Towards Local and Epistemic Uncertainty Quantification with Conformal Prediction

Luben M. C. Cabezas (UFSCar/USP/Inria)

Mateus P. Otto
(Stats/UW)

Rafael B. Stern
(IME-USP)

Rafael Izbicki
(UFSCar)

Vagner S. Santos
(UFSCar)

Thiago R. Ramos
(UFSCar)

Luben M. C. Cabezas
(UFSCar/ICMC-USP/Inria)

Presentation Outline

- Motivation
- Introduction to Conformal prediction
- Locally-adaptive Conformal Prediction
    - Context
    - Our approach
- Epistemic Conformal Prediction
    - Context
    - Our approach

Motivation

Prediction tasks

- Main interest i predictions
- **Supervised le** es X through model $g(x)$.
- Today: models
- But even best

Which one is the dog?

Misleading conclusions

- Predictions alone may lead to misleading conclusions
- Can also be dangerous to use in high-stakes contexts alone

The central question is not only:

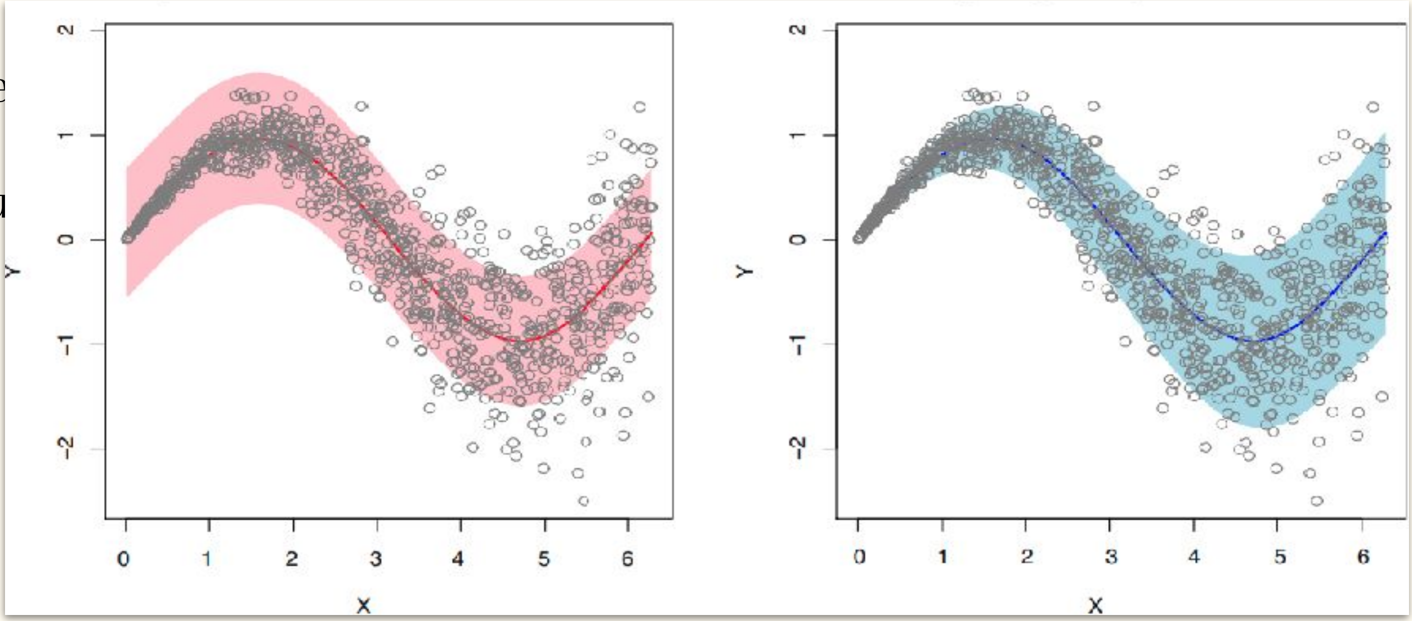
"What is the prediction?"

but also:

"How uncertain is it?"

From point predictions to uncertainty sets

- Inte
- **Usu**



- Several methods, but same problems: strong data assumptions for $g(x)$

  for $C(X)$ to be valid: $\mathbb{P}(Y \in C(X)) = 1 - \alpha$

Conformal Prediction

Main solution

- How to obtain symmetric and **valid** interval around $g(x)$ without strong data assumptions? $\mathbb{P}(Y \in C(X)) = 1 - \alpha$
- **Possible answer:** Conformal Prediction (Vovk et. al 2005, Shafer and Vovk 2008, Lei et. al 2018)
- Only assumption for validity: exchangeability (or i.i.d) (Vovk et al., 2005)
- Based in **conformal scores** (residuals): $s(X, Y) = |g(X) - Y|$

Split Conformal Prediction (split CP)

D is the dataset of interest and $1-\alpha$ the miscoverage level:

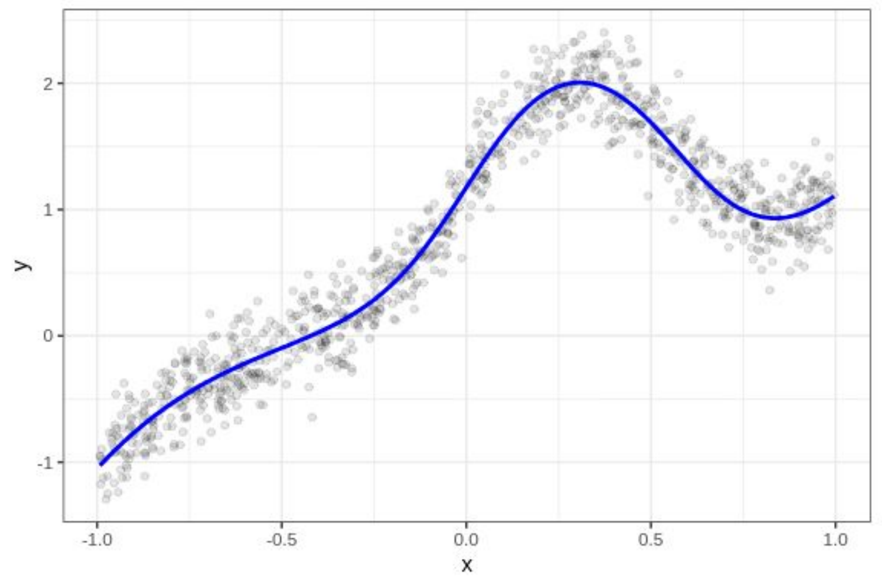
1. **Split** D in training and calibration sets. Train g in the training set.
2. **Compute** conformal scores $s(X,Y) = |g(X) - Y|$ in the calibration set.
3. **Sort** and compute the scores $1-\alpha$ quantile $\hat{t}_{1-\alpha}$.

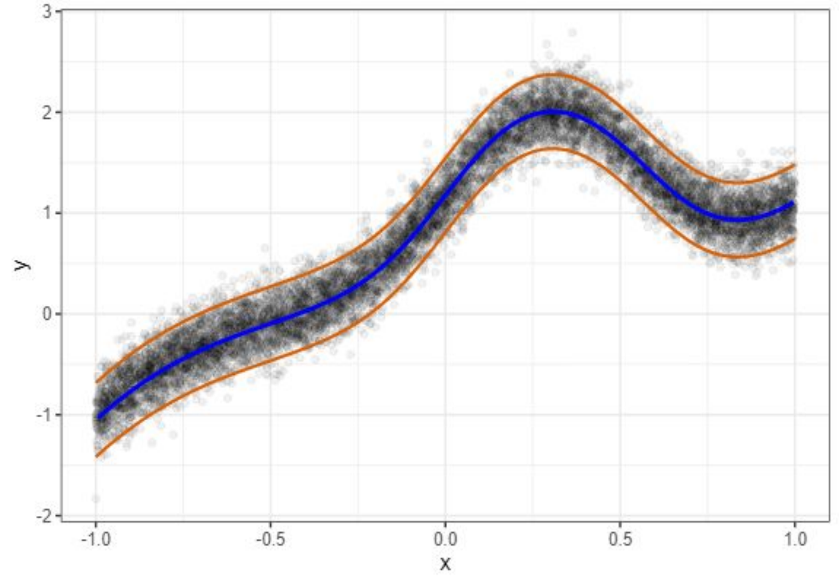
For x_{n+1} assign:

