On a definition of a functional depth through finite projections: a FPCA approach

Sara Armaut, Roland Diel, Thomas Laloë

ON A DEFINITION OF A FUNCTIONAL DEPTH THROUGH FINITE PROJECTIONS: A FPCA APPROACH

SARA ARMAUT & ROLAND DIEL & THOMAS LALOË

Laboratoire J.A. Dieudonné, UMR CNRS 7351,
Université Côte d'Azur, Parc Valrose,
06108 Nice, Cedex 2, France.
E-mail address: armaut@unice.fr, diel@unice.fr, laloe@unice.fr

Abstract. In this paper, we present a new functional depth called *Principal Component functional Depth* (PCD) for square-integrable processes X over a compact set. This depth involves a generic multivariate depth function which is evaluated at the projection of the function on the basis formed by the first J vectors of the Karhunen-Loève decomposition of X. We first investigate whether our Principal Component Decomposition (PCD) satisfies the desirable properties of statistical functional depths, following the axiomatization of [21]. In a second step, we present a consistent estimator for our PCD and establish its uniform consistency with a convergence rate. Finally, we complement our study with simulations and various real-world applications in supervised classification, demonstrating that our maximum PCD classifier equals or outperforms conventional competitors.

Contents

Key words and phrases. Depth function; FPCA; functional data, Karhunen-Loève decomposition.

Introduction

In multivariate data analysis, a depth function measures the extent to which a point is 'central' in a given d-variate data cloud, $d \geq 1$. Different notions of data depth in $\mathbb{R}^d$ have been proposed as tools generalizing those of ranks, orderings, outliers and quantiles to the multivariate setup [35, 51, 56]. A formal definition of statistical depth function has been provided by Zuo and Serfling [56]: they have list four desirable properties that a d-variate depth should satisfy, namely: affine-invariance, maximality at center provided that this center exists, monotonicity on rays and vanishing at infinity. For a general account on depth in the widely studied setup of finite-dimensional spaces, see [56] and the references therein.

For functional data, that is, when the data lies in an infinite-dimensional space such as a functions space, it is more delicate to determine a general definition of the classical multivariate depth and to establish which properties a functional depth should satisfy. In a recent paper, Nieto-Reyes and Battey [44] provide a general definition for a statistical depth function for functional data. This first formal definition has been revised by Gijbels and Nagy [21]. In this work [21], the authors present an extensive update of the survey of [44], in which they establish adaptations of the depth properties and prove that they are more easily met by common functional depths.

In the present contribution, we provide a new definition of a specific functional depth called *Principal Component Functional Depth* based on functional principal component analysis (FPCA). It uses projections to reduce the problem to a multivariate one. These finite projections are obtained by representing the data in terms of the eigenfunctions of the covariance operator and the associated functional principal components (FPCs) as usually done in FPCA.

Let us point out that FPCA is a field of study that has proved to be a significant technique in the development of functional data analysis. Indeed, Jones & Rice [30] presented an interesting proposition to characterize samples of random curves through principal component scores. Next, [29] built up an extension of FPCA which aimed at estimating harmonics from fragments of curves. Further, [54] estimated functional principal component scores for sparse longitudinal data by developping a FPCA procedure via a conditional expectation approach. More recently, [27] suggested graphical tools in order to vizualise functional data and detect functional outliers. References, along with theoretical and algorithmical details, can be found in the work of [53, 50, 16].

On the other hand, we further study our PCD depth in combination with the basic analysis of depth properties and uniform consistency results. For the latter, we upstream provide a natural estimator of PCD, and its consistency is unsurprisingly linked to that of the estimated covariance operator, hence that of the principal values and principal factors, as well as of the involved multivariate depth. Originally, asymptotic distributions have already been derived by [12] for the eigenvalues and eigenfunctions, as well as for the empirical projectors onto any eigenspace. In recent years, [23] developed asymptotic Taylor expansions of the estimated eigenfunctions in terms of the operator norm error between the theoretical and estimated covariance operator. Further, [22] investigated the properties of FPCA and provided insights into methodology and convergence speed-rates for FPCA. Also, when the sample curves are assumed to be observed, [26] gave explicit convergence rates for the differences between estimated and true eigenelements. For more detailed limiting results upon the FPCs estimators and associated eigenelements, see [12, 3, 2, 46] and references therein.

Among the wide functional applications and those stated above, an interesting one is to compare the FPCs characterizing two given samples. Such a case study may be seen as a functional generalization of the concept of "common functional principal components" as

introduced by Flury [17] in multivariate analysis. This is notably of considerable importance in implied volatility dynamics. Implied volatility is obtained from the pricing model introduced by Black and Scholes [5] and is a key factor for quoting options prices. Such an application is discussed in [2] over two log-returns datasets. In this work, they also show that $J = 3$ FPCs suffice to explain 98.2 % of the variability of the sample functions. This is one example where usefulness of low-dimensional components is exhibited and hence our interest in defining a functional depth through J finite projections using a generic J-variate depth.

The interest of our FPCA-approach lies first in its ability to capture maximal information from the underlying process while projecting it onto a finite set of elements. Moreover, our depth estimator has demonstrated adaptability in functional classification problems, through the selection of the parameter J via cross-validation. Indeed, this adaptability has proven to enhance performance in practical scenarios when compared to existing maximum-depth classifiers and other conventional ones in the literature. This study is detailed in Section 3.

The paper is organized as follows. In Section 1, we recall the properties of multivariate and functional statistical depths following the definitions introduced by Gijbels and Nagy [21], then Section 2 is devoted to our main contribution: Subsection 2.1 introduces the *Principal Component Depth* (PCD) for functional data, in Subsection 2.2 we check that the PCD satisfies the classical properties of functional depth, while Subsection 2.3 is dedicated to the construction of our depth estimator along with the study of its asymptotic behavior in a general setting. Illustrations and simulations are presented in Section 3 along with a real world application. Section 4 gives a conclusion and some perspectives.

1. Preliminary: multivariate and functional data depths

For the sake of completeness, we further recall the definition of statistical functional depth as introduced by [21] below. This definition itself is a generalization to the functional case of the multivariate definition proposed in [56].

Fix a compact subset $\mathcal{V}$ of $\mathbb{R}^d$ and consider a space $\mathscr{F}$ of functions $x : \mathcal{V} \to \mathbb{R}$ endowed with a norm $\|\cdot\|$ and its associated metric d (for example, $(\mathscr{C}(\mathcal{V}), \|\cdot\|_\infty)$ or $(\mathbb{L}^2(\mathcal{V}), \|\cdot\|_2)$). Denote by $\mathscr{P} := \mathcal{P}(\mathscr{F})$ the set of Borel probability measures on $\mathscr{F}$ and, for a random variable X on $\mathscr{F}$, denote by P_X its probability distribution. Remark that the following definition encompasses the multivariate case ([56]) by choosing a finite set $\mathcal{V}$.

