# Conformal prediction for multi-dimensional time series by ellipsoidal sets

Chen Xu, Hanyang Jiang, Yao Xie

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

# Motivation and objective

- Construct prediction regions for multivariate time-series.

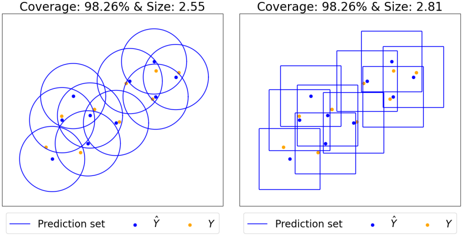
- Conformal prediction (CP) has been gaining popularity (distribution-free and model-free).

# Motivation and objective

- Construct prediction regions for multivariate time-series.

- Conformal prediction (CP) has been gaining popularity (distribution-free and model-free).

- Existing approaches can be conservative.

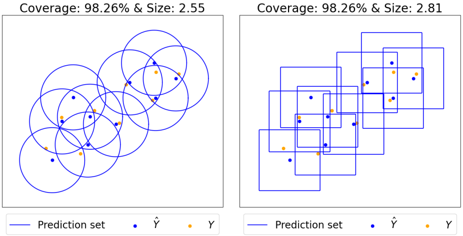


# Motivation and objective

- Construct prediction regions for multivariate time-series.

- Conformal prediction (CP) has been gaining popularity (distribution-free and model-free).

- Existing approaches can be conservative.




**Goal:** Develop sequential CP methods for multivariate time-series that jointly capture the dependency among time-series.

# Related works

- Univariate sequential CP: leverage feedback during prediction (Xu & Xie, 2021; Gibbs & Candes, 2021; Xu & Xie, 2023; Angelopoulos et al., 2024).
- Multivariate CP (Messoudi et al., 2021, 2022; Johnstone & Ndiaye, 2022; Sun & Yu, 2024).
- Probabilistic forecasting via quantile regression: (Salinas et al., 2020; Lim et al., 2021)

# Related works

- Univariate sequential CP: leverage feedback during prediction (Xu & Xie, 2021; Gibbs & Candes, 2021; Xu & Xie, 2023; Angelopoulos et al., 2024).

- Multivariate CP (Messoudi et al., 2021, 2022; Johnstone & Ndiaye, 2022; Sun & Yu, 2024).

- Probabilistic forecasting via quantile regression: (Salinas et al., 2020; Lim et al., 2021)

**Challenge:** effectively account for dependency across space and time.

# Proposed `MultiDimSPCI`

- In essence, `MultiDimSPCI` handles dependency in

(1) Space through ellipsoidal prediction sets.
(2) Time through the sequential `SPCI` (Xu & Xie, 2023).

# Proposed `MultiDimSPCI`

• In essence, `MultiDimSPCI` handles dependency in
(1) Space through ellipsoidal prediction sets.
(2) Time through the sequential `SPCI` (Xu & Xie, 2023).

• Data $(X_t, Y_t), Y_t \in \mathbb{R}^p$ arrive sequentially and we train a point predictor $\hat{f}$ and obtain *prediction* residuals $\hat{\epsilon}$ on first $T$ samples (Papadopoulos et al., 2007; Xu & Xie, 2021).

# Proposed `MultiDimSPCI`

• In essence, `MultiDimSPCI` handles dependency in
(1) Space through ellipsoidal prediction sets.
(2) Time through the sequential `SPCI` (Xu & Xie, 2023).

• Data $(X_t, Y_t), Y_t \in \mathbb{R}^p$ arrive sequentially and we train a point predictor $\hat{f}$ and obtain *prediction* residuals $\hat{\epsilon}$ on first $T$ samples (Papadopoulos et al., 2007; Xu & Xie, 2021).

• Let $\widehat{\Sigma} \in \mathbb{R}^{p \times p}$ be the empirical covariance of $\hat{\epsilon}$ and $\widehat{\Sigma}_\rho$ be the low-rank approximation of $\widehat{\Sigma}$.

