POLITECNICO
MILANO 1863
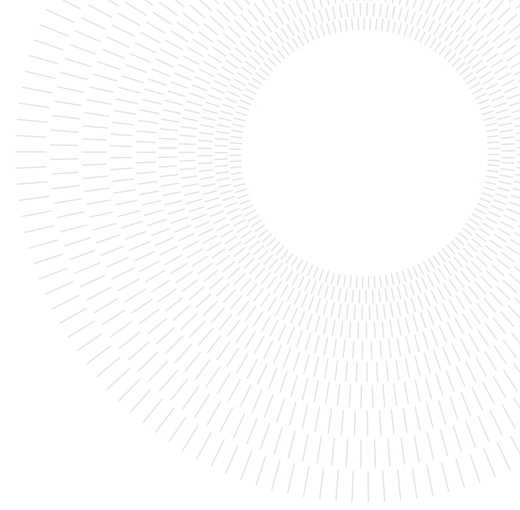
Conformal prediction and copula based methods for profile monitoring

Niccolò Donadini, Student ID: 10622527

Advisor:
Prof. Simone Vantini

Co-advisors:
Teresa Bortolotti

Academic year:
2023-2024

Abstract: In many industrial processes the intrinsic complexity has increased over years and, in parallel, the necessity to monitor the processes, in order to detect anomalous patterns has become more and more essential. For this reason, profile monitoring, a recent field of research in Statistical Process Control, is attracting the interest of many researchers. In this work, our aim is to propose an additional tool, a novel conformal based p-value function, that integrates the traditional approach used in profile monitoring. Our methodology, based on Conformal prediction and Copula estimation, is able to recognize not only if an anomaly is present, but also which part of the domain is responsible for that strange behaviour. To check the goodness of our method, we go through applications to literature datasets and simulations study. A real case study, namely the analysis on Vertical Density Profiles (VDP), is presented in order to illustrate the potential of our procedure in real applications.

1. Introduction

The continuous evolution of technology in the industrial field provides the possibility to collect an increasing amount of information about a process of interest, giving the opportunity to study it in a deeper and more precise way. The high level of complexity, characterizing the majority of the industrial operations, leads to organize all the gathered information in dataset with a huge number of columns; it's the case of functional dataset, where each row corresponds to a function that takes values on a specific interval, for more details we refer to Ramsay, Silverman (2005) and/or Horvath, Kokoszka (2012). This allows us to use various tools in the framework of functional analysis to develop a new method for anomaly detection.

The main purpose of our research is to identify strange, unexpected behaviour in the resulting product of an industrial process of interest and, consequently, to asses its overall strength. In the literature, this type of analysis is called profile monitoring and it is divided in 2 parts: during the first one (usually named phase I, or design phase) a set of in-control profiles is used to build a control chart with related threshold, while phase II is aimed at monitoring new profiles by comparing them with the previously obtained control chart.

In this paper we propose a new method for profile monitoring based on conformal prediction. Given a starting functional dataset, which can be seen as the set of in-control profiles, we associate to every new profile a specific function, the so-called p-value function, that assumes values between zero and one. The interpretation of the value of this function in a specific point of the domain is the following: a high value, close to one, indicates the presence of a standard behaviour while a low value, near zero, highlights an unusual pattern and allows us

to label that point as an anomaly. This new p-value function is built starting from the conformal prediction bands proposed in Diquigiovanni, Fontana and Vantini (2021). We are aware that a similar method already exits in the literature. While the statistical interpretation is the same of our p-value function, the way in which it is constructed is completely different: for a specific point of the domain, its corresponding value is computed focusing only on the values of the in-control functions *on that point*. Instead, in our proposal, the value in a specific point is influenced by what is happening in the rest of the domain, thanks to the use of the conformal prediction bands. We see our new solution as a first extension, over the domain, of the benchmark case which we refer to as the not adjusted case.

After having tested the p-value function on some of the famous dataset available in the literature, by comparing it with the not adjusted case, a weak aspect of this procedure comes out. Indeed, the p-value function is helpful in detecting anomalous behavior related to the values of a specific function, but it is useless if the strange pattern occurs with an higher order of derivative. A clever way to solve this problem is to consider not only the functions alone, but also the functions' derivatives. Inspired by the work of Messoudi, Destercke and Rousseau (2021), we adopt the copula approach to ensure a specific joint coverage, given the marginal coverage used to build the conformal regions respectively for the functions and their derivatives. This is the second extension, done at the codomain level, that is going to be discuss in our studies.

The final p-value function, including both the first and the second extensions, is applied, first of all, to the datasets of the literature and, successively, to a real case study, the Vertical density profile (VDP). A generic VDP profile measures the density over the vertical axes of a given particleboard. As a consequence, since there's a strict relation between the mechanical properties of the particleboard and the VDP curve, it is a common practice to detect possible failures of the final product by analyzing the behaviour of the corresponding VDP profile.

In detail the paper is structured as follows. Section 2 presents the methodology of the p-value function, focusing on the first extension. In Section 3, we compare the performance of our method with respect to the not-adjusted method, using famous datasets of the literature on Functional Data Analysis (e.g. Ramsay and Silverman, 2005) and a simulation study. Section 4 introduces the methodology behind the second extensions, while in Section 5 we show the importance of taking into consideration higher order derivatives, following the copula adjustment, by applying the procedure to the well-known datasets from the literature and by presenting a second simulation study. Section 6 discusses the real case study of VDP, while section 7 concludes the paper.

2. Methodology

A fundamental part of every industrial process is the one focused to control its quality and to find out strange patterns as soon as they arise. The methodology proposed belongs to the well known field of profile monitoring and presents a new way of checking the behaviour of a given industrial operation with respect to an in-control sample. Our method is based on Conformal prediction, that is a non parametric approach aimed at building prediction sets, under the weak assumption of exchangeable data. During recent years, many authors (L´opez-Pintado and Romo 2009, Lei et al. 2015) have stated out that in the functional framework, the one of our analysis, the prediction set should be a particular one, namely the prediction band. Formally, a band is defined as

$$\{y \in \mathcal{Y}(\mathcal{T}) : y(t) \in B_n(t), \quad \forall t \in \mathcal{T}\} \tag{1}$$

where $B_n(t) \subset \mathbb{R}$ is an interval for each $t \in \mathcal{T}$. In the framework of Conformal prediction (Vovk et al. 2005), following the notation of Diquigiovanni, Fontana and Vantini (2021), we consider an initial dataset made of n *i.i.d* random functions $Y_1, \ldots, Y_n \sim P$ and an independent function $Y_{n+1} \sim P$ and we build a prediction band $C_{n,1-\alpha} := C_{n,1-\alpha}(Y_1, \ldots, Y_n)$ for Y_{n+1}. The constructed prediction band is a set such that

$$\mathbb{P}(Y_{n+1} \in C_{n,1-\alpha}) \geq 1 - \alpha \tag{2}$$

for any significance level $\alpha \in (0, 1)$. In the rest of the paper we will use the term *coverage* to indicate $\mathbb{P}(Y_{n+1} \in C_{n,1-\alpha})$ and, by using the Conformal approach, we are guaranteed to build prediction band $C_{n,1-\alpha}$ with an associated coverage greater or equal to $1 - \alpha$, for any $\alpha \in (0, 1)$. Precisely, we will consider the Semi-Off-Line Inductive Conformal framework, well known as *Split Conformal*, that is more efficient from a computational point of view than the Transductive Conformal method, also called *Full Conformal* (firstly proposed in Papadopoulos et al. 2002).

Let us explain how the prediction band $C_{n,1-\alpha}$ is obtained under the *Split Conformal* framework, following the work of Diquigiovanni, Fontana and Vantini (2021). An initial dataset made of n functions $y_1, \ldots, y_n$ is randomly split in 2 parts: the training set $\{y_h : h \in \mathcal{I}_1\}$ and the calibration set $\{y_h : h \in \mathcal{I}_2\}$, where $\mathcal{I}_1$ and $\mathcal{I}_2$ are two set of indexes obtained by the random split of $\{1, \ldots, n\}$. Let us define $|\mathcal{I}_1| = m$ and $|\mathcal{I}_2| = l$, of course $m + l = n$. Now we define the *nonconformity measure* as a measurable function $A(\{y_h : h \in \mathcal{I}_1\}, y)$ taking

values in $\mathbb{R}$, whose quantify how different $y \in \mathcal{Y}(\mathcal{T})$ is from the training set and the *nonconformity scores* as $R_j := A(\{y_h : h \in \mathcal{I}_1\}, y_j)$ for $j \in \mathcal{I}_2$. We are left with the decision of which measurable function we should use as *nonconformity measure* to continue our studies. Coherently, we follow what Diquigiovanni, Fontana and Vantini propose in their paper and we set:

$$A(\{y_h : h \in \mathcal{I}_1\}, y) = \sup_{t \in \mathcal{T}} \left| \frac{y(t) - g_{\mathcal{I}_1}(t)}{s_{\mathcal{I}_1}(t)} \right| \tag{3}$$

where $g_{\mathcal{I}_1}(t) : \mathcal{T} \to \mathbb{R}$ is a point predictor summarizing the information of the training set $\{y_h : h \in \mathcal{I}_1\}$, in particular it is the sample functional mean. $s_{\mathcal{I}_1}(t) : \mathcal{T} \to \mathbb{R}_{>0}$, based on the training set $\{y_h : h \in \mathcal{I}_1\}$, is the so called *modulation function*, that forces the prediction bands to adapt their width according to the local variability of the *training functions*. Consequently, the *nonconformtity scores* of the calibration set are the following:

$$R_J^s := \sup_{t \in \mathcal{T}} \left| \frac{y_j(t) - g_{\mathcal{I}_1}(t)}{s_{\mathcal{I}_1}(t)} \right| \tag{4}$$

with $j \in \mathcal{I}_2$. Therefore, the prediction band, obtained by applying the *Split conformal* algorithm with the previous mentioned choice of *nonconformity measure*, is:

$$C_{n,1-\alpha}^s := \{y \in \mathcal{Y}(\mathcal{T}) : y(t) \in [g_{\mathcal{I}_1}(t) - k^s s_{\mathcal{I}_1}(t); g_{\mathcal{I}_1}(t) + k^s s_{\mathcal{I}_1}(t)]\} \tag{5}$$

where k^s is the $[(l+1)(1-\alpha)]^{th}$ smallest value in the set of the *nonconformity scores* of the calibration set, i.e. $\{R_h^s : h \in \mathcal{I}_2\}$. For the rest of our discussion let us set $C_{n,1-\alpha}^s = C_{n,1-\alpha}$ because we will keep fixed the modulation function $s_{\mathcal{I}_1}$ and equal to the sample functional standard deviation of the training set $\{y_h : h \in \mathcal{I}_1\}$. Now that we have explained how to build the *Conformal prediction band* for a specific value of $\alpha \in (0,1)$, let's proceed with the cornerstone of our research: the *p-value function*. Given a dataset $\mathcal{D}$, made of n functions $y_1, \ldots, y_n$ defined on $\mathcal{T} \to \mathbb{R}$, and a completely new function $y_{n+1} : \mathcal{T} \to \mathbb{R}$, our objective is to compute $p : \mathcal{T} \to [0,1]$, the *p-value function* of the function y_{n+1} with respect to the dataset $\mathcal{D}$. This *p-value function* is a new way to quantify how unusual is the function y_{n+1} compared to $y_1, \ldots, y_n$. In particular, for a specific point of the domain $t \in \mathcal{T}$, a high value, close to one, of $p(t)$ means that in that point the function y_{n+1} behaves like the other functions in $\mathcal{D}$. On the other hand, if $p(t)$ is close to zero then the function y_{n+1} will be labeled as anomalous in t. The *p-value function* may play a crucial role in the framework of profile monitoring, since it tells not only if a given function is an anomaly, but also which parts of the function's domain are responsible of such unexpected behaviour and how much that parts are strange, on a scale between [0,1].

