



MOX-Report No. 46/2021

Conformal Prediction Bands for Multivariate Functional Data

Diquigiovanni, J.; Fontana, M.; Vantini, F.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<http://mox.polimi.it>

Conformal Prediction Bands for Multivariate Functional Data

Jacopo Diquigiovanni^{1,*}, Matteo Fontana^{2,3}, Simone Vantini²

1 Department of Statistical Sciences, University of Padova, Italy

2 MOX - Department of Mathematics, Politecnico di Milano, Italy

3 now at Joint Research Centre - European Commission, Ispra (VA), Italy

* jacopo.diquigiovanni@phd.unipd.it

Abstract

Motivated by the pressing request of methods able to create prediction sets in a general regression framework for a multivariate functional response and pushed by new methodological advancements in non-parametric prediction for functional data, we propose a set of conformal predictors that produce finite-sample either valid or exact multivariate simultaneous prediction bands under the mild assumption of exchangeable regression pairs. The fact that the prediction bands can be built around any regression estimator and that can be easily found in closed form yields a very widely usable method, which is fairly straightforward to implement. In addition, we first introduce and then describe a specific conformal predictor that guarantees an asymptotic result in terms of efficiency and inducing prediction bands able to modulate their width based on the local behavior and magnitude of the functional data. The method is investigated and analyzed through a simulation study and a real-world application in the field of urban mobility.

Keywords: Functional data; Conformal Prediction; Prediction band; Exact prediction set; Distribution-free prediction set; Finite-sample prediction set

1 Introduction

Functional Data Analysis (FDA, Ramsay and Silverman 2005) is now a fairly established, but still very ebullient field of statistics whose goal is to develop theory and methods to treat datasets composed of smooth functions. Since the first seminal paper by Jim O. Ramsay (Ramsay 1982), many standard multivariate tools have been translated to the functional realm: among those Functional Principal Component Analysis (Ramsay and Silverman 2005, Chapter 10), Functional Linear Regression (Ramsay and Silverman 2005, Chapter 12) and functional boxplots (Sun and Genton 2011), just to give a very partial and non-exhaustive list.

A crucial challenge in FDA is the issue of uncertainty quantification in prediction. Intuitively, we are interested in creating prediction sets, namely subsets of the sample space including a new functional observation with a certain nominal confidence level

$1 - \alpha$. Only very recent works in FDA provide some knowledge into this theoretical (but yet full of applied repercussions) issue, all of them focusing on the univariate setting (i.e. a framework in which the functional observation consists of a single real-valued function defined over a domain). These approaches can be classified in three groups: the first one consists of works principally based on parametric bootstrapping techniques (e.g., Degras 2011, Cao et al. 2012), the second one is characterized by the application of dimensionality reduction techniques to manage the naturally infinite dimensionality (e.g., Hyndman and Shahid Ullah 2007, Antoniadis et al. 2016). These first two groups carry obvious drawbacks since they are either based on not easily provable distributional assumptions and/or on asymptotic results. In addition, the first class of approaches is computationally demanding, whereas the second one relies on the approximations induced by basis projection. The third group is based on a novel approach to forecasting in the framework of Conformal Prediction (CP) (Diquigiovanni et al. 2021). This approach is able to output either exact or valid prediction bands under minimal distributional assumptions and in an efficient way, thus bypassing the methodological shortcomings identified in the previous literature. However, this is done in the setting of univariate i.i.d. functional data. The objective of the present work is to build from that contribution by extending the method to multivariate functional data and to a regressive framework.

Formally, we will consider independent and identically distributed regression pairs $\mathbf{Z}_1, \dots, \mathbf{Z}_n \sim P$, with $\mathbf{Z}_i = (\mathbf{X}_i, \mathbf{Y}_i)$ consisting of a multivariate functional response variable \mathbf{Y}_i and a set of (not necessarily scalar) covariates $\mathbf{X}_i \forall i = 1, \dots, n$. Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{ip})$ be a multivariate random function such that its j -th component Y_{ij} ($j = 1, \dots, p$) is a random function taking values in $L^\infty(\mathcal{T}_j)$, which is the family of limited functions $y : \mathcal{T}_j \rightarrow \mathbb{R}$ with \mathcal{T}_j closed and bounded subset of \mathbb{R}^{d_j} , $d_j \in \mathbb{N}_{>0}$. For the sake of brevity, later in the discussion we will indicate the space $L^\infty(\mathcal{T}_1) \times \dots \times L^\infty(\mathcal{T}_p)$ in which \mathbf{Y}_i takes values as $\prod_{j=1}^p L^\infty(\mathcal{T}_j)$. Note that the framework considered is extremely wide since both the domain \mathcal{T}_j and the image of Y_{ij} are allowed to be very different when j varies. $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ is a set of covariates such that its element related to the j -th component X_{ij} (which is a set of covariates itself) belongs to a measurable space and can be very general: for example, X_{ij} can be the usual vector of predictors, or it can be a set of functional covariates allowing for a functional-on-functional regression model, or it can contain both scalar and functional predictors. Let $\mu^j(x_{ij}) = \mathbb{E}(Y_{ij}|X_{ij} = x_{ij})$ denote the regression function for the j -th component of the i -th observation, and consistently with this notation let us define the scalar value $[\mu^j(x_{ij})](t) = \mathbb{E}(Y_{ij}(t)|X_{ij} = x_{ij})$.

The aim of the article is to build a procedure able to output exact (or at least valid) multivariate functional prediction bands under no assumptions on P and $\mu^1(\cdot), \dots, \mu^p(\cdot)$ other than i.i.d. regression pairs. A multivariate functional prediction band is a specific kind of prediction set that can be defined, consistently with the well-known definition of univariate functional prediction band (López-Pintado and Romo 2009, Degras 2017), as

$$\left\{ \mathbf{y} = (y_1, \dots, y_p) \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : y_j(t) \in B_j(t), \quad \forall j \in 1, \dots, p, \quad \forall t \in \mathcal{T}_j, \right\}$$

with $B_j(t)$ interval $\forall j, t$. Prediction bands are so relevant in the functional set prediction framework due to their conceptual simplicity and because they can be plotted in parallel coordinates (Inselberg 1985). A detailed discussion of the topic

is provided by Diquigiovanni et al. (2021). For the sake of simplicity, later in the discussion the term *prediction band* will be used to indicate a multivariate functional prediction band, unless otherwise specified.

The terms *valid prediction set* and *exact prediction set* are instead used to indicate the coverage ensured by a prediction set.

Valid prediction set Consistently with the notation of Lei et al. (2018), a valid prediction set for $\mathbf{Z}_{n+1} = (\mathbf{X}_{n+1}, \mathbf{Y}_{n+1})$ - which is independent from and identically distributed to $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ - is the set $\mathcal{C}_{n,1-\alpha}$ based on $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ such that

$$\mathbb{P}(\mathbf{Y}_{n+1} \in \mathcal{C}_{n,1-\alpha}(\mathbf{X}_{n+1})) \geq 1 - \alpha \quad (1)$$

for any significance level $\alpha \in (0, 1)$, with $\mathcal{C}_{n,1-\alpha}(\mathbf{x}) = \{\mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : (\mathbf{x}, \mathbf{y}) \in \mathcal{C}_{n,1-\alpha}\}$

Exact prediction set An exact prediction set for $\mathbf{Z}_{n+1} = (\mathbf{X}_{n+1}, \mathbf{Y}_{n+1})$ - which is independent from and identically distributed to $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ - is the set $\mathcal{C}_{n,1-\alpha}$ based on $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ such that

$$\mathbb{P}(\mathbf{Y}_{n+1} \in \mathcal{C}_{n,1-\alpha}(\mathbf{X}_{n+1})) = 1 - \alpha \quad (2)$$

for any significance level $\alpha \in (0, 1)$ and with $\mathcal{C}_{n,1-\alpha}(x)$ defined as above.

It is important to notice that the left side of Inequality (1) and of Equality (2) refers to the unconditional coverage reached by the prediction set, i.e. the probability is taken over the i.i.d. draws $\mathbf{Z}_1, \dots, \mathbf{Z}_{n+1}$. In view of this, later in the discussion the term *coverage* will be used to indicate the unconditional coverage and the term *empirical coverage* will be used to indicate an estimate of the coverage.

The article is organized as follows: in Section 2 we introduce the CP framework; in Section 3 we present the method developed; in Section 4 we discuss three simulation studies aimed at investigating different aspects of the method; in Section 5 we apply our method to a real-world application; in Section 6 we provide an overview of the main findings and sketch directions of future research.

2 The Conformal Prediction Framework

Conformal Prediction is an innovative method to build either valid or exact prediction sets under no assumptions other than exchangeable data (Vovk et al. 2005). Moreover, the CP framework ensures that valid/exact prediction sets are obtained regardless the sample size n (i.e. not only asymptotically), a fact that allows Conformal Prediction to be used in an extremely wide range of different scenarios. In this article we consider the Semi-Off-Line Inductive Conformal framework, also known as *Split Conformal* (Papadopoulos et al. 2002), which represents a computationally and methodologically convenient alternative to the original Transductive framework. Split Conformal approach is characterized by two sub-frameworks: Non-Smoothed Split Conformal framework and Smoothed Split Conformal framework¹. The two procedures are defined below.

¹since the term ‘Split Conformal’ itself is used to indicate ‘Non-Smoothed Split Conformal’, later in the discussion we will use the following two terms to indicate the two sub-frameworks: Split Conformal, Smoothed Split Conformal

Split Conformal method Let $\mathbf{z}_1, \dots, \mathbf{z}_n$ be realizations of $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, and let $\{1, \dots, n\}$ be randomly splitted into two sets $\mathcal{I}_1, \mathcal{I}_2$ of size m and l respectively such that $n = m + l$, $m, l \in \mathbb{N}_{>0}$. Let us also define the set $\{\mathbf{z}_h : h \in \mathcal{I}_1\}$ as *training set*, the set $\{\mathbf{z}_d : d \in \mathcal{I}_2\}$ as *calibration set* and the *nonconformity measure*, which represents the key aspect of Conformal Prediction, as any measurable function $A(\{\mathbf{z}_h : h \in \mathcal{I}_1\}, \mathbf{z})$ taking values in \mathbb{R} . The Split Conformal approach defines the prediction set for \mathbf{Y}_{n+1} as $\mathcal{C}_{n,1-\alpha}(\mathbf{x}_{n+1}) := \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : \delta_{\mathbf{y}} > \alpha \right\}$, with

$$\delta_{\mathbf{y}} := \frac{|\{d \in \mathcal{I}_2 \cup \{n+1\} : R_d \geq R_{n+1}\}|}{l+1},$$

and *nonconformity scores* $R_d := A(\{\mathbf{z}_h : h \in \mathcal{I}_1\}, \mathbf{z}_d)$ for $d \in \mathcal{I}_2$, $R_{n+1} := A(\{\mathbf{z}_h : h \in \mathcal{I}_1\}, (\mathbf{x}_{n+1}, \mathbf{y}))$. Intuitively, nonconformity score R_d (R_{n+1} respectively) scores how different \mathbf{z}_d ($(\mathbf{x}_{n+1}, \mathbf{y})$ respectively) is from the training set, and so $\delta_{\mathbf{y}}$ indicates the conformity of $(\mathbf{x}_{n+1}, \mathbf{y})$ to the training set compared to the conformity of the elements of the calibration set to the same training set (i.e. it is the p-value of $(\mathbf{x}_{n+1}, \mathbf{y})$, Vovk et al. 2005).

The Split Conformal method is particularly appealing since it outputs - by construction - finite-sample, valid prediction sets by only assuming exchangeable data. In fact, Diquigiovanni et al. (2021) show that, under the mild assumption that $\{R_d : d \in \mathcal{I}_2\}$ have a continuous joint distribution (an assumption that we will make hereafter), the coverage ensured by Split Conformal prediction set is equal to an easy-to-compute fixed quantity, i.e. $\mathbb{P}(\mathbf{Y}_{n+1} \in \mathcal{C}_{n,1-\alpha}(\mathbf{X}_{n+1})) = 1 - \frac{\lfloor (l+1)\alpha \rfloor}{l+1}$, and it is not only greater than or equal to $1 - \alpha$. As a consequence, exact (and not only valid) prediction sets are automatically obtained whenever $\lfloor (l+1)\alpha \rfloor = (l+1)\alpha$.

Smoothed Split Conformal method Moving from the Split Conformal framework, let us consider a single realization of a uniform random variable in $[0, 1]$, called τ_{n+1} . The Smoothed Split Conformal approach defines the prediction set for \mathbf{Y}_{n+1} as $\mathcal{C}_{n,1-\alpha,\tau_{n+1}}(\mathbf{x}_{n+1}) := \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : \delta_{\mathbf{y},\tau_{n+1}} > \alpha \right\}$, with

$$\delta_{\mathbf{y},\tau_{n+1}} := \frac{|\{d \in \mathcal{I}_2 : R_d > R_{n+1}\}| + \tau_{n+1} |\{d \in \mathcal{I}_2 \cup \{n+1\} : R_d = R_{n+1}\}|}{l+1}.$$

By introducing the element of randomization τ_{n+1} , the Smoothed Split Conformal method outputs finite-sample, exact prediction sets by only assuming exchangeable data (Vovk et al. 2005).

In order to avoid redundancy, later in the discussion we will mainly focus on Split Conformal method, but the generalization of the main findings of this article to the Smoothed Split Conformal method is reported in Appendix A.

