Ellipsoidal conformal inference for Multi-Target Regression

Soundouss MESSOUDI, Sebastien DESTERCKE, Sylvain ROUSSEAU

COPA 2022 - Brighton, UK

Plan

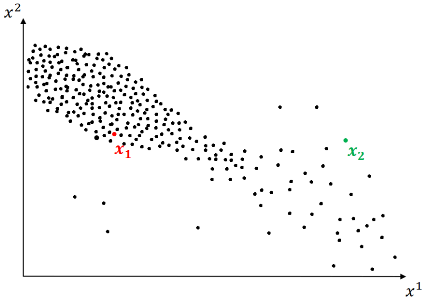
- Introduction
- Copula-based Conformal MTR
  - Hyper-rectangles for Conformal MTR
  - How to obtain a global validity using copulas ?
  - Results
- Ellipsoidal Conformal MTR
  - SGE
  - NLE
  - Local covariance matrix estimation
  - Overall approach
- Results
  - Results on synthetic data
  - Results on real data
- Conclusion

Introduction

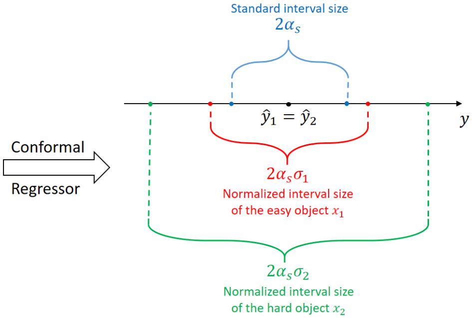
There are many papers on conformal prediction that provide NCMs for single-output regression.

Introduction

There are many papers on conformal prediction that provide NCMs for single-output regression.
σ_i can be used to get an individual interval size based on the difficulty to predict.



Visualisation of a 2D projection of the object space



Conformal interval sizes of two objects with the same regressor prediction

Introduction

How about when we have **more** than one target to predict **at once** ?

Introduction

How about when we have **more** than one target to predict **at once** ?
Within the MTR setting, each x_i is associated to multiple outputs such that y_i is a m-dimensional real valued target $y_i = (y_i^1, \ldots, y_i^m) \in Y$ with $Y = \mathbb{R}^m$.

Introduction

How about when we have **more** than one target to predict **at once** ?
Within the MTR setting, each x_i is associated to multiple outputs such that y_i is a m-dimensional real valued target $y_i = (y_i^1, \ldots, y_i^m) \in Y$ with $Y = \mathbb{R}^m$.
The conformal prediction method should thus include conformal regions for all targets at once, and provide a global validity guarantee as required by the user.

Introduction

How about when we have **more** than one target to predict **at once** ?
Within the MTR setting, each x_i is associated to multiple outputs such that y_i is a m-dimensional real valued target $y_i = (y_i^1, \ldots, y_i^m) \in Y$ with $Y = \mathbb{R}^m$.
The conformal prediction method should thus include conformal regions for all targets at once, and provide a global validity guarantee as required by the user.

Objectives

Propose a new flexible NCM that takes into consideration all m targets for MTR.
Obtain a global confidence level instead of an individual one for each target.

Plan

- Introduction
- Copula-based Conformal MTR
  - Hyper-rectangles for Conformal MTR
  - How to obtain a global validity using copulas ?
  - Results
- Ellipsoidal Conformal MTR
  - SGE
  - NLE
  - Local covariance matrix estimation
  - Overall approach
- Results
  - Results on synthetic data
  - Results on real data
- Conclusion

Copula-based Conformal MTR [1]

Hyper-rectangles for Conformal MTR

- Within the MTR setting, each x_i is associated to multiple outputs such that y_i is a m-dimensional real valued target $y_i = (y_i^1, \dots, y_i^m) \in Y$ with $Y = \mathbb{R}^m$.

Copula-based Conformal MTR [1]

Hyper-rectangles for Conformal MTR

- Within the MTR setting, each x_i is associated to multiple outputs such that y_i is a m-dimensional real valued target $y_i = (y_i^1, \ldots, y_i^m) \in Y$ with $Y = \mathbb{R}^m$.
- For a new instance x_{n+1} , let $\underline{\hat{y}}^j_{n+1}, \overline{\hat{y}}^j_{n+1}$ be respectively the lower and upper bounds of their individual interval predictions for each target y^j.