After having introduced the p-value function, we can explain how it is actually derived from the previous mentioned conformal regions. We recall that we are considering a functional dataset $\mathcal{D}$ and a function y_{n+1} for which we want to compute the *p-value function* p. The idea at the base is quite simple: given a specific value of $\alpha \in [0,1]$, let us construct the associated conformal prediction band $C_{n,1-\alpha}$ from the dataset D and compare it with the function y_{n+1}. Whether we find some points over the domain, such that y_{n+1} intersects the upper or the lower bound of $C_{n,1-\alpha}$, then we will assign to those points a p-value of α. Of course we must repeat that procedure $\forall \alpha \in [0,1]$. Formally, we define the entire method as follows:

$$\forall t \in \mathcal{T} \quad p(t) = \min_{\alpha \in [0,1]} \alpha \quad s.t.$$
$$(u_{1-\alpha} - y_{n+1})^+(t) = 0 \quad \lor \quad (y_{n+1} - l_{1-\alpha})^+(t) = 0 \tag{6}$$

where $u_{1-\alpha}$ and $l_{1-\alpha}$ are respectively the upper and the lower bound of the conformal region $C_{n,1-\alpha}$.

It is appropriate to remark that our proposal is not a completely new one. Indeed a similar technique is already present in the literature. We hereafter refer to the literature benchmark as *not adjusted* p-value function. For a given point $t \in \mathcal{T}$, the corresponding value of the *not adjusted* p-value function, $p_{not\ adjusted}(t)$, is obtained by taking into consideration $y_{n+1}(t)$ and the values that the other functions, $y_1, \ldots, y_n$, assume in t. In other words, $p_{not\ adjusted}(t)$ indicates how strange is the function y_{n+1} from the others only focusing on what is happening in t, it does not care about the global behaviour over the domain, an aspect, on the contrary, that is very important in our version of the p-value function. For this reason, from now on, we will refer to our proposal with the name *adjusted* p-value function, highlighting the fact that an "adjustment" at a global level is always present, and to the literature's variant with the name *not adjusted* p-value function, putting evidence on the fact that the domain adjustment, induced by Conformal prediction, is not consider.

Concerning the not adjusted p-value function, its formal definition is the following:

$$\forall t \in \mathcal{T} \quad p_{not\ adjusted}(t) = \min_{\alpha \in [0,1]} \alpha \quad s.t.$$
$$(u_{1-\alpha} - y_{n+1})^+(t) = 0 \quad \lor \quad (y_{n+1} - l_{1-\alpha})^+(t) = 0 \tag{7}$$

3

with $u_{1-\alpha}$ and $l_{1-\alpha}$ that are, respectively, the upper and the lower bound of the *not adjusted* region for a given $\alpha \in [0, 1]$. It is worth underlining that these upper/lower bound are computed point after point by taking into consideration only the local trend.

3. Application and simulation study

3.1. Application to literature datasets

The next step of our analysis is to make a comparison between the two version of the *p-value function*. In order to show the improvements, that characterize our proposal with respect to the literature's one, we are going to apply both methods to some famous functional datasets, namely the Berkeley Growth Study data and the CanadianWeather data. Fig. 1 shows a plot of theese two data, giving an idea of the type of functions we are going to use for our aim.
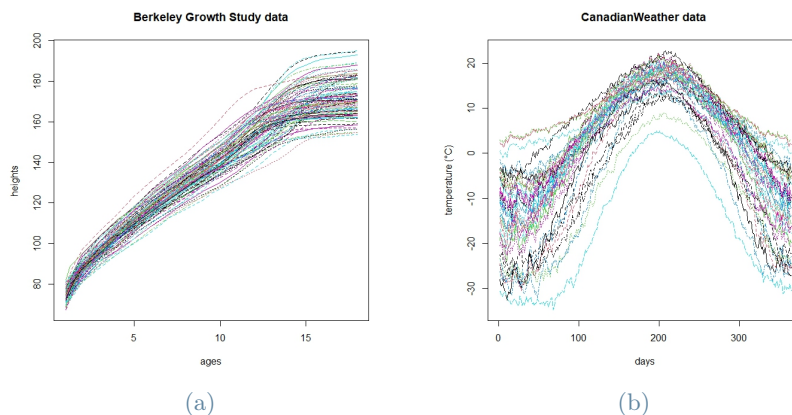

(a)   (b)

Figure 1: An overview of the data used to compare the p-value function adjusted with respect to the p-value function point_wise: Berkeley Growth Study data (a) and CanadianWeather data (b)

The first one we deal with is the Berkeley Growth Study data, taken from the R library fda; it contains the heights of 39 boys and 54 girls and the corresponding ages, included in a range from 1 to 18 years, at which the 31 measurements were collected. Fig.2 shows the results obtained considering the function relative to the boy05, i.e. the fifth curve of the boys dataset. For visual reasons, we decide to display in one plot the boy05's function over the conformal regions and in another one the relative p-value function, for both the adjusted and not adjusted cases. The chosen function assumes quite low values in the central part of the domain, as a consequence the two p-value functions have lower values in that part. In general, it is clear that the *adjusted* p-value function takes higher values compared to the not adjusted one. The reason is that the not adjusted regions tend to be narrower with respect to the adjusted ones.
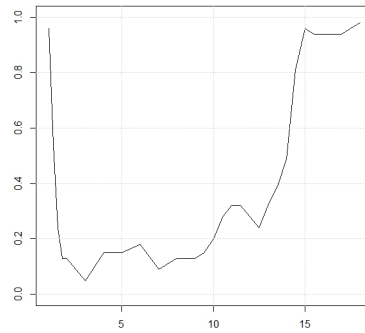
4
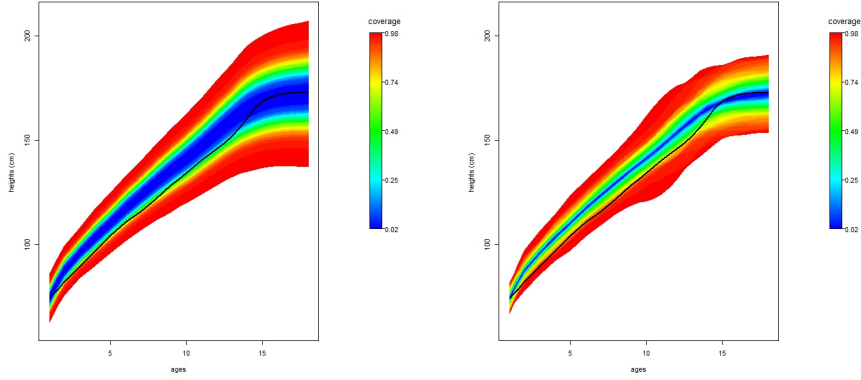
(a) boy05's function over the adjusted conformal regions

(b) boy05's function over the not adjusted conformal regions
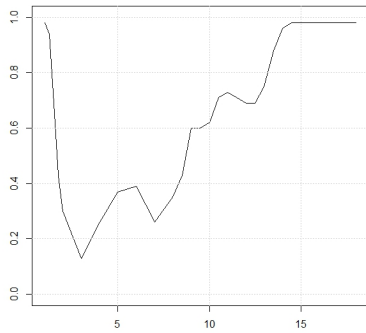
(c) p-value function adjusted

(d) p-value function not adjusted

Figure 2: Berkeley Growth Study data, we consider the function of the boy05, that is plotted over the conformal regions in the two plot at the top (panel a,b).The two corresponding p-value functions are shown in the plots at the bottom: adjusted case(c), not adjusted case (d)

Another dataset used to test the differences between the 2 versions is the CanadianWeather dataset, part of the library fda; it contains the daily measurements of temperature and precipitation at 35 different locations in Canada, averaged in the years between 1960 and 1994.For the sake of simplicity, we will consider only the data about temperature. Accordingly, our dataset is made of 35 functions that assume 365 temperature values each, one for every day in a year. Looking at the Fig.3, we notice that, in both cases, the function intersects the outer regions, characterized by the hot colors, at the extremes of the domain. As a consequence, these part are associated to small values of the two p-value functions. On the other hand, we observe a peak in the central part of the two p-value functions, higher in the adjusted case, meaning that in this part the function encounter the inner regions, highlighted by cold colors. As before, we register that the *adjusted* p-value function takes higher values, globally, than the *not adjusted* one. In addition, we want to remark that the *not adjusted* version, since it considers only the point-wise evaluation of the function and not its behaviour over the entire domain, is more sensible to small local deviations, that not necessarily correspond to abnormal patterns, with respect to the *adjusted* p-value function.

(a) Inuvik's function over the adjusted conformal regions

(b) Inuvik's function over the not adjusted conformal regions

(c) p-value function adjusted
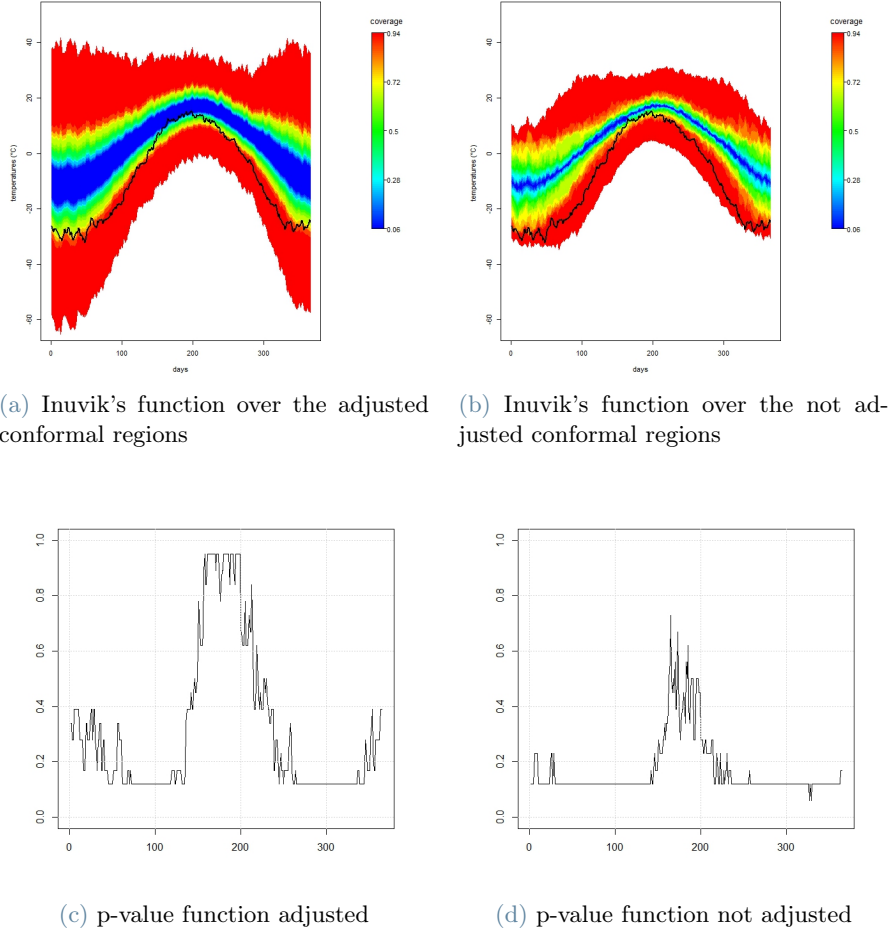
(d) p-value function not adjusted

Figure 3: CanadianWeather dataset, we consider the function related to Inuvik (northwest of Canada), that is plotted over the conformal regions in the two plot at the top (panel a,b).The two corresponding p-value functions are shown in the plots at the bottom: adjusted case (c), not adjusted case (d)

Lastly, we present here a peculiar case of anomaly. It is build from the Berkeley Growth Study data, considering only the functions related to the boys. What makes this case so particular is the fact that we are considering an anomalous function, that we get by adding noise, over the entire domain, to the boy04's function. By doing this, we are interested in studying how the two versions of the p-value functions react to an unusual pattern. Fig. 4 shows that in the adjusted case the black function always stays in the cold colored regions, implying high values of the related p-value functions, which goes under the value of 0.8 just for a total of three times. Passing to the not adjusted case, one can easily observe that the regions are, as usual, narrower and more irregular. Consequently, the anomaly touches the outer regions very often and this is the reason why the *not adjusted* p-value function goes repeatedly from high to low values and vice versa. Between the two p-value functions, we prefer the adjusted one because, by taking high values overall, it puts evidence on the fact that the function, even though anomalous, conforms to the others, under the aspect of the assumed values. Nevertheless, we can not be completely satisfied by the *adjusted* p-value function. Indeed, its values, being always very high, do not detect the presence of an anomalous profile. For this reason, we propose a further extension to this methodology, in order to solve issues of this kind.