Definition 1.1. *A mapping $D : \mathscr{F} \times \mathscr{P} \to \mathbb{R}$ is a* statistical functional depth *if it satisfies the following properties **P-1** to **P-6**:*

P-1: ***Distance invariance.***
Let $f : \mathscr{F} \to \mathscr{F}$ be a mapping s.t. for any $x, y \in \mathscr{F}$, $d(f(x), f(y)) = a_f d(x, y)$, for some fixed $a_f > 0$. Then, for any random variable X on $\mathscr{F}$ and any $x \in \mathscr{F}$, $D(f(x), P_{f(X)}) = D(x, P_X)$.

P-2: ***Maximality at center.***
For any distribution $P \in \mathscr{P}$ possessing a unique center of symmetry $\theta \in \mathscr{F}$ w.r.t some notion of symmetry, $D(\theta, P) = \sup_{x\in\mathscr{F}} D(x, P)$.

P-3: ***Decreasing w.r.t the deepest point.***
For any $P \in \mathscr{P}$ and $z \in \mathscr{F}$ s.t. $D(z, P) = \sup_{x\in\mathscr{F}} D(x, P)$ we have that $D(z, P) > \inf_{x\in\mathscr{F}} D(x, P)$ and for all $x \in \mathscr{F}$ and $\alpha \in [0, 1]$, $D(x, P) \leq D(z + \alpha(x - z), P)$.

P-4: ***Vanishing at infinity.***
For any $P \in \mathscr{P}$,ve $\lim_{\|x\|\to\infty} D(x, P) = \inf_{x\in\mathscr{F}} D(x, P)$.

P-5: ***Upper semi-continuity in x.***

Let $P \in \mathscr{P}$ and $x \in \mathscr{F}$. For any $\varepsilon > 0$, there exists $\delta > 0$ s.t.

$$\sup_{\substack{y \in \mathscr{F} \\ d(x,y)<\delta}} D(y,P) \leq D(x,P) + \varepsilon.$$

P-6: ***Continuity in P.***

Let $P \in \mathscr{P}$ and $x \in \mathscr{F}$. For any $\varepsilon > 0$, there exists $\delta > 0$ s.t.

$$P\text{-a.s. for all } Q \in \mathscr{P},\ d_{\mathscr{P}}(P,Q) < \delta \Longrightarrow |D(x,P) - D(x,Q)| < \varepsilon,$$

where $d_{\mathscr{P}}$ is a metric associated to the topology of weak convergence.

In practice, depth functions rarely satisfy all the properties stated in the definition; Property **P-2**, in particular, depends on the chosen notion of symmetry. However, this remains an objective for any new proposal of a depth function. We comment below on these different properties and present alternatives for some.

Understanding the importance of property **P-1** and its relation to its multivariate counterpart, led [38] and [9] to distinguish two different further weaker properties of affine invariance:

P-1S: Scalar-affine invariance.

Let $f : \mathscr{F} \to \mathscr{F}$ be a scalar-affine mapping, that is, for some $b \in \mathscr{F}$ and $a \in \mathbb{R}\backslash\{0\}$,

$$\forall x \in \mathscr{F},\ f(x) = ax + b\,.$$

Then, $D(f(x), P_{f(X)}) = D(x, P_X)$ for any random variable X and $x \in \mathscr{F}$.

P-1F: Function-affine invariance.

Let $f : \mathscr{F} \to \mathscr{F}$ be a function-affine mapping, that is, for some $b \in \mathscr{F}$ and $a \in \mathscr{F}$ where $a(t) \neq 0$ for all $t \in \mathcal{V}$ and $ax \in \mathscr{F}$

$$\forall x \in \mathscr{F},\ f(x) = ax + b\,.$$

Then, $D(f(x), P_{f(X)}) = D(x, P_X)$ for any random variable X and $x \in \mathscr{F}$.

Note that, in **P-1S** and **P-1F**, point-wise product of functions is intended in the sense that $f(x) = ax + b$ means $f(x)(t) = ax(t) + b(t)$ or $f(x)(t) = a(t)x(t) + b(t)$ depending on the property at hand.

As far as Property **P-2** is concerned, it can be more meaningful to consider the behaviour of D for particular P for which many notions of centre of symmetry coincide at θ. In $\mathbb{R}^d$, such P are the zero-mean Gaussian distributions, for which the mean is a center of central and halfspace symmetry (e.g. [56]). This inspection is intuitively extended to the functional setting with the Gaussian process in mind.

P-2G: Maximality at Gaussian process mean.

For P a zero-mean, stationary, almost-surely continuous Gaussian process on $\mathcal{V}$, $D(\theta, P) = \sup_{x \in \mathscr{F}} D(x,P) \neq \inf_{x \in \mathscr{F}} D(x,P)$, where the function $\theta \equiv 0$ over $\mathcal{V}$.

Having at hand a reasonable notion of symmetry for random functions, a stronger alternative than **P-2G** is proposed in Lemma 1 of [21]:

P-2C: Maximality at center (central symmetry).

For any centrally symmetric $P \in \mathscr{P}$, we have that $D(\theta, P) = \sup_{x \in \mathscr{F}} D(x,P)$ if and only if P is centrally symmetric around $\theta \in \mathscr{F}$ in the sense that $X - \theta = -(X - \theta)$ in distribution for $\mathscr{F} = \mathbb{L}^2[0,1]$ or $\mathcal{C}$.

Geometrically, **P-3** means that the upper level sets of the depth D are star-shaped (and thus connected) in $\mathscr{F}$ relative to the deepest point.

Of course, Property **P-4** is one natural functional counterpart of the vanishing at infinity property in the multivariate setting [21]. Regarding Property **P-5**, this is rather a technical

requirement and is the functional counterpart of multivariate upper semi-continuity, which is equivalent to stating that the upper level sets are closed.

Continuity of D in the distribution argument given in Property **P-6** is of great interest regardless of the precise nature of the application. Indeed, it first delivers almost sure convergence of the empirical depth to its population counterpart. Further, it tacitly addresses the intrinsic problem of partially observed functional data. The delicate challenge of P_n being inaccessible in its entirety entails reconstructing the functional data object by some interpolation or smoothing means. Then provided the reconstruction $\tilde{P}_n$ is such that $\tilde{P}_n \to P$, P-a.s., Property **P-6** ensures the desired consistency of the functional depth. Last but not least, the fulfilment of **P-6** produces a *qualitatively robust* (see Theorem 2, Section 2.2, in [25]) empirical depth and thus confirms that conclusions are not exceedingly affected by outliers.

Recently, [21] has proposed a uniform extension of Property **P-6** as follows:

P-6U: Uniform continuity in P.

For any $\varepsilon > 0$, there exists $\delta > 0$ s.t.

$$\text{for any } P, Q \in \mathscr{P},\ \mathrm{d}_{\mathscr{P}}(P, Q) < \delta \Longrightarrow \sup_{x \in \mathscr{F}} |D(x, P) - D(x, Q)| < \varepsilon,$$

where we recall $\mathrm{d}_{\mathscr{P}}$ metricises the topology of weak convergence.

