ICML
International Conference
On Machine Learning

2026

RSA-CP: Efficient Conformal Prediction in Small-Sample Regimes via Random Score Alignment

Pankaj Bhagwat, Zhixian Yang, Yihao Wang, Bei Jiang, Linglong Kong




UNIVERSITY
OF ALBERTA

Conformal Prediction

Split Conformal prediction (SCP) sets a threshold from calibration scores.

1. Fit any predictor

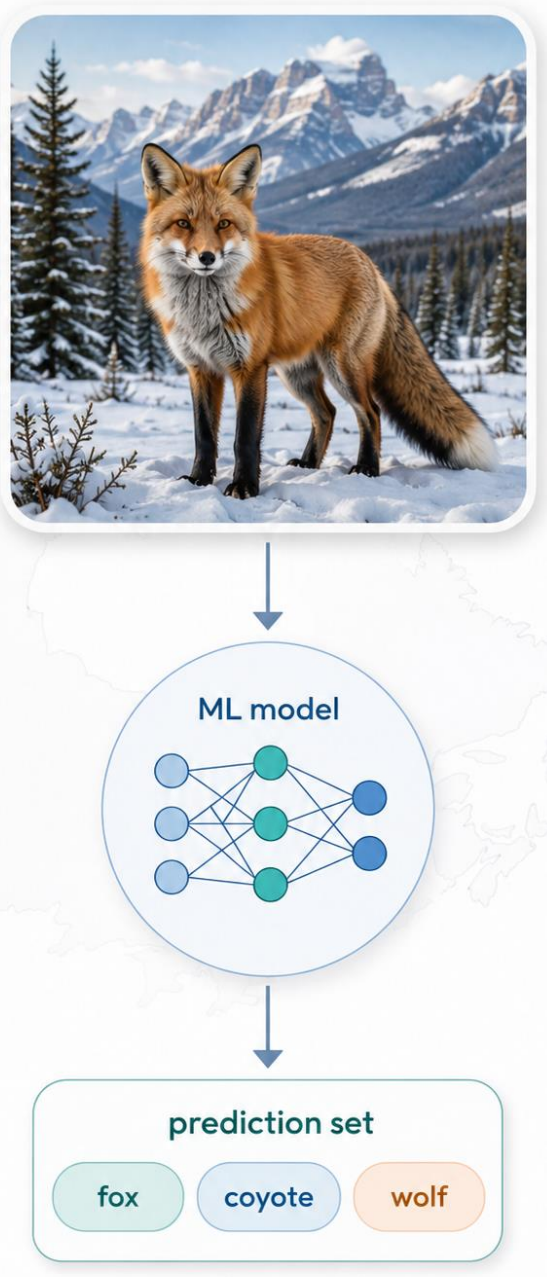
Train a model on training data.

2. Calibration

Compute nonconformity scores $S_i = \mathrm{score}(x_i, y_i)$.

3. Predict a set

Choose a quantile threshold q and include labels with score ≤ q.