(a) anomalous function over the adjusted conformal regions

(b) anomalous function over the not adjusted conformal regions

(c) p-value function adjusted
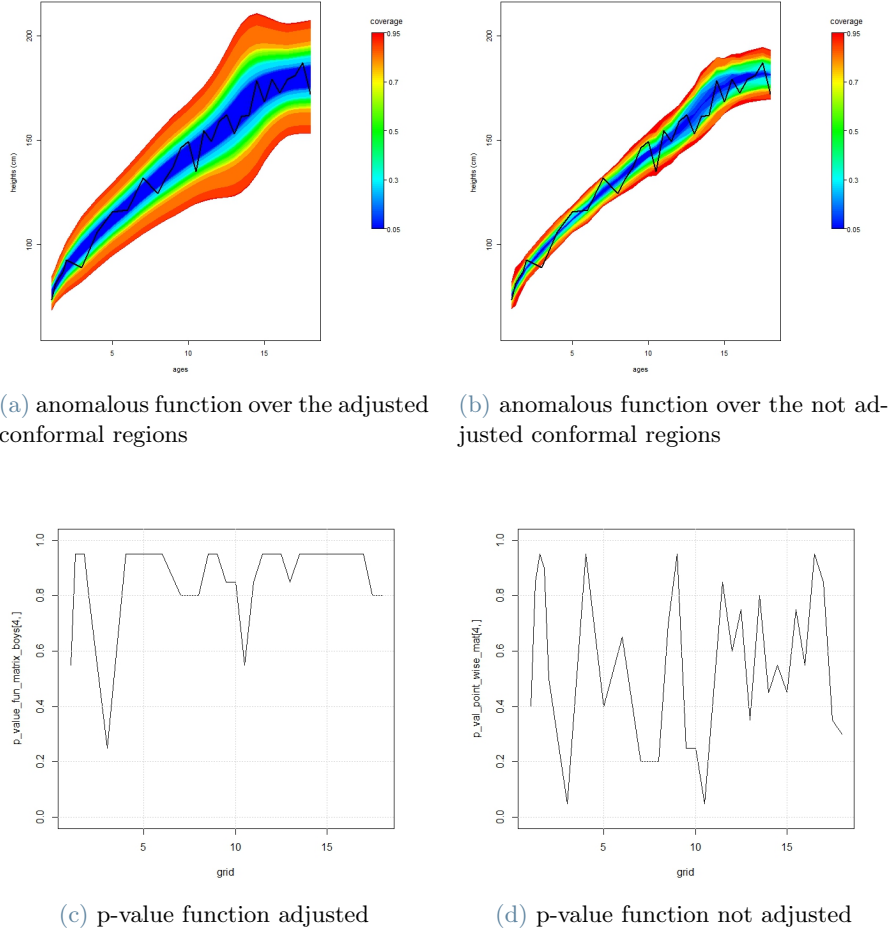
(d) p-value function not adjusted

Figure 4: Berkeley Growth Study data, but only the boys functions. We build an anomaly by adding noise to the boy04 function. This anomaly is plotted over the conformal regions in the two plot at the top (panel a,b).The two corresponding p-value functions are shown in the plots at the bottom: adjusted case (c), not adjusted case (d)

3.2. Simulation study

The comparison between the two versions of the p-value function goes on with a simulation study, which is focused in analyzing the validity of the two methods. In particular, we expect that the *adjusted* p-value function will ensure the typical coverage of a conformal prediction method: for a fixed value of coverage $1 - \alpha$, the percentage of *adjusted* p-value function with all values grater or equal to α must be near to $1 - \alpha$; those *adjusted* p-value function corresponds to the functions that are always inside the $1 - \alpha$ conformal prediction region that, from the theory, are known to be $(1 - \alpha)\%$. On the other hand, since the conformal prediction adjustment is missing, there are no guarantees to reach a right coverage by the *not adjusted* p-value function. In the first place, we consider the case of sinusoidal functions with amplitude and phase variation. More precisely, we build a fixed dimension test set, made of 200 functions. For each of them, we compute the corresponding p-value function with respect to a training set, the set used to compute the prediction bands, with an increasing dimension. The aim is to show the differences between the adjusted and not adjusted case, in term of validity, when the dimension of the calibration set gets bigger. After having obtained all the p-value functions of the test set, we perform the following evaluation: for a given value of coverage $1 - \alpha$, (consider for example $1 - \alpha = 0.8$), we calculate the number of p-value functions in the test set with all the values greater or equal to 0.2; we expect this number to be nearer to the 80% of the test set dimension as the calibration set increases. Obviously, we take into consideration different values of $1 - \alpha$ in the interval [0,1]. For what concerns the training and the test set, the sinusoidal functions are generated by the same model:

$$y_m(t) = A_m \sin(t + \phi_m) + \epsilon_m(t), \qquad m = 1, \dots, M, \quad t \in [0, 2\pi] \qquad (8)$$

where $A_m \sim \mathcal{U}(0, 100)$ and $\phi_m \sim \mathcal{U}(-1, 1)$ are, respectively, the random amplitude and the random phase

terms; $\epsilon_m(t) \sim \mathcal{N}(0, 0.2)$ is the noise term. M is fixed and equal to 200 for the test set, while it increases on a logarithmic scale, i.e. $M = 4, 8, 16, 32, 64, 128$, in the case of the training set. In order to obtain significant estimates, we repeat the procedure $B = 100$ times for each training set and we take the mean over the resulting values. The results are shown in fig.5 where for each value of M we have a separated plot. Concerning the adjusted case, it is clear that the estimates become more and more accurate as M gets bigger, the expected behaviour is reached at $M = 32$. On the other side, the coverage is underestimate in a bad way and the increasing of the dimension does not help to reach acceptable values. Indeed, the green triangles lies far below the red line, which represent the target, also for the highest values of M, indicating that the coverage, in the not adjusted case, doesn't get better when more functions are available. Additionally, for the lowest value of M, let's say $M = 4, 8, 16$, it is easy to see that the blue circles are positioned at the reachable levels of the corresponding M, i.e. $\{\frac{1}{l+1}, \frac{2}{l+1}, \dots, 1\}$ with $l = \frac{M}{2}$ and this is coherent with the fact that our method is based on conformal prediction bands.
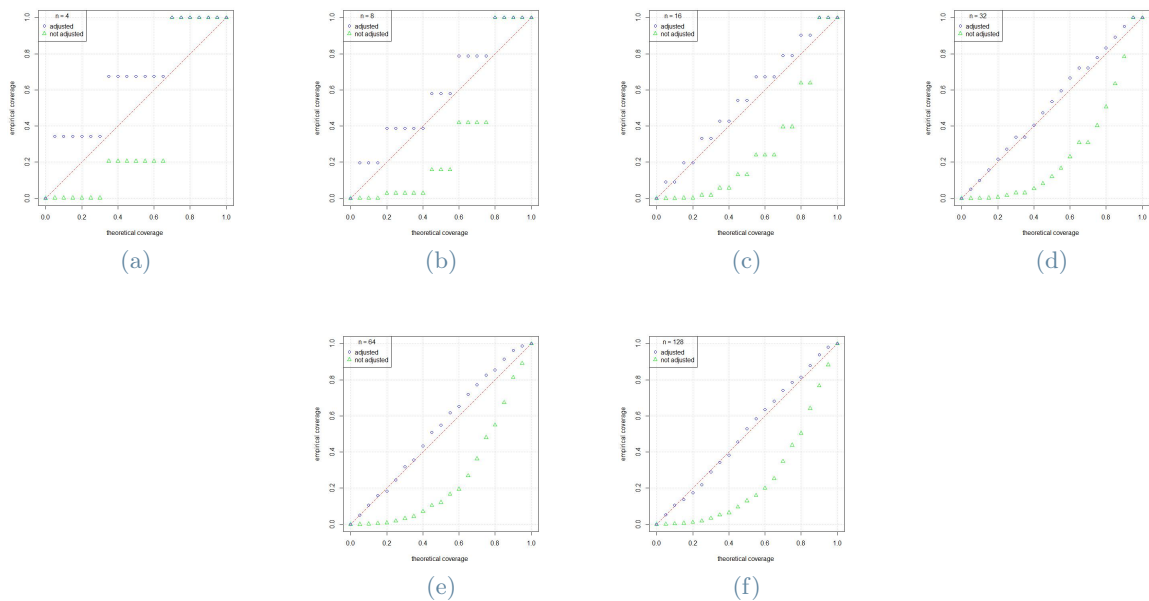

Figure 5: Simulation study considering amplitude and phase variation, the test is fixed at 200 functions, while the dimension of the calibration set increases on the logarithm scale: $M = 4, 8, 16, 32, 64, 128$. Each figure represent the behaviour of the two versions of the p-value function for a specific value of M, following the scale above. In particular, the blue circles refer to the adjusted case, while the green triangles indicate the estimate in the not adjusted version.

A second simulation case is presented, this time taking into consideration splines, that are piecewise polynomials continuous at interior nodes. For our purpose, we set four nodes at 0.2, 0.4, 0.6, 0.8 in the interval $[0, 1]$. The degree is equal to 3, which give us cubic splines, and the coefficients that multiply each basis are sampled from a uniform distribution with varying range. By using this trick, we manage to generate functions with different variability over the domain. The computed quantities are the same as the previous case, the general specifications doesn't change too. Obviously, the difference is in the model we used to generate the functions:

$$y_s(t) = \sum_{j=1}^{k} c_j^s \mathcal{P}_j(t), \qquad s = 1, \dots, S, \quad t \in [0, 1] \tag{9}$$

where $k = 8$ is the number of basis considered, which is related to the previous mentioned choices of knots and degree. $c_j^s \sim \mathcal{U}(0, b_j)$ is the coefficient associated to the jth basis. Let's underline that the second extreme of the above uniform distribution is different for every basis, in this way we modify the variability over the domain. $\mathcal{P}_j(t)$ is the jth basis evaluated in t. S will be equal to 200 for the test set, while it will increase on a logarithm scale in the case of the training set. Both the two *p-value function* versions have the general behaviour observed in the previous simulation study. Indeed, the fig. 6 shows that the coverage of the adjusted case gets better as S increases, reaching the expected level when just 32 functions are considered in the training set. On the contrary, the not adjusted case does not improve when S gets bigger, its estimate remain far below from the desirable target for every S. The reachable level, for low values of S, are highlighted also in this simulation study by the blue circles, that represent the estimates of the adjusted case.
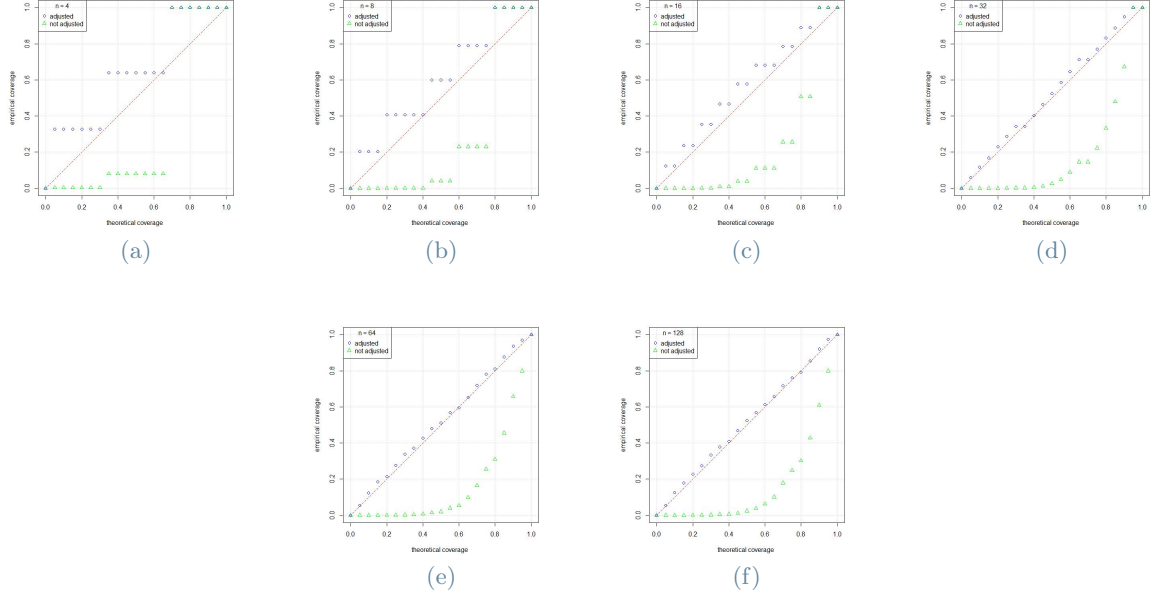
Figure 6: Simulation study considering splines with four nodes at 0.2, 0.4, 0.6, 0.8 in the interval [0,1] and degree equal to 3. Each generated function is the linear combination of the spline basis, with coefficients sampled from a uniform distribution with different range for every basis. The test set is fixed at 200 functions, while the dimension of the training set increases on the logarithm scale: $S = 4, 8, 16, 32, 64, 128$. Each figure represent the behaviour of the two versions of the p-value function for a specific value of S. In particular, the blue circles refer to the adjusted case, while the green triangles indicate the estimate in the not adjusted version.

