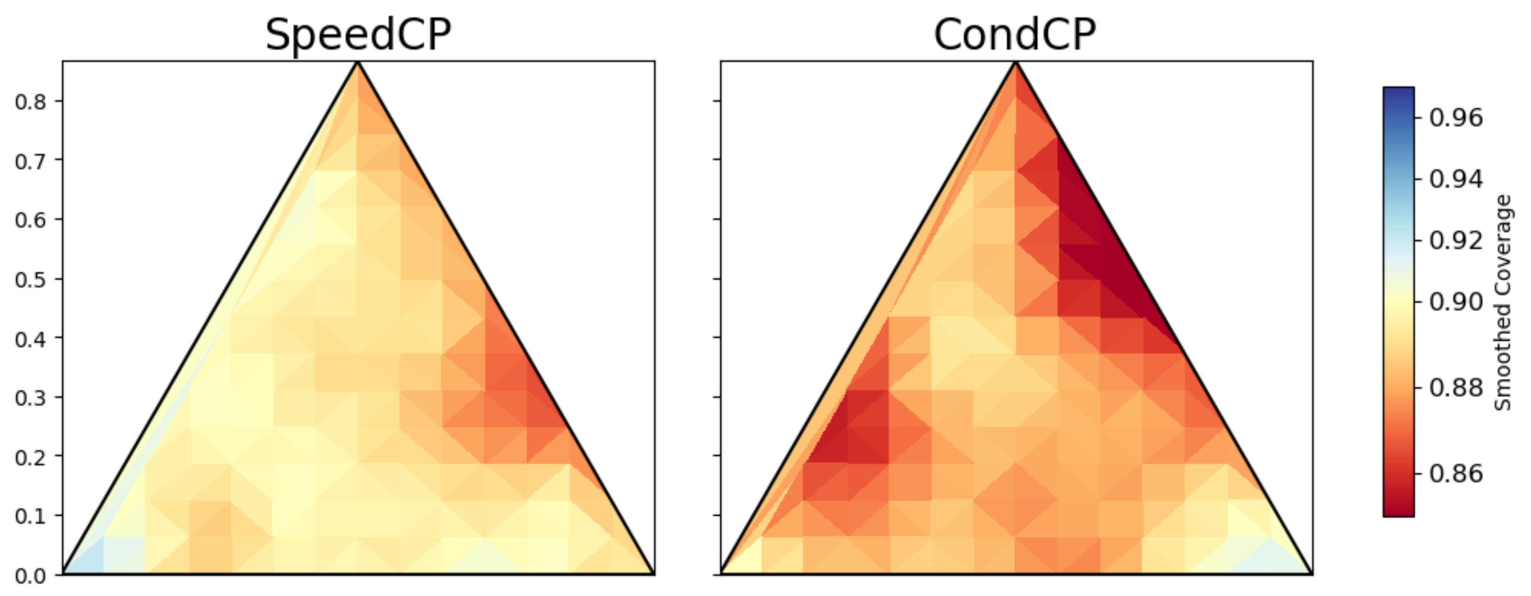
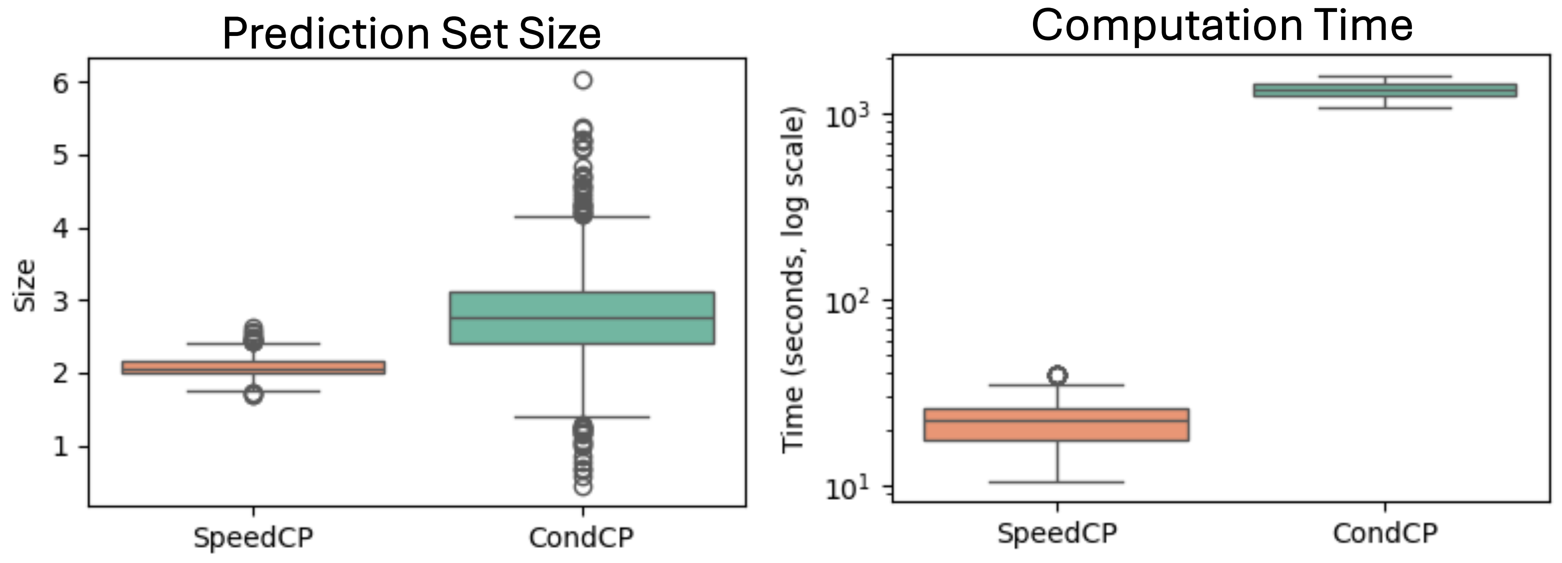
SpeedCP: Fast Kernel-based Conditional Conformal Prediction

