CLEAR: Calibrated Learning for Epistemic and Aleatoric Risk

| ICLR 2026 |

Ilia Azizi[1*]

Juraj Bodik[1,2*]

Jakob Heiss[2*]

Bin Yu[2]

Unil.

[1]University of Lausanne, Switzerland

[2]UC Berkeley, USA

[*]Equal contribution

UC Berkeley

Why is uncertainty quantification still hard?

Given i.i.d. data (X_i, Y_i), we want a prediction interval $C(X_{n+1})$ for Y_{n+1}.

Standard target: marginal validity

$$\mathbb{P}\big(Y_{n+1} \in C(X_{n+1})\big) \geq 1 - \alpha.$$

Given i.i.d. data (X_i, Y_i), we want a prediction interval $C(X_{n+1})$ for Y_{n+1}.

Standard target: marginal validity

$$\mathbb{P}\big(Y_{n+1} \in C(X_{n+1})\big) \geq 1 - \alpha.$$

Practical target: conditional reliability

$$\mathbb{P}\big(Y_{n+1} \in C(X_{n+1}) \mid X_{n+1} = x\big) \approx 1 - \alpha.$$

Why is uncertainty quantification still hard?

Given i.i.d. data (X_i, Y_i), we want a prediction interval $C(X_{n+1})$ for Y_{n+1}.

Standard target: marginal validity

$$\mathbb{P}\big(Y_{n+1} \in C(X_{n+1})\big) \geq 1 - \alpha.$$

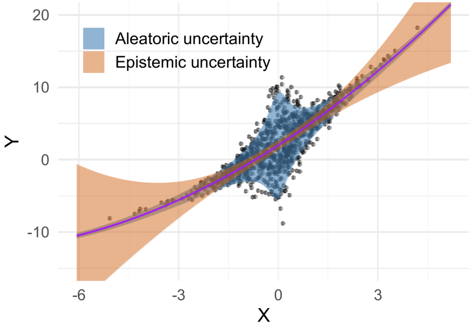
Practical target: conditional reliability

$$\mathbb{P}\big(Y_{n+1} \in C(X_{n+1}) \mid X_{n+1} = x\big) \approx 1 - \alpha.$$

- Conditional guarantees are impossible to obtain without additional strong assumptions.
- Conformal methods guarantee only marginal validity.

Uncertainty has many sources:

- **Aleatoric**: irreducible noise,
- **Epistemic**: model uncertainty.



	More obs.	Better prior	More cov.	Less noise	Algorithms
Epistemic	↓	↓	↑/↓	↓	Ensembles, PCS, GPR
Aleatoric	–	–	↓	↓↓	Quantile Reg., ML for (μ, σ)

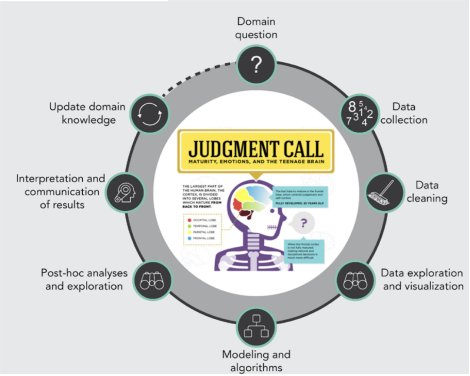
PCS framework: one culture

Three principles of data science:

(P)redictability [ML and Stats]

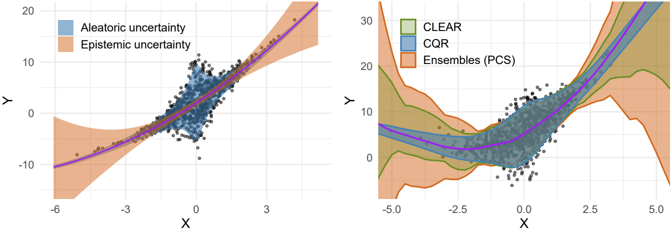
(C)omputability [ML]