4. Methodology, second part

Up to now, we can infer that our version of the p-value function is better in detecting strange pattern than the one already available in the literature, that is influenced in a too strong way from what is happening point-wise. Nevertheless, we showed, at the end of Section 3.1, that the *adjusted* p-value function has an important limit: it doesn't recognize a function as anomalous if the unusual behaviour is at higher order of derivatives. Going back to Figure 4a, one can see how strange the considered function is: it goes rapidly up and down while the other functions have a monotonically increasing trend. Although, the corresponding p-value function, Figure 4c, has high values all over the domain, not putting evidence on the fact that we are in presence of an anomaly. It is clear that something is missing, for this reason we propose a further extension to our method, at a codomain level. In particular, we decided to compute the p-value function also for higher order of derivatives, in this way we are able to spot unexpected trend at the 1st order, such as the one presented in the previous discussion, or greater. By extending the codomain of our method, we need to adjust the p-value function a second time, to take into consideration the fact that we are building multivariate conformal prediction bands with a specific *global* coverage $1 - \alpha_g$. Let $\mathbf{D}$ be a multivariate functional dataset composed by n multivariate random functions $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. For each $i = 1, \ldots, n$, $\mathbf{Y}_i = (Y_i^0, Y_i^1, \ldots, Y_i^m)$ where the first, the second and the m+1-th component are, respectively, the random function (derivate of order 0), its first derivative and its derivative of order m. Let us indicate with $\mathbf{Y}_{n+1} = (Y_{n+1}^0, Y_{n+1}^1, \ldots, Y_{n+1}^m)$ a new independent observation; the multivariate conformal prediction region $\mathbf{C}_{n,1-\alpha_g}$ is such that:

$$\mathbb{P}(\mathbf{Y}_{n+1} \in \mathbf{C}_{n,1-\alpha_g}) \geq 1 - \alpha_g \qquad i.e.$$
$$\mathbb{P}(Y_{n+1}^0 \in C_{n,1-\alpha_0}, Y_{n+1}^1 \in C_{n,1-\alpha_1}, \ldots, Y_{n+1}^m \in C_{n,1-\alpha_m}) \geq 1 - \alpha_g \qquad (10)$$

where $C_{n,1-\alpha_0}$ is the univariate conformal prediction region for the functions, built using $Y_1^0, \ldots, Y_n^0$; $C_{n,1-\alpha_1}$ is the univariate conformal prediction region associated to the first derivatives $Y_1^1, \ldots, Y_n^1$ and $C_{n,1-\alpha_m}$ is the univariate conformal prediction region related to the derivative of order m: $Y_1^m, \ldots, Y_n^m$

Thus, the problem is now to find the marginal values $\alpha_0, \alpha_1, \ldots, \alpha_m$ that ensure a joint coverage of $1 - \alpha_g$. A solution is available thank to the well know theory of copulae (e.g. Sempi and Durante, Principles of Copula Theory, 2021), a mathematical tool used to model the dependence between multi-variate random variables, that we are going to introduce briefly. Let's consider $(Q^0, \ldots, Q^m)$, an m+1-dimensional random vector with

9

joint cumulative distribution function $F = F_Q : \mathbb{R}^m + 1 \to [0, 1]$. Accordingly, for each $Q^j, \quad j = 0, \ldots, m$ the associated marginal distribution is $F_j = F_j(q^j) = \mathbb{P}(Q^j \leq q^j)$. A copula is a function C that is able to model the dependence structure between the Q^j, regardless of the specific marginal distribution of each variable, that are assumed to be standard uniforms $U^j, \quad j = 0, \ldots, m$. For this reason, we can think to an m+1-dimensional copula as a c.d.f $C : [0, 1]^m + 1 \to [0, 1]$ with standard uniforms marginals. A copula C is characterized by the following properties:

1. $C(u^0, \ldots, u^{j-1}, 0, u^{j+1}, \ldots, u^m) = 0$
2. $C(1, \ldots, 1, u, 1, \ldots, 1) = u$
3. C is m-non-decreasing, i.e. $\int_B dC(u) \geq 0 \quad \forall B = \prod_{j=0}^m [a_j, b_j] \subseteq [0, 1]^m + 1$

It is important to say that, using a suitable copula, is possible to produce a multivariate distribution function from a set of univariate marginals distributions. This is stated in the Sklar's theorem as following:

Theorem 4.1 (Sklar's theorem). *For any m+1-dimensional cumulative distribution function (c.d.f.) F with marginal distributions $F_0, \ldots, F_m$ there exist a copula $C : [0, 1]^m + 1 \to [0, 1]$ such that:*

$$F(\boldsymbol{q}) = F(q^0, \ldots, q^m) = C(F_0(q^0), \ldots, F_m(q^m)) \qquad \boldsymbol{q} \in \mathbb{R}^m + 1 \tag{11}$$

If F_j is continuous for all $j \in \{0, \ldots, m\}$ then C is unique.

By denoting with $[F_j]^{-1}$ the inverse of F_j we get

$$C(\mathbf{u}) = C(u^0, \ldots, u^m) = F([F_0]^{-1}(u^0), \ldots, [F_m]^{-1}(u^m)) \tag{12}$$

At this point the previous problem, of finding the individual level $\alpha_0, \alpha_1, \alpha_m$ so that a global coverage of $1 - \alpha_g$ is guaranteed, can be tackled. Applying the Sklar's theorem to (10), we get this important relation:

$$C(1 - \alpha_0, 1 - \alpha_1, \ldots, 1 - \alpha_m) = 1 - \alpha_g \tag{13}$$

Clearly, if we knew the copula C, then we could find out which are the marginal values $\alpha_0, \alpha_1, \ldots, \alpha_m$ that provide a joint coverage of $1 - \alpha_g$. Unfortunately, such copula C is unknown so we have to estimate it. Following the work of Messoudi, Destercke and Rousseau (2021), we estimate the desired copula from the matrix of the non conformity scores generated from the calibration set, using the *Split conformal* method presented in section 2. Calling R_i^j the non conformity scores associated to the ith function in the calibration set, whose dimension will be indicated with l, taken at the jth order of derivative, the matrix has the following form:

$$R = \begin{bmatrix} R_1^0 & R_1^1 & \ldots & R_1^m \\ \vdots & \vdots & & \vdots \\ R_l^0 & R_l^1 & \ldots & R_l^m \end{bmatrix} \tag{14}$$

Without any prior knowledge about the distribution of the data, we prefer to adopt non-parametric approaches to estimate the copula. In particular, we are referring to the empirical and kernel copula.

Let us now give a brief overview about these two estimation methods. The empirical copula is a way of estimating the marginals directly from the observations; it is a step-wise function with the height of the jumps that is inversely proportional to the sample size. Formally, it is defined as follows:

$$\hat{C}_E(u^0, \ldots, u^m) = \frac{1}{l} \sum_{i=1}^l \mathbb{1}_{\mathbf{u}_i \leq \mathbf{u}} = \frac{1}{l} \sum_{i=1}^l \prod_{j=0}^m \mathbb{1}_{u_i^j \leq u^j} \tag{15}$$

where $u_i^j = \hat{F}_j(R_i^j)$, $\hat{F}_j$ is the empirical cumulative distribution, i.e $\hat{F}_j(r) = \frac{1}{l} \sum_{i=1}^l \mathbb{1}_{R_i^j \leq r}, j = 0, \ldots, m$ and $\mathbb{1}_A$ is the indicator function of event A.

On the other side, the kernel copula can be seen as a smoothed version of $\hat{C}_E$, with the particular choice of the Gaussian kernel density as smoothing term. It is defined as follows:

$$\hat{C}_K(u^0, \ldots, u^m) = \frac{1}{l} \sum_{i=1}^l \prod_{j=0}^m \frac{1}{h^j} \Phi\left(\frac{v^j - R_i^j}{h^j}\right) \tag{16}$$

where $v^j = \left[\hat{F}_j^s\right]^{-1}(u^j)$ and $\hat{F}_j^s(r) = \frac{1}{lh^j} \sum_{i=1}^l \Phi\left(\frac{r - R_i^j}{h^j}\right), j = 0, \ldots, m$ is a smoothed version of the empirical cumulative distribution, called kernel density estimate (KDE). Φ is the cumulative distribution function of the standard normal, while $h^j := \sqrt{[\mathbf{H}]_{jj}}$, where $\mathbf{H}$ is known in the literature as bandwith matrix. In particular, $\mathbf{H}$ is computed using the Silverman's rule of thumb.

After having introduced these two copula estimation methods, let us now present three well known copulae, that we will use to better understand the behaviour of our estimates:

- the product (or independence) copula, which model independence: $\Pi(u^0 \ldots, u^m) = \prod_{j=0}^m u^j$;
- The Fréchet-Höffding upper bound copula, which model positive dependence:
  $M(u^0, \ldots, u^m) = \min_{0 \le j \le m}\{u^j\}$;
- The Fréchet-Höffding lower bound copula, which model negative dependence:
  $W(u^0, \ldots, u^m) = \max\left\{\sum_{j=0}^m u^j - m + 1, 0\right\}$.

Back to the main focus of this part of the study, the problem is finding the joint coverage $1 - \alpha_g$ when the marginal levels are $\alpha_0, \alpha_1, \ldots, \alpha_m$. Without loss of generality, we put every marginal level $\alpha_0, \alpha_1, \ldots, \alpha_m$ equal to the value of α_t, meaning that the conformal prediction bands will have the same coverage $1 - \alpha_t$ for every order of derivative. By using the empirical and/or the kernel copula, starting from the matrix of the non conformity scores (14), we manage to map every marginal coverage $1 - \alpha_t$ into its corresponding joint value, i.e $1 - \alpha_g$. The result of this procedure is shown in Fig. 7, where the Berkeley Study Growth data and the Canadian weather data are used to compute the copula estimation. We set m equal to 1, meaning that only the first order of derivative is considered, in addition to the functions themselves. The empirical and kernel copulae are compared not only with each other, but also with the product, upper and lower bound copula. This is the simplest way to check the presence of independence, positive or negative dependence in our estimates. For what concerns the Berkeley Study Growth data, first row of Fig. 7, we can see that the empirical and kernel estimates are similar to each other. Furthermore, they are very close to the independence copula, the blue line, indicating that there is no dependence between the "strangeness" of a function and the "strangeness" of its first derivative, i.e. if a function is unusual with respect to the others, it doesn't imply that its first derivative will be unusual w.r.t the other derivatives too. In the second row the results about the Canadian weather are displayed. The lack of observation, in this case only 16 NCS are available, entails that the empirical copula is not a reliable estimator. Additionally, for the high values of marginal coverage, the empirical copula exceed the red line, which represent the upper bound copula, leading us to prefer the kernel copula.
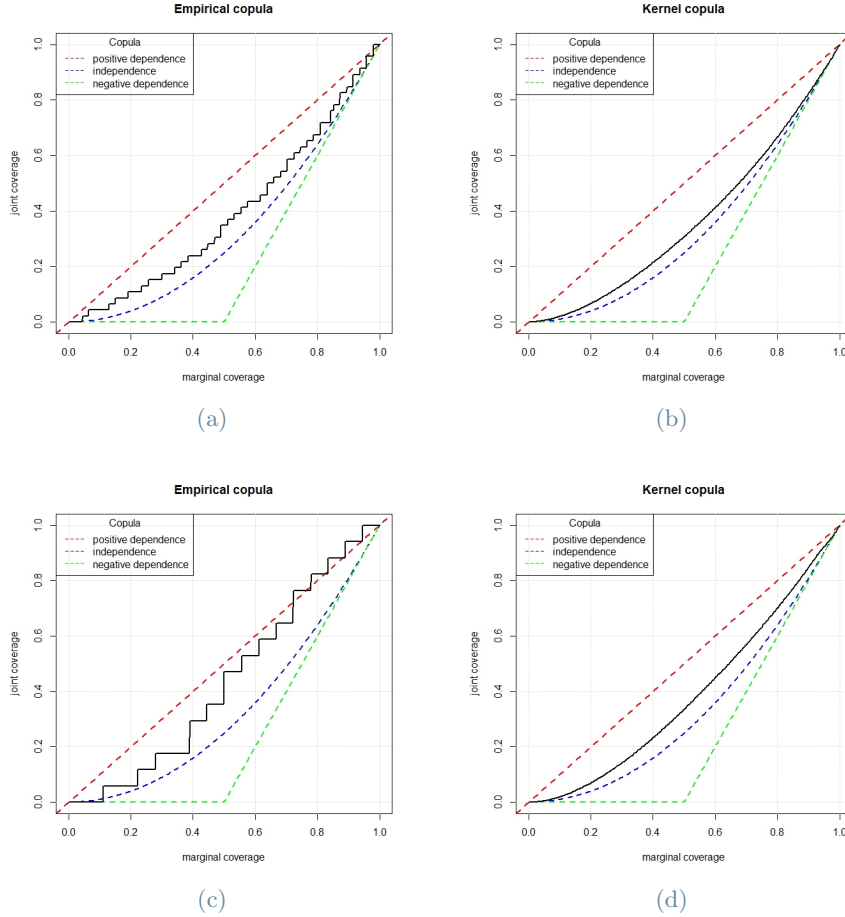
Figure 7: Empirical and kernel copula estimates using functions and their first derivatives taken from Berkeley Study Growth data (first row) and Canadian weather (second row). The Empirical copula, panel (a),(c), and the kernel copula, panel (b),(d), are represented by the black line, while the red,blue and green lines are, respectively, the positive dependence, independence and negative dependence copula.