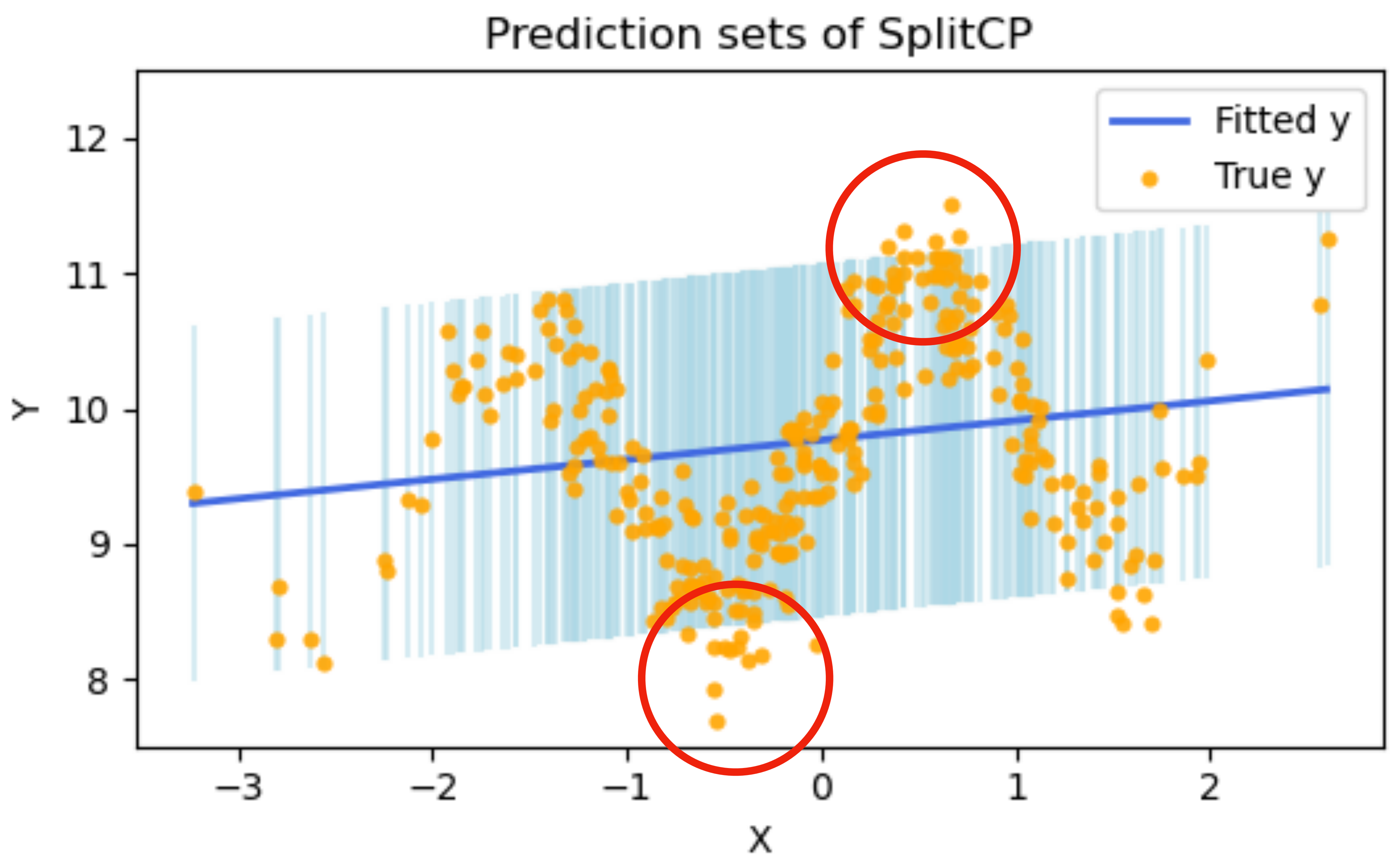
Yating Liu[1], Yeo Jin Jung[1], Sowon Jeong[2], Zixuan Wu[1], Claire Donnat[1]

