Ellipsoidal conformal inference for Multi-Target Regression

Soundouss Messoudi SOUNDOUSS.MESSOUDI@HDS.UTC.FR
Sébastien Destercke SEBASTIEN.DESTERCKE@HDS.UTC.FR
Sylvain Rousseau SYLVAIN.ROUSSEAU@HDS.UTC.FR
HEUDIASYC - UMR CNRS 7253, Université de Technologie de Compiègne, 57 avenue de Landshut, 60203 COMPIEGNE CEDEX - FRANCE

Editor: Ulf Johansson, Henrik Bostrm, Khuong An Nguyen, Zhiyuan Luo and Lars Carlsson

Abstract

Quantifying the uncertainty of a predictive model output is of essential importance in learning scenarios involving critical applications. As the learning task becomes more complex, so does uncertainty quantification. In this paper, we consider the task of multi-target regression and propose a method to output ellipsoidal confidence regions whose shapes are tailored to each instance to predict. We also guarantee that those confidence regions are well-calibrated, i.e., that they cover the ground truth with a specified probability. To achieve such a feat, we propose a conformal prediction method outputting ellipsoidal prediction regions. Experiments on both simulated and real-world data sets show that our methods outperform existing ones.

Keywords: Inductive conformal prediction, Ellipsoids, Multi-target regression.

1. Introduction

Uncertainty quantification is an important aspect of Machine Learning. It can be as crucial as the prediction when it comes to sensitive tasks such as when the safety of humans, such as pedestrians and drivers in an autonomous vehicle application (Pan et al. (2020)) or patients in a medical case (Pereira et al. (2020)), is at risk. In these situations, uncertainty estimates can help in the decision-making process by identifying instances where an expert's opinion is better than a wrongly over-confident model's prediction. Having multiple outputs to predict at once, especially when correlated, makes the estimation even tougher. Some examples of multi-target regression problems are predicting the financial cost and the price of a house (Rafiei and Adeli (2016)), or predicting latitude, longitude and altitude in an indoor localization problem (Torres-Sospedra et al. (2014)).

One of the methods that generates uncertainty estimates for any Machine Learning model is conformal prediction (Vovk et al. (2005)), a distribution-free predictive inference that has gained popularity in the recent years. Instead of point predictions, this theoretically-proven framework produces a set (in classification) or an interval (in regression) as predictions with a statistical guarantee based on a probability error chosen by the user.

Most research studies on conformal prediction for regression problems focus on single-output problems. However, in the last few years, some papers working with a multivariate output space have emerged. Neeven and Smirnov (2018) extend a conformal single-output k-nearest neighbor to a multi-target weather forecasting problem. Kuleshov et al. (2018a) provide a theoretical conformal prediction framework for manifold learning. Our previous work Messoudi et al. (2021) produces a

hyper-rectangle conformal prediction region fitted to a multi-target regression problem, and Johnstone and Cox (2021) exploits an ellipsoidal conformal prediction region for robust optimization.

In this paper, we propose a new conformal prediction method for multi-target regression that provides k-dimensional ellipsoidal uncertainty regions specific to each instance. Section 2 gives an overview of our setting by introducing multi-target regression and inductive conformal regression. Section 3 presents our method of ellipsoidal conformal prediction applied to multi-target regression, and prove that it is theoretically well-calibrated. Section 4 presents the results of experiments that show that our method provide tighter prediction sets, while maintaining the desired probability error.

2. Setting

In this section, we present the setting of our work by briefly defining the multi-target regression problem, reviewing frequency-calibrated prediction methods, before introducing inductive conformal regression, the basic framework used in our research.

2.1. Multi-target regression (MTR)

Let X be the feature space and Y be the target space. The training set is constructed by drawing n i.i.d. pairs (x_i, y_i) from a probability distribution on $X \times Y$, where each instance $x_i \in X$ is associated with a target y_i. In a single target regression setting, we have one real-valued output $y_i \in Y = \mathbb{R}$. In a multi-target regression setting, each x_i is associated to multiple outputs such that y_i is a k-dimensional real valued target $y_i = (y_i^1, \ldots, y_i^k) \in Y$ with $Y = \mathbb{R}^k$. The objective of multi-target regression is to learn a function $H : X \to Y$, i.e. to predict multiple outputs based on the input features characterizing the data set, which is a straightforward generalization of single-output supervised learning.

2.2. Frequency-calibrated predictions

When learning a model H that concerns some critical application such as medical diagnosis or reliability analysis, it is common to ask the predictions of H to come with an uncertainty quantification that is well-calibrated.

This calibration can be expressed as the requirement that a probabilistic prediction $\hat{p}(y)$ of H should converge to the true probability $p(y)$. This is often done through the use of a post-calibration procedure (Kuleshov et al. (2018b)) or by the use of proper loss functions (Gneiting and Raftery (2007)).

Another popular means of quantifying uncertainty in a probabilistic setting is to perform a Bayesian analysis (Kendall and Gal (2017)), yet such analysis does not come with statistical guarantees.

