Hierarchical Copula-based Conformal Prediction and Exact Validity via Nested Prediction Regions

Bruce Cyusa Mukama BRUCE.CYUSA-MUKAMA@HDS.UTC.FR
Soundouss Messoudi SOUNDOUSS.MESSOUDI@HDS.UTC.FR
Sylvain Rousseau SYLVAIN.ROUSSEAU@HDS.UTC.FR
Sébastien Destercke SEBASTIEN.DESTERCKE@HDS.UTC.FR
Université de technologie de Compiègne, CNRS, Heudiasyc, Compiègne, France

Editor: Khuong An Nguyen, Zhiyuan Luo, Harris Papadopoulos, Tuwe Löfström, Lars Carlsson and Henrik Boström

Abstract

Empirical, Archimedean and vine copulas have been repeatedly investigated and leveraged to infer conformal prediction regions for multivariate predictions, but they do not provide finite-size guarantees when the estimated copula is biased or misspecified. To address this limitation, we start with copula-based conformal prediction regions that are always nested and we leverage this property to counteract this copula-estimation bias, via an additional conformal re-calibration step. Furthermore, we introduce a simpler class of semi-parametric copulas (i.e., hierarchical Archimedean copulas) as an alternative to the more complex vine copulas for which incorporating prior knowledge is difficult. Using synthetic data sets, we compare biased and debiased copula-based conformal prediction methods, and we report the impact of the data size and the impact of the number of output dimensions. Using real data, we leverage prior knowledge via this simpler class of copulas. In these experiments, we observe that this additional re-calibration step effectively eliminates the estimation bias of empirical and semi-parametric copulas when its computations are precise (enough) and the data size is large enough. The debiased hierarchical Archimedean copulas yield performances that are comparable to the results of debiased vine copulas.

Keywords: exact validity, multivariate, conformal prediction, hierarchical copulas, knowledge representation

1. Introduction

In safety-critical applications and in other contexts that require rigorous statistical guarantees, most machine learning models cannot be safely deployed without additional calibration steps. For this purpose, univariate conformal prediction (Vovk et al., 2022, Chapter 2) provides finite-size statistical guarantees without assuming the specific distribution of the prediction model's errors: it is sufficient to assume that the prediction errors are exchangeable. As such, these procedures are distribution-free and model-agnostic because they do not depend on the details of the underlying prediction algorithm. Moreover, inductive conformal prediction (Papadopoulos et al., 2002) can be efficiently applied to any memoryless machine learning model, in a post-hoc calibration scheme.

However, in the case of multivariate outputs, conformal prediction (CP) cannot be directly applied when joint statistical guarantees are required. To generalize conformal

prediction to such cases, many approaches (Dheur et al., 2025) have been developed and copula-based conformal prediction (Messoudi et al., 2021) is one of the most straightforward, and one of the earliest.

Copula-based conformal prediction (CoCP) consists in estimating a copula $\widehat{\mathcal{C}}$, i.e., a dependency model of the prediction errors, and in leveraging this copula to determine the dimension-wise confidence levels $(1-\alpha_1, \ldots, 1-\alpha_m)$ with which univariate conformal prediction should be applied to each dimension $d \in \{1, \ldots, m\}$ of the output $\widehat{\boldsymbol{y}} \in \mathbb{R}^m$ in order to yield a user specified joint confidence level $1-\alpha$. The effectiveness of this method has been shown experimentally (Messoudi et al., 2021; Zhang et al., 2023) using estimated copula models $\widehat{\mathcal{C}}$ that are non-parametric $\mathcal{C}_n$ or a semi-parametric $\mathcal{C}_\theta$.

Since the estimated copula $\widehat{\mathcal{C}}$ can be biased, for instance when the sample size is (very) limited, the provided joint statistical guarantees are not provably exactly valid: when the sample size is finite, despite the impressive performance of CoCP, the user-specified joint confidence level $1-\alpha$ is not theoretically guaranteed because there remains an estimation bias $\Delta(\boldsymbol{u}) = \widehat{\mathcal{C}}(\boldsymbol{u}) - \mathcal{C}(\boldsymbol{u})$, with $\boldsymbol{u} \in [0,1]^m$ and $\mathcal{C}$ as the actual (true) copula. In other words, as initially developed by Messoudi et al. (2021), CoCP only provides finite-size joint guarantees if the estimated copula $\widehat{\mathcal{C}}$ is equal everywhere to the true copula $\mathcal{C}$.

Furthermore, although semi-parametric copulas $\mathcal{C}_\theta$ can also be used to robustly infer prediction regions, the classes of such copulas that have been investigated remain limited. Messoudi et al. (2021) introduced simple Archimedean copulas (Genest and Favre, 2007), but this class of copulas can only model symmetrical dependencies between the dimensions of the output. Park and Cho (2025) introduced vine copulas (Czado and Nagler, 2022), but their structure can be too complex for the incorporation of prior knowledge.

This limitation is all the more important because the estimation of semi-parametric models can be improved by leveraging (accurate & informative) prior knowledge. This practice is popular when the data is limited. Moreover, the incorporation of prior knowledge can also increase expert acceptance of the (more interpretable) method and allow those same experts to potentially extract additional knowledge from the statistical analysis. In view of the aforementioned, we introduce the class of hierarchical Archimedean copulas (Hofert et al., 2013) as an alternative to vine copulas and to (simple) Archimedean copulas. More specifically, our contributions are as follows:

- to address the residual copula estimation bias, we introduce an additional conformal re-calibration (or debiasing) step that provably ensures finite-size joint coverage guarantees for any non-parametric or semi-parametric copula model,
- to ease the incorporation of prior knowledge in semi-parametric copula-based conformal prediction with asymmetric dependencies, we introduce the class of hierarchical Archimedean copulas,
- we compare biased CoCP methods and debiased CoCP methods with synthetic & real data, and the experiments show the effectiveness of the introduced re-calibration step.

The remainder of this article is organised as follows: Section 2 presents the preexisting research works that are related to our contributions. Sections 3 & 4 present the methods that are investigated in this study. Section 5 details our experiments and discusses the results. Finally, Section 6 concludes the study and draws recommendations for future developments.

2. Background

Multivariate conformal prediction There are many approaches to generalizing CP to multivariate regression setting. CP has been applied to multivariate regression using dimensionality reduction (Kuleshov et al., 2018; Andéol et al., 2023), multiple hypothesis testing (Angelopoulos et al., 2025; Timans et al., 2025), high predictive density regions (Izbicki et al., 2022; Sampson and Chan, 2024), Monge-Kantorovich (multivariate) quantiles (Thurin et al., 2025; Klein et al., 2025) and estimated copulas (Messoudi et al., 2021; Sun and Yu, 2024). In this article, we shall focus on copula-based multivariate CP approaches, thus we refer the readers to Dheur et al. (2025) for the details of the other approaches.

Copula-based conformal prediction This technique was introduced by Messoudi et al. (2021) as a different approach to generalizing conformal prediction to multi-target regression, and it has been adopted and extended several times. To improve the efficiency of CoCP by reducing the size of the regions that are inferred, Zhang et al. (2023) reformulated CoCP as a constrained optimization problem where the goal is to infer the smallest prediction regions and the constraints enforce the user-specified joint confidence level $1-\alpha$. Park and Cho (2025) used a vine copula with kernels (Geenens, 2014) and influence functions, i.e., targeted learning (Van der Laan et al., 2011) to correct the (plugin) bias in the quantiles of CoCP. Those approaches improve the size of the predicted regions, but they do not provide finite-size joint statistical guarantees for biased or misspecified copulas.

Regarding the preexisting works that inspired our contributions, similarly to Zhang et al. (2023), Sun and Yu (2024) used a constrained (re)formulation of CoCP in order to define a mulitivariate quantile function that would ensure finite-size joint statistical guarantees. However, this approach requires the use of the non-parametric copula $\mathcal{C}_n$ and this model is cursed by the number of the output dimensions m. Contrary to their approach, our recalibration approach is compatible with any type of copula. Lastly, as detailed in Section 4, our bias correction method relies on a geometry-based interpretation of inductive CP that was provided by Gupta et al. (2022).