Copula-based Conformal MTR [1]

Hyper-rectangles for Conformal MTR

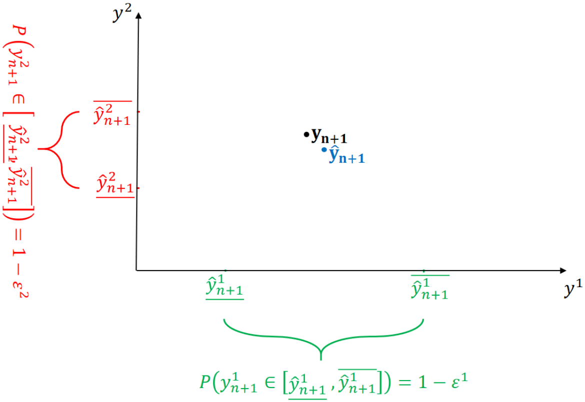
- Within the MTR setting, each x_i is associated to multiple outputs such that y_i is a m-dimensional real valued target $y_i = (y_i^1, \ldots, y_i^m) \in Y$ with $Y = \mathbb{R}^m$.
- For a new instance x_{n+1} , let $\underline{\hat{y}}^j_{n+1}, \overline{\hat{y}}^j_{n+1}$ be respectively the lower and upper bounds of their individual interval predictions for each target y^j.
- We define the hyper-rectangle $[\hat{y}_{n+1})]$, to which a global ground truth y_{n+1} of a new example x_{n+1} should belong to get a valid prediction, as :

$$[\hat{y}_{n+1})] = \times_{j=1}^{m} [\underline{\hat{y}}^j_{n+1}, \overline{\hat{y}}^j_{n+1}].$$

Copula-based Conformal MTR

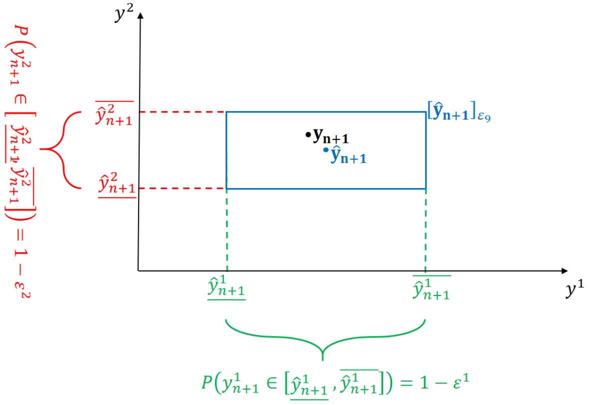
Hyper-rectangles for Conformal MTR

Suppose that $m = 2$



Copula-based Conformal MTR
Hyper-rectangles for Conformal MTR

Suppose that $m = 2$



Copula-based Conformal MTR

How to obtain a global validity using copulas ?

- Instead of choosing individual significance levels $\epsilon^1, \ldots, \epsilon^m$ for each target, we define a **global** significance level ϵ_g.
- For this global confidence level $1 - \epsilon_g$, we have

$$P(y_{n+1} \in [\hat{y}_{n+1}]) \geq 1 - \epsilon_g.$$

Copula-based Conformal MTR

How to obtain a global validity using copulas ?

- Instead of choosing individual significance levels $\epsilon^1, \ldots, \epsilon^m$ for each target, we define a **global** significance level ϵ_g.
- For this global confidence level $1 - \epsilon_g$, we have

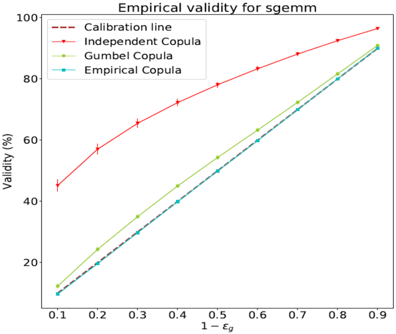
$$P(y_{n+1} \in [\hat{y}_{n+1}]) \geq 1 - \epsilon_g.$$

