Conformal Prediction for Time-series Forecasting with Change Points

Sophia Sun, Rose Yu, NeurIPS 2025

Conformal prediction

- Conformal prediction is useful for obtaining stable coverage for time-series, even under distribution shift.
  - See: *A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification* by Anastasios N Angelopoulos and Stephen Bates, 2022

Baselines / Context

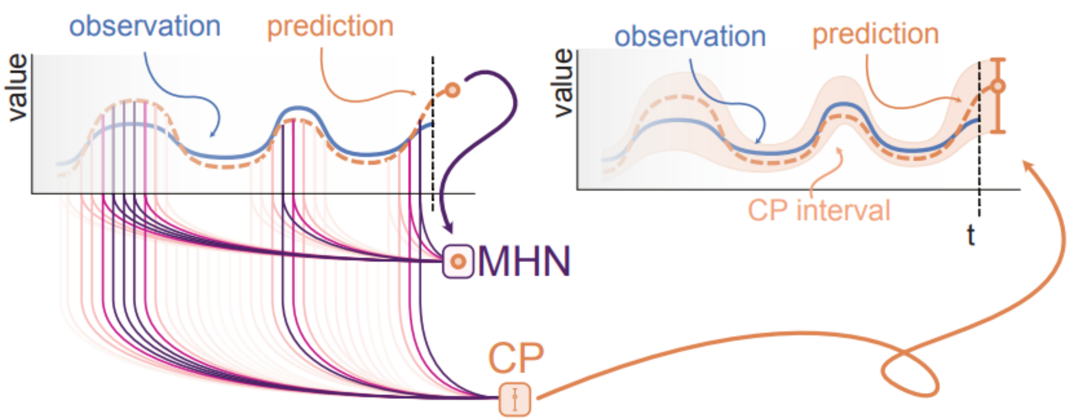


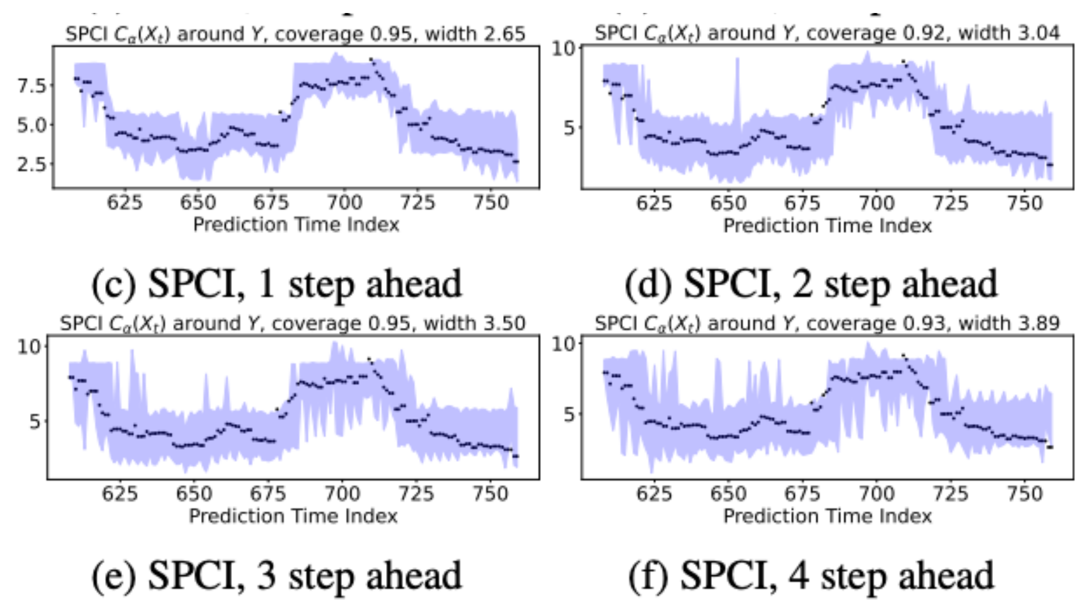
Figure 1: Schematic illustration of HopCPT. The Modern Hopfield Network (MHN) identifies regimes similar to the current one and up-weights them (colored lines). The weighted information enriches the conformal prediction (CP) procedure so that prediction intervals can be derived.

