CONTRA: Conformal Prediction Region via Normalizing Flow Transformation

Zhenhan Fang[1] Aixin Tan[1] Jian Huang[2]

[1]The University of Iowa

[2]The Hong Kong Polytechnic University

Uncertainty quantification

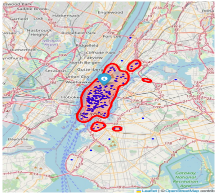
Our Goals
Given $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$ and $\mathbf{x}_{n+1}$, form a prediction region $C(\mathbf{x}_{n+1})$ for $\mathbf{y}_{n+1}$ s.t.

$$\mathbb{P}(\mathbf{y}_{n+1} \in C(\mathbf{x}_{n+1})) \geq 1 - \alpha\,.$$

- Prediction region for multi-dimensional $\mathbf{y}$ with coverage guarantee.
- Prediction region with corresponding density shape.

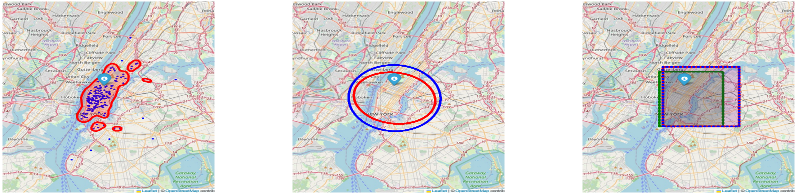
Our solution:
CONTRA — a combination of conformal prediction and normalizing flows.

Prediction Region for NYC Taxi Drop-off Location.



(a) KDE (No Coverage Guarantee)

Prediction Region for NYC Taxi Drop-off Location.

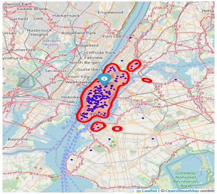


(a) KDE (No Coverage Guarantee)

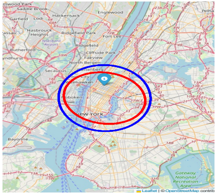
(b) NLE(R), RCP(B)

(c) MCQR(R), Dist-Split(G), $\mathrm{CQR}_{\mathrm{bon}}$(B)

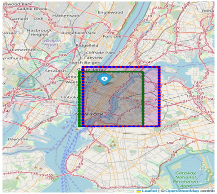
Prediction Region for NYC Taxi Drop-off Location.



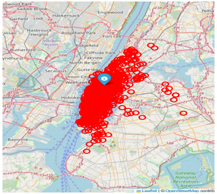
(a) KDE (No Coverage Guarantee)



(b) NLE(R), RCP(B)

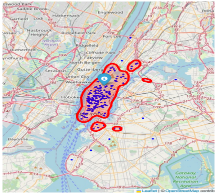


(c) MCQR(R), Dist-Split(G), CQR_{bon}(B)

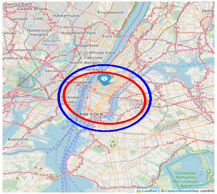


(d) PCP

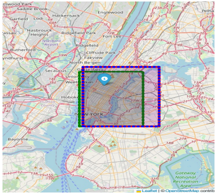
Prediction Region for NYC Taxi Drop-off Location.



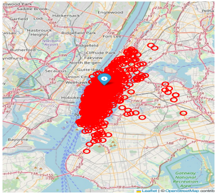
(a) KDE (No Coverage Guarantee)



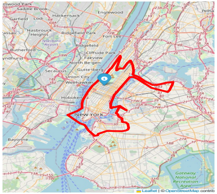
(b) NLE(R), RCP(B)



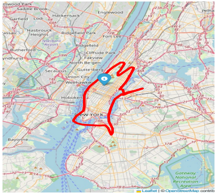
(c) MCQR(R), Dist-Split(G), $\mathrm{CQR}_{\mathrm{bon}}$(B)