To conclude this section, let us define in a formal way our final output: given a new observation $\mathbf{Y}_{n+1}$, its *p-value function* with respect to a multivariate functional dataset $\mathbf{Y}_1 \ldots \mathbf{Y}_n$ is defined as follows:

$$\forall t \in \mathcal{T}, \quad \forall j = 0, \ldots, m, \quad p^j(t) = \min_{\alpha \in \mathcal{A}_g} \alpha \quad s.t.$$

$$(u^j_{1-\alpha} - Y^j_{n+1})^+(t) = 0 \quad \vee \quad (Y^j_{n+1} - l^j_{1-\alpha})^+(t) = 0 \tag{17}$$

where $\mathcal{A}_g$ is the set with all the corrected values of α, obtained by mapping every α_t in [0,1] into its corresponding joint value. $u^j_{1-\alpha}$ and $l^j_{1-\alpha}$ are, respectively, the upper and the lower bound of the conformal region with coverage $1 - \alpha$ associated to the jth order of derivative. $p^j(t) \in [0,1]$ is the value in t of the jth order *p-value function*.

5. Applications and simulation study, second part

5.1. Applications to literature datasets

The ultimate version of the *p-value function* is now tested, with respect to our first proposal, on some of the literature datasets already seen in Section 3.1. To avoid any confusion, we will call it *copula adj* p-value function. It is a simple way to underline the presence of a codomain adjustment, induced by the copula theory, in addition to the domain adjustment, obtained by the conformal prediction bands, that is always considered. On the other hand, the version presented in Section 2 will be named *not copula adj* p-value function, putting evidence on the fact that the codomain adjustment is missing, in this case only the domain adjustment is applied. For the sake of simplicity, we consider the case of $m = 1$. As a consequence, a generic bivariate function $\mathbf{Y}_j$ is such that

$\mathbf{Y}_j = (Y_j^0, Y_j^1)$ where Y_j^0 and Y_j^1 are, respectively, the function and its *first* derivative. The first example is built from the Berkeley Study Growth data, using the empirical as copula estimation method, a choice justified due to the high dimension of the data available. The first derivatives are computed manually, using the incremental ratio formula. The output is shown in Fig. 8. We are considering a function that lies in the outer regions for the most of the domain, accordingly, the corresponding p-value function assumes very low values, in both the copula adj and no copula cases. On the other hand, the first derivatives always stays in the cold colored regions, except in the central part. This behaviour is underlined by its p-value function that, in both the two versions, is close to one for the entire domain, with the exception of the middle segment, where it decrease rapidly. In general, it is easy to observe that in the copula adjusted version the p-value function has slightly higher values, with respect to the no copula one, but this is an expected property. Indeed, the joint coverage levels, obtained by using the empirical copula, are smaller than the marginal coverage ones; as a consequence, the values of the copula adjusted version are greater, since they are computed on the base of bigger values of α.

(a) boy02 function over the no copula adj conf. regions

(b) boy02 function over the copula adj conf. regions

(c) boy02 first derivative over the no copula adj conf. regions

(d) boy02 first derivative over the copula adj conf. regions

(e) p-value function no copula adj, function

(f) p-value function copula adj, function

(g) p-value function no copula adj, first derivative

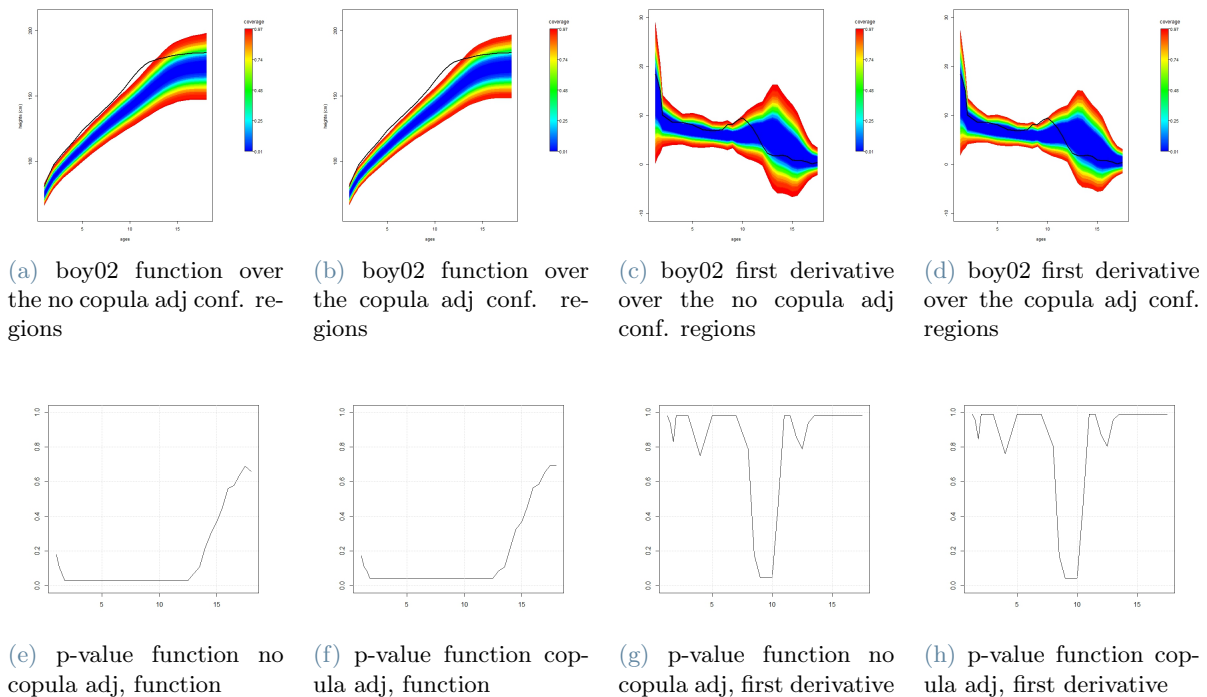(h) p-value function copula adj, first derivative

Figure 8: Berkeley Growth Study data, we consider the function of the boy02 and its corresponding first derivative. Specifically, the function is plotted over the conformal regions in the no copula case (a) and copula adj case (b). the corresponding p-value functions are shown in (e, f). The same scheme is followed to present the result of the first derivative in panels (c, d, g, h)

A second example is presented, it refers to the temperature measurements, taken from the Canadian whether dataset. Differently from the previous case, the first derivatives are now computed after having applied a smoothing (Fourier basis) to the original functions. Furthermore, despite of the low number of observation, we decide to opt for the empirical as copula estimation method, for the purpose of underlining the adjustment, that is more evident with respect to the kernel technique. The results are shown in Fig. 9. The function under analysis is the same that we have discussed in Section 3.1, the only difference is the presence of the smoothing. Both the two versions of the p-value function have high values in the middle and low values at the extremes of the domain, a trend that is confirmed by the fact that the function crosses the outer regions at the beginning and at the end, while it touches inner (cold colored) regions in the center part. Let's focus now abut the first derivative: it goes thorough both the outer and inner regions repeatedly along the domain; accordingly, its p-value function has the characteristic of fluctuating between zero and one continuously. We put evidence on the fact that, also in this example, the *copula adjusted* p-value function has higher values than the no copula case.

(a) Inuvik function over the no copula adj conf. regions

(b) Inuvik function over the copula adj conf. regions

(c) Inuvik first derivative over the no copula adj conf. regions

(d) Inuvik first derivative over the copula adj conf. regions

(e) p-value function no copula adj, function

(f) p-value function copula adj, function

(g) p-value function no copula adj, first derivative
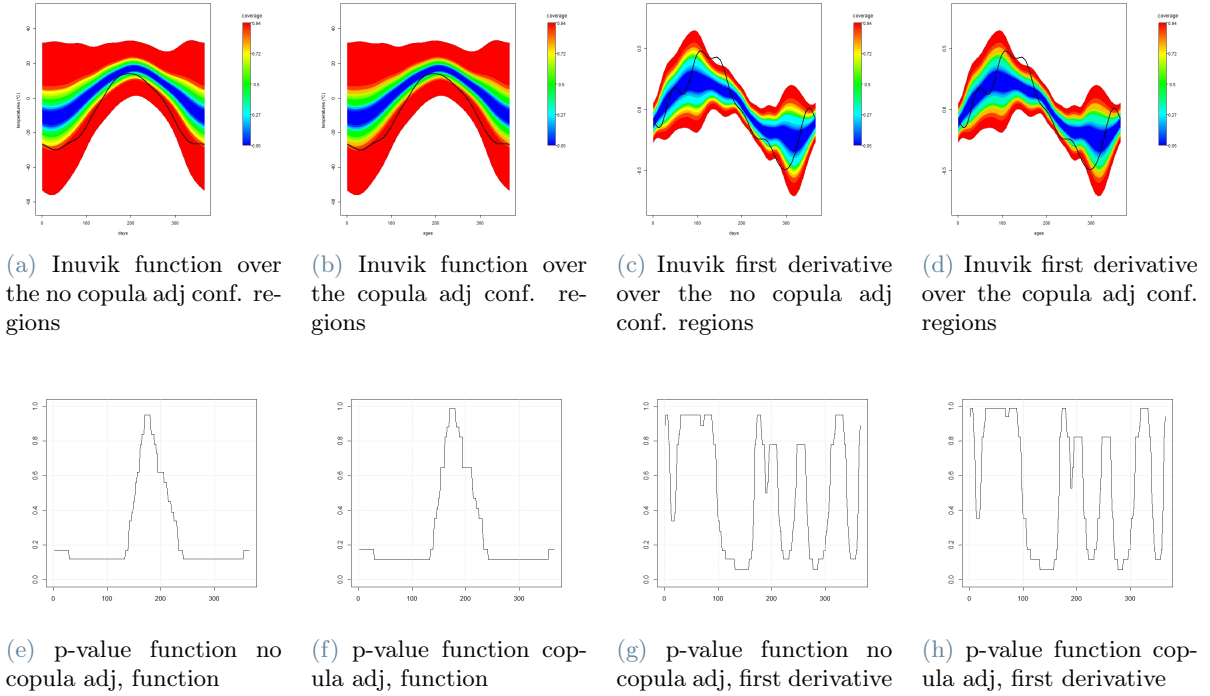
(h) p-value function copula adj, first derivative

Figure 9: Canadian weather data - temperatures, we consider the function related to the location of Inuvik and its corresponding first derivative. Specifically, the function is plotted over the conformal regions in the no copula case (a) and copula adj case (b). the corresponding p-value functions are shown in (e, f). The same scheme is followed to present the result of the first derivative in panels (c, d, g, h)