$$C^{(split)}(x_{n+1}) = \{y \in \mathcal{Y} : s(x_{n+1}, y) \leq \hat{t}_{1-\alpha}\}$$

$$= (g(x_{n+1}) - \hat{t}_{1-\alpha}, g(x_{n+1}) + \hat{t}_{1-\alpha})$$

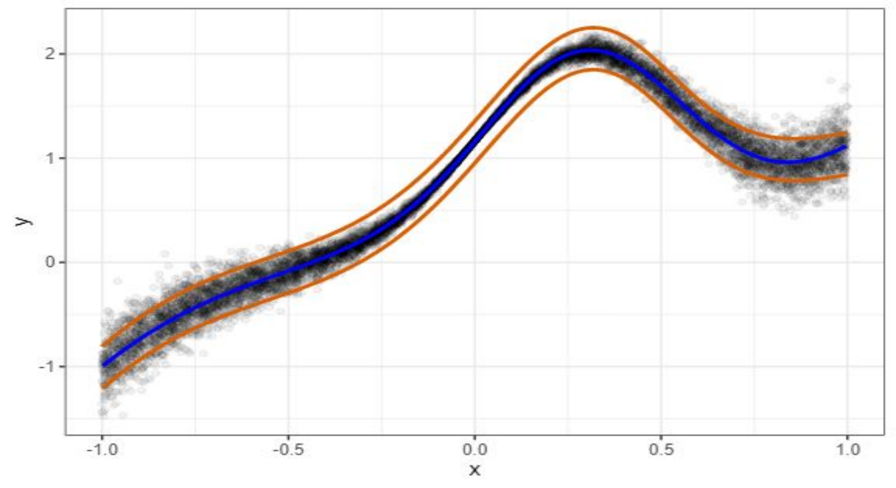
Example



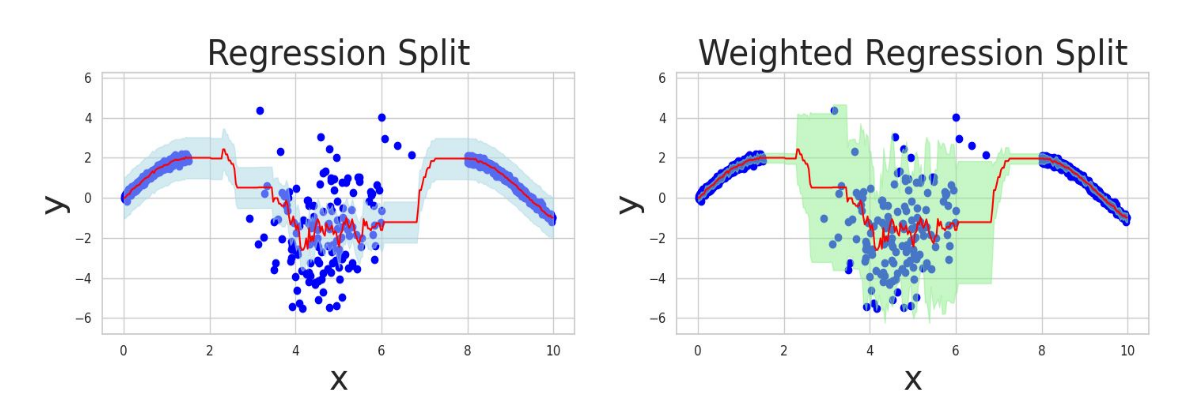


Source: https://www.tidymodels.org/learn/models/conformal-regression/

Typical problems:



Conditional coverage and adaptivity



Epistemic uncertainty and scarcity

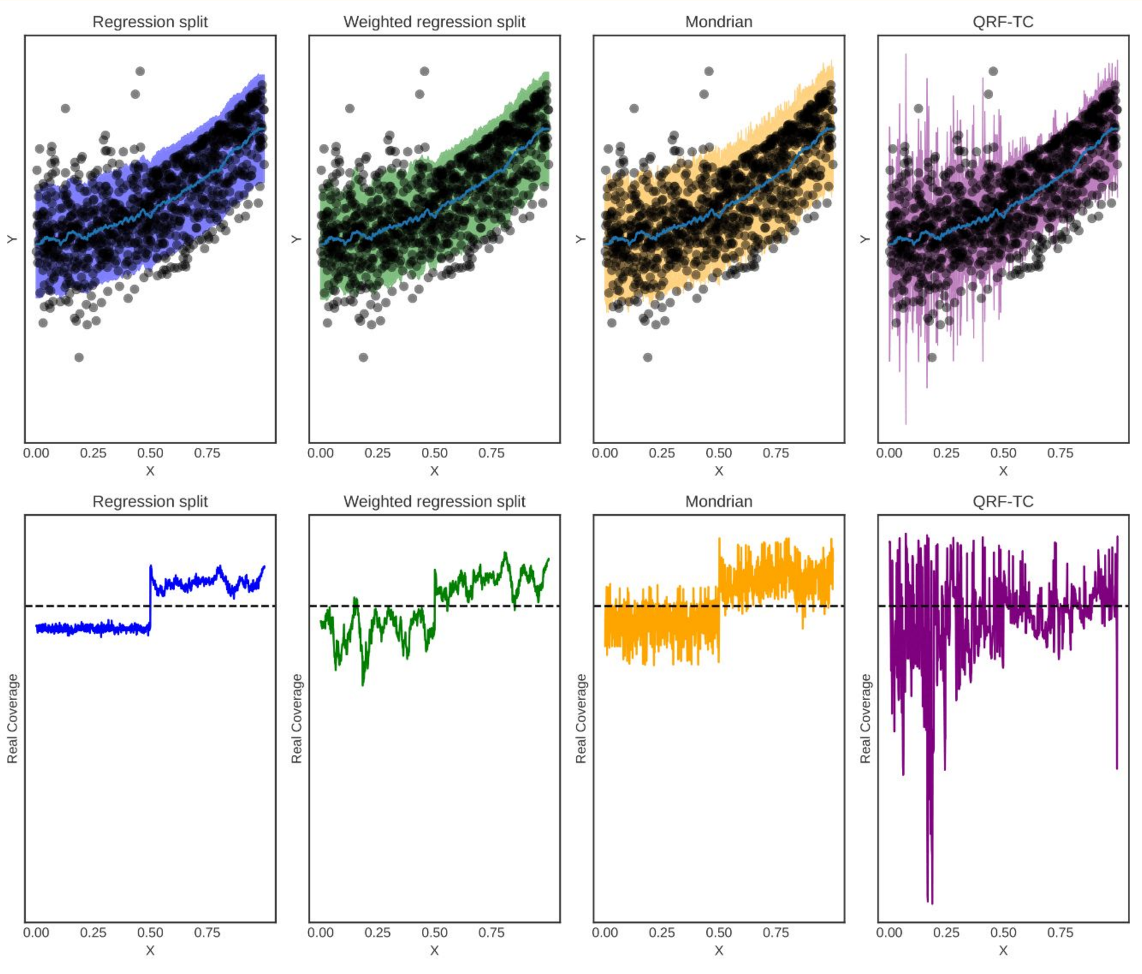
Part 1: Locally-adaptive Conformal Prediction

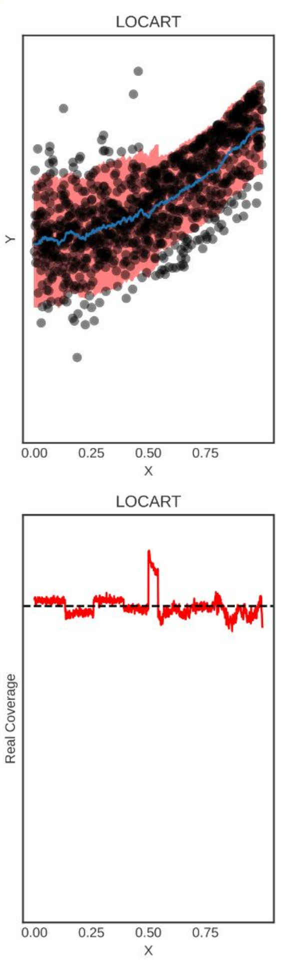
Main limitations

- CP does not have controlled conditional cover. $\mathbb{P}(Y \in C(X)|X = x) = 1 - \alpha$
- Prediction intervals may have problem on locality/adaptivity
- **Other strategies:** change the prediction model/conformal score (Romano et. al 2019, Izbicki et. al 2022) $s(X, Y) = \max\{\widehat{q}_{\alpha_1}(X) - Y, Y - \widehat{q}_{\alpha_2}(X)\}$
- Focus on asymp. conditional coverage, but scores not based on original $g(x)$
- Further strategies: can still fail to give adaptive intervals in several contexts.