A. Auer, M. Gauch, D. Klotz, and S. Hochreiter. *Conformal prediction for time series with modern hopfield networks*. NeurIPS 2023.

Uses **Modern Hopfield Network** to learn temporal patterns

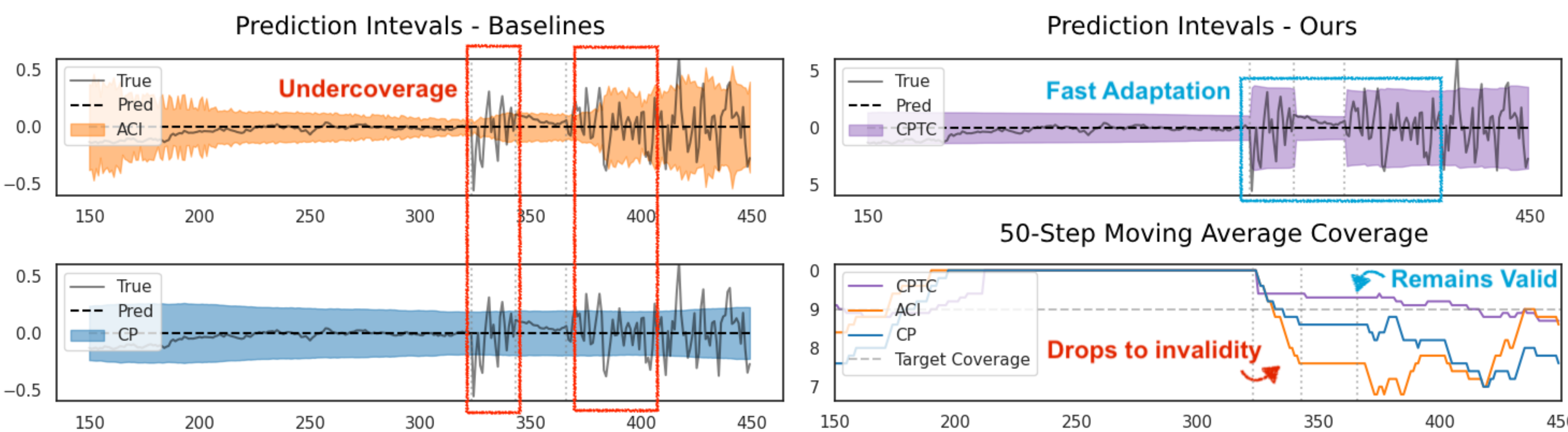
C. Xu and Y. Xie. Sequential predictive conformal inference for time series. ICML 2023