3 Proposed Methodology

3.1 Nonconformity Measure

Moving from Diquigiovanni et al. (2021), we propose the following nonconformity measure and nonconformity scores:

$$A^s(\{\mathbf{z}_h : h \in \mathcal{I}_1\}, \tilde{\mathbf{z}}) = \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{\tilde{y}_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(\tilde{x}_j)](t)}{s_{j,\mathcal{I}_1}(t)} \right| \right) \quad (3)$$

$$\begin{aligned}
R_d^s &= \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)}{s_{j, \mathcal{I}_1}(t)} \right| \right), \quad d \in \mathcal{I}_2 \\
R_{n+1}^s &= \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t)}{s_{j, \mathcal{I}_1}(t)} \right| \right)
\end{aligned} \tag{4}$$

with $\tilde{\mathbf{z}} = (\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_p)$, $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_p)$, y_j the j -th component of \mathbf{y} , $[\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)$ estimate of $[\mu^j(x_{dj})](t)$ based on $\{\mathbf{z}_h : h \in \mathcal{I}_1\}$, $s_{\mathcal{I}_1} = \{s_{j, \mathcal{I}_1}\}_{j=1}^p$ set of modulation functions with $s_{j, \mathcal{I}_1} : \mathcal{T}_j \rightarrow \mathbb{R}_{>0}$ a (strictly positive) function belonging to $L^\infty(\mathcal{T}_j)$ based on $\{\mathbf{z}_h : h \in \mathcal{I}_1\}$ called modulation function, and with the superscript s introduced in order to emphasize the role of $s_{\mathcal{I}_1}$. It is fundamental to notice that no specific assumptions are made on the estimators $[\hat{\mu}_{\mathcal{I}_1}^1(\cdot)](t), \dots, [\hat{\mu}_{\mathcal{I}_1}^p(\cdot)](t)$ (considered in this case as random variables instead of observed values) since the Conformal framework only requires the nonconformity scores R_d^s and R_{n+1}^s to be computed on the basis of the observations belonging to the training set and on \mathbf{z}_d and $(\mathbf{x}_{n+1}, \mathbf{y})$ respectively. As a consequence, finite-sample, either valid or exact prediction sets are obtained regardless the choice of the regression estimators, allowing Conformal Inference to be satisfactorily performed also when the underlying model is completely misspecified.

By considering the Split Conformal method and the nonconformity measure (3), if $\alpha \in (0, 1/(l+1))$ then $\mathcal{C}_{n, 1-\alpha}^s(\mathbf{x}_{n+1}) = \prod_{j=1}^p L^\infty(\mathcal{T}_j)$ since $\delta_{\mathbf{y}}^s$ is always greater or equal than $1/(l+1)$. If $\alpha \in [1/(l+1), 1)$ (representing the scenario on which we will focus on hereafter), then

$$\begin{aligned}
\mathcal{C}_{n, 1-\alpha}^s(\mathbf{x}_{n+1}) := \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : y_j(t) \in \left[[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - k^s \cdot s_{j, \mathcal{I}_1}(t), \right. \right. \\
\left. \left. [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + k^s \cdot s_{j, \mathcal{I}_1}(t) \right] \right. \\
\left. \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j \right\}, \tag{5}
\end{aligned}$$

with k^s the $\lceil (l+1)(1-\alpha) \rceil$ th smallest value in the set $\{R_d^s : d \in \mathcal{I}_2\}$. The computation needed to find analytically $\mathcal{C}_{n, 1-\alpha}^s(\mathbf{x}_{n+1})$ is provided in Appendix A.1, together with the definition of $\mathcal{C}_{n, 1-\alpha, \tau_{n+1}}^s(\mathbf{x}_{n+1})$, i.e. the Smoothed Split Conformal prediction set induced by nonconformity measure (3).

From a practical point of view, first of all the observed sample $\mathbf{z}_1, \dots, \mathbf{z}_n$ is used to compute k^s and $s_{1, \mathcal{I}_1}, \dots, s_{p, \mathcal{I}_1}$, and after that the prediction set is built around the regression estimates $[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t)$, $j \in \{1, \dots, p\}$. Despite the fact that no specific constraints on $[\hat{\mu}_{\mathcal{I}_1}^j(\cdot)](t)$ are required by the Split Conformal framework, the choice of the regression estimators is fundamental in providing small prediction sets, a key topic that will be investigated in Section 3.2: indeed, intuitively one is justified in expecting prediction sets to be smaller when improved regression estimators are chosen since they typically provide smaller nonconformity scores and so a smaller value of k^s (Lei et al. 2018). However, later in the discussion (and specifically in Section 4 and Section 5) we will always consider the regression estimators as given by the application at hand: in fact, our aim is to construct valid/exact prediction sets in general and arbitrary prediction scenarios and not only in specific, well informed frameworks.

Under the exchangeability assumption of the regression pairs and regardless the choice of $s_{\mathcal{I}_1}$ and $[\hat{\mu}_{\mathcal{I}_1}^j(\cdot)](t)$, the prediction sets induced by nonconformity measure (3)

- are either finite-sample valid (Split Conformal method) or finite-sample exact (Smoothed Split Conformal method) for any distribution P ;
- are in closed form;
- are bands;
- are scalable ².

Note that nonconformity measure (3) ensures multivariate simultaneous bands, i.e. bands guaranteeing the desired coverage globally (i.e. for the multivariate random function \mathbf{Y}_{n+1}). Proper multivariate simultaneous coverage represents a leap forward with respect to univariate simultaneous coverage (i.e. coverage holding for $Y_{n+1,j}$) and pointwise coverage (i.e. coverage holding for $Y_{n+1,j}(t)$). Conformal prediction bands for multivariate functional data (5) can be proven to be a superset of the multivariate functional bands found by concatenating the p univariate prediction bands obtained by applying the nonconformity measure $\sup_{t \in \mathcal{T}_j} |(y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_j)](t)) / s_{j,\mathcal{I}_1}(t)|$ to the p components separately (Diquigiovanni et al. 2021), and also a superset of the multivariate functional bands found by concatenating the pointwise prediction intervals obtained by applying the pointwise nonconformity measure $|(y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_j)](t)) / s_{j,\mathcal{I}_1}(t)| \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j$ (see Appendix A.1 for the proof). In other words, multivariate functional simultaneous bands (5) ensure also both univariate simultaneous and pointwise validity, while the converse is not guaranteed. The topic is further addressed by means of a simulation study in Section 4.2.

Alongside the choice to base the nonconformity measure on the supremum metric, the set $s_{\mathcal{I}_1}$ of (strictly positive) modulation functions s_{j,\mathcal{I}_1} represents the core of our approach. First of all, one can notice that prediction bands induced by $\{s_{j,\mathcal{I}_1}\}_{j=1}^p$ and by $\{\lambda \cdot s_{j,\mathcal{I}_1}\}_{j=1}^p$ coincide $\forall \lambda \in \mathbb{R}_{>0}$ (see Appendix A.1 for the proof), and so later in the discussion we will consider, for any equivalence class, the set of modulation functions such that $\sum_{j=1}^p \int_{\mathcal{T}_j} s_{j,\mathcal{I}_1}(t) dt = 1$. In the next Section, we detail the role of $s_{\mathcal{I}_1}$ by highlighting its impact on the efficiency (i.e. the size) of the prediction bands and we propose a specific set of modulation functions able to guarantee an asymptotic result in terms of efficiency.

3.2 The Choice of the Set of Modulation Functions

Intuitively, in addition to the appealing properties presented in Section 3.1, a prediction band should modulate its width over $\mathcal{T}_1, \dots, \mathcal{T}_p$ according to the local variability of the data. Specifically, the aim is to obtain prediction bands able to properly manage the fact that: focusing on the j -th component, the pointwise evaluations of functional data may be characterized by highly different variability

²Indeed, conditional on the computational cost required to calculate the regression estimates and the set of modulation functions (a set that can be chosen to be computationally parsimonious), and by keeping the ratio l/n fixed when n grows, the time required to compute k^s (and therefore to output the prediction set) increases linearly with l , and so linearly with n

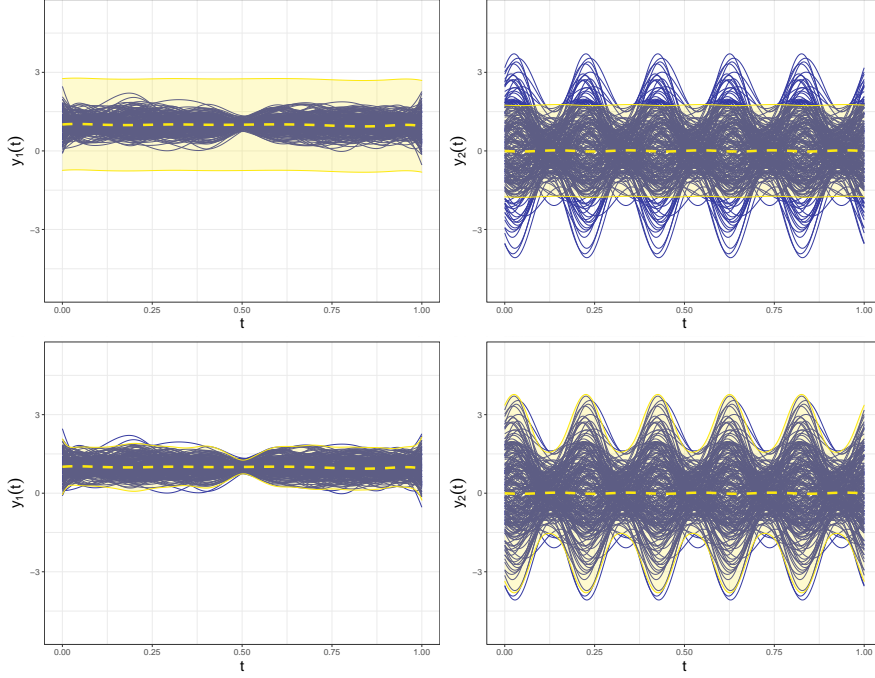


Figure 1. Split Conformal multivariate prediction band for $\mathbf{Y}_{201} = (Y_{201,1}, Y_{201,2})$ obtained by considering $\{s_j^0\}_{j=1}^2$ (at the top) and $\{s_{j, \mathcal{I}_1}^\sigma\}_{j=1}^2$ (at the bottom) as set of modulation functions. The dashed yellow lines represent the regression estimates. $\alpha = 0.25$, $n = 200$, $m = l = 100$.

when $t \in \mathcal{T}_j$ varies; the p components may be characterized by different magnitude. In order to achieve these two purposes, a careful choice of a data-driven set of modulation functions $s_{\mathcal{I}_1}$ is recommended. In order to clarify this concept, let us consider the following example: let $p = 2$ with $\mathbf{y}_1, \dots, \mathbf{y}_{200}$ independent realizations of $\mathbf{Y}_1, \dots, \mathbf{Y}_{200}$ such that $Y_{i1}(t) = \beta_1(t) + \varepsilon_{i1}(t)$ and $Y_{i2}(t) = \beta_2(t) + \varepsilon_{i2}(t)$ ($i = 1, \dots, 200$, $\mathcal{T}_1 = \mathcal{T}_2 = [0, 1]$), with the systematic components defined simply as $\beta_1(t) = 1$, $\beta_2(t) = 0 \forall t \in [0, 1]$ and the independent functional error components $\{\varepsilon_{i1}\}_{i=1}^{200}$ ($\{\varepsilon_{i2}\}_{i=1}^{200}$ respectively) obtained by means of a B-spline basis expansion (Fourier basis expansion respectively) with normally distributed random vectors as coefficients. In full generality, we consider $[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) = \hat{\beta}_j(t)$, $j = 1, 2$, with $\hat{\beta}_1(t), \hat{\beta}_2(t)$ the estimates (based on $\{\mathbf{z}_h : h \in \mathcal{I}_1\}$) obtained by fitting the two concurrent functional-on-functional linear models (Ramsay and Silverman 2005). This example represents the simplest, almost trivial regression scenario which allows to - hopefully - easily understand the crucial role of $s_{\mathcal{I}_1}$, but the discussion presented hereafter naturally holds also when decidedly more complex regression functions and regression estimators are taken into account. Figure 1 shows the multivariate prediction band for $\mathbf{Y}_{201} = (Y_{201,1}, Y_{201,2})$ obtained by considering two different sets of modulation functions: the two panels at the top of the Figure 1 show the multivariate prediction band obtained by not modulating (i.e. by setting $s_{1, \mathcal{I}_1}(t) = s_{2, \mathcal{I}_1}(t) = 1 / \sum_{j=1}^2 |\mathcal{T}_j| \propto 1 \forall t \in [0, 1]$), whereas the two panels at the bottom of the same Figure show the prediction band obtained by considering the two standard deviation functions of the functional residuals as modulation functions (after normalization in order to meet the condition $\sum_{j=1}^2 \int_{\mathcal{T}_j} s_{j, \mathcal{I}_1}(t) dt = 1$). In order to distinguish the two sets of modulation functions, later in the discussion we will denote the first

set by $s^0 := \{s_j^0\}_{j=1}^p$ (whose notation excludes the subscript \mathcal{I}_1 to remark its lack of dependence on the training set) and the second one by $s_{\mathcal{I}_1}^\sigma := \{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^p$. The prediction sets are obtained by considering the Split Conformal framework and by setting $\alpha = 0.25$, $m = l = 100$. Focusing on the two panels at the top of Figure 1, it is possible to notice that the two univariate prediction bands are far from desirable: specifically, the univariate prediction band related to $Y_{201,1}$ is large along all the domain \mathcal{T}_1 , whereas the one related to $Y_{201,2}$ contains almost all the pointwise evaluations of the functional data in the low-variance parts of \mathcal{T}_2 but excludes many pointwise evaluations in the other, high-variance parts of the domain. In this specific case, the absence of a modulation process does not allow to take into account: first of all, the different variability of the data over \mathcal{T}_1 and \mathcal{T}_2 respectively; secondly, the different magnitude that characterizes the two components. In so doing, one is justified in expecting that a procedure based on $\{s_j^0\}_{j=1}^p$, although able to output a valid prediction band, may be of limited practical use in real applications. Vice versa, the set of modulation functions $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^p$ properly adapts the width of the prediction band according to the local variability of functional data, allowing for a meaningful, interpretable and useful prediction band.