Already in $\mathbb{R}^d$, requirement **P-6U** is demanding in full generality. In addition, it is often too strict to assume uniform convergence over the whole space $\mathscr{F}$. But one appropriate weaker condition can be local uniformity over compact sets in $\mathscr{F}$.

Remark: when Property **P-1S** is satisfied, Property **P-6** implies the continuity of the depth in x (and therefore it implies Property **P-5**). Indeed, for any r.v. X and any $x, h \in \mathscr{F}$,

$$\|D(x + h, P_X) - D(x, P_X)\| = \|D(x, P_{X-h}) - D(x, P_X)\|.$$

As P_{X-h} converges weakly to P_X when $\|h\|$ goes to 0, this gives the continuity with respect to x. The same argument shows that **P-1S** and **P-6U** imply the uniform continuity of the depth in the variable x.

2. A projection-type functional depth

In the present section, we introduce a new type of functional depth using mainly two tools: the Karhunen-Loève (K-L) decomposition of a square-integrable stochastic process X [31] and a multivariate depth function D_J defined on $\mathbb{R}^J$, $J \geq 1$, where J is the number of principal components to be chosen. This will be discussed further in the paper. The K-L expansion is achieved analytically by projecting the process, considered over a compact time interval, onto a deterministic orthonormal basis obtained from the covariance Hilbert-Schmidt operator's eigenfunctions, which correspond to positive eigenvalues.

2.1. The Principal Component Functional Depth.

We consider $\left(\mathbb{L}^2[0,1], \langle \cdot, \cdot \rangle\right)$ the Hilbert space of square integrable functions over $[0,1]$ w.r.t. Lebesgue measure on $\mathbb{R}$. This space is equipped with the inner product

$$\langle f, g \rangle = \int_0^1 f(t) g(t) dt, \quad f, g \in \mathbb{L}^2[0,1]\,,$$

which associated norm is given by

$$\|f\| = \int_0^1 f(t)^2 dt, \quad f \in \mathbb{L}^2[0,1]\,.$$

We denote by $\mathcal{P}_2(\mathbb{L}^2[0,1])$ the set of probability measures on $\mathbb{L}^2[0,1]$ such that a random process $(X(t))_{t \in [0,1]}$ with probability measure $P_X \in \mathcal{P}_2(\mathbb{L}^2[0,1])$ satisfies the following conditions:

- the process X is square-integrable: for any $t \in [0,1]$, $X(t) \in \mathbb{L}^2(\Omega, \mathscr{A}, \mathbb{P})$
- its covariance function K_X, defined by

$$\forall s,t \in [0,1], K_X(s,t) = \mathbb{E}\big[X(s)X(t)\big] - \mathbb{E}\big[X(s)\big]\mathbb{E}\big[X(t)\big],$$

  is continuous.

We will also denote by $m_X(t) := \mathbb{E}[X(t)],\ t \in [0,1]$ the mean function of X. The function K_X above then defines a linear operator $\mathcal{K}_X$ in the following way:

$$\begin{array}{rccl} \mathcal{K}_X & : & \mathbb{L}^2[0,1] & \to & \mathbb{L}^2[0,1] \\ & & f & \mapsto & \mathcal{K}_X f(\cdot) := \displaystyle\int_0^1 K_X(\cdot,s) f(s) ds. \end{array}$$

Theorem 2.1 (Karhunen-Loève theorem [31])**.**
Consider a zero-mean random process $(X(t))_{t\in[0,1]}$ such that $P_X \in \mathcal{P}_2(\mathbb{L}^2[0,1])$. There exists $(e_j)_{j\geq 1}$ an orthonormal basis on $\mathbb{L}^2[0,1]$ formed by eigenfunctions of $\mathcal{K}_X$ with respective eigenvalues $(\lambda_j)_{j\geq 1}$. And the process $X(t)$ admits the following representation

$$X(t) = \sum_{j=1}^{\infty} Z_j e_j(t).$$

where, for all $j \geq 1$,

$$Z_j = \int_0^1 X(t) e_j(t) dt = \langle X, e_j \rangle .$$

Moreover, the r.v. Z_j have zero-mean, are uncorrelated and have variance λ_j: for any $i,j \geq 1$,

$$\mathbb{E}[Z_j] = 0, \quad \textit{and} \quad \operatorname{cov}(Z_i, Z_j) = \mathbb{E}[Z_i Z_j] = \lambda_j \delta_{ij}.$$

One of the main interests of the K-L decomposition is that the first J principal components provide a "best basis" for approximating the process in terms of the total mean squared error:

$$X^J := \sum_{j=1}^{J} Z_j e_j = \underset{\substack{\mathrm{Y}^J \text{ orth. projection of } X \text{ over} \\ \text{a subspace of dimension } J \text{ of } \mathbb{L}^2[0,1]}}{\arg\min} \mathbb{E}\big[\|X - \mathrm{Y}^J\|^2_{\mathbb{L}^2[0,1]}\big].$$

Notice that if there exists $j_0 \geq 1$ such that $\lambda_{j_0} = \lambda_{j_0+1}$, meaning λ_{j_0} is of multiplicity two, then the choice of the basis $(e_j)_{j\geq 1}$ is not unique even up to a sign.

This theorem lets us to introduce the notion of principal component functional depth.

In order to ensure readability, we start with some notations. For any $x \in \mathbb{L}^2[0,1]$ and some integer $J \geq 1$, we set

$$\mathbf{x}^{(J)} := (\langle x - m_X, e_j \rangle)_{1\leq j\leq J} \in \mathbb{R}^J , \tag{2.1}$$

$$x^J := m_X + \sum_{j=1}^{J} \langle x - m_X, e_j \rangle e_j \in \mathbb{L}^2[0,1] . \tag{2.2}$$

The parameter J denotes the maximal number of components to be considered in the basis.

Definition 2.2 (Principal Component functional depth)**.** *Fix an integer $J \geq 1$ and D_J a J-variate statistical depth function. Let X a random process with distribution $P_X \in \mathcal{P}_2(\mathbb{L}^2[0,1])$. Based on a K-L decomposition $(\lambda_j, e_j)_{j\geq 1}$ of the centered process $X - m_X$, we define the principal component-functional depth as follows:*

$$\forall x \in \mathbb{L}^2[0,1],\ \mathrm{PCD}^J(x, P_X) := \frac{D_J\big(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}\big)}{1 + \Big(\|x - x^J\|^2 / \sum_{i=1}^{\infty} \lambda_j\Big)^{\beta}} \tag{2.3}$$

where $\mathbf{x}^{(J)}$ *and* x^J *are defined in* (2.1) *and* (2.2) *respectively and* $P_{\mathbf{X}^{(J)}}$ *is the* J*-variate distribution of the centered random vector* $\mathbf{X}^{(J)} := (\langle X - m_X, e_1\rangle, \ldots, \langle X - m_X, e_J\rangle)$.