Uses **Quantile Random Forest** to learn temporal patterns

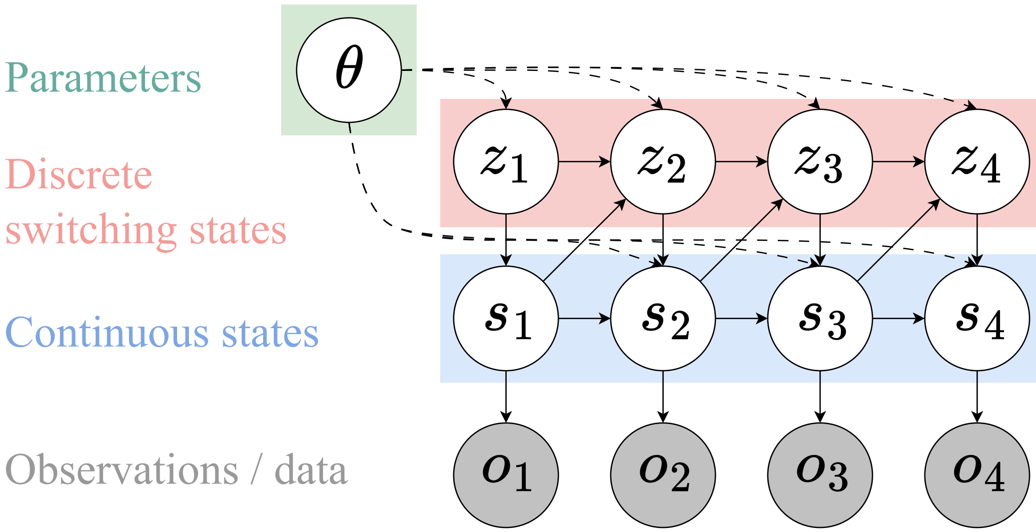


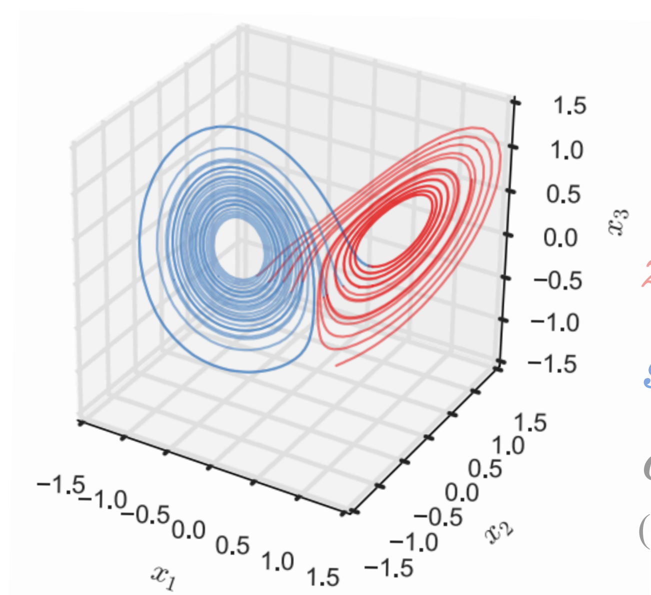
Conformal prediction

- Baseline *react*, or use *regression* at test time to learn correlations.
- What happens when we can *anticipate* distribution shifts?



Switching Dynamical Systems





$z_t \in \{L, R\}$

$s_t \in \mathbb{R}^3$

$o_t \in \mathbb{R}^d$

(observation dimension)

Algorithm

Idea: K regimes, K concurrent ACI processes

Algorithm 1: Conformal Prediction for Time series with Change points (CPTC)

Input: nonconformity score function A, probabilistic state model $\hat{p}(z_t|x_{0:t})$ with prior $\hat{p}(z_0)$, forecaster model $\hat{f} : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$, confidence level $1 - \alpha$, learning rate γ

Output: Prediction intervals $\Gamma(x_t)$ for $t \geq 1$

for $z \in \mathcal{Z}$ **do**
    Initialize $\mathcal{S}_z \leftarrow \{\}$, $\alpha_{z,0} \leftarrow \alpha$
end for
if *exists warm start data* $\{(x_t, y_t)\}_{t=-w}^{0}$ **then**
    for $t \in [-w, 0]$ **do**
        $\mathcal{S}_{z_t} \leftarrow \mathcal{S}_{z_t} \cup \{A(x_t, y_t)\}$, where $z_t \sim \hat{p}(z_t = z|x_{0:t})$; `// warm start scores`
    end for
end if
for $t \in [1, T]$ **do**
    for $z \in \mathcal{Z}$ **do**
        $\Gamma_{z,t}(x_t) \leftarrow \{y : A(x_t, y) \leq Q^{1-\alpha_{z,t}}(\mathcal{S}_z \cup \{\infty\})\}$; `// state-specific CP`
    end for
    $\Gamma_t(x_t) \leftarrow$ Aggregate the $\Gamma_{z,t}(x_t)$s by Eqn 10 or Eqn 11;
    Output: $\Gamma_t(x_t)$
    Sample state $\hat{z}_t \sim \hat{p}(z_t|x_{0:t})$; `// state-specific coverage target tracking`
    Update $\alpha_{\hat{z}_t,t+1} \leftarrow \alpha_{\hat{z}_t,t} + \gamma \cdot (\alpha - err_t)$, where $err_t = \mathbb{1}\{y_t \notin \Gamma_t(x_t)\}$;
    Update scores $\mathcal{S}_{\hat{z}_t} \leftarrow \mathcal{S}_{\hat{z}_t} \cup \{A(x_t, y_t)\}$;
end for

