Statistica Sinica **35** (2025), 853-871
doi:https://doi.org/10.5705/ss.202022.0087

THE IMPORTANCE OF BEING A BAND: FINITE-SAMPLE EXACT DISTRIBUTION-FREE PREDICTION SETS FOR FUNCTIONAL DATA

Jacopo Diquigiovanni*[1], Matteo Fontana[2,3] and Simone Vantini[2]

[1] *University of Padova,* [2] *Politecnico di Milano and* [3] *Royal Holloway, University of London*

Abstract: Functional Data Analysis represents a field of growing interest in statistics. Despite several studies have been proposed leading to fundamental results, the problem of obtaining valid and efficient prediction sets has not been thoroughly covered. Indeed, the great majority of methods currently in the literature rely on strong distributional assumptions (e.g., Gaussianity), dimension reduction techniques and/or asymptotic arguments. We propose a new nonparametric approach in the field of Conformal Prediction, based on a new family of nonconformity measures inducing conformal predictors able to create closed-form finite-sample valid or exact prediction sets for functional data under very minimal distributional assumptions. In addition, our proposal ensures that the prediction sets obtained are bands, an essential feature in the functional setting that allows the visualization and interpretation of such sets. The procedure is also fast, scalable, does not rely on functional dimension reduction techniques and allows the user to select different nonconformity measures depending on the problem at hand always obtaining valid bands. Within this family of measures, we propose also a specific measure leading to prediction bands asymptotically no less efficient than those with constant width.

Key words and phrases: Conformal Prediction, distribution-free prediction, exact prediction set, functional data, prediction band, uncertainty quantification.

1. Introduction

One of the main roles of statistics in our new, data-rich world is to provide scientists, businesspeople, and policymakers with tools able to deal with an increasing amount of data of increasing complexity. Automated sensor arrays and measuring systems now provide huge quantities of high-frequency and high-dimensional data about all sorts of social or physical phenomena.

Among the most popular toolboxes that have the capacity to deal with this kind of complex data, one can find Functional Data Analysis (FDA, Ramsay and Silverman, 2005). FDA is an ebullient field of statistics which aim is to develop theory and methods to deal with data sets made of functions defined over a domain, either uni- or multidimensional, and usually characterized by some

*Corresponding author. E-mail: jacopodiquigiovanni93@gmail.com

degree of smoothness. In the following, we will indicate with $\mathcal{Y}(\mathcal{T})$ the family of functions $y : \mathcal{T} \to \mathbb{R}$ belonging to $L^\infty(\mathcal{T})$ with $\mathcal{T}$ closed and bounded subset of $\mathbb{R}^d$, $d \in \mathbb{N}_{>0}$, and with $y_1, \ldots, y_n$ possible realizations of n i.i.d. random functions $Y_1, \ldots, Y_n \sim P$ taking values in $\mathcal{Y}(\mathcal{T})$. Without loss of generality, hereafter we will consider $d = 1$ since it is the most common practical case. Despite being born in relatively recent times (Ramsay, 1982), a plethora of standard multivariate tools have ported to the functional realm: among others Functional Principal Component Analysis (Ramsay and Silverman, 2005, Chap. 10), Functional Linear Regression (Ramsay and Silverman, 2005, Chap. 10) and Functional Boxplots (Sun and Genton, 2011).

A problem that, perhaps surprisingly, has not been covered in a satisfactory way in the FDA literature is the issue of uncertainty quantification in prediction and forecasting. In a more formal way, the interest is in the creation of prediction sets, namely subsets of $\mathcal{Y}(\mathcal{T})$ that include a new function Y_{n+1} (i.i.d. to $Y_1, \ldots, Y_n$) with a certain nominal confidence level $1 - \alpha$. In particular, the aim is to obtain either exact—i.e., ensuring a coverage equal to the nominal confidence level—or at least valid—i.e., ensuring a coverage no less than the nominal confidence level—prediction sets. Recent works in FDA provide novel insights into this very meaningful applied and theoretical issue. These attempts can be broadly classified in two classes: The first one is composed of works based mainly on bootstrapping techniques, either parametric (Degras, 2011; Cao, Yang and Todem, 2012) or, via the use of functional quantiles, via nonparametric bootstrap techniques (Cuevas, Febrero and Fraiman, 2006; Berg et al., 2017; Schüssler and Trede, 2016). The first two references are involved with construction of simultaneous confidence bands for the mean of functional data, but it should be noted that in the case of Gaussian functional data this problem and the issue of forecasting a new functional observation are essentially equivalent, and one can transform the simultaneous confidence bands for the mean into simultaneous prediction bands via a simple rescaling. The second class is represented by cases in which a dimensionality reduction technique is applied to render the naturally infinite-dimensional problem more tractable by projecting it on a finite dimensional functional basis (Hyndman and Shahid Ullah, 2007; Antoniadis et al., 2016). These approaches carry some shortcomings: the first group of techniques is computationally intensive, thus requiring long calculation times, while the second ones rely on the approximations introduced by basis projection. Both of them, in any case, either rely on not easily provable distributional assumptions and/or on asymptotic results.

The framework of this manuscript is Conformal Prediction (Vovk, Gammerman and Shafer, 2005; Shafer and Vovk, 2008), a novel method of forecasting firstly developed in the Machine Learning community as a way to define prediction intervals for Support Vector Machines (Gammerman, Vovk and Vapnik, 1998).

The interested reader can find a recent review in Fontana, Zeni and Vantini (2023). In univariate setting, Conformal Prediction is able to generate distribution-free, valid prediction intervals and it has also been used as a data exploration tool for Functional Data (Lei, Rinaldo and Wasserman, 2015), via the use of a truncated basis approach.

In this article, we build on top of the literature about set prediction for functional data and Conformal Prediction, by introducing several theoretical and methodological innovations.

1. After introducing the importance in interpretative terms of obtaining functional prediction sets having a specific shape (i.e., prediction bands), Section 2.2, functional prediction sets and introduces the Semi-Off-Line Inductive Conformal framework, also known simply as Split Conformal. Specifically, we contribute in two ways to the Conformal Prediction literature: first, by enriching the results about the validity of split conformal prediction sets by making the exact probability reached by them explicit (Thm. 1); and second, by providing what is to the best of our knowledge, the first formal proof of the exactness of smoothed split conformal prediction sets (Supplementary Material S1.1).

2. In Section 2.3, we propose a nonconformity measure inducing a conformal predictor able to create closed-form finite-sample either valid or exact prediction bands of constant amplitude, under minimal distributional assumptions. The procedure is fast, scalable and does not rely on widespread functional dimension reduction techniques.

3. In Section 2.4, we propose a family of nonconformity measures (to which the nonconformity measure introduced in Section 2.3 belongs) indexed by modulation function $s_{\mathcal{I}_1}$ that allows for prediction bands with non-constant width, but able to keep all the aforementioned appealing properties. As a consequence, prediction bands induced by the nonconformity measures belonging to this family can be compared on the basis of features other than validity, such as efficiency (i.e., the size).