By construction, the $(\lambda_j)_{j\geq 1}$ decay to zero as $j \to \infty$ and are summable, that is $\sum_{j=1}^{\infty} \lambda_j < \infty$. Notice further that the covariance operator of the centered process $X - m_X$ is that of X since the covariance is translation-invariant. Another important observation is that the variance of $\|X - m_X\|$ is nothing but the sum of the eigenvalues of the covariance operator of X, that is :

$$\mathbb{E}[\|X - m_X\|^2] = \int_0^1 \mathrm{var}(X(t))dt = \sum_{j=1}^{\infty} \lambda_j \,.$$

In particular, as long as the choice of the fixed integer $J \geq 1$ is concerned, one can determine $J := J_\gamma$ such that the J-truncated expansion explains a proportion

$$\frac{\sum_{j=1}^{J} \lambda_j}{\sum_{j=1}^{\infty} \lambda_j} \geq 1 - \gamma$$

of the variance with $\gamma \in [0, 1]$ small.

The scalar

$$1 + \left(\|x - x^J\|^2 / (\sum_{i=1}^{\infty} \lambda_j)\right)^\beta = 1 + \left(\|x - x^J\|^2 / \mathbb{E}\big[\|X - m_X\|^2\big]\right)^\beta$$

is a "penalizing" coefficient in the sense that it decreases the value of the depth when x is far from its projection x^J compared to the typical fluctuation of $\|X - m_X\|$. Observe further that, for any $P \in \mathcal{P}_2(\mathbb{L}^2[0, 1])$ and any $x \in \mathbb{L}^2[0, 1]$, $\mathrm{PCD}^J(x, P) \leq \mathrm{PCD}^J(x^J, P)$ with equality if and only if $x = x^J$.

A first important property is that the expression of PCD^J above does not depend on the choice of the K-L basis $(e_j)_j$ whenever $\lambda_J > \lambda_{J+1}$ (which implies that the associated eigenspace is not "cut through") as soon as the J-variate depth D_J is isometry-invariant. Also, denoting by E_j the eigenspace associated to λ_j, we set

$$L_J := \dotplus_{j=1}^{J} E_j$$

as the sum eigenspaces. As $\lambda_J > \lambda_{J+1}$, L_J is of dimension J. And as only λ_J is supposed to be strictly larger than λ_{J+1}, for $j < J$, we can have $E_j = E_{j+1}$.

Proposition 2.3. *If* $\lambda_J > \lambda_{J+1}$ *and the multivariate depth* D_J *is invariant under isometry (Property* ***P-1****), then the principal component functional depth* PCD^J *(in Definition 2.2) does not depend on the choice of the K-L basis. Moreover, any orthonormal basis of* L_J*, even if not a K-L basis, will give the same result.*

Proof. Let $(e_j)_{j\geq 1}$ and $(\tilde{e}_j)_{j\geq 1}$ be two orthonormal basis of $\mathbb{L}^2[0, 1]$ such that

$$X(t) = \sum_{j=1}^{\infty} Z_j e_j(t) = \sum_{j=1}^{\infty} \tilde{Z}_j \tilde{e}_j(t),$$

such that $(e_j)_{j\leq J}$ and $(\tilde{e}_j)_{j\leq J}$ are basis of L_J. (As $\lambda_{J+1} < \lambda_J$, any K-L basis is such a basis.) Our aim is to show that

$$D_J\left(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}\right) = D_J\left(\tilde{\mathbf{x}}^{(J)}, P_{\tilde{\mathbf{X}}^{(J)}}\right),$$

The bases $(e_j)_{j\geq 1}$ and $(\tilde{e}_j)_{j\geq 1}$ are linked through some linear isometry, that is there exists a linear norm-preserving mapping $\Phi : \mathbb{L}^2[0,1] \to \mathbb{L}^2[0,1]$ such that $\tilde{e}_j = \Phi(e_j)$. As the restriction of the J first elements of the basis generates L_J, Φ induces an isometry on L_J. Then, it holds

$$x^{(J)} = m_X + \sum_{j=1}^{J} \langle x - m_X, e_j \rangle \, e_j = m_X + \sum_{j=1}^{J} \langle x - m_X, \tilde{e}_j \rangle \, \tilde{e}_j,$$

and with $\mathbf{x}^{(J)} := (\langle x - m_X, e_j \rangle)_{j=1,\dots,J} \in \mathbb{R}^J$ and $\tilde{\mathbf{x}}^{(J)} := (\langle x - m_X, \tilde{e}_j \rangle)_{j=1,\dots,J} \in \mathbb{R}^J$, the isometry-invariance of the multivariate depth D_J leads to

$$D_J\left(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}\right) = D_J\left(\tilde{\mathbf{x}}^{(J)}, P_{\tilde{\mathbf{X}}^{(J)}}\right),$$

which proves the proposition. □

2.2. **Properties of PCD depth.** We study here the functional depth properties presented in Section 1 satisfied by the PCD^J depth (2.3). These properties happen to be inherited from the classical multivariate depth properties of D_J. This is summed up in the following proposition as well as in Table 1.

Adherence of PCD^J to functional depth properties									
P-1	**P-1S**	**P-1F**	**P-2C**	**P-2G**	**P-3**	**P-4**	**P-5**	**P-6**	**P-6U**
✓	✓	✗	✓	✓	✓	✓	✓	✓*	✓**

✓* : if D_J satisfies **P-1**, **P-6** and Q is the empirical measure (Theorem 2.6)
✓** : if D_J satisfies **P-1**, **P-5**, **P-6U** and Q is the empirical measure (Theorem 2.6)

TABLE 1. Summary of the theoretical functional depths properties (**P-1**)-(**P-6U**) for PCD^J. Fulfilment or failure of the depth properties are respectively indicated by ✓ and ✗.