3. Copula-based conformal prediction

3.1. Introduction to CoCP

Multivariate conformal regression infers a prediction region $\mathbf{\Gamma}_i^\alpha$ for a ground truth vector $\boldsymbol{y}_{n+1} = [y_{n+1,1}, \ldots, y_{n+1,m}] \in \mathbb{R}^m$, such that the values on each dimension of $\boldsymbol{y}_{n+1}$ are simultaneously covered with a user-specified joint confidence level $1-\alpha$:

$$P(\boldsymbol{y}_{n+1} \in \mathbf{\Gamma}_{n+1}^\alpha) = P(y_{n+1,1} \in \Gamma_{n+1,1}^\alpha, \ldots, y_{n+1,m} \in \Gamma_{n+1,m}^\alpha) \geq 1-\alpha \tag{1}$$

where $\Gamma_{n+1,d}^\alpha$ designates the conformal prediction interval that is inferred from the d-th dimension of the predicted vector $\widehat{\boldsymbol{y}}_{n+1} \in \mathbb{R}^m$. The above guarantee is achieved with a calibration data set $\mathcal{D}_{\text{Cal}} = \{(\boldsymbol{y}_i, \widehat{\boldsymbol{y}}_i)\}_{i=1}^n$ and with a non-conformity measure such as:

$$\boldsymbol{s}_i = |\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i| = (|y_{i,1} - \widehat{y}_{i,1}|, \ldots, |y_{i,m} - \widehat{y}_{i,m}|) = (s_{i,1}, \ldots, s_{i,m}) \in \mathbb{R}^m. \tag{2}$$

Given the marginal cumulative distribution functions $(F_1, \ldots, F_m)$ of the multidimensional non-conformity scores $\{\boldsymbol{s}_i\}_{i=1}^n$, the dependencies between the dimensions of the random vector $\boldsymbol{s} = (s_1, \ldots, s_m)$ can be modelled via a unique copula function $\mathcal{C}$, under mild conditions,

as proven by Sklar's theorem (Sklar, 1959):

$$P(\boldsymbol{s} \leq \boldsymbol{s}_i) = \mathcal{C}(F_1(s_{i,1}), \ldots, F_m(s_{i,m})) \tag{3}$$

$$\text{with } \boldsymbol{s} \leq \boldsymbol{s}_i \iff s_d \leq s_{i,d}\ \forall d \in \{1, \ldots, m\}. \tag{4}$$

As shown by Messoudi et al. (2021), if the non-conformity scores $\{\boldsymbol{s}_i\}_{i=1}^{\infty}$ are exchangeable, copulas (Genest and Favre, 2007) can be leveraged to jointly apply (univariate) inductive CP (Papadopoulos et al., 2002) to the dimensions of the output $\widehat{\boldsymbol{y}}_i \in \mathbb{R}^m$. Hence, the joint confidence region $\boldsymbol{\Gamma}_i^{\alpha}$ can be a hyper-rectangle that is defined by the Cartesian product of m dimension-wise conformal intervals $\Gamma_{i,d}^{\alpha}$:

$$\boldsymbol{\Gamma}_i^{\alpha} = \Gamma_{i,1}^{\alpha} \times \cdots \times \Gamma_{i,m}^{\alpha} \tag{5}$$

$$\Gamma_{i,d}^{\alpha} = [\widehat{y}_{i,d} - s_{\alpha,d}\ ,\ \widehat{y}_{i,d} + s_{\alpha,d}] \tag{6}$$

$$\text{with } s_{\alpha,d} = F_d^{-1}\left((1-\alpha_d)(1+1/n)\right),\ \forall d \in \{1, \ldots, m\} \tag{7}$$

where $1-\alpha_d$ is the confidence level of the d-th dimension of the output. Following Deheuvels (1979), Messoudi et al. (2021) estimate an empirical copula $\mathcal{C}_n$ as well as m empirical marginal distribution functions $(\widehat{F}_1, \ldots, \widehat{F}_m)$, and they determine the required dimension-wise confidence levels $(1-\alpha_1, \ldots, 1-\alpha_m)$ as follows:

$$\boldsymbol{u}_i = (u_{i,1}, \ldots, u_{i,m}) = \left(\widehat{F}_1(s_{i,1}), \ldots, \widehat{F}_m(s_{i,m})\right) \tag{8}$$

$$\mathcal{C}_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{u_{i,1} \leq u_1, \ldots, u_{i,m} \leq u_m\right\},\ \ \forall \boldsymbol{u} \in [0,1]^m \tag{9}$$

$$1-\alpha_1 = 1-\alpha_2 = \cdots = 1-\alpha_m = \min\left\{u \in [0,1] : \mathcal{C}_n(u, u, \ldots, u) \geq 1-\alpha\right\} \tag{10}$$

where $\mathbb{1}\{\ldots\}$ takes the value 1 if all the inequalities are verified and 0 otherwise. Here, all the dimensions-wise confidence levels $1-\alpha_d$ are equal to a single scalar $u \in [0,1]$. Initially, Messoudi et al. (2021) introduced this constraint to limit the number of feasible solutions, but we shall show this constraint can also be used to infer prediction regions that are always nested (see Figure 1) and to provide finite-size coverage guarantees by leveraging this property (via an additional calibration step).

Furthermore, this constraint essentially defines a single direction, the line $u_1 = u_2$ in Figure 1, but this direction is somewhat arbitrary because other directions could be considered if they yield smaller prediction regions, for example. Thus, in a bivariate case, the constraint in Equation 10 can be weakened and generalized as follows:

$$(1-\alpha_1, 1-\alpha_2) = \boldsymbol{u} \in [0,1]^2 \text{ s.t. } \mathcal{C}_n(u_1, u_2) \geq 1-\alpha \text{ and } \frac{u_1}{u_2} = a \in \mathbb{R}^+ \tag{11}$$

where $a \in \mathbb{R}^+$ is the direction-defining constant (which is equal to 1 in Equation 10). Regarding the provided statistical guarantees, Deheuvels (1979) shows that the estimator $\widehat{F}(\boldsymbol{s}) = \mathcal{C}_n(\widehat{F}_1(s_1), \ldots, \widehat{F}_m(s_m))$ converges almost surely to the true joint distribution function $F(\boldsymbol{s})$ as n increases $(n \to \infty)$. Therefore, the non-parametric method in (Messoudi et al., 2021) can only guarantee the user-specified joint confidence level $1-\alpha$ asymptotically: it does not provide finite-size marginal coverage guarantees, i.e., exact validity.

Furthermore, the convergence of the empirical copula model $\mathcal{C}_n$ can require more data as the numbers of the dimensions of the non-conformity scores $\boldsymbol{s}_i$ increases. On one hand, there exists other copula models that are not subject to the curse of dimensionality: namely, vine copulas and Archimedean copulas, as respectively shown by Nagler and Czado (2016) and Hofert et al. (2013). On the other hand, these (scalable) semi-parametric alternatives can be subject to specification errors.

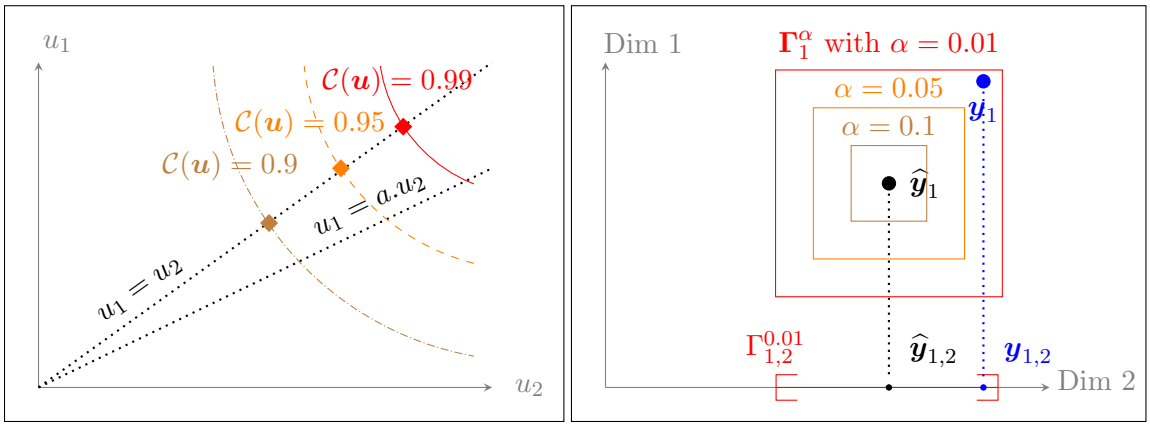

Figure 1: bivariate copula-based nested prediction regions for $1 - \alpha \in \{0.9, 0.95, 0.99\}$. The solutions of the problem in Equation 10 are shown on the left as diamonds. The line $u_1 = a.u_2$ shows an alternative nested-set-generating constraint. The right side shows the resulting (nested) conformal prediction regions and one valid conformal interval on the $2nd$ dimension ($\Gamma_{1,2}^{0.01}$).