The last function we present is an anomalous one, built from the Berkeley Study Growth data. Specifically, we add noise, distributed as a Gaussian random variable with $\mu = 0$ and $\sigma = 3$, to the central part of the girl05's function. The aim is to find out if the p-value function is able to detect this particular pattern, focusing on both function and first derivative aspects. Following the first example of the section, in which we used the same data, we are going to employ the empirical copula in order to find the joint coverage levels. By looking at Fig. 10, we can see that the anomalous function always lies in the cold colored region. Accordingly, its p-value function is close to one in both the copula adjusted and no copula cases. Therefore, the presence of a strange pattern, inducted by the noise, is not recorded if we limit to consider only the function. On the other side, the first derivative's p-value function, in both cases, oscillates between high and low values in its central part, indicating that something unusual it's happening. The plot of the first derivative over the conformal regions, panel c and d of Fig.10, shows a completely different trend, with respect to the other functions, confirming what the p-value function points out. As for the previous cases, also here the p-value function takes higher values when the copula adjustment is considered. This last example justifies, in our opinion, the importance of computing the p-value function also for higher order of derivatives. Sometimes, an anomalous behaviour is not detected by looking only at the function. However, the p-value function at an higher order of derivative, first derivative in our example, can easily underline the existence of an unusual pattern, which is entirely missed at a function level.
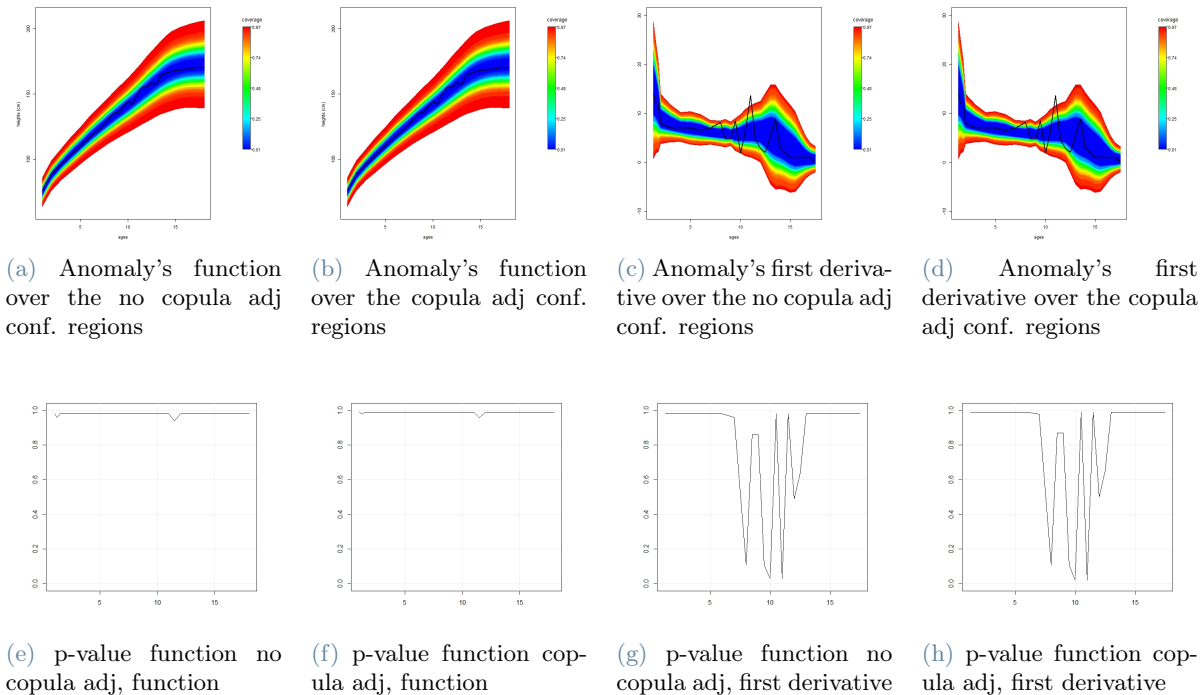
14

(a) Anomaly's function over the no copula adj conf. regions

(b) Anomaly's function over the copula adj conf. regions

(c) Anomaly's first derivative over the no copula adj conf. regions

(d) Anomaly's first derivative over the copula adj conf. regions

(e) p-value function no copula adj, function

(f) p-value function copula adj, function

(g) p-value function no copula adj, first derivative

(h) p-value function copula adj, first derivative

Figure 10: Berkeley Study Data, The anomaly is built by appling noise to the central part of the girl05's function. Specifically, the function is plotted over the conformal regions in the no copula case (a) and copula adj case (b). the corresponding p-value functions are shown in (e, f). The same scheme is followed to present the result of the first derivative in panels (c, d, g, h)

5.2. Simulation study

The characteristics of the *copula adjusted* p-value function are investigated with the help of a simulation study. The general settings are the same of the simulation study presented in Section 3.2, the important difference is that now we are comparing the validity of the copula adjusted version with respect to the not copula adjusted one. As in the previous section, we will set $m = 1$, limiting our analysis up to the first order of derivative. Consequently, we will compute the empirical coverage in three distinct cases: firstly, we will focus only on the functions, computing the probability of a function to be inside a certain region of level $1 - \alpha$, i.e. the functions marginal coverage. In the second place, we will do the same considerations but referring just to the first derivatives, obtaining the first derivatives marginal coverage. Lastly, we will calculate the probability that both the function and its first derivative are always inside a specific conformal prediction band, i.e. the joint coverage. Obviously, all these estimates are done using the p-value function copula adjusted and not copula adjusted of the functions and first derivatives.

The two simulated scenarios are the same we have used in Section 3.2: amplitude and phase variation applied to sinusoidal functions and splines with different variability over the domain. In each scenario, we test the validity induced by the p-value functions of 200 bivariate functions that composed our test set. The training set, instead, is characterized by an increasing dimension on the following logarithm scale: 4,8,16,32,64,128. In order to have reliable estimates, the measurements are repeated for a total of $B = 100$ times for each training set's dimension and the mean over all the estimates is taken.

Let's now focus the attention on the first scenario, where we consider sinusoidal function with amplitude and phase variation. The chosen model to generate the functions, for both test and training set, is the same we have used in Section 3.2, in particular, we are referring to equation (8). The first derivatives are computed accordingly by the incremental ratio formula. The resulting output of marginal and joint coverage are shown in Fig. 11 and Fig. 12; the difference between these two figures lies in the copula estimation: in the first one we use the empirical copula, while in the second case the result is carried out by the kernel copula. The aim is to understand if the way in which the copula is obtained influence strongly the validity results. The main difference between the empirical and the kernel method can be found in the first columns of the two above mentioned figures; indeed, when N is equal to 4,8,16 the black squares, that indicate the behaviour in the copula adjusted case, does not lie near the first reachable coverage level in the empirical case, while they are always there if we look at the kernel copula plots. The reason is that the empirical copula generates rough estimates when a

15

small collection of observations is available and, consequently, it happens that the first coverage level is skipped. Generally, we notice that the kernel coverage levels are smaller than the empirical ones. For what concerns the marginal and joint coverage, we do not observe a big difference; the cause lies in the fact that there is a strict relation in the behaviour between the function and the first derivatives, i.e if a function is strange with respect to the others, its first derivative will be strange too, compared to the rest of the first derivatives. To conclude, we would like to point out that the expected behaviour is observed, since the joint coverage empirical estimates lies in proximity of the theoretical ones, indicated by the red dotted line.
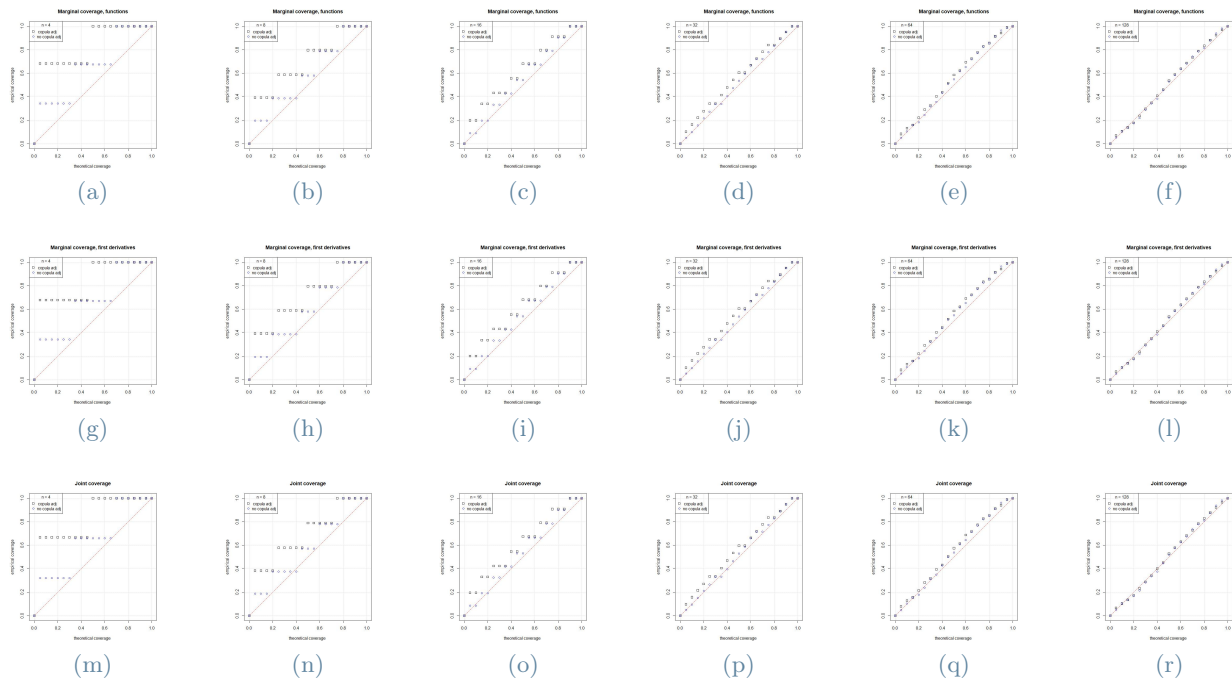

Figure 11: Simulation study considering amplitude and phase variation, the test is fixed at M=200 functions, while the dimension of the calibration set increases on the logarithm scale: 4,8,16,32,64,128. The corresponding first derivatives are computed by the incremental ratio formula and the **empirical** copula estimation method is used. The general layout can be understood easily by referring to a matrix: the first row shows the marginal coverage of the functions, the second row the marginal coverage of the first derivatives, while the the third one present the joint coverage. In each column we have the output for a specific and fixed value on the logarithm scale. In every subfigures, the black squares refer to the copula adjusted case, while the blue circles indicate the estimate in the not copula adjusted version.
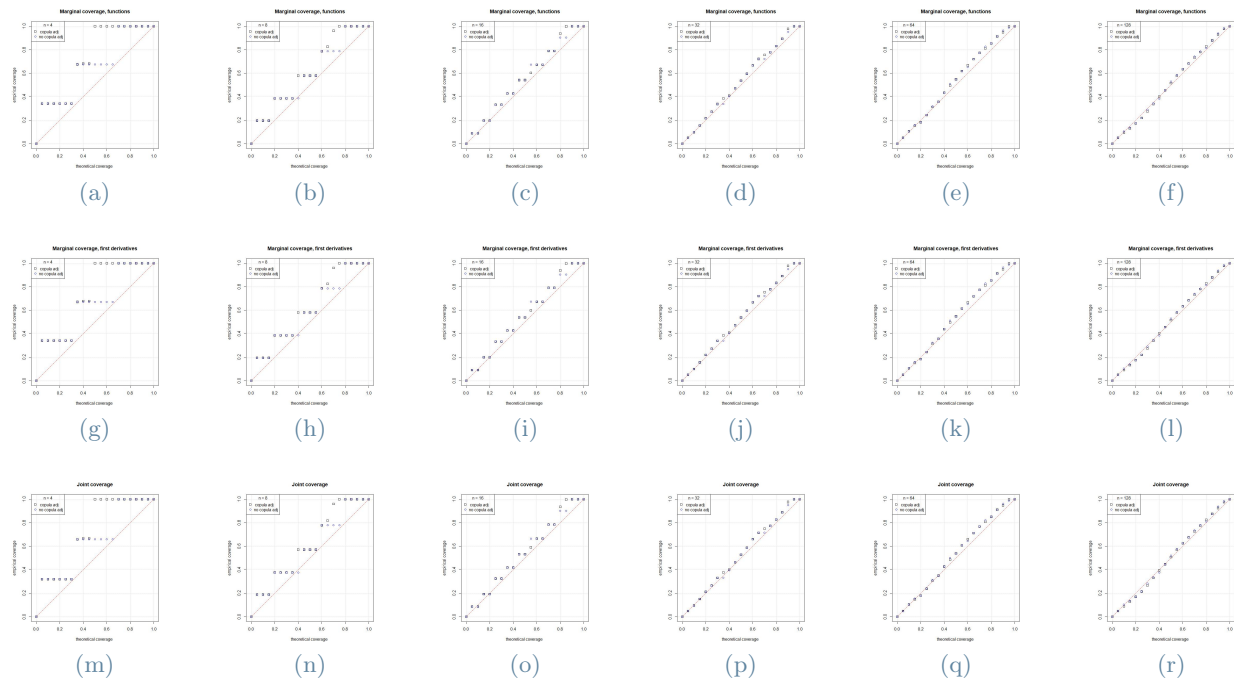
Figure 12: Simulation study considering amplitude and phase variation, the test is fixed at M=200 functions, while the dimension of the calibration set increases on the logarithm scale: $N = 4, 8, 16, 32, 64, 128$. The corresponding first derivatives are computed by the incremental ratio formula and the **kernel** copula estimation method is used. The general layout can be understood easily by referring to a matrix: the first row shows the marginal coverage of the functions, the second row the marginal coverage of the first derivatives, while the the third one present the joint coverage. In each column we have the output for a specific and fixed value on the logarithm scale. In every subfigures, the black squares refer to the copula adjusted case, while the blue circles indicate the estimate in the not copula adjusted version.

