The Importance of Being a Band: Finite-Sample Exact Distribution-Free Prediction Sets for Functional Data

Jacopo Diquigiovanni[1,4], Matteo Fontana[2,3], Simone Vantini[2]

[1]Department of Statistical Sciences, University of Padova, Italy.

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

[3]now at Royal Holloway, University of London

[4]Corresponding Author: jacopo.diquigiovanni@phd.unipd.it

S1 Technical Proofs

S1.1 Proofs of Section 2.2

Proof of Theorem 1.

Since $\mathcal{C}_{n,1-\alpha} := \{y \in \mathcal{Y}(\mathcal{T}) : \delta_y > \alpha\}$, then $\mathcal{C}_{n,1-\alpha} := \{y \in \mathcal{Y}(\mathcal{T}) : (l+1)\delta_y > (l+1)\alpha\}$. Under the hypothesis of the theorem, $(l+1)\delta_Y \sim U\{1, 2, \ldots, l+1\}$ holds. As a consequence:

$$
\begin{aligned}
\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha}\right) &= \mathbb{P}\left((l+1)\delta_Y > (l+1)\alpha\right) \\
&= 1 - \mathbb{P}\left((l+1)\delta_Y \leq (l+1)\alpha\right) \\
&= 1 - \frac{\lfloor (l+1)\alpha \rfloor}{l+1}.
\end{aligned}
$$

In addition, since

$$\frac{\lfloor (l+1)\alpha \rfloor}{l+1} \le \frac{(l+1)\alpha}{l+1} = \alpha$$

then $\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha}\right) \ge 1-\alpha$, i.e. $\mathcal{C}_{n,1-\alpha}$ is valid. Finally, since

$$\frac{\lfloor (l+1)\alpha \rfloor}{l+1} > \frac{(l+1)\alpha - 1}{l+1} = \alpha - \frac{1}{l+1}$$

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

Proof that smoothed split conformal prediction sets are exact.

Let us consider the hypothesis of Theorem 1. Let us notice that

$$\begin{aligned}\delta_{y,\tau_{n+1}} :=& \frac{\left|\{j \in \mathcal{I}_2 : R_j > R_{n+1}\}\right| + \tau_{n+1}\left|\{j \in \mathcal{I}_2 \cup \{n+1\} : R_j = R_{n+1}\}\right|}{l+1} \\ =& \frac{\tau_{n+1}}{l+1} + \frac{\left|\{j \in \mathcal{I}_2 : R_j \ge R_{n+1}\}\right|}{l+1}.\end{aligned}$$

Under the hypothesis of Theorem 1, $\left|\{j \in \mathcal{I}_2 : R_j \ge R_{n+1}\}\right| \sim U\{0,1,\ldots,l\}$ holds. As a consequence:

$$\begin{aligned}\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha,\tau_{n+1}} | \tau_{n+1}\right) &= \mathbb{P}\left(\delta_{Y,\tau_{n+1}} > \alpha | \tau_{n+1}\right) \\ &= \mathbb{P}\left(\left|\{j \in \mathcal{I}_2 : R_j \ge R_{n+1}\}\right| > (l+1)\alpha - \tau_{n+1} | \tau_{n+1}\right) \\ &= 1 - \mathbb{P}\left(\left|\{j \in \mathcal{I}_2 : R_j \ge R_{n+1}\}\right| \le (l+1)\alpha - \tau_{n+1} | \tau_{n+1}\right) \\ &= 1 - \frac{\lfloor (l+1)\alpha - \tau_{n+1} \rfloor + 1}{l+1}.\end{aligned}$$

Let us call $f(\tau_{n+1}) = 1 \cdot \mathbb{1}\{\tau_{n+1} \in [0,1]\}$. Then

$$\begin{aligned}\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha,\tau_{n+1}}\right) =& \int_0^1 \mathbb{P}\left(Y_{n+1} \in \mathcal{C}_{n,1-\alpha,\tau_{n+1}} | \tau_{n+1}\right) f(\tau_{n+1}) d\tau_{n+1} \\ =& 1 - \\ & \left(\int_0^{(l+1)\alpha - \lfloor (l+1)\alpha \rfloor} \frac{\lfloor (l+1)\alpha - \tau_{n+1} \rfloor + 1}{l+1} d\tau_{n+1} + \right. \\ & \left. \int_{(l+1)\alpha - \lfloor (l+1)\alpha \rfloor}^1 \frac{\lfloor (l+1)\alpha - \tau_{n+1} \rfloor + 1}{l+1} d\tau_{n+1} \right).\end{aligned}$$

Let us consider $\int_0^{(l+1)\alpha-\lfloor(l+1)\alpha\rfloor} \frac{\lfloor(l+1)\alpha-\tau_{n+1}\rfloor+1}{l+1}d\tau_{n+1}$. Since if $\tau_{n+1} \leq (l+1)\alpha-\lfloor(l+1)\alpha\rfloor$ then $\lfloor(l+1)\alpha-\tau_{n+1}\rfloor = \lfloor(l+1)\alpha\rfloor$, we can notice that

$$\begin{aligned}
&\int_0^{(l+1)\alpha-\lfloor(l+1)\alpha\rfloor} \frac{\lfloor(l+1)\alpha-\tau_{n+1}\rfloor+1}{l+1}d\tau_{n+1} \\
=&\int_0^{(l+1)\alpha-\lfloor(l+1)\alpha\rfloor} \frac{\lfloor(l+1)\alpha\rfloor+1}{l+1}d\tau_{n+1} \\
=&\frac{\lfloor(l+1)\alpha\rfloor+1}{l+1} \cdot \left((l+1)\alpha-\lfloor(l+1)\alpha\rfloor\right).
\end{aligned}$$

Let us consider $\int_{(l+1)\alpha-\lfloor(l+1)\alpha\rfloor}^1 \frac{\lfloor(l+1)\alpha-\tau_{n+1}\rfloor+1}{l+1}d\tau_{n+1}$. Since if $\tau_{n+1} > (l+1)\alpha-\lfloor(l+1)\alpha\rfloor$ then $\lfloor(l+1)\alpha-\tau_{n+1}\rfloor = \lfloor(l+1)\alpha\rfloor-1$, we can notice that

$$\begin{aligned}
&\int_{(l+1)\alpha-\lfloor(l+1)\alpha\rfloor}^1 \frac{\lfloor(l+1)\alpha-\tau_{n+1}\rfloor+1}{l+1}d\tau_{n+1} \\
=&\int_{(l+1)\alpha-\lfloor(l+1)\alpha\rfloor}^1 \frac{\lfloor(l+1)\alpha\rfloor}{l+1}d\tau_{n+1} \\
=&\frac{\lfloor(l+1)\alpha\rfloor}{l+1} \cdot \left(1-\left((l+1)\alpha-\lfloor(l+1)\alpha\rfloor\right)\right).
\end{aligned}$$

Then

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

S1.2 Proofs of Section 2.3

Proof that the concatenation of pointwise prediction intervals leads to a prediction band that is a subset of the simultaneous prediction band (2.4).