4. In Section 2.4, we focus on a specific nonconformity measure belonging to this family, which leads to valid prediction bands that are asymptotically no less efficient than those obtained without modulation (Thm. 2 and 3).

Finally, in Section S2 in the Supplementary Material, we propose a simulation study to compare our method with four alternatives, and in Section 3, we apply our approach to the Berkeley Growth Study data set (Tuddenham and Snyder, 1954). Section 4 provides an overview of the main results.

2. Conformal Prediction Bands

2.1. The importance of being a band

Set prediction is an issue of key importance in the statistical community. Specifically, three main features characterize a prediction set: shape, coverage, and size. We start by tackling, in this section, the first issue, while the last two are explored in Section 2.2. In the classical multivariate statistical setting, elliptic regions have been and are still considered as the standard shapes for prediction sets. Differently, in the functional context many authors (López-Pintado and Romo, 2009; Lei, Rinaldo and Wasserman, 2015) note how the focus should be on a particular type of prediction set, commonly known as *prediction band*. Formally, a band is defined as

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

with $B_n(t) \subseteq \mathbb{R}$ interval for each $t \in \mathcal{T}$ (López-Pintado and Romo, 2009; Degras, 2017). The focus on this type of sets, that can be defined as the Cartesian product of the (infinitely many) intervals $\{B_n(t) : t \in \mathcal{T}\}$, comes from the fact that, differently from a generic region of $\mathcal{Y}(\mathcal{T})$, such a shape can be easily visualized on a plot (i.e., it is a band, in parallel coordinates, as noted by López-Pintado and Romo, 2009) and thus interpreted with respect to the domain $\mathcal{T}$.

It should also be noted that, differently from prediction sets characterized by other shapes, prediction bands always coincide with (and are not only a subset of) their envelope. In view of this, the development of a method that necessarily outputs prediction bands, instead of more general prediction sets, represents the starting point of this work.

2.2. Conformal prediction

The framework we use to develop our prediction sets is *Conformal Prediction*, a nonparametric approach proposed in the multivariate literature for the first time by Gammerman, Vovk and Vapnik (1998) and thoroughly described in Vovk, Gammerman and Shafer (2005), that can be used to construct finite-sample either valid or exact prediction sets under no assumptions other than *i.i.d.* data. For a review of the topic see, e.g., Lei et al. (2018) and Fontana, Zeni and Vantini (2023). Even though the theory holds also under the weaker assumption of exchangeable data, in this manuscript we will focus on the case of *i.i.d.* data which is a very common case in applications and in particular on the case of *i.i.d.* functional data taking values in $\mathcal{Y}(\mathcal{T})$.

Following the notation of Vovk, Gammerman and Shafer (2005), given a set of *i.i.d.* random functions $Y_1, \ldots, Y_n \sim P$ and an independent random function $Y_{n+1} \sim P$, a valid prediction set $\mathcal{C}_{n,1-\alpha} := \mathcal{C}_{n,1-\alpha}(Y_1, \ldots, Y_n)$ for Y_{n+1} is a set such that

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha}\right) \geq 1-\alpha \tag{2.1}$$

for any significance level $\alpha \in (0,1)$ and with $\mathbb{P}$ the probability corresponding to the product measure induced by P (Lei, Rinaldo and Wasserman, 2015). If the inequality in (2.1) is replaced by the equality, the prediction set is also said to be exact. In order to avoid ambiguity, later in the discussion the term *coverage* (or *unconditional coverage*) will be used to refer to $\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha}\right)$, the term *conditional coverage* will be used to refer to $\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha} | \mathcal{C}_{n,1-\alpha}\right)$ and the terms *empirical coverage* and *empirical conditional coverage* will be used to refer to the estimate, from simulated data, of the coverage and conditional coverage, respectively.

Specifically, we focus on the Semi-Off-Line Inductive Conformal framework, also known simply as *Split Conformal*, which is a computationally efficient modification of the original Transductive Conformal method first proposed by Papadopoulos et al. (2002). In order to present this approach, let us consider the following procedure: given data $y_1, \ldots, y_n$, let $\{1, \ldots, n\}$ be randomly divided into two sets $\mathcal{I}_1, \mathcal{I}_2$ and let us define the training set as $\{y_h : h \in \mathcal{I}_1\}$ and the calibration set as $\{y_h : h \in \mathcal{I}_2\}$, with $|\mathcal{I}_1| = m$, $|\mathcal{I}_2| = l$ and $m, l \in \mathbb{N}_{>0}$ such that $n = m + l$. Let us also define *nonconformity measure* as any measurable function $A(\{y_h : h \in \mathcal{I}_1\}, y)$ taking values in $\bar{\mathbb{R}}$ whose aim is to score how different $y \in \mathcal{Y}(\mathcal{T})$ is from the training set. The split conformal prediction set constructed on the basis of the observed sample $y_1, \ldots, y_n$ is defined as $\mathcal{C}_{n,1-\alpha} := \{y \in \mathcal{Y}(\mathcal{T}) : \delta_y > \alpha\}$, with

$$\delta_y := \frac{|\{j \in \mathcal{I}_2 \cup \{n+1\} : R_j \geq R_{n+1}\}|}{l+1}$$

and *nonconformity scores* $R_j := A(\{y_h : h \in \mathcal{I}_1\}, y_j)$ for $j \in \mathcal{I}_2$, $R_{n+1} := A(\{y_h : h \in \mathcal{I}_1\}, y)$. In particular, hereafter we will focus on nonconformity scores $\{R_h : h \in \mathcal{I}_2\}$ having a continuous joint distribution, an assumption generally satisfied in the functional context.

The essential result, due to Vovk, Gammerman and Shafer (2005), traditionally evoked when dealing with the Conformal approach concerns the validity of split prediction sets. Namely, under the exchangeability assumption (a direct consequence of having i.i.d. data) δ_Y is uniformly distributed over $\{1/(l+1), 2/(l+1), \ldots, 1\}$, and then (2.1) holds. Theorem 1 proves and enriches such known result by making the exact probability reached by split prediction sets explicit. The proof is given in Supplementary Material S1.1.

Theorem 1. *Let $\mathcal{C}_{n,1-\alpha}$ be a split conformal prediction set. If $Y_1, \ldots, Y_{n+1}$ are i.i.d. and $\{R_h : h \in \mathcal{I}_2\}$ have a continuous joint distribution, then*

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha}\right) = 1 - \frac{\lfloor (l+1)\alpha \rfloor}{l+1}.$$

Specifically, $\mathcal{C}_{n,1-\alpha}$ always satisfies

$$1-\alpha \leq \mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha}\right) < 1-\alpha+\frac{1}{l+1}. \tag{2.2}$$

A natural consequence of the first part of Theorem 1 is that when $\lfloor (l+1)\alpha \rfloor = (l+1)\alpha$ the procedure automatically outputs exact prediction sets: in practice, since in most cases both α and l are given by the application in hand, such property should be simply considered as an useful by-product that may occur in some circumstances. More generally, Theorem 1 states that the Conformal approach ensures an easy-to-compute precise coverage for split prediction sets, and not only their validity. Furthermore, the second part of Theorem 1 suggests that the coverage provided by split conformal prediction sets is no less than $1-\alpha$ and over-coverage is basically avoided when sample size is large. In particular, inequality (2.2) represents a minimal modification of Theorem 2 of Lei et al. (2018): the only difference, besides notation, is the change of "$\leq$" with "$<$" in the upper bound of (2.2).