# Proposed `MultiDimSPCI`

- In essence, `MultiDimSPCI` handles dependency in

(1) Space through ellipsoidal prediction sets.
(2) Time through the sequential `SPCI` (Xu & Xie, 2023).

- Data $(X_t, Y_t), Y_t \in \mathbb{R}^p$ arrive sequentially and we train a point predictor $\hat{f}$ and obtain *prediction* residuals $\hat{\boldsymbol{\epsilon}}$ on first $T$ samples (Papadopoulos et al., 2007; Xu & Xie, 2021).

- Let $\widehat{\Sigma} \in \mathbb{R}^{p \times p}$ be the empirical covariance of $\hat{\boldsymbol{\epsilon}}$ and $\widehat{\Sigma}_\rho$ be the low-rank approximation of $\widehat{\Sigma}$.

- Define non-conformity score $e(Y)$ for a candidate value $Y$ as

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

where $\hat{\varepsilon} = Y - \hat{f}(X)$ and $\bar{\varepsilon}$ is the mean of $\hat{\boldsymbol{\epsilon}}$.

# Proposed `MultiDimSPCI` (cont.)

- The ellipsoidal prediction region $\widehat{C}_{t-1}(X_t)$ at level $\alpha$ is

$$
\begin{aligned}
&\{Y : \widehat{Q}_t(\hat{\beta}) \le \hat{e}(Y) \le \widehat{Q}_t(1-\alpha+\hat{\beta})\} && (1)\\
=&\hat{f}(X_t) + \mathcal{B}(\sqrt{\widehat{Q}_t(1-\alpha+\hat{\beta})}, \bar{\varepsilon}, \widehat{\Sigma}_\rho) \setminus \mathcal{B}(\sqrt{\widehat{Q}_t(\hat{\beta})}, \bar{\varepsilon}, \widehat{\Sigma}_\rho)\\
\hat{\beta} =&\arg\min_{\beta\in[0,\alpha]} V(\widehat{\Sigma}_\rho, \widehat{Q}_t(1-\alpha+\beta)) - V(\widehat{\Sigma}_\rho, \widehat{Q}_t(\beta)) && (2)
\end{aligned}
$$

where $\widehat{Q}_t$ is the quantile regressor and $V$ denotes the volume of an ellipsoid $\mathcal{B}(r, \bar{\varepsilon}, \widehat{\Sigma}_\rho) = \{x \in \mathbb{R}^p : (x-\bar{\varepsilon})^T \widehat{\Sigma}_\rho^{-1}(x-\bar{\varepsilon}) \le r\}$.

# Proposed `MultiDimSPCI` (cont.)

- The ellipsoidal prediction region $\widehat{C}_{t-1}(X_t)$ at level $\alpha$ is

$$\begin{aligned}
&\{Y : \widehat{Q}_t(\hat{\beta}) \leq \hat{e}(Y) \leq \widehat{Q}_t(1-\alpha+\hat{\beta})\} \quad (1)\\
=&\hat{f}(X_t) + \mathcal{B}(\sqrt{\widehat{Q}_t(1-\alpha+\hat{\beta})}, \bar{\varepsilon}, \widehat{\Sigma}_\rho) \setminus \mathcal{B}(\sqrt{\widehat{Q}_t(\hat{\beta})}, \bar{\varepsilon}, \widehat{\Sigma}_\rho)\\
\hat{\beta} =& \underset{\beta\in[0,\alpha]}{\arg\min}\ V(\widehat{\Sigma}_\rho, \widehat{Q}_t(1-\alpha+\beta)) - V(\widehat{\Sigma}_\rho, \widehat{Q}_t(\beta)) \quad (2)
\end{aligned}$$

where $\widehat{Q}_t$ is the quantile regressor and $V$ denotes the volume of an ellipsoid $\mathcal{B}(r, \bar{\varepsilon}, \widehat{\Sigma}_\rho) = \{x \in \mathbb{R}^p : (x-\bar{\varepsilon})^T \widehat{\Sigma}_\rho^{-1}(x-\bar{\varepsilon}) \leq r\}$.

- Regions in (1) are sequentially constructed on updated $\hat{\boldsymbol{\epsilon}}$, using adaptively re-fitted $\widehat{Q}_t$.