Let $\mathcal{U}_{n,1-\alpha}$ be the pointwise prediction set. Let us define $\tilde{R}_j(t) := |y_j(t) - g_{\mathcal{I}_1}(t)|\ \forall t \in \mathcal{T}, j \in \mathcal{I}_2,\ \tilde{R}_{n+1}(t) := |y(t) - g_{\mathcal{I}_1}(t)|$ for a given $y \in \mathcal{Y}(\mathcal{T})$ and $\tilde{k}(t)$ the $\lceil (l+1)(1-\alpha) \rceil$th smallest value in the set $\{\tilde{R}_h(t) : h \in \mathcal{I}_2\}$. By construction $R_j = \operatorname{ess\,sup}_{t \in \mathcal{T}} \tilde{R}_j(t)$, and so $R_j \geq \tilde{R}_j(t)\ \forall t \in \mathcal{T}, j \in \mathcal{I}_2$ and then $k \geq \tilde{k}(t)\ \forall t \in \mathcal{T}$. Let us consider $y \in \mathcal{U}_{n,1-\alpha}$, i.e. $y(t) \in [g_{\mathcal{I}_1}(t) - \tilde{k}(t), g_{\mathcal{I}_1}(t) + \tilde{k}(t)] \quad \forall t \in \mathcal{T}$. Since $k \geq \tilde{k}(t)$, also $y(t) \in [g_{\mathcal{I}_1}(t) - k, g_{\mathcal{I}_1}(t) + k] \quad \forall t \in \mathcal{T}$, i.e. $y \in \mathcal{C}_{n,1-\alpha}$.

Since the converse is not necessarily true (in the sense that $y \in \mathcal{C}_{n,1-\alpha}$ does not imply $y \in \mathcal{U}_{n,1-\alpha}$), we conclude that $\mathcal{U}_{n,1-\alpha} \subseteq \mathcal{C}_{n,1-\alpha}$.

S1.3 Proofs of Section 2.4

Proof of the prediction set induced by the nonconformity measure $A(\{y_h : h \in \mathcal{I}_1\}, y) = \operatorname{ess\,sup}_{t \in \mathcal{T}} \left| \frac{y(t) - g_{\mathcal{I}_1}(t)}{s_{\mathcal{I}_1}(t)} \right|$**.**

For a given $y \in \mathcal{Y}(\mathcal{T})$, let us define

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

The split conformal prediction set is defined as $\mathcal{C}_{n,1-\alpha}^s := \left\{ y \in \mathcal{Y}(\mathcal{T}) : \delta_y^s > \alpha \right\}$. As a consequence, $y \in \mathcal{C}_{n,1-\alpha}^s \iff R_{n+1}^s \leq k^s$, with k^s the $\lceil (l+1)(1-\alpha) \rceil$th smallest value in the set $\{R_h^s : h \in \mathcal{I}_2\}$. Then:

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

Therefore, the split conformal prediction set is

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

Proof of Remark 7.

Let us define $\mathcal{C}^{\lambda\cdot s}_{n,1-\alpha}$ the prediction set obtained by considering the modulation function $\lambda \cdot s_{\mathcal{I}_1}$. The nonconformity scores are

$$R^{\lambda\cdot s}_j = \operatorname*{ess\,sup}_{t\in\mathcal{T}} \left| \frac{y_j(t) - g_{\mathcal{I}_1}(t)}{\lambda \cdot s_{\mathcal{I}_1}(t)} \right| = \frac{1}{\lambda} R^s_j, \quad j \in \mathcal{I}_2$$
$$R^{\lambda\cdot s}_{n+1} = \operatorname*{ess\,sup}_{t\in\mathcal{T}} \left| \frac{y(t) - g_{\mathcal{I}_1}(t)}{\lambda \cdot s_{\mathcal{I}_1}(t)} \right| = \frac{1}{\lambda} R^s_{n+1}.$$

Let us also define

$$\delta^{\lambda\cdot s}_y := \frac{\left|\left\{j \in \mathcal{I}_2 \cup \{n+1\} : R^{\lambda\cdot s}_j \geq R^{\lambda\cdot s}_{n+1}\right\}\right|}{l+1}.$$

The split conformal prediction set is defined as $\mathcal{C}^{\lambda\cdot s}_{n,1-\alpha} := \left\{y \in \mathcal{Y}(\mathcal{T}) : \delta^{\lambda\cdot s}_y > \alpha\right\}$. As a consequence, $y \in \mathcal{C}^{\lambda\cdot s}_{n,1-\alpha} \iff R^{\lambda\cdot s}_{n+1} \leq k^{\lambda\cdot s}$, with $k^{\lambda\cdot s}$ the $\lceil (l+1)(1-\alpha)\rceil$th smallest value in the set $\{R^{\lambda\cdot s}_h : h \in \mathcal{I}_2\}$. In addition, since $R^{\lambda\cdot s}_j = R^s_j/\lambda \ \forall j \in \mathcal{I}_2$, then $k^{\lambda\cdot s} = k^s/\lambda$. Then:

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

and since $y \in \mathcal{C}^s_{n,1-\alpha} \iff R^s_{n+1} \leq k^s$, then $\mathcal{C}^{\lambda\cdot s}_{n,1-\alpha} = \mathcal{C}^s_{n,1-\alpha}$.

Adjustment procedure of $\bar{s}^c_{\mathcal{I}_1}$ and $\bar{s}_{\mathcal{I}_1}$

If $\max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| = 0$ for at least one value t but the condition $\int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt \neq 0$ still holds, in order to ensure that $\bar{s}^c_{\mathcal{I}_1}(t) > 0 \ \forall t \in \mathcal{T}$ it is sufficient to add an arbitrarily (small) positive value to $\bar{s}^c_{\mathcal{I}_1}(t)$ $\forall t \in \mathcal{T}$ and to adjust the normalization constant accordingly. The pathological case in which $\int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt = 0$ is addressed only when $y_j(t) = g_{\mathcal{I}_1}(t)$ $\forall j \in \mathcal{H}_2$ and almost every $t \in \mathcal{T}$ and it represents a case of no practical interest.

Should $\exists\, t \in \mathcal{T}$ such that $\max_{j \in \mathcal{H}_1} |y_j(t) - g_{\mathcal{I}_1}(t)| = 0$, the same procedure is developed.

Proof of Theorem 2.

Let us focus on $\bar{s}_{\mathcal{I}_1}(t)$. Since $m/n = \theta$ with $0 < \theta < 1$, if $n \to +\infty$ then $m \to +\infty$. By definition, the scalar γ is the empirical quantile of order $\lceil (m+1)(1-\alpha) \rceil)$ of $\{\operatorname{ess\,sup}_{t \in \mathcal{T}} |y_h(t) - g_{\mathcal{I}_1}(t)| : h \in \mathcal{I}_1\}$. First of all note that

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

and since

$$\frac{(m+1)\alpha - 1}{m} \leq \frac{\lfloor (m+1)\alpha \rfloor}{m} \leq \frac{(m+1)\alpha}{m} \quad \forall m \in \mathbb{N}$$

and

$$\lim_{m \to +\infty} \frac{(m+1)\alpha - 1}{m} = \lim_{m \to +\infty} \frac{(m+1)\alpha}{m} = \alpha$$

then by the squeeze theorem (also known as the sandwich theorem) we obtain that

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

and then

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

As a consequence, γ is the empirical quantile of order $1 - \alpha$ when $m \to +\infty$.