Conformal inference is a field of deep interest as minimal assumptions are required on P to obtain prediction sets satisfying (2.1) for any finite sample size n, a property particularly appealing in the functional context. A slight modification (Vovk, Gammerman and Shafer, 2005) of the aforementioned procedure even allows to obtain a stronger version of Theorem 1. To present it, first of all let us introduce an element of randomization τ_{n+1}, realization of a uniform random variable in $[0,1]$. The smoothed split conformal prediction set is defined as $\mathcal{C}_{n,1-\alpha,\tau_{n+1}} := \left\{y \in \mathcal{Y}(\mathcal{T}) : \delta_{y,\tau_{n+1}} > \alpha\right\}$, with

$$(l+1)\cdot\delta_{y,\tau_{n+1}} := \left|\{j \in \mathcal{I}_2 : R_j > R_{n+1}\}\right| + \tau_{n+1}\left|\{j \in \mathcal{I}_2 \cup \{n+1\} : R_j = R_{n+1}\}\right|.$$

Smoothed split conformal prediction sets are, by construction, exact for any α, l, i.e., $\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha,\tau_{n+1}}\right) = 1-\alpha$. To the best of our knowledge, in the literature there is no formal proof of this well-established result, due to Vovk, Gammerman and Shafer (2005), and so a proof is given in Supplementary Material S1.1.

Remark 1. Our discussion was limited to the split setting because our work only focuses on it, but the results of this section are very general and require just little changes to be applied to the Transductive/Full Conformal framework. In addition, as highlighted by Vovk, Gammerman and Shafer (2005) and briefly mentioned at the beginning of this section, Theorem 1 and the result about exactness of smoothed prediction sets hold even when the weaker assumption of exchangeability is formulated instead of the traditional hypothesis of i.i.d. data.

Remark 2. The division of data into the training and calibration sets always induces an element of randomness into the procedure, also in the non-smoothed scenario. A possible approach to limit the effect of this evidence consists of combining prediction sets obtained from different splits, but the results provided

by Lei et al. (2018) suggest to perform a single split. As a consequence, in this article the aforementioned single-split process is considered. The computation of the effect of splitting, as well as the impact of the specific values m and l, on the procedure has not yet been properly analyzed in the Conformal Prediction literature (Fontana, Zeni and Vantini, 2023), but it is a topic worth of further research.

Remark 3. The Conformal approach can be also successfully applied to regression and classification problems. A detailed presentation is not included hereafter being out of scope, but an exhaustive discussion can be found in Vovk, Gammerman and Shafer (2005).

Remark 4. Although we focus on the functional setting, the Conformal framework has initially been developed in the traditional univariate and multivariate settings and so all arguments and results presented in this section can also be applied to univariate variables and random vectors.

2.3. The nonconformity measure

Although some authors proposed different approaches to find prediction bands under the Gaussian assumption (Yao, Müller and Wang, 2005) and through finite dimensional projection (Lei, Rinaldo and Wasserman, 2015), to the best of our knowledge, no method to create valid prediction bands by only assuming i.i.d. functional data and by avoiding dimension reduction is available in the literature.

In light of this and of the discussion in Section 2.1, we propose a fast and scalable split conformal predictor that outputs closed-form finite-sample valid (or even exact) prediction bands under only the i.i.d. assumption. Indeed, the Conformal framework ensures, by construction, that the prediction sets obtained are always valid, but other features such as shape and size depend on the specific nonconformity measure used: as a consequence, the core of the Conformal approach is represented by the choice of such measure.

In particular, the nonconformity measure we propose automatically allows to obtain prediction bands and is based on the essential supremum:

$$A(\{y_h : h \in \mathcal{I}_1\}, y) = \operatorname*{ess\,sup}_{t \in \mathcal{T}} |y(t) - g_{\mathcal{I}_1}(t)|, \tag{2.3}$$

with $g_{\mathcal{I}_1} : \mathcal{T} \to \mathbb{R}$ a function belonging to $L^\infty(\mathcal{T})$ based on $\{y_h : h \in \mathcal{I}_1\}$ and acting as a point predictor of the new observation. Although valid prediction bands are obtained regardless the specific $g_{\mathcal{I}_1}$ involved, a careful choice of this function helps to obtain small prediction bands, a desirable property from an application point of view which will be investigated in Section 2.4 (Lei et al., 2018). In view of this, $g_{\mathcal{I}_1}$ is typically a point predictor summarizing information provided by $\{y_h : h \in \mathcal{I}_1\}$, e.g., the sample functional mean. However, since the purpose of the article is to construct either valid or exact prediction bands

starting from any point predictor in order to obtain a widely usable procedure, later in the discussion, we will always consider $g_{\mathcal{I}_1}$ as given—and properly chosen by the expert according to the specific framework considered. Focusing on the non-smoothed scenario (the minor changes needed for the smoothed case are introduced in Supplementary Material S1.4), first of all it is possible to notice that if $\alpha \in (0, 1/(l+1))$, then $\mathcal{C}_{n,1-\alpha} = \mathcal{Y}(\mathcal{T})$, since δ_y can not be less than $1/(l+1)$. For this reason, later in the discussion we will always consider $\alpha \in [1/(l+1), 1)$, unless otherwise stated. If $\alpha \in [1/(l+1), 1)$, the definition of $\mathcal{C}_{n,1-\alpha}$ and δ_y implies that $y \in \mathcal{C}_{n,1-\alpha} \iff R_{n+1} \leq k$, with k the $\lceil (l+1)(1-\alpha) \rceil$th smallest value in the set $\{R_h : h \in \mathcal{I}_2\}$. Then

$$\begin{aligned} \operatorname*{ess\,sup}_{t \in \mathcal{T}} |y(t) - g_{\mathcal{I}_1}(t)| \leq k \iff \\ |y(t) - g_{\mathcal{I}_1}(t)| \leq k \quad \forall t \in \mathcal{T} \iff \\ y(t) \in [g_{\mathcal{I}_1}(t) - k, g_{\mathcal{I}_1}(t) + k] \quad \forall t \in \mathcal{T}. \end{aligned}$$

Therefore, the split conformal prediction set induced by the nonconformity measure (2.3) is

$$\mathcal{C}_{n,1-\alpha} := \{y \in \mathcal{Y}(\mathcal{T}) : y(t) \in [g_{\mathcal{I}_1}(t) - k, g_{\mathcal{I}_1}(t) + k] \quad \forall t \in \mathcal{T}\}. \tag{2.4}$$