[1]Department of Statistics, [2]Booth Business School
The University of Chicago

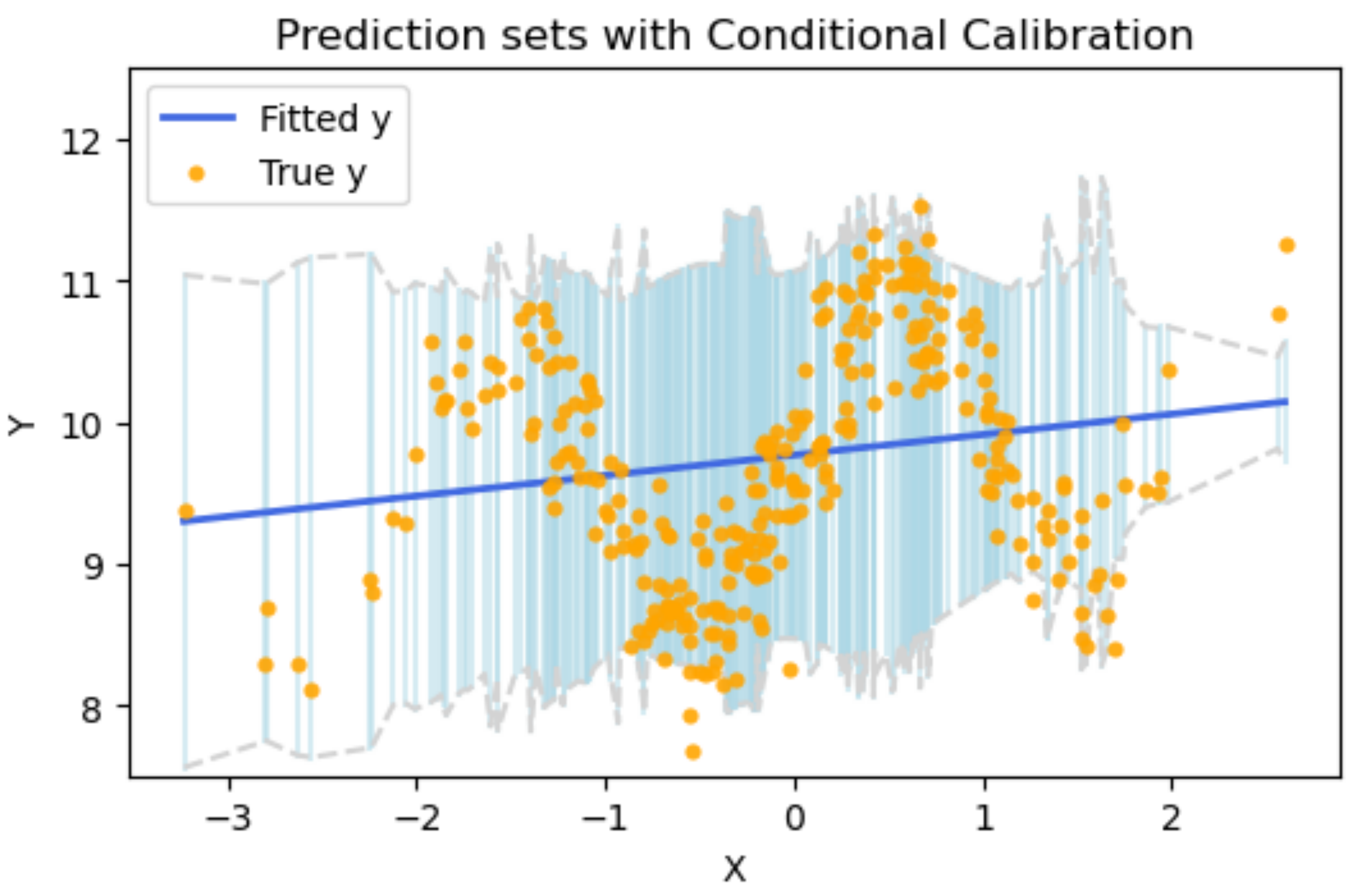




Marginal vs Conditional Coverage



90% coverage on average (marginal)



90% coverage everywhere (conditional)

- We want conditional coverage of a new test point X_{n+1}
- Coverage can fail badly in subregions of the feature space

Conditional coverage in RKHS

- Conditional coverage:

$$\mathbb{P}(Y_{n+1} \in \hat{C}_n(X_{n+1}) \,|\, X_{n+1} = x) = 1 - \alpha \quad \text{for all } x$$

$$\Longleftrightarrow$$

$$\mathbb{E}[g(X_{n+1}) \cdot \big(\mathbf{1}\{Y_{n+1} \in \hat{C}_n(X_{n+1})\} - (1 - \alpha)\big)] = 0 \quad \text{for any measurable function } g$$

Conditional coverage in RKHS

- Relaxed conditional coverage (Gibbs et al., 2025):

$$\mathbb{E}[g(X_{n+1}) \cdot \left(\mathbf{1}\{Y_{n+1} \in \hat{C}_n(X_{n+1})\} - (1-\alpha)\right)] = 0 \quad \text{for } g \in \text{some function class } \mathscr{G}$$

$\mathscr{G}$ = **Reproducing Kernel Hilbert space (RKHS)** → *"90% coverage on complex functions of X"*