Proposition 2.4. *Let D_J be a J-variate depth and* PCD^J *be the associated principal component functional depth of Definition 2.2.*

(i) ***P-1****: If D_J is distance-invariant, then* PCD^J *depth is distance-invariant.*
(ii) ***P-2C****: If D_J meets multivariate maximality at center w.r.t. central symmetry, then* PCD^J *depth is maximal at the center w.r.t. central symmetry.*
(iii) ***P-3****: If D_J is decreasing w.r.t. deepest point, then* PCD^J *is decreasing w.r.t. deepest point.*
(iv) ***P-4****: If D_J is vanishing at infinity, then* PCD^J *is.*
(v) ***P-5****: If D_J is upper semi-continuous in its first argument, then* PCD^J *is.*
(vi) If D_J is distance-invariant and satisfies ***P-6*** *then* PCD^J *depth is continuous in the space variable x. And if D_J is scaling-invariant and satisfies* ***P-6U*** *then* PCD^J *depth is uniformly continuous in the space variable x.*

Proof.
(i) **P-1. Distance invariance.** Let f be a mapping s.t. $\|f(x) - f(y)\| = a_f \|x - y\|$, for some $a_f > 0$. This implies that f is of the form $f(x) = a_f g(x) + f(0)$, with g a linear isometry. For the sake of clarity, we denote in the current section $(\lambda_j^X, e_j^X)_j$ the eigenvalues and eigenfunctions of the process X. So, we have immediately, for any $j \geq 1$, $\lambda_j^{f(X)} = a_f^2 \lambda_j^{g(X)}$ and $e_j^{f(X)} = e_j^{g(X)}$. Next, using the K-L decomposition $X = m_X + \sum_{j=1}^{\infty} \left\langle X - m_X, e_j^X \right\rangle e_j^X$ and applying further the continuous linear isometry g to X, we obtain

$$g(X) = g(m_X) + \sum_{j=1}^{\infty} \langle X - m_X, e_j^X \rangle \, g(e_j^X) = g(m_X) + \sum_{j=1}^{\infty} \langle g(X - m_X), g(e_j^X) \rangle \, g(e_j^X).$$

Hence,$(g(e_j^X))_{j\geq 1}$ is a basis associated to the covariance operator of the process $g(X)$ in its K-L decomposition and we can choose $e_j^{f(X)} = e_j^{g(X)} = g(e_j^X)$ for any $j \geq 1$.

On one hand, remark that, when working under the distribution $P_{f(X)}$, one clearly has

$$\boldsymbol{f}(\mathbf{x})^{(J)} = \left\langle f(x) - m_{f(X)}, e_j^{f(X)} \right\rangle_j = a_f \left\langle g(x - m_X), g(e_j^X) \right\rangle_j = a_f \left\langle x - m_X, e_j^X \right\rangle_j \in \mathbb{R}^J, \tag{2.4}$$

where the last equality holds as g preserves the inner product. On the other hand, using again the linear isometry g along with some computations, it yields

$$\begin{aligned}\left(\frac{\|f(x) - (f(x))^J\|^2}{\sum_{j=1}^\infty \lambda_j^{f(X)}}\right)^\beta &= \frac{a_f^{2\beta}\|g(x) - g(m_X) - \sum_{j=1}^J \left\langle g(x - m_X), g(e_j^X)\right\rangle g(e_j^X)\|^{2\beta}}{\left(\sum_{i=1}^\infty a_f^2 \lambda_j^X\right)^\beta} \\ &= \frac{\|g(x) - g\left(m_X + \sum_{j=1}^J \left\langle x - m_X, e_j^X \right\rangle e_j^X\right)\|^{2\beta}}{\left(\sum_{i=1}^\infty \lambda_j^X\right)^\beta} \\ &= \left(\frac{\|x - x^J\|^2}{\sum_{i=1}^\infty \lambda_j^X}\right)^\beta.\end{aligned}$$

Eventually, all together with the fact that $P_{\boldsymbol{f}(\mathbf{X})^{(J)}} = P_{a_f \mathbf{X}^{(J)}}$ based on (2.4), we obtain, by scale-invariance of the multivariate depth,

$$D_J\left(\boldsymbol{f}(\mathbf{x})^{(J)}, P_{\boldsymbol{f}(\mathbf{X})^{(J)}}\right) = D_J(a_f \cdot \mathbf{x}^{(J)}, P_{a_f \cdot \mathbf{X}^{(J)}}) = D_J\left(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}\right).$$

Therefore, $\mathrm{PCD}^J(f(x), P_{f(X)}) = \mathrm{PCD}^J(x, P_X)$.

(ii) **P-2C. Maximality at center w.r.t central symmetry.** Let X be a centrally symmetric process w.r.t $\theta \in \mathbb{L}^2[0,1]$, that is $X - \theta = -(X - \theta)$ in distribution. This implies $m_X = \theta$ almost everywhere over $[0,1]$. In particular, $\theta = \theta^J$ and hence $\boldsymbol{\theta}^{(J)} = 0 \in \mathbb{R}^J$ in the definition of PCD^J. We want to show that

$$\mathrm{PCD}^J(\theta, P_X) = \sup_{x \in \mathbb{L}^2[0,1]} \mathrm{PCD}^J(x, P_X) \tag{2.5}$$

First, it is clear that $\mathrm{PCD}^J(\theta, P_X) = D_J(\boldsymbol{\theta}^{(J)}, P_{\mathbf{X}^{(J)}}) = D_J(0, P_{(\langle X - \theta, e_j\rangle)_{j=1,\dots,J}})$. Also, the right-hand side of (2.5) rewrites as

$$\begin{aligned}\sup_{x \in \mathbb{L}^2[0,1]} \mathrm{PCD}^J(x, P_X) &= \sup_{y \in \mathbb{L}^2[0,1]} D_J(\langle y, e_j\rangle_{j=1,\dots,J}, P_{(\langle X - \theta, e_j\rangle)_{j=1,\dots,J}}) \\ &= \sup_{z \in \mathbb{R}^J} D_J(z, P_{(\langle X - \theta, e_j\rangle)_{j=1,\dots,J}})\end{aligned}$$

as one may write $\mathbb{L}^2[0,1] = \mathrm{Span}(e_j)_{1\leq j\leq J} \oplus \mathrm{Span}(e_j)_{1\leq j\leq J}^\perp$, where $\mathrm{Span}(e_j)_{1\leq j\leq J}^\perp$ is the orthogonal space associated to $\mathrm{Span}(e_j)_{1\leq j\leq J}$, and the projection onto the orthogonal space $\mathrm{Span}(e_j)_{j=1,\dots,J}^\perp$ provides obviously null inner product w.r.t. e_j for any $j = 1, \dots J$.

Since $X - \theta = \theta - X$ in distribution, then the centered random vector $(\langle X - \theta, e_j\rangle)_{j=1,\dots,J} = -(\langle X - \theta, e_j\rangle)_{j=1,\dots,J}$ in distribution. Thereupon, from maximality at the center property of the multivariate depth D_J, it holds

$$D_J(0, P_{(\langle X - \theta, e_j\rangle)_{j=1,\dots,J}}) = \sup_{z \in \mathbb{R}^J} D_J(z, P_{(\langle X - \theta, e_j\rangle)_{j=1,\dots,J}}),$$

hence the desired result (2.5).