3.2. Semi-parametric conformal prediction with vine copulas

Vine copulas

As detailed by Czado and Nagler (2022), Sklar's theorem (Equation 3) can be reformulated by derivation as follows:

$$f(s_{i,1}, \ldots, s_{i,m}) = c(F_1(s_{i,1}), \ldots, F_m(s_{i,m}))f_1(s_{i,1}) \ldots f_m(s_{i,m}) \tag{12}$$

where $f(\boldsymbol{s}_i)$ is the joint density function and $f_d(s_{i,d})$ is the marginal density function of the d-th dimension, and c represents the density function of the copula $\mathcal{C}$. Furthermore, the conditional density function $f_{a|b}$ of a dimension a given a dimension b and the corresponding conditional CDF $F_{a|b}$ can be expressed as follows:

$$f_{a|b}(s_{i,a}|s_{i,b}) = c_{ab}(F_a(s_{i,a}), F_b(s_{i,b}))f_b(s_{i,b}) \tag{13}$$

$$F_{a|b}(s_{i,a}|s_{i,b}) = \frac{\partial}{\partial F_b(s_{i,b})}\mathcal{C}_{ab}(F_a(s_{i,a}), F_b(s_{i,b})) \tag{14}$$

by leveraging the properties in Equations 13 & 14, Equation 12 can be transformed into a product of bivariate copulas and marginal densities. In the case of a four-dimensional non-conformity scores ($m = 4$), we can obtain the following:

$$
\begin{aligned}
f(s_{i,1},\ldots,s_{i,4}) = c_{14|23}(&F_{1|23}(s_{i,1}|s_{i,2},r_3), F_{4|23}(s_{i,4}|s_{i,2},s_{i,3}) \mid s_{i,2},s_{i,3}) \\
&\times c_{13|2}(F_{1|3}(s_{i,1}|s_{i,3}), F_{3|2}(s_{i,3}|s_{i,2}) \mid s_{i,2}) \\
&\times c_{24|3}(F_{2|3}(s_{i,2}|s_{i,3}), F_{4|3}(s_{i,4}|s_{i,3}) \mid s_{i,3}) \\
&\times c_{34}(F_3(s_{i,3}), F_4(s_{i,4})) \times c_{23}(F_2(s_{i,2}), F_3(s_{i,3})) \\
&\times c_{12}(F_1(s_{i,1}), F_2(s_{i,2})) \times f_1(s_{i,1}) f_2(s_{i,2}) f_3(s_{i,3}) f_4(s_{i,4})
\end{aligned}
\tag{15}
$$

Where $c_{ab|cd}$ is the conditional copula density of the dimensions (a, b) given the dimensions (c, d). The above factorization enables a graphical representation named "vine copula" (Bedford and Cooke, 2001): a sequence of trees where each node represents a variable and each edge represents a bivariate copula that models the dependency between two nodes. Although there can be many possible permutations of variables in Equation 15, each permutation does not necessarily produce a valid factorization of the joint density function.

Regular vine copula

To ensure that a vine (a factorization) truly represents a density function, Bedford and Cooke (2001) defined a class of structures called "regular vines" (R-vine) copulas: sequences of trees where the edges in a previous tree become nodes in the next tree (see Figure 2). More specifically, the sequence of trees $\{T_1, \ldots, T_{m-1}\}$ is an actual m-dimensional R-vine if the following statements are verified:

- the tree T_1 is composed of nodes $N_1 = \{1, \ldots, m\}$ and edges E_1,
- for all $j \geq 2$, the tree T_j is composed of nodes $N_j = E_{j-1}$ and edges E_j,
- for all $j = 2, \ldots, m-1$, $\{a, b\} \in E_j$ iff $|a \cap b| = 1$.

These characteristics guarantee that the graphical model $\{T_1, \ldots, T_{m-1}\}$ represents a valid copula (Czado and Nagler, 2022; Bedford and Cooke, 2001), but they also reduce the flexibility of the graphical model and this can hinder the incorporation of prior knowledge, especially when the user specifies that information in the form of a single knowledge tree.

Vine-copula based conformal prediction (VcoCP)

Starting from Equation 10, one can replace the empirical copula with a better copula model. For instance, Park and Cho (2025) use an empirical vine copula with transformation local likelihood kernel estimators (Geenens, 2014) and asymptotically guarantee the user specified joint confidence level $1 - \alpha$ for 3 to 16 output dimensions.

3.3. Semi-parametric conformal prediction with HAC(s)

In this article, we introduce Hierarchical Archimedean Copulas (HAC) (Górecki and Okhrin, 2024; Okhrin and Ristig, 2014) to ease the incorporation of domain knowledge into the estimation of the copula because accurate prior knowledge can reduce the amount of necessary calibration data. HAC(s) are also more flexible than (simple) Archimedean copulas: the basic Archimedean copulas (Gumbel, Clayton, etc.) can only model symmetrical joint cumulative distribution functions.

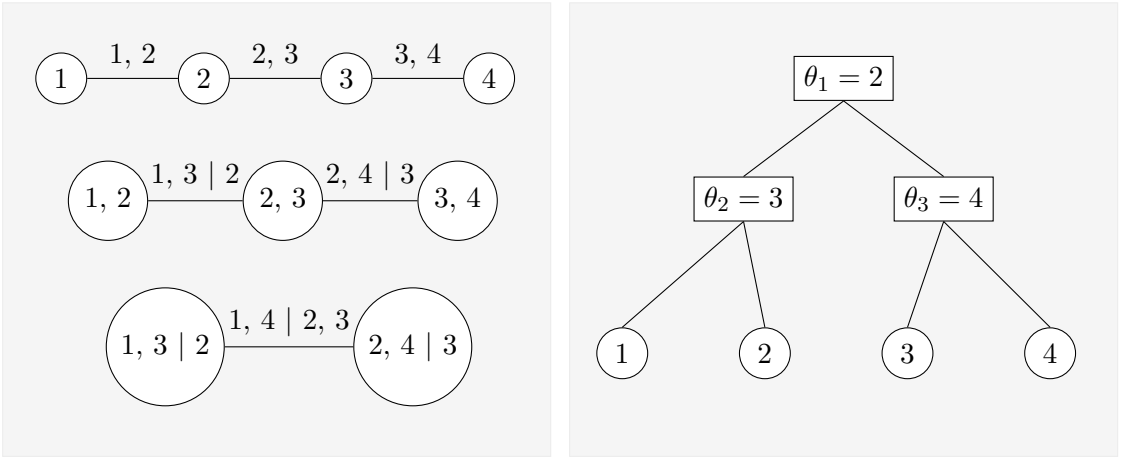

Figure 2: An illustration a regular vine copula composed of three trees (on the left) and a partially nested hierarchical Archimedean copula (on the right).

Archimedean copulas

Archimedean copulas are simple closed-form representations that are used to model a non-elliptical cumulative distribution function that might exhibit some tail dependency:

$$\mathcal{C}_\theta(\boldsymbol{u}) = \phi_\theta\left(\phi_\theta^{-1}(u_1) + \cdots + \phi_\theta^{-1}(u_m)\right), \quad \boldsymbol{u} \in [0,1]^m, \tag{16}$$

$$\text{with } \phi_\theta : [0,\infty) \to [0,1],\ \phi_\theta(0) = 1,\ \phi_\theta(\infty) = 0, \tag{17}$$

$$\phi_\theta(x) = \exp\left(-x^{\frac{1}{\theta}}\right), \quad \text{for Gumbel copulas.} \tag{18}$$

Additionally, the derivatives $(-1)^j \phi_\theta^{(j)}(x)$ must be non-decreasing and convex for all $x \geq 0$ and all $j \in \mathbb{N}$. The generator function $\phi(x)$ determines the family of the estimated copula (e.g., Equation 18 for Gumbel Archimedean copulae). Since Archimedean copulas are commonly defined with one to three parameters (Genest and Favre, 2007), Archimedean copulas with a single parameter θ can be alluring because of their simplicity, but they can only model cumulative distribution functions that are symmetrical, i.e., with the same dependency structure between all the dimensions of $\boldsymbol{u} \in [0,1]^m$.