Beyond these common-sense considerations, a criterion that is both reasonable and well-established in Conformal Prediction to discriminate between procedures able to guarantee validity is the minimization of the size of the prediction sets outputted (also known, in the Conformal framework, as maximization of efficiency, Balasubramanian et al. 2014): this choice is due to the fact that desirable prediction sets should include subsets of the sample space where the probability mass is highly concentrated (Lei et al. 2013). In the context of our article, the aim would be to find the nonconformity measure $A^s(\{\mathbf{z}_h : h \in \mathcal{I}_1\}, \cdot)$ (and so, practically, the set of modulation functions $s_{\mathcal{I}_1}$) inducing the smallest prediction bands. The first, fundamental step in assessing the size of a prediction band for multivariate functional data is the definition of the concept of ‘size’, a nontrivial task if compared to the traditional univariate and multivariate statistical settings. By generalizing the definition given in Diquigiovanni et al. (2021) to the multivariate case, we define the size of a multivariate prediction band as the sum of the p areas between the upper and lower bound of the p univariate prediction bands:

$$\mathcal{Q}(s_{\mathcal{I}_1}) := \sum_{j=1}^p \int_{\mathcal{T}_j} 2 \cdot k^s \cdot s_{j,\mathcal{I}_1}(t) dt = 2 \cdot k^s. \quad (6)$$

Since $\mathcal{Q}(s_{\mathcal{I}_1})$ is a random variable depending on $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, the task of finding the set of modulation functions minimizing the risk functional $\mathbb{E}[\mathcal{Q}(s_{\mathcal{I}_1})]$ is unfeasible in the case of no assumptions on P . A simplification of such a complex task consists of considering the quantity to be minimized $k^s(\propto \mathcal{Q}(s_{\mathcal{I}_1}))$ as an observed value depending on $\mathbf{z}_1, \dots, \mathbf{z}_n$ instead of on $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ according to the empirical risk minimization principle (Vapnik 1992). In so doing, the optimization problem is certainly simplified, but its resolution still remains unfeasible due to the specific structure of k^s . Indeed, k^s is a specific empirical quantile of $\{R_d^s : d \in \mathcal{I}_2\}$, and R_d^s (see Equation (4)) depends by construction both on the training set through $\{\mathbf{z}_h : h \in \mathcal{I}_1\}$ and on the calibration set through \mathbf{z}_d . Since by construction the set of modulation functions $s_{\mathcal{I}_1}$ depends only on the training set (as its dependence on the calibration set would imply not to obtain closed-form valid prediction bands),

no rule minimizing k^s only by combining the elements of the training set (i.e. by varying $s_{\mathcal{I}_1}$) can be found for general $\mathbf{z}_1, \dots, \mathbf{z}_n$.

In view of this, we propose an alternative, unconventional strategy to build a set of modulation functions able to guarantee an asymptotic result in terms of efficiency. Specifically, the purpose is to find a couple of sets of functions $(\bar{s}_{\mathcal{I}_1}, \bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$ such that:

- $\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c := \{\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c\}_{j=1}^p$ is a set of functions such that $\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c$ meets the definition of modulation function, but depends also on the calibration set through $\{\mathbf{z}_d : d \in \mathcal{I}_2\}, \forall j \in \{1, \dots, p\}$
- prediction bands obtained by using $\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c$ as set of modulation functions are smaller than or equal to (in terms of Equation (6)) those induced by the set of modulation functions s^0 for every possible value of n and for every possible observed sample $\mathbf{z}_1, \dots, \mathbf{z}_n$
- $\bar{s}_{\mathcal{I}_1} = \{\bar{s}_{j, \mathcal{I}_1}\}_{j=1}^p$ is a set of modulation functions such that $\bar{s}_{j, \mathcal{I}_1}$ is equal to $\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c$, but in which the dependence on $\{\mathbf{z}_d : d \in \mathcal{I}_2\}$ is replaced by the dependence on $\{\mathbf{z}_h : h \in \mathcal{I}_1\} \forall j \in \{1, \dots, p\}$
- $\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c$ and $\bar{s}_{j, \mathcal{I}_1}$ converge to the same function when $m, l \rightarrow +\infty, \forall j \in \{1, \dots, p\}$

In so doing, prediction bands induced by the set of modulation functions $\bar{s}_{\mathcal{I}_1}$ are characterized by all the appealing properties presented in Section 3.1 (including validity) and are asymptotically not wider than those induced by s^0 regardless the specific sample $\mathbf{z}_1, \dots, \mathbf{z}_n$. In order to find $(\bar{s}_{\mathcal{I}_1}, \bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$ satisfying the aforementioned conditions, let us consider the structure of k^s : operationally, k^s computes a summary of the multivariate functional residual for every observation in the calibration set, and selects the $[(l+1)(1-\alpha)]$ th smallest value among them. In particular: the summary is naturally induced by the specific nonconformity measure used, which searches the greatest value of the absolute value of the modulated multivariate functional residual over the p domains $\mathcal{T}_1, \dots, \mathcal{T}_p$; k^s is not affected by the $l - [(l+1)(1-\alpha)]$ greatest values of $\{R_d^s : d \in \mathcal{I}_2\}$. In view of this, a proper candidate for $\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c$ should ignore the elements of $\{\mathbf{z}_d : d \in \mathcal{I}_2\}$ leading to the $l - [(l+1)(1-\alpha)]$ greatest values of $\{R_d^s : d \in \mathcal{I}_2\}$ and should modulate data based on the most extreme value observed $\forall t \in \mathcal{T}_j, j \in \{1, \dots, p\}$.

Therefore, the couple of sets of functions $(\bar{s}_{\mathcal{I}_1}, \bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$ we propose - which represents a generalization of the finding of Diquigiovanni et al. (2021) in the univariate case - is defined below. Formally, the set of functions $\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c$ is such that

$$\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c(t) := \frac{\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)|}{\sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt}$$

$\forall j = 1, \dots, p, t \in \mathcal{T}_j$ with

$$\mathcal{H}_2 := \left\{ d \in \mathcal{I}_2 : \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \leq k \right\}$$

and $k = k^{s^0} / \sum_{j=1}^p |\mathcal{T}_j|$ the $[(l+1)(1-\alpha)]$ th smallest value in the set

$$\left\{ \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) : d \in \mathcal{I}_2 \right\}.$$

For the sake of simplicity, we assumed $\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \neq 0 \forall j \in \{1, \dots, p\}, t \in \mathcal{T}_j$. If this condition does not hold for at least one couple (t, j) but the condition $\sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt \neq 0$ still holds (the case in which $\sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt = 0$ represents a pathological case of no practical interest), in order to have that $\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c(t) > 0 \forall j = 1, \dots, p, t \in \mathcal{T}_j$ it is sufficient to add a small, positive value to $\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c(t)$ and to normalize accordingly.

The set of modulation functions $\bar{s}_{\mathcal{I}_1}$ is such that

$$\bar{s}_{j, \mathcal{I}_1}(t) := \frac{\max_{h \in \mathcal{H}_1} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)|}{\sum_{j=1}^p \int_{\mathcal{T}_j} \max_{h \in \mathcal{H}_1} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)| dt}$$

$\forall j = 1, \dots, p, t \in \mathcal{T}_j$ with $\mathcal{H}_1 = \mathcal{I}_1$ if $\lceil (m+1)(1-\alpha) \rceil > m$, otherwise

$$\mathcal{H}_1 := \left\{ h \in \mathcal{I}_1 : \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)| \right) \leq \gamma \right\}$$

and γ the $\lceil (m+1)(1-\alpha) \rceil$ th smallest value in the set

$$\left\{ \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)| \right) : h \in \mathcal{I}_1 \right\}.$$

If $\exists(t, j)$ such that $\max_{h \in \mathcal{H}_1} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)| = 0$, the adjustment used for $\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c$ is implemented.

Specifically, the fact that the set of modulation functions $\bar{s}_{\mathcal{I}_1}$ depends on α (through γ) allows for a procedure able to modulate data according to the specific value $1 - \alpha$, i.e. the desired nominal coverage. In addition, such an unconventional set of modulation functions is particularly useful when functional residuals show a non-standard behavior (e.g. there are outliers). The following two theorems show that $(\bar{s}_{\mathcal{I}_1}, \bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$ satisfies the aforementioned conditions.

Theorem 1. Let $m/n = \theta$ with $0 < \theta < 1$ and let $\text{Var}[[\hat{\mu}_{\mathcal{I}_1}^j(X_{ij})](t)] \rightarrow 0 \forall i \in \{1, \dots, n\}, \forall t \in \mathcal{T}_j, \forall j \in \{1, \dots, p\}$ when $m \rightarrow +\infty$. Then $\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c$ and $\bar{s}_{j, \mathcal{I}_1}$ converge to the same function $\forall j \in \{1, \dots, p\}$ when $n \rightarrow +\infty$.

Theorem 2. If at least one of the functions $\{\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c(t)\}_{j=1}^p$ is not constant almost everywhere over its domain, then $\mathcal{Q}(s^0) > \mathcal{Q}(\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$. Otherwise, $\mathcal{Q}(s^0) = \mathcal{Q}(\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$.

See Appendix A.2 for both proofs, together with the generalization of $(\bar{s}_{\mathcal{I}_1}, \bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$, Theorem 1 and Theorem 2 to the Smoothed Split Conformal framework. Due to the very mild conditions required by the two theorems to hold, the set of modulation functions $\bar{s}_{\mathcal{I}_1}$ can be used in many general frameworks and provides a new, we believe appealing data-driven alternative to other solutions (e.g. $s_{\mathcal{I}_1}^\sigma$). In the next Section, the set of modulation functions $\bar{s}_{\mathcal{I}_1}$ is compared to other sets of modulation functions in different simulated scenarios.

4 Simulation Study

In this Section we perform three simulation studies aimed at evaluating different practical aspects of the method presented in Section 3. Since, to our knowledge, there

are no methods dealing with building prediction bands in a multivariate functional setting, the simulations will focus on exploring the empirical properties of our method. In Section 4.1, the empirical coverage provided by the prediction bands is evaluated in different scenarios, considering different sample sizes and different kinds of model misspecification. In Section 4.2, we compare the multivariate prediction bands obtained by the method presented in this article with those obtained by concatenating the p univariate prediction bands induced by the Conformal approach of Diquigiovanni et al. (2021). Finally, in Section 4.3, the three sets of modulation functions presented in Section 3 ($\{s_j^0\}_{j=1}^p, \{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^p, \{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^p$) are compared in terms of efficiency in order to highlight their strengths and weaknesses.

In all simulation studies, some quantities are kept fixed: $p = 2$, $\mathcal{T}_1 = \mathcal{T}_2 = [0, 1]$, $\alpha = 0.10$. Three possible sample sizes are taken into account: $n = 20, n = 200, n = 2000$. We focus on the Split Conformal method and since the coverage reached by Split Conformal prediction set is $1 - \lfloor (l+1)\alpha \rfloor / (l+1)$ (see Section 2), the size of the calibration set is set equal to $l = 9, l = 99, l = 999$ respectively in order to obtain $1 - \lfloor (l+1)\alpha \rfloor / (l+1) = 1 - \alpha$ and consequently to facilitate the readability of the results. A possible alternative would be to consider a different value of l (e.g. $n/2$) and to evaluate the empirical coverage taking into account the coverage $1 - \lfloor (l+1)\alpha \rfloor / (l+1)$. Each combination of simulation study, scenario, sample size, regression estimators and set of modulation functions is evaluated based on $N = 5000$ replications. Specifically, for each replication, a sample $\mathbf{z}_1, \dots, \mathbf{z}_{n+1}$ is generated and n randomly chosen elements are assigned to the training and calibration sets, whereas the remaining element is considered as the one we aim to predict (however, for the sake of simplicity, hereafter we will simply define the two sets as $\{\mathbf{z}_i\}_{i=1}^n$ and \mathbf{z}_{n+1}). All simulations are computed using the R Programming Language (R Core Team 2020).

4.1 Simulation Study 1: Coverage

The aim of the simulation study in this Section is to evaluate the empirical coverage (computed as the fraction of the $N = 5000$ replications in which \mathbf{y}_{n+1} belongs to $\mathcal{C}_{n,1-\alpha}(\mathbf{x}_{n+1})$) reached by the method presented in Section 3 in different scenarios and for different values of n .