Create conformal prediction set for each possible regime

Algorithm

Idea: K regimes, K concurrent ACI processes

Algorithm 1: Conformal Prediction for Time series with Change points (CPTC)

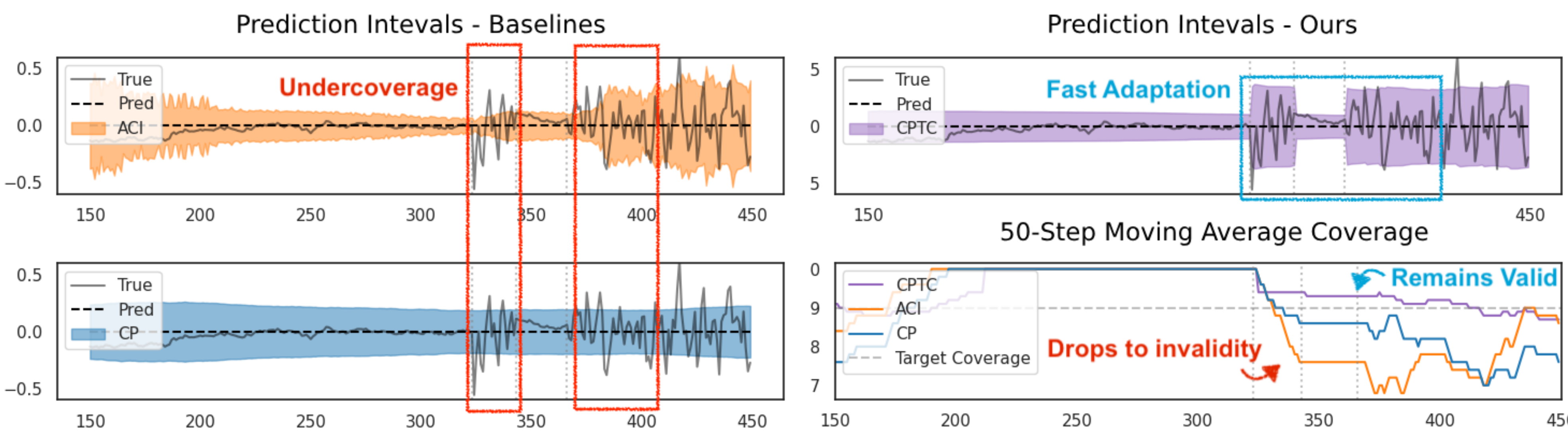
Input: nonconformity score function A, probabilistic state model $\hat{p}(z_t|x_{0:t})$ with prior $\hat{p}(z_0)$, forecaster model $\hat{f} : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$, confidence level $1 - \alpha$, learning rate γ

Output: Prediction intervals $\Gamma(x_t)$ for $t \geq 1$

for $z \in \mathcal{Z}$ **do**
  Initialize $\mathcal{S}_z \leftarrow \{\}$, $\alpha_{z,0} \leftarrow \alpha$
end for
if *exists warm start data* $\{(x_t, y_t)\}_{t=-w}^{0}$ **then**
  for $t \in [-w, 0]$ **do**
    $\mathcal{S}_{z_t} \leftarrow \mathcal{S}_{z_t} \cup \{A(x_t, y_t)\}$, where $z_t \sim \hat{p}(z_t = z|x_{0:t})$; `// warm start scores`
  end for
end if
for $t \in [1, T]$ **do**
  for $z \in \mathcal{Z}$ **do**
    $\Gamma_{z,t}(x_t) \leftarrow \{y : A(x_t, y) \leq Q^{1-\alpha_{z,t}}(\mathcal{S}_z \cup \{\infty\})\}$; `// state-specific CP`
  end for
  $\Gamma_t(x_t) \leftarrow$ Aggregate the $\Gamma_{z,t}(x_t)$s by Eqn 10 or Eqn 11;
  Output: $\Gamma_t(x_t)$
  Sample state $\hat{z}_t \sim \hat{p}(z_t|x_{0:t})$; `// state-specific coverage target tracking`
  Update $\alpha_{\hat{z}_t,t+1} \leftarrow \alpha_{\hat{z}_t,t} + \gamma \cdot (\alpha - err_t)$, where $err_t = \mathbb{1}\{y_t \notin \Gamma_t(x_t)\}$;
  Update scores $\mathcal{S}_{\hat{z}_t} \leftarrow \mathcal{S}_{\hat{z}_t} \cup \{A(x_t, y_t)\}$;
end for