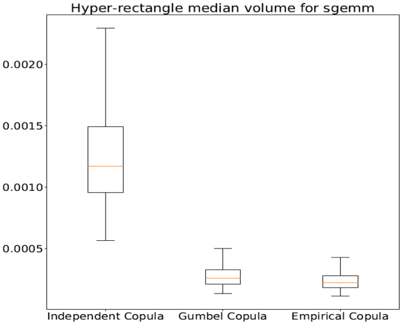
- We use **copulas** C to take advantage of the dependence structure between the m targets by giving ϵ_g in function of $\epsilon^1, \ldots, \epsilon^m$, following Sklar's theorem :

$$C(1 - \epsilon^1, \ldots, 1 - \epsilon^m) = 1 - \epsilon_g.$$

Copula-based Conformal MTR

Results on "sgemm" data set (copulas)





→ Results are great, but hyper-rectangles are not very flexible. We can use a better shape.

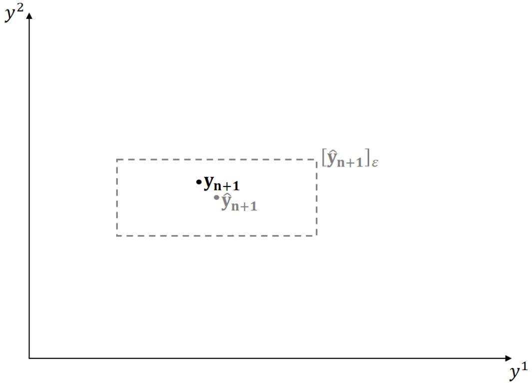
Plan

- Introduction
- Copula-based Conformal MTR
  - Hyper-rectangles for Conformal MTR
  - How to obtain a global validity using copulas ?
  - Results
- Ellipsoidal Conformal MTR
  - SGE
  - NLE
  - Local covariance matrix estimation
  - Overall approach
- Results
  - Results on synthetic data
  - Results on real data
- Conclusion

Ellipsoidal Conformal MTR

How about ellipsoids ?

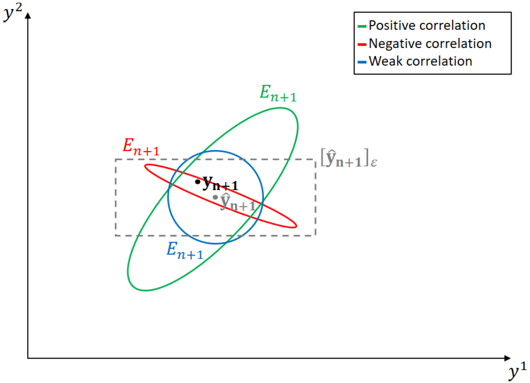
Suppose that $m = 2$



Ellipsoidal Conformal MTR

How about ellipsoids ?

Suppose that $m = 2$



Ellipsoidal Conformal MTR

Standard Global Ellipsoidal NCM (SGE)

Based on an ellipsoid shape, we can define a standard global ellipsoidal NCM [2] as follows :

$$\alpha_i = \sqrt{(y_i - \hat{y}_i)^T \hat{\Sigma}^{-1} (y_i - \hat{y}_i)},$$

where $y_i - \hat{y}_i$ is a vector and $\hat{\Sigma}^{-1}$ is the sample inverse-covariance matrix globally estimated from the training data's errors.

Ellipsoidal Conformal MTR

Standard Global Ellipsoidal NCM (SGE)

Based on an ellipsoid shape, we can define a standard global ellipsoidal NCM [2] as follows :

$$\alpha_i = \sqrt{(y_i - \hat{y}_i)^T \hat{\Sigma}^{-1} (y_i - \hat{y}_i)},$$

where $y_i - \hat{y}_i$ is a vector and $\hat{\Sigma}^{-1}$ is the sample inverse-covariance matrix globally estimated from the training data's errors.

$\rightarrow$ Standard approach, with the non-conformity score not tailored for each instance.