One more way to ensure calibration is to accept set-valued predictions from H, typically in the form of intervals $[\underline{y}, \overline{y}]$ for univariate regression, that have a guaranteed coverage probability. More precisely, if $H(x) = \hat{Y}$ denotes a set-valued prediction and y the observed quantity, we require that:

$$\mathbb{P}(y \in \hat{Y}) \geq 1 - \epsilon$$

with a specified error ϵ. This can be achieved by using conformal inference, a flexible framework that can be added to any Machine Learning algorithm to produce well-calibrated predictions, and

that only requires exchangeable data (a weaker assumption than i.i.d.). The next subsection will explain this method for single-output regression in the inductive setting.

2.3. Inductive conformal regression

In a single-output regression problem, conformal prediction provides a prediction interval instead of a point prediction, allowing for the uncertainty to be quantified depending on the interval size. This prediction interval can be obtained using a transductive setting (Vovk et al. (2005)), where all instances are used for each new instance to get its prediction, or an inductive setting (Papadopoulos and Haralambous (2011)), where the Machine Learning model is trained once on a part of the training data with the other part used as calibration for conformal inference.

Let $\mathbf{Z} = \{(x_i, y_i),\ i = 1, \ldots, n\}$ be a sequence of exchangeable pairs for our regression problem where $|\mathbf{Z}| = n$. Each $x_i \in \mathbf{X}$ is an object and $y_i \in \mathbf{Y}$ its corresponding real-valued label. The inductive conformal regression steps to obtain a prediction interval are as follows:

1. Divide $\mathbf{Z}$ into two disjoint sets: a *proper training* set $\mathbf{Z}^{\text{tr}}$ with $|\mathbf{Z}^{\text{tr}}| = l$ and a *calibration* set $\mathbf{Z}^{\text{cal}}$ with $|\mathbf{Z}^{\text{cal}}| = n - l = q$.

2. Train a regression model H on $\mathbf{Z}^{\text{tr}}$ to predict y and set a *non-conformity measure (NCM)* $f(z) = |y - \hat{y}|$ with $\hat{y} = H(x)$ that evaluates how much an instance $z = (x, y)$ is strange compared to those in $\mathbf{Z}^{\text{tr}}$.

3. Apply the non-conformity measure $f(z)$ to each element $z_i \in \mathbf{Z}^{\text{cal}}$ to obtain the calibration scores $\alpha_{l+1}, \ldots, \alpha_n$, with $\alpha_i = f(z_i) = |y_i - \hat{y}_i|$.

4. Set a *significance level* $\epsilon \in (0, 1)$, sort calibration scores $\alpha_{l+1}, \ldots, \alpha_n$ in a descending order and get the index of the $(1 - \epsilon)$-percentile of the non-conformity score α_s such that $\mathbb{P}(|y_i - \hat{y}_i| \leq \alpha_s) \geq 1 - \epsilon$. Thus,
$$\alpha_s := \alpha_{\lceil (1-\epsilon)\cdot q \rceil}$$

5. For a new instance x_{n+1}, get its prediction interval
$$\hat{\mathbf{y}}_{n+1} = [\hat{y}_{n+1} - \alpha_s, \hat{y}_{n+1} + \alpha_s]$$

By using the NCM $\alpha_i = |y_i - \hat{y}_i|$, all prediction intervals for all instances have the same size $2\alpha_s$. Instead of using this standard approach, a *normalized NCM* was proposed to get personalized prediction intervals depending on the models difficulty of outputting an instances prediction Papadopoulos and Haralambous (2011). Hence, if an instance is easy to predict, its prediction interval should be smaller than another hard instance. This difficulty estimation σ_i is then used to scale the standard NCM such as $\alpha_i = |y_i - \hat{y}_i|/\sigma_i$. In this case, the prediction interval becomes
$$\hat{\mathbf{y}}_{n+1} = [\hat{y}_{n+1} - \alpha_s\sigma_{n+1}, \hat{y}_{n+1} + \alpha_s\sigma_{n+1}]$$

The two desirable properties from a conformal prediction method are *validity*, meaning that the confidence level always satisfies the probability error ϵ chosen by the user, and *efficiency*, i.e., the prediction sets/intervals are the smallest possible in its predictive sense. These two notions will be used to assess the non-conformity measure performances. For other criteria of efficiency, the reader can refer to Vovk et al. (2017).

3. Ellipsoidal multivariate conformal regression

In a multi-target regression problem, it is possible that the outputs share some information between them and therefore, these outputs are likely correlated. Our objective is to exploit this relationship to get a generalized conformal inference model with boosted performance results (Caruana (1993)).

Our goal here is to propose a method that is both efficient, localized and offers some flexibility to capture output dependencies. None of the methods we cited have these three features. Using and learning manifold (Kuleshov et al. (2018a)) is quite complex, hyper-rectangles cannot capture covariance information, and (Johnstone and Cox (2021)) consider one global ellipsoid and do not propose a learning method to get the covariance matrix, as they work in a robust optimization setting, not a learning one. Nevertheless, as their method can be readily applied to learning and compared to ours, we will recall it in detail.