For convenience, let us define $x_i := \operatorname{ess\,sup}_{t \in \mathcal{T}} |y_i(t) - g_{\mathcal{I}_1}(t)|$ $\forall\, i \in \mathcal{I}_1$. The random variables $\{X_h : h \in \mathcal{I}_1\}$ from which $\{x_h : h \in \mathcal{I}_1\}$ are drawn are continuous and they are asymptotically i.i.d. as $\mathrm{Var}[g_{\mathcal{I}_1}(t)] \to 0$. The Glivenko-Cantelli theorem ensures that the empirical distribution function of these variables converges uniformly (and almost surely pointwise) to its distribution function, and then also the empirical quantiles converge in distribution (and so in probability) to the corresponding theoretical quantiles, as shown for example by Van der Vaart (2000, chap. 21). Specifically, empirical quantile γ converges to $q_{1-\alpha}$, the theoretical quantile of order $1-\alpha$. As a consequence, when $m \to +\infty$:

$$\mathcal{H}_1 := \{j \in \mathcal{I}_1 : \operatorname*{ess\,sup}_{t \in \mathcal{T}} |y_j(t) - g_{\mathcal{I}_1}(t)| \leq q_{1-\alpha}\}$$

with $q_{1-\alpha}$ deterministic quantity. Let us focus on the numerator of $\bar{s}_{\mathcal{I}_1}(t)$ since the denominator is just a normalizing constant. $\forall t \in \mathcal{T}$, the sequence $\{\max_{j \in \mathcal{H}_1} |y_j(t) - g_{\mathcal{I}_1}(t)|\}_m$ is eventually bounded by $q_{1-\alpha}$ and is eventually increasing since $\{|\mathcal{H}_1|\}_m$ is eventually increasing. By the monotone convergence theorem, the sequence converges to its supremum.

In order to prove the convergence of the numerator of $\bar{s}^c_{\mathcal{I}_1}$ to the same limit function, it is sufficient to consider the previous computations by noting that if $n \to +\infty$ then $l = n(1-\theta) \to +\infty$ and 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 $g_{\mathcal{I}_1}$ that is naturally not substituted by $g_{\mathcal{I}_2}$). Since the numerators of $\bar{s}_{\mathcal{I}_1}$ and $\bar{s}^c_{\mathcal{I}_1}$ converge to the same function, also the two normalizing constants converge to the same quantity. In view of this and since $\mathcal{C}^{\bar{s}}_{n,1-\alpha}$ and $\mathcal{C}^{\bar{s}^c}_{n,1-\alpha}$ are defined as

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

$$\mathcal{C}^{\bar{s}^c}_{n,1-\alpha} := \left\{y \in \mathcal{Y}(\mathcal{T}) : y(t) \in [g_{\mathcal{I}_1}(t) - k^{\bar{s}^c}\bar{s}^c_{\mathcal{I}_1}(t), g_{\mathcal{I}_1}(t) + k^{\bar{s}^c}\bar{s}^c_{\mathcal{I}_1}(t)] \quad \forall t \in \mathcal{T}\right\}$$

then $\lim_{n\to+\infty} \mathcal{C}^{\bar{s}}_{n,1-\alpha} = \lim_{n\to+\infty} \mathcal{C}^{\bar{s}^c}_{n,1-\alpha}$.

Proof of Theorem 3.

The proof consists of two steps. At the first step we show that $k^{\bar{s}^c} = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$, a fundamental result to obtain, at the second step, the proof of the theorem.

I step

In order not to overcomplicate the proof, first of all let us consider the case in which $|\mathcal{H}_2| = \lceil (l+1)(1-\alpha) \rceil$. It is important to notice that under the assumption concerning the continuous joint distribution of $\{R_h : h \in \mathcal{I}_2\}$ made in Section 2.2 such condition is always satisfied. However, the result proved at this first step holds also when this assumption is violated, and its proof requires just minor changes. Therefore, for the sake of completeness such proof is addressed below.

- $\forall i \in \mathcal{H}_2$ the following relationship holds $\forall t \in \mathcal{T}$:

$$\begin{aligned}
&\left| \frac{y_i(t) - g_{\mathcal{I}_1}(t)}{\bar{s}^c_{\mathcal{I}_1}(t)} \right| \\
&= \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt \cdot \frac{|y_i(t) - g_{\mathcal{I}_1}(t)|}{\max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|} \\
&\leq \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt,
\end{aligned}$$

  and then

$$R_i^{\bar{s}^c} := \operatorname*{ess\,sup}_{t \in \mathcal{T}} \left| \frac{y_i(t) - g_{\mathcal{I}_1}(t)}{\bar{s}^c_{\mathcal{I}_1}(t)} \right| \leq \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt.$$

  Specifically, $\exists\ \underline{i} \in \mathcal{H}_2$ such that $R_{\underline{i}}^{\bar{s}^c} = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$ since $\forall t \in \mathcal{T}$ at least one function $y_{\underline{i}}$ satisfies $|y_{\underline{i}}(t) - g_{\mathcal{I}_1}(t)| = \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|$.

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

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

  If t_i^* is not unique, it is randomly chosen from the values that satisfy that condition. $\forall i \in \mathcal{CH}_2$, by definition of $\mathcal{H}_2$ we obtain that $|y_i(t_i^*) - g_{\mathcal{I}_1}(t_i^*)| > \max_{j \in \mathcal{H}_2} |y_j(t_i^*) - g_{\mathcal{I}_1}(t_i^*)|$

and so the following relationship holds:

$$
\begin{aligned}
&\left|\frac{y_i(t_i^*) - g_{\mathcal{I}_1}(t_i^*)}{\bar{s}^c_{\mathcal{I}_1}(t_i^*)}\right| \\
&= \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt \cdot \frac{|y_i(t_i^*) - g_{\mathcal{I}_1}(t_i^*)|}{\max_{j \in \mathcal{H}_2} |y_j(t_i^*) - g_{\mathcal{I}_1}(t_i^*)|} \\
&> \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt.
\end{aligned}
$$

As a consequence,

$$
R_i^{\bar{s}^c} := \operatorname*{ess\,sup}_{t \in \mathcal{T}} \left|\frac{y_i(t) - g_{\mathcal{I}_1}(t)}{\bar{s}^c_{\mathcal{I}_1}(t)}\right| > \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt.
$$

Since:

- $|\mathcal{H}_2| = \lceil (l+1)(1-\alpha) \rceil$
- $\forall i \in \mathcal{H}_2$ $R_i^{\bar{s}^c} \leq \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$ and $\exists\, \underline{i} \in \mathcal{H}_2$ such that $R_{\underline{i}}^{\bar{s}^c} = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$
- $\forall i \in \mathcal{C}\mathcal{H}_2$ $R_i^{\bar{s}^c} > \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$

we conclude that $k^{\bar{s}^c} = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$, with $k^{\bar{s}^c}$ the $\lceil (l+1)(1-\alpha) \rceil$th smallest value in the set $\{R_h^{\bar{s}^c} : h \in \mathcal{I}_2\}$.

If $|\mathcal{H}_2| > \lceil (l+1)(1-\alpha) \rceil$, then $R_i^{\bar{s}^c} = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$ is valid $\forall i \in \mathcal{H}_2$ such that $\operatorname{ess\,sup}_{t \in \mathcal{T}} |y_i(t) - g_{\mathcal{I}_1}(t)| = k$ and in the same way we can conclude that $k^{\bar{s}^c} = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$.