Finite-sample guarantee

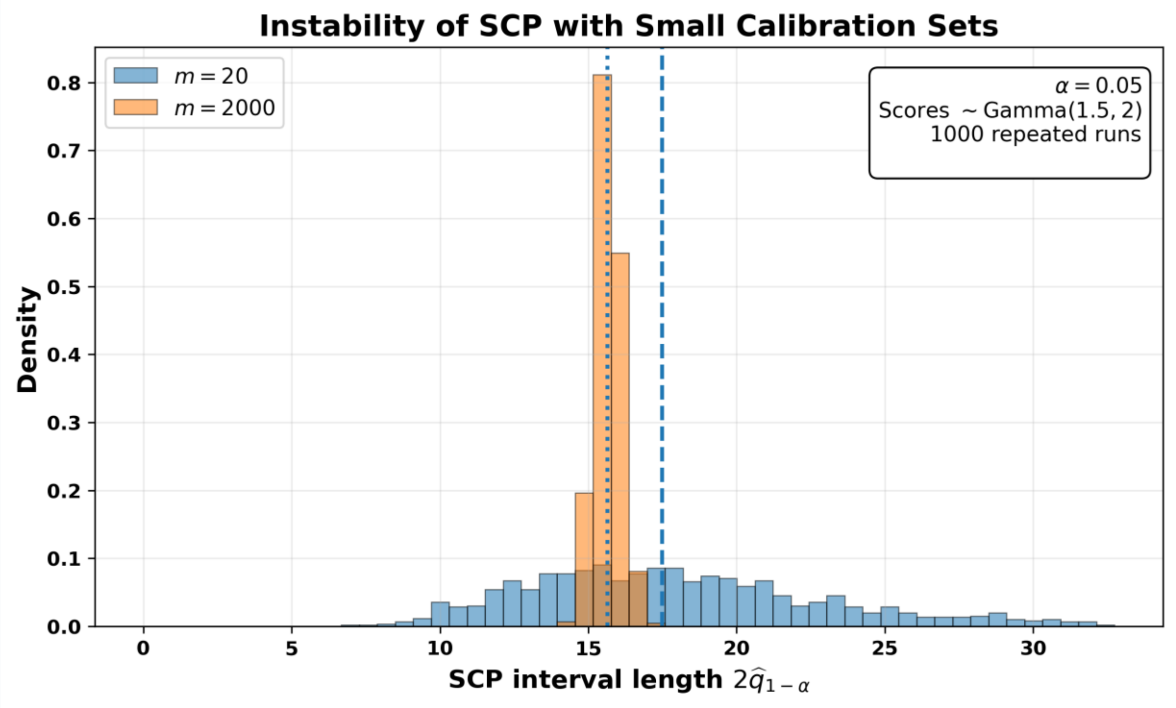
Under exchangeability, the prediction set covers the true output with probability at least 1 − α. (**High Coverage**)

Why small calibration sets are a problem?

Split Conformal prediction (SCP) sets a threshold from calibration scores.

Only a few calibration scores

Threshold choices are coarse.
The method may jump to a very large set.

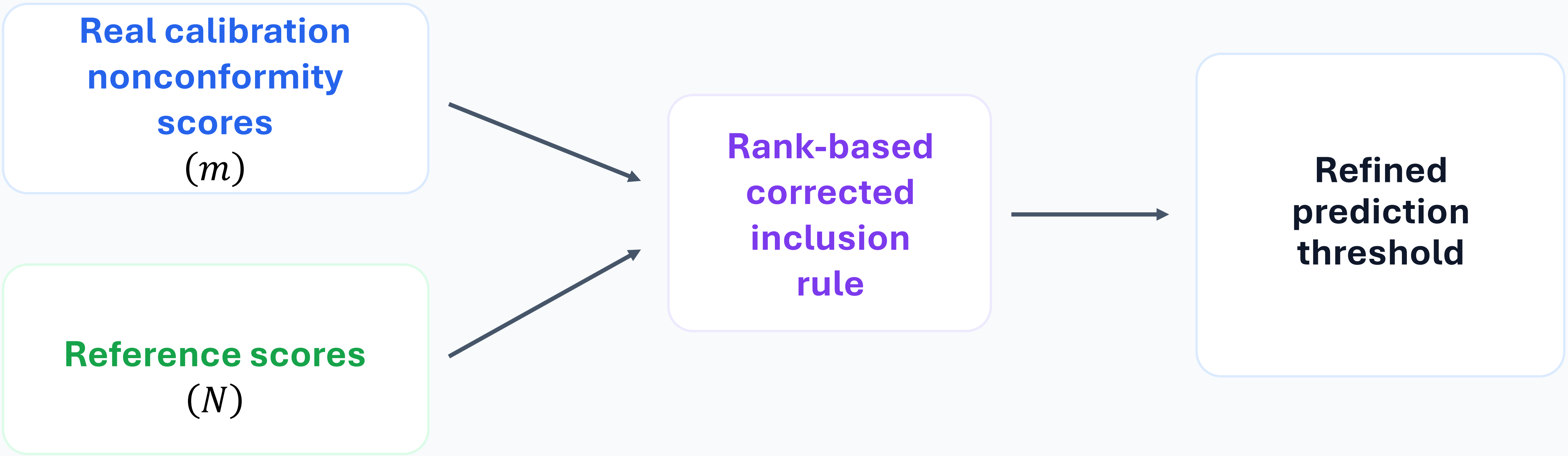


Consequence

Prediction sets become larger than necessary, or even uninformative.

Main idea: augment additional scores, but carefully!

