPCI` against a wide range of CP methods and existing probabilistic forecasting approaches based on deep neural networks (NN). We consistently observed that `MultiDimSPCI` can maintain valid empirical coverage at $1 - \alpha$ and generate prediction regions that have significantly smaller volumes than baselines, especially in high dimensions.

Table 2. 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

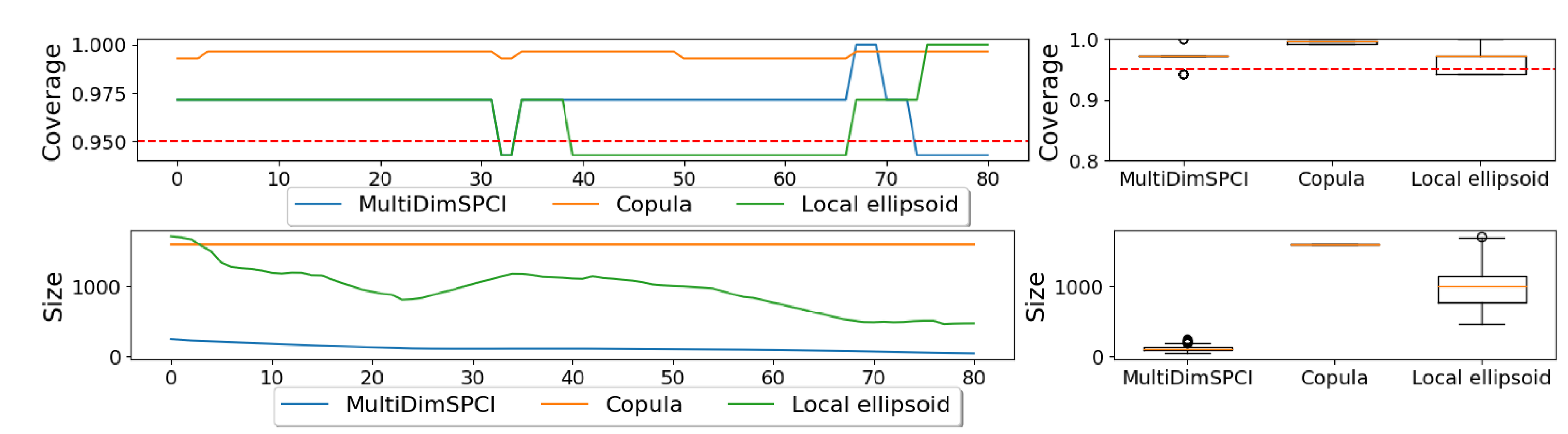

Figure 2. Real-data comparison of rolling coverage (target coverage is 95%) and size of prediction sets at $p = 8$ for the wind data. In each subplot of (a)-(c), the top row plots rolling coverage over prediction time indices (red dashed line is the target coverage) and as boxplots, and the bottom row shows results for rolling sizes.

References

[1] 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.

[2] Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, 120:108101, 2021.

[3] Sophia Huiwen Sun and Rose Yu. 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`.

[4] Chen Xu and Yao Xie. Sequential predictive conformal inference for time series. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 38707–38727. PMLR, 23–29 Jul 2023.