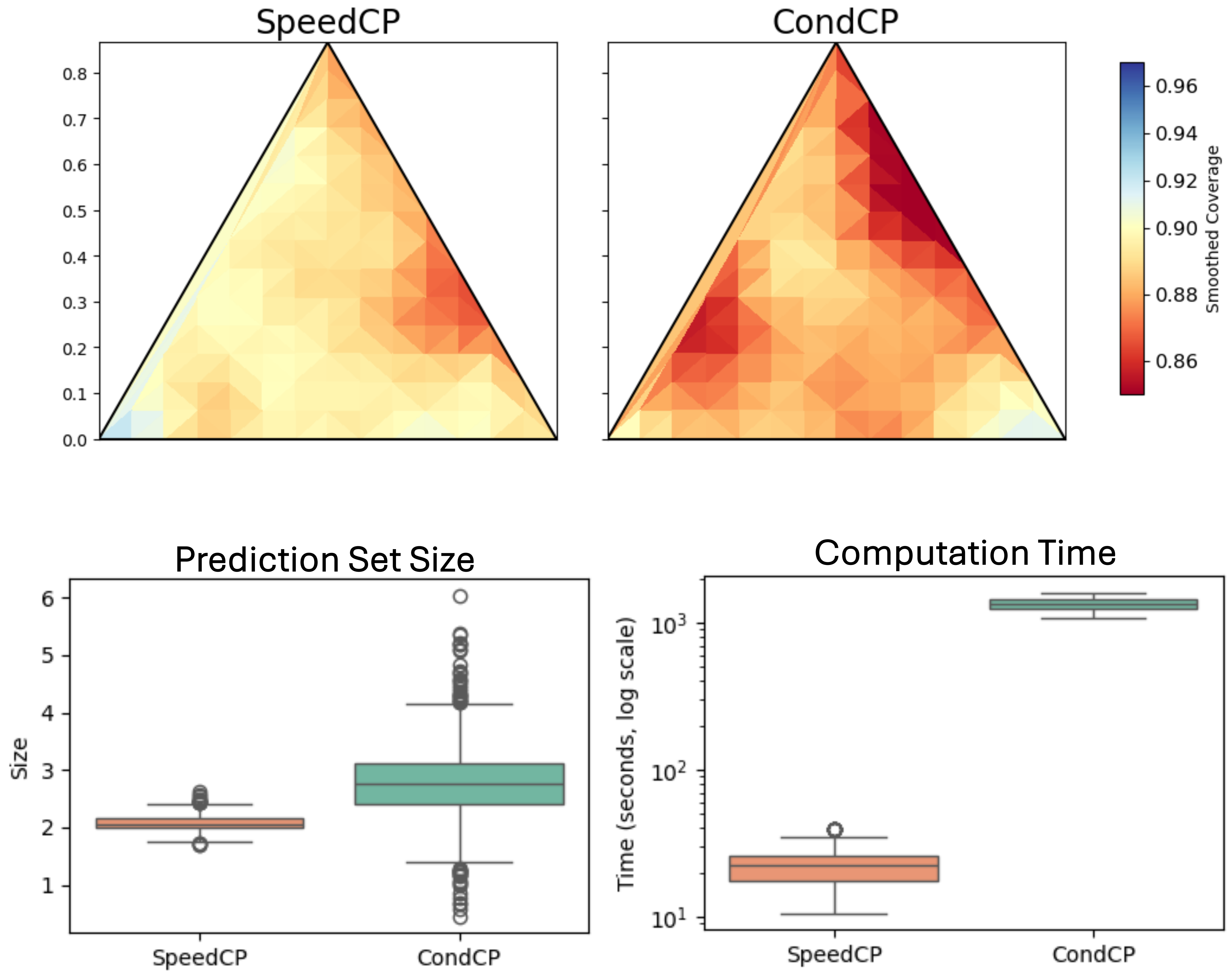
Closer to perfect conditional coverage

1) Conformal prediction with conditional guarantees (Gibbs, Cherian, & Candès, JRSS-B, 2025)

SpeedCP: A better solver for RKHS Conditional CP

Practical Challenges in RKHS setting

- Slow and unstable algorithm
- Kernel methods on high-dimensional data can be problematic



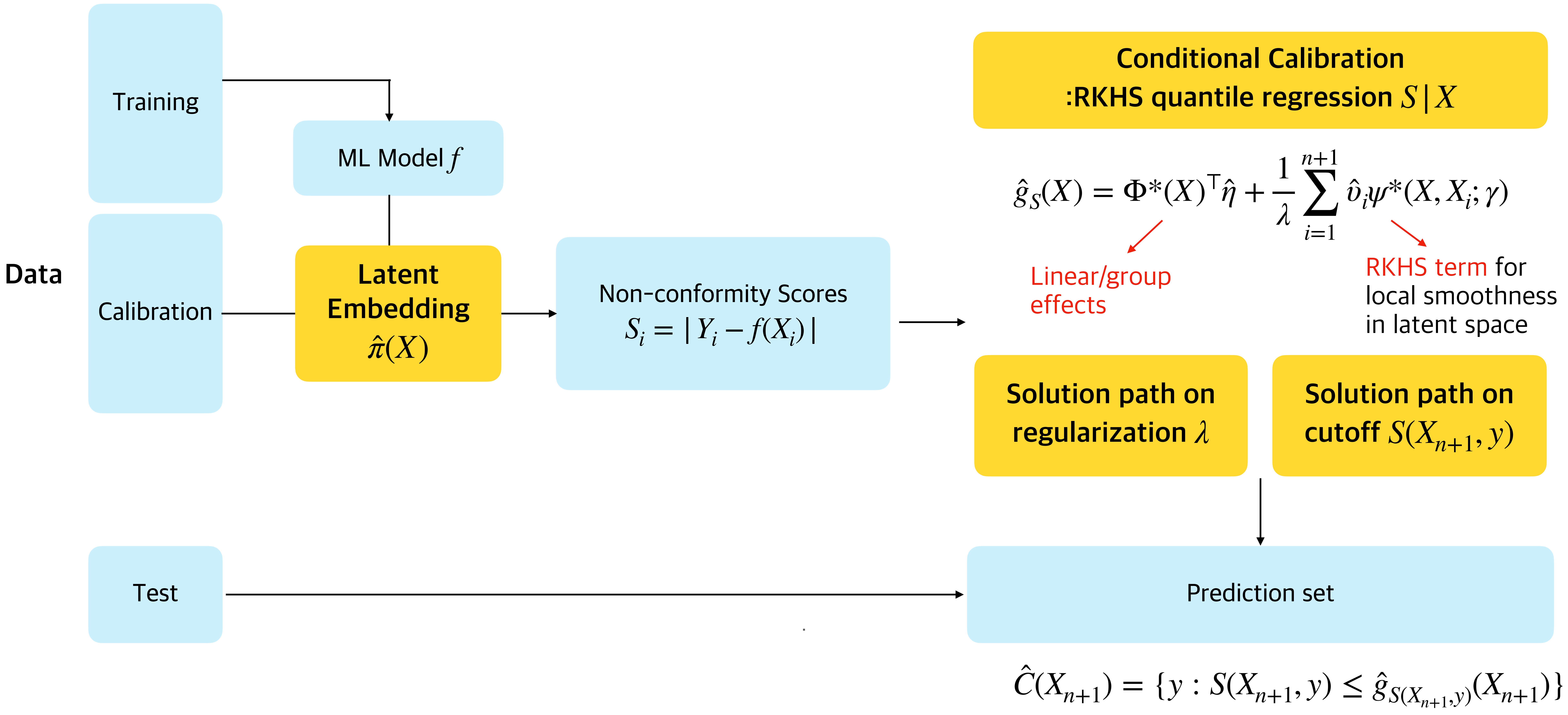
SpeedCP: A better solver for RKHS Conditional CP