(d) PCP



(e) **CONTRA**



(f) **ResCONTRA**

Split Conformal Prediction

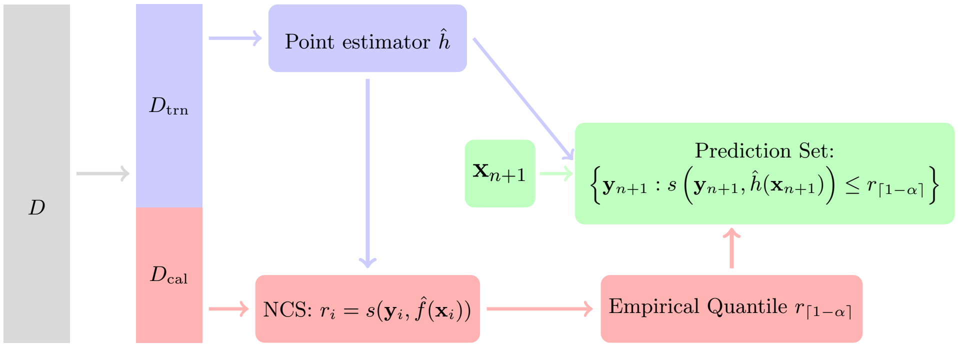


Figure 1: Split conformal prediction

A popular choice of non-conformity score(NCS) function is $s_i = |y_i - \hat{y}_i|$.

CONTRA

- Train $\hat{t}^{-1}$ on D_{trn} such that $\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)$ and $\{\hat{\mathbf{z}}_i\} \approx \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Obtain set $\{\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)\}, i \in D_{\text{cal}}$.
- Define non-conformity score $r_i = \|\hat{\mathbf{z}}_i\|$.
- Set threshold $r_{\lceil 1-\alpha \rceil}$ as the $(1-\alpha)$-empirical quantile of $\{r_i \mid i \in D_{\text{cal}}\}$.
- For new input $\mathbf{x}_{n+1}$, construct the prediction region:

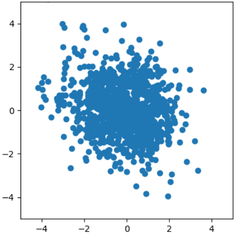
$$C(\mathbf{x}_{n+1}) = \hat{t}(\hat{E}, \mathbf{x}_{n+1}), \quad \text{where} \quad \hat{E} = \{\mathbf{z} : \|\mathbf{z}\| \leq r_{\lceil 1-\alpha \rceil}\}$$

Figure 2: Workflow of CONTRA.

CONTRA

- Train $\hat{t}^{-1}$ on D_{trn} such that $\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)$ and $\{\hat{\mathbf{z}}_i\} \approx \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Obtain set $\{\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)\}, i \in D_{\text{cal}}$.
- Define non-conformity score $r_i = \|\hat{\mathbf{z}}_i\|$.
- Set threshold $r_{\lceil 1-\alpha \rceil}$ as the $(1-\alpha)$-empirical quantile of $\{r_i \mid i \in D_{\text{cal}}\}$.
- For new input $\mathbf{x}_{n+1}$, construct the prediction region:

$$C(\mathbf{x}_{n+1}) = \hat{t}(\hat{E}, \mathbf{x}_{n+1}), \quad \text{where} \quad \hat{E} = \{\mathbf{z} : \|\mathbf{z}\| \leq r_{\lceil 1-\alpha \rceil}\}$$

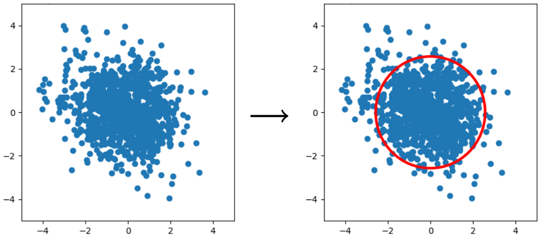


(a) $\{\hat{\mathbf{z}}_i\}, i \in D_{\text{cal}}$

CONTRA

- Train $\hat{t}^{-1}$ on D_{trn} such that $\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)$ and $\{\hat{\mathbf{z}}_i\} \approx \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Obtain set $\{\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)\}, i \in D_{\text{cal}}$.
- Define non-conformity score $r_i = \|\hat{\mathbf{z}}_i\|$.
- Set threshold $r_{\lceil 1-\alpha \rceil}$ as the $(1-\alpha)$-empirical quantile of $\{r_i \mid i \in D_{\text{cal}}\}$.
- For new input $\mathbf{x}_{n+1}$, construct the prediction region:

$$C(\mathbf{x}_{n+1}) = \hat{t}(\hat{E}, \mathbf{x}_{n+1}), \quad \text{where} \quad \hat{E} = \{\mathbf{z} : \|\mathbf{z}\| \leq r_{\lceil 1-\alpha \rceil}\}$$