(iii) **P-3. Monotonicity w.r.t. deepest point.** Let $\alpha \in [0,1]$, $P \in \mathscr{P}$ and $z \in \mathbb{L}^2[0,1]$ s.t. $\mathrm{PCD}^J(z,P) = \sup_{x\in\mathbb{L}^2[0,1]} \mathrm{PCD}^J(x,P)$. Letting $x \in \mathbb{L}^2[0,1]$, our goal is to show

$$\mathrm{PCD}^J(x,P) \leq \mathrm{PCD}^J(x+\alpha(z-x),P) := \mathrm{PCD}^J(x_\alpha,P). \tag{2.6}$$

Notice first that inequality (2.6) holds true if one shows that

$$\begin{aligned} &\|x_\alpha - x_\alpha^J\|^{2\beta} \leq \|x - x^J\|^{2\beta} \quad \text{and} \\ &D_J(\mathbf{x}_\alpha^{(J)}, P_{\mathbf{X}^{(J)}}) \geq D_J(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}), \end{aligned} \tag{2.7}$$

where we recall that $\mathbf{x}_\alpha^{(J)} = (\langle x_\alpha - m_X, e_j\rangle)_{j=1,\dots,J}$, $\mathbf{x}^{(J)} = (\langle x - m_X, e_j\rangle)_{j=1,\dots,J}$ as in (2.1) and x^J, x_α^J are defined accordingly in (2.2). By definition of z and PCD^J depth, one may write $z = m_X + z_0$ for some $z_0 \in \mathrm{Span}(e_1,\dots,e_J)$, where m_X is the mean function of the process with distribution P. From definition (2.2) , we have

$$\begin{aligned} x_\alpha - x_\alpha^J &= (1-\alpha)(x-m_X) + \alpha z_0 - \sum_{j=1}^J \langle (1-\alpha)(x-m_X) + \alpha z_0, e_j\rangle\, e_j \\ &= (1-\alpha)(x-x^J), \end{aligned}$$

where, in the before-last equality, we use the fact that $\sum_{j=1}^J \langle z_0, e_j\rangle\, e_j = z_0$ since $z_0 \in \mathrm{Span}(e_1,\dots,e_J)$. Hence, we obtain the first desired inequality in (2.7):

$$\|x_\alpha - x_\alpha^J\|^{2\beta} \leq \|x - x^J\|^{2\beta}.$$

Handling next the second equation in (2.7), we have

$$D_J(\mathbf{x}_\alpha^{(J)}, P_{\mathbf{X}^{(J)}}) = D_J(\alpha\, \langle z_0, e_j\rangle_j + (1-\alpha)\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}).$$

Therefore, it remains to show that

$$(\langle z_0, e_j\rangle)_{j=1,\dots,J} = \arg\sup_{y\in\mathbb{R}^J} D_J(y, P_{\mathbf{X}^{(J)}}),$$

so that one immediately derives the needed second inequality (2.7) from the multivariate monotonicity along rays property of D_J. Recall that $\mathrm{PCD}^J(z,P) = \sup_{x\in\mathbb{L}^2[0,1]} \mathrm{PCD}^J(x,P)$ and that $\sum_{j=1}^J \langle z_0, e_j\rangle\, e_j = z_0$, so

$$\begin{aligned} D_J(\langle z_0, e_j\rangle_j, P_{\mathbf{X}^{(J)}}) &= \sup_{x\in\mathbb{L}^2[0,1]} \frac{1}{1 + \left(\frac{\|x-x^J\|^2}{\sum_{j=1}^\infty \lambda_j}\right)^\beta} D_J(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}) \\ &= \sup_{x\in\mathrm{Span}(e_1,\dots,e_J)} D_J(\langle x - m_X, e_j\rangle_j, P_{\mathbf{X}^{(J)}}) \\ &= \sup_{y\in\mathbb{R}^J} D_J(y, P_{\mathbf{X}^{(J)}}) \end{aligned}$$

and the desired result (2.6) follows.

(iv) **P-4. Vanishing at infinity.** Fix some $\epsilon > 0$. Using boundedness of the multivariate depth D_J, we can find some $M_1 > 0$ such that if $\|x - x^J\|^2 > M_1$, $|\mathrm{PCD}^J(x,P)| \leq \epsilon$. And, since D_J is vanishing at infinity and $\left(1 + \left(\frac{\|x-x^J\|^2}{\sum_{j=1}^\infty \lambda_j}\right)^\beta\right)^{-1}$ is bounded, we can find some $M_2 > 0$ such that if $\|\mathbf{x}^{(J)}\|_J = \|x^J - m_X\| > M_2$, $\mathrm{PCD}^J(x,P) \leq \epsilon$. Hence, as for any $x \in \mathbb{L}^2[0;1]$, $\|x - x^J\|^2 + \|x^J\|^2 = \|x\|^2$, if $\|x\| > \sqrt{2}\max(M_1, M_2)$ then $|\mathrm{PCD}^J(x,P)| \leq \epsilon$ which gives the result.

(v) **P-5. Upper semi-continuity in x.** We show that for any function $x_0 \in \mathbb{L}^2[0,1]$,

$$\limsup_{\substack{x\to x_0 \\ \text{in } \mathbb{L}^2[0,1]}} \mathrm{PCD}^J(x,P) \leq \mathrm{PCD}^J(x_0,P). \tag{2.8}$$

As PCD^J is the product of two positive functions, it holds

$$\begin{aligned}\limsup_{\substack{x\to x_0 \\ \text{in } \mathbb{L}^2[0,1]}} \mathrm{PCD}^J(x,P) &= \limsup_{x\to x_0} \frac{1}{1+\left(\frac{\|x-x^J\|^2}{\sum_{j=1}^\infty \lambda_j}\right)^\beta} D_J(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}) \\ &\leq \limsup_{x\to x_0} \frac{1}{1+\left(\frac{\|x-x^J\|^2}{\sum_{j=1}^\infty \lambda_j}\right)^\beta} \limsup_{x\to x_0} D_J(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}).\end{aligned}$$

Hence, by continuity of the inner product and the associated norm, we obtain

$$\limsup_{x\to x_0} \mathrm{PCD}^J(x,P) \leq \frac{1}{1+\left(\frac{\|x_0-x_0^J\|^2}{\sum_{j=1}^\infty \lambda_j}\right)^\beta} \cdot \limsup_{\mathbf{x}^{(J)}\to \mathbf{x}_0^{(J)}} D_J(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}})$$

where $\mathbf{x}_0^{(J)} = \langle x_0 - m_X, e_j\rangle_j \in \mathbb{R}^J$. Under the assumption of upper semi-continuity of the multivariate depth one obtains directly

$$\limsup_{\mathbf{x}^{(J)}\to \mathbf{x}^{(J)}{}_0} D_J(\mathbf{x}^{(J)}, P_{\mathbf{X}^{(J)}}) \leq D_J(\mathbf{x}^{(J)}{}_0, P_{\mathbf{X}^{(J)}})$$

and thus the inequality (2.8).