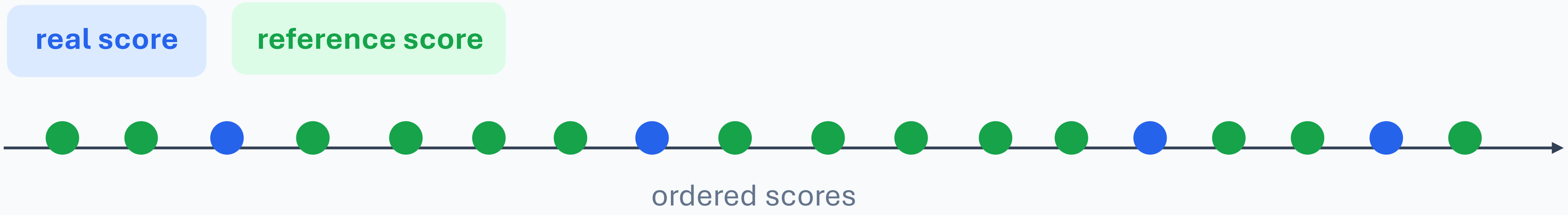
RSA-CP supplements real calibration scores with reference scores.



Reference sources can include simulators, fitted models, previous studies, or random generators.

Main idea: How does RSA-CP work?

The method works with order, not absolute score values.



1. Analyze the distribution of m ranks in the ensemble of $(m + N + 1)$

→

2. Compute a rank window

→

3. Choose a threshold with coverage control

The correction is conservative enough to protect coverage, but flexible enough to exploit informative reference scores.

What the theory says?

Coverage bounds

Coverage bounds remain computable for any reference-score choice.

$$LB(m,\beta) \leq \mathbb{P}\left(y \in \hat{C}_{RSA}(x)\right) \leq UB(m,\beta)$$

Robustness

Mismatch is not ignored; its effect appears in distribution-robust bounds.

$$\left|\mathbb{P}\left(y \in \hat{C}_{RSA}(x)\right) - (1 - \alpha)\right| \leq \frac{1}{m+N+1} + \varepsilon(TV)$$

Efficiency Gains

If reference scores resemble real scores, prediction sets can shrink.

Takeaway: RSA-CP provides coverage bounds for any reference-score augmentation, while informative reference scores can yield more efficient prediction sets.

Score Alignment via Optimal Transport (OT)

Monotone Transform: real score → aligned real score = T(real score)

Before alignment

Reference scores may be shifted or stretched relative to real calibration scores.

→

OT map

Use the monotone 1D map $T(s)$ to transport real scores.