Specifically, the simulation study consists of two scenarios. In the first one, the systematic component generating data is linear and, in addition to the case in which the model is correctly specified, two different kinds of model misspecification are taken into account: misspecification due to omitted relevant variable and misspecification due to inclusion of irrelevant variable (see Rao 1971). In the second scenario, a third kind of model misspecification is evaluated, i.e. functional form misspecification (see Wooldridge 1994). The two scenarios are formally defined as follows:

- *Scenario 1*

$$\begin{aligned} Y_{i1}(t) &= \beta_0(t) + \beta_1(t)w_i + \varepsilon_{i1}(t), & i \in \{1, \dots, n+1\}, t \in [0, 1] \\ Y_{i2}(t) &= \beta_0(t) + \beta_2(t)w_i^2 + \varepsilon_{i2}(t), & i \in \{1, \dots, n+1\}, t \in [0, 1] \end{aligned}$$

with $w_i = i/(n+1)$, $\beta_0(t), \beta_1(t), \beta_2(t)$ generated by means of a B-spline basis expansion of order four, with six basis functions, equally spaced knots,

	Scenario 1		
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	0.894[0.886,0.903]	0.896[0.888,0.905]	0.904[0.896,0.912]
$n = 200$	0.902[0.894,0.910]	0.894[0.885,0.903]	0.901[0.893,0.909]
$n = 2000$	0.899[0.890,0.907]	0.906[0.898,0.914]	0.902[0.894,0.911]
	Scenario 2		
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	0.907[0.899,0.915]	0.899[0.890,0.907]	0.904[0.896,0.913]
$n = 200$	0.899[0.891,0.907]	0.898[0.890,0.907]	0.901[0.893,0.909]
$n = 2000$	0.893[0.884,0.901]	0.893[0.884,0.901]	0.899[0.891,0.908]

Table 1. Simulation study 1: empirical coverage and related 95% confidence interval in brackets for each combination of scenario, sample size and set of covariates. $\alpha = 0.10$, set of modulation functions $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$.

coefficients generated independently by a standard normal random variable and $\varepsilon_{i1}(t), \varepsilon_{i2}(t)$ independent functional errors obtained by means of the same B-spline basis expansion with independent standard normal random variables as coefficients. It is important to notice that regression coefficient functions $\beta_0, \beta_1, \beta_2$ are generated only once, i.e. they do not vary between the $N = 5000$ replications.

- *Scenario 2*

$$Y_{i1}(t) = \exp(\beta_0(t) + \beta_1(t)w_i + \varepsilon_{i1}(t)), \quad i \in \{1, \dots, n+1\}, t \in [0, 1]$$

$$Y_{i2}(t) = \exp(\beta_0(t) + \beta_2(t)w_i^2 + \varepsilon_{i2}(t)), \quad i \in \{1, \dots, n+1\}, t \in [0, 1]$$

with $w_i, \beta_0(t), \beta_1(t), \beta_2(t), \varepsilon_{i1}(t), \varepsilon_{i2}(t)$ defined as in Scenario 1.

Both scenarios are evaluated considering the following three regression estimates:

- *Set of Covariates 1.* $[\hat{\mu}_{\mathcal{I}_1}^1(x_{i,1} = \{1\})](t) = [\hat{\mu}_{\mathcal{I}_1}^2(x_{i,2} = \{1\})](t) = \hat{\beta}_0(t)$
- *Set of Covariates 2.* $[\hat{\mu}_{\mathcal{I}_1}^1(x_{i,1} = \{1, w_i\})](t) = \hat{\beta}_0(t) + \hat{\beta}_1(t)w_i$ and $[\hat{\mu}_{\mathcal{I}_1}^2(x_{i,2} = \{1, w_i^2\})](t) = \hat{\beta}_0(t) + \hat{\beta}_2(t)w_i^2$
- *Set of Covariates 3.* $[\hat{\mu}_{\mathcal{I}_1}^1(x_{i,1} = \{1, w_i, w_i^2\})](t) = [\hat{\mu}_{\mathcal{I}_1}^2(x_{i,2} = \{1, w_i, w_i^2\})](t) = \hat{\beta}_0(t) + \hat{\beta}_1(t)w_i + \hat{\beta}_2(t)w_i^2$

with $\hat{\beta}_0(t), \hat{\beta}_1(t), \hat{\beta}_2(t)$ the estimates (based on $\{z_h : h \in \mathcal{I}_1\}$) obtained by fitting each time the corresponding functional-on-scalar linear model. Focusing on Scenario 1, ‘Set of Covariates 1’ represents the omitted relevant variable case, ‘Set of Covariates 2’ represents the case in which the model is correctly specified and ‘Set of Covariates 3’ represents the case in which an irrelevant variable is included, whereas Scenario 2 is characterized by functional form misspecification.

Table 1 shows the empirical coverage \hat{p} , as well as the 95% confidence interval $[\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/N}]$, obtained for each combination of scenario, sample size and set of covariates considering the set of modulation functions $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$. The results

are decidedly satisfactory, as the empirical coverages are really close to $1 - \alpha = 0.90$ and the observed confidence intervals always include the desired coverage regardless the specific combination of scenario, sample size and set of covariates considered. Specifically, the method presented in Section 3 is able to guarantee the desired coverage also when the sample size is small and the model misspecified.

4.2 Simulation Study 2: Univariate and Multivariate Prediction Bands

The simulation study of this Section is aimed at comparing the Multivariate Prediction Bands outputted by the method presented in this article (*MPB method*) to the bands obtained by Concatenating the p Univariate prediction Bands provided by the Conformal approach presented in Diquigiovanni et al. (2021) (*CUB method*, see Section 3.1 for further details). We focus on two aspects: empirical coverage and efficiency. The empirical coverage is evaluated as described in Section 4.1, whereas for each of the $N = 5000$ replications the size of the observed prediction band $\mathcal{C}_{n,1-\alpha}(\mathbf{x}_{n+1})$ is defined as the average value $\mathcal{Q}(\cdot)/2$ (see Equation (6)).

Two different scenarios are considered: in the first one, the two components $Y_{i,1}, Y_{i,2}$ share the systematic component, but they are characterized by independent error terms; in the second one, the two components both share the systematic component and the error term in the first half of the domain. In so doing, one is justified in expecting the method presented in this article not to be affected by the different specification of the error terms in terms of coverage, while the CUB method to provide different empirical coverages according to the scenario considered. The two scenarios are:

- *Scenario 1*

$$\begin{aligned} Y_{i1}(t) &= \beta_0(t) + \beta_1(t)w_i + \beta_2(t)w_i^2 + \varepsilon_{i1}(t), & i \in \{1, \dots, n+1\}, t \in [0, 1] \\ Y_{i2}(t) &= \beta_0(t) + \beta_1(t)w_i + \beta_2(t)w_i^2 + \varepsilon_{i2}(t), & i \in \{1, \dots, n+1\}, t \in [0, 1] \end{aligned}$$

with $w_i, \beta_0(t), \beta_1(t), \beta_2(t), \varepsilon_{i1}(t), \varepsilon_{i2}(t)$ defined as in Section 4.1.

- *Scenario 2*

$$\begin{aligned} Y_{i1}(t) &= \beta_0(t) + \beta_1(t)w_i + \beta_2(t)w_i^2 + \eta_{i1}(t), & i \in \{1, \dots, n+1\}, t \in [0, 1] \\ Y_{i2}(t) &= \beta_0(t) + \beta_1(t)w_i + \beta_2(t)w_i^2 + \eta_{i2}(t), & i \in \{1, \dots, n+1\}, t \in [0, 1] \end{aligned}$$

with $\eta_{i1}(t) = \varepsilon_{i1}(t)$,

$$\eta_{i2}(t) = \begin{cases} \varepsilon_{i1}(t) & t \in [0, 0.5] \\ \varepsilon_{i2}(t) & t \in (0.5, 1] \end{cases}$$

and $w_i, \beta_0(t), \beta_1(t), \beta_2(t), \varepsilon_{i1}(t), \varepsilon_{i2}(t)$ defined as in Section 4.1.

As in the previous simulation study, three regression estimates are considered:

- *Set of Covariates 1.* $[\hat{\mu}_{\mathcal{Z}_1}^1(x_{i,1} = \{1\})](t) = [\hat{\mu}_{\mathcal{Z}_1}^2(x_{i,2} = \{1\})](t) = \hat{\beta}_0(t)$

Scenario 1			
<i>MPB method</i>			
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	0.902[0.894,0.910]	0.897[0.889,0.906]	0.902[0.893,0.910]
$n = 200$	0.907[0.899,0.915]	0.909[0.901,0.917]	0.899[0.890,0.907]
$n = 2000$	0.899[0.891,0.908]	0.897[0.888,0.905]	0.904[0.896,0.913]
<i>CUB method</i>			
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	0.812[0.801,0.822]	0.811[0.800,0.822]	0.813[0.802,0.824]
$n = 200$	0.820[0.809,0.830]	0.823[0.813,0.834]	0.805[0.794,0.816]
$n = 2000$	0.808[0.797,0.819]	0.801[0.790,0.812]	0.815[0.804,0.826]
Scenario 2			
<i>MPB method</i>			
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	0.899[0.890,0.907]	0.888[0.879,0.897]	0.905[0.894,0.915]
$n = 200$	0.900[0.892,0.908]	0.895[0.887,0.904]	0.893[0.882,0.904]
$n = 2000$	0.906[0.898,0.914]	0.899[0.891,0.908]	0.906[0.895,0.916]
<i>CUB method</i>			
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	0.838[0.828,0.848]	0.829[0.819,0.840]	0.844[0.834,0.854]
$n = 200$	0.853[0.843,0.863]	0.843[0.833,0.853]	0.846[0.836,0.856]
$n = 2000$	0.857[0.848,0.867]	0.852[0.843,0.862]	0.863[0.854,0.873]

Table 2. Simulation study 2: empirical coverage and related 95% confidence interval in brackets for each combination of scenario, method, sample size and set of covariates. $\alpha = 0.10$, set of modulation functions $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$.

- *Set of Covariates 2.* $[\hat{\mu}_{\mathcal{I}_1}^1(x_{i,1} = \{1, w_i\})](t) = [\hat{\mu}_{\mathcal{I}_1}^2(x_{i,2} = \{1, w_i\})](t) = \hat{\beta}_0(t) + \hat{\beta}_1(t)w_i$
- *Set of Covariates 3.* $[\hat{\mu}_{\mathcal{I}_1}^1(x_{i,1} = \{1, w_i, w_i^2\})](t) = [\hat{\mu}_{\mathcal{I}_1}^2(x_{i,2} = \{1, w_i, w_i^2\})](t) = \hat{\beta}_0(t) + \hat{\beta}_1(t)w_i + \hat{\beta}_2(t)w_i^2$

Note that ‘Set of Covariates 3’ represents the case in which the model is correctly specified, while the other two sets of covariates represent a case of misspecification.

Table 2 shows the empirical coverage \hat{p} , together with the 95% confidence interval defined as in Section 4.1, obtained for each combination of scenario, method, sample size and set of covariates considering the set of modulation functions $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$. In accordance with the results provided by the first simulation study, the MPB method presented in this article ensures empirical coverages very close to $1 - \alpha$, with only two confidence intervals not including the target value $1 - \alpha$. As regards the CUB method, in both scenarios the empirical coverages are far from $1 - \alpha$ as expected, and moving from Scenario 1 to Scenario 2 they grow due to the fact that the error terms are dependent. In particular, in the first scenario almost all the confidence intervals include the value $(1 - \alpha)^2 = 0.81$, that is the coverage we expect from the CUB method since the error terms are independent. In view of this, the simulation study fully confirms the quite obvious conjecture that a carefully chosen multivariate approach must be considered in order to obtain proper multivariate simultaneous bands.

	Scenario 1		
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	5.606[4.940,6.539]	5.373[4.704,6.322]	6.007[5.106,7.212]
$n = 200$	4.270[4.137,4.407]	3.820[3.706,3.943]	3.831[3.714,3.954]
$n = 2000$	4.166[4.126,4.206]	3.707[3.673,3.741]	3.702[3.666,3.737]
	Scenario 2		
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	5.279[4.623,6.186]	5.063[4.357,6.005]	5.637[4.788,6.814]
$n = 200$	4.133[3.996,4.279]	3.690[3.571,3.815]	3.698[3.575,3.823]
$n = 2000$	4.032[3.992,4.073]	3.583 [3.547,3.618]	3.578[3.543,3.614]
	Scenario 3		
	Set of Cov. 1	Set of Cov. 2	Set of Cov. 3
$n = 20$	4.820[4.184,5.649]	4.568[3.932,5.446]	5.052[4.273,6.141]
$n = 200$	3.852[3.708,4.000]	3.442[3.319,3.565]	3.450[3.327,3.578]
$n = 2000$	3.773[3.731,3.817]	3.354[3.315,3.391]	3.347[3.310,3.385]

Table 3. Simulation study 2: median size (first and third quartile in brackets) for each combination of scenario, sample size and set of covariates. MPB method, $\alpha = 0.10$, set of modulation functions $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$.

Since the CUB method does not guarantee the desired coverage, hereafter we will only focus on the efficiency of the prediction bands outputted by the MPB method. Table 3 shows the median size (together with the first and third quartile in brackets) of the prediction bands analyzed in Table 2. In addition to the two scenarios considered so far, Table 3 also analyzes a third scenario in which $\varepsilon_{i1}(t) = \varepsilon_{i2}(t) \forall i, t$ (and so $Y_{i1} = Y_{i2}$): despite its limited practical utility, this scenario represents an edge case that can provide useful information. First of all, for each combination of scenario and set of covariates, the size decreases when n grows, both because improved regression estimates generally provide smaller nonconformity scores and because the value of k^s is less dependent on random fluctuations. Focusing now on each combination of sample size and set of covariates, it is possible to notice that Scenario 1 typically provides the biggest prediction bands, whereas Scenario 3 the smallest. From a practical point of view, this evidence is due to the nonconformity measure used: indeed, it searches the most extreme value of $|(y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t))/s_{j,\mathcal{I}_1}(t)|$ over $\mathcal{T}_1, \mathcal{T}_2$, and so the nonconformity score computed when $y_{d1} = y_{d2}$ as in Scenario 3 will always be less than or equal to that computed when $y_{d1} \neq y_{d2}$. Scenario 2 represents an intermediate case between Scenario 1 and Scenario 3 as regards the structure of the error terms, and this is confirmed by the evidence provided by Table 3.