(vi) The result is a direct consequence of the remark at the end of the first section of the paper and of the uniform continuity of function:

$$x \to \left(1+\left(\|x-x^J\|^2 / \sum_{j=1}^\infty \lambda_j\right)^\beta\right)^{-1}.$$

□

2.3. **Sample depth consistency.** Let $P_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ be the empirical measure based on the i.i.d. sample $X_1,\dots,X_n$ of $\mathbb{L}^2[0,1]$-valued processes with distribution $P \in \mathcal{P}_2(\mathbb{L}^2[0,1])$. A natural estimator of $\mathrm{PCD}^J = \mathrm{PCD}^J(\cdot,P)$ (2.3) is given by

$$\mathrm{PCD}_n^J(x) := \mathrm{PCD}^J(x,P_n) = \frac{1}{1+\left(\frac{\|x-x_n^J\|^2}{\sum_{i=1}^\infty \lambda_j^n}\right)^\beta} D_J\left(\mathbf{x}_n^{(J)}, P_{n,\mathbf{X}_n^{(J)}}\right) \tag{2.9}$$

where $(\lambda_j^n)_j$ and $(e_j^n)_j$ are respectively eigenvalues and eigenfunctions of a K-L decomposition associated to P_n. They are driven by the empirical covariance operator:

$$\widehat{\mathcal{K}}_n(f)(\cdot) := \int_0^1 \hat{K}_n(\cdot,s)f(s)ds \quad \text{with} \quad \hat{K}_n(s,t) = \frac{1}{n}\sum_{i=1}^n (X_i(s)-\overline{X}_n(s))(X_i(t)-\overline{X}_n(t)).$$

and $\overline{X}_n(t) = \frac{1}{n}\sum_{j=1}^n X_i(t)$ denotes the empirical mean process. One may further highlight that the empirical covariance operator $\widehat{\mathcal{K}}_n$ may be written

$$\widehat{\mathcal{K}}_n(f) = \sum_{j\geq 1} \lambda_j^n \left\langle e_j^n, f\right\rangle e_j^n$$

with $\lambda_j^n = 0$ for all $j \geq n+1$.

Let also introduce the following quantities

$$\mathbf{x}_n^{(J)} := \left(\left\langle x-\overline{X}_n, e_1^n\right\rangle, \dots, \left\langle x-\overline{X}_n, e_J^n\right\rangle\right) \in \mathbb{R}^J \quad \text{and}$$

$$x_n^J(\cdot) := \overline{X}_n(\cdot) + \sum_{j=1}^{J} \left\langle x - \overline{X}_n, e_j^n \right\rangle e_j^n(\cdot) \in \mathbb{L}^2[0,1]$$

which are respectively the empirical analogues of (2.1) and (2.2). Eventually, the empirical measure associated to $P_{\mathbf{X}^{(J)}} := P_{(\langle X - m_X, e_j\rangle)_{j=1,\dots,J}} \in \mathscr{P}(\mathbb{R}^J)$ is the probability measure

$$P_{n,\mathbf{X}_n^{(J)}} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\left(\langle X_i - \overline{X}_n, e_1^n\rangle, \dots, \langle X_i - \overline{X}_n, e_J^n\rangle\right)} \in \mathscr{P}(\mathbb{R}^J).$$

The empirical counterpart of Proposition 2.3 is provided in the next proposition. This result is driven by consistency of the estimated eigenvalues $(\lambda_j^n)_j$. Indeed, assumption $\lambda_{J+1} < \lambda_J$ implies that $\lambda_{J+1}^n < \lambda_J^n$ for n large enough since $\lambda_{J+1}^n \xrightarrow[n\to\infty]{a.s.} \lambda_{J+1}$ and $\lambda_J^n \xrightarrow[n\to\infty]{a.s.} \lambda_J$ (see Proposition 2 of [12]). Hence, as the theoritical one, the empirical eigenspace is not cut through for large n and the estimated depth does not depend on the choice of the basis.

Proposition 2.5. *If $\lambda_{J+1} < \lambda_J$ and the multivariate depth D_J is invariant under isometries (Property **P-1**), then, a.s., for n large enough, PCD_n^J does not depend on the choice of orthonormal basis of the subspace of dimension J, $L_J^n := +_{j=1,\dots,J} E_j^n$, where the E_j^n are the empirical eigenspace associated to λ_j^n, $1 \le j \le J$.*

Taking into account the multiplicity of the $\mathcal{K}$-eigenvalues λ_j, $j = 1, \dots, J$, plays an essential role in the consistency of the empirical eigenvalues, eigenfunctions or projectors.

Distinctly, as long as the PCD^J depth consistency is concerned, the uniform regularity requirement **P-6** upon the multivariate depth D_J is needed as seen below in Theorem 2.6.

Theorem 2.6. *Let $J \ge 1$ and D_J be a J-multivariate depth. Consider the associated functional depth PCD^J given in Definition 2.2. Let $P \in \mathcal{P}_2(\mathbb{L}^2[0,1])$ satisfying $\lambda_J > \lambda_{J+1}$.*

*(1) if D_J satisfies Assumptions **P-1** and **P-6** then*

$$\forall x \in \mathbb{L}^2[0,1], \quad \mathrm{PCD}^J(x, P_n) \xrightarrow[n\to\infty]{a.s.} \mathrm{PCD}^J(x, P).$$

*(2) if D_J satisfies Assumptions **P-1**, **P-4** and **P-6U** then*

$$\sup_{x \in \mathbb{L}^2[0,1]} \left|\mathrm{PCD}^J(x, P) - \mathrm{PCD}^J(x, P_n)\right| \xrightarrow[n\to\infty]{a.s.} 0.$$

Property **P-6** of continuity in the distribution argument, contestably, exhibits a relatively weak formulation, and within the extensive body of literature, a more robust and uniform, though demanding, version of it is embodied by Property **P-6U** as introduced in [21]. Additionally, it is established that **P-6**/**P-6U** is fulfilled by most of the multivariate depths as discussed thereafter.

The Mahalanobis depth adheres to **P-6** for distributions with invertible covariance matrix. With a minor regularity condition as outlined by [7], zonoid depth meets **P-6** for distributions with finite first moment: $\mathbb{E}[\|X\|] < \infty$. The latter holds true in our case, since our setting requires, at least, the existence of a finite second moment for the process under consideration. A similar behavior is observed for $\mathbb{L}^p$-depth, $p \ge 1$. In contrast, halfspace and simplicial depths are generally non-continuous in both their arguments, however they are when the distribution admits Lebesgue-density on $\mathbb{R}^J$. This characteristic is established in works by [36] (Th. 5) and [40] (Prop. 1). Based on the summarized work of [42] (Table 2), we restitute in Table 2 continuity in the distribution properties **P-6**/**P-6U** of some multivariate depths commonly encountered in the literature (the reader may refer to Appendix B for a precise definitions of these different depth).

Adherence of D_J to continuity in the distribution							
Property	MHD	$D_{\mathbb{L}^p}$	D_{Proj}	D_{HS}	D_{Simp}	D_{Zon}	D_{Spa}
P-6	✓	✓	✓	✓[1]	✓[1]	✓	✓
P-6U	✓*	✓*	✓*				

✓[1] : for absolutely continuous multivariate distributions.
✓* : when the probability space is equipped with the Wasserstein distance as metric $\mathrm{d}_{\mathscr{P}}$.