→

After alignment

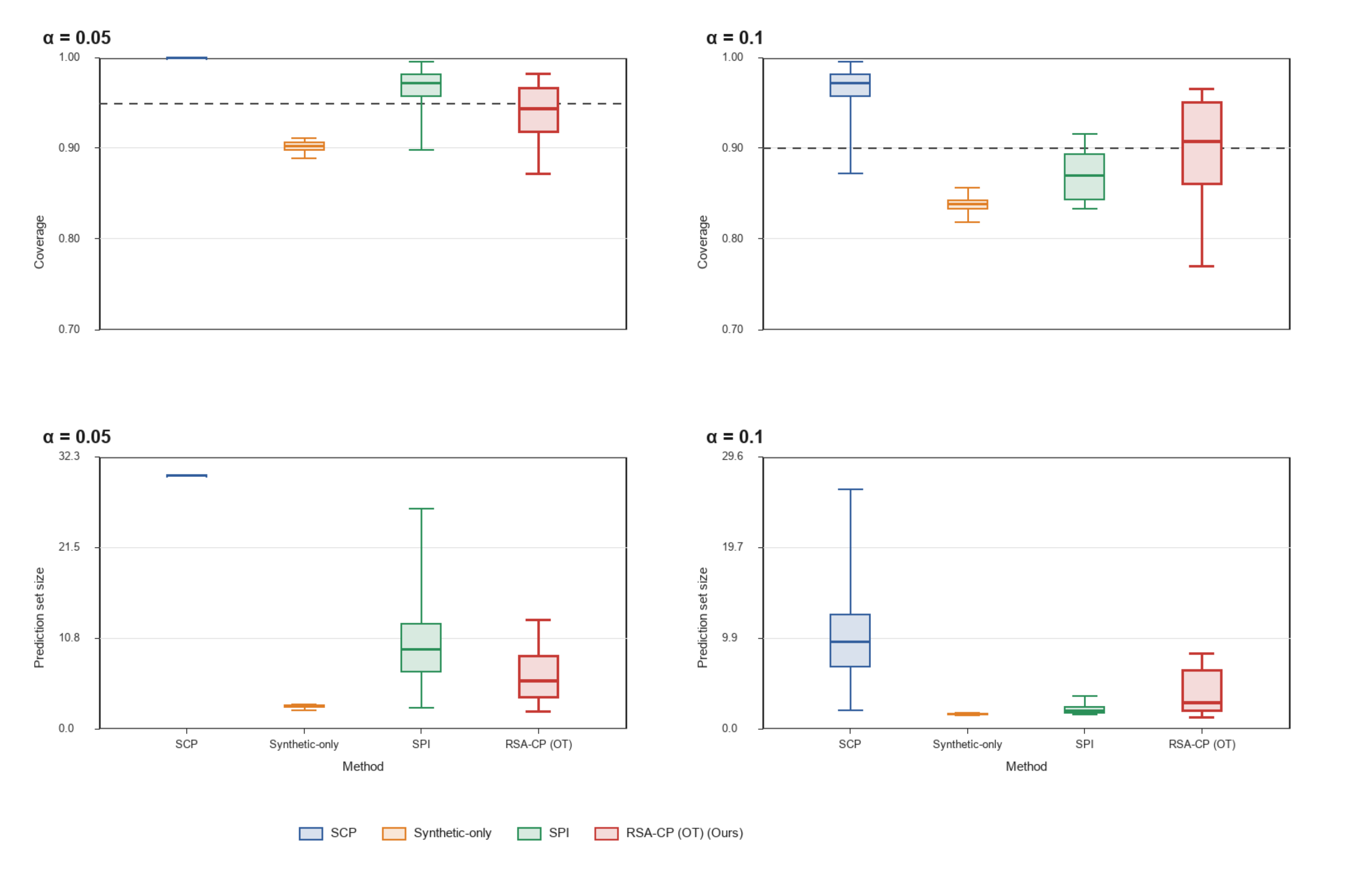
Aligned real scores are more comparable to reference scores.

Then use inside RSA-CP:

reference scores → OT alignment → aligned real scores → rank correction → prediction set

Takeaway: RSA-CP (OT) provides better score alignment leading to improved efficiency, while satisfying coverage bounds.