Besides having the shape of a a band, the introduced prediction set can be found in closed form, an appealing property that incredibly speeds up computation time. In addition, the Conformal framework and the simplicity of the nonconformity measure ensure highly scalable prediction bands as, on top of the cost needed to build the point predictor $g_{\mathcal{I}_1}$, the time required to find k increases linearly with l. Then, if a particularly sophisticated predictor is chosen for $g_{\mathcal{I}_1}$, one is justified in expecting the total computation cost to be dominated by the calculation of such point predictor. Moreover, as usual in the prediction framework the band is built around a "central" object ($g_{\mathcal{I}_1}$ in this case), a fact that further suggests to define this function as a data-driven point predictor. Finally, the prediction bands defined in (2.4) are simultaneous by construction, i.e., bands ensuring the desired coverage globally (in addition to the pointwise validity). Similarly to the multivariate setting, a simple concatenation of pointwise prediction intervals based on the pointwise nonconformity score $|y(t) - g_{\mathcal{I}_1}(t)|$ for all $t \in \mathcal{T}$ would lead to a prediction band. This is a subset of the simultaneous prediction band (2.4) (the proof is given in Supplementary Material S1.2), which guarantees pointwise coverage for all $t \in \mathcal{T}$, but whose simultaneous coverage over the domain $\mathcal{T}$ can be dramatically lower than the desired one.

Remark 5. In application scenarios where data are characterized by specific features (e.g., positivity, monotonicity etc.), the approach presented in this Section allows to remove portions of the observed prediction bands that violate

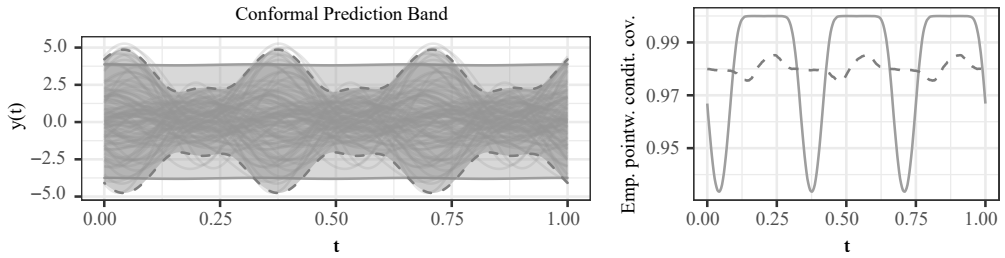

Figure 1. The left panel shows the split conformal prediction band computed as in (2.4) (solid light grey band) and that computed as in (2.6) by considering the standard deviation function as $s_{\mathcal{I}_1}$ (dashed band). For visualization, a random subsample of $y_1, \ldots, y_{198}$ is plotted. The right panel shows the empirical pointwise conditional coverage reached by the first band (solid light grey line) and by the second one (dashed line). $\alpha = 0.1$.

such known characteristics, without affecting the coverage. An example of this band trimming procedure is given in Section 3. This possibility is a desirable implication which derives from using a fully nonparametric approach to prediction, since this takes away the burden of an explicit and possibly non-trivial modeling of the existing constraints.

2.4. Improving efficiency: The choice of the modulation function

It can be easily noted that the width of (2.4) over $\mathcal{T}$ is constant and equal to $2k$ but, intuitively, prediction bands that do not adapt their width according to the local variability of functional data, even though theoretically sound, may be of limited interest in real applications. For this reason it is of key importance to create prediction bands whose width can be adapted to the local variability of functional data.

Let us consider the following running example: let $y_1, \ldots, y_{198}$ be independent realizations of the random function $Y(t) := X_1 + X_2 \cos(6\pi t) + X_3 \sin(6\pi t)$, with $t \in [0, 1]$ and (X_1, X_2, X_3) being a Gaussian random vector such that $\mathrm{E}(X_i) = 0$, $\mathrm{Var}(X_i) = 1$, $\mathrm{Cov}(X_i, X_j) = 0.6$ for $i, j = 1, 2, 3$, $i \neq j$. The solid light grey band in the left panel of Figure 1 shows the prediction band obtained by the procedure presented in Section 2.3 considering $\alpha = 0.1$, $m = n/2$ and $g_{\mathcal{I}_1}$ sample functional mean of the training set: given the different variability of functional data over $\mathcal{T}$, in the low-variance parts of the domain the prediction band is dramatically large containing all the pointwise evaluations of the functional data (see, for example, $t = 0.5$ and nearby points).

A possible solution to this drawback consists of defining the following nonconformity measure and nonconformity scores:

$$A(\{y_h : h \in \mathcal{I}_1\}, y) = \operatorname*{ess\,sup}_{t \in \mathcal{T}} \left| \frac{y(t) - g_{\mathcal{I}_1}(t)}{s_{\mathcal{I}_1}(t)} \right|, \tag{2.5}$$

$$R_j^s := \operatorname*{ess\,sup}_{t\in\mathcal{T}} \left| \frac{y_j(t) - g_{\mathcal{I}_1}(t)}{s_{\mathcal{I}_1}(t)} \right|,$$
$$R_{n+1}^s := \operatorname*{ess\,sup}_{t\in\mathcal{T}} \left| \frac{y(t) - g_{\mathcal{I}_1}(t)}{s_{\mathcal{I}_1}(t)} \right|,$$

with $j \in \mathcal{I}_2$ and $s_{\mathcal{I}_1} := s(\{y_h : h \in \mathcal{I}_1\}) : \mathcal{T} \to \mathbb{R}_{>0}$ a function which belongs to $L^\infty(\mathcal{T})$ based on $\{y_h : h \in \mathcal{I}_1\}$. At the interpretative level, the new nonconformity measure (2.5) can be suitably considered as the nonconformity measure (2.3) taking the transformed functions $y^s(t) := y(t)/s_{\mathcal{I}_1}(t)$ and $g^s_{\mathcal{I}_1}(t) = g_{\mathcal{I}_1}(t)/s_{\mathcal{I}_1}(t)$ $\forall t \in \mathcal{T}$ as input instead of the original functions $y(t)$, $g_{\mathcal{I}_1}(t)$. It is important to notice that, since $s_{\mathcal{I}_1}(t) > 0$ $\forall t \in \mathcal{T}$, the function $s_{\mathcal{I}_1}$ modulates the original data without altering the order of the functions at each point t: for this reason, later in the discussion the term *modulation function* will be used to refer to $s_{\mathcal{I}_1}$.