3.1. Standard global ellipsoidal non-conformity measure

Knowing that the ellipsoid[1] is one of the most flexible geometrical shapes, Johnstone and Cox (2021) proposed for a robust optimization problem an ellipsoidal NCM:

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

where $y_i - \hat{y}_i$ is a vector of univariate non-conformity scores (here, the regressor's error rate) and $\hat{\Sigma}^{-1}$ is the sample inverse-covariance matrix of the observed errors.

3.2. Normalized local ellipsoidal non-conformity measure

The standard global ellipsoidal NCM in (1) uses a sample inverse-covariance matrix $\hat{\Sigma}^{-1}$ globally-estimated from the training data's errors. Therefore, it resembles the standard approach in the univariate case where the non-conformity score is not tailored for each instance. We propose to use instead a normalized inverse-covariance matrix $\hat{\Sigma}_i^{-1}$ for each instance x_i, to take into consideration a locally-estimated covariance matrix of the instance. The normalized NCM in this case is:

$$\alpha_i = \sqrt{(y_i - \hat{y}_i)^T \hat{\Sigma}_i^{-1} (y_i - \hat{y}_i)} \tag{2}$$

This NCM enables us to define the conformal prediction region as an ellipsoid E_i given by the following theorem:

Theorem 1 *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.

1. Note that the term "ellipsoid" used in this paper refers to a k-dimensional ellipsoid.

Proof Let E_i be the ellipsoid to which the ground truth y_i of an instance x_i should belong. To satisfy the validity of a conformal predictor considering a significance level ϵ, we have:

$$\begin{aligned}
&\mathbb{P}(y_i \in E_i) \geq 1-\epsilon \\
&\iff \mathbb{P}(\alpha_i \leq \alpha_s) \geq 1-\epsilon. \\
&\iff \mathbb{P}\left(\sqrt{(y_i-\hat{y}_i)^T\hat{\Sigma}_i^{-1}(y_i-\hat{y}_i)} \leq \alpha_s\right) \geq 1-\epsilon. \\
&\iff \mathbb{P}\left(\sqrt{(y_i-\hat{y}_i)^T\frac{\hat{\Sigma}_i^{-1}}{\alpha_s^2}(y_i-\hat{y}_i)} \leq 1\right) \geq 1-\epsilon.
\end{aligned}$$

Following an ellipsoid's definition, we can see that:

$$E_i = \left\{ y_i \in \mathbb{R}^k : (y_i-\hat{y}_i)^T\frac{\hat{\Sigma}_i^{-1}}{\alpha_s^2}(y_i-\hat{y}_i) \leq 1 \right\}.$$

E_i is the ellipsoid given by the center $\hat{y}_i$ and the matrix $\frac{\hat{\Sigma}_i^{-1}}{\alpha_s^2}$. ■

The volume of E_i, its predictive efficiency, is equal to the standard volume of an ellipsoid:

$$\text{Vol}(E_i) = \alpha_s^k \det\left(\hat{\Sigma}_i\right)^{1/2} \text{Vol}(B_k), \tag{3}$$

where $B_k = \{y \in \mathbb{R}^k : ||y||_2 \leq 1\}$ is the unit ball.

3.3. Local covariance matrix estimation

As mentioned above, the NCM given by (2) is mainly composed of a normalized inverse-covariance matrix $\hat{\Sigma}_i^{-1}$ for each instance x_i, meaning that we need to estimate locally a covariance matrix $\hat{\text{Cov}}_i$ for x_i. Thus, we propose to calculate $\hat{\Sigma}_i$ by:

$$\hat{\Sigma}_i = \lambda\hat{\text{Cov}}_i + (1-\lambda)\hat{\Sigma}, \tag{4}$$

where $\hat{\Sigma}$ is the global covariance matrix estimated from error rates over all training instances in $\mathbf{Z}^{\text{tr}}$, that is from the observed errors $(y_i - \hat{y}_i)$, $\hat{\text{Cov}}_i$ is the local covariance matrix of instance x_i, and λ is a parameter to control the trade-off between $\hat{\Sigma}$ and $\hat{\text{Cov}}_i$ in order to get a positive definite matrix $\hat{\Sigma}_i$. Its value should be high enough to ensure that the influence of $\hat{\text{Cov}}_i$ is predominant compared to $\hat{\Sigma}$.

To estimate $\hat{\text{Cov}}_i$, we propose to use a kNN model to get the neighboring instances of x_i from $\mathbf{Z}^{\text{tr}}$, and then use the observed error rates on these instances to estimate $\hat{\text{Cov}}_i$. That is, given x_i, we consider the k nearest instances $x_j \in \mathbf{Z}^{\text{tr}}$ from x_i, and estimate $\hat{\text{Cov}}_i$ from the observed errors $(y_j - \hat{y}_j)$ for instances x_j.