Ellipsoidal Conformal MTR

Normalized Local Ellipsoidal NCM (NLE)

Theorem

The ellipsoid E_i given by the NCM

$$\alpha_i = \sqrt{(y_i - \hat{y}_i)^T \hat{\Sigma}_i^{-1} (y_i - \hat{y}_i)},$$

which center is the regressor's prediction $\hat{y}_i$, and covariance matrix is $\frac{\hat{\Sigma}_i^{-1}}{\alpha_s^2}$, is a conformal valid prediction.

Ellipsoidal Conformal MTR
Normalized Local Ellipsoidal NCM (NLE)

Theorem

The ellipsoid E_i given by the NCM

$$\alpha_i = \sqrt{(y_i - \hat{y}_i)^T \hat{\Sigma}_i^{-1} (y_i - \hat{y}_i)},$$

which center is the regressor's prediction $\hat{y}_i$, and covariance matrix is $\frac{\hat{\Sigma}_i^{-1}}{\alpha_s^2}$, is a conformal valid prediction.

The volume of E_i, its predictive efficiency, is equal to the standard volume of an ellipsoid : $Vol(E_i) = \alpha_s^m \det\left(\hat{\Sigma}_i\right)^{1/2} Vol(B_m)$, where $B_m = \{y \in \mathbb{R}^m : ||y||_2 \leq 1\}$ is the unit ball.

Ellipsoidal Conformal MTR
Local covariance matrix estimation

$$\hat{\Sigma}_i = \lambda \hat{Cov}_i + (1 - \lambda)\hat{\Sigma}$$

$\hat{\Sigma}$: the global covariance matrix estimated from error rates over all training instances in $\mathbf{Z}^{\text{tr}}$,
$\hat{Cov}_i$: the local covariance matrix of instance x_i,
λ : a parameter to control the trade-off between $\hat{\Sigma}$ and $\hat{Cov}_i$.

Ellipsoidal Conformal MTR

Local covariance matrix estimation

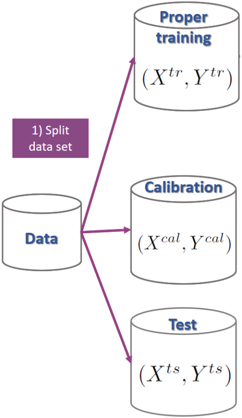
$$\hat{\Sigma}_i = \lambda \hat{Cov}_i + (1 - \lambda)\hat{\Sigma}$$

$\hat{\Sigma}$: the global covariance matrix estimated from error rates over all training instances in $\mathbf{Z}^{\text{tr}}$,
$\hat{Cov}_i$: the local covariance matrix of instance x_i,
λ : a parameter to control the trade-off between $\hat{\Sigma}$ and $\hat{Cov}_i$.

$\rightarrow$ To estimate $\hat{Cov}_i$, we use a kNN model to get the k nearest instances $x_j \in \mathbf{Z}^{\text{tr}}$ from x_i, and estimate $\hat{Cov}_i$ from the observed errors $(y_j - \hat{y}_j)$ for instances x_j.