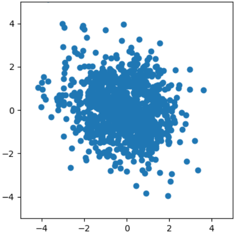
(a) $\{\hat{\mathbf{z}}_i\}, i \in D_{\text{cal}}$

(b) Draw $\hat{E}$

CONTRA

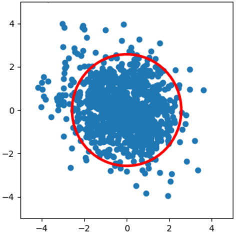
- Train $\hat{t}^{-1}$ on D_{trn} such that $\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)$ and $\{\hat{\mathbf{z}}_i\} \approx \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Obtain set $\{\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)\}, i \in D_{\text{cal}}$.
- Define non-conformity score $r_i = \|\hat{\mathbf{z}}_i\|$.
- Set threshold $r_{\lceil 1-\alpha \rceil}$ as the $(1-\alpha)$-empirical quantile of $\{r_i \mid i \in D_{\text{cal}}\}$.
- For new input $\mathbf{x}_{n+1}$, construct the prediction region:

$$C(\mathbf{x}_{n+1}) = \hat{t}(\hat{E}, \mathbf{x}_{n+1}), \quad \text{where} \quad \hat{E} = \{\mathbf{z} : \|\mathbf{z}\| \le r_{\lceil 1-\alpha \rceil}\}$$



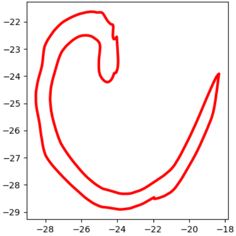
(a) $\{\hat{\mathbf{z}}_i\}, i \in D_{\text{cal}}$

$\longrightarrow$



(b) Draw $\hat{E}$

$\longrightarrow$

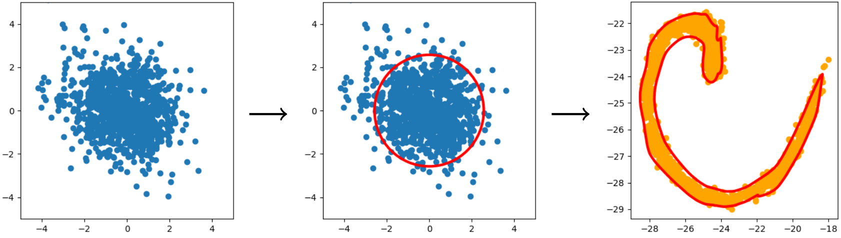


(c) $C(\mathbf{x}_{n+1})$ for new $\mathbf{x}_{n+1}$

CONTRA

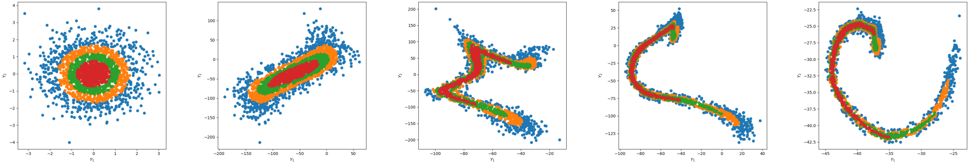
- Train $\hat{t}^{-1}$ on D_{trn} such that $\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)$ and $\{\hat{\mathbf{z}}_i\} \approx \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Obtain set $\{\hat{\mathbf{z}}_i = \hat{t}^{-1}(\mathbf{y}_i, \mathbf{x}_i)\}, i \in D_{\text{cal}}$.
- Define non-conformity score $r_i = \|\hat{\mathbf{z}}_i\|$.
- Set threshold $r_{\lceil 1-\alpha \rceil}$ as the $(1-\alpha)$-empirical quantile of $\{r_i \mid i \in D_{\text{cal}}\}$.
- For new input $\mathbf{x}_{n+1}$, construct the prediction region:

$$C(\mathbf{x}_{n+1}) = \hat{t}(\hat{E}, \mathbf{x}_{n+1}), \quad \text{where} \quad \hat{E} = \{\mathbf{z} : \|\mathbf{z}\| \leq r_{\lceil 1-\alpha \rceil}\}$$