Baseline approaches

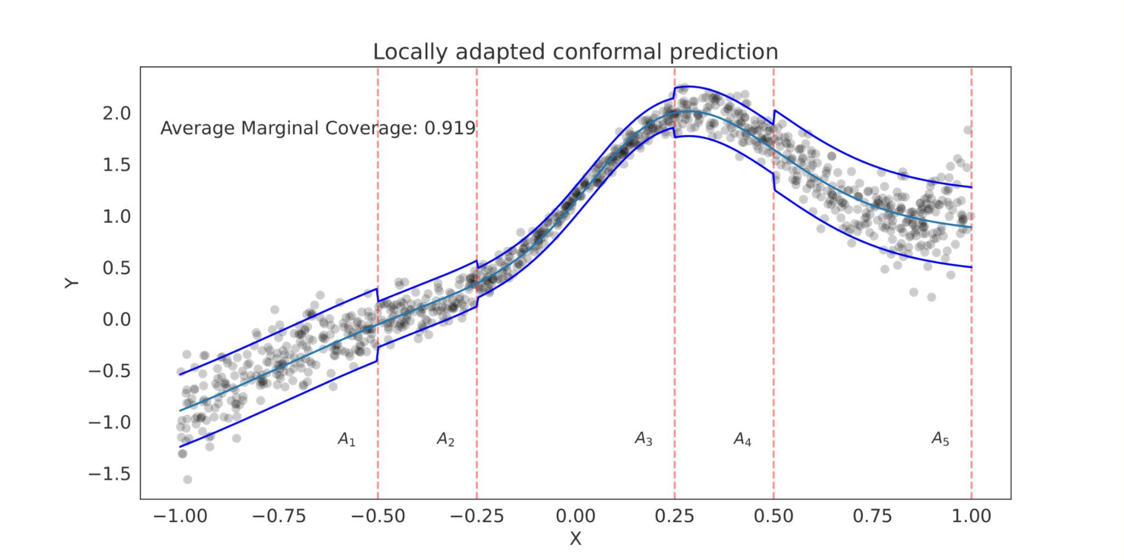
Our approach





LOCART (Local Regression Trees)

Rationale of our app



- Core idea: build conformal in
- In the partition $\mathcal{A} = \{A_1, \ldots,$
- **Outcome:** different cutoffs fo
- **Local Coverage Guarantee:**

$$\mathbb{P}[Y_{n+1} \in C_{local}(X_{n+1}) | X_{n+1} \in A_j] \geq 1 - \alpha, \forall A_j \in \mathcal{A}$$

Challenges

- Creating partition $\mathcal{A}$ is non-trivial:

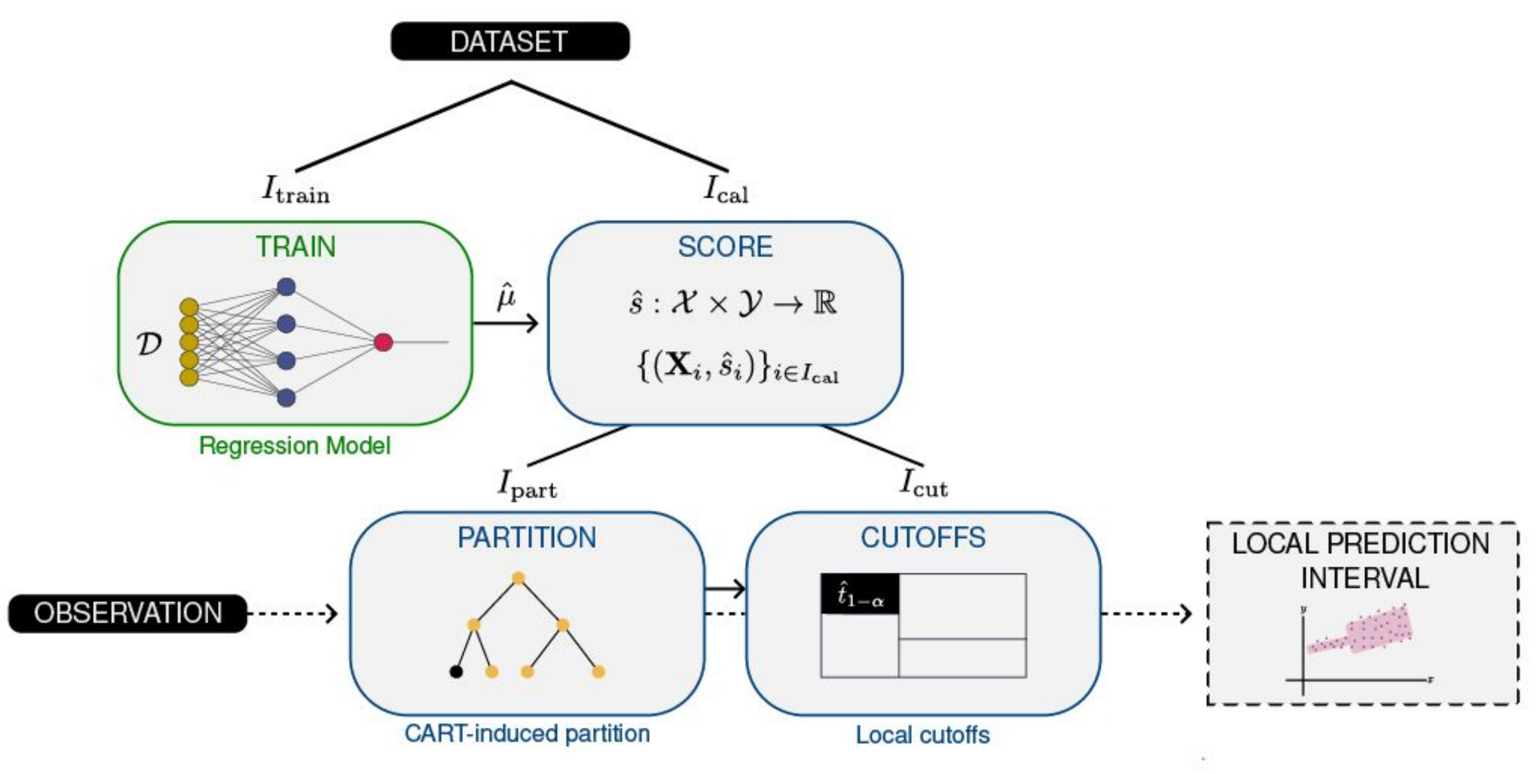
  Thin partition: sparsely populated -> poorly estimated cutoffs.