Hierarchical Archimedean copulas (HAC)

Hierarchical Archimedean copulas (Okhrin and Ristig, 2014) break the symmetry of mono-parametric Archimedean copulas by nesting multiple (simple) Archimedean copulas. For instance, the partially nested HAC in Figure 2 is defined as follows:

$$\mathcal{C}_{\boldsymbol{\theta}}(\boldsymbol{u}) = \mathcal{C}_{\theta_1}\left(\mathcal{C}_{\theta_2}(u_1, u_2), \mathcal{C}_{\theta_3}(u_3, u_4)\right), \quad \boldsymbol{u} \in [0,1]^4, \tag{19}$$

$$\mathcal{C}_{\theta_1}(\boldsymbol{u}) = \phi_{\theta_1}\left(\phi_{\theta_1}^{-1} \circ \mathcal{C}_{\theta_2}(u_1, u_2) + \phi_{\theta_1}^{-1} \circ \mathcal{C}_{\theta_3}(u_3, u_4)\right) \tag{20}$$

Thus, the corresponding joint cumulative distribution function is not symmetrical if the parameters (θ_2, θ_3) are different. Furthermore, the nested copulas can be non-bivariate and

they can be drawn from different Archimedean families (e.g., Gumbel, Clayton, Frank). However, the resulting HAC is not always a valid cummulative distribution function: the compositions of the generative functions ($\phi_{\theta_1} \circ \phi_{\theta_2}(x)$, $\phi_{\theta_1} \circ \phi_{\theta_3}(x)$, etc.) must also be valid generator functions (Rezapour, 2015). In the case of nested copulas that are drawn from the same family, the HAC is a valid cummulative distribution function if the values of the parameters ($\theta_1, \theta_2, \theta_3$, etc.) increase from the highest to the lowest hierarchical level (Okhrin and Ristig, 2014), as in Figure 2. Thus, this class of copulas is readily compatible with prior knowledge that is specified in the form of a single knowledge tree.

More concretely, in an object detection use case, the nodes ① and ③ could represent the errors w.r.t. the predicted latitude and longitude of an object (i.e., its location), and the nodes ② and ④ could represent the errors w.r.t. the predicted height and width of the object (i.e., its size). In this case, the tree on the right side of Figure 2 would inform or enforce that the errors on the latitude and the width are more correlated together than they are correlated with the errors w.r.t. the longitude and height. Here, the user would only have to specify the structure of the tree, and a suitable algorithm (Hofert et al., 2013) would determine the values of θ_1, θ_2 and θ_3.

HAC-based conformal prediction (HcoCP)

Similarly to the VcoCP method, we can replace the empirical copula in Equation 10 with an HAC. As opposed to vine-based methods, here the user can easily use prior knowledge to define the graphical structure of the HAC without knowing the exact values of the parameters $\boldsymbol{\theta}$. Specifically, the user must be able to organise the dimensions of the non-conformity score $\boldsymbol{s}_i \in \mathbb{R}^m$ into a tree where the nodes in the lower levels of the hierarchy are more correlated than the nodes in the upper levels.

Incorporating such knowledge representations can improve the estimation of the copula, but their specification can also be demanding (but not as much as vine copulas) if the number of child nodes is significantly high. Also, similarly to the previous copula-based conformal prediction methods, this method only provides asymptotic coverage guarantees if the copula (the HAC) converges to the true copula of the scores $\boldsymbol{s}_i \in \mathbb{R}^m$.

Thus, whenever the non-parametric or semi-parametric copula is estimated, finite-size coverage guarantees cannot hold unless the estimated copula has fully converged to the true copula. Otherwise, there would still be a coverage gap due to the remaining copula bias.

On the advantages of HcoCP

As repeatedly evoked in this article, hierachical copula-based conformal prediction (HcoCP) provides many advantages. This section briefly compares it to the other CoCP approaches:

- compared to CoCP with (simple) Archimedian copulas, HcoCP has the advantage of capturing dependency structures that are asymmetrical (i.e., more complex);
- compared to CoCP with vine copulas, HcoCP relies on a (tree-like) copula model that is significantly more interpretable, easier to specify and to justify. This model also requires fewer parameters and its estimation cost can be reduced by constraining (i.e., by specifying) the underlying tree structure;

- compared to CoCP with empirical copulas, HcoCP relies on a semi-parametric dependency model that scales better with the number of output dimensions. Furthermore, the debiasing method in the next section shows that its estimation bias (or specification errors) can be corrected with a finite-size sample.

4. Copula bias correction

To expedite the proof of Lemma 2, we introduce an explanatory view of conformal prediction that is based on "ordered nested sets" (Gupta et al., 2022). From this geometric perspective, with $\mathscr{T}$ as the range of possible values of a parameter t, the inference of an interval Γ_i^α such that $P(y_i \in \Gamma_i^\alpha) \geq 1-\alpha$ can be seen as the selection of the smallest interval among all the prediction intervals $\{\mathscr{F}_t\}_{t\in\mathscr{T}}$ that are nested in each other, ordered according to t, and centered at the predicted value $\widehat{y}_i \in \mathbb{R}$; Specifically:

1. Gupta et al. (2022) require that $\mathscr{F}_{Inf\{\mathscr{T}\}} = \emptyset$ and $\mathscr{F}_{Sup\{\mathscr{T}\}} = \mathbb{R}$, thus $y_{i,m} \in \bigcup_{t\in\mathscr{T}} \mathscr{F}_t$,
2. Gupta et al. (2022) also require that $\mathscr{F}_{t_2} \subseteq \mathscr{F}_{t_1}$ for all $t_1, t_2 \in \mathscr{T}$ such that $t_1 \geq t_2$ (larger values of t should yield larger intervals),
3. given a calibration set $\mathcal{D}_{\text{Cal}} = \{(y_i, \widehat{y}_i)\}_{i=1}^n$, one can geometrically construct random sets around the prediction $\widehat{y}_i$ such as $\{\mathscr{F}_t = [\widehat{y}_i - t, \widehat{y}_i + t] : t \in \mathscr{T}\}$,
4. from the random sets $\{\mathscr{F}_t : t \in \mathscr{T}\}$ and the calibration set $\mathcal{D}_{\text{Cal}}$, one can compute the scalar scores $\mathscr{S}_i = Inf\{t \in \mathscr{T} : y_i \in \mathscr{F}_t\}$,
5. for any unseen target y_{n+1}, Gupta et al. (2022) prove that $P(y_{n+1} \in \mathscr{F}_{t_\alpha}) \geq 1-\alpha$ if t_α is the $(1-\alpha)(1+1/n)$ quantile of the scores $\{\mathscr{S}_i\}_{i=1}^n$.

We leverage this nested CP method to de-bias any estimated copula $\widehat{\mathcal{C}}(\boldsymbol{u})$ thanks to an additional re-calibration or debiasing data set $\mathcal{D}_{\text{Bias}}$. We denote the size of this debiasing data set by $n' < \infty$ and by n the size of the (primary) calibration data set $\mathcal{D}_{\text{Cal}}$, and we consider the non-conformity scores $\{\boldsymbol{s}_i\}_{i=1}^{n+n'}$ to be exchangeable. Under that condition, we define the secondary non-conformity scores $\mathscr{S}_i$ and the resulting de-biased joint confidence level $1-\widehat{\alpha}$ as follows:

$$\mathscr{S}_i = Inf\left\{ (1-\alpha') \in [0,1] : \left(y_{i,d} \in \Gamma_{i,d}^{\alpha'}, \forall d \in \{1,\ldots,m\}\right) \right\} \tag{21}$$

$$(1-\widehat{\alpha}) = (1-\alpha)(1+1/n')\text{-quantile of the secondary scores } \{\mathscr{S}_i\}_{i=n+1}^{n+n'} \tag{22}$$

Lemma 1 *The copula-based conformal prediction regions $\boldsymbol{\Gamma}_i^\alpha$ in Equation 5 are "ordered nested sets" and their parameter is the joint confidence level $1-\alpha$.*

Proof Given that copulas are non-decreasing functions, the constraint $1-\alpha_1 = 1-\alpha_2 = \cdots = 1-\alpha_m$ in Equation 10 induces a complete order between the dimension-wise confidence levels $(1-\alpha_d)$ and $(1-\beta_d)$ that are associated to two joint confidence levels $(1-\alpha)$ and $(1-\beta)$:

$$(1-\alpha) \geq (1-\beta) \implies (1-\alpha_d) \geq (1-\beta_d) \quad \forall d \in \{1,\ldots,m\} \tag{23}$$

Given that the dimension-wise quantile functions are also non-decreasing and because the dimension-wise interval $\Gamma^{\alpha}_{i,d}$ is always centered at the prediction $\widehat{y}_{i,d}$, the dimension-wise confidence intervals are also nested and ordered as follows:

$$(1-\alpha_d) \geq (1-\beta_d) \implies \Gamma^{\beta}_{i,d} \subseteq \Gamma^{\alpha}_{i,d} \quad \forall d \in \{1,\ldots,m\} \tag{24}$$

From the definition of the (multidimensional) confidence regions $\mathbf{\Gamma}^{\alpha}_{i}$ in Equation 5, we get that these prediction regions are ordered and nested as follows:

$$(1-\alpha_d) \geq (1-\beta_d) \implies \mathbf{\Gamma}^{\beta}_{i} \subseteq \mathbf{\Gamma}^{\alpha}_{i} \tag{25}$$

Thus the regions $\mathbf{\Gamma}^{\boldsymbol{\alpha}}_{\boldsymbol{i}}$ are "ordered nested sets". ■

Lemma 2 *The prediction regions $\mathbf{\Gamma}^{\widehat{\alpha}}_{i}$ that are computed at the de-biased joint confidence level $1-\widehat{\alpha}$ (Equation 22) ensure exact validity (i.e., finite-size coverage) at the user specified joint confidence level $1-\alpha$, for any biased copula estimator $\widehat{\mathcal{C}}(\boldsymbol{u}) = \mathcal{C}(\boldsymbol{u}) + \Delta(\boldsymbol{u})$.*

$\mathcal{C}(\boldsymbol{u})$ denotes the true copula of the (primary) non-conformity scores $\boldsymbol{s}_i$ in Equation 2 and $\Delta(\boldsymbol{u})$ denotes the copula estimation bias. As outlined by Algorithm 1, any non-parametric or semi-parametric copula can be estimated in a first step using the (primary) calibration data set $\mathcal{D}_{\text{Cal}}$ and de-biased in a second step using an additional finite-size data set $\mathcal{D}_{\text{Bias}}$.