(a) $\{\hat{\mathbf{z}}_i\}, i \in D_{\text{cal}}$ (b) Draw $\hat{E}$ (c) $C(\mathbf{x}_{n+1})$ with density points

Normalizing Flow



$\mathbf{z}$ | $\mathbf{z}^{(1)} = t_1(\mathbf{z})$ | $\mathbf{z}^{(2)} = t_2(\mathbf{z}^{(1)})$ | $\mathbf{z}^{(3)} = t_3(\mathbf{z}^{(2)})$ | $\mathbf{y} = t_4(\mathbf{z}^{(3)})$

Figure 3: Transformation process through sequential functions t_1, t_2, t_3, t_4 for given $\mathbf{x}$.

Composition

$$\mathbf{y} = t(\mathbf{z}) = t_k \circ t_{k-1} \circ \ldots \circ f_1(\mathbf{z})$$

ResCONTRA

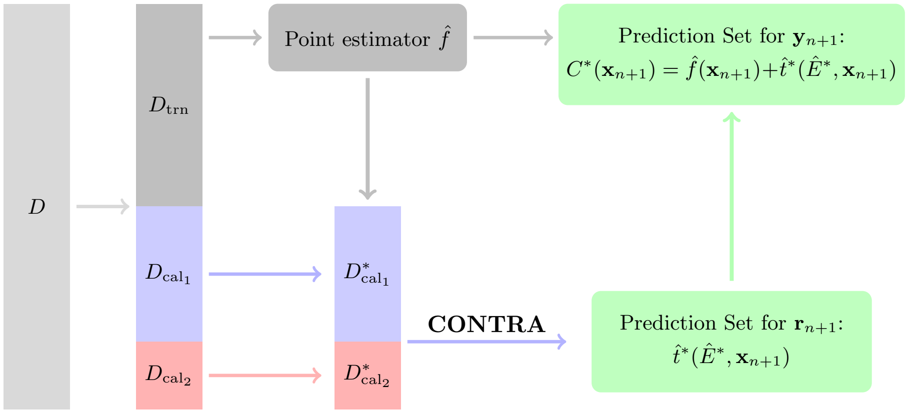


Figure 4: ResCONTRA

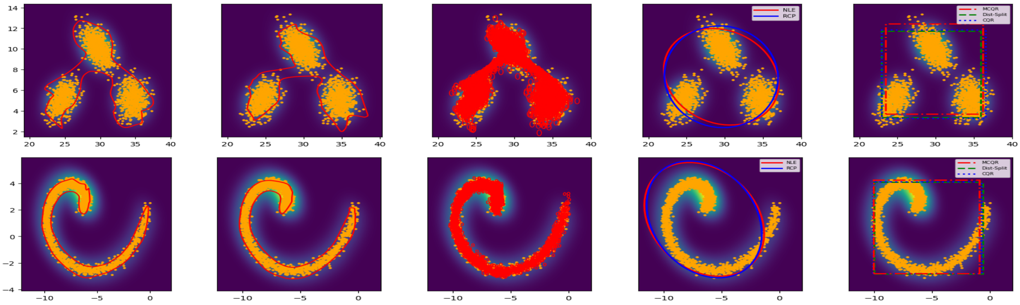
$\mathbf{r}_i = \mathbf{y}_i - \hat{f}(\mathbf{x}_i)$ and $D^* = \{(\mathbf{x}_i, \mathbf{r}_i)\}$

Synthetic data examples ($\alpha = 0.1$)

CONTRA provides compact prediction regions with flexible shapes and smooth boundaries.