4.3 Simulation Study 3: Efficiency

The aim of the simulation study of this Section is to compare the three sets of modulation functions ($\{s_j^0\}_{j=1}^p, \{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^p, \{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^p$) in terms of efficiency. To do that, three different scenarios are taken into account: focusing just for now on the error terms and ignoring the systematic components, in the first scenario the error terms are characterized by a constant variability over the domains, in the second scenario the variability differs whereas in the third scenario the presence of outliers further complicates their specification. Formally, the three scenarios are:

- *Scenario 1.* The two systematic components are defined as in the first scenario of Section 4.1, while the independent functional errors $\varepsilon_{i1}(t), \varepsilon_{i2}(t)$ are defined as follows:

$$\begin{aligned}\varepsilon_{ij}(t) = & B_{i+(n+1)(j-1),1} + \\ & B_{i+(n+1)(j-1),2} \cos(10\pi(t + U_{i+(n+1)(j-1)})) + \\ & B_{i+(n+1)(j-1),3} \sin(10\pi(t + U_{i+(n+1)(j-1)}))\end{aligned}$$

$\forall i \in \{1, \dots, n+1\}, j \in \{1, 2\}, t \in [0, 1]$, with i.i.d. random vectors $\mathbf{B}_1, \dots, \mathbf{B}_{2(n+1)} \sim N_3(\mathbf{0}, \Sigma)$, Σ having the entries on the main diagonal equal to 1 and the entries outside the main diagonal equal to 0.7, i.i.d. random variables $U_1, \dots, U_{2(n+1)} \sim U[-0.5, 0.5]$.

- *Scenario 2.* The two systematic components are defined as in the first scenario of Section 4.1, while the independent functional errors $\varepsilon_{i1}(t), \varepsilon_{i2}(t)$ are obtained by means of a B-spline basis expansion of order four, with 13 basis functions, equally spaced knots and normal random vectors as vectors of coefficients. Specifically, the $2 \cdot (n+1)$ (observed) vectors of coefficients are independent realizations of $\mathbf{C} = (C_1, \dots, C_{13}) \sim N_{13}(\mathbf{0}, \Sigma)$ with Σ diagonal matrix such that $\text{Var}[C_a] = 0.001 \forall a \neq 7, \text{Var}[C_7] = 9 \cdot 10^{-6}$.
- *Scenario 3*

$$\begin{aligned}Y_{i1}(t) &= \beta_0(t) + \eta_{i1}(t), \quad i \in \{1, \dots, n+1\}, t \in [0, 1] \\ Y_{i2}(t) &= \beta_0(t) + \eta_{i2}(t), \quad i \in \{1, \dots, n+1\}, t \in [0, 1]\end{aligned}$$

with $\beta_0(t) = 0 \forall t \in [0, 1]$,

$$\eta_{ij}(t) = \beta_1(t)w_{ij} + \varepsilon_{ij}(t), \quad i \in \{1, \dots, n+1\}, j \in \{1, 2\}, t \in [0, 1]$$

with $\beta_1(t)$ obtained by means of a B-spline basis expansion of order four, with 13 basis, equally spaced knots and all coefficients equal to 0 but the seventh equal to 0.5, $\varepsilon_{ij}(t)$ defined as in Scenario 2, and if $n = 20$ then $w_{ij} = 0 \forall \{i, j\} \neq \{1, 1\}, w_{1,1} = 1$, whereas if $n \in \{200, 2000\}$ then

$$w_{ij} = \begin{cases} 1 & \text{if } i \in \{j + 40 \cdot \zeta : \zeta \in \{0, 1, 2, \dots, \frac{n}{40} - 1\}\} \\ 0 & \text{otherwise} \end{cases}$$

Despite the complex notation, the introduction of w_{ij} is aimed at obtaining that $\sim 5\%$ of the multivariate functions $\mathbf{y}_1, \dots, \mathbf{y}_{n+1}$ (i.e. 1 out of 21 when $n = 20$, 10 out of 201 when $n = 200$, 100 out of 2001 when $n = 2000$) is characterized, in one of the two components, by the anomalous behavior induced by $\beta_1(t)$. We propose such an unconventional structure for the error terms to simulate, for example, a regression framework in which relevant variables are not available.

All three scenarios are evaluated considering only one set of covariates each, namely the case in which the corresponding model is correctly specified. Figure 2 shows,

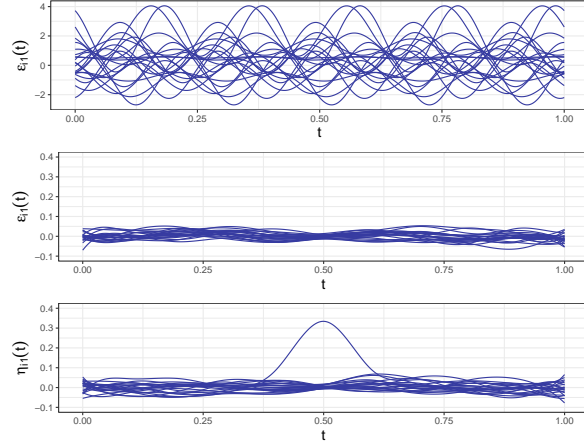


Figure 2. Example of realization of the error term related to $\{Y_{i1}\}_{i=1}^{n+1}$. First scenario at the top, second scenario in the middle, third scenario at the bottom. $n = 20$.

for each scenario, a realization of the error terms $\{\varepsilon_{i1}\}_{i=1}^{n+1}$ ($\{\eta_{i1}\}_{i=1}^{n+1}$ for Scenario 3) when $n = 20$.

Table 4 shows the median size (defined as in Section 4.2; first and third quartile in brackets) of the $N = 5000$ prediction bands obtained for each combination of scenario, sample size and set of modulation functions. All three scenarios share the evidence that the prediction bands induced by $\{s_j^0\}_{j=1}^2$ are typically smaller than those induced by the other two sets of modulation functions when the sample size is very small ($n = 20$). This is due to the fact that regression estimates obtained with a small training set size likely provide an unreliable (and potentially misleading) set of modulation functions, leading to a preference for a set of modulation functions not depending on \mathcal{I}_1 . As proof of that, it is not surprising that the two data-driven sets of modulation functions $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$, $\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$ deliver the worst performance in the most complex Scenario, i.e. Scenario 3. Focusing on the other two sample sizes, in Scenario 1 the choice of not modulating seems appropriate due to the equal magnitude of the two components and the constant variability over \mathcal{T}_1 , \mathcal{T}_2 , but, as expected, the difference between the three alternative sets of modulation functions decreases when n grows. Differently from Scenario 1, Scenario 2 is characterized by multivariate residuals showing a lower variability in the central portion of \mathcal{T}_1 and \mathcal{T}_2 : as a consequence, $\{s_j^0\}_{j=1}^2$ provides large prediction bands since it is not able to adapt the width of the band according to the local variability of the residuals, whereas $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$ is particularly effective since it induces a modulation process based on the two standard deviation functions. Finally, $\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$ represents the best solution in Scenario 3 given its ability to focus on the ‘least extreme’ $\sim (1 - \alpha) \cdot 100\%$ of data: indeed, differently from $\{s_j^0\}_{j=1}^2$ it is able to reduce the width of the band in the central part of the domains, and differently from $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$ it does not uselessly enlarge the band in the same subinterval of \mathcal{T}_1 , \mathcal{T}_2 . Consequently, the simulation study seems to confirm the statistical intuition given in Section 3.2 that the newly launched set of modulation functions $\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^p$ represents an interesting solution when functional residuals show a non-standard behavior and a modulation process driven by the value $1 - \alpha$ is needed.

	Scenario 1		
	$\{s_j^0\}_{j=1}^2$	$\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$	$\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$
$n = 20$	9.599[8.348,11.116]	12.205[10.121,14.853]	14.241[11.730,17.798]
$n = 200$	8.658[8.289,9.077]	8.835[8.442,9.246]	9.315[8.892,9.784]
$n = 2000$	8.568[8.449,8.692]	8.587[8.469,8.712]	8.681[8.561,8.806]
	Scenario 2		
	$\{s_j^0\}_{j=1}^2$	$\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$	$\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$
$n = 20$	0.168[0.152,0.188]	0.190[0.167,0.221]	0.213[0.186,0.249]
$n = 200$	0.148[0.144,0.153]	0.126[0.123,0.130]	0.139[0.135,0.144]
$n = 2000$	0.146[0.145,0.148]	0.122[0.121,0.123]	0.134[0.133,0.136]
	Scenario 3		
	$\{s_j^0\}_{j=1}^2$	$\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$	$\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$
$n = 20$	0.201[0.157,0.667]	0.294[0.212,1.767]	0.407[0.277,1.869]
$n = 200$	0.162[0.155,0.170]	0.167[0.161,0.172]	0.151[0.145,0.158]
$n = 2000$	0.160[0.157,0.162]	0.161[0.160,0.163]	0.145[0.143,0.147]

Table 4. Simulation Study 3: median size (first and third quartile in brackets) for each combination of scenario, sample size and set of modulation functions. $\alpha = 0.10$.

5 Case Study: Analysis of Bike Mobility in the City of Milan

In order to illustrate the application potential of the method presented in this article, in this section we focus on a case study concerning urban mobility, and specifically the usage of a bike-sharing system in the Italian city of Milan. Moving from the raw data and the context presented in Torti et al. (2021), the aim is to study the behavior of subscribers of Bikemi, a bike sharing system active in the city in which bikes are picked up and dropped off in specific docking stations located through the city. Starting from raw data providing various information about picked up bikes (simply *pickups* hereafter) and dropped off bikes (simply *dropoffs* hereafter) for each day considered, and focusing our attention - as an example - on the Duomo district only (i.e. the area in which Milan's cathedral is), the multivariate functional response variable $\mathbf{y}_i = (y_{i1}, y_{i2})$ representing the rate of dropoffs (y_{i1}) and pickups (y_{i2}) is obtained via a standard kernel density estimation smoothing method (Hastie et al. 2009). In so doing, $y_{i1}(t)$ ($y_{i2}(t)$) represents the dropoff (pickup) rate at time t , with t ranging from 7 a.m. day i to 1 a.m. the next day (consequently, we assume that day i ends at 1 a.m. the next day). The period considered starts on 25 January 2016 and ends on 6 March 2016: due to an error in the data collection, 25 February is removed from the dataset in accordance with Torti et al. (2021), and so the sample size is $n = 41$. Data are shown in the two top panels of Figure 3.

Like in Torti et al. (2021), the regression estimates are obtained by fitting a concurrent functional-on-functional linear model (Ramsay and Silverman 2005). The model hereby used includes as covariates a functional intercept, the temperature function (after subtracting the average daily temperature function of the period considered) in degrees Celsius, and a dummy variable indicating whether day i is a weekday or not. Since the rates cannot be negative in any subinterval of the domain, the predicted functions are truncated to 0. However, as discussed in Section 3.1, the

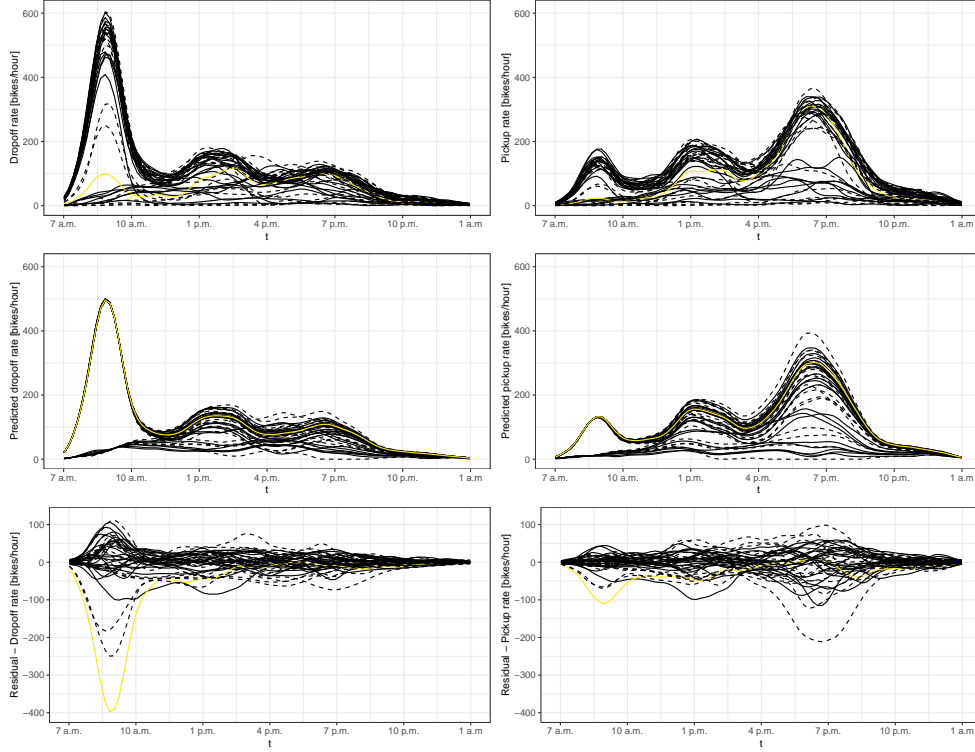


Figure 3. Dropoff and pickup rates (top left, top right respectively), corresponding functional predictions (center left, center right) and functional residuals (bottom left, bottom right). Yellow curves refer to 29 February; continuous curves refer to the observations in the training set, dashed curves to those in the calibration set.

purpose is to construct valid, meaningful and interpretable prediction bands also when simple regression estimators are specified, and so the choice of the covariates, as well as the functional form of the model, represents an aspect of limited interest in the framework considered.