(S)tability [Stats, control theory, numerical analysis]



Reference: https://vdsbook.com

CLEAR: separate calibration of aleatoric and epistemic uncertainty



$$C = \left[\hat{f} - \boxed{\gamma_1} \underbrace{\hat{q}^{\text{ale}}_{\alpha/2}}_{\text{est. via QR}} - \boxed{\gamma_2} \underbrace{\hat{q}^{\text{epi}}_{\alpha/2}}_{\text{est. via PCS-UQ}}, \; \hat{f} + \boxed{\gamma_1} \underbrace{\hat{q}^{\text{ale}}_{1-\alpha/2}}_{\text{est. via QR}} + \boxed{\gamma_2} \underbrace{\hat{q}^{\text{epi}}_{1-\alpha/2}}_{\text{est. via PCS-UQ}} \right]$$

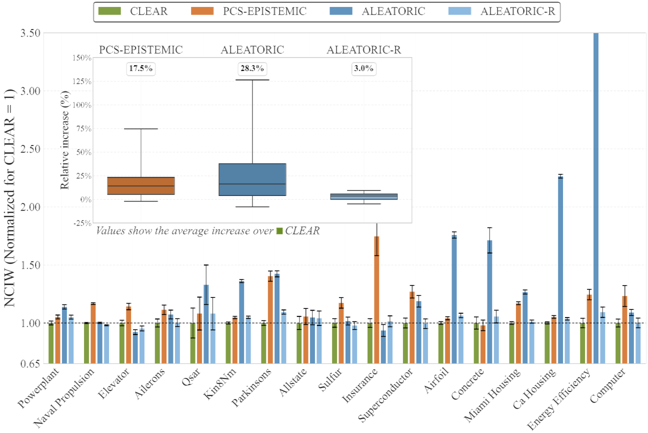
with $\gamma_2 = \lambda\gamma_1$. We estimate γ_1 and λ such that:

1. Marginal coverage is valid
2. Loss is minimized on a validation set

Real-world results - evaluation on 17 datasets

Average width reduction of

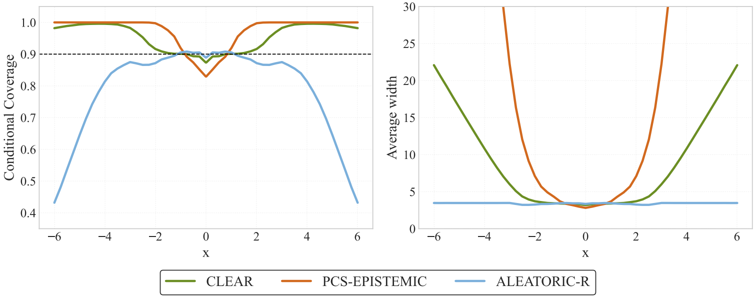
- 17% compared to PCS-UQ
- 28% compared to CQR



Normalized Calibrated Interval Width (NCIW)

Conditional coverage

- Experimental and theoretical justifications of improved conditional coverage



Example: Ames housing with few vs many covariates

Question: How do interval methods behave when we use only 2 strong predictors versus all 80 predictors?

Setting	Method	NCIW	QL	Avg. Width ($)	Cov.
2 features	PCS	0.21	381	107880	0.87
	CQR	0.18	345	104741	**0.90**
	CLEAR	**0.17**	**313**	**95177**	0.89
80 features	PCS	**0.10**	**192**	57594	**0.89**
	CQR	0.12	219	62398	0.88
	CLEAR	**0.10**	**192**	**55910**	0.88

Takeaway:

- Aleatoric methods (CQR) are competitive for few features.
- Epistemic methods (PCS) are competitive with many features.
- **CLEAR is adaptive to both!**

CLEAR

Code, paper & project page

https://unc03892.github.io/clear/