	Metric	CONTRA	ResCONTRA	PCP	NLE	RCP	MCQR	Dist-split$_{\text{bon}}$	CQR$_{\text{bon}}$
Mixt.	Coverage	**0.91(0.003)**	0.90(0.003)	0.91(0.003)	0.89(0.003)	0.90(0.002)	0.89(0.003)	0.90(0.002)	0.90(0.002)
	Volume	**64.92(1.015)**	90.94(1.727)	71.32(0.845)	129.88(10.299)	135.12(12.378)	109.51(0.412)	112.38(0.575)	112.30(0.563)
Spiral	Coverage	0.91(0.003)	0.90(0.003)	**0.91(0.002)**	0.91(0.003)	0.91(0.002)	0.91(0.003)	0.91(0.003)	0.91(0.002)
	Volume	23.70(1.622)	41.73(2.275)	**23.22(1.154)**	68.97(0.917)	68.16(0.943)	64.69 (0.199)	65.82(0.157)	65.42(0.202)

Table 1: Average coverage and volume of prediction regions with $\alpha = 0.1$.



CONTRA ResCONTRA PCP Ellipses Rectangles

Figure 5: 90% Prediction regions of (y_1, y_2) given a specific $\mathbf{x}$ value from the test set.

Six Real Data Examples

CONTRA achieves the smallest prediction regions in four cases.

	Metric	CONTRA	ResCONTRA	PCP	NLE	RCP	MCQR	Dist-split$_{\text{bon}}$	CQR$_{\text{bon}}$
Bio	Coverage	0.90(0.002)	0.90(0.002)	**0.90(0.003)**	\	\	\	0.89(0.004)	0.90(0.002)
	Volume	13.23(0.383)	12.64(0.178)	**12.54(0.072)**	\	\	\	13.49(0.132)	12.55(0.076)
Taxi	Coverage	**0.89(0.002)**	0.89(0.002)	0.89(0.002)	0.90(0.001)	0.90(0.002)	0.89(0.002)	0.90(0.002)	0.91(0.002)
	Volume($\times 10^{-3}$)	**8.71(0.290)**	9.06(0.281)	8.95(0.132)	10.77(0.188)	12.21(0.262)	10.86(0.283)	14.65(0.378)	12.41(0.266)
Energy	Coverage	0.87(0.006)	0.87(0.009)	**0.88(0.006)**	0.86(0.006)	0.85(0.008)	0.84(0.007)	0.84(0.010)	0.87(0.008)
	Volume	18.24(1.269)	22.76(4.238)	**16.40(0.901)**	19.14(1.982)	26.04(2.888)	25.39(2.238)	32.12(2.383)	27.73(2.343)
2D RF	Coverage	**0.91(0.002)**	0.90(0.002)	0.91(0.002)	0.91(0.002)	0.90(0.002)	0.91(0.001)	0.92(0.001)	0.92(0.002)
	Volume	**5.29(0.074)**	10.34(0.207)	7.19(0.164)	15.50(0.307)	23.85(0.644)	11.92(0.185)	12.14(0.187)	12.30(0.190)
SCM20D	Coverage	**0.89(0.001)**	0.89(0.003)	0.89(0.002)	0.90(0.002)	0.89(0.002)	0.89(0.002)	0.90(0.003)	0.91(0.002)
	Volume($\times 10^{4}$)	**6.30(0.120)**	8.91(0.207)	6.97(0.154)	10.37(0.599)	6.52(0.153)	7.10(0.165)	8.82(0.231)	8.50(0.206)
4D RF	Coverage	**0.90 (0.003)**	0.89 (0.002)	0.89 (0.003)	0.89 (0.002)	0.90 (0.002)	0.89(0.002)	0.91(0.001)	0.92(0.002)
	Volume($\times 10^{2}$)	**0.59(0.043)**	1.54(0.098)	1.13 (0.057)	26.10(4.020)	52.93(11.403)	10.92(0.597)	20.40 (1.219)	12.68(0.665)

Table 2: Coverage and volume for multi-dimensional 90% conformal regions in six real datasets. Each table entry is the average of results over 20 random splits, with standard error in the parentheses. The method that achieved the smallest volume is in boldface.

Thank You !