The method presented in Section 3 is performed by considering the three sets of modulation functions $\{s_j^0\}_{j=1}^2$, $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$, $\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$, $\alpha = 0.25$ and $m = 22$, $l = 19$ in order to assign, as in the simulation studies, about half of the observations to the training set and to obtain the value $1 - \lfloor (l+1)\alpha \rfloor / (l+1)$ equal to $1 - \alpha$. To remain as neutral as possible, we will consider the case in which - after having labeled the days considered with numbers from 1 to 41 - the observations referring to an odd day are assigned to the training set and those referring to an even day to the calibration set, with the observation related to day 20 assigned to the training set to satisfy $m = 22$. Two possible prediction scenarios are taken into account for the scope of visualization: in the first, we construct the multivariate prediction band for a weekday having the average temperature function of the period as temperature function; in the second, we construct it for a warmer than usual weekday (see Figure B.1 in Appendix B for a graphical representation of the two functional covariates, together with those observed). Figure 4 shows, for each of the three sets of modulation functions ($\{s_j^0\}_{j=1}^2$ in the first row, $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$ in the second row, $\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$ in the third row), the prediction sets induced by the two scenarios (first set of covariates in the first column, second in the second column). In particular, each panel shows the prediction band for the dropoff rate (light blue band) and the pickup rate (red band), with

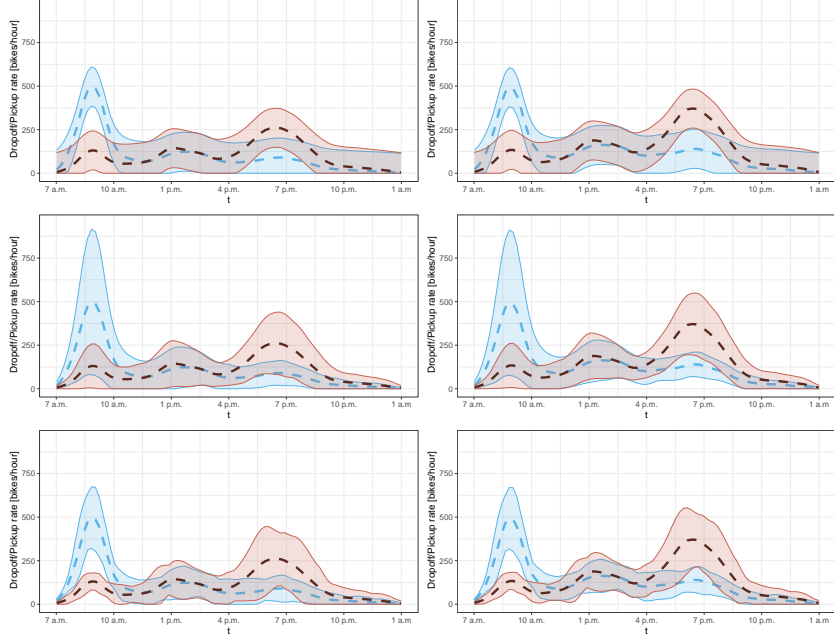


Figure 4. Prediction bands for the dropoff rate (light blue band) and the pickup rate (red band). Each panel refers to a combination of set of modulation functions ($\{s_j^0\}_{j=1}^2$ at the top, $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$ in the middle, $\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$ at the bottom) and scenario (first set of covariates on the left, second on the right). The dashed lines indicate the corresponding regression estimates. $\alpha = 0.25$. Split into calibration/training set: even/odd(+ day 20) days.

the two dashed lines representing the corresponding regression estimates. As for the predicted functions, the prediction bands are truncated to 0, as the rates cannot be negative in any subinterval of the domain. Note that this truncation does not involve any kind of drawback since the coverage reached by the prediction sets remains unchanged if a null probability portion of the bands is removed from the prediction bands. It is evident that the prediction bands for dropoffs induced by $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$ are quite large in the initial portion of the domain compared to those obtained by not modulating (i.e. $\{s_j^0\}_{j=1}^2$) and by the proposed set $\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$. In order to clarify this aspect, let us consider Figure 3. Focusing on the residual functions of the dropoff rates (i.e. the panel at the bottom left of the figure), it is easily noticeable that the yellow curve (referring to weekday 35, i.e. 29 February, which is assigned to the training set) shows an anomalous behavior in the initial part of the domain. The panel at the top left of the same figure suggests that this is due to the fact that day 35 was characterized by an unusually low dropoff rate compared to that observed in the other weekdays (which are the curves showing a pick around 9 a.m.). Consequently, by using $\{s_{j,\mathcal{I}_1}^\sigma\}_{j=1}^2$ it is natural to obtain prediction bands for dropoffs extremely wide in the first portion of the domain since this outlier has a huge impact on the modulation process, while the corresponding prediction bands obtained by not modulating are not adversely affected as $\{s_j^0\}_{j=1}^2$ does not modulate the width of the band according to the local variability of the residuals. In view of this, the set of modulation functions $\{\bar{s}_{j,\mathcal{I}_1}\}_{j=1}^2$ represents an intriguing solution since, in addition to modulate the width of the band along the domains, induces a modulation process which is not misled by the anomalous behavior of day 35. However, similar

considerations would have been made also if other observations than the one related to day 35 had been assigned to the training set, as can be noticed by analyzing the functional residuals of the observations assigned to the calibration set in the two panels at the bottom of Figure 3 (dashed curves for the calibration set; continuous curves for the training set). Despite the small sample size, the prediction sets of Figure 4 can provide profitable information: first of all, subscribers of Bikemi seem to mainly use bikes to go to Duomo in the morning, whereas in the early evening the bike flow is reversed. Moving from the first set of covariates (weekday-temperature equal to the mean temperature of the period) to the second one (weekday-warm day), we notice that a higher temperature does not strongly affect people’s behavior in the morning, whereas it involves a moderate increase in dropoffs and, at the same time, a big increase in pickups in the period of time around 7 p.m.. The information provided by the prediction bands can be indeed very useful to fleet managers in identifying the periods of time in which the imbalance between pickups and dropoffs could become critical based on the day of the week, the temperature function and other possible carefully chosen covariates.

6 Conclusion and Further Developments

In the present work we have developed a procedure aimed at creating prediction bands for multivariate functional data in a regression framework. Despite the paramount importance of this topic both from the methodological and applied point of view, to the best of our knowledge our method represents the first proposal in this direction. Moving from the approach proposed by Diquigiovanni et al. (2021) for univariate i.i.d. functional data, the method presented in this article builds finite-sample either exact or valid prediction bands under the only assumption of exchangeable regression pairs with multivariate functional response. These properties, together with the fact that the procedure is scalable and the bands can be easily found in closed form, allow to obtain meaningful prediction bands regardless the regression estimator used, leading to a methodology which can be applied in a wide range of application scenarios. Moreover, we have introduced a specific set of modulation functions (namely $\{\bar{s}_{j,\mathcal{X}_1}\}_{j=1}^p$) achieving an asymptotic result in terms of efficiency regardless the sample observed $\mathbf{z}_1, \dots, \mathbf{z}_n$ and inducing prediction bands whose width varies along the domains and across the components according to the local behavior. The simulation study and the real-world application provided in Section 4 and 5 respectively confirm the potential of the approach. Nevertheless, many possible directions still remain unexplored. Among these, we plan to modify the methodology in order to apply it when regression data are dependent (as in the case, for example, of a functional time series); and, we plan to explore the impact of the regression estimator on the size of the prediction sets.

Supplementary material

A Technical Proofs

A.1 Proof of Section 3.1

Computation to find $C_{n,1-\alpha}^s(\mathbf{x}_{n+1})$

Since

$$\delta_{\mathbf{y}}^s = \frac{|\{d \in \mathcal{I}_2 \cup \{n+1\} : R_d^s \geq R_{n+1}^s\}|}{l+1},$$

$$\mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1}) = \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : \delta_{\mathbf{y}}^s > \alpha \right\},$$

if $\alpha \in [1/(l+1), 1)$, then $\mathbf{y} \in \mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1}) \iff R_{n+1}^s \leq k^s$, with k^s the $[(l+1)(1-\alpha)]$ th smallest value in the set $\{R_d^s : d \in \mathcal{I}_2\}$. Then

$$\sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t)}{s_{j,\mathcal{I}_1}(t)} \right| \right) \leq k^s$$

$$\iff \left| \frac{y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t)}{s_{j,\mathcal{I}_1}(t)} \right| \leq k^s \quad \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j$$

$$\iff y_j(t) \in [[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - k^s \cdot s_{j,\mathcal{I}_1}(t),$$

$$[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + k^s \cdot s_{j,\mathcal{I}_1}(t)] \quad \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j.$$

As a consequence, the Split Conformal prediction set is

$$\mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1}) := \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : y_j(t) \in [[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - k^s \cdot s_{j,\mathcal{I}_1}(t),$$

$$[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + k^s \cdot s_{j,\mathcal{I}_1}(t)]$$

$$\forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j \right\}.$$

Computation to find $\mathcal{C}_{n,1-\alpha,\tau_{n+1}}^s(\mathbf{x}_{n+1})$

Consistently with the Split Conformal scenario, let us define

$$\delta_{\mathbf{y},\tau_{n+1}}^s := \frac{|\{d \in \mathcal{I}_2 : R_d^s > R_{n+1}^s\}| + \tau_{n+1} |\{d \in \mathcal{I}_2 \cup \{n+1\} : R_d^s = R_{n+1}^s\}|}{l+1}$$

$$\mathcal{C}_{n,1-\alpha,\tau_{n+1}}^s(\mathbf{x}_{n+1}) := \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : \delta_{\mathbf{y},\tau_{n+1}}^s > \alpha \right\}.$$

By definition, $\mathcal{C}_{n,1-\alpha,1}^s(\mathbf{x}_{n+1}) = \mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$.

Since $\delta_{\mathbf{y},\tau_{n+1}}^s \in [\tau_{n+1}/(l+1), (l+\tau_{n+1})/(l+1)]$, we will focus on the scenario in which $\alpha \in [\tau_{n+1}/(l+1), (l+\tau_{n+1})/(l+1)]$. Let us define w^s the $[l+\tau_{n+1} - (l+1)\alpha]$ th smallest value in the set $\{R_d : d \in \mathcal{I}_2\}$, and r_n^s (v_n^s respectively) the number of elements in the set $\{R_d : d \in \mathcal{I}_2\}$ that are equal to w^s and that are to the right (left respectively) of w^s in the sorted version of the set. Note that $r_n^s = v_n^s = 0$ when the assumption about the continuous joint distribution of $\{R_d : d \in \mathcal{I}_2\}$ is satisfied, but generally speaking we will consider $r_n^s, v_n^s \in \mathcal{N}_{\geq 0}$ such that $r_n^s + v_n^s \leq l-1$. By replicating calculations similar to those performed in the Split Conformal framework, we obtain that:

- if

$$\tau_{n+1} > \frac{(l+1)\alpha - \lfloor (l+1)\alpha - \tau_{n+1} \rfloor + r_n^s}{r_n^s + v_n^s + 2}$$

then $\mathbf{y} \in \mathcal{C}_{n,1-\alpha,\tau_{n+1}}^s(\mathbf{x}_{n+1}) \iff R_{n+1}^s \leq w^s$ and so

$$\mathcal{C}_{n,1-\alpha,\tau_{n+1}}^s(\mathbf{x}_{n+1}) := \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : y_j(t) \in \left[[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - w^s \cdot s_{j,\mathcal{I}_1}(t), \right. \right. \\ \left. \left. [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + w^s \cdot s_{j,\mathcal{I}_1}(t) \right] \right. \\ \left. \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j \right\}.$$

• if

$$\tau_{n+1} \leq \frac{(l+1)\alpha - \lfloor (l+1)\alpha - \tau_{n+1} \rfloor + r_n^s}{r_n^s + v_n^s + 2}$$

then $\mathbf{y} \in \mathcal{C}_{n,1-\alpha,\tau_{n+1}}^s(\mathbf{x}_{n+1}) \iff R_{n+1} < w^s$ and so

$$\mathcal{C}_{n,1-\alpha,\tau_{n+1}}^s(\mathbf{x}_{n+1}) := \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : y_j(t) \in \left([\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - w^s \cdot s_{j,\mathcal{I}_1}(t), \right. \right. \\ \left. \left. [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + w^s \cdot s_{j,\mathcal{I}_1}(t) \right) \right. \\ \left. \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j \right\}.$$

Proof that the concatenation of the p univariate prediction bands obtained by applying the nonconformity measure $\sup_{t \in \mathcal{T}_j} |(y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_j)](t)) / s_{j,\mathcal{I}_1}(t)|$ to the p components separately is a subset of (5)

Let us define $\mathcal{U}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$ as the multivariate prediction band obtained by concatenating the p univariate prediction bands induced by applying the nonconformity measure $\sup_{t \in \mathcal{T}_j} |(y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_j)](t)) / s_{j,\mathcal{I}_1}(t)|$ to the p components separately. Let us define, with a slight abuse of notation,

$$\tilde{R}_{dj}^s := \sup_{t \in \mathcal{T}_j} |(y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)) / s_{j,\mathcal{I}_1}(t)|, \quad \forall d \in \mathcal{I}_2, \forall j \in \{1, \dots, p\}$$

and \tilde{k}_j^s the $\lceil (l+1)(1-\alpha) \rceil$ th smallest value in the set $\{\tilde{R}_{dj}^s : d \in \mathcal{I}_2\}$. By construction $R_d^s = \sup_{j \in \{1, \dots, p\}} \tilde{R}_{dj}^s$, and so $R_d^s \geq \tilde{R}_{dj}^s \forall j \in \{1, \dots, p\}, d \in \mathcal{I}_2$ and then $k^s \geq \tilde{k}_j^s \forall j \in \{1, \dots, p\}$. In view of this, if $\mathbf{y} \in \mathcal{U}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$, i.e.

$$y_j(t) \in \left[[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - \tilde{k}_j^s \cdot s_{j,\mathcal{I}_1}(t), \right. \\ \left. [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + \tilde{k}_j^s \cdot s_{j,\mathcal{I}_1}(t) \right] \quad \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j,$$

then

$$y_j(t) \in \left[[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - k^s \cdot s_{j,\mathcal{I}_1}(t), \right. \\ \left. [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + k^s \cdot s_{j,\mathcal{I}_1}(t) \right] \quad \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j,$$

i.e. $\mathbf{y} \in \mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$. As $\mathbf{y} \in \mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$ does not necessarily imply $\mathbf{y} \in \mathcal{U}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$, then $\mathcal{U}_{n,1-\alpha}^s \subseteq \mathcal{C}_{n,1-\alpha}^s$.