4. Experiments

In this section, we present our experimental approach, conduct it on synthetic data and benchmark data sets and analyze the obtained results. The experiments were conducted using Python and the code is available in Github [2].

2. https://github.com/M-Soundouss/EllipsoidalConformalMTR

4.1. Experimental protocol

The objective of our experiments is to compare between four non-conformity measures:

- **Standard Empirical Copula (SEC)** Messoudi et al. (2021): produces same-size hyper-rectangle conformal regions by non-parametrically estimating the dependence structure between calibration non-conformity scores α_i using the empirical copula. Although not present in our previous paper, we can easily derive it by considering the standard NCM $\alpha_i = |y_i - \hat{y}_i|$.
- **Normalized Empirical Copula (NEC)** Messoudi et al. (2021): generates personalized hyper-rectangle conformal regions with a difficulty estimator $\sigma_i = \exp(\mu_i) + \beta$, where a Multi-Layer Perceptron (MLP) is used to learn $\mu_i = \ln(|y_i - \hat{y}_i|)$.
- **Standard Global Ellipsoid (SGE)** Johnstone and Cox (2021): gets same-size ellipsoidal conformal regions using the global inverse-covariance matrix $\hat{\Sigma}^{-1}$ from proper training data $\mathbf{Z}^{\text{tr}}$ following Equation (1).
- **Normalized Local Ellipsoid (NLE - Ours)**: constructs individual ellipsoidal conformal regions following Equation (2) by exploiting a local covariance matrix specific to each instance, which is obtained as explained in subsection 3.3.

Table 1 summarizes the features of the different methods. We can see from it that local ellipses can capture both covariance structures and local variations, hence providing possibly better results.

	Is Local	Captures Covariance
SEC	x	x
NEC	✓	x
SGE	x	✓
NLE - Ours	✓	✓

Table 1: Properties of the used non-conformity measures.

The experiments are conducted with a 10-fold cross validation following these steps:

1. Split the data set to get proper training $\mathbf{Z}^{\text{tr}}$, calibration $\mathbf{Z}^{\text{cal}}$ and test $\mathbf{Z}^{\text{ts}}$ sets, by allocating 10% of the data (i.e. 1 fold) to $\mathbf{Z}^{\text{ts}}$, and splitting the remaining 90% of the instances (i.e. 9 folds) into proper training and calibration, with a $\mathbf{Z}^{\text{cal}}$ size equal to 10% of the training instances.
2. Train the underlying algorithm on $\mathbf{Z}^{\text{tr}}$, here a multi-output random forest using Scikit Learn's "MultiOutputRegressor", calculate $\hat{\Sigma}$ of the errors made on $\mathbf{Z}^{\text{tr}}$ instances, and get the regressor's predictions for $\mathbf{Z}^{\text{cal}}$ and $\mathbf{Z}^{\text{ts}}$.
3. For the normalized NCMs, train the normalizing models on $\mathbf{Z}^{\text{tr}}$: an MLP for NEC to get σ_i by learning μ_i, and a kNN for our method NLE to get the neighboring instances of x_i, which number is equal to 5% of the number of $\mathbf{Z}^{\text{tr}}$ instances, to calculate $\hat{\Sigma}_i$.
4. For each instance in $\mathbf{Z}^{\text{cal}}$ and $\mathbf{Z}^{\text{ts}}$, get σ_i values for NEC using the MLP, and $\hat{\Sigma}_i$ values for NLE using the kNN with $\lambda = 0.95$, a high value that favors the impact of $\hat{\text{Cov}}_i$ in (4).

5. Choose the significance level's value ϵ.
6. Obtain the non-conformity scores for the different NCMs for $\mathbf{Z}^{\text{cal}}$, sort $\alpha_1, \ldots, \alpha_q$ in a descending order and get the index s of the $(1-\epsilon)$-percentile of the non-conformity score α_s.
7. For a new object x_{n+1} in $\mathbf{Z}^{\text{ts}}$, get its hyper-rectangle or ellipsoidal prediction region depending on the used NCM.

4.2. On synthetic data

The synthetic data set aims at assessing the performance of all NCMs by controlling the dependence structure between the outputs. It contains 50000 instances and is created as follows:

- x^1 and x^2 are generated independently using a uniform distribution with values between -5 and 5.
- y is a 2-dimensional linear transformation of x plus some Gaussian centered noise:

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

The covariance matrix S_x of ε depends on x and is a weighted average of four covariance matrices at four different points μ_i:

$$S_x = \sum_{i=1}^{4} \frac{\Delta(x, \mu_i)}{\sum_{j=1}^{4} \Delta(x, \mu_j)} \times \text{Cov}_{\mu_i} \tag{5}$$

where Δ is an inverse distance calculated as follows :

$$\Delta(x, \mu_i) = \left(\frac{1}{d(x, \mu_i) + \xi} \right)^4 \tag{6}$$

with ξ a small value used to avoid a division by 0. Figure 1 shows the four μ_i points that are placed at the extremities of our distribution and have the distinct covariance matrices Cov_{μ_i} :

$$\text{Cov}_{\mu_1} = \text{Cov}_{\mu_4} = \begin{bmatrix} 0.1 & -0.09 \\ -0.09 & 0.1 \end{bmatrix} \quad \text{and} \quad \text{Cov}_{\mu_2} = \text{Cov}_{\mu_3} = \begin{bmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{bmatrix}$$