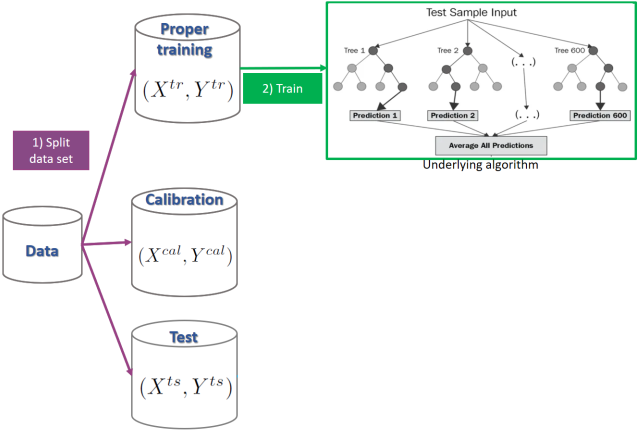
Ellipsoidal Conformal MTR

Overall approach

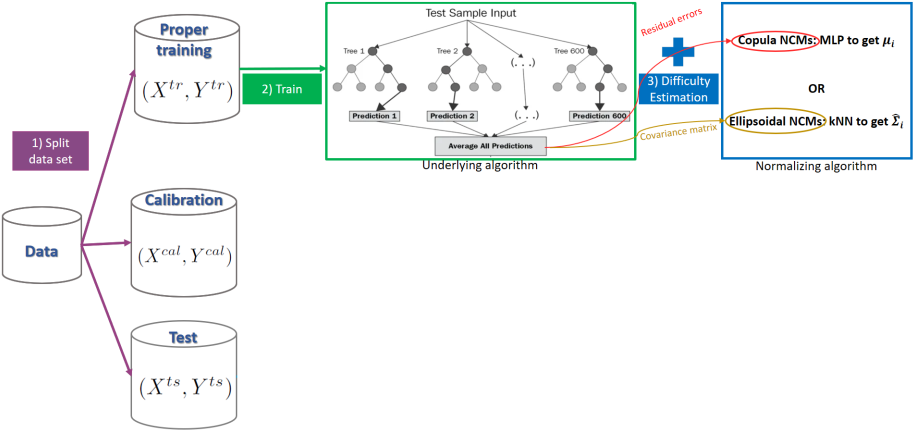


Ellipsoidal Conformal MTR

Overall approach

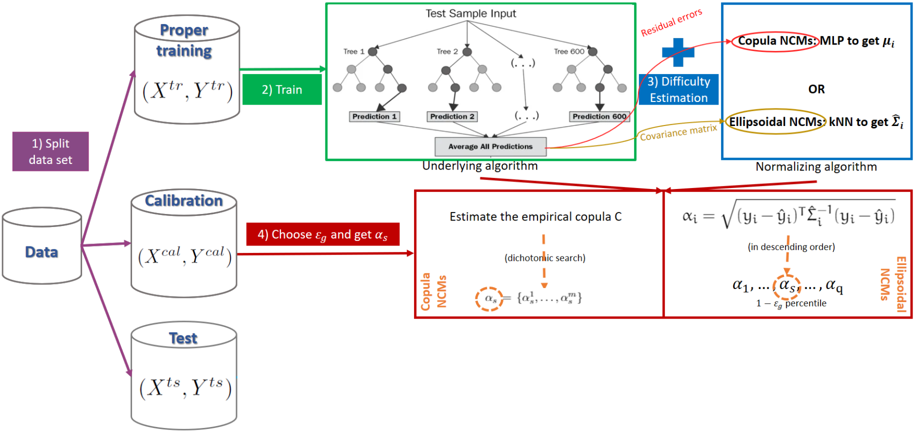


Ellipsoidal Conformal MTR
Overall approach



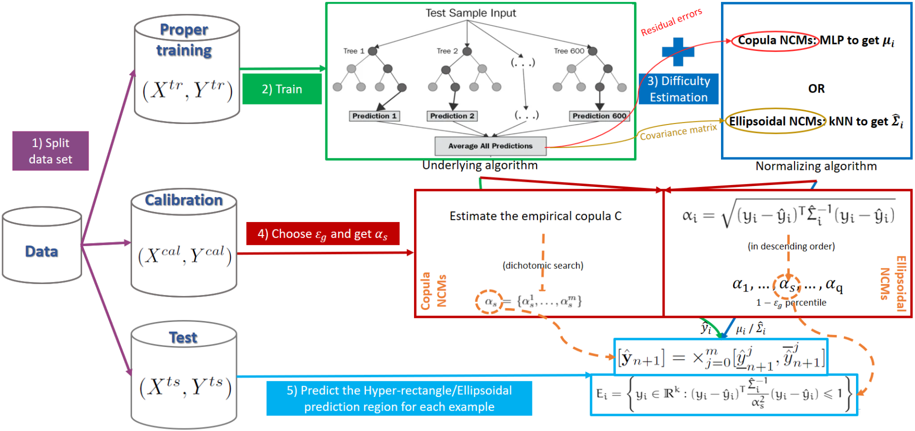
Ellipsoidal Conformal MTR

Overall approach



Ellipsoidal Conformal MTR

Overall approach



Plan

- Introduction
- Copula-based Conformal MTR
  - Hyper-rectangles for Conformal MTR
  - How to obtain a global validity using copulas ?
  - Results
- Ellipsoidal Conformal MTR
  - SGE
  - NLE
  - Local covariance matrix estimation
  - Overall approach
- Results
  - Results on synthetic data
  - Results on real data
- Conclusion