TABLE 2. Analytical properties (**P-6**)-(**P-6U**) of some multivariate depths D_J, $J \geq 2$ (Table 2 of [42]). Fulfilment of the involved property is indicated by ✓.

In Theorem 2.7 below, we provide a rate of convergence for our PCD estimator. For that, we need further assumptions upon the multivariate depth D_J:

- *Power decay of the multivariate depth D_J with respect to the space variable:* there is a continuous function $M : \mathscr{P}(\mathbb{R}^J) \to]0;\infty[$ s.t. for any $P \in \mathscr{P}(\mathbb{R}^J)$,
$$\forall z \in \mathbb{R}^J,\ D_J(z,P) \leq \frac{M(P)}{1+\|z\|_J^{2\beta}}. \tag{A1}$$
where β is the constant appearing in the definition of the PCD depth 2.2 and $\|\cdot\|_J$ is the euclidean norm on $\mathbb{R}^J$.
- *Local Lipschitz regularity with respect to the probability measure:* fix some $p \geq 1$ and consider W_p the p-Wasserstein distance on $\mathscr{P}_p(\mathbb{R}^J)$, the set of p-integrable probability measure on $\mathbb{R}^J$, defined by
$$W_p(P,Q) := \inf_{X\sim P, \mathrm{Y}\sim Q} \mathbb{E}[\|X-\mathrm{Y}\|^p]^{\frac{1}{p}}.$$
We suppose that for any $P \in \mathscr{P}_p(\mathbb{R}^J)$ there is an open neighborhood $U_P \subset \mathscr{P}_p(\mathbb{R}^J)$ of P and a constant K_P such that
$$\forall Q \in U_P,\ \sup_{z\in\mathbb{R}^J} |D_J(z,P) - D_J(z,Q)| \leq K_P W_p(P,Q)\,. \tag{A2}$$

Theorem 2.7. *Let $J \geq 1$ and $p \geq 1$ and consider D_J be a J-multivariate depth which satisfies **P-1**, **P-4** and Assumptions (**A1**) and (**A2**) for the chosen p.*

Fix some $q > 2p$ and denote by $\mathcal{P}_q(\mathbb{L}^2[0,1])$ the set of probability measures in $\mathcal{P}_2(\mathbb{L}^2[0,1])$ such that
$$\int_{\mathbb{L}^2[0;1]} \|x\|^q P(dx) < \infty.$$
Let $P \in \mathcal{P}_q(\mathbb{L}^2[0,1])$ such that $\lambda_J > \lambda_{J+1}$. Then, for any $\epsilon > 0$,
$$n^{\left(\frac{1}{2}\wedge\frac{p}{J}\wedge\beta\right)-\epsilon}\left|\mathrm{PCD}^J(x,P) - \mathrm{PCD}^J(x,P_n)\right|$$
and
$$n^{\left(\frac{\beta}{1+2\beta}\wedge\frac{p}{J}\right)-\epsilon} \sup_{x\in\mathbb{L}^2[0,1]} \left|\mathrm{PCD}^J(x,P) - \mathrm{PCD}^J(x,P_n)\right|$$
converge in probability to 0.

Remark 2.8. *The proofs of Theorem 2.6 and Theorem 2.7 complement one another and may be found in Appendix A.*

Several common depth functions satisfy Assumptions (**A1**) and (**A2**) and therefore Theorem 2.7. Concerning assumption (**A1**), it can be readily verified that Mahalanobis depth satisfies (**A1**) for any $0 < \beta \leq 1$, while $0 < \beta \leq 1/2$ holds for $\mathbb{L}^p$-depth for any $1 \leq p < \infty$ (due to the equivalence of all norms on $\mathbb{R}^J$). Additionally, for the projection depth, taking $u = z/\|z\|$

in its definition results in $D_{\text{Proj}}(z,Q) \leq C_Q/(1+\|z\|)$, hence any $0<\beta\leq 1/2$ holds. When the depth is halfspace or simplicial, one may establish upper bounds by employing Markov's inequality, and this holds true for any β such that $\|X\|^{2\beta}$ is integrable.

Moving on to (**A2**), using a triangular inequality and Hölder's inequality, it is straightforward that $|D_{\mathbb{L}^p}(z,P)-D_{\mathbb{L}^p}(z,Q)|\leq \mathbb{E}_{\mathbf{X}\sim P,\mathbf{Y}\sim Q}[\|\mathbf{X}-\mathbf{Y}\|_p]$. Thus, by Jensen's inequality for the concave function $t\mapsto t^{1/p}$, $p\geq 1$, it yields $\sup_{z\in R^J}|D_{\mathbb{L}^p}(z,P)-D_{\mathbb{L}^p}(z,Q)|\leq KW_p(P,Q)$. Consequently, $\mathbb{L}^p$-depth satisfies (**A2**). Another example is Mahalanobis depth, which meets (**A2**) with $p=2$. Indeed, using the upper bound derived in [1] for MHD, one may write:

$$\|\text{MHD}_P-\text{MHD}_Q\|_\infty \leq \|\Sigma_P^{-1}-\Sigma_Q^{-1}\|+\|\Sigma_Q^{-1/2}\|^2\|m_Q-m_P\|\left(\|m_Q-m_P\|+\|\Sigma_P^{1/2}\|\right).$$

Notice when Q converges to P in the weak sense, by continuity, we have for some constant $K>0$ depending on P,

$$\begin{aligned}\|\text{MHD}_P-\text{MHD}_Q\|_\infty &\leq \|\Sigma_P^{-1}-\Sigma_Q^{-1}\|+C\|m_Q-m_P\|\\ &\leq K\left(\|\Sigma_P^{-1}(\Sigma_P-\Sigma_Q)\Sigma_Q^{-1}\|+\mathbb{E}[\|\mathbf{X}-\mathbf{Y}\|\right)\\ &\leq K\left(\|\Sigma_P-\Sigma_Q\|+\mathbb{E}[\|\mathbf{X}-\mathbf{Y}\|\right).\end{aligned}$$

When P and Q have zero mean, some classical computations based on the matrix norm over $\mathbb{R}^J$ yield

$$\|\Sigma_P-\Sigma_Q\|\leq C\sqrt{\mathbb{E}[\|\mathbf{X}-\mathbf{Y}\|^2]},$$

and the same holds true when P and Q are not centered, since $\mathbf{X}-\mathbb{E}_P[\mathbf{X}]$ and $\mathbf{Y}-\mathbb{E}_Q[\mathbf{Y}]$ are. Thus, $\|\text{MHD}_P-\text{MHD}_Q\|_\infty\leq CW_2(P,Q)$.

Considering further the projection depth, as demonstrated by [55], we generally have, for some constant $C>0$ depending on P as Q converges to P,

$$\|D_{\text{Proj},P}-D_{\text{Proj},Q}\|_\infty \leq C\sup_{\|u\|=1}