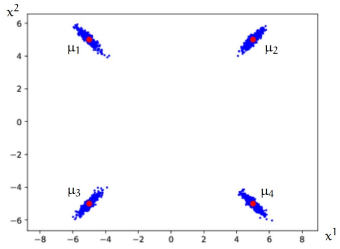
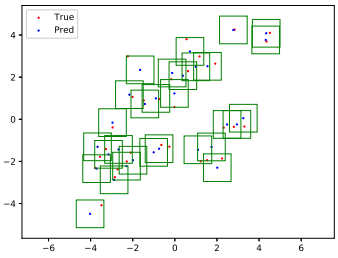


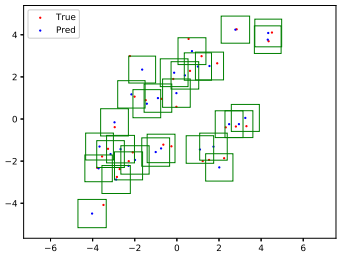
Figure 1: Representation of μ_i points and their covariance matrices.

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	$\mathbf{2.85 \pm 0.07}$	$\mathbf{1.81 \pm 0.02}$	$\mathbf{1.39 \pm 0.02}$	$\mathbf{1.14 \pm 0.01}$	$\mathbf{0.97 \pm 0.01}$

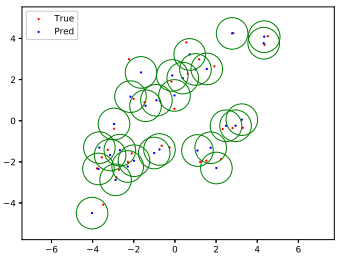
Table 2: Validity and efficiency results for synthetic data.



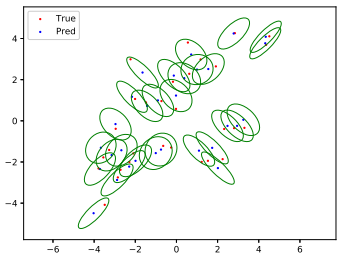
(*a*) Standard Empirical Copula



(*b*) Normalized Empirical Copula



(*c*) Standard Global Ellipsoid



(*d*) Normalized Local Ellipsoid

Figure 2: Results' visualization for synthetic data with $\epsilon = 0.1$.

For synthetic data, we calculated the mean validity and efficiency (surface) of all non-conformity measures for different ϵ values. Results are summed up in Table 2. They show that all NCMs are exactly valid, with a slight difference between them. However, when it comes to efficiency, our method NLE outperforms the others, giving the best results with tighter conformal regions.

For visualization purposes, we drew prediction regions (a rectangle for empirical copula NCMs and an ellipse for ellipsoidal NCMs) in Figure 2. We notice that both empirical copula NCMs give almost the same results, which explains why their efficiency results in Table 2 are almost equal. Another remark is that the drawn ellipses for NLE in subfigure 2d follow the choice of μ_i values and their covariance matrices used to generate synthetic data, which shows that our method respects local covariance, thus explaining the gain in efficiency going from a rectangle to an ellipse.

4.3. On real data

We use 19 data sets with various numbers of targets to compare between the four different NCMs. These data sets are from Mulan and the UCI repository. Their information is summed up in Table 3.

Names	Instances	Features	Targets	Source
residential building	372	105	2	Rafiei and Adeli (2016)
enb	768	8	2	Tsoumakas et al. (2011)
music origin	1059	68	2	Zhou et al. (2014)
bias correction	7750	25	2	Cho et al. (2020)
jura	359	15	3	Tsoumakas et al. (2011)
scpf	1137	23	3	Tsoumakas et al. (2011)
indoor localization	21049	520	3	Torres-Sospedra et al. (2014)
sgemm	241600	14	4	Nugteren and Codreanu (2015)
atp1d	337	411	6	Tsoumakas et al. (2011)
atp7d	296	411	6	Tsoumakas et al. (2011)
rf1	9125	64	8	Tsoumakas et al. (2011)
rf2	9125	576	8	Tsoumakas et al. (2011)
osales	639	413	12	Tsoumakas et al. (2011)
wq	1060	16	14	Tsoumakas et al. (2011)
scm1d	9803	280	16	Tsoumakas et al. (2011)
scm20d	8966	61	16	Tsoumakas et al. (2011)
oes10	403	298	16	Tsoumakas et al. (2011)
oes97	334	263	16	Tsoumakas et al. (2011)
community crime	2215	125	18	Redmond (2009)

Table 3: Information on the used multi-target regression data sets.