II step

Let us define $\forall i \in \mathcal{I}_2$

$$
R_i^{s^0} := \operatorname*{ess\,sup}_{t \in \mathcal{T}} \left|\frac{y_i(t) - g_{\mathcal{I}_1}(t)}{s^0(t)}\right| = |\mathcal{T}| \operatorname*{ess\,sup}_{t \in \mathcal{T}} |y_i(t) - g_{\mathcal{I}_1}(t)|\,.
$$

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

$$\begin{aligned} k^{s^0} &= |\mathcal{T}| \max_{j \in \mathcal{H}_2} \left(\operatorname*{ess\,sup}_{t \in \mathcal{T}} |y_j(t) - g_{\mathcal{I}_1}(t)| \right) \\ &= |\mathcal{T}| \operatorname*{ess\,sup}_{t \in \mathcal{T}} \left(\max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| \right). \end{aligned}$$

Since at the first step we proved that $k^{\bar{s}^c} = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$, we obtain that

$$k^{s^0} - k^{\bar{s}^c} = |\mathcal{T}| \operatorname*{ess\,sup}_{t \in \mathcal{T}} \left(\max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| \right) - \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt.$$

Since the right side of the equation is greater than or equal to 0 by the integral mean value theorem, then $\mathcal{Q}(s^0) \geq \mathcal{Q}(\bar{s}^c_{\mathcal{I}_1})$.

The same theorem ensures that

$$\begin{aligned} &|\mathcal{T}| \operatorname*{ess\,sup}_{t \in \mathcal{T}} \left(\max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| \right) = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt \\ \iff &\max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| \quad \text{is constant almost everywhere,} \end{aligned}$$

i.e. if and only if $\bar{s}^c_{\mathcal{I}_1}(t) = \bar{s}^0(t)$ almost everywhere.

Proof of Theorem 4.

We have already shown at the first step of the previous proof that $k^{\bar{s}^c} = \int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$. Since by assumption $s^d_{\mathcal{I}_1}(t_i^*) \leq \bar{s}^c_{\mathcal{I}_1}(t_i^*)$ $\forall i \in \mathcal{CH}_2$ and $|\mathcal{H}_2| = \lceil (l+1)(1-\alpha) \rceil$, let us define $a_i \geq 0$ $\forall i \in \mathcal{CH}_2$ the value such that $s^d_{\mathcal{I}_1}(t_i^*) = \bar{s}^c_{\mathcal{I}_1}(t_i^*) - a_i$.

- *Case 1*: If $\exists\ x \in \mathcal{CH}_2$ s.t. $a_x > 0$, $\exists\ \underline{i} \in \mathcal{H}_2$ such that

$$
\begin{aligned}
&\left|\frac{y_{\underline{i}}(t_x^*) - g_{\mathcal{I}_1}(t_x^*)}{s^d_{\mathcal{I}_1}(t_x^*)}\right| \\
=&\left|\frac{y_{\underline{i}}(t_x^*) - g_{\mathcal{I}_1}(t_x^*)}{\bar{s}^c_{\mathcal{I}_1}(t_x^*) - a_x}\right| \\
=&\int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|dt \quad \times \\
&\frac{|y_{\underline{i}}(t_x^*) - g_{\mathcal{I}_1}(t_x^*)|}{\max_{j\in\mathcal{H}_2} |y_j(t_x^*) - g_{\mathcal{I}_1}(t_x^*)| - a_x \cdot \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|dt} \\
>&\int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|dt
\end{aligned}
$$

since $\forall t \in \mathcal{T}$ (and specifically for t_x^*) at least one function $y_{\underline{i}}$ satisfies $|y_{\underline{i}}(t) - g_{\mathcal{I}_1}(t)| = \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|$.

Case 2: If $a_i = 0\ \forall i \in \mathcal{CH}_2$, there exist at least two values $t_\downarrow, t_\uparrow \in \mathcal{T}^*$ such that $s^d_{\mathcal{I}_1}(t_\downarrow) < \bar{s}^c_{\mathcal{I}_1}(t_\downarrow)$ and $s^d_{\mathcal{I}_1}(t_\uparrow) > \bar{s}^c_{\mathcal{I}_1}(t_\uparrow)$ since otherwise $s^d_{\mathcal{I}_1}(t) = \bar{s}^c_{\mathcal{I}_1}(t)\ \forall t \in \mathcal{T}^*$. Let us define $a_\downarrow > 0$ the value such that $s^d_{\mathcal{I}_1}(t_\downarrow) = \bar{s}^c_{\mathcal{I}_1}(t_\downarrow) - a_\downarrow$. Therefore $\exists\ \underline{i} \in \mathcal{H}_2$ such that

$$
\begin{aligned}
&\left|\frac{y_{\underline{i}}(t_\downarrow) - g_{\mathcal{I}_1}(t_\downarrow)}{s^d_{\mathcal{I}_1}(t_\downarrow)}\right| \\
=&\left|\frac{y_{\underline{i}}(t_\downarrow) - g_{\mathcal{I}_1}(t_\downarrow)}{\bar{s}^c_{\mathcal{I}_1}(t_\downarrow) - a_\downarrow}\right| \\
=&\int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|dt \quad \times \\
&\frac{|y_{\underline{i}}(t_\downarrow) - g_{\mathcal{I}_1}(t_\downarrow)|}{\max_{j\in\mathcal{H}_2} |y_j(t_\downarrow) - g_{\mathcal{I}_1}(t_\downarrow)| - a_\downarrow \cdot \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|dt} \\
>&\int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|dt
\end{aligned}
$$

since $\forall t \in \mathcal{T}$ (and specifically for $t_\downarrow$) at least one function $y_{\underline{i}}$ satisfies $|y_{\underline{i}}(t) - g_{\mathcal{I}_1}(t)| = \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)|$.

As a consequence, in both cases ($\exists x \in \mathcal{CH}_2$ s.t. $a_x > 0$ and $a_i = 0\ \forall i \in \mathcal{CH}_2$) we obtain that $\exists\ \underline{i} \in \mathcal{H}_2$ such that

$$R_{\underline{i}}^{s^d} := \operatorname*{ess\,sup}_{t\in\mathcal{T}} \left| \frac{y_{\underline{i}}(t) - g_{\mathcal{I}_1}(t)}{s^d_{\mathcal{I}_1}(t)} \right| > \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt.$$

- $\forall i \in \mathcal{CH}_2$, by definition of $\mathcal{H}_2$ we obtain that $|y_i(t_i^*) - g_{\mathcal{I}_1}(t_i^*)| > \max_{j\in\mathcal{H}_2} |y_j(t_i^*) - g_{\mathcal{I}_1}(t_i^*)|$ and so the following relationship holds:

$$\begin{aligned} & \left| \frac{y_i(t_i^*) - g_{\mathcal{I}_1}(t_i^*)}{s^d_{\mathcal{I}_1}(t_i^*)} \right| \\ = & \left| \frac{y_i(t_i^*) - g_{\mathcal{I}_1}(t_i^*)}{\bar{s}^c_{\mathcal{I}_1}(t_i^*) - a_i} \right| \\ = & \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt \quad \times \\ & \frac{|y_i(t_i^*) - g_{\mathcal{I}_1}(t_i^*)|}{\max_{j\in\mathcal{H}_2} |y_j(t_i^*) - g_{\mathcal{I}_1}(t_i^*)| - a_j \cdot \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt} \\ > & \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt. \end{aligned}$$

  As a consequence,

$$R_i^{s^d} := \operatorname*{ess\,sup}_{t\in\mathcal{T}} \left| \frac{y_i(t) - g_{\mathcal{I}_1}(t)}{s^d_{\mathcal{I}_1}(t)} \right| > \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt.$$

Since:

- $|\mathcal{H}_2| = \lceil (l+1)(1-\alpha) \rceil$
- $\exists\ \underline{i} \in \mathcal{H}_2$ such that $R_{\underline{i}}^{s^d} > \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$
- $\forall i \in \mathcal{CH}_2\ R_i^{s^d} > \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$

we conclude that $k^{s^d} > \int_{\mathcal{T}} \max_{j\in\mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt$, i.e. $k^{s^d} > k^{\bar{s}^c}$, with k^{s^d} the $\lceil (l+1)(1-\alpha) \rceil$th smallest value in the set $\{R_h^{s^d} : h \in \mathcal{I}_2\}$.