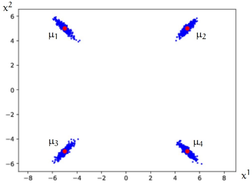
Results

Synthetic data set

The synthetic data set aims at assessing the performance of all NCMs by controlling the dependence structure between 2-dimensional outputs with :

$$y = Ax + \varepsilon \quad \text{with } A = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix} \text{ and } \varepsilon \sim \mathcal{N}(0, S_x).$$

with $S_x = \sum_{i=1}^{4} \frac{\Delta(x,\mu_i)}{\sum_{j=1}^{4} \Delta(x,\mu_j)} \times Cov_{\mu_i}$



Results

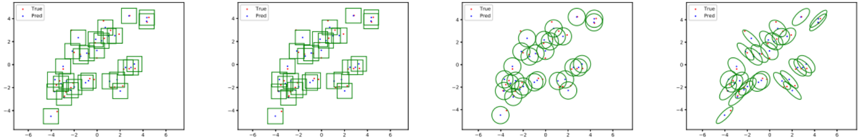
Results on synthetic data

synthetic (k = 2)		$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.15$	$\epsilon = 0.2$
Validity	SEC	99.21 ± 0.12	95.11 ± 0.23	90.25 ± 0.35	85.07 ± 0.45	80.07 ± 0.79
	NEC	99.24 ± 0.11	95.08 ± 0.31	90.26 ± 0.39	84.99 ± 0.41	80.12 ± 0.81
	SGE	98.97 ± 0.17	95.02 ± 0.37	90.00 ± 0.50	84.95 ± 0.49	79.94 ± 0.73
	NLE - Ours	99.01 ± 0.15	94.89 ± 0.28	90.02 ± 0.48	84.91 ± 0.33	80.00 ± 0.45
Efficiency	SEC	3.95 ± 0.13	2.32 ± 0.04	1.75 ± 0.03	1.39 ± 0.02	1.16 ± 0.02
	NEC	3.99 ± 0.14	2.32 ± 0.04	1.76 ± 0.03	1.38 ± 0.02	1.16 ± 0.02
	SGE	4.19 ± 0.17	2.51 ± 0.05	1.84 ± 0.04	1.46 ± 0.02	1.20 ± 0.02
	NLE - Ours	$\mathbf{2.85} \pm \mathbf{0.07}$	$\mathbf{1.81} \pm \mathbf{0.02}$	$\mathbf{1.39} \pm \mathbf{0.02}$	$\mathbf{1.14} \pm \mathbf{0.01}$	$\mathbf{0.97} \pm \mathbf{0.01}$

TABLE – Validity and efficiency results for synthetic data.

Results

Results’ visualization on synthetic data



(a) Standard Empirical Copula (b) Normalized Empirical Copula (c) Standard Global Ellipse (d) Normalized Local Ellipse

Results

Real data sets

Names	Instances	Features	Targets
residential building	372	105	2
enb	768	8	2
music origin	1059	68	2
bias correction	7750	25	2
jura	359	15	3
scpf	1137	23	3
indoor localization	21049	520	3
sgemm	241600	14	4
atp1d	337	411	6
atp7d	296	411	6
rf1	9125	64	8
rf2	9125	576	8
osales	639	413	12
wq	1060	16	14
scm1d	9803	280	16
scm20d	8966	61	16
oes10	403	298	16
oes97	334	263	16
community crime	2215	125	18