Therefore, the split conformal prediction band induced by the nonconformity measure (2.5), obtained by replicating the computations of Section 2.3 (see Supplementary Material S1.3 for the proof), is

$$\mathcal{C}^s_{n,1-\alpha} := \{y \in \mathcal{Y}(\mathcal{T}) : y(t) \in [g_{\mathcal{I}_1}(t) - k^s s_{\mathcal{I}_1}(t), g_{\mathcal{I}_1}(t) + k^s s_{\mathcal{I}_1}(t)] \ \forall t \in \mathcal{T}\}, \quad (2.6)$$

with k^s the $\lceil (l+1)(1-\alpha) \rceil$th smallest value in the set $\{R^s_h : h \in \mathcal{I}_2\}$. In other words, the procedure presented in this section consists of modulating the data, computing the prediction band (2.4) by using the transformed data and back-transforming it in the non-modulated space: in so doing, prediction bands adapt their width according to the specific modulation function chosen and their validity is guaranteed by the Conformal framework. A similar consideration has been highlighted also in the scalar regression setting by Lei et al. (2018), who proposed a locally weighted Split Conformal method to vary the width of the prediction sets over the covariates $x \in \mathbb{R}^p$.

In order to understand the modification introduced by the modulation function, let us consider the aforementioned running example and specifically the left panel of Figure 1: in this case, the band obtained by considering the standard deviation function (Ramsay and Silverman, 2005) as $s_{\mathcal{I}_1}$ (dashed band) is deeply different from the one in the top panel and it seems to better adapt to the variability of the data over $\mathcal{T}$. Intuitively, one is justified in accepting the bands to become wider in the parts of the domain where data show high variability in order to obtain narrower and more informative prediction bands in those parts characterized by low variability.

Remark 6. Replacing function $s_{\mathcal{I}_1}$ with $s_{\mathcal{I}_2}$ does not allow to obtain closed-form valid prediction bands. This is due to the fact that their dependence on the calibration set involves $\{R^s_h : h \in \mathcal{I}_2 \cup \{n+1\}\}$ not being exchangeable, and consequently validity not being guaranteed.

Remark 7. Prediction bands induced by the modulation functions $s_{\mathcal{I}_1}$ and $\lambda \cdot s_{\mathcal{I}_1}$, with $\lambda \in \mathbb{R}_{>0}$, are identical. The proof is given in Supplementary Material S1.3. As a consequence, an equivalence relation naturally arises and so for each specific equivalence class (made up of modulation functions equal up to a multiplicative factor) we will consider the modulation function whose integral is equal to 1. In view of this, the original nonconformity measure (2.3) can be interpreted as the nonconformity measure induced by the modulation function $s^0(t) := 1/|\mathcal{T}|$ $\forall t \in \mathcal{T}$, whose notation does not include the subscript $\mathcal{I}_1$ to underline the lack of dependence of this function on the training set.

Remark 8. One of the aims of the introduction of $s_{\mathcal{I}_1}$ is to reduce the variability of the pointwise miscoverage over $\mathcal{T}$. In order to clarify this concept, let us consider the right panel of Figure 1. The solid light grey (dashed respectively) line shows the empirical pointwise conditional coverage of the solid light grey (dashed respectively) prediction band showed in the left panel of the same figure, that was obtained by setting $\alpha = 0.1$. The empirical conditional coverage has been computed considering the number of times that 200,000—independent from and identically distributed to the original sample—new functions belong to the two prediction bands over $\mathcal{T}$. As expected, the absence of modularization involves the empirical pointwise converage being highly variable over $\mathcal{T}$, whereas the use of the standard deviation function as modulation function leads to an empirical pointwise coverage concentrated around 0.98.

However, in absence of an optimality criterion there are no formal reasons to prefer a specific modulation function over another, as the Conformal approach ensures valid prediction sets regardless the choice of $s_{\mathcal{I}_1}$. In this regard, a criterion that naturally arises in the prediction framework to discriminate between modulation functions is maximization of efficiency, i.e., minimization of the size of prediction sets (Vovk, Gammerman and Shafer, 2005) . The reason of this choice is very intuitive: since prediction bands are, by construction, valid, one is justified in seeking small prediction bands because they include subregions of the sample space where the probability mass is concentrated (Lei, Robins and Wasserman, 2013). In view of this, first of all it is essential to define what the size of a prediction band is, a nontrivial topic in the functional framework. The definition we will consider is simply the area between the upper and lower bound of the prediction band:

$$\mathcal{Q}(s_{\mathcal{I}_1}) := \int_{\mathcal{T}} 2 \cdot k^s \cdot s_{\mathcal{I}_1}(t)dt = 2 \cdot k^s, \tag{2.7}$$

that is equal to k^s up to a constant and proportional to $2k^s/|\mathcal{T}|$, i.e., the average width of the prediction band over the domain $\mathcal{T}$.

Formally, in the usual finite-dimensional setting the aim would be to find the optimal modulation function that minimizes the risk functional $\mathrm{E}[k^s]$. Unfortu-

nately, in the functional setting even the concept of probability density function is generally not well defined since there is no σ-finite dominating measure (Delaigle and Hall, 2010), and so that minimization is not feasible for general P. As a consequence, the minimization problem must be simplified: by considering k^s as a non-random quantity depending on observed functions $y_1, \ldots, y_n$ instead of random functions $Y_1, \ldots, Y_n$, the aim becomes the direct minimization of k^s. Although initially it may seem like an oversimplification to some readers, it is important to underline that this approach is made possible by a well-established principle representing the core idea of many algorithms and methods (e.g., machine learning techniques) known as empirical risk minimization principle (Vapnik, 1992).

The proposed adjustment reduces the complexity of the optimization task, but the problem still presents tricky aspects. Indeed, not only the minimization can not be analytically addressed by calculus of variations given the complexity of k^s, but also the optimal modulation function can not be uniquely determined given the specific structure of $R_h^s, h \in \mathcal{I}_2$. In fact, the dependency of $s_{\mathcal{I}_1}$ only on the functions of the training set and of the numerator of R_h^s (i.e., $|y_h(t) - g_{\mathcal{I}_1}(t)|, \quad h \in \mathcal{I}_2$) also on the functions of the calibration set makes the optimization unfeasible for all P and the general problem ill-posed.

In such a non-standard context, the line of reasoning must necessarily be changed. Therefore, in the discussion below we focus on finding a function (called c-function hereafter for the sake of simplicity) satisfying the definition of modulation function but depending also on the calibration set through $\{y_h : h \in \mathcal{I}_2\}$ and such that

1. For $m, l \to +\infty$ it converges to a given function and its training counterpart (i.e., the function, called t-function hereafter, equal to the c-function but whose dependence on $\{y_h : h \in \mathcal{I}_2\}$ is replaced by the dependence on the training set through $\{y_h : h \in \mathcal{I}_1\}$) converges to the same function;

2. it leads to prediction bands that are not wider (in the sense of (2.7)) than those obtained by not modulating (i.e., by using s^0).

If these two conditions are met, the use of the t-function as modulation function ensures that valid prediction bands are obtained (due to its dependence only on $\{y_h : h \in \mathcal{I}_1\}$) and that asymptotically the second condition is satisfied. Specifically, that condition represents a desirable and appealing property since, if violated, the modulation process could represent a meaningless complication compared to the original nonconformity measure (2.3).