Proof that Theorem 4 does not imply Theorem 3.

Theorem 4 does not imply Theorem 3 since s^0 may not fulfill $s^0(t_i^*) \leq \bar{s}^c_{\mathcal{I}_1}(t_i^*)$ $\forall i \in \mathcal{CH}_2$. In fact, $\forall i \in \mathcal{CH}_2$:

$$s^0(t_i^*) \leq \bar{s}^c_{\mathcal{I}_1}(t_i^*) \iff \frac{\int_{\mathcal{T}} \max_{j \in \mathcal{H}_2} |y_j(t) - g_{\mathcal{I}_1}(t)| dt}{|\mathcal{T}|} \leq \max_{j \in \mathcal{H}_2} |y_j(t_i^*) - g_{\mathcal{I}_1}(t_i^*)|$$

and the condition on the right side is not always satisfied because no constraints are imposed on $y_j(t_i^*)$, with $j \in \mathcal{H}_2$, $i \in \mathcal{CH}_2$.

S1.4 Proofs about Smoothed Conformal Predictor

Proof of the smoothed conformal prediction set

By considering the notation of Section 2, first of all let us notice that, by definition, $\mathcal{C}_{n,1-\alpha,1} = \mathcal{C}_{n,1-\alpha}$.

Since $\delta_{y,\tau_{n+1}}$ can not be less than $\tau_{n+1}/(l+1)$ and can not be greater than $(l+\tau_{n+1})/(l+1)$, we consider the case in which $\alpha \in [\tau_{n+1}/(l+1), (l+\tau_{n+1})/(l+1))$. Let us define w the $\lceil l + \tau_{n+1} - (l+1)\alpha \rceil$th smallest value in the set $\{R_h : h \in \mathcal{I}_2\}$, and r_n (v_n respectively) the number of elements in the set $\{R_h : h \in \mathcal{I}_2\}$ that are equal to w and that are to the right (left respectively) of w in the sorted version of the set. Under the assumption concerning the continuous joint distribution of $\{R_h : h \in \mathcal{I}_2\}$ made in Section 2.2 $r_n = v_n = 0$ holds, but generally speaking we assume $r_n, v_n \in \mathcal{N}_{\geq 0}$ such that $r_n + v_n \leq l - 1$. By performing calculations similar to those needed in the non-randomized scenario, we obtain that:

- if

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

  then $y \in \mathcal{C}_{n,1-\alpha,\tau_{n+1}} \iff R_{n+1} \leq w$ and so

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

- if

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

then $y \in \mathcal{C}_{n,1-\alpha,\tau_{n+1}} \iff R_{n+1} < w$ and so

$$\mathcal{C}_{n,1-\alpha,\tau_{n+1}} = \{y \in \mathcal{Y}(\mathcal{T}) : y(t) \in \big(g_{\mathcal{I}_1}(t) - w, \\ g_{\mathcal{I}_1}(t) + w\big) \quad \forall t \in \mathcal{T}\}.$$

Also the introduction of the modulation function presented in Section 2.4 can be easily generalized in the smoothed conformal context. Let us define for a given $y \in \mathcal{Y}(\mathcal{T})$

$$\delta^s_{y,\tau_{n+1}} := \frac{\left|\left\{j \in \mathcal{I}_2 : R^s_j > R^s_{n+1}\right\}\right| + \tau_{n+1}\left|\left\{j \in \mathcal{I}_2 \cup \{n+1\} : R^s_j = R^s_{n+1}\right\}\right|}{l+1}$$

$$\mathcal{C}^s_{n,1-\alpha,\tau_{n+1}} := \left\{y \in \mathcal{Y}(\mathcal{T}) : \delta^s_{y,\tau_{n+1}} > \alpha\right\}.$$

By reconsidering the previous computations and by substituting $\delta_{y,\tau_{n+1}}$ with $\delta^s_{y,\tau_{n+1}}$, w with w^s, R_h with R^s_h, r_n with r^s_n and v_n with v^s_n it is possible to notice that

- if

$$\tau_{n+1} > \frac{(l+1)\alpha - \lfloor (l+1)\alpha - \tau_{n+1} \rfloor + r^s_n}{r^s_n + v^s_n + 2}$$

then

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

- if

$$\tau_{n+1} \leq \frac{(l+1)\alpha - \lfloor (l+1)\alpha - \tau_{n+1} \rfloor + r^s_n}{r^s_n + v^s_n + 2}$$

then

$$\mathcal{C}^s_{n,1-\alpha,\tau_{n+1}} = \{y \in \mathcal{Y}(\mathcal{T}) : y(t) \in \big(g_{\mathcal{I}_1}(t) - w^s s_{\mathcal{I}_1}(t), \\ g_{\mathcal{I}_1}(t) + w^s s_{\mathcal{I}_1}(t)\big) \quad \forall t \in \mathcal{T}\}.$$

Proof of Remark 10.

The functions $\bar{s}^c_{\mathcal{I}_1}$ and $\bar{s}_{\mathcal{I}_1}$ are defined as in Section 2.4 except for k (γ respectively) that is the $\lceil l + \tau_{n+1} - (l+1)\alpha \rceil$th ($\lceil m + \tau_{n+1} - (m+1)\alpha \rceil$th respectively) smallest value in the corresponding set; similarly, if $\lceil m + \tau_{n+1} - (m+1)\alpha \rceil > m$ then $\mathcal{H}_1 = \mathcal{I}_1$ and if $\lceil m + \tau_{n+1} - (m+1)\alpha \rceil \leq 0$ we arbitrarily set $\bar{s}_{\mathcal{I}_1} = s^0$. The theorems of Section 2.4 still hold by substituting $\lceil (l+1)(1-\alpha) \rceil, \lceil (m+1)(1-\alpha) \rceil$ with $\lceil l + \tau_{n+1} - (l+1)\alpha \rceil, \lceil m + \tau_{n+1} - (m+1)\alpha \rceil$.

S2 Simulation Study

S2.1 Study Design

In this section, we summarize the results of a two-stage simulation study comparing our approach with four alternative methods from the literature that will be detailed in the following: Naive, Band Depth, Modified Band Depth, Extremal depth and Bootstrap. In Section S2.2 the empirical coverage is evaluated for each approach in three different scenarios, whereas in Section S2.3 the prediction bands obtained by the methods that guarantee a proper coverage are compared in terms of efficiency. The simulation study has been mainly performed in the R programming language using the `conformalInference.fd` package (Diquigiovanni et al. 2022). The code to reproduce the simulations and the analyses of the test case is available upon request to the authors. The hierarchical structure of the simulation study reflects the "nested" nature of the two features we are considering, i.e. coverage and size: indeed, the size of a prediction set should be investigated only after verifying that the method which outputted that specific prediction set guarantees the desired coverage, which represents the primary aspect when assessing prediction sets.