Key Ideas

- Calibrate in latent space $\hat{\pi}(X)$: neighborhoods are denser, less noisy, and more meaningful for conditional coverage.
- Fit RKHS quantile regression on nonconformity scores using two piecewise affine paths:

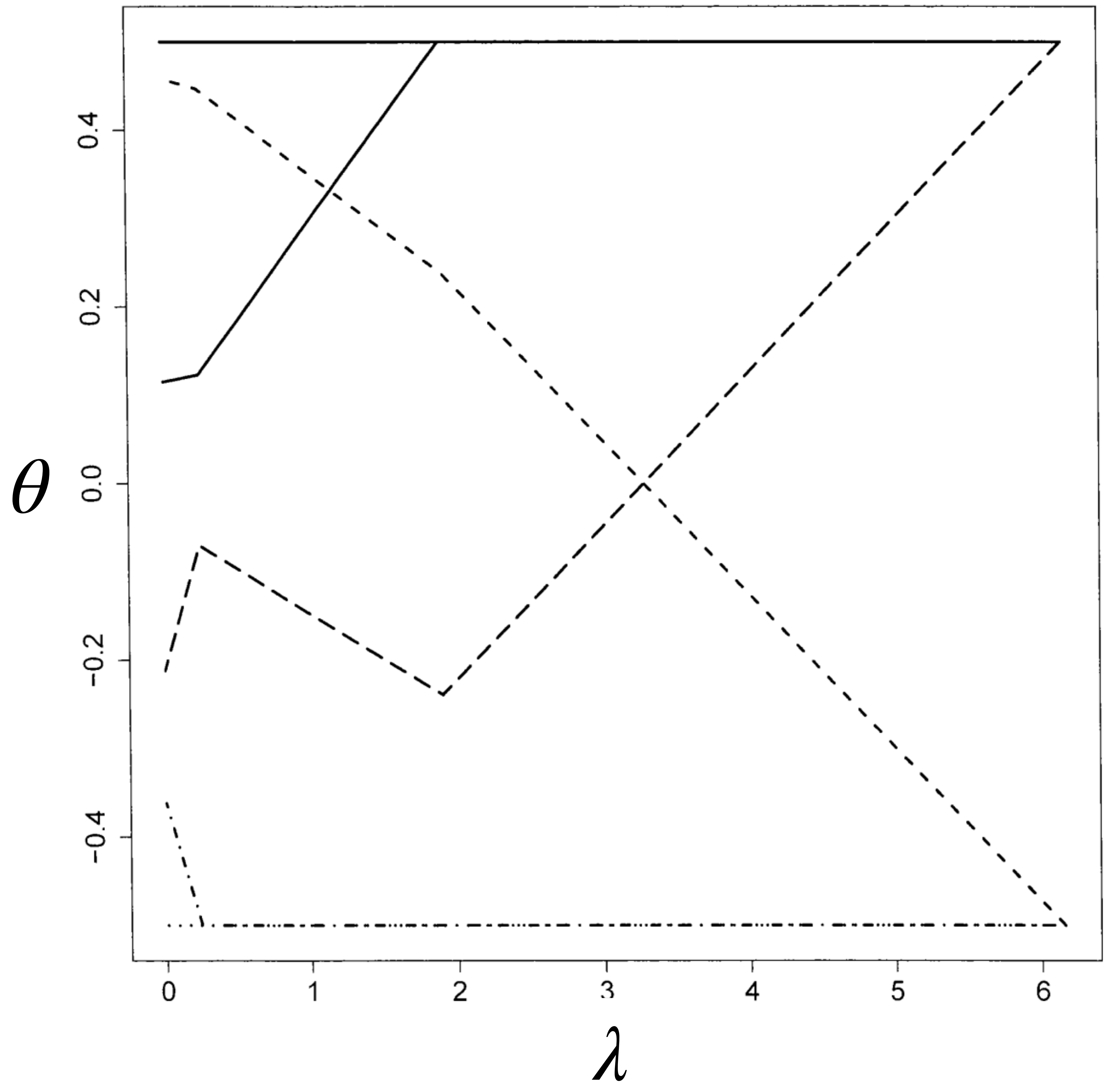
    1. λ-path for regularization hyperparameter tuning
    2. S-path for prediction set construction

SpeedCP: A better solver for RKHS Conditional CP



Affine piecewise solution paths

- λ**-path**: traces quantile regression parameters as λ changes, allowing us to find best level of smoothness
- S**-path**: traces parameters as the score cutoff S changes, allowing us to find the largest valid cutoff
- **Events:** the path changes only when a point enters or leaves the elbow set
- **Key benefit**: only need to solve for λ and S at events



Adapted from active-set algorithm (Li, Y., Liu, Y., & Zhu, J. (2007))

2) Quantile regression in reproducing kernel Hilbert spaces (Li, Y., Liu, Y., & Zhu, J., JASA, 2007)

Why is SpeedCP faster?

	CondCP	Our Method
Choosing smoothness hyperparameters	• Cross validation to choose λ • No optimization on γ	• Derive piecewise solution path of λ for each γ
Getting prediction sets	• Binary search on $[S_{\min}, S_{\max}]$	• Derive piecewise solution path of S

Why is it slow? On each iteration, refit RKHS quantile regression for $n+1$ points

Conditional coverage guarantee

Takeaway: SpeedCP does not claim unrestricted pointwise coverage; it gives finite-sample guarantees for the chosen latent / kernel / group structure.