In order to construct a c-function able to meet these two conditions, it is important to focus on what k^s is. Ignoring just for now the contribution of the modulation function, k^s is a quantity derived by the $\lceil (l+1)(1-\alpha) \rceil$th least conforming function between those in the calibration set, in which the concept of "conformity" is induced by the metric used, being deemed as "conforming"

a function whose distance from g_{I_1} is particularly high. In light of this, the guidelines we decided to follow in the construction of a meaningful c-function are two. First of all, the behavior of the $l - \lceil (l+1)(1-\alpha) \rceil$ most extreme functions in the calibration set should not be taken into account since they do not affect the value of k^s. Secondly, given the specific nonconformity measure considered, the c-function should modulate data considering the remaining $\lceil (l+1)(1-\alpha) \rceil$ functions on the basis of the most extreme value observed $\forall t \in \mathcal{T}$.

Inspired by these guidelines, we propose the following c-function:

$$\bar{s}^c_{\mathcal{I}_1}(t) := \frac{\max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|}{\int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt} \tag{2.8}$$

with

$$\mathcal{H}_2 := \{j \in \mathcal{I}_2 : \operatorname*{ess\,sup}_{t \in \mathcal{T}} |y_j(t) - g_{\mathcal{I}_1}(t)| \leq k\}$$

and k defined as in Section 2.3, i.e., the $\lceil (l+1)(1-\alpha) \rceil$th smallest value in the set $\{R_h : h \in \mathcal{I}_2\}$. The corresponding t-function is

$$\bar{s}_{\mathcal{I}_1}(t) := \frac{\max_{j \in \mathcal{H}_1} |y_j(t) - g_{\mathcal{I}_1}(t)|}{\int_{\mathcal{T}} \max_{j \in \mathcal{H}_1} |y_j(t) - g_{\mathcal{I}_1}(t)| dt} \tag{2.9}$$

with $\mathcal{H}_1 = \mathcal{I}_1$ if $\lceil (m+1)(1-\alpha) \rceil > m$, otherwise

$$\mathcal{H}_1 := \{j \in \mathcal{I}_1 : \operatorname*{ess\,sup}_{t \in \mathcal{T}} |y_j(t) - g_{\mathcal{I}_1}(t)| \leq \gamma\}$$

with γ the $\lceil (m+1)(1-\alpha) \rceil$th smallest value in the set $\{\operatorname{ess\,sup}_{t \in \mathcal{T}} |y_h(t) - g_{\mathcal{I}_1}(t)| : h \in \mathcal{I}_1\}$.

In order not to overcomplicate the notation, in the definition of $\bar{s}^c_{\mathcal{I}_1}$ and $\bar{s}_{\mathcal{I}_1}$ we quietly assumed that both numerators are different from 0 $\forall t \in \mathcal{T}$ almost surely. If not, the adjustment described in Supplementary Material S1.3 is developed. From an operational point of view, t-function $\bar{s}_{\mathcal{I}_1}(t)$ ignores the most extreme functions (i.e., the functions belonging to $\mathcal{I}_1 \setminus \mathcal{H}_1$) and modulates data on the basis of the remaining non-extreme functions. Specifically, the dependence of γ on α allows to provide carefully chosen modulation process according to the specific level $1 - \alpha$ chosen for the prediction set.

The fulfillment of the two aforementioned conditions by the function (2.8) is proved by the following two theorems.

Theorem 2. *Let $m/n = \theta$ with $0 < \theta < 1$ and let $\mathrm{Var}\{g_{\mathcal{I}_1}(t)\} \to 0$ when $m \to +\infty$. Then $\bar{s}^c_{\mathcal{I}_1}$ and $\bar{s}_{\mathcal{I}_1}$ converge to the same function when $n \to +\infty$ and $\lim_{n \to +\infty} \mathcal{C}^{\bar{s}}_{n,1-\alpha} = \lim_{n \to +\infty} \mathcal{C}^{\bar{s}^c}_{n,1-\alpha}$ $\forall$ $\alpha \in (0,1)$.*

Theorem 3. *$\mathcal{Q}(s^0) \geq \mathcal{Q}(\bar{s}^c_{\mathcal{I}_1})$. Specifically, $\mathcal{Q}(s^0) = \mathcal{Q}(\bar{s}^c_{\mathcal{I}_1})$ if and only if $\max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|$ is constant almost everywhere.*

Both proofs are given in Supplementary Material S1.3. It is important to notice that Theorem 2 requires very mild conditions, an evidence that allows it to hold in many general contexts.

In light of this, the function (2.9) represents an outstanding candidate in the choice of the modulation function since the Conformal setting and the nonconformity measure (2.5) guarantee valid prediction bands, as well as all the other desirable properties highlighted in Section 2.3, and at the same time to asymptotically obtain prediction bands no less efficient than those induced by s^0.

Remark 9. The fact that $\bar{s}^c_{\mathcal{I}_1}(t)$ leads to prediction bands that are not wider than those obtained by not modulating is not the only relevant result that is possible to obtain. The following Theorem shows that prediction bands induced by $\bar{s}^c_{\mathcal{I}_1}$ are also smaller than those induced by the functions belonging to a specific group. This theorem provides a further theoretical justification for preferring function (2.9) to other possible modulation functions.

Theorem 4. *Let us define* $\mathcal{CH}_2 := \mathcal{I}_2 \setminus \mathcal{H}_2$ *and let* t_i^* *be the value such that*

$$|y_i(t_i^*) - g_{\mathcal{I}_1}(t_i^*)| = \operatorname*{ess\,sup}_{t\in\mathcal{T}} |y_i(t) - g_{\mathcal{I}_1}(t)| \quad \forall i \in \mathcal{I}_2. \tag{2.10}$$

If t_i^* *is not unique, it is randomly chosen from the values that satisfy* (2.10).

Let $s^d_{\mathcal{I}_1}$ *be a modulation function such that:*

1. $s^d_{\mathcal{I}_1} \neq \bar{s}^c_{\mathcal{I}_1}$ *in the sense of Lebesgue, i.e.,* $\exists\ \mathcal{T}^* \subseteq \mathcal{T}$ *such that* $s^d_{\mathcal{I}_1}(t) \neq \bar{s}^c_{\mathcal{I}_1}(t)$ $\forall t \in \mathcal{T}^*$ *and* $\mu(\mathcal{T}^*) > 0$*, with* μ *the Lebesgue measure*
2. $s^d_{\mathcal{I}_1}(t_i^*) \leq \bar{s}^c_{\mathcal{I}_1}(t_i^*)\ \forall i \in \mathcal{CH}_2$

If $|\mathcal{H}_2| = \lceil (l+1)(1-\alpha) \rceil$*, then* $\mathcal{Q}(s^d_{\mathcal{I}_1}) > \mathcal{Q}(\bar{s}^c_{\mathcal{I}_1})$.