Specifically, the three scenarios allow to compare the methods in three different frameworks:

when data show a constant variability over the domain (Scenario 1), when data show a different variability over the domain (Scenario 2) and when data are characterized by outliers (Scenario 3). The data generating processes of the three scenarios are:

- Scenario 1. $\forall i = 1, \ldots, n$

$$y_i(t) = x_{i1} + x_{i2}\cos(6\pi(t + u_i)) + x_{i3}\sin(6\pi(t + u_i))$$

with $\mathcal{T} = [0, 1]$, $(x_{11}, x_{12}, x_{13})^T, \ldots, (x_{n1}, x_{n2}, x_{n3})^T$ i.i.d. realizations of

$$X \sim N_3\left(\mathbf{0}, \begin{bmatrix} 1 & 0.6 & 0.6 \\ 0.6 & 1 & 0.6 \\ 0.6 & 0.6 & 1 \end{bmatrix}\right)$$

and $u_1, \ldots, u_n$ i.i.d. realizations of

$$U \sim \text{Unif}\left[-\frac{1}{6}, \frac{1}{6}\right].$$

- Scenario 2. $\forall i = 1, \ldots, n$

$$y_i(t) = \sum_{j=1}^{13} c_{ij} B_j^{\omega}(t)$$

with $\mathcal{T} = [0, 1]$, $B_j^{\omega}(t)$ the b-spline basis system of order 4 with interior knots $\omega = (0.1, 0.2, \ldots, 0.9)$ and $(c_{1,1}, \ldots, c_{1,13})^T, \ldots, (c_{n,1}, \ldots, c_{n,13})^T$ i.i.d. realizations of $C = (C_1, \ldots, C_{13}) \sim N_{13}(\mathbf{0}, \Sigma)$ such that $\text{Var}[C_i] = 0.03^2$ $\forall i \neq 7$, $\text{Var}[C_7] = 0.003^2$ and $\text{Cov}[C_i, C_j] = 0$ for $i, j = 1, \ldots, 13$, $i \neq j$.

- Scenario 3. The scenario is the previous one after contamination with outliers. Formally, $(c_{1,1}, \ldots, c_{1,13})^T, \ldots, (c_{n,1}, \ldots, c_{n,13})^T$ are i.i.d. realizations of a vector random variable whose probability density function is a Gaussian mixture density with weights $(1-\beta, \beta)$, shared mean vector $\mathbf{0}$, the covariance matrix defined as in Scenario 2 for the first group and such that $\text{Var}[C_7] = 0.3^2$ instead of $\text{Var}[C_7] = 0.003^2$ for the second group.

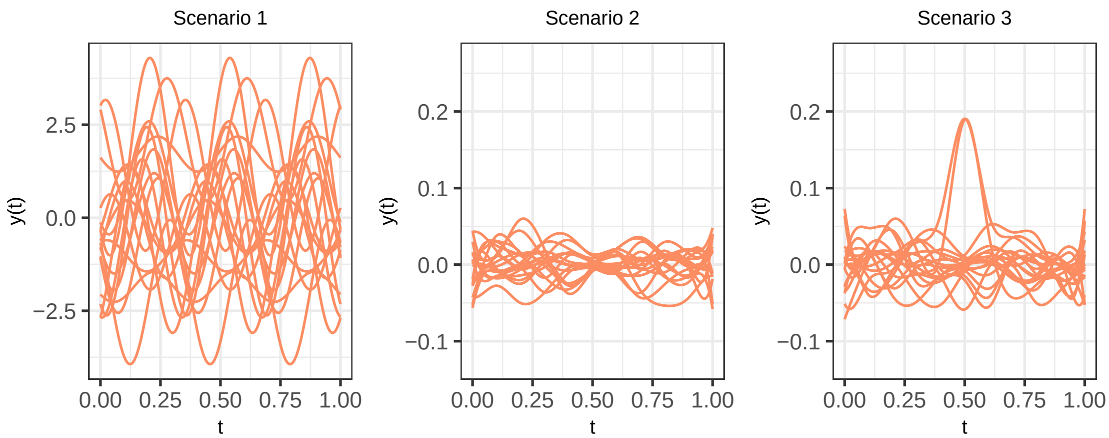

Figure S2.1: Graphical representation of the scenarios. The sample size is $n = 18$.

A graphical representation of a replication for each scenario with $n = 18$ is provided in Figure S2.1. The Conformal approach presented in Section 2 is evaluated in the non-smoothed framework and considering three different modulation functions: s^0, the normalised pointwise standard deviation function $s^{\sigma}_{\mathcal{I}_1}$ as natural representative of functions that capture data variability, and $\bar{s}_{\mathcal{I}_1}$. Since the focus of the work is not on the construction of sophisticated point predictors $g_{\mathcal{I}_1}$ but rather on the construction of valid prediction bands around any point predictor $g_{\mathcal{I}_1}$, we hereby simply set $g_{\mathcal{I}_1}(t) = \bar{y}_{\mathcal{I}_1}(t)$.

The performance of our approach is compared to four alternative methods. These are: *Naive* method, which outputs prediction bands defined as $\{y \in \mathcal{Y}(\mathcal{T}) : y(t) \in [q_{\frac{\alpha}{2}}(t), q_{1-\frac{\alpha}{2}}(t)]$ $\forall t \in \mathcal{T}\}$ with $q_{\alpha}(t)$ empirical quantile of order α for $(y_1(t), \dots, y_n(t))$. Such approach represents a very naive solution to the prediction task we are considering and we expect it to suffer greatly from undercoverage; *BD* and *MBD* methods, which output the sample $(1-\alpha)$ central region induced by the band depth (BD) and the modified band depth (MBD) respectively (Sun & Genton 2011); *Extremal* which output the sample $(1-\alpha)$ central region induced by the extremal depth (Narisetty & Nair 2016); *Boot.* method, which outputs the band based on 2500 bootstrap samples, as proposed by Degras (2011). We consider $\alpha = 0.1$, $\beta = 0.06$ and three different sample sizes: $n = 18$, $n = 198$, $n = 1998$. In order not to overcomplicate the simulation study, the ratio $\rho = l/n$ is kept fixed and equal to 0.5 as commonly suggested in the Conformal literature. A deeper investigation about the possible effect of the ratio $\rho = l/n$ on efficiency - even though possibly interesting - is out of the scope of this work. The atypical values of n in the simulations have been simply chosen to have a miscoverage exactly equal to α (indeed in these cases $\lfloor (l+1)\alpha \rfloor/(l+1) = \alpha$) and consequently making the simulation results easier to read. Similar results would have been attained with rounded values of n (e.g. $n = 20$, $n = 200$, $n = 2000$) by evaluating the empirical miscoverage considering the theoretical one: $\lfloor (l+1)\alpha \rfloor/(l+1)$ (see Theorem 1). The simulations are achieved by using the R Programming Language (R Core Team 2018) and the computation of the band depth and the modified band depth by *roahd* package (Tarabelloni et al. 2018). Finally, every combination of scenario and sample size is evaluated considering $N = 500$ replications.