Results

Results on real data

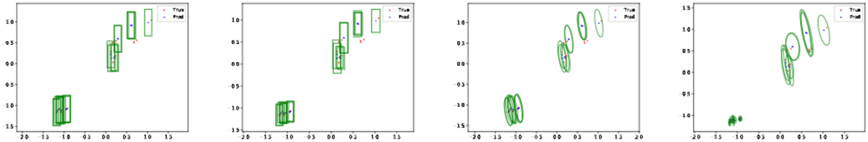
2*$\epsilon = 0.1$	Validity				Efficiency			
	SEC	NEC	SGE	NLE - Ours	SEC	NEC	SGE	NLE - Ours
res building	83.59 ± 6.73	85.48 ± 4.02	85.75 ± 5.08	89.54 ± 5.11	0.30 ± 0.07	0.22 ± 0.06	0.31 ± 0.08	**0.17 ± 0.05**
enb	83.98 ± 6.37	84.89 ± 5.99	87.23 ± 4.15	89.57 ± 5.22	0.13 ± 0.03	0.08 ± 0.02	0.11 ± 0.02	**0.04 ± 0.02**
music origin	88.76 ± 3.59	88.66 ± 2.64	89.70 ± 4.80	89.05 ± 4.30	10.98 ± 1.11	16.54 ± 3.30	10.60 ± 1.41	**9.55 ± 0.73**
bias corr	89.98 ± 0.93	90.42 ± 1.32	90.24 ± 1.06	90.29 ± 1.48	1.47 ± 0.06	1.32 ± 0.05	1.37 ± 0.05	**1.19 ± 0.07**
jura	86.93 ± 6.20	85.81 ± 4.69	86.64 ± 5.22	88.30 ± 8.03	24.29 ± 14.14	12.79 ± 7.57	12.89 ± 4.26	**10.17 ± 5.60**
scpf	84.52 ± 5.18	84.26 ± 4.95	87.86 ± 5.02	87.87 ± 4.75	$3.77^{10} \pm 6.83^{10}$	$1.26^{10} \pm 2.81^{10}$	$4.87^{7} \pm 8.71^{6}$	**69.59 ± 89.96**
indoor loc	90.32 ± 0.48	90.32 ± 0.58	90.37 ± 0.96	90.11 ± 0.89	0.06 ± 0.01	**0.05 ± 0.01**	0.07 ± 0.01	0.29 ± 0.03
sgemm	90.04 ± 0.17	90.05 ± 0.27	89.98 ± 0.20	90.03 ± 0.17	$8.45^{-5} \pm 2.56^{-6}$	$7.34^{-5} \pm 3.15^{-6}$	$1.84^{-5} \pm 4.93^{-7}$	**$1.15^{-5} \pm 3.41^{-7}$**
atp1d	72.09 ± 11.19	66.74 ± 10.73	85.18 ± 7.08	85.78 ± 5.50	6.25 ± 3.47	1.02 ± 1.15	8.17 ± 5.63	**0.47 ± 0.45**
atp7d	72.29 ± 11.82	68.54 ± 12.96	81.82 ± 10.41	86.13 ± 10.83	8.07 ± 10.73	**0.64 ± 0.35**	6.84 ± 9.23	4.11 ± 7.57
rf1	89.10 ± 2.14	89.13 ± 1.99	90.04 ± 1.50	90.20 ± 1.53	$5.75^{-7} \pm 3.79^{-7}$	$3.60^{-7} \pm 2.16^{-7}$	$6.20^{-7} \pm 5.27^{-7}$	**$4.14^{-8} \pm 3.34^{-8}$**
rf2	90.18 ± 1.53	89.79 ± 1.76	90.26 ± 1.31	89.98 ± 1.36	$7.50^{-7} \pm 4.40^{-7}$	$3.13^{-7} \pm 2.38^{-7}$	$6.49^{-7} \pm 4.96^{-7}$	**$4.44^{-8} \pm 2.41^{-8}$**
osales	81.39 ± 7.02	78.86 ± 7.88	86.71 ± 3.95	88.12 ± 4.64	$5.60^{6} \pm 9.55^{6}$	$1.20^{15} \pm 3.45^{15}$	$1.47^{5} \pm 2.46^{5}$	**$8.08^{4} \pm 1.26^{5}$**
wq	72.74 ± 4.43	78.49 ± 2.76	89.15 ± 4.91	87.55 ± 3.91	$1.31^{10} \pm 6.27^{9}$	$3.93^{17} \pm 1.18^{18}$	**$1.16^{9} \pm 7.06^{8}$**	$1.35^{9} \pm 1.08^{9}$
scm1d	89.70 ± 1.27	89.60 ± 1.61	90.07 ± 1.27	90.42 ± 1.25	$6.86^{4} \pm 2.92^{4}$	$1.12^{3} \pm 6.22^{2}$	$4.21^{2} \pm 1.79^{2}$	**8.86 ± 3.42**
scm20d	88.23 ± 1.14	88.72 ± 1.64	89.26 ± 1.02	89.45 ± 1.15	$7.98^{5} \pm 6.57^{5}$	$7.02^{4} \pm 3.85^{4}$	$3.21^{3} \pm 2.27^{3}$	**$5.52^{2} \pm 3.54^{2}$**
oes10	74.15 ± 13.92	66.13 ± 19.95	87.57 ± 8.13	88.58 ± 6.84	$7.62^{5} \pm 2.23^{6}$	$1.05^{5} \pm 2.86^{5}$	$1.83^{6} \pm 4.83^{6}$	**$8.25^{3} \pm 1.26^{4}$**
oes97	69.11 ± 8.18	62.94 ± 15.06	89.22 ± 6.47	87.99 ± 7.41	**$1.45^{5} \pm 4.16^{5}$**	$3.86^{6} \pm 8.38^{6}$	$6.90^{6} \pm 1.53^{7}$	$1.18^{7} \pm 1.82^{7}$
com crime	86.18 ± 3.51	84.61 ± 3.04	90.21 ± 3.66	89.03 ± 2.34	$5.19^{9} \pm 9.74^{9}$	$7.05^{4} \pm 7.61^{4}$	$1.17^{5} \pm 1.49^{5}$	**2.80 ± 7.01**