The proof is given in Supplementary Material S1.3, along with the demonstration that Theorem 3 is not a direct consequence of Theorem 4 since s^0 may not fulfill $s^0(t_i^*) \leq \bar{s}^c_{\mathcal{I}_1}(t_i^*)\ \forall i \in \mathcal{CH}_2$. Also in this case, the field of application of Theorem 4 is particularly wide since the condition about the cardinality of $|\mathcal{H}_2|$ is always met under the assumption concerning the continuous joint distribution of $\{R_h : h \in \mathcal{I}_2\}$ made in Section 2.2.

Remark 10. The definitions of functions (2.8), (2.9) and Theorems 2, 3 and 4 can be easily generalized to hold also in the Smoothed Conformal framework. Technical details are provided in Supplementary Material S1.4.

Remark 11. As it is the case with the mean function g_{I_1} that is chosen according to the specific applicative problem at hand, the choice of the modulation function s_{I_1} has to be performed in a similar fashion. Indeed it has to be selected on a case-by-case basis by taking into consideration the specific characteristics of the modelling task at hand (e.g., homoscedasticity/heteroscedasticity along the domain or presence/absence of outliers).

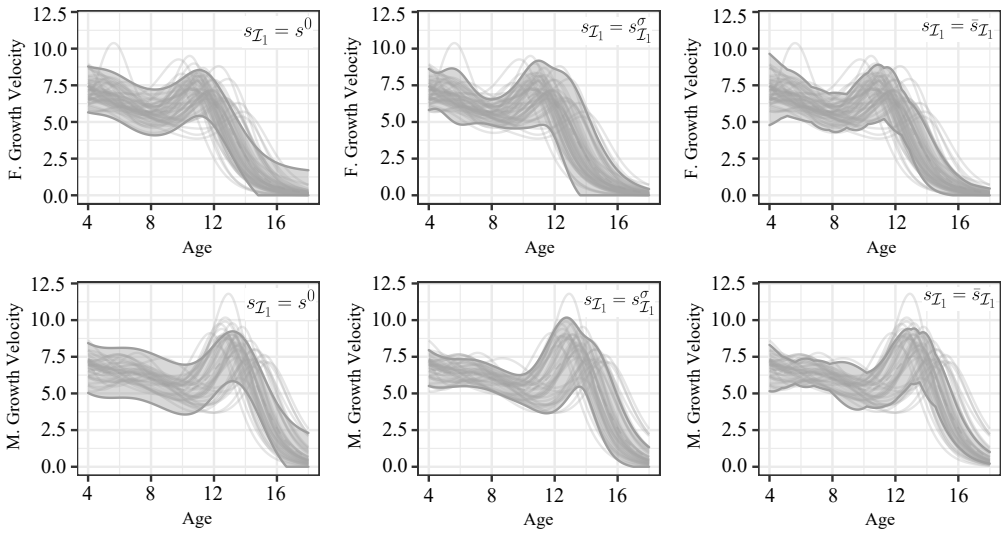

Figure 2. Berkeley Growth Study data: each panel shows the prediction band obtained considering a different modulation function (s^0 on the left, the normalised pointwise standard deviation function $s^\sigma_{\mathcal{I}_1}$ in the middle, $\bar{s}_{\mathcal{I}_1}$ on the right). In all cases, the dashed line represents $g_{\mathcal{I}_1}$. Predictions for girls at the top and predictions for boys at the bottom.

3. Application

In order to show the wide generality of our approach, in this section we apply our Conformal approach to a well known data set in the FDA community (i.e., the Berkeley Growth Study data set (Tuddenham and Snyder, 1954)) that is characterized by features that cannot be trivially framed in a standard probabilistic parametric model, such as heteroscedasticity along the functional domain, phase misalignment, presence of outlier curves, and positivity constraint. The specific data set contains in detail the heights (in cm) of 54 female and 39 male children measured quarterly from 1 to 2 years, annually from 2 to 8 years and biannually from 8 to 18 years. We focus on the first derivative of the growth curves, which are estimated in a standard fashion by the R function *smooth.monotone* of *fda* package (Ramsay et al., 2020) implementing monotonic cubic regression splines (Ramsay and Silverman, 2005, Chap. 6). Specifically, the prediction bands here reported refer to the growth velocity curves between 4 and 18 years for girls and boys separately comparing, in the Non-Smoothed Conformal framework, the three modulation functions analyzed in Section S2 and with $g_{\mathcal{I}_1}$ being simply for each group the corresponding functional sample mean, $\alpha = 0.5$, $m = 27$ for girls, $m = 20$ for boys.

The prediction bands are shown in Figure 2. Note that since the application at hand does not allow the functions to be negative in any subset of the domain, the prediction bands can be (and are indeed) truncated to 0 without decreasing their coverage.

Table 1. Berkeley Growth Study data: average width of the prediction bands.

	s^0	$s^{\sigma}_{\mathcal{I}_1}$	$\bar{s}_{\mathcal{I}_1}$
Females	2.904	3.244	2.811
Males	3.334	3.107	2.690

Table 2. Berkeley Growth Study data: average width of the prediction bands for different point predictors, using $\bar{s}_{\mathcal{I}_1}$ as the modulation function.

	Mean	Median	Trimmed mean
Females	2.811	3.614	2.910
Males	2.690	4.362	3.266

Focusing on Figure 2, the graphical representation of the prediction bands highlights the well-known different growth path between girls and boys, in which the latter group typically starts to grow later but achieves higher growth velocities. In terms of the role of modulation functions, their impact on female growth velocity prediction seems to be less than the one on the male bands. From a prediction point of view, girls' curves represent a simpler scenario in which the variance is lower along the domain, while boys' curves represent a more complex scenario with strong heteroskedasticity of the functions over $\mathcal{T}$ (due to the joint presence of misalignment of data and a very localized high peak around 13 years of age). As expected from these considerations, the prediction bands for a new girl's velocity curve obtained using the different modulation functions are relatively similar, with the prediction band associated to $\bar{s}_{\mathcal{I}_1}$ being slightly narrower due to the presence of outliers. Instead focusing on boys' curves, the strong heteroskedasticity forces the prediction band induced by s^0 to be uselessly large in some parts of the domain, whereas in general the prediction band induced by $s^{\sigma}_{\mathcal{I}_1}$ seems to be smoother than that induced by $\bar{s}_{\mathcal{I}_1}$, whose "bumps" are idiosyncratic and caused by the specific modulation function used which is not pointwise related to an average but on the specific value assumed by one of the functions. This creates narrower but less smooth bands. Both for boys and girls $\bar{s}_{\mathcal{I}_1}$ outputs the smallest prediction band, as shown in Table 1 where the quantity $\mathcal{Q}(\cdot)/|\mathcal{T}|$ is reported.