Following the same experiment protocol described above, we calculate validity and efficiency values for all NCMs with $\epsilon = 0.1$. These results are summarized in Table 4.

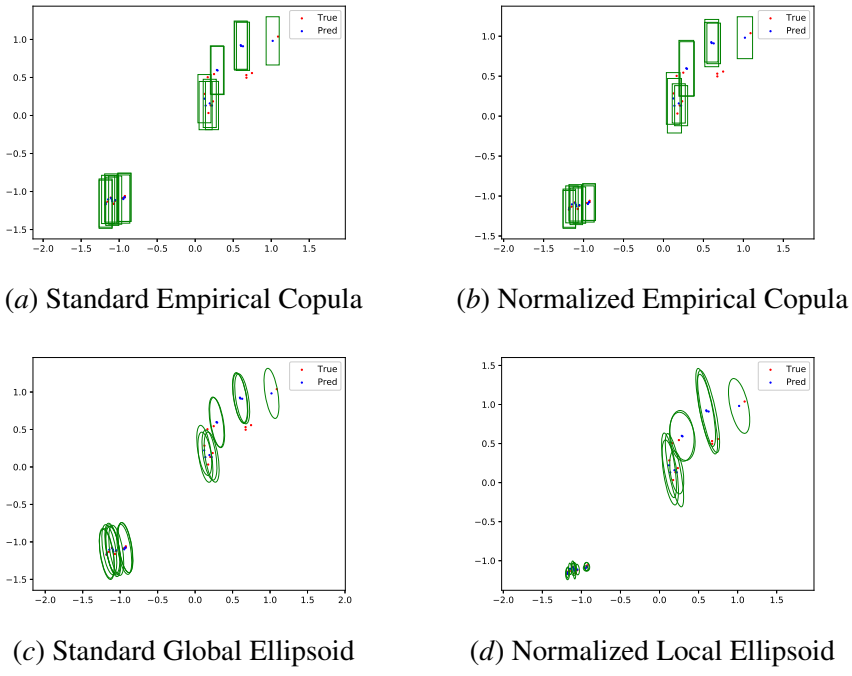
Validity results show that the ellipsoid NCMs perform better than the empirical copula NCMs, especially when it comes to small data sets (with less than 1000 instances). For these data sets, the empirical copula NCMs are invalid with higher variance. This can be explained by the fact that these NCMs are non-parametric, mostly relying on the size of calibration data which is insufficient for smaller data sets. This simply confirms that good performances are achieved when there is a good trade-off between the inductive bias (the amount of made hypothesis) of the method and the amount of data at disposal. Empirical copulas, that make very little assumptions, perform badly when having only a small amount of $\mathbf{Z}^{\text{cal}}$. Even with our theoretical guaranties, the effect of low calibration data can still be noticed for our method (as in "atp7d"), however its impact is less visible.

Efficiency results show that our method gives the tightest volumes in most times, with a massive difference when compared to other NCMs in some cases such as "scpf" and "osales", "scm1d", "scm20d" and "community crime". This confirms that our NCM takes advantage of the flexibility of an ellipsoid shape in order to give tailored prediction regions for each instance depending on its local covariance. For the few data sets where an empirical copula approach gives a tighter volume, this NCM tends to be invalid (except for "indoor localization").