Combine sets based on regime's probability

Results


Prediction Intevals - Baselines
Prediction Intevals - Ours
Undercoverage
Fast Adaptation
50-Step Moving Average Coverage
Remains Valid
Drops to invalidity
True
Pred
ACI
CP
CPTC
Target Coverage
0.5
0.0
−0.5
150
200
250
300
350
400
450

Results

Table 1: Performance on synthetic and real-world datasets with target confidence $1-\alpha=0.9$ (for horizon $T=200$ for synthetic datasets, 300 for electricity and traffic, and 60 for bee, mean $\pm$ standard deviation of the test samples). Methods that are *invalid* (coverage below 90%) are grayed out. Our method achieves a high level of calibration (coverage is close to 90%) consistently.

		RED-SDS	CP	ACI	SPCI	HopCPT	Ours
Bouncing Ball obs.	Cov	16.90 ± 12.96	87.67 ± 6.54	89.82 ± 2.95	87.13 ± 4.79	79.45 ± 12.51	90.15 ± 1.19
	Width	1.60 ± 0.07	10.54 ± 3.19	3.85 ± 2.60	2.82 ± 0.41	1.98 ± 0.90	3.71 ± 0.98
Bouncing Ball dyn.	Cov	13.20 ± 11.50	86.45 ± 12.28	89.38 ± 2.56	89.16 ± 3.47	81.12 ± 7.85	90.47 ± 2.30
	Width	1.37 ± 0.05	11.49 ± 4.39	2.27 ± 1.65	2.26 ± 0.28	2.95 ± 1.12	1.76 ± 0.71
3-Mode System	Cov	93.02 ± 3.04	63.98 ± 5.04	89.90 ± 2.25	84.90 ± 5.41	90.85 ± 5.90	94.96 ± 1.96
	Width	2.08 ± 0.16	1.46 ± 0.50	6.40 ± 1.92	3.70 ± 0.89	9.07 ± 1.23	2.45 ± 0.72
Traffic	Cov	23.15 ± 10.95	87.06 ± 3.71	90.01 ± 0.87	84.38 ± 2.47	88.91 ± 3.55	92.38 ± 1.24
	Width	0.05 ± 0.02	21.91 ± 10.42	27.81 ± 54.08	5.32 ± 1.95	7.50 ± 10.09	7.92 ± 2.98
Electricity	Cov	62.85 ± 13.67	85.69 ± 6.17	89.79 ± 0.83	84.10 ± 2.77	86.50 ± 2.66	91.22 ± 1.29
	Width	162.67 ± 811.73	366.05 ± 2280.45	45.74 ± 279.21	228.14 ± 1207.58	155.71 ± 130.43	139.75 ± 620.44
Dancing Bees	Cov	84.92 ± 6.84	79.86 ± 20.73	86.25 ± 8.10	79.20 ± 0.30	72.11 ± 1.84	92.64 ± 3.19
	Width	0.25 ± 0.02	1.65 ± 0.58	4.79 ± 4.02	1.77 ± 1.38	1.06 ± 0.51	0.79 ± 0.27

Strengths

- Finite Sample miscoverage Guarantee
- Very fast compared to regression-based methods. (No learning during test time.)
- Adapts to drifts (via ACI update) and shifts (regime changes), adapts faster to the later case.
- Handles the case of misspecified regimes.

Strengths

- Finite Sample miscoverage Guarantee
- Very fast compared to regression-based methods. (No learning during test time.)
- Adapts to drifts (via ACI update) and shifts (regime changes), adapts faster to the later case.
- Handles the case of misspecified regimes.

Limitations

- Requires a explicit discrete state-space model and state probability
- In very long test time-series, CPTC performs worse than regression-based methods.

Thank you!

See you at the poster / email me at shs066@ucsd.edu