Algorithm 1 Copula-based conformal prediction with bias correction

Require: the user-specified joint confidence level $1-\alpha$, the calibration data set $\mathcal{D}_{\text{Cal}}$, and the de-biasing data set $\mathcal{D}_{\text{Bias}}$, a non-parametric or semi-parametric copula model $\widehat{\mathcal{C}}$.

1: following Equation 2, use the calibration data set $\mathcal{D}_{\text{Cal}}$ to compute the (primary) non-conformity scores $\{\boldsymbol{s}_i\}_{i=1}^{n}$,
2: use the (primary) non-conformity scores $\{\boldsymbol{s}_i\}_{i=1}^{n}$ to fit the copula model $\widehat{\mathcal{C}}$,
3: following Equation 21, use the estimated copula $\widehat{\mathcal{C}}$ and the re-calibration data set $\mathcal{D}_{\text{Bias}}$ to evaluate the secondary non-conformity scores $\{\mathscr{S}_i\}_{i=n+1}^{n+n'}$,
4: following Equation 22, determine the de-biased confidence level $1-\widehat{\alpha}$ from the secondary non-conformity scores $\{\mathscr{S}_i\}_{i=n+1}^{n+n'}$,
5: infer predictions regions $\mathbf{\Gamma}^{\widehat{\alpha}}_{i}$ that are exactly valid at the user-specified joint confidence level $1-\alpha$ by inserting the estimated copula $\widehat{\mathcal{C}}$ and the adjusted joint confidence level $1-\widehat{\alpha}$ into Equations 10 & 5.

5. Experiments

The experiments compare the following CoCP methods: the baseline with the empirical copula "ec", the (simple) Achimedian copula baseline "ac", the HAC baseline "hac", the independence copula baseline "ic", the vine copula baseline "vac"; and their respective debiased versions: "db-ec", "db-ac", "db-hac", "db-ic" and "db-vac".

5.1. Implementation details

High-dimensional data synthesis

To (finely) control the number of output dimensions m and the sizes of the necessary data sets $\mathcal{D}_{\text{Cal}}$ & $\mathcal{D}_{\text{Bias}}$, we generate a synthetic data set by automatically constructing a known (D-vine) copula $\mathcal{C}$ of the non-conformity scores $\boldsymbol{s}_i = |\widehat{\boldsymbol{b}}_i - \boldsymbol{b}_i| \in \mathbb{R}^m$ depending on the number of output dimensions m, as shown by Algorithm 2. Moreover, choosing the bivariate copulas that compose the m-dimensional copula $\mathcal{C}$ from multiple families also allows us to test CoCP on a data set with dependency structures that are more complex than hierarchical Archimedean Copulas, the class of copulas that is introduced in this study.

Furthermore, the marginal distribution of the non-conformity scores $s_{i,d}$ on each dimension $d \in \{1, \ldots, m\}$ is either $\mathcal{N}(1, 2)$, $U(-3, 4)$, $\mathcal{N}(5, 6)$, or $U(-7, 8)$. The vines and HAC(s) are implemented using R-packages: *rvinecopulib*[1] and *HAC*[2], respectively.

Performance metrics

With $\mathcal{D}_{\text{Test}}$ as the test data set, we use the marginal empirical coverage rate to evaluate the validity of the inferred prediction regions and a custom efficiency metric to evaluate their sizes without ignoring the number of output dimensions:

$$\text{Coverage} = \frac{1}{|\mathcal{D}_{\text{Test}}|} \sum_{i \in \mathcal{D}_{\text{Test}}} \mathbb{1}\{\boldsymbol{y}_i \in \boldsymbol{\Gamma}_i^{\alpha}\} \tag{26}$$

$$\text{Efficiency} = \frac{1}{|\mathcal{D}_{\text{Test}}|} \sum_{i \in \mathcal{D}_{\text{Test}}} \sqrt[m]{|\boldsymbol{\Gamma}_i^{\alpha}|} = \sqrt[m]{\prod_{d=1}^{m} s_{\alpha,d}} \tag{27}$$

Algorithm 2 Generating the copula $\mathcal{C}$ of the non-conformity scores

Require: m the number of output dimensions

$\mathcal{F} \leftarrow$ (*Gumbel*, *Frank*, *Clayton*), # *Diverse candidate copula families*
$\boldsymbol{\theta} \leftarrow$ (5, 10, 25), # *Large candidate copula parameters values*
for each of the $m - 1$ trees **do**
    randomly select the edges (the bivariate copulas) from $\mathcal{F} \times \boldsymbol{\theta}$,
end for
form the m-dimensional D-vine $\mathcal{C}$ with the $m - 1$ trees (as in Figure 2),
return the vine copula $\mathcal{C}$

5.2. Results

On the impact of the number of output dimensions

To empirically assess the impact of the number of the output dimensions m on the marginal coverage results, we fix the size of the calibration data at $|\mathcal{D}_{\text{Cal}}| = |\mathcal{D}_{\text{Bias}}| = 10^4$ and we generate the non-conformity scores from multiple copulae, by varying the number of output

1. https://cran.r-project.org/web/packages/rvinecopulib/index.html
2. https://cran.r-project.org/web/packages/HAC/index.html

dimensions m. We try two setups for Algorithm 2: $m = 4$ and $m = 64$. We repeat each experiment 10 times with the same amount of test data ($|\mathcal{D}_{\text{Test}}| = 10^4$). In Table 1, the best results are in boldface and their format is "mean(Coverage) ± standard deviation".

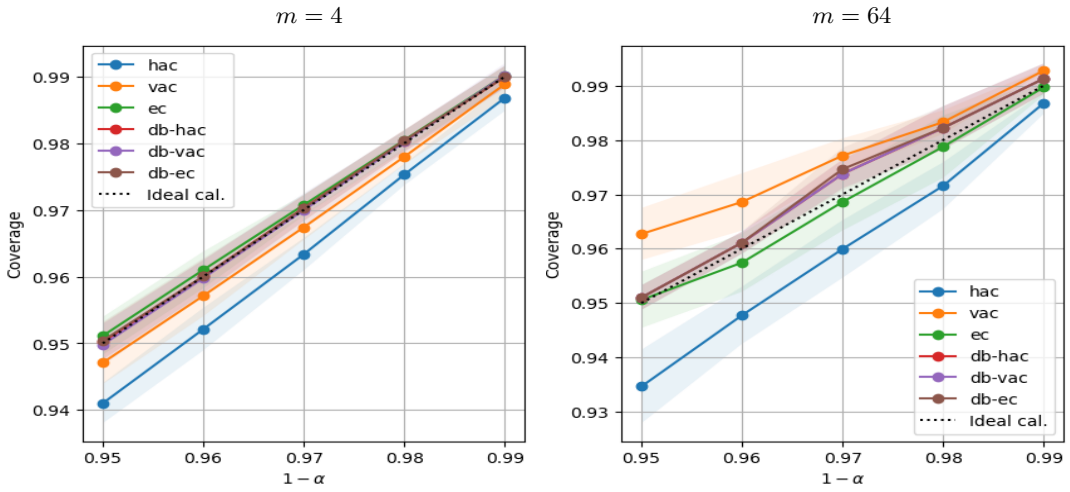

Figure 3: The impact of the number of output dimensions m on marginal coverage rates.