The second and last simulation scenario regard splines over the interval [0,1] with four internal knots at 0.2,0.4,0.6,0.8. The variability over the domain is changed by sampling the coefficients, that are going to multiply the spline basis, from a uniform distribution with varying range, as we did in the first part of the paper (Section 3.2). The functions are obtained by the model in equation (9), while the first derivatives are computed by the linear combination between the *same* coefficients used for the functions and the first derivative of the splines basis. As before, we display the final estimates of marginal/joint coverage in two separated figures, namely Fig. 13 and Fig. 14, to put evidence on the differences that could come out if different copula estimators are used, in our case empirical vs kernel copula. Overall, one can notice that the empirical copula adjustment leads to greater estimates, represented by the black squares, with respect to the kernel copula scenario. Concerning the empirical copula plot, as for the previous simulation case, the small value of N, displayed in the first three columns of the figures, are characterized by the absence of the black squares near the first level of coverage. this particular trend can be connected to the copula estimator itself, which is not reliable if few non conformity scores are used. The marginal coverage, of both functions and first derivatives, follows the expected target, i.e the red dotted line, as the number of observation in the training set increases, whether in the copula adjusted or in the no copula adjusted case. On the other hand, an unusual behaviour comes out from the joint coverage, displayed in the last row of Fig 13 and 14. Indeed, for high values of N, the empirical coverage goes below the bisector. This is a predictable pattern when the copula adjustment is missing (blue circles) while it is difficult to explain when the copula adjustment is considered (black square). As a matter of fact, the main reason for which we applied the copula correction is to reach the theoretical joint coverage. We will reserve the right to better understand what is happening in this case, in order to find the problem, which with high probability is at coding level.
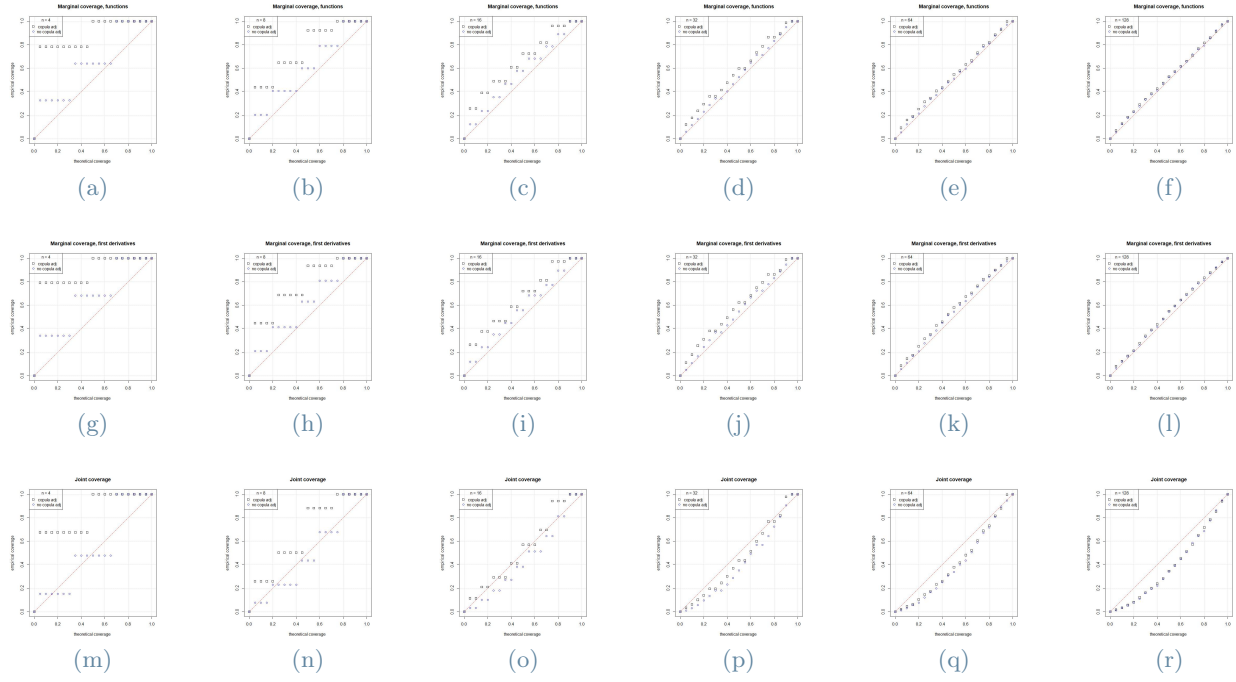
(a)   (b)   (c)   (d)   (e)   (f)

(g)   (h)   (i)   (j)   (k)   (l)

(m)   (n)   (o)   (p)   (q)   (r)

Figure 13: Simulation study considering splines over the interval [0,1] with four internal knots at 0.2,0.4,0.6,0.8. The different variability over the domain is obtained by sampling the coefficients of the linear combination from a uniform distribution with varying range. The test is fixed at M=200 functions, while the dimension of the calibration set increases on the logarithm scale: 4,8,16,32,64,128. The **empirical** copula estimation method is used. The general layout can be understood easily by referring to a matrix: the first row shows the marginal coverage of the functions, the second row the marginal coverage of the first derivatives, while the the third one present the joint coverage. In each column we have the output for a specific and fixed value on the logarithm scale. In every subfigures, the black squares refer to the copula adjusted case, while the blue circles indicate the estimate in the not copula adjusted version.
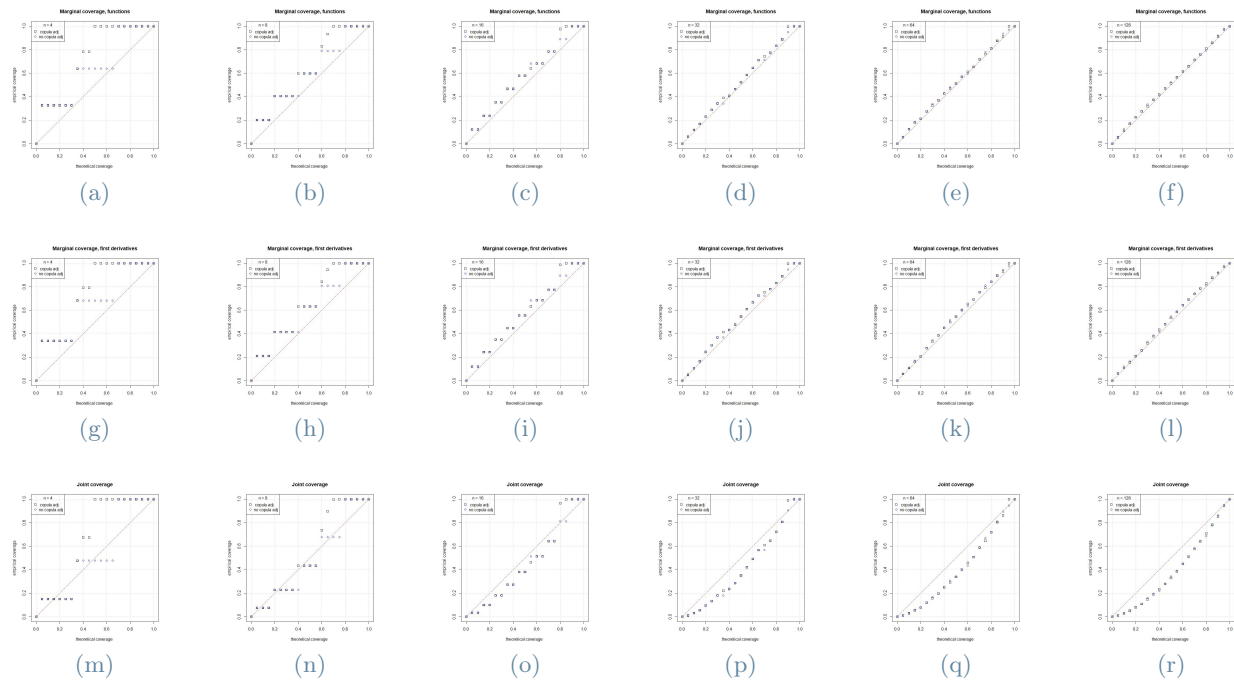
Figure 14: Simulation study considering splines over the interval [0,1] with four internal knots at 0.2,0.4,0.6,0.8. The different variability over the domain is obtained by sampling the coefficients of the linear combination from a uniform distribution with varying range. The test is fixed at M=200 functions, while the dimension of the calibration set increases on the logarithm scale: 4,8,16,32,64,128. The **kernel** copula estimation method is used. The general layout can be understood easily by referring to a matrix: the first row shows the marginal coverage of the functions, the second row the marginal coverage of the first derivatives, while the the third one present the joint coverage. In each column we have the output for a specific and fixed value on the logarithm scale. In every subfigures, the black squares refer to the copula adjusted case, while the blue circles indicate the estimate in the not copula adjusted version.

6. Real case study

In this section the result of our methodology, i.e. the *copula adjusted* p-value function, is applied to a real case study: the Vertical Density Profile (VDP), that has been treated by many authors over the years(Colosimo, Meneses and Semeraro 2013, Walker and Wright 2002). VDPs play a fundamental role in measuring the quality of particleboards, widely used in housing and furniture, that are build through a complex manufacturing process. Fortunately, there's a strict relation between process condition and the VDP curve, in particular, the mechanical characteristic of the final product are affected by VDP curve's changes. As a consequence, possible failure of the final product can be easily detect by spotting anomalous behaviour in the VDP curve and that's the reason why VDP profile monitoring is so important.

The profile monitoring traditional approaches are divided in two parts, namely phase I and phase II. In the first one, also called design phase, a set of in-control profiles is used to build a control chart that, in the second (monitoring) phase, is adopted to detect any kind of change of new profiles with respect to the in-control set. Our proposed methodology, based on conformal prediction bands and copula estimation, goes under this framework perfectly. Indeed, With the help of our *copula adjusted* p-value function, it is possible to recognize strange pattern in a new profile, and in its higher derivatives, with respect to a in-control set in a fast and easy way.

For what concerns the dataset we are going to deal with, it is composed of $N = 263$ functions, the VDP profiles. Each profile measures the density (kg/m^3) over the vertical axes of a given particleboards, considering a grid of $p = 189$ equally spaced locations, each measurement is 0.09mm apart. In line with the rest of the paper, we investigate the behaviour up to the first order of derivative. Concerning the copula adjustment, we limit to use the empirical copula estimator, that is thought to produce reliable values because of the huge number of observations in the VDP data. Fig. 15 shows the data we are going to deal with and the empirical copula estimate in two separated panels. In particular, by looking at the empirical copula, one can notice that the

black line lies in the middle, between the positive dependence, red line, and the independence, blue line.
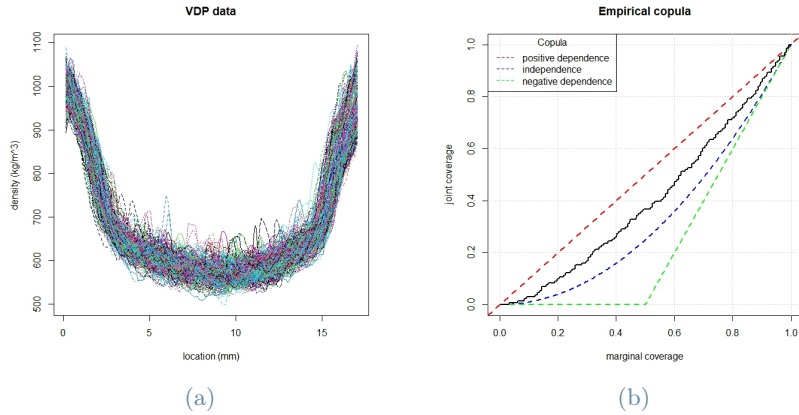

(a)
(b)

Figure 15: An overview of VDP data (a) and the plot of the empirical copula estimate compared with the three copula that model independence, positive/negative dependence (b)

Of course, a smoothing is applied to the functions in order to have noise free first derivatives. In the following, we will discuss two different cases: a in-control profile, for which we expect to observe an high valued p-value function copula adj, and a profile out of control, characterized, hopefully, by a low valued p-value function.