Additionally, we have explored the role that the pointwise predictor covers with respect to the prediction perfomance in this applied case. The explored methods are, similarly to the simulation study, a baseline method, represented by the sample mean (stylised "*Mean*", accompanied by the functional median (stylised "*Median*"), where the point predictor is represented by the deepest curve of the sample, according to MBD. and by a trimmed functional mean, computed excluding the 10% of the shallowest curves in the sample, again according to MBD. The the modulation function selected is $\bar{s}_{\mathcal{I}_1}$.

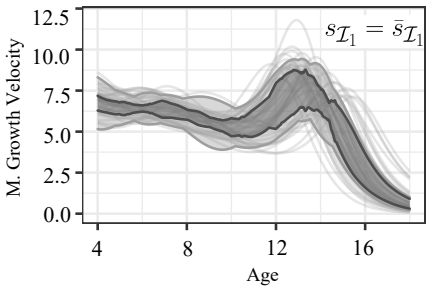

Figure 3. Berkeley Growth Study data: the prediction band represented at the bottom right of Figure 2 (light grey) and the corresponding pointwise conformal prediction band (dark grey).

Some useful information can be also provided by the comparison between the proposed approach and its pointwise counterpart, in which the prediction band is constructed by applying a coherent univariate Conformal approach at each point t separately. Indeed, by construction the former creates prediction bands larger or equal than those obtained by the latter, but on the other hand it guarantees simultaneous (and not pointwise) validity and of course it interprets a function as a whole, a key aspect in the functional context. In order to clarify this concept, let us consider Figure 3, in which the pointwise prediction band (dark grey) is overlaid to the bottom-right panel of Figure 2. As expected, the pointwise prediction band is simply modulated by the local variability of the 50% central curves. Differently, the prediction bands here proposed instead take also into consideration the behavior of the functions along the domain $\mathcal{T}$ with the effect of generating narrower or wider bands also in presence of similar local variabilities and so not just obtaining a simple expansion of the pointwise prediction band.

4. Conclusion

The creation of prediction sets for functional data is still an open problem of paramount importance in statistical methodology research. In order to define and compute them, the great majority of methods currently presented in the literature rely on non-provable distributional assumption, dimension reduction techniques and/or asymptotic arguments. On the contrary, the approach proposed in this article represents an innovative proposal in this field. Specifically, the Conformal framework ensures that finite-sample either valid or exact prediction sets are obtained under minimal distributional assumptions, whereas the specific family of nonconformity measures introduced guarantees, besides prediction sets that are bands, also a fast, scalable and closed-form solution. Moreover, despite the fact that our approach works regardless the specific choice of $s_{\mathcal{I}_1}$ (which can be chosen, for example, a priori), we proposed a specific data-driven modulation function, namely $\bar{s}_{\mathcal{I}_1}$, which leads to prediction bands asymptotically no less efficient than

those obtained by not modulating. The focus of this article was on i.i.d. data, but we envision an extension of the procedure to regression and classification problems.

Our procedure is able to achieve encouraging results and could represent a promising starting point for future developments, but at least two aspects, among others, should be carefully investigated. First of all, the division of data into the training and calibration sets induces an intrinsic element of randomness into the method and, although this phenomenon is well known in the Conformal literature, a quantification of the effect of the split process, and also of the values m and l, on the procedure has not yet been properly analyzed. Secondly, the prediction sets proposed in this article are purposely shaped as functional bands. This geometrical characterization in most application scenarios can be considered well suited. Nevertheless, one can think at more complicated scenarios (e.g., functional mixtures) where a prediction set made of multiple bands could be considered more suited from an application point of view. This possible extension will be the object of future work.

Supplementary Material

The supporting material contains the technical proofs and the supplementary material for the simulation study.

Acknowledgments

The authors would like to thank Giulio Pegorer for fruitful discussions. Prof. Vantini and Dr. Fontana acknowledge the financial support from Accordo Quadro ASI-POLIMI "Attività di Ricerca e Innovazione" n. 2018-5-HH.0, collaboration agreement between the Italian Space Agency and Politecnico di Milano.

References

Antoniadis, A., Brossat, X., Cugliari, J. and Poggi, J.-M. (2016). A prediction interval for a function-valued forecast model: Application to load forecasting. *Int. J. Forecast.* **32**, 939–947.

Berg, J., Oikarinen, E., Järvisalo, M. and Puolamäki, K. (2017). Minimum-width confidence bands via constraint optimization. In *Principles and Practice of Constraint Programming* (Edited by J. C. Beck), 443–459. Springer International Publishing, Cham.

Cao, G., Yang, L. and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *J. Nonparametr. Stat.* **24**, 359–377.

Cuevas, A., Febrero, M. and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics & Data Analysis* **51**, 1063–1074.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21**, 1735–1765.

Degras, D. (2017). Simultaneous confidence bands for the mean of functional data. *Wiley Interdiscip. Rev. Comput. Stat.* **9**, e1397.

Delaigle, A. and Hall, P. (2010). Defining probability density for a distribution of random functions. *Ann. Statist.* **38**, 1171–1193.

Fontana, M., Zeni, G. and Vantini, S. (2023). Conformal prediction: A unified review of theory and new challenges. *Bernoulli* **29**, 1–23.

Gammerman, A., Vovk, V. and Vapnik, V. (1998). Learning by transduction. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* UAI'98, 148–155. Morgan Kaufmann Publishers Inc., San Francisco.

Hyndman, R. J. and Shahid Ullah, M. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Statist. Data Anal.* **51**, 4942–4956.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J. and Wasserman, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113**, 1094–1111.

Lei, J., Rinaldo, A. and Wasserman, L. (2015). A conformal prediction approach to explore functional data. *Ann. Math. Artif. Intell.* **74**, 29–43.

Lei, J., Robins, J. and Wasserman, L. (2013). Distribution-free prediction sets. *J. Amer. Statist. Assoc.* **108**, 278–287.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104**, 718–734.

Papadopoulos, H., Proedrou, K., Vovk, V. and Gammerman, A. (2002). Inductive confidence machines for regression. In *European Conference on Machine Learning*, 345–356. Springer.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika* **47**, 379–396.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* 2nd Edition. Springer, New York.

Ramsay, J. O., Wickham, H., Graves, S. and Hooker, G. (2020). *fda: Functional Data Analysis.* R package version 2.4.8.1.

Schüssler, R. and Trede, M. (2016). Constructing minimum-width confidence bands. *Economics Letters* **145**, 182–185.

Shafer, G. and Vovk, V. (2008). A tutorial on conformal prediction. *J. Mach. Learn. Res.* **9**, 371–421.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *J. Comput. Graph. Statist.* **20**, 316–334.

Tuddenham, R. D. and Snyder, M. M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publications in Child Development* **1**, 183–364.

Vapnik, V. (1992). Principles of risk minimization for learning theory. In *Advances in Neural Information Processing Systems* **4**, 831–838.

Vovk, V., Gammerman, A. and Shafer, G. (2005). *Algorithmic Learning in a Random World.* Springer Science & Business Media.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577–590.

(Received March 2022; accepted June 2023)