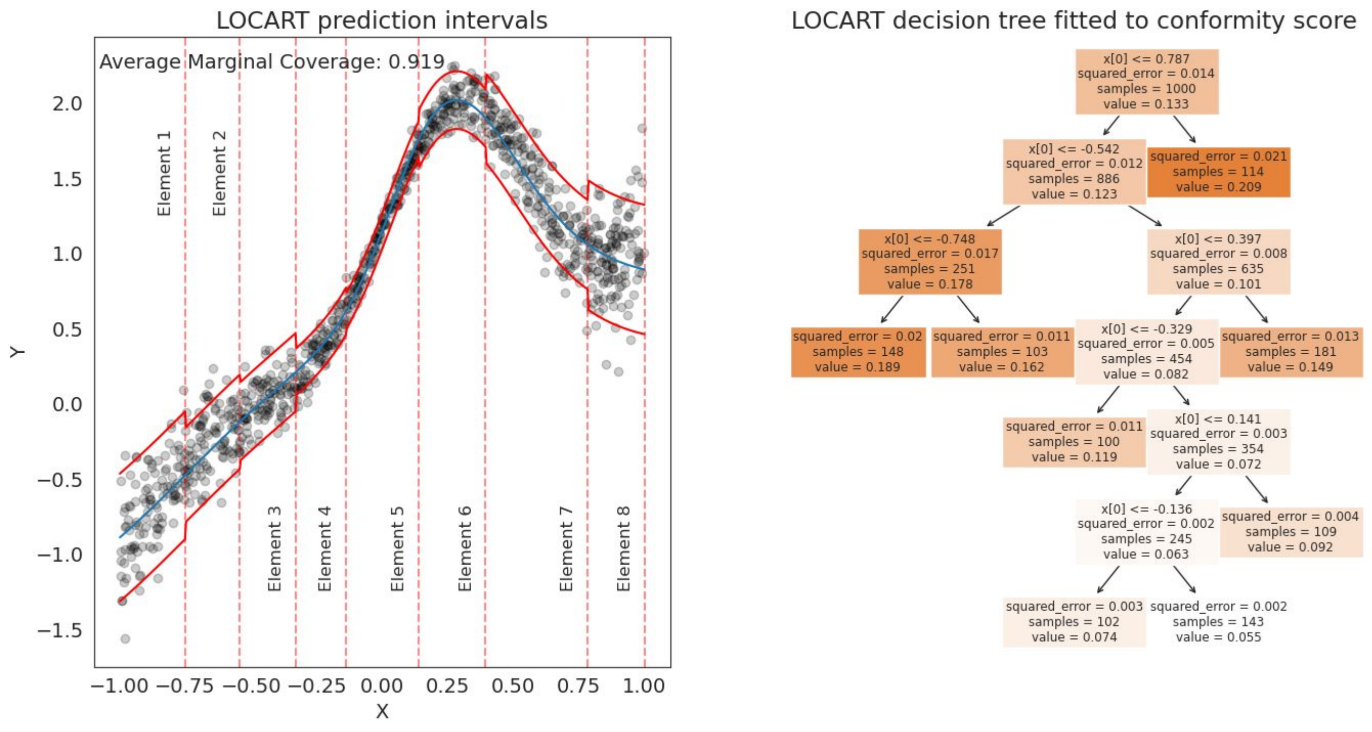
  Coarse partition: locally lacking -> maladaptive intervals.

- **Our idea:** use decision trees to derive an "optimal" partition.

$$x_1, x_2 \in A \in \mathcal{A} \leftrightarrow s(X,Y)|X = x_1 \sim s(X,Y)|X = x_2 \ .$$

$$x_1, x_2 \in A \in \mathcal{A} \leftrightarrow s(X,Y)|X = x_1 \sim s(X,Y)|X = x_2 \ .$$



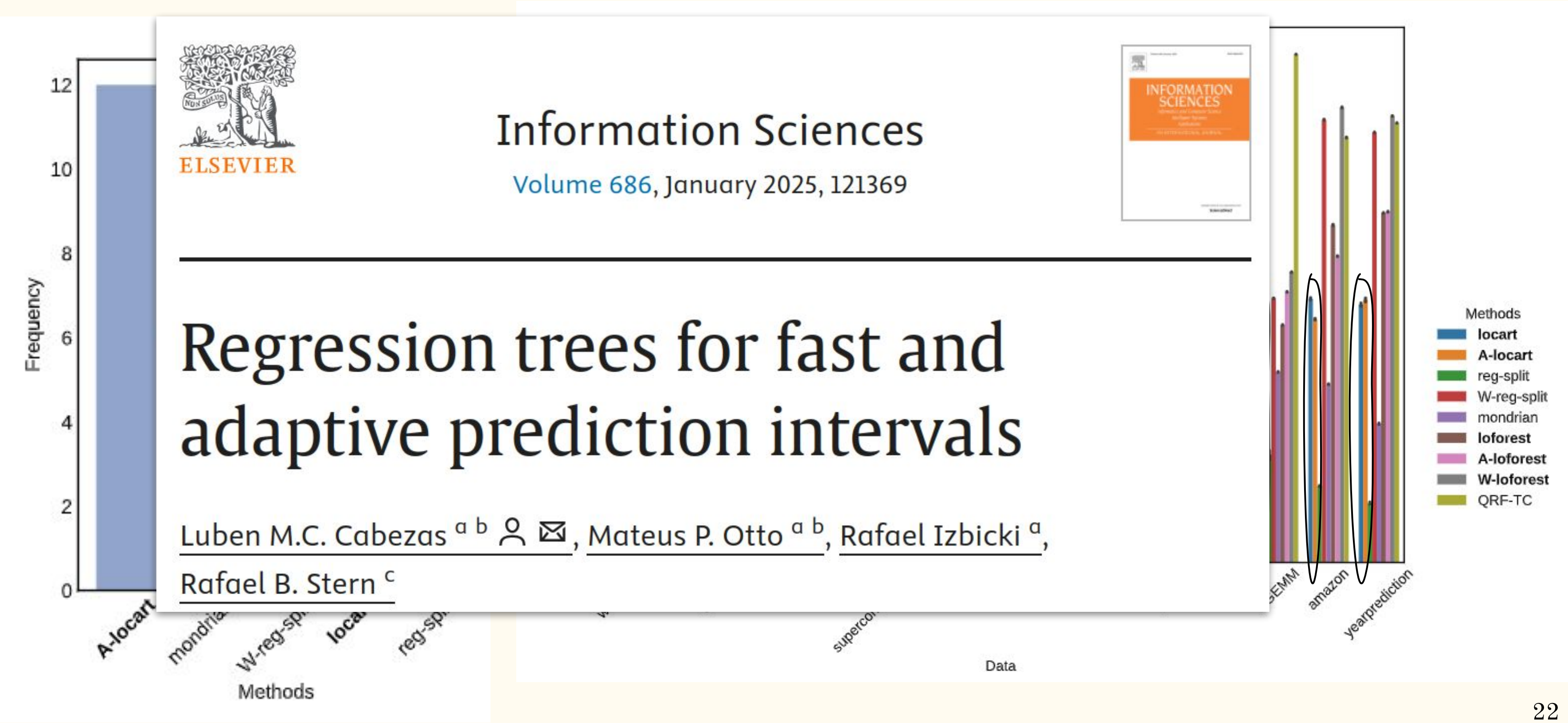


LOCART prediction intervals
Average Marginal Coverage: 0.919
Element 1
Element 2
Element 3
Element 4
Element 5
Element 6
Element 7
Element 8
Y
X
LOCART decision tree fitted to conformity score
x[0] <= 0.787
squared_error = 0.014
samples = 1000
value = 0.133
x[0] <= -0.542
squared_error = 0.012
samples = 886
value = 0.123
squared_error = 0.021
samples = 114
value = 0.209
x[0] <= -0.748
squared_error = 0.017
samples = 251
value = 0.178
x[0] <= 0.397
squared_error = 0.008
samples = 635
value = 0.101
squared_error = 0.02
samples = 148
value = 0.189
squared_error = 0.011
samples = 103
value = 0.162
x[0] <= -0.329
squared_error = 0.005
samples = 454
value = 0.082
squared_error = 0.013
samples = 181
value = 0.149
squared_error = 0.011
samples = 100
value = 0.119
x[0] <= 0.141
squared_error = 0.003
samples = 354
value = 0.072
x[0] <= -0.136
squared_error = 0.002
samples = 245
value = 0.063
squared_error = 0.004
samples = 109
value = 0.092
squared_error = 0.003
samples = 102
value = 0.074
squared_error = 0.002
samples = 143
value = 0.055

Some properties

- **Marginal Coverage** $\mathbb{P}(Y \in C(X)) = 1 - \alpha$
- **Local coverage** $\mathbb{P}[Y_{n+1} \in C_{local}(X_{n+1}) | X_{n+1} \in A_j] \geq 1 - \alpha, \forall A_j \in \mathcal{A}$
- **Asymp. Conditional coverage** $\lim_{|D_{\text{calib}}| \to \infty} \mathbb{P}\left(Y \in C(X) \mid X = x\right) = 1 - \alpha$
- Smart and lightweight partitioning