$\epsilon = 0.1$	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	$\mathbf{0.17 \pm 0.05}$
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	$\mathbf{0.04 \pm 0.02}$
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	$\mathbf{9.55 \pm 0.73}$
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	$\mathbf{1.19 \pm 0.07}$
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	$\mathbf{10.17 \pm 5.60}$
scpf	84.52 ± 5.18	84.26 ± 4.95	87.86 ± 5.02	87.87 ± 4.75	$3.77^{10} \pm 6.83^{10}$	$1.26^{10} \pm 2.81^{10}$	$4.87^{7} \pm 8.71^{6}$	$\mathbf{69.59 \pm 89.96}$
indoor loc	90.32 ± 0.48	90.32 ± 0.58	90.37 ± 0.96	90.11 ± 0.89	0.06 ± 0.01	$\mathbf{0.05 \pm 0.01}$	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	$8.45^{-5} \pm 2.56^{-6}$	$7.34^{-5} \pm 3.15^{-6}$	$1.84^{-5} \pm 4.93^{-7}$	$\mathbf{1.15^{-5} \pm 3.41^{-7}}$
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	$\mathbf{0.47 \pm 0.45}$
atp7d	72.29 ± 11.82	68.54 ± 12.96	81.82 ± 10.41	86.13 ± 10.83	8.07 ± 10.73	$\mathbf{0.64 \pm 0.35}$	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	$5.75^{-7} \pm 3.79^{-7}$	$3.60^{-7} \pm 2.16^{-7}$	$6.20^{-7} \pm 5.27^{-7}$	$\mathbf{4.14^{-8} \pm 3.34^{-8}}$
rf2	90.18 ± 1.53	89.79 ± 1.76	90.26 ± 1.31	89.98 ± 1.36	$7.50^{-7} \pm 4.40^{-7}$	$3.13^{-7} \pm 2.38^{-7}$	$6.49^{-7} \pm 4.96^{-7}$	$\mathbf{4.44^{-8} \pm 2.41^{-8}}$
osales	81.39 ± 7.02	78.86 ± 7.88	86.71 ± 3.95	88.12 ± 4.64	$5.60^{6} \pm 9.55^{6}$	$1.20^{15} \pm 3.45^{15}$	$1.47^{5} \pm 2.46^{5}$	$\mathbf{8.08^{4} \pm 1.26^{5}}$
wq	72.74 ± 4.43	78.49 ± 2.76	89.15 ± 4.91	87.55 ± 3.91	$1.31^{10} \pm 6.27^{9}$	$3.93^{17} \pm 1.18^{18}$	$\mathbf{1.16^{9} \pm 7.06^{8}}$	$1.35^{9} \pm 1.08^{9}$
scm1d	89.70 ± 1.27	89.60 ± 1.61	90.07 ± 1.27	90.42 ± 1.25	$6.86^{4} \pm 2.92^{4}$	$1.12^{3} \pm 6.22^{2}$	$4.21^{2} \pm 1.79^{2}$	$\mathbf{8.86 \pm 3.42}$
scm20d	88.23 ± 1.14	88.72 ± 1.64	89.26 ± 1.02	89.45 ± 1.15	$7.98^{5} \pm 6.57^{5}$	$7.02^{4} \pm 3.85^{4}$	$3.21^{3} \pm 2.27^{3}$	$\mathbf{5.52^{2} \pm 3.54^{2}}$
oes10	74.15 ± 13.92	66.13 ± 19.95	87.57 ± 8.13	88.58 ± 6.84	$7.62^{5} \pm 2.23^{6}$	$1.05^{5} \pm 2.86^{5}$	$1.83^{6} \pm 4.83^{6}$	$\mathbf{8.25^{3} \pm 1.26^{4}}$
oes97	69.11 ± 8.18	62.94 ± 15.06	89.22 ± 6.47	87.99 ± 7.41	$\mathbf{1.45^{5} \pm 4.16^{5}}$	$3.86^{6} \pm 8.38^{6}$	$6.90^{6} \pm 1.53^{7}$	$1.18^{7} \pm 1.82^{7}$
com crime	86.18 ± 3.51	84.61 ± 3.04	90.21 ± 3.66	89.03 ± 2.34	$5.19^{9} \pm 9.74^{9}$	$7.05^{4} \pm 7.61^{4}$	$1.17^{5} \pm 1.49^{5}$	$\mathbf{2.80 \pm 7.01}$

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

Table 4: Validity and efficiency results for all real data sets.
For validity, in red are mean values lower than 80% and in orange are mean values between 80% and 85%. For efficiency, tighter volumes are in bold. Reported results are mean values of medians over all folds.



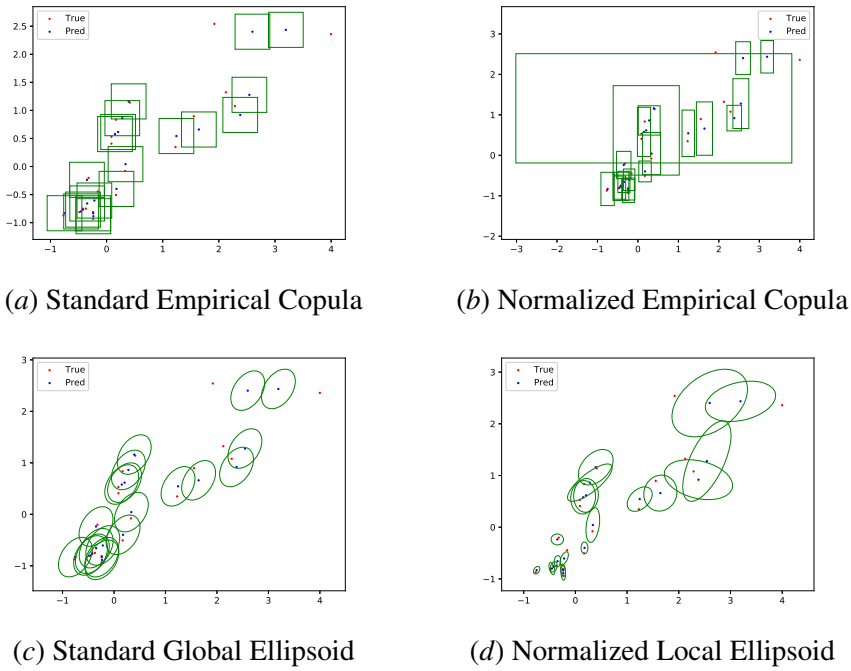
(*a*) Standard Empirical Copula (*b*) Normalized Empirical Copula

(*c*) Standard Global Ellipsoid (*d*) Normalized Local Ellipsoid

Figure 3: Results' visualization for the data set "enb" with $\epsilon = 0.1$.