# Remarks on `MultiDimSPCI`

- **Benefits**:
  - `MultiDimSPCI` can be used with any base model $\hat{f}$ and quantile regressor $\widehat{Q}_t$.
  - The ellipsoids have simple forms and are efficiently constructed.

# Remarks on `MultiDimSPCI`

- **Benefits**:
  - `MultiDimSPCI` can be used with any base model $\hat{f}$ and quantile regressor $\widehat{Q}_t$.
  - The ellipsoids have simple forms and are efficiently constructed.
- **Differences**:
  - Versus copula-based CP methods: smaller prediction regions with less design choices.
  - Versus Prob. forecasting methods: improved performance with theoretical guarantees.

# Theoretical results

- Extend analyses in EnbPI (Xu & Xie, 2021) to $Y_t \in \mathbb{R}^p$.

# Theoretical results

- Extend analyses in EnbPI (Xu & Xie, 2021) to $Y_t \in \mathbb{R}^p$.

- We impose additional assumptions on the error process, so that $\widehat{\Sigma}$ converges to $\Sigma$ in the operator norm.

# Theoretical results

- Extend analyses in EnbPI (Xu & Xie, 2021) to $Y_t \in \mathbb{R}^p$.

- We impose additional assumptions on the error process, so that $\widehat{\Sigma}$ converges to $\Sigma$ in the operator norm.

- We then obtain finite-sample bound on the absolute coverage gap $|\mathbb{P}(Y_t \in \widehat{C}_{t-1}(X_t)|X_t) - (1-\alpha)|$.

# Theoretical results

- Extend analyses in EnbPI (Xu & Xie, 2021) to $Y_t \in \mathbb{R}^p$.

- We impose additional assumptions on the error process, so that $\widehat{\Sigma}$ converges to $\Sigma$ in the operator norm.

- We then obtain finite-sample bound on the absolute coverage gap $|\mathbb{P}(Y_t \in \widehat{C}_{t-1}(X_t)|X_t) - (1-\alpha)|$.

- Guarantee bounds the worst-case deviation, while empirical coverage is almost always $\geq 1-\alpha$.

# Experiments (simulation)

- Compare `MultiDimSPCI` against `SPCI` (Xu & Xie, 2023), which is applied entry-wise with $\tilde{\alpha} = 1 - (1 - \alpha)^{1/p}$ per dimension.

*Table 3.* Simulation results by both methods. Target coverage is 90%. Standard deviation is computed over ten independent trials in which training and test data are regenerated.

(a) Independent AR($w$)

| $p$ | 8 | | 10 | | 16 | | 20 | |
|---|---|---|---|---|---|---|---|---|
| Method | `MultiDim SPCI` | `SPCI` (entry-wise) | `MultiDim SPCI` | `SPCI` (entry-wise) | `MultiDim SPCI` | `SPCI` (entry-wise) | `MultiDim SPCI` | `SPCI` (entry-wise) |
| Coverage | 90.0% (0.31) | 89.9% (0.30) | 89.8% (0.25) | 89.8% (0.27) | 89.9% (0.24) | 89.9% (0.23) | 90.0% (0.26) | 89.8% (0.30) |
| Size | 1.30e+5 (1.43e+3) | 3.68e+5 (6.44e+3) | 2.65e+6 (4.79e+4) | 1.22e+7 (1.61e+5) | 2.23e+10 (5.61e+8) | 5.84e+11 (1.39e+10) | 9.15e+12 (2.97e+11) | 8.67e+14 (2.90e+13) |

(b) VAR($w$)

| $p$ | 8 | | 10 | | 16 | | 20 | |
|---|---|---|---|---|---|---|---|---|
| Method | `MultiDim SPCI` | `SPCI` (entry-wise) | `MultiDim SPCI` | `SPCI` (entry-wise) | `MultiDim SPCI` | `SPCI` (entry-wise) | `MultiDim SPCI` | `SPCI` (entry-wise) |
| Coverage | 90.0% (0.23) | 91.6% (0.18) | 89.9% (0.23) | 90.7% (0.31) | 89.9% (0.20) | 91.0% (0.19) | 90.0% (0.25) | 90.9% (0.19) |
| Size | 7.16e+4 (7.25e+2) | 9.27e+6 (1.46e+5) | 3.63e+7 (4.79e+5) | 3.24e+9 (6.09e+7) | 8.55e+12 (1.45e+11) | 1.91e+17 (5.38e+15) | 1.14e+16 (2.11e+14) | 7.41e+22 (1.68e+21) |