Real data



ELSEVIER
Information Sciences
Volume 686, January 2025, 121369

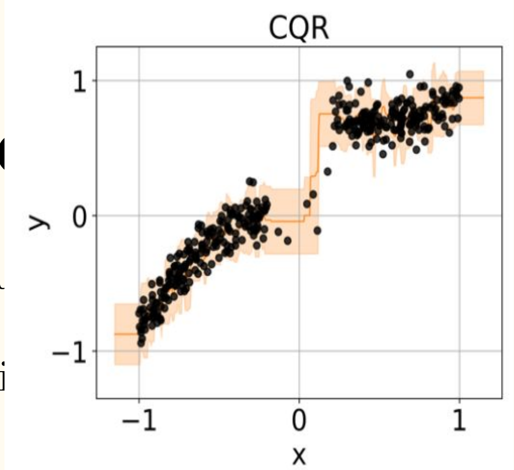
Regression trees for fast and adaptive prediction intervals

Luben M.C. Cabezas [a] [b], Mateus P. Otto [a] [b], Rafael Izbicki [a], Rafael B. Stern [c]

Part 2: Epistemic Conformal Prediction

A different kind (

- Standard CP: mostly captu
- Another source of uncertai (Hüllermeier & Waegeman 2021).



Aleatoric (irreducible)

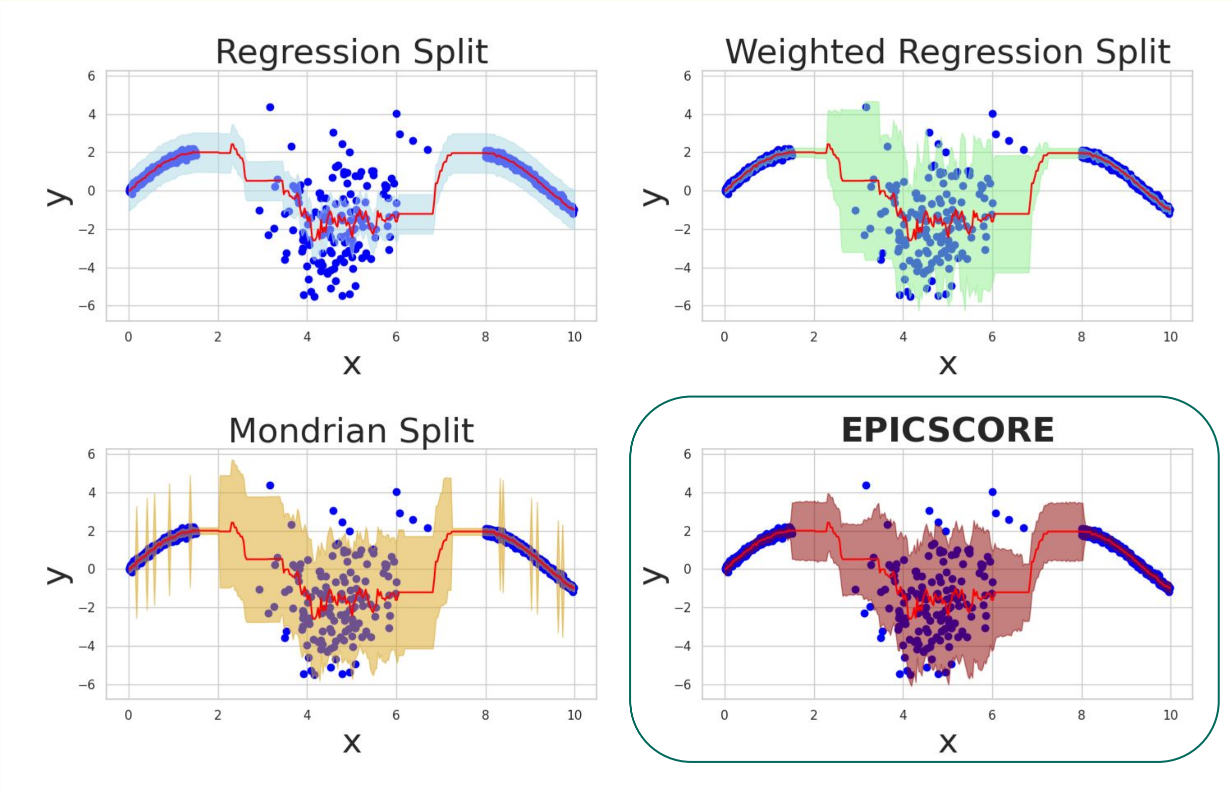
Inherent randomness in Y given X.

More data does NOT help.

Epistemic (reducible)

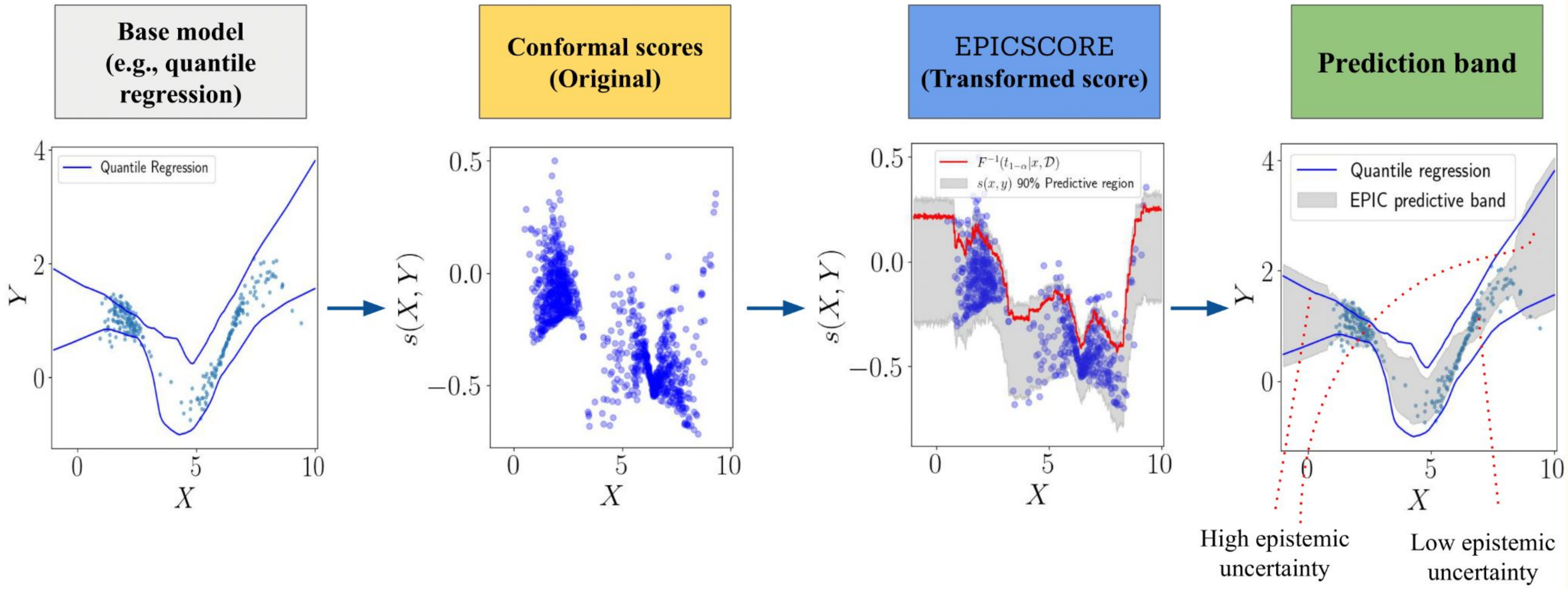
Uncertainty from lack of data or model knowledge.

More data CAN help.