S2.2 Coverage

In this section we focus on the sample mean and the standard deviation of the empirical conditional coverage provided by the prediction bands generated by each method for each combination of sample size and scenario (see Table S2.1). Specifically, the empirical conditional coverage of a given prediction band (i.e. the empirical coverage obtained conditioning on the prediction band obtained by the observed data) is computed as the fraction of times that 10,000 new functions - independent from and identically distributed to the original sample - belong to such prediction band. The purpose of this scheme is twofold: first of all, by averaging the $N = 500$ empirical conditional coverages obtained for each combination of scenario and sample size it is possible to

		Conformal Method			Alternative Methods				
		s^0	$s^{\sigma}_{\mathcal{I}_1}$	$\bar{s}_{\mathcal{I}_1}$	Naive	MBD	BD	Ext. Depth	Boot.
$n = 18$	Scenario 1	0.902	0.900	0.900	0.409	0.504	0.547	0.498	0.875
		(0.088)	(0.085)	(0.087)	(0.092)	(0.109)	(0.111)	(0.107)	(0.064)
	Scenario 2	0.901	0.910	0.909	0.048	0.123	0.145	0.119	0.922
		(0.089)	(0.081)	(0.083)	(0.021)	(0.044)	(0.051)	(0.042)	(0.042)
	Scenario 3	0.904	0.904	0.907	0.049	0.124	0.148	0.122	0.932
		(0.084)	(0.089)	(0.085)	(0.023)	(0.049)	(0.055)	(0.048)	(0.061)
$n = 198$	Scenario 1	0.901	0.902	0.901	0.625	0.861	0.900	0.826	0.865
		(0.029)	(0.030)	(0.031)	(0.031)	(0.028)	(0.028)	(0.028)	(0.019)
	Scenario 2	0.901	0.899	0.900	0.189	0.733	0.788	0.678	0.897
		(0.029)	(0.031)	(0.029)	(0.019)	(0.036)	(0.032)	(0.033)	(0.015)
	Scenario 3	0.897	0.900	0.899	0.197	0.742	0.798	0.688	0.892
		(0.031)	(0.030)	(0.031)	(0.020)	(0.034)	(0.030)	(0.033)	(0.020)
$n = 1998$	Scenario 1	0.900	0.899	0.900	0.666	0.942	0.918	0.885	0.866
		(0.010)	(0.010)	(0.010)	(0.011)	(0.006)	(0.008)	(0.008)	(0.008)
	Scenario 2	0.900	0.900	0.899	0.233	0.958	0.971	0.858	0.899
		(0.009)	(0.010)	(0.010)	(0.007)	(0.006)	(0.005)	(0.008)	(0.008)
	Scenario 3	0.900	0.899	0.900	0.240	0.959	0.973	0.859	0.884
		(0.010)	(0.010)	(0.010)	(0.008)	(0.006)	(0.005)	(0.008)	(0.007)

Table S2.1: For each combination of sample size and scenario, the first line shows the sample mean of the empirical conditional coverage, the second line the sample standard deviation in brackets.

obtain the empirical coverage, which is an estimate of the (unconditional) coverage. Secondly, this scheme allows to evaluate the variability of the conditional coverage when the observed sample varies, a particularly useful indication in real applications.

The simulation study fully confirms the theoretical property concerning the validity of split conformal prediction sets with 53 out of the 54 99%-confidence intervals associated to conformal bands including the nominal value $1-\alpha$. The evidence provided is particularly appealing since the desired coverage is guaranteed also when a very small sample size ($n = 18$) is considered, a framework in which such property is traditionally hard to obtain. Vice versa, in almost all cases the alternative methods do not ensure the desired coverage with some estimates dramatically far from $1-\alpha$, especially for small sample sizes (i.e., $n = 18$). In view of this, in Section S2.3 only the efficiency of the Conformal methods is evaluated and compared.

S2.3 Efficiency

In this section the sample mean and the standard deviation of the size defined as in (2.7) of the prediction bands computed in the previous section are evaluated for each combination of modulation function, sample size and scenario (see Table S2.2). First of all, it is noticeable that when $n = 18$ the absence of modulation (i.e. s^0) seems to provide smaller prediction bands than those induced by $s^{\sigma}_{\mathcal{I}_1}$ and $\bar{s}_{\mathcal{I}_1}$, conceivably because the extremely low number of functions belonging to the training set ($m = 9$) leads to an unstable and possibly misleading modulation function supporting the statistical intuition that for small sample sizes simpler modulation functions should be preferred.

More deeply, focusing now on each scenario separately and considering the remaining sample sizes, Scenario 1 represents a framework in which a constant width prediction band is the ideal candidate since the horizontal shift due to the random variable U induces constant variance along the domain. As a consequence, the pointwise evaluations $Y(t)$ are equally distributed $\forall t \in \mathcal{T}$ and so one is justified in expecting $s^{\sigma}_{\mathcal{I}_1}$ and $\bar{s}_{\mathcal{I}_1}$ to be of no practical use. The results confirm this conjecture, but the differences between the three modulation functions seems to

		s^0		$s^{\sigma}_{\mathcal{I}_1}$		$\bar{s}_{\mathcal{I}_1}$	
		Mean	*st.dev*	*Mean*	*st.dev*	*Mean*	*st.dev*
$n = 18$	Scenario 1	8.113	(2.044)	10.088	(3.618)	11.638	(4.309)
	Scenario 2	0.142	(0.025)	0.165	(0.041)	0.185	(0.049)
	Scenario 3	0.246	(0.192)	0.448	(0.550)	0.505	(0.633)
$n = 198$	Scenario 1	7.175	(0.560)	7.295	(0.608)	7.556	(0.647)
	Scenario 2	0.127	(0.006)	0.109	(0.005)	0.120	(0.006)
	Scenario 3	0.139	(0.013)	0.139	(0.013)	0.137	(0.020)
$n = 1998$	Scenario 1	7.059	(0.179)	7.065	(0.176)	7.128	(0.184)
	Scenario 2	0.125	(0.002)	0.106	(0.001)	0.117	(0.002)
	Scenario 3	0.136	(0.003)	0.137	(0.004)	0.131	(0.003)

Table S2.2: Size of the prediction bands.

decrease as the sample size grows (see, for example, the difference between s^0 and $\bar{s}_{\mathcal{I}_1}$ when n increases from 198 to 1998).

Scenario 2 represents a completely different setting, in which a modulation process is appropriate since the curves highlight a reduction of variability in the central part of the domain. As expected, s^0 induces larger predictions bands (on average) than those obtained by $s^{\sigma}_{\mathcal{I}_1}$ and $\bar{s}_{\mathcal{I}_1}$ and it forces the band to be unnecessary large around $t = 0.5$. On the other hand, the other two modulation functions (especially $s^{\sigma}_{\mathcal{I}_1}$) provide a better performance since they allow the band width to be adapted according to the behavior of data over $\mathcal{T}$.

Scenario 3 is obtained by contaminating Scenario 2 with outliers. Table S2.2 suggests that $\bar{s}_{\mathcal{I}_1}$ outperforms both s^0 and - unlike Scenario 2 - also $s^{\sigma}_{\mathcal{I}_1}$. In order to clarify this evidence, let