1 Localized RKHS guarantee

Condition on a kernel neighborhood of the test embedding:

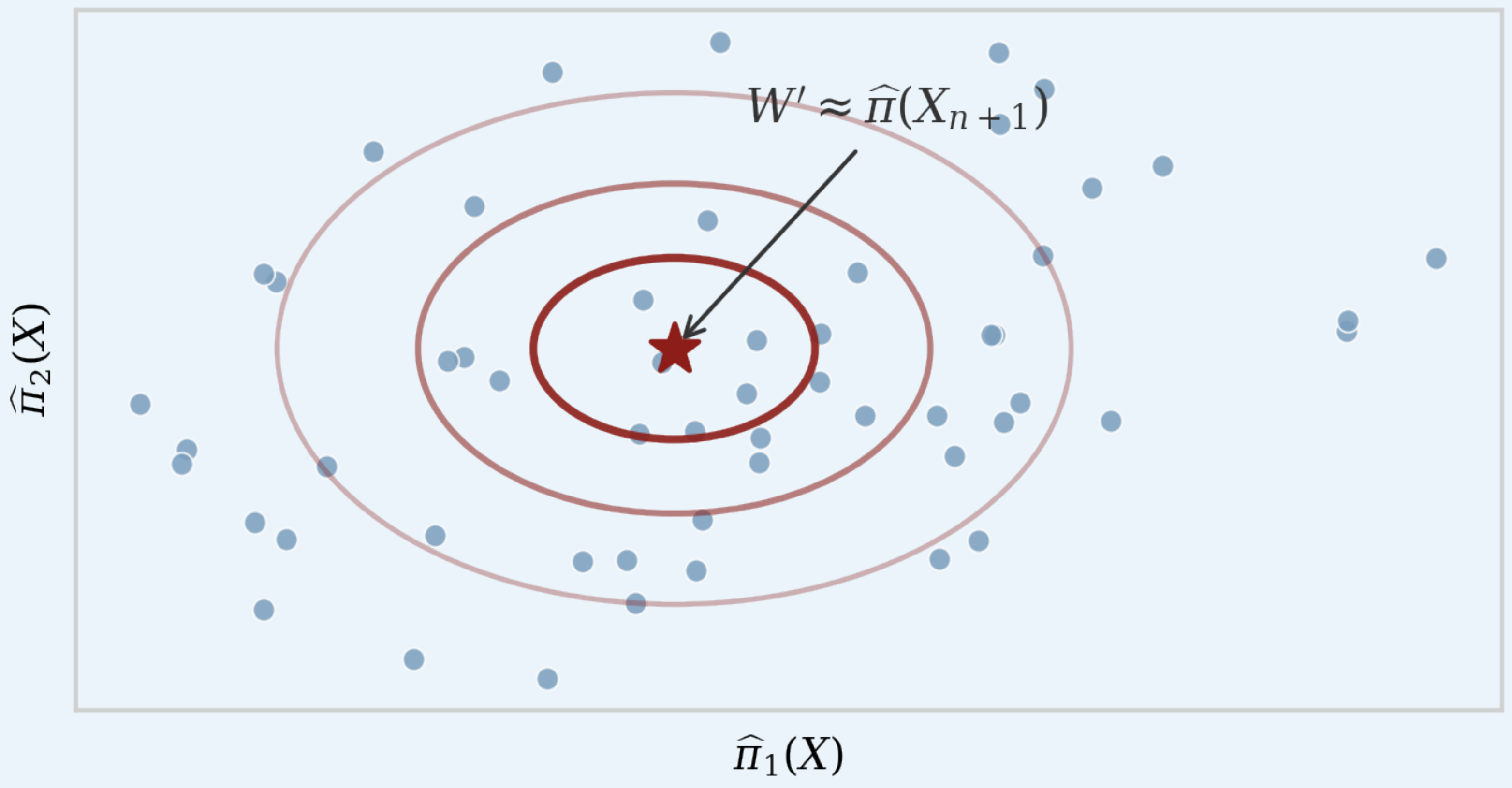
$$\mathbb{P}\{Y_{n+1} \in \widehat{C}^*_{\text{rand}}(X_{n+1}) \mid W'\} = 1 - \alpha - \text{Gap}(W')$$

where $W' \mid X_{n+1} = x$ is drawn from $\psi^*_W(\widehat{\pi}(x), \cdot)$

The gap is an RKHS residual term:

$$\text{Gap}(W') = \frac{\mathbb{E}\left[\sum_{i=1}^{n+1} \hat{v}_i \, \psi^*_W(W', \widehat{\pi}(X_i))\right]}{\mathbb{E}\left[\psi^*_W(W', \widehat{\pi}(X))\right]}$$

Kernel neighborhood in latent space



Local in the learned representation; not exact at every raw x.