m	$1-\alpha$	db-ec	db-hac	db-vac	ec	hac	vac
4	0.99	**0.990** ± 0.002	**0.990** ± 0.002	**0.990** ± 0.002	**0.990** ± 0.001	0.987 ± 0.002	0.989 ± 0.002
	0.98	**0.980** ± 0.002	**0.980** ± 0.002	**0.980** ± 0.002	**0.980** ± 0.001	0.975 ± 0.002	0.978 ± 0.001
	0.97	**0.970** ± 0.002	**0.970** ± 0.002	**0.970** ± 0.002	0.971 ± 0.002	0.963 ± 0.002	0.967 ± 0.002
	0.96	**0.960** ± 0.003	**0.960** ± 0.002	**0.960** ± 0.003	0.961 ± 0.003	0.952 ± 0.003	0.957 ± 0.003
	0.95	**0.950** ± 0.003	**0.950** ± 0.003	**0.950** ± 0.003	0.951 ± 0.003	0.941 ± 0.003	0.947 ± 0.003
64	0.99	0.991 ± 0.003	0.991 ± 0.003	0.991 ± 0.003	**0.990** ± 0.001	0.987 ± 0.002	0.993 ± 0.000
	0.98	**0.982** ± 0.004	**0.982** ± 0.004	**0.982** ± 0.004	0.979 ± 0.006	0.972 ± 0.004	0.983 ± 0.002
	0.97	0.975 ± 0.004	**0.974** ± 0.003	**0.974** ± 0.003	0.969 ± 0.005	0.960 ± 0.005	0.977 ± 0.003
	0.96	**0.961** ± 0.002	**0.961** ± 0.002	**0.961** ± 0.002	0.957 ± 0.005	0.948 ± 0.005	0.969 ± 0.005
	0.95	**0.951** ± 0.002	**0.951** ± 0.002	**0.951** ± 0.002	**0.951** ± 0.005	0.935 ± 0.007	0.963 ± 0.005

Table 1: The impact of the number of output dimensions on marginal coverage rates.

As shown by Figure 3 and detailed by Table 1, when the number of output dimension is $m = 4$, the debiasing procedure transforms the biased methods ("ec", "hac", "vac") into well calibrated methods ("db-ec", "db-hac", "db-vac") regardless of how under-conservative or over-conservative they were before the debiasing step. However, when the number of dimensions is increased to $m = 64$, while the size of the calibration data is kept at $|\mathcal{D}_{\text{Cal}}| = |\mathcal{D}_{\text{Bias}}| = 10^4$, the debiased methods ("db-ec", "db-hac", "db-vac") become slightly over-conservative. This is not an inherent limitation of the method: it is the result of a compromise that we made to speed-up the re-calibration procedure, during implementation.

To avoid solving Equation 21 with a binary search for each of the (10^4) calibration items, we tried a discretized set with 200 possible values of $1 - \alpha'$ (see Equation 21). Thus,

the step size in this discretization was too big for 64 dimensions: the computed secondary non-conformity scores $\mathscr{S}_i$ were not actual infinima. These results are nonetheless valuable because they highlight the computational cost of the proposed debiasing procedure.

Interestingly, as shown by Figure 4, the efficiency results of these methods in high dimensions ($m = 64$) are close to the efficiency results in low-dimensions ($m = 4$). This suggests that the values of the copula-based conformal quantiles $s_{\alpha,d}$ in Equation 7 don't increase as fast as the number of output dimensions.

Efficiency results for different numbers of output dimensions

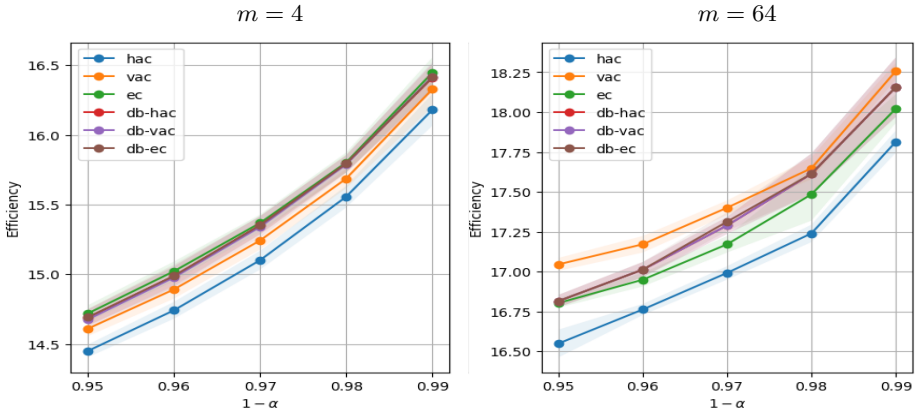

Figure 4: The impact of the number of output dimensions m on the size of the regions.

On the impact of the size of the data

To empirically assess the impact of the size of the calibration data sets $\mathcal{D}_{\text{Cal}}$ and $\mathcal{D}_{\text{Bias}}$ on the marginal coverage results, we generate the non-conformity scores $\boldsymbol{s}_i$ from a four dimensional copula (we set $m = 4$ in Algorithm 2) and we experiment with two data sizes: $|\mathcal{D}_{\text{Cal}}| = |\mathcal{D}_{\text{Bias}}| = 10^4$ and $|\mathcal{D}_{\text{Cal}}| = |\mathcal{D}_{\text{Bias}}| = 400$. Thus, we consistently use the same amount of data to fit the copulae and to debias them, but we test with a large sample in both cases: $|\mathcal{D}_{\text{Test}}| = 10^4$. In Table 2, the best results are in boldface and their format is "mean(Coverage) ± standard deviation".

$\lvert\mathcal{D}_{\text{Bias}}\rvert$	$1-\alpha$	db-ec	db-hac	db-vac	ec	hac	vac
10^4	0.99	**0.990** ± 0.002	**0.990** ± 0.002	**0.990** ± 0.002	**0.990** ± 0.001	0.987 ± 0.002	0.989 ± 0.002
	0.98	**0.980** ± 0.002	**0.980** ± 0.002	**0.980** ± 0.002	**0.980** ± 0.001	0.975 ± 0.002	0.978 ± 0.001
	0.97	**0.970** ± 0.002	**0.970** ± 0.002	**0.970** ± 0.002	0.971 ± 0.002	0.963 ± 0.002	0.967 ± 0.002
	0.96	**0.960** ± 0.003	**0.960** ± 0.002	**0.960** ± 0.003	0.961 ± 0.003	0.952 ± 0.003	0.957 ± 0.003
	0.95	**0.950** ± 0.003	**0.950** ± 0.003	**0.950** ± 0.003	0.951 ± 0.003	0.941 ± 0.003	0.947 ± 0.003
400	0.99	**0.992** ± 0.004	0.993 ± 0.003	0.993 ± 0.003	0.993 ± 0.003	0.993 ± 0.003	0.993 ± 0.003
	0.98	0.986 ± 0.006	0.986 ± 0.006	**0.980** ± 0.008	0.983 ± 0.005	0.983 ± 0.005	0.985 ± 0.003
	0.97	0.979 ± 0.005	0.979 ± 0.005	0.972 ± 0.006	0.974 ± 0.008	**0.970** ± 0.005	0.976 ± 0.006
	0.96	0.970 ± 0.009	0.970 ± 0.009	0.968 ± 0.011	0.962 ± 0.006	**0.960** ± 0.005	0.963 ± 0.006
	0.95	0.957 ± 0.008	0.957 ± 0.008	**0.950** ± 0.012	0.955 ± 0.010	0.951 ± 0.011	0.956 ± 0.010

Table 2: The impact of the size of the calibration data on marginal coverage rates.

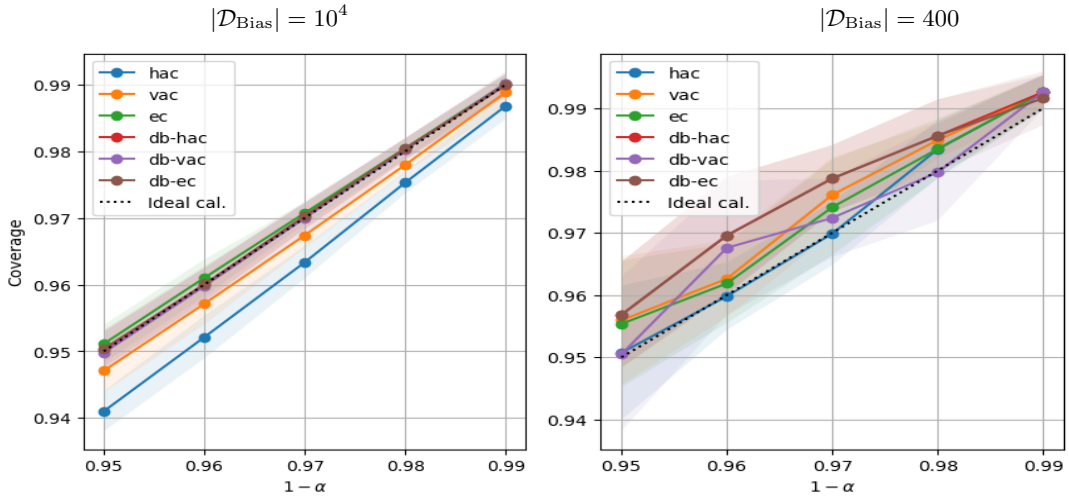

Figure 5: The impact of the size of the calibration data on marginal coverage rates.