Proof that the concatenation of the pointwise prediction intervals obtained by applying the pointwise nonconformity measure $| (y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_j)](t)) / s_{j,\mathcal{I}_1}(t) |$ $\forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j$ is a subset of (5)

Let us define $\mathcal{U}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$ as the multivariate prediction band obtained by concatenating the pointwise prediction intervals obtained by applying the pointwise nonconformity measure $|(y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_j)](t)) / s_{j,\mathcal{I}_1}(t)| \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j$. Let us define, with a slight abuse of notation,

$$\tilde{R}_{dj}^s(t) := |(y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)) / s_{j,\mathcal{I}_1}(t)|, \quad \forall d \in \mathcal{I}_2, \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j$$

and $\tilde{k}_j^s(t)$ the $[(l+1)(1-\alpha)]$ th smallest value in the set $\{\tilde{R}_{dj}^s(t) : d \in \mathcal{I}_2\}$. By construction $R_d^s = \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \tilde{R}_{dj}^s(t) \right)$, and so $R_d^s \geq \tilde{R}_{dj}^s(t) \forall j \in \{1, \dots, p\}, d \in \mathcal{I}_2, t \in \mathcal{T}_j$ and then $k^s \geq \tilde{k}_j^s(t) \forall j \in \{1, \dots, p\}, t \in \mathcal{T}_j$. In view of this, if $\mathbf{y} \in \mathcal{U}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$, i.e.

$$\begin{aligned} y_j(t) \in & \left[[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - \tilde{k}_j^s(t) \cdot s_{j,\mathcal{I}_1}(t), \right. \\ & \left. [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + \tilde{k}_j^s(t) \cdot s_{j,\mathcal{I}_1}(t) \right] \quad \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j, \end{aligned}$$

then

$$\begin{aligned} y_j(t) \in & \left[[\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) - k^s \cdot s_{j,\mathcal{I}_1}(t), \right. \\ & \left. [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t) + k^s \cdot s_{j,\mathcal{I}_1}(t) \right] \quad \forall j \in \{1, \dots, p\}, \forall t \in \mathcal{T}_j, \end{aligned}$$

i.e. $\mathbf{y} \in \mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$. As $\mathbf{y} \in \mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$ does not necessarily imply $\mathbf{y} \in \mathcal{U}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$, then $\mathcal{U}_{n,1-\alpha}^s \subseteq \mathcal{C}_{n,1-\alpha}^s$.

Proof that prediction bands induced by $\{s_{j,\mathcal{I}_1}\}_{j=1}^p$ and by $\{\lambda \cdot s_{j,\mathcal{I}_1}\}_{j=1}^p$ coincide $\forall \lambda \in \mathbb{R}_{>0}$

Let $\mathcal{C}_{n,1-\alpha}^{\lambda \cdot s}(\mathbf{x}_{n+1})$ be the prediction band induced by the set of modulation functions $\{\lambda \cdot s_{j,\mathcal{I}_1}\}_{j=1}^p$. The nonconformity scores are:

$$\begin{aligned} R_d^{\lambda \cdot s} &= \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)}{\lambda \cdot s_{j,\mathcal{I}_1}(t)} \right| \right) = \frac{1}{\lambda} R_d^s, \quad d \in \mathcal{I}_2 \\ R_{n+1}^{\lambda \cdot s} &= \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{y_j(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{n+1,j})](t)}{\lambda \cdot s_{j,\mathcal{I}_1}(t)} \right| \right) = \frac{1}{\lambda} R_{n+1}^s. \end{aligned}$$

Moreover, let us define:

$$\delta_{\mathbf{y}}^{\lambda \cdot s} := \frac{|\{d \in \mathcal{I}_2 \cup \{n+1\} : R_d^{\lambda \cdot s} \geq R_{n+1}^{\lambda \cdot s}\}|}{l+1},$$

with, as usual, $\mathcal{C}_{n,1-\alpha}^{\lambda \cdot s}(\mathbf{x}_{n+1}) := \left\{ \mathbf{y} \in \prod_{j=1}^p L^\infty(\mathcal{T}_j) : \delta_{\mathbf{y}}^{\lambda \cdot s} > \alpha \right\}$. As a consequence, $\mathbf{y} \in \mathcal{C}_{n,1-\alpha}^{\lambda \cdot s}(\mathbf{x}_{n+1}) \iff R_{n+1}^{\lambda \cdot s} \leq k^{\lambda \cdot s}$, with $k^{\lambda \cdot s}$ the $[(l+1)(1-\alpha)]$ th smallest value in the set $\{R_d^{\lambda \cdot s} : d \in \mathcal{I}_2\}$. Since $R_d^{\lambda \cdot s} = R_d^s / \lambda \forall d \in \mathcal{I}_2$, then $k^{\lambda \cdot s} = k^s / \lambda$. Then:

$$\begin{aligned} R_{n+1}^{\lambda \cdot s} &\leq k^{\lambda \cdot s} \\ \iff \frac{1}{\lambda} R_{n+1}^s &\leq \frac{k^s}{\lambda} \\ \iff R_{n+1}^s &\leq k^s, \end{aligned}$$

and since $\mathbf{y} \in \mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1}) \iff R_{n+1}^s \leq k^s$, then $\mathcal{C}_{n,1-\alpha}^{\lambda \cdot s}(\mathbf{x}_{n+1})$ coincides with $\mathcal{C}_{n,1-\alpha}^s(\mathbf{x}_{n+1})$.

A.2 Proof of Section 3.2

Proof of Theorem 1

Let us consider $\bar{s}_{j,\mathcal{I}_1}(t)$, with $j \in \{1, \dots, p\}$. Since $m/n = \theta$ with $0 < \theta < 1$, if $n \rightarrow +\infty$ then $m \rightarrow +\infty$. The scalar γ is the empirical quantile of order $\lceil (m+1)(1-\alpha) \rceil$ of $\{\sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)| \right) : h \in \mathcal{I}_1\}$. First of all note that

$$\lim_{m \rightarrow +\infty} \frac{\lceil (m+1)(1-\alpha) \rceil}{m} = \lim_{m \rightarrow +\infty} \frac{m+1 - \lfloor (m+1)\alpha \rfloor}{m}$$

and since

$$\begin{aligned} \frac{(m+1)\alpha - 1}{m} &\leq \frac{\lfloor (m+1)\alpha \rfloor}{m} \leq \frac{(m+1)\alpha}{m} \quad \forall m \in \mathbb{N}, \\ \lim_{m \rightarrow +\infty} \frac{(m+1)\alpha - 1}{m} &= \lim_{m \rightarrow +\infty} \frac{(m+1)\alpha}{m} = \alpha \end{aligned}$$

then by the squeeze theorem we know that

$$\lim_{m \rightarrow +\infty} \frac{\lfloor (m+1)\alpha \rfloor}{m} = \alpha$$

and then

$$\lim_{m \rightarrow +\infty} \frac{\lceil (m+1)(1-\alpha) \rceil}{m} = 1 - \alpha.$$

Consequently, γ is the empirical quantile of order $1 - \alpha$ when $m \rightarrow +\infty$.

Let us define $w_h := \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)| \right) \forall h \in \mathcal{I}_1$. The random variables $\{W_h : h \in \mathcal{I}_1\}$ from which $\{w_h : h \in \mathcal{I}_1\}$ are drawn are continuous and after $\text{Var}[\hat{\mu}_{\mathcal{I}_1}^j(X_{hj})](t) \rightarrow 0 \forall j \in \{1, \dots, p\}$ they become i.i.d.. The Glivenko-Cantelli theorem guarantees that the empirical distribution function of these variables converges uniformly and almost surely pointwise to its distribution function, and so also the empirical quantiles converge in distribution - and so in probability - to the corresponding theoretical quantiles (see, for example, Van der Vaart 2000, chap. 21). In so doing, empirical quantile γ converges to $q_{1-\alpha}$, the theoretical quantile of order $1 - \alpha$. As a consequence:

$$\mathcal{H}_1 := \left\{ h \in \mathcal{I}_1 : \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)| \right) \leq q_{1-\alpha} \right\}$$

when $m \rightarrow +\infty$, with $q_{1-\alpha}$ non-random quantity. Let us consider the numerator of $\bar{s}_{j,\mathcal{I}_1}(t) \forall j \in \{1, \dots, p\}$ as the denominator is a normalizing constant. $\forall t \in \mathcal{T}_j$, the sequence $\{\max_{h \in \mathcal{H}_1} |y_{hj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)|\}_m$ is eventually bounded by $q_{1-\alpha}$ and is eventually increasing since $\{|\mathcal{H}_1|\}_m$ is eventually increasing. Therefore the sequence converges to its supremum by the monotone convergence theorem.

As regards $\bar{s}_{j,\mathcal{I}_1,\mathcal{I}_2}^c$, first of all it is possible to notice that if $n \rightarrow +\infty$ then $l = n(1-\theta) \rightarrow +\infty$. In order to show the convergence of the numerator of $\bar{s}_{j,\mathcal{I}_1,\mathcal{I}_2}^c$ to the same limit function, it is sufficient to consider the previous calculations by substituting γ with k , m with l , \mathcal{H}_1 with \mathcal{H}_2 and \mathcal{I}_1 with \mathcal{I}_2 (except for $[\hat{\mu}_{\mathcal{I}_1}^j(x_{hj})](t)$ that is not substituted by $[\hat{\mu}_{\mathcal{I}_2}^j(x_{hj})](t)$). Finally, as the numerators of $\bar{s}_{j,\mathcal{I}_1,\mathcal{I}_2}$ and $\bar{s}_{j,\mathcal{I}_1}^c$

converge to the same function $\forall j \in \{1, \dots, p\}$, also the two normalizing constants converge to the same value.

Proof of Theorem 2

For the sake of simplicity, let us focus on the case in which $|\mathcal{H}_2| = \lceil (l+1)(1-\alpha) \rceil$. Under the assumption concerning the continuous joint distribution of $\{R_d : d \in \mathcal{I}_2\}$ made in Section 2 such condition is always satisfied, but for the sake of completeness the proof when this assumption is violated is addressed below.

- $\forall d \in \mathcal{H}_2, \forall j \in \{1, \dots, p\}$ the following relationship holds $\forall t \in \mathcal{T}_j$:

$$\begin{aligned} & \left| \frac{y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)}{\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c(t)} \right| \\ &= \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt \cdot \frac{|y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)|}{\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)|} \\ &\leq \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt, \end{aligned}$$

and then

$$R_d^{\bar{s}^c} := \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)}{\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c(t)} \right| \right) \leq \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt.$$

Specifically, $\exists \underline{d} \in \mathcal{H}_2$ such that $R_{\underline{d}}^{\bar{s}^c} = \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt$ since $\forall j \in \{1, \dots, p\}$ and $\forall t \in \mathcal{T}_j$ at least one function $y_{\underline{d}, j}$ satisfies $|y_{\underline{d}, j}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{\underline{d}, j})](t)| = \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)|$.

- Let us define $\mathcal{CH}_2 := \mathcal{I}_2 \setminus \mathcal{H}_2$ and let (t_d^*, j_d^*) be the couple of values such that

$$\left| y_{dj_d^*}(t_d^*) - [\hat{\mu}_{\mathcal{I}_1}^{j_d^*}(x_{dj_d^*})](t_d^*) \right| = \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \quad \forall d \in \mathcal{I}_2.$$

If (t_d^*, j_d^*) is not unique, it is randomly chosen from the couples satisfying that condition.