local mass ↑ → gap ↓

γ controls locality

λ controls penalty

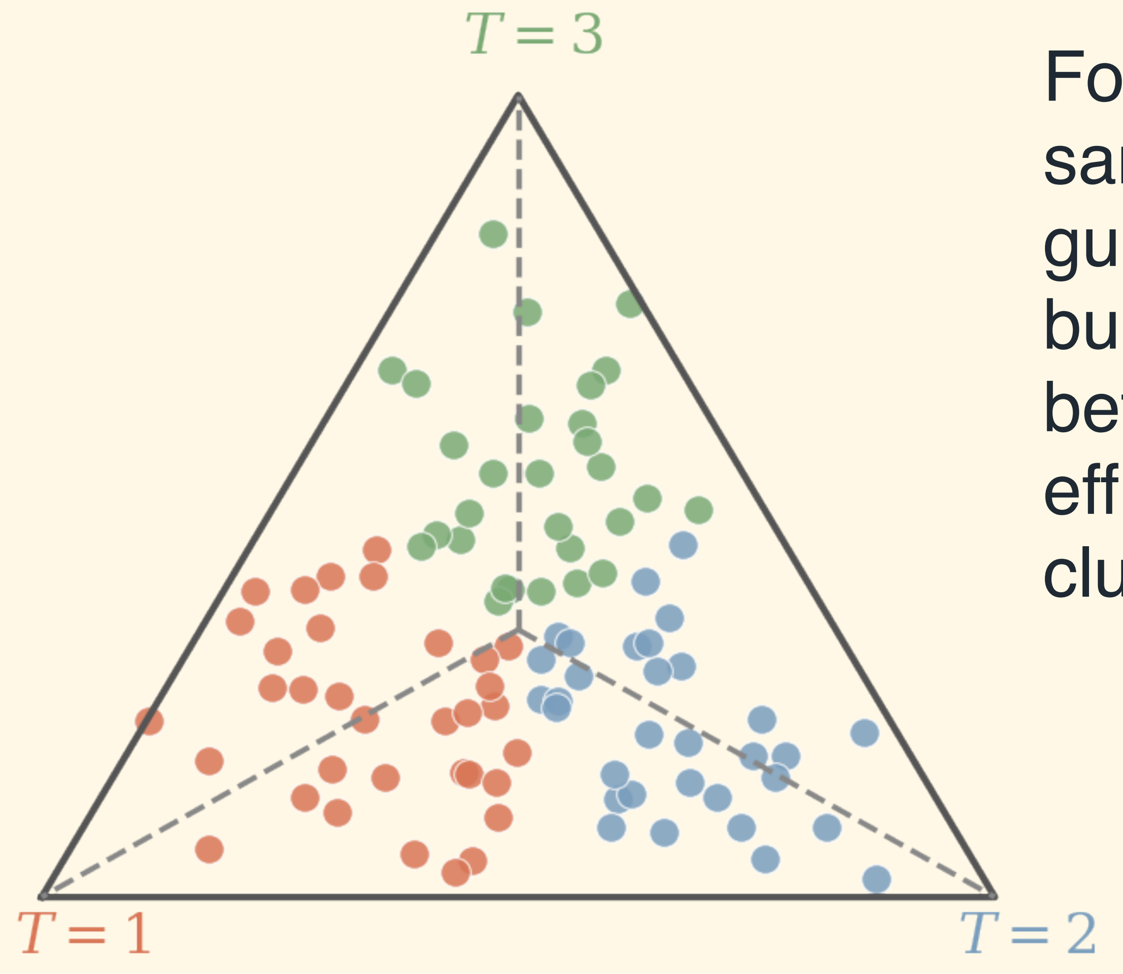
2 Exact group guarantee in latent clusters

If the oracle latent cluster T(X) is known and encoded as linear features:

$$\Phi^*(X) = (\, 1\{T(X)=1\}, \ldots, 1\{T(X)=K\} \,)^\top$$

$$\mathbb{P}\{Y_{n+1} \in \widehat{C}^*_{\text{rand}}(X_{n+1}) \mid T(X_{n+1}) = k\} = 1 - \alpha$$