Empirical results: RSA-CP (OT) leads to informative prediction sets

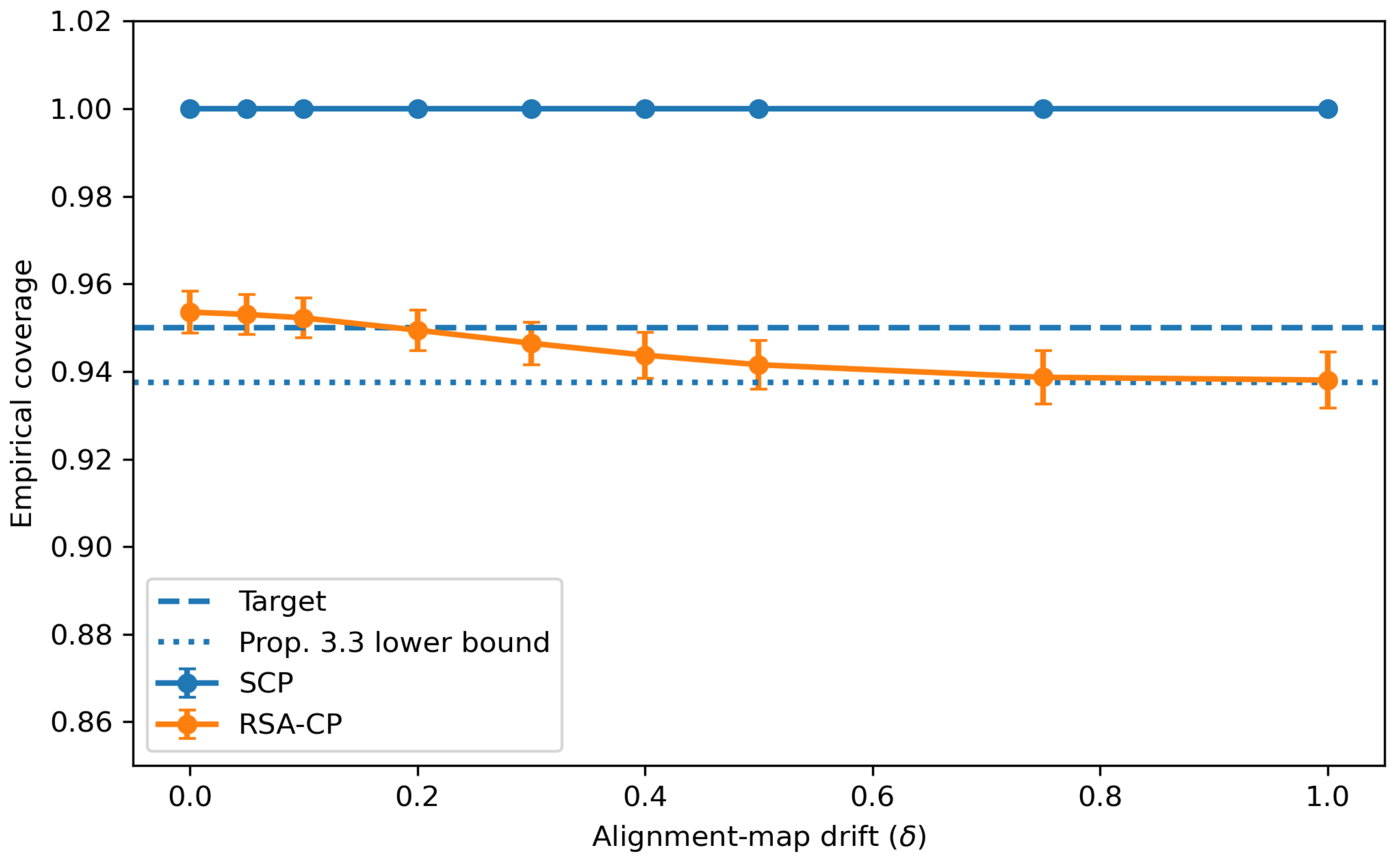


RSA-CP (OT) coverage remains close to the targeted level.

ImageNet Image Classification Task

Prediction set size: smaller is better when coverage is controlled.

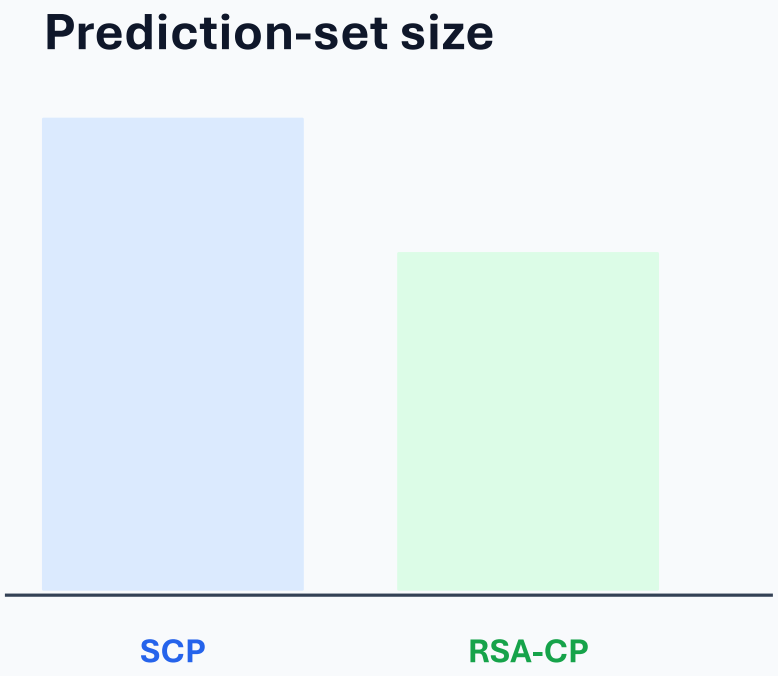
Empirical results: RSA-CP (OT) satisfies theoretical coverage bounds even under mismatch



RSA-CP (OT) satisfies coverage bounds even when OT alignment is not perfect!

Simulation study

In Summary,



smaller is better
when coverage is controlled

- When calibration data are scarce, RSA-CP can reduce conservativeness.
- RSA-CP (OT) satisfies distribution-free deterministic coverage bounds.
- Efficiency gains are significant when reference and real score distributions are close.

Read more here

Get in touch ...

pbhagwat@ualberta.ca

LinkedIn

SCAN ME

Thank you!