As shown by Figure 5 and detailed by Table 2, when the size of the calibration data is small ($|\mathcal{D}_{\text{Bias}}| = 400$), the debiased methods ("db-ec", "db-hac", "db-vac") yield empirical coverage rates that are mostly greater than the user-specified joint confidence levels. They are over-conservative, the variance is higher, and the difference between the observed (marginal) coverage rate and the user-specified joint confidence $(1-\alpha)$ reaches 0.01.

This is due to two factors. First, according to Gupta et al. (2022), the (observable) marginal coverage rate is limited by the number ($|\mathscr{S}|$) of distinct secondary non-conformity scores $\mathscr{S}_i$ (see Equation 21) that are computed from the debiasing data set $\mathcal{D}_{\text{Bias}}$:

$$1-\alpha+\frac{1}{|\mathscr{S}|+1} \geq P(\boldsymbol{y}_{n+1} \in \boldsymbol{\Gamma}^{\alpha}_{n+1}) \geq 1-\alpha. \tag{28}$$

Second, using a discretized set of possible values of $1-\alpha'$ in Equation 21 yields secondary non-conformity scores $\mathscr{S}_i$ that are not actual infinima and reduces the number of distinct values of $\mathscr{S}_i$ by producing duplicates. Therefore, when $m = 4$ and $|\mathcal{D}_{\text{Bias}}| = 400$, the sample size is simply too small and discretizing the search space of the scores $\mathscr{S}_i$ is not a safe strategy for reducing the computational cost of the debiasing procedure.

Regarding the efficiency results, Figure 6 shows that the sizes of the inferred prediction regions $\boldsymbol{\Gamma}^{\alpha}_i$ slightly increase as the size of the calibration data $|\mathcal{D}_{\text{Cal}}| = |\mathcal{D}_{\text{Bias}}|$ decreases. Also, the debiased vine based method ("db-vac") seems to produce smaller prediction regions compared to the other debiased methods, but this remains an extreme over-conservative case with a high variance.

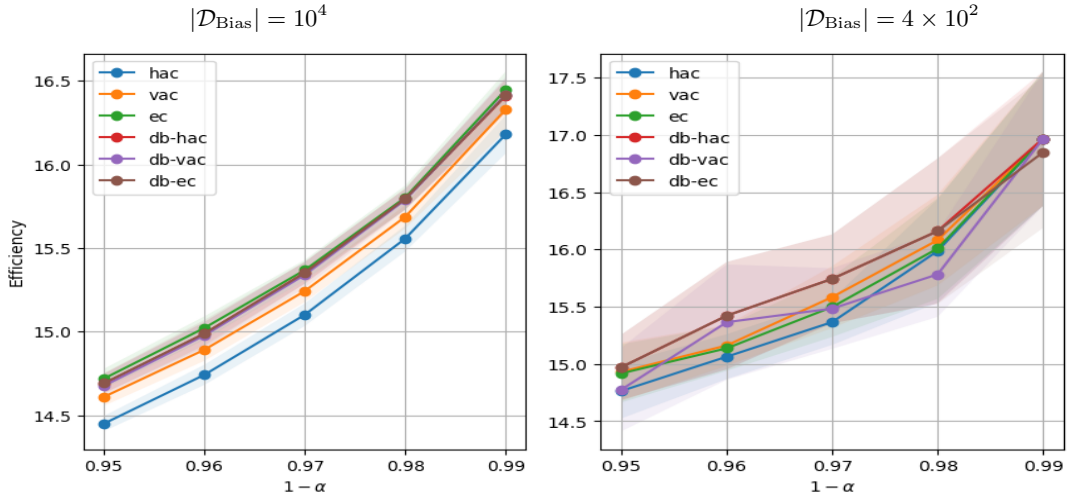

Figure 6: The impact of the size of the calibration data on the size of the regions.

In both the experiments on the number of output dimensions and the experiments on the size of the calibration data, the results of the debiased semi-parametric methods ("db-hac" and "db-vac") were mostly similar to the results of the debiased non-parametric method ("db-ec"). In view of this, one can expect similar performances if the semi-parametric are not estimated, but rather fully specified using (fairly) accurate prior knowledge. In this case, the amount of necessary data would be halved as $\mathcal{D}_{\text{Cal}}$ would not be required.

Furthermore, to reduce the difficulty of specifying the copula (using pior-knowledge), HAC(s) can be preferred as they are simpler asymmetric copula models: they are more accessible to non-expert users. Unlike the vine copulas, the HAC(s) do not involve conditional copulas and the number of their sub-copulas can be chosen by the user independently of the number of the output dimensions.

A practical application

To evaluate the baselines on real data, we use the Berkley DeepDrive benchmark (BDD100K). BDD100K (Yu et al., 2020) is a large, diverse, and crowd-sourced autonomous driving data set. We use it to predict the bounding boxes of traffic signs and traffic lights. The annotated part[3] is split as follows: 15562 instances are used for calibration, 15563 for de-biasing, 31069 for testing, and the rest is used to train the underlying detector[4]. To evaluate the variability of our results, the experiment is repeated 10 times. In Table 3, all the values are in percentages, the best results are shown in boldface and their format is "mean(Coverage) $\pm$ standard deviation".

As shown on the left of Figure 7 and in Table 3, the "ic" and the "ac" baselines are the most over-conservative approaches, and the "ec" baseline is the least over-conservative. In fact, the over conservativeness (i.e., the positive coverage gap) increases as the dependency

3. See https://doc.bdd100k.com/download.html#k-images
4. See https://docs.ultralytics.com/models/yolov8/

model get simpler and less flexible. Although the methods above the calibration line seem to verify Equation 1, they provide no finite-size guarantees and their inferred bounding box regions are larger, as shown on the right of Figure 7.

In contrast, the bounding box regions that are inferred with the de-biased baselines are significantly smaller and provably valid. Surprisingly, the sizes of the regions that are inferred with de-biased methods are very close despite the difference in complexity of the re-calibrated copulas. This is not a feature of the proposed de-biasing method: there is no proof that this would always be observed on any data set.

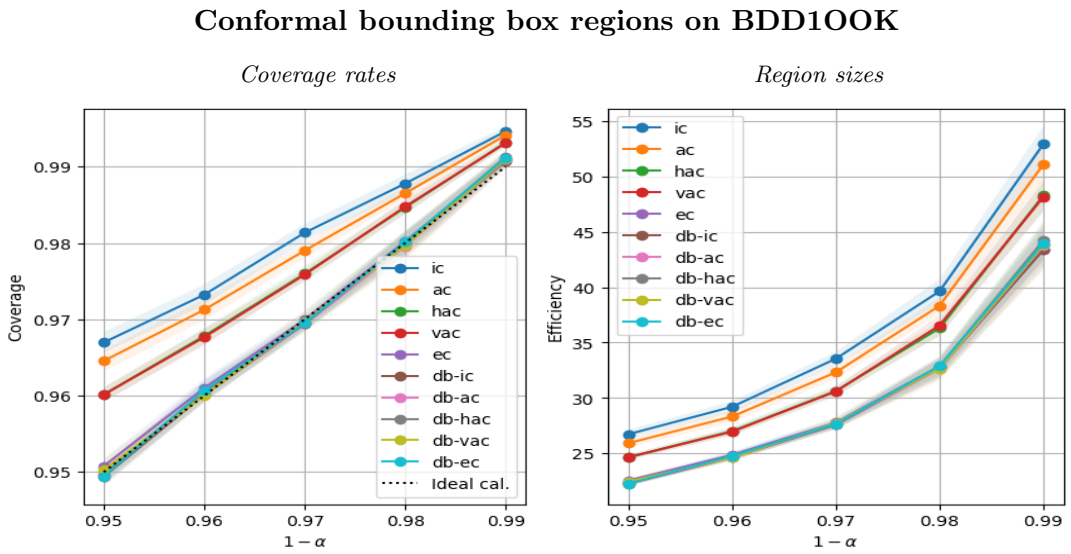

Figure 7: The empirical marginal coverage rates (left) and the efficiency (right).