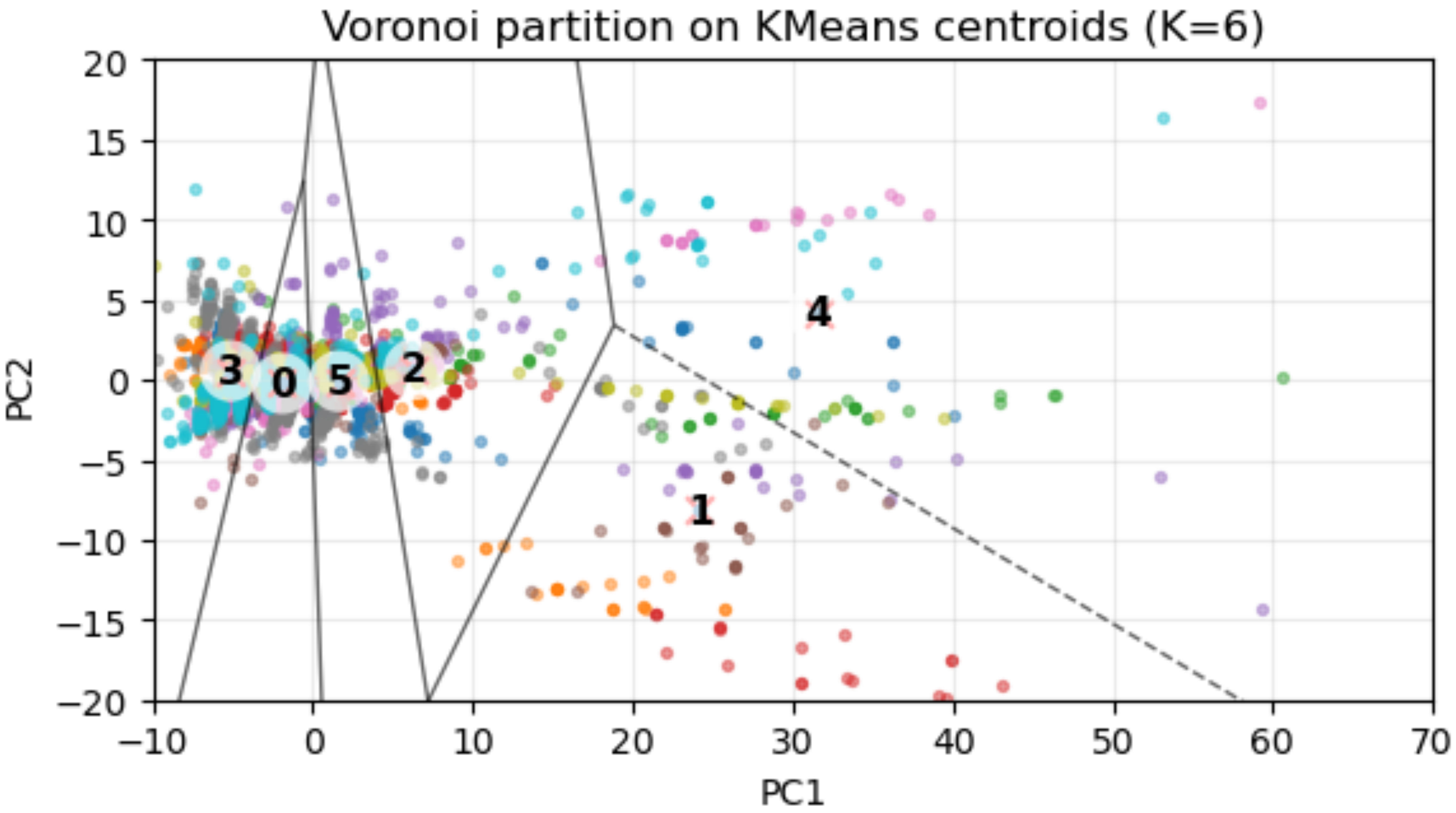
Oracle latent clusters



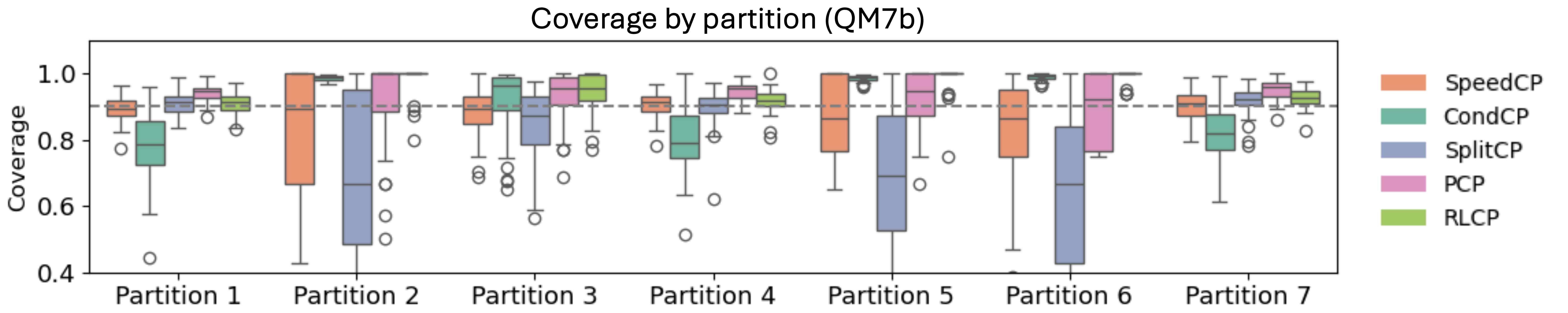
For estimated clusters $\hat{T}(X)$, the same group-conditional guarantee applies with $\Phi^*(X)$ built from $\hat{T}$, while mismatch between $\hat{T}$ and the true T affects efficiency / transfer to true clusters.

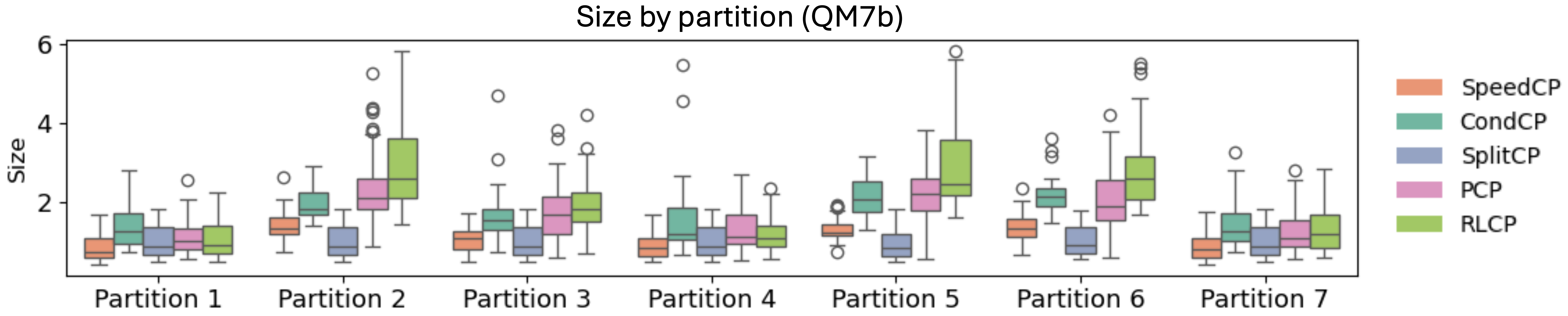
Experiment: predicting molecule polarizability

- Goal: 90% coverage on each partition of latent space (each point = test set molecule)
- Fit GNN and use PCA on last layer as latent embeddings
- QM7b Molecule graphs dataset (Training: 1000, Calibration: 500, Test: 500)



Experiment: predicting molecule polarizability





SpeedCP: stable coverage across latent space + tighter prediction sets

Takeaways: SpeedCP

- **Conditional validity**: provides finite sample coverage guarantees in the low rank setting
- **High-dimensional adaptivity**: condition on low rank embeddings for local calibration in sparse regions
- **Piecewise affine λ and S-paths**: avoid slow repeated RKHS refitting and yield tighter prediction sets