Figures 3 and 4 illustrate prediction regions for all NCMs for "enb" and "residential building" data sets. For "enb", we can clearly see that even if the dependence structure is roughly the same

for all instances, our method NLE enables us to have adjustable ellipsoid sizes. For "residential building", the different dependence structures in each space region emphasizes the importance of using an approach that takes into consideration local covariance matrices for each instance, and thus, give a noticeable advantage to our method, especially when compared to NEC.



(*a*) Standard Empirical Copula (*b*) Normalized Empirical Copula

(*c*) Standard Global Ellipsoid (*d*) Normalized Local Ellipsoid

Figure 4: Results' visualization for the data set "residential building" with $\epsilon = 0.1$.

5. Conclusion

In this paper, we proposed a new conformal inference approach for a multi-dimensional regression setting that provides individual ellipsoidal uncertainty regions based on the local covariance matrix of the instance. This method showed that it can give tighter volumes compared to other non-conformity measures, and thus better efficiency results while maintaining a validity defined by the required confidence level.

For future research studies, we would like to explore other ways to estimate the local covariance matrix, for example by taking into consideration a density estimation. We would also like to adapt our method to other uncertainty ones such as jack-knife+ (Barber et al. (2021)) and its leave-one-out approach that allows to use all data in the training, especially for small data sets.

References

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1):486–507, 2021.

Rich Caruana. Multitask learning: A knowledge-based source of inductive bias icml. *Google Scholar Google Scholar Digital Library Digital Library*, 1993.

Dongjin Cho, Cheolhee Yoo, Jungho Im, and Dong-Hyun Cha. Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. *Earth and Space Science*, 7(4):e2019EA000740, 2020.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Chancellor Johnstone and Bruce Cox. Conformal uncertainty sets for robust optimization. In *Conformal and Probabilistic Prediction and Applications*, pages 72–90. PMLR, 2021.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NIPS*, 2017.

Alexander Kuleshov, Alexander Bernstein, and Evgeny Burnaev. Conformal prediction in manifold learning. In *Conformal and Probabilistic Prediction and Applications*, pages 234–253, 2018a.

Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate uncertainties for deep learning using calibrated regression. In *International Conference on Machine Learning*, pages 2796–2804. PMLR, 2018b.

Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, page 108101, 2021.

Jelmer Neeven and Evgueni Smirnov. Conformal stacked weather forecasting. In *Conformal and Probabilistic Prediction and Applications*, pages 220–233, 2018.

Cedric Nugteren and Valeriu Codreanu. Cltune: A generic auto-tuner for opencl kernels. In *2015 IEEE 9th International Symposium on Embedded Multicore/Many-core Systems-on-Chip*, pages 195–202. IEEE, 2015.

Hujie Pan, Zining Wang, Wei Zhan, and Masayoshi Tomizuka. Towards better performance and more explainable uncertainty for 3d object detection of autonomous vehicles. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–7. IEEE, 2020.

Harris Papadopoulos and Haris Haralambous. Reliable prediction intervals with regression neural networks. *Neural Networks*, 24(8):842–851, 2011.

Telma Pereira, Sandra Cardoso, Manuela Guerreiro, Sara C Madeira, Alzheimers Disease Neuroimaging Initiative, et al. Targeting the uncertainty of predictions at patient-level using an ensemble of classifiers coupled with calibration methods, venn-abers, and conformal predictors: A case study in ad. *Journal of biomedical informatics*, 101:103350, 2020.

Mohammad Hossein Rafiei and Hojjat Adeli. A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management*, 142(2): 04015066, 2016.

Michael Redmond. Communities and crime data set. *UCI Machine Learning Repository. In website: https://archive.ics.uci.edu/ml/datasets/communities +and+crime*, 2009.

Joaquín Torres-Sospedra, Raúl Montoliu, Adolfo Martínez-Usó, Joan P Avariento, Tomás J Arnau, Mauri Benedito-Bordonau, and Joaquín Huerta. Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems. In *2014 international conference on indoor positioning and indoor navigation (IPIN)*, pages 261–270. IEEE, 2014.

Grigorios Tsoumakas, Eleftherios Spyromitros-Xioufis, Jozef Vilcek, and Ioannis Vlahavas. Mulan: A java library for multi-label learning. *Journal of Machine Learning Research*, 12(71):2411–2414, 2011. URL http://jmlr.org/papers/v12/tsoumakas11a.html.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer Science & Business Media, 2005.

Vladimir Vovk, Ilia Nouretdinov, Valentina Fedorova, Ivan Petej, and Alex Gammerman. Criteria of efficiency for set-valued classification. *Annals of Mathematics and Artificial Intelligence*, 81 (1):21–46, 2017.

Fang Zhou, Q Claire, and Ross D King. Predicting the geographical origin of music. In *2014 IEEE International Conference on Data Mining*, pages 1115–1120. IEEE, 2014.