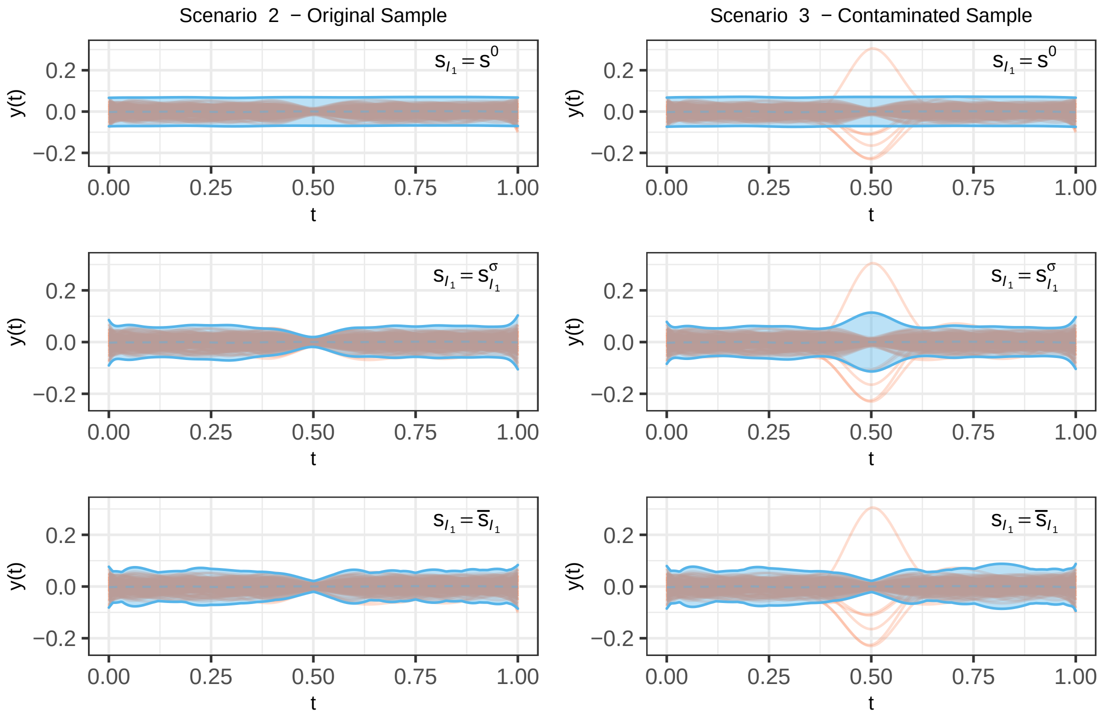

Figure S2.2: The prediction bands obtained considering a combination of modulation functions (s^0 at the top, $s^{\sigma}_{\mathcal{I}_1}$ in the middle, $\bar{s}_{\mathcal{I}_1}$ at the bottom) and sample (the original one on the left, the contaminated one on the right). In all cases, the dashed line represents $g_{\mathcal{I}_1}$.

us consider a sample $y_1, \ldots, y_{198}$ generated as in Scenario 2 that, after being created, is exposed to a contamination process in which each function $y_i, i = 1, \ldots, 198$, becomes an outlier as described in Scenario 3 with probability $\beta = 0.06$. Figure S2.2 shows examples of prediction bands induced by the three modulation functions (s^0 at the top, $s^{\sigma}_{\mathcal{I}_1}$ in the middle, $\bar{s}_{\mathcal{I}_1}$ at the bottom) obtained by considering the original sample (on the left) and the contaminated one (on the right). Moving from Scenario 2 to Scenario 3 and focusing on $s^{\sigma}_{\mathcal{I}_1}$, it is possible to notice that the increased variability in the central part of the domain due to the contamination process involves an increase in the band width around $t = 0.5$. This behavior, although not surprising, is counterproductive since the purpose of the method is to create prediction bands with coverage at the level $1 - \alpha = 0.9$ and in this specific case $\sim 94\%$ of the functions tends to be highly concentrated around $g_{\mathcal{I}_1}$ in the central part of the domain, and not overdispersed. By contrast, $\bar{s}_{\mathcal{I}_1}$ by construction removes the most extreme (in terms of measure (2.3)) functions and properly modulates data on the basis of the non-extreme functions keeping the band shape unchanged. From a methodological point of view, this is due to the dependency of $\bar{s}_{\mathcal{I}_1}$ on α which allows only a portion of the training set - chosen according to the specific level $1 - \alpha$ - to be taken into account and the trend of the "misleading" functions to be completely ignored. Overall, the evidence provided by this example - together with the results provided by Table S2.2 - suggests that s^0 is not affected by the contamination process (pro) but does not modulate (con), $s^{\sigma}_{\mathcal{I}_1}$ modulates (pro) but overreacts to the contamination process (con), whereas $\bar{s}_{\mathcal{I}_1}$ is able to simultaneously modulate (pro) and manage the contamination process (pro).

In short, the three scenarios seem to highlight that s^0 is an outstanding candidate when the sample size is very small, whereas a modulation process is useful in the very common case in which the variability over $\mathcal{T}$ varies and the sample size is either moderate or large. Specifically, $\bar{s}_{\mathcal{I}_1}$ provides encouraging results in some complex scenarios as it focuses on the specific behavior of the central (according to the level $1 - \alpha$) portion of data.

As an additional step, and to further evaluate the robustness of the proposed prediction method with respect to the use of different point forecasting methods, we inspect the sample mean and the standard deviation of the size, defined as in (2.7), of the prediction bands computed using different point predictors. Namely, we propose a table similar to S2.2, with the same declination in different sample sizes and different scenarios, but we explicitly explore different point prediction methods (S2.3. The explored methods are a baseline case, represented by the sample mean (stylised "*Mean*", already used as a candidate in all the previous simulations. The baseline is accompanied by two less standard cases, represented by a functional median case (stylised "*Median*"), where the point predictor is represented by the deepest curve of the sample, according to MBD. The third case is instead represented by a trimmed mean, computed excluding the 10% of the shallowest curves in the sample, again according to MBD. In all the simulations the modulation function selected is $\bar{s}_{\mathcal{I}_1}$.

In our specific simulation scenario, the use of more complex methods does not seem to be justified by a statistically significant increase in prediction performance, nevertheless deeper explorations of this important and relatively overlooked topic are in order.

		Mean		Median		Trimmed Mean 90%	
		Mean	*st.dev*	*Mean*	*st.dev*	*Mean*	*st.dev*
$n = 18$	Scenario 1	11.749	(4.458)	11.849	(4.269)	11.749	(4.458)
	Scenario 2	0.183	(0.046)	0.195	(0.050)	0.183	(0.046)
	Scenario 3	0.491	(0.605)	0.506	(0.604)	0.491	(0.605)
$n = 198$	Scenario 1	7.509	(0.648)	7.582	(0.663)	7.522	(0.660)
	Scenario 2	0.120	(0.006)	0.130	(0.007)	0.120	(0.006)
	Scenario 3	0.138	(0.023)	0.149	(0.023)	0.138	(0.026)
$n = 1998$	Scenario 1	7.134	(0.186)	7.144	(0.188)	7.133	(0.188)
	Scenario 2	0.117	(0.002)	0.128	(0.006)	0.117	(0.002)
	Scenario 3	0.131	(0.003)	0.143	(0.007)	0.131	(0.003)

Table S2.3: Size of the prediction bands for different point predictors, modulated using $\bar{s}_{\mathcal{I}_1}$

Bibliography

Degras, D. A. (2011), 'Simultaneous confidence bands for nonparametric regression with functional data', *Statist. Sinica* **21**(4), 1735–1765.

Diquigiovanni, J., Fontana, M., Solari, A., Vantini, S. & Vergottini, P. (2022), *conformalInference.fd: Tools for Conformal Inference for Regression in Multivariate Functional Setting*. R package version 1.1.1.
URL: *https://CRAN.R-project.org/package=conformalInference.fd*

Narisetty, N. N. & Nair, V. N. (2016), 'Extremal depth for functional data and applications', *Journal of the American Statistical Association* **111**(516), 1705–1714.
URL: *https://doi.org/10.1080/01621459.2015.1110033*

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Sun, Y. & Genton, M. G. (2011), 'Functional Boxplots', *J. Comput. Graph. Statist.* **20**(2), 316–334.

Tarabelloni, N., Arribas-Gil, A., Ieva, F., Paganoni, A. M. & Romo, J. (2018), *roahd: Robust Analysis of High Dimensional Data*. R package version 1.4.1.
URL: *https://CRAN.R-project.org/package=roahd*

Van der Vaart, A. W. (2000), *Asymptotic statistics*, Vol. 3, Cambridge university press.