Regression Split
Weighted Regression Split
Mondrian Split
EPICSCORE
y
x

Existing approaches

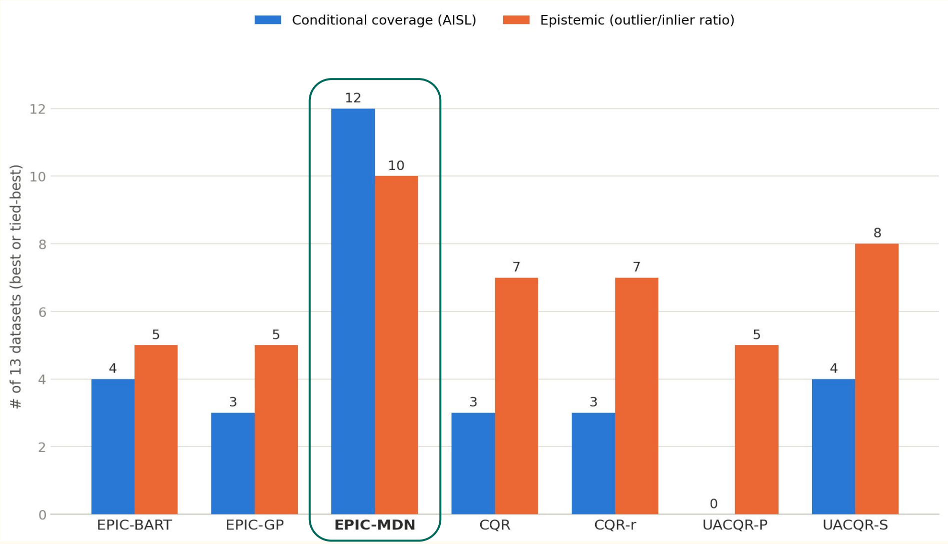
- Few strategies focus on integrating epistemic uncertainty more directly.
- Existing approaches are restricted to specific prediction models (quantile or regression).
- Our approach (EPICSCORE):
  - Integrate epistemic uncertainty on **any** conformal score;
  - Bayesian approaches to **model epistemic uncertainty** of $s(\mathbf{X}, Y)$;
  - **Maintains** marginal coverage guarantees (achieves asymp conditional);
  - **Enhance** rather than **replace** the conformal score;



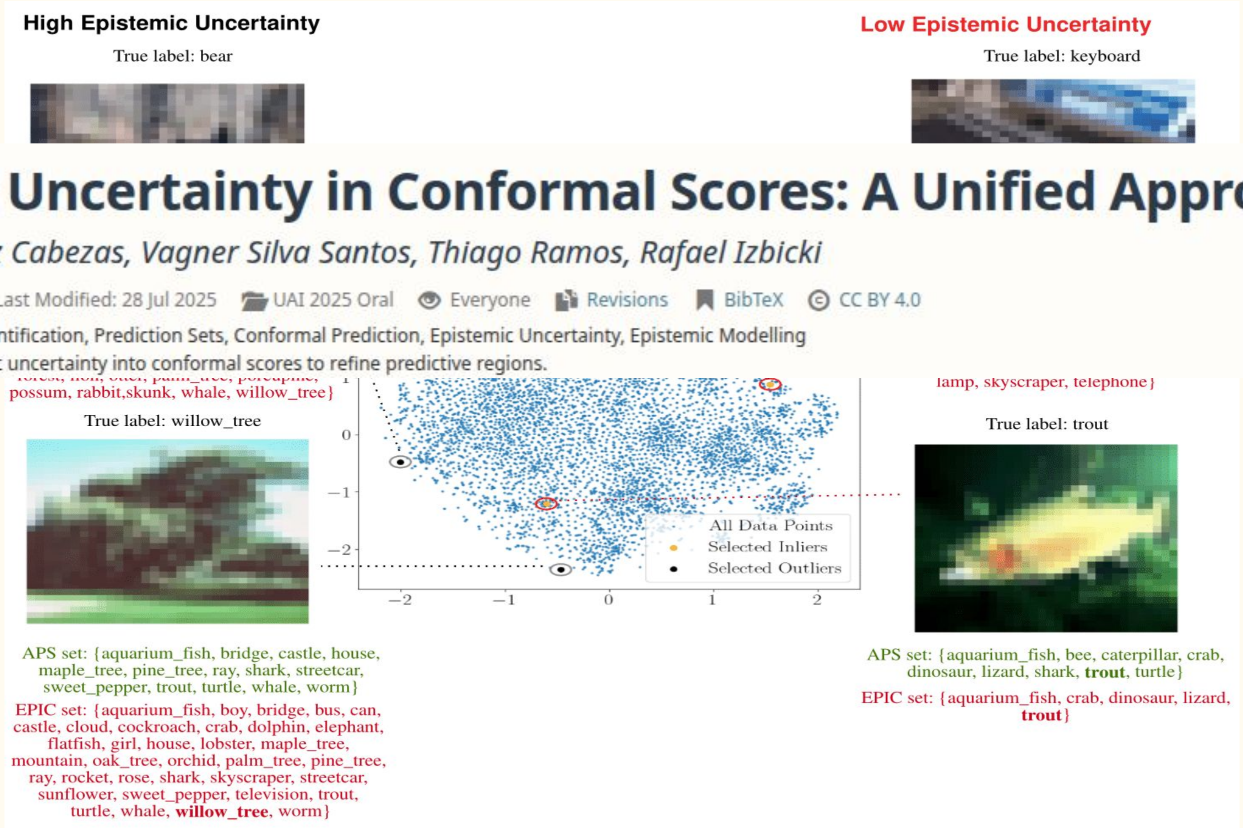
- **Marginal Coverage** guarantee
- **Asymp. conditional coverage** if predictive distribution converges.

Real data comparisons

- Comparisons over 13 real datasets
- Mainly two metrics:
  - Average Interval Score Loss (again)
  - Outlier to inlier interval ratio
- Comparing in regression and quantile regression contexts



Application to image data



Epistemic Uncertainty in Conformal Scores: A Unified Approach

Luben Miguel Cruz Cabezas, Vagner Silva Santos, Thiago Ramos, Rafael Izbicki

Published: 07 May 2025, Last Modified: 28 Jul 2025 · UAI 2025 Oral · Everyone · Revisions · BibTeX · CC BY 4.0

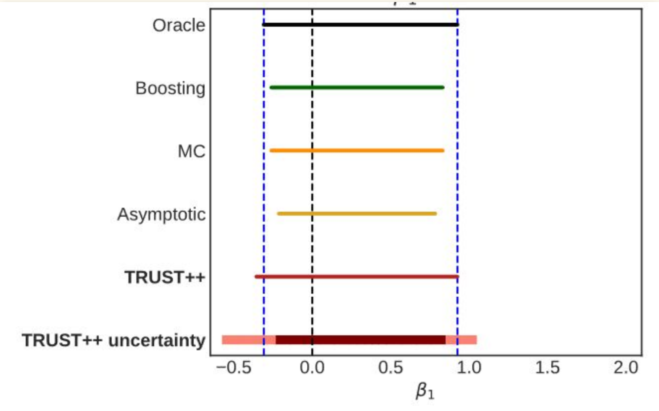
Keywords: Uncertainty Quantification, Prediction Sets, Conformal Prediction, Epistemic Uncertainty, Epistemic Modelling

TL;DR: Integrating epistemic uncertainty into conformal scores to refine predictive regions.