$1-\alpha$	ac	db-ac	db-ec	db-hac	db-ic	db-vac	ec	hac	ic	vac
99	99.4 ± 0.1	**99.1** ± 0.1	**99.1** ± 0.1	**99.1** ± 0.1	**99.1** ± 0.1	**99.1** ± 0.1	**99.1** ± 0.1	99.3 ± 0.0	99.4 ± 0.1	99.3 ± 0.0
98	98.6 ± 0.1	**98.0** ± 0.1	**98.0** ± 0.1	**98.0** ± 0.1	**98.0** ± 0.1	**98.0** ± 0.2	**98.0** ± 0.1	98.5 ± 0.1	98.8 ± 0.1	98.5 ± 0.1
97	97.9 ± 0.1	96.9 ± 0.1	96.9 ± 0.1	96.9 ± 0.1	96.9 ± 0.1	**97.0** ± 0.1	**97.0** ± 0.1	97.6 ± 0.1	98.1 ± 0.1	97.6 ± 0.1
96	97.1 ± 0.1	**96.0** ± 0.1	96.1 ± 0.1	**96.0** ± 0.1	**96.0** ± 0.1	**96.0** ± 0.1	96.1 ± 0.1	96.8 ± 0.1	97.3 ± 0.1	96.8 ± 0.1
95	96.5 ± 0.1	**95.0** ± 0.1	94.9 ± 0.1	**95.0** ± 0.1	94.9 ± 0.1	**95.0** ± 0.2	95.1 ± 0.1	96.0 ± 0.1	96.7 ± 0.1	96.0 ± 0.1

Table 3: The detailed marginal coverage rates of CoCP approaches on BDD100K.

Regarding, the performance of the class of copulae (HAC) that is introduced in CoCP through this study, we can observe that its performance (empirical marginal coverage and efficiency) is extremely close to that of the vine-based baseline ("vac"), before and after the re-calibration process. This complex computer vision task (i.e., bounding box prediction for object detection) shows that vine-copulas do not always yield better performance compared to Hierarchical Archimedean Copulas. When the performance is comparable, HAC(s) can be preferable because they are simpler, easier to estimate and to easier interpret.

6. Conclusion

We presented a new approach to addressing the residual copula estimation (or specification) bias that has limited the statistical guarantees of copula-based conformal prediction to asymptotic validity. This new approach leverages nested prediction regions to transform copula-based conformal predictors into conformal predictors with finite-size joint coverage guarantees, i.e., with exact validity. Unlike the previous approach, our method can also be applied to semi-parametric copulas such as the hierarchical Archimedean copulas (HAC) that are firstly introducing in conformal prediction in this study.

The experimental results show that this debiasing procedure successfully transforms copula-based conformal prediction regions into exactly valid prediction regions for a variety of copula models and for multiple numbers of output dimensions. However, this recalibration procedure is sensitive to the size of the (additional) debiasing data set $\mathcal{D}_{\text{Bias}}$ and to the precision with which its secondary non-conformity scores $\mathscr{S}_i$ are computed. Compared to debiased vine copulas, debiased HAC(s) yield similar performances, but they are simpler dependency models because their interpretation and their specification are far more accessible to non-expert users.

Although this study has mainly focused on reaching exact (joint) validity with copula-based conformal prediction, future developments can generalize this debiasing method to semi-parametric dependency models that are not copulas, e.g., weighted averages of the dimension-wise confidence levels $(1-\alpha_1, \ldots, 1-\alpha_m)$. To reduce the size of the inferred regions, this method could also be (re)formulated as a minimization of the size of the inferred regions under the joint probability (or copula) constraint, following Zhang et al. (2023), with an additional nestedness constraint. To obtain adaptable prediction regions, copula-based outlier detection (Li et al., 2020) could be explored to design non-conformity scores that reflect the "strangeness" of each prediction. To formally study the concentration properties of copula-based statistical guarantees, empirical-copula processes $\sqrt{n}\left(\mathcal{C}_n(\boldsymbol{u}) - \mathcal{C}(\boldsymbol{u})\right)$ could be exploited (Fermanian et al., 2004; Bücher and Dette, 2010).

7. Acknowledgements

This research was supported by the UTC Foundation, the Safe AI Chair, and the Hauts-de-France region.

References

Léo Andéol, Thomas Fel, Florence De Grancey, and Luca Mossina. Confident object detection via conformal prediction and conformal risk control: an application to railway signaling. In *Conformal and Probabilistic Prediction with Applications*, pages 36–55. PMLR, 2023.

Anastasios N Angelopoulos, Stephen Bates, Emmanuel J Candès, Michael I Jordan, and Lihua Lei. Learn then test: Calibrating predictive algorithms to achieve risk control. *The Annals of Applied Statistics*, 19(2):1641–1662, 2025.

Tim Bedford and Roger M Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32:245–268, 2001.

Axel Bücher and Holger Dette. A note on bootstrap approximations for the empirical copula process. *Statistics & probability letters*, 80(23-24):1925–1932, 2010.

Claudia Czado and Thomas Nagler. Vine copula based modeling. *Annual Review of Statistics and Its Application*, 9(1):453–477, 2022.

Paul Deheuvels. La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance. *Bulletins de l'Académie Royale de Belgique*, 65(1):274–292, 1979.

Victor Dheur, Matteo Fontana, Yorick Estievenart, Naomi Desobry, and Souhaib Ben Taieb. A unified comparative study with generalized conformity scores for multi-output conformal regression. In *Forty-second International Conference on Machine Learning*, 2025.

Jean-David Fermanian, Dragan Radulovic, and Marten Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860, 2004.

Gery Geenens. Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association*, 109(505):346–358, 2014.

Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368, 2007.

Jan Górecki and Ostap Okhrin. *Hierarchical Archimedean Copulas*. Springer Nature, 2024.

Chirag Gupta, Arun K Kuchibhotla, and Aaditya Ramdas. Nested conformal prediction and quantile out-of-bag ensemble methods. *Pattern Recognition*, 127:108496, 2022.

Marius Hofert, Martin Mächler, and Alexander J McNeil. Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique*, 154(1):25–63, 2013.

Rafael Izbicki, Gilson Shimizu, and Rafael B Stern. Cd-split and hpd-split: Efficient conformal regions in high dimensions. *Journal of Machine Learning Research*, 23(87):1–32, 2022.

Michal Klein, Louis Bethune, Eugene Ndiaye, and Marco Cuturi. Multivariate conformal prediction using optimal transport. *arXiv preprint arXiv:2502.03609*, 2025.

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

Zheng Li, Yue Zhao, Nicola Botta, Cezar Ionescu, and Xiyang Hu. Copod: copula-based outlier detection. In *2020 IEEE international conference on data mining (ICDM)*, pages 1118–1123. IEEE, 2020.

Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, 120:108101, 2021.

Thomas Nagler and Claudia Czado. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, 151: 69–89, 2016.

Ostap Okhrin and Alexander Ristig. Hierarchical archimedean copulae: the hac package. *Journal of Statistical Software*, 58:1–20, 2014.

Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *Machine learning: ECML 2002: 13th European conference on machine learning Helsinki, Finland, August 19–23, 2002 proceedings 13*, pages 345–356. Springer, 2002.

Ji Won Park and Kyunghyun Cho. Semiparametric conformal prediction. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

Mohsen Rezapour. On the construction of nested archimedean copulas for d-monotone generators. *Statistics & Probability Letters*, 101:21–32, 2015.

Max Sampson and Kung-Sik Chan. Flexible conformal highest predictive conditional density sets. *arXiv preprint arXiv:2406.18052*, 2024.

M Sklar. Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, volume 8, pages 229–231, 1959.

Sophia Huiwen Sun and Rose Yu. Copula conformal prediction for multi-step time series prediction. In *The Twelfth International Conference on Learning Representations*, 2024.

Gauthier Thurin, Kimia Nadjahi, and Claire Boyer. Optimal transport-based conformal prediction. *arXiv preprint arXiv:2501.18991*, 2025.

Alexander Timans, Christoph-Nikolas Straehle, Kaspar Sakmann, Christian A Naesseth, and Eric Nalisnick. Max-rank: Efficient multiple testing for conformal prediction. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

Mark J Van der Laan, Sherri Rose, et al. *Targeted learning: causal inference for observational and experimental data*, volume 4. Springer, 2011.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer Nature, 2022.

Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020.

Ruiyao Zhang, Ping Zhou, and Tianyou Chai. Improved copula-based conformal prediction for uncertainty quantification of multi-output regression. *Journal of Process Control*, 129: 103036, 2023.