# Experiments (real data)

*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

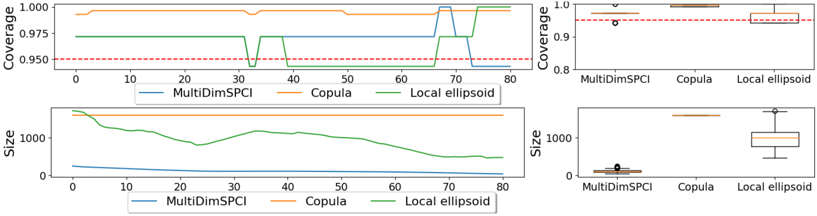
| Method | $p=2$ coverage | $p=2$ size | $p=4$ coverage | $p=4$ size | $p=8$ coverage | $p=8$ size |
|---|---|---|---|---|---|---|
| `MultiDimSPCI` | 0.96 | 1.68 | 0.96 | **2.89** | 0.97 | **4.97** |
| CopulaCPTS (Sun & Yu, 2024) | 0.99 | 4.36 | 0.99 | 37.56 | 0.99 | 3.28e+3 |
| Local ellipsoid (Messoudi 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

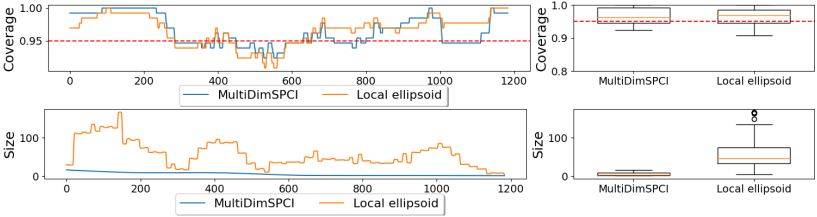
| Method | $p=2$ coverage | $p=2$ size | $p=4$ coverage | $p=4$ size | $p=8$ coverage | $p=8$ size |
|---|---|---|---|---|---|---|
| `MultiDimSPCI` | 0.96 | **1.31** | 0.96 | **1.93** | 0.96 | **2.98** |
| CopulaCPTS (Sun & Yu, 2024) | 0.95 | 1.70 | 0.94 | 3.15 | 0.95 | 14.10 |
| Local ellipsoid (Messoudi 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 |

# Experiments (real data)

- Similar rolling coverage with significantly narrower and more stable sizes of prediction regions.




(a) Wind data




(b) Solar data

# Summary

- The main novelty of `MultiDimSPCI` lies in jointly capturing spatial and temporal dependency in multivariate time-series.

- Against existing multivariate CP and probabilistic forecasting approaches, `MultiDimSPCI` returns much smaller prediction regions with no coverage loss.

- In the future, we will test the approach on more datasets with improved quantile estimation approaches.

# References I


Angelopoulos, A., Candes, E., and Tibshirani, R. J. Conformal pid control for time series prediction. *Advances in Neural Information Processing Systems*, 36, 2024.

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

Johnstone, C. and Ndiaye, E. Exact and approximate conformal inference in multiple dimensions. *arXiv preprint arXiv:2210.17405*, 2022.

Lim, B., Arık, S. Ö., Loeff, N., and Pfister, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

Messoudi, S., Destercke, S., and Rousseau, S. Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, 120:108101, 2021.

# References II

Messoudi, S., Destercke, S., and Rousseau, S. Ellipsoidal conformal inference for multi-target regression. In *Conformal and Probabilistic Prediction with Applications*, pp. 294–306. PMLR, 2022.

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.

Salinas, D., Flunkert, V., Gasthaus, J., and Januschowski, T. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

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`.

# 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. 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 2023.