[Submitted on 28 Nov 2024 (v1), last revised 25 Sep 2025 (this version, v2)]

Conformal Calibration of Statistical Confidence Sets

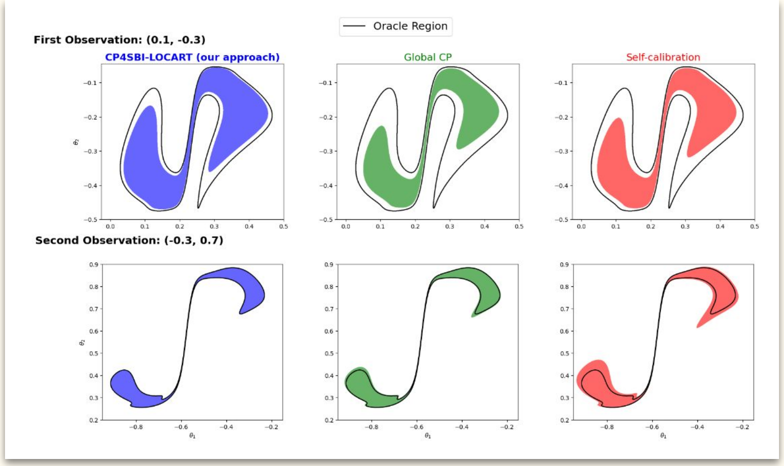
Luben M. C. Cabezas, Guilherme P. Soares, Thiago R. Ramos, Rafael B. Stern, Rafael Izbicki



[Submitted on 23 Aug 2025 (v1), last revised 27 Aug 2025 (this version, v2)]

CP4SBI: Local Conformal Calibration of Credible Sets in Simulation-Based Inference

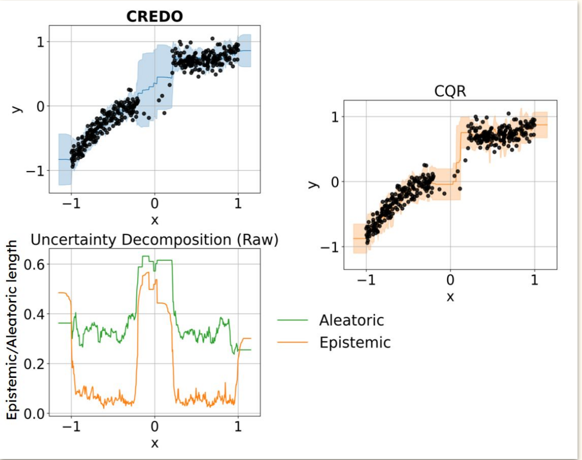
Luben M. C. Cabezas, Vagner S. Santos, Thiago R. Ramos, Pedro L. C. Rodrigues, Rafael Izbicki



[Submitted on 6 Mar 2026]

CREDO: Epistemic-Aware Conformalized Credal Envelopes for Regression

Luben M. C. Cabezas, Sabina J. Sloman, Bruno M. Resende, Fanyi Wu, Michele Caprio, Rafael Izbicki



Thank you!


ORDEM E PROGRESSO


UNIVERSIDADE FEDERAL DE SÃO CARLOS
ufscar


ICMC

References

- Bishop, C. M. (1994). Mixture density networks.
- Boström, H., & Johansson, U. (2020, August). Mondrian conformal regressors. In Conformal and probabilistic prediction and applications (pp. 114-133). PMLR.
- Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees.
- Cocheteux, M., Moreau, J., & Davoine, F. (2025, February). Uncertainty-aware online extrinsic calibration: A conformal prediction approach. In 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (pp. 6167-6176). IEEE.
- Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning (pp. 1050-1059). PMLR.
- Guan, L. (2023). Localized conformal prediction: A generalized inference framework for conformal prediction. Biometrika, 110(1), 33-50.
- Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Machine learning, 110(3), 457-506.

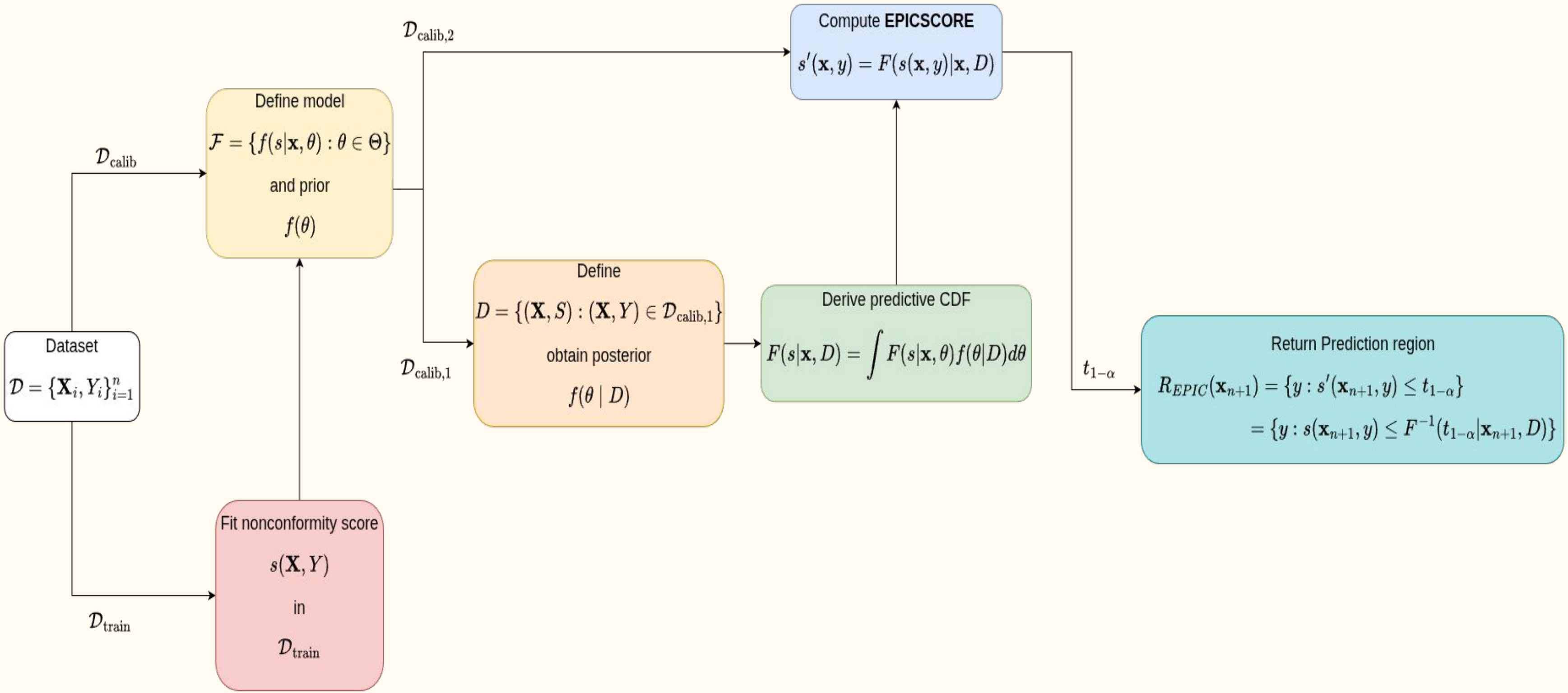
References

- Izbicki, R., Shimizu, G., & Stern, R. B. (2022). Cd-split and hpd-split: Efficient conformal regions in high dimensions. Journal of Machine Learning Research, 23(87), 1-32.
- Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018). Distribution-free predictive inference for regression. Journal of the American Statistical Association, 113(523), 1094-1111.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models.
- Romano, Y., Patterson, E., & Candes, E. (2019). Conformalized quantile regression. Advances in neural information processing systems, 32.
- Rossellini, R., Barber, R. F., & Willett, R. (2024, April). Integrating uncertainty awareness into conformalized quantile regression. In International Conference on Artificial Intelligence and Statistics (pp. 1540-1548). PMLR.
- Shafer, G., & Vovk, V. (2008). A tutorial on conformal prediction. Journal of Machine Learning Research, 9(3).

References

- Williams, C. K., & Rasmussen, C. E. (2006). Gaussian processes for machine learning (Vol. 2, No. 3, p. 4). Cambridge, MA: MIT press.
- Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic learning in a random world. Boston, MA: Springer US.

EPICSCORE



Dataset
$\mathcal{D} = \{\mathbf{X}_i, Y_i\}_{i=1}^n$
$\mathcal{D}_{\text{calib}}$
$\mathcal{D}_{\text{train}}$
Fit nonconformity score
$s(\mathbf{X}, Y)$
in
$\mathcal{D}_{\text{train}}$
Define model
$\mathcal{F} = \{f(s|\mathbf{x}, \theta) : \theta \in \Theta\}$
and prior
$f(\theta)$
$\mathcal{D}_{\text{calib},2}$
$\mathcal{D}_{\text{calib},1}$
Define
$D = \{(\mathbf{X}, S) : (\mathbf{X}, Y) \in \mathcal{D}_{\text{calib},1}\}$
obtain posterior
$f(\theta \mid D)$
Derive predictive CDF
$F(s|\mathbf{x}, D) = \int F(s|\mathbf{x}, \theta) f(\theta|D) d\theta$
Compute **EPICSCORE**
$s'(\mathbf{x}, y) = F(s(\mathbf{x}, y)|\mathbf{x}, D)$
$t_{1-\alpha}$
Return Prediction region
$R_{EPIC}(\mathbf{x}_{n+1}) = \{y : s'(\mathbf{x}_{n+1}, y) \leq t_{1-\alpha}\}$
$= \{y : s(\mathbf{x}_{n+1}, y) \leq F^{-1}(t_{1-\alpha}|\mathbf{x}_{n+1}, D)\}$

Improving LOCART

Partitioning could be improved with extra information about Y|X

- A-LOCART: **augment** the feature space with **extra** info
- Example: estimated $\mathbb{V}[Y|X]$ as an additional feature

Some model details

- Practical model choices for predictive distribution:
  - BART (Chipman et al. 2012);
  - Gaussian Process (Williams & Rassmussen 2006);
  - Mixture Density Network with MC-Dropout (Bishop 1994, Gal & Gharamani 2016)

Real data

Table 3
Real-world data information. For each dataset, n is the sample size, and p is the number of features.