The result of the in-control profile are shown in the Fig. 16. Both the two *copula adjusted* p-value functions are extremely close to one all over the domain, indicating that the profile considered doesn't have anomalous pattern at any derivative order. Similarly, by looking at the profile itself, one can observe that it is always inside the inner cold colored region, confirming that both the function and the first derivative are behaving regularly. On the other hand, we refer to Fig. 17 for the discussion about the out-of control profile. The unusual pattern is obtained by manually modifying the amplitude of the VDP254 profile, i.e. the 254-th profile of the dataset, in the central part of the domain. Particularly, nearby the point of abscissa equal to 5, we force the profile to grow until a maximum and, then, to decrease until a minimum, as shown in panel (a). The profile's *copula adjusted* p-value function is close to the highest value, except for 2 points in the middle region, in which it goes below the 0.2; these points are the previous mentioned maximum and minimum, where the function crosses the outer regions. Concerning the first derivative, we can see that the *copula adjusted* p-value function report that something unexpected is happening exactly in the central part; indeed, it decrease badly at three distinct points, that are, accordingly, the points in which the profile's first derivative goes thorough the hot colored regions. Putting together all the considerations, we can understand that in the middle part of the domain not only we have two anomalous values of the profile, but also an anomalous trend, in term of increasing/decreasing, is present, as the first derivative's p-value function suggests.
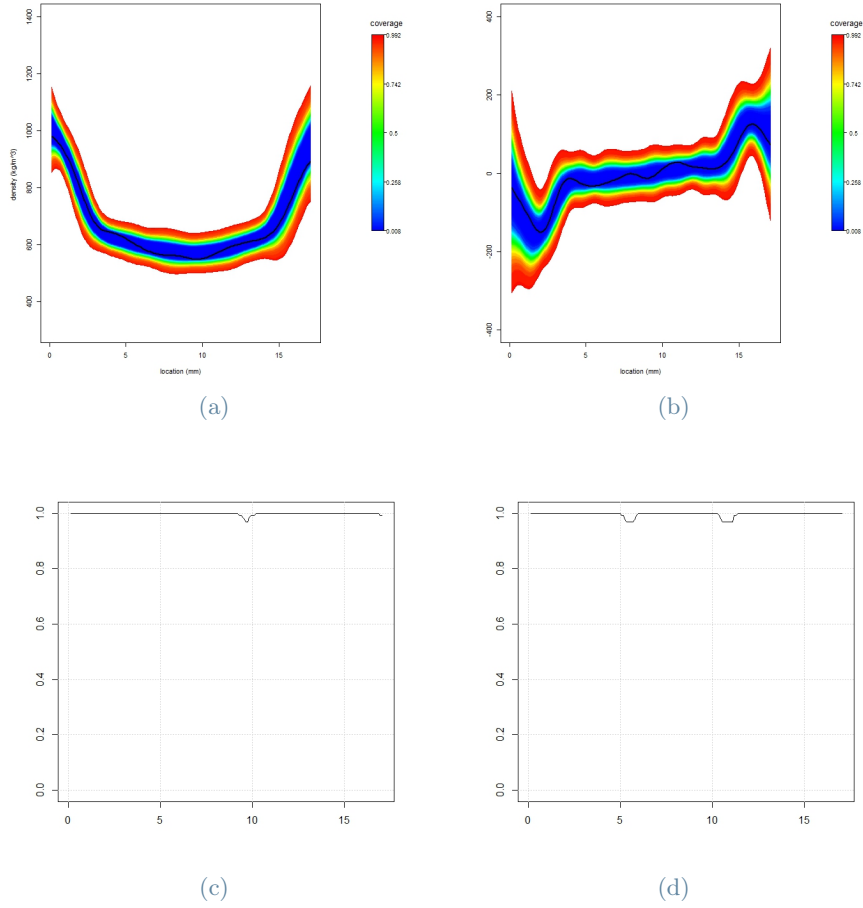
Figure 16: VDP dataset, in particular the function considered is an in-control one, the VDP53 profile. In the first row the function and its first derivative are plotted over the copula adjusted conformal prediction bands, respectively in panel (a) and (b). The second row displays the 2 p-value function copula adj, for the function, (c), and for its first derivative, (d).
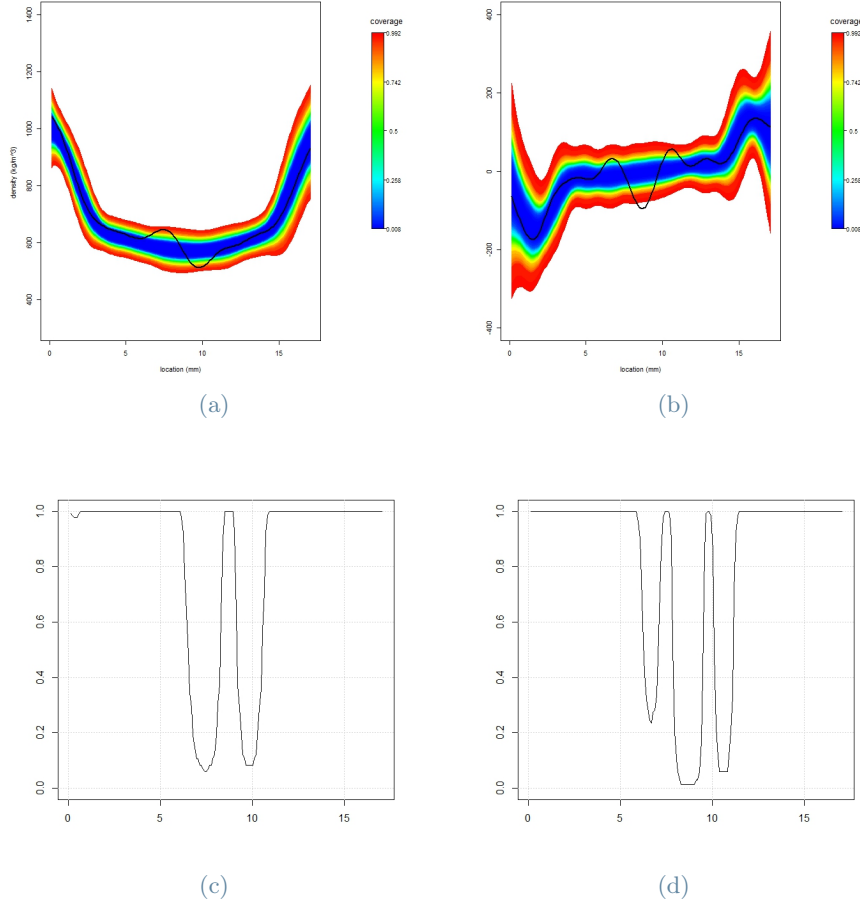
(a)

(b)

(c)

(d)

Figure 17: VDP dataset, the function considered is an out of control one, obtained manually by modifying the amplitude of the VDP254 in the central part of the domain. In the first row the function and its first derivative are plotted over the copula adjusted conformal prediction bands, respectively in panel (a) and (b). The second row displays the 2 p-value function copula adj, for the function, (c), and for its first derivative, (d).

7.   Conclusion

Our proposed methodology is thought to be an additional tool that can integrate the traditional approach in the field of profile monitoring. Indeed, by scanning the various conformal prediction band, the p-value function can recognize if unexpected values are present in new functions (unexpected with respect to the values taken by the functions in a reference dataset). However, if an anomaly occurs at higher order of derivative, for example in the way the function increase/decrease, it could be missed, as in the case of a slightly noisy function, a weakness that is widely analyzed in Fig. 4. For this reason, the p-value function is extended to include higher order of derivatives. Consequently, an adjustment is needed to ensure joint coverage and our intuition is using the copula theory to reach our goal. Since in general no prior information about the data is available, we adopt two non parametric methods, the empirical and kernel copula, in order to estimate the copula and, successively, to obtain the desired correction of the coverage levels. Both the two methodology steps are supported by a simulation study and an application to well-known literature dataset. In the first part of the paper, the attention is focused on the differences between the adjusted p-value function and the version already available in the literature, i.e the not adjusted p-value function. Section 3 underlines the advantages of our proposed methodology in terms of legibility and validity. In the second part, after having explained why a codomain adjustment is needed, we go into the pro and cons of considering a copula adjustment or not. The results, presented in Section 5, demonstrate that the copula adjusted p-value function takes higher values and, in general, guarantees bigger estimates of the joint/marginal coverage levels. Nevertheless, the splines simulation study in Figures 13 and 14 shows the existence of a problematic underestimation of the joint coverage, also in the copula adjusted version. We will reserve the right to investigate deeply this problem and, hopefully, we will manage to present a corrected version that will reach the theoretical coverage.

All considered, Our proposed methodology can be an additional and useful tool in the field of profile monitoring, and more in general, could find application in every situation in which the detection of anomalous pattern is required.

References

[1] J. Diquigiovanni, M. Fontana and S. Vantini, (2021), The Importance of Being a Band: Finite-Sample Exact Distribution-Free Prediction Sets for Functional Data, arXiv:2102.06746v2

[2] S. Messoudi, S. Destercke and S. Rousseau, (2021), Copula-based conformal prediction for multi-target regression, Pattern Recognition, Université de Technologie de Compiègne

[3] B.M. Colosimo, M. Meneses and Q. Semeraro, (2013), Vertical density profile monitoring using mixed-effects model, ScienceDirect, p. 2.

[4] Walker, E., Wright, S., (2002), Comparing curves using additive models, Journal of Quality Technology

[5] J.O. Ramsay and B. Silverman, (2005), Functional Data Analysis, Springer

[6] L. Horvath and P. Kokoszka, (2012), Inference for Functional Data With Applications, Springer

[7] Lòpez-Pintado, S. and Romo, J. (2009), On the concept of depth for functional data, Journal of the American Statistical Association

[8] Lei, J., Rinaldo, A. and Wasserman, L. (2015), A conformal prediction approach to explore functional data, Annals of Mathematics and Artificial Intelligence

[9] Vovk, V., Gammerman, A. and Shafer, G. (2005), Algorithmic learning in a random world, Springer Science & Business Media

[10] Papadopoulos, H., Proedrou, K., Vovk, V. and Gammerman, A. (2002), Inductive confidence machines for regression, in 'European Conference on Machine Learning', Springer

[11] C. Sempi and F. Durante, (2021), Principles of Copula Theory, Routledge

Abstract in lingua italiana

In molti processi industriali la complessità è aumentata nel corso degli anni e, parallelamente, la necessità di monitorare i processi, al fine di trovare comportamenti anomali, è diventata sempre più essenziale. Per questo motivo, il monitoraggio dei profili, un recente campo di ricerca nel controllo dei processi statistici, sta attirando l'interesse di molti ricercatori. Con questa tesi, il nostro obiettivo è proporre uno strumento aggiuntivo, la cosiddetta p-value function, che andrà ad integrare l'approccio tradizionale utilizzato nel monitoraggio dei profili. In particolare, la nostra metodologia, basata sulla Conformal prediction e sulla teoria delle copule, è in grado di riconoscere non solo se è presente un'anomalia, ma anche quale parte del dominio è responsabile di quel particolare comportamento. Per verificare I punti di forza e/o debolezza del nostro metodo, la discussione è accompagnata da applicazioni a dataset presenti in letteratura e da uno studio su dati simulati. Un caso reale, vale a dire i Profili di Densità Verticale (VDP), è presentato per illustrare il potenziale della nostra procedura in applicazioni reali.

Parole chiave: monitoraggio di profili, dati funzionali, conformal prediction, copulae, funzione p-value.