We note X^Y the value $X \times 10^Y$.

TABLE – Validity & efficiency results. *For validity, in red are mean values lower than 80% and in orange are mean values between 80% and 85%.*

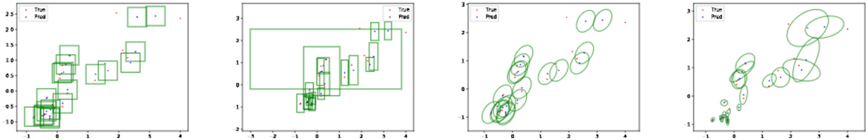
Results

Results' visualization on real data



(a) Standard Empirical Copula (b) Normalized Empirical Copula (c) Standard Global Ellipse (d) Normalized Local Ellipse

Results' visualization for the data set "enb" with $\epsilon = 0.1$.



(a) Standard Empirical Copula (b) Normalized Empirical Copula (c) Standard Global Ellipse (d) Normalized Local Ellipse

Results' visualization for the data set "residential building" with $\epsilon = 0.1$.

Plan

- Introduction
- Copula-based Conformal MTR
  - Hyper-rectangles for Conformal MTR
  - How to obtain a global validity using copulas ?
  - Results
- Ellipsoidal Conformal MTR
  - SGE
  - NLE
  - Local covariance matrix estimation
  - Overall approach
- Results
  - Results on synthetic data
  - Results on real data
- Conclusion

Conclusion :

- New non-conformity measures for multi-target regression with a more flexible ellipsoidal shape based on the local covariance matrix of the instance.
- Tighter volumes compared to other NCMs while maintaining a validity defined by the required global confidence level.

Conclusion :

- New non-conformity measures for multi-target regression with a more flexible ellipsoidal shape based on the local covariance matrix of the instance.
- Tighter volumes compared to other NCMs while maintaining a validity defined by the required global confidence level.

Perspectives :

- Look for other ways to estimate the local covariance matrix (density estimation,. . .).
- Adapt our approach to other methods such as jack-knife+ [3].
- Work on other Multi-Task Learning problems.

Thank you for your attention !
Questions ?

[1] Messoudi, S., Destercke, S. and Rousseau, S., 2021. Copula-based conformal prediction for multi-target regression. Pattern Recognition, 120, p.108101.

[2] Johnstone, C. and Cox, B., 2021. Conformal uncertainty sets for robust optimization. In Conformal and Probabilistic Prediction and Applications pp. 72-90. PMLR.

[3] Barber, R.F., Candes, E.J., Ramdas, A. and Tibshirani, R.J., 2021. Predictive inference with the jackknife+. The Annals of Statistics, 49(1), pp.486-507.