$\forall b \in \mathcal{CH}_2$, by definition of \mathcal{H}_2 it is possible to notice that $\left| y_{bj_b^*}(t_b^*) - [\hat{\mu}_{\mathcal{I}_1}^{j_b^*}(x_{bj_b^*})](t_b^*) \right| > \max_{d \in \mathcal{H}_2} |y_{dj_b^*}(t_b^*) - [\hat{\mu}_{\mathcal{I}_1}^{j_b^*}(x_{dj_b^*})](t_b^*)|$ and so the following relationship holds:

$$\begin{aligned} & \left| \frac{y_{bj_b^*}(t_b^*) - [\hat{\mu}_{\mathcal{I}_1}^{j_b^*}(x_{bj_b^*})](t_b^*)}{\bar{s}_{j_b^*, \mathcal{I}_1, \mathcal{I}_2}^c(t_b^*)} \right| \\ &= \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt \cdot \frac{|y_{bj_b^*}(t_b^*) - [\hat{\mu}_{\mathcal{I}_1}^{j_b^*}(x_{bj_b^*})](t_b^*)|}{\max_{d \in \mathcal{H}_2} |y_{dj_b^*}(t_b^*) - [\hat{\mu}_{\mathcal{I}_1}^{j_b^*}(x_{dj_b^*})](t_b^*)|} \\ &> \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt. \end{aligned}$$

Consequently,

$$R_b^{\bar{s}^c} := \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{y_{bj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{bj})](t)}{\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c(t)} \right| \right) > \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt.$$

Since:

- $|\mathcal{H}_2| = \lceil (l+1)(1-\alpha) \rceil$
- $\forall d \in \mathcal{H}_2$ $R_d^{\bar{s}^c} \leq \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt$ and $\exists \underline{d} \in \mathcal{H}_2$ such that $R_{\underline{d}}^{\bar{s}^c} = \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt$
- $\forall b \in \mathcal{CH}_2$ $R_b^{\bar{s}^c} > \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt$

we conclude that $k^{\bar{s}^c} = \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt$, with $k^{\bar{s}^c}$ the $\lceil (l+1)(1-\alpha) \rceil$ th smallest value in the set $\{R_d^{\bar{s}^c} : d \in \mathcal{I}_2\}$.

If $|\mathcal{H}_2| > \lceil (l+1)(1-\alpha) \rceil$, then $R_d^{\bar{s}^c} = \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt$ is valid $\forall d \in \mathcal{H}_2$ such that $\sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) = k$ and we can conclude also in this case that $k^{\bar{s}^c} = \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt$.

Focusing now on the set of modulation functions s^0 , $\forall d \in \mathcal{I}_2$:

$$R_d^{s^0} := \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left| \frac{y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)}{s_j^0(t)} \right| \right) = \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \cdot \sum_{j=1}^p |\mathcal{T}_j|.$$

Since k^{s^0} is the $\lceil (l+1)(1-\alpha) \rceil$ th smallest value in the set $\{R_d^{s^0} : d \in \mathcal{I}_2\}$, by definition of \mathcal{H}_2 we can notice that

$$\begin{aligned} k^{s^0} &= \max_{d \in \mathcal{H}_2} R_d^{s^0} \\ &= \max_{d \in \mathcal{H}_2} \left(\sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \right) \cdot \sum_{j=1}^p |\mathcal{T}_j| \\ &= \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \right) \cdot \sum_{j=1}^p |\mathcal{T}_j|. \end{aligned}$$

Since by the integral mean value theorem we know that $\forall j \in \{1, \dots, p\}$

$$\sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \cdot |\mathcal{T}_j| \geq \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt,$$

then the following relationship is valid:

$$\sum_{j=1}^p \sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \cdot |\mathcal{T}_j| \geq \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt. \quad (7)$$

In addition, by definition $\forall j \in \{1, \dots, p\}$

$$\sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \right) \geq \sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right)$$

and so:

$$\begin{aligned}
& \sum_{j=1}^p \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \right) \cdot |\mathcal{T}_j| \\
&= \sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \right) \cdot \sum_{j=1}^p |\mathcal{T}_j| \\
&\geq \sum_{j=1}^p \sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \cdot |\mathcal{T}_j|.
\end{aligned} \tag{8}$$

By combining (7) and (8) we can notice that

$$\sup_{j \in \{1, \dots, p\}} \left(\sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \right) \cdot \sum_{j=1}^p |\mathcal{T}_j| \geq \sum_{j=1}^p \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt,$$

i.e. $k^{s^0} \geq k^{\bar{s}^c}$. Then, $\mathcal{Q}(s^0) \geq \mathcal{Q}(\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$.

Specifically, the integral mean value theorem guarantees that $\forall j \in \{1, \dots, p\}$

$$\begin{aligned}
\sup_{t \in \mathcal{T}_j} \left(\max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \right) \cdot |\mathcal{T}_j| &= \int_{\mathcal{T}_j} \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| dt \\
&\iff \max_{d \in \mathcal{H}_2} |y_{dj}(t) - [\hat{\mu}_{\mathcal{I}_1}^j(x_{dj})](t)| \quad \text{is constant almost everywhere,}
\end{aligned}$$

i.e. if and only if $\bar{s}_{j, \mathcal{I}_1, \mathcal{I}_2}^c(t)$ is constant almost everywhere over \mathcal{T}_j . Consequently, if at least one of the functions $\bar{s}_{1, \mathcal{I}_1, \mathcal{I}_2}^c(t), \dots, \bar{s}_{p, \mathcal{I}_1, \mathcal{I}_2}^c(t)$ is not constant almost everywhere over its domain then the left side of (7) is strictly greater than the right side (implying $\mathcal{Q}(s^0) > \mathcal{Q}(\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$); otherwise, $\mathcal{Q}(s^0) = \mathcal{Q}(\bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$.

Generalization of $(\bar{s}_{\mathcal{I}_1}, \bar{s}_{\mathcal{I}_1, \mathcal{I}_2}^c)$, Theorem 1 and Theorem 2 to the Smoothed Split Conformal framework

The functions $\bar{s}_{\mathcal{I}_1}^c$ and $\bar{s}_{\mathcal{I}_1}$ are defined as in the Split Conformal framework, except for: k (γ respectively) that is the $[l + \tau_{n+1} - (l + 1)\alpha]$ th ($[m + \tau_{n+1} - (m + 1)\alpha]$ th respectively) smallest value in the corresponding set; similarly to the Split Conformal framework, if $[m + \tau_{n+1} - (m + 1)\alpha] > m$ then $\mathcal{H}_1 = \mathcal{I}_1$ and if $[m + \tau_{n+1} - (m + 1)\alpha] \leq 0$ we arbitrarily set $\bar{s}_{j, \mathcal{I}_1} = s_j^0$. Theorem 1 and Theorem 2 still hold by substituting $[(l + 1)(1 - \alpha)]$, $[(m + 1)(1 - \alpha)]$ with $[l + \tau_{n+1} - (l + 1)\alpha]$, $[m + \tau_{n+1} - (m + 1)\alpha]$.

B Supplementary Figures

Acknowledgements

Prof. Vantini and Dr. Fontana acknowledge the financial support from Accordo Quadro ASI-POLIMI ‘‘Attivit  di Ricerca e Innovazione’’ n. 2018-5-HH.0, collaboration agreement between the Italian Space Agency and Politecnico di Milano. The authors are deeply grateful to Clear Channel Italia S.p.A, which provided the data for the case study, and to Agostino Torti for providing part of the code used in the case study.

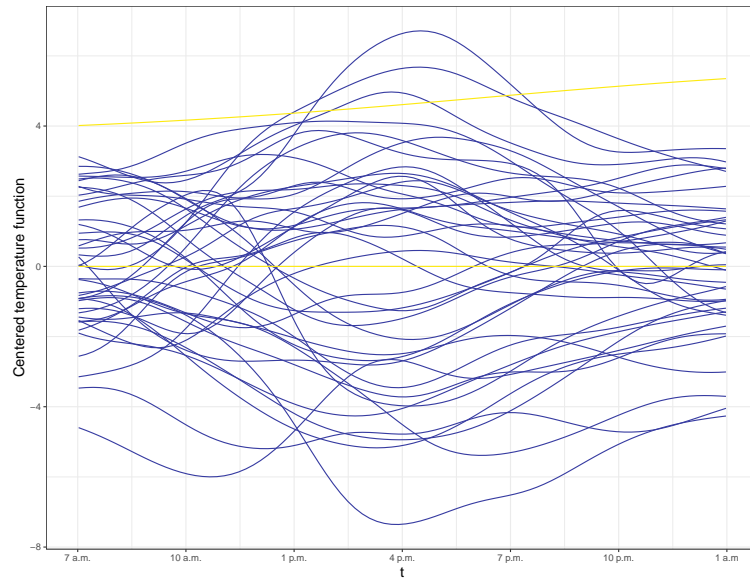


Figure B.1. Temperature function (after subtracting the average daily temperature function of the period considered) in degrees Celsius for the observed days (blue curves) and for the two hypothetical days (yellow curves).

References

- Antoniadis, A., Brossat, X., Cugliari, J., Poggi, J.M., 2016. A prediction interval for a function-valued forecast model: Application to load forecasting. *Int. J. Forecast.* 32, 939–947.
- Balasubramanian, V., Ho, S.S., Vovk, V., 2014. Conformal prediction for reliable machine learning: theory, adaptations and applications. Newnes.
- Cao, G., Yang, L., Todem, D., 2012. Simultaneous Inference For The Mean Function Based on Dense Functional Data. *J. Nonparametr. Stat.* 24, 359–377.
- Degras, D.A., 2017. Simultaneous confidence bands for the mean of functional data. *Wiley Interdiscip. Rev. Comput. Stat.* 9.
- Degras, D.A., 2011. Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* 21.
- Diquigiovanni, J., Fontana, M., Vantini, S., 2021. The importance of being a band: Finite-sample exact distribution-free prediction sets for functional data. [arXiv:2102.06746](https://arxiv.org/abs/2102.06746).
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Hyndman, R.J., Shahid Ullah, M., 2007. Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Statist. Data Anal.* 51, 4942–4956.
- Inselberg, A., 1985. The plane with parallel coordinates. *Vis Comput* 1, 69–91.

- Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R.J., Wasserman, L., 2018. Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* 113, 1094–1111.
- Lei, J., Robins, J., Wasserman, L., 2013. Distribution-free prediction sets. *J. Amer. Statist. Assoc.* 108, 278–287.
- López-Pintado, S., Romo, J., 2009. On the concept of depth for functional data. *J. Amer. Statist. Assoc.* 104, 718–734.
- Papadopoulos, H., Proedrou, K., Vovk, V., Gammerman, A., 2002. Inductive confidence machines for regression, in: *European Conference on Machine Learning*, Springer. pp. 345–356.
- R Core Team, 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Ramsay, J.O., 1982. When the data are functions. *Psychometrika* 47, 379–396.
- Ramsay, J.O., Silverman, B.W., 2005. *Functional data analysis*. Springer series in statistics. second edition ed., Springer, New York, NY.
- Rao, P., 1971. Some notes on misspecification in multiple regressions. *Amer. Statist.* 25, 37–39.
- Sun, Y., Genton, M.G., 2011. Functional Boxplots. *J. Comput. Graph. Statist.* 20, 316–334.
- Torti, A., Pini, A., Vantini, S., 2021. Modelling time-varying mobility flows using function-on-function regression: Analysis of a bike sharing system in the city of milan. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 70, 226–247.
- Van der Vaart, A. W., 2000. *Asymptotic statistics*, Vol.3. Cambridge University press.
- Vapnik, V., 1992. Principles of risk minimization for learning theory, in: *Advances in neural information processing systems*, pp. 831–838.
- Vovk, V., Gammerman, A., Shafer, G., 2005. *Algorithmic learning in a random world*. Springer Science & Business Media.
- Wooldridge, J.M., 1994. A simple specification test for the predictive ability of transformation models. *Rev. Econ. Stat.* 76, 59–65.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 45/2021** Diquigiovanni, J.; Fontana, M.; Vantini, S.
Distribution-Free Prediction Bands for Multivariate Functional Time Series: an Application to the Italian Gas Market
- 44/2021** Gentili, G.G.; Khosronejad, M.; Bernasconi, G.; Perotto, S.; Micheletti, S.
Efficient Modeling of Multimode Guided Acoustic Wave Propagation in Deformed Pipelines by Hierarchical Model Reduction
- 43/2021** Salvador, M.; Fedele, M.; Africa, P.C.; Sung, E.; Dede', L.; Prakosa, A.; Chrispin, J.; Trayanov
Electromechanical modeling of human ventricles with ischemic cardiomyopathy: numerical simulations in sinus rhythm and under arrhythmia
- 42/2021** Calissano, A.; Fontana, M.; Zeni, G.; Vantini, S.
Conformal Prediction Sets for Populations of Graphs
- 41/2021** Costa, G., Cavinato, L., Maschi, C., Fiz, F., Sollini, M., Politi, L. S., Chiti, A., Balzarini, L., Ag
Virtual Biopsy for Diagnosis of Chemotherapy-Associated Liver Injuries and Steatohepatitis: A Combined Radiomic and Clinical Model in Patients with Colorectal Liver Metastases
- 39/2021** Barnafi, N.; Di Gregorio, S.; Dede', L.; Zunino, P.; Vergara, C.; Quarteroni, A.
A multiscale poromechanics model integrating myocardial perfusion and systemic circulation
- 40/2021** Martinolli, M.; Cornat, F.; Vergara, C.
Computational Fluid-Structure Interaction Study of a new Wave Membrane Blood Pump
- 38/2021** Giusteri, G. G.; Miglio, E.; Parolini, N.; Penati, M.; Zambetti, R.
Simulation of viscoelastic Cosserat rods based on the geometrically exact dynamics of special Euclidean strands
- 36/2021** Parolini, N.; Dede', L.; Antonietti, P. F.; Ardenghi, G.; Manzoni, A.; Miglio, E.; Pugliese, A.; V
SUIHTER: A new mathematical model for COVID-19. Application to the analysis of the second epidemic outbreak in Italy
- 37/2021** Dassi, F.; Fumagalli, A.; Mazzieri, I.; Scotti, A.; Vacca, G.
A Virtual Element Method for the wave equation on curved edges in two dimensions