Dataset	n	p	Source	Dataset	n	p	Source
Concrete	1030	8	Concrete (UCI)	Meps19	15781	141	Meps19 (clover repository)
Airfoil	1503	5	Airfoil (UCI)	Superconductivity	21263	81	Superconductivity (UCI)
Wine white	1599	11	Wine white (UCI)	News	39644	59	News (UCI)
Star	2161	48	Star (ACPI repository)	Protein	45730	8	Protein (UCI)
Wine red	4898	11	Wine red (UCI)	Wave energy converter	54000	49	WEC (UCI)
Cycle	9568	4	Cycle (UCI)	Amazon	130000[a]	500	Amazon (Kaggle)
Electric	10000	12	Electric (UCI)	SGMM	241600	14	SGEMM (UCI)
Bike	10886	12	Bike (Kaggle)	Year Prediction	463715	90	Year Prediction (UCI)

[a] In each iteration, we subsampled from a large text dataset with 393,930 instances.

Some properties (LOCART)

- **Marginal Coverage** $\mathbb{P}(Y \in C(X)) = 1 - \alpha$
- **Local coverage** $\mathbb{P}[Y_{n+1} \in C_{local}(X_{n+1}) | X_{n+1} \in A_j] \geq 1 - \alpha, \forall A_j \in \mathcal{A}$
- Oracle Consistency $|t^*(X_{n+1}) - \hat{t}_{1-\alpha}(X_{n+1})| = o_{\mathbb{P}}(1)$
- **Asymp. Conditional coverage** $\lim_{|D_{\text{calib}}| \to \infty} \mathbb{P}\left(Y \in C(X) \mid X = x\right) = 1 - \alpha$
- Smart and lightweight partitioning

Dataset	EPIC-BART	EPIC-GP	EPIC-MDN	CQR	CQR-r	UACQR-P	UACQR-S
airfoil	**1.028 (0.034)**	**1.01 (0.018)**	**1.048 (0.04)**	**1.005 (0.013)**	**1.004 (0.012)**	**1.002 (0.01)**	**1.005 (0.016)**
bike	0.991 (0.007)	**1.01 (0.007)**	0.989 (0.007)	0.985 (0.007)	0.985 (0.007)	0.99 (0.005)	0.981 (0.01)
concrete	0.983 (0.033)	0.995 (0.016)	**1.043 (0.029)**	0.998 (0.012)	0.998 (0.012)	0.998 (0.008)	0.998 (0.012)
cycle	0.953 (0.014)	0.966 (0.014)	0.949 (0.014)	**0.997 (0.005)**	**0.997 (0.005)**	**1.0 (0.004)**	**0.998 (0.007)**
electric	**1.01 (0.004)**	**1.01 (0.004)**	**1.012 (0.008)**	**1.008 (0.004)**	**1.007 (0.003)**	**1.006 (0.003)**	**1.01 (0.005)**
homes	1.116 (0.019)	1.101 (0.016)	**1.173 (0.027)**	1.088 (0.015)	1.081 (0.014)	1.039 (0.006)	1.094 (0.014)
meps19	**1.052 (0.029)**	**1.042 (0.025)**	**1.044 (0.028)**	**1.04 (0.025)**	**1.04 (0.025)**	**1.021 (0.015)**	**1.045 (0.026)**
protein	1.001 (0.002)	1.001 (0.003)	**1.006 (0.002)**	1.002 (0.001)	1.002 (0.001)	1.001 (0.001)	1.002 (0.001)
star	**0.994 (0.006)**	**0.994 (0.003)**	**0.999 (0.007)**	**0.994 (0.003)**	**0.993 (0.003)**	**0.997 (0.003)**	**0.997 (0.003)**
superconductivity	1.021 (0.007)	0.987 (0.006)	1.029 (0.008)	**1.041 (0.007)**	**1.041 (0.007)**	1.023 (0.004)	**1.049 (0.008)**
WEC	**1.012 (0.006)**	1.003 (0.008)	**1.02 (0.01)**	**1.016 (0.003)**	**1.016 (0.003)**	1.011 (0.002)	**1.021 (0.004)**
winered	1.008 (0.012)	1.007 (0.008)	**1.072 (0.022)**	0.993 (0.008)	0.992 (0.008)	0.996 (0.005)	0.992 (0.009)
winewhite	1.031 (0.007)	1.024 (0.006)	**1.053 (0.014)**	1.018 (0.007)	1.018 (0.007)	1.018 (0.006)	**1.034 (0.01)**

Dataset	EPIC-BART	EPIC-GP	EPIC-MDN	CQR	CQR-r	UACQR-P	UACQR-S
airfoil	19.361 (0.234)	19.704 (0.27)	**18.799 (0.29)**	20.521 (0.234)	20.535 (0.236)	23.021 (0.337)	20.188 (0.3)
bike $\times(10^1)$	44.722 (0.297)	47.818 (0.320)	**43.858 (0.326)**	45.628 (0.256)	45.638 (0.258)	53.413 (0.376)	**43.815 (0.385)**
concrete	**42.765 (0.723)**	45.276 (0.764)	44.442 (0.8)	46.882 (0.681)	46.896 (0.683)	52.789 (1.097)	47.324 (1.349)
cycle	34.435 (0.142)	35.054 (0.131)	**34.077 (0.129)**	39.218 (0.134)	39.408 (0.136)	43.775 (0.181)	35.346 (0.197)
electric	0.099 (<0.001)	0.096 (<0.001)	**0.082 (<0.001)**	0.102 (0.001)	0.102 (0.001)	0.111 (0.001)	0.097 (<0.001)
homes $\times(10^5)$	7.739 (0.066)	8.098 (0.072)	**7.225 (0.049)**	8.360 (0.075)	8.433 (0.078)	11.427 (0.131)	8.544 (0.107)
meps19	**65.085 (1.469)**	**64.907 (1.56)**	**64.3 (1.528)**	**64.239 (1.56)**	**64.239 (1.56)**	71.015 (1.763)	**63.737 (1.461)**
protein	17.687 (0.019)	18.096 (0.037)	**17.417 (0.019)**	17.7 (0.015)	17.7 (0.016)	18.149 (0.015)	17.691 (0.015)
star $\times(10^1)$	**98.466 (0.768)**	**98.033 (0.750)**	**98.725 (0.754)**	**97.770 (0.725)**	**97.791 (0.724)**	99.782 (0.647)	99.809 (0.968)
superconductivity	74.37 (0.222)	80.278 (0.266)	**70.212 (0.196)**	75.496 (0.219)	75.508 (0.218)	87.929 (0.513)	73.971 (0.404)
WEC $\times(10^5)$	2.925 (0.009)	2.665 (0.012)	**2.374 (0.010)**	3.138 (0.009)	3.142 (0.009)	3.517 (0.010)	3.046 (0.010)
winered	**3.007 (0.058)**	**3.009 (0.059)**	**2.977 (0.05)**	**2.979 (0.069)**	**2.978 (0.069)**	3.059 (0.069)	**2.999 (0.063)**
winewhite	3.334 (0.03)	3.327 (0.034)	**3.219 (0.03)**	3.316 (0.036)	3.315 (0.036)	3.378 (0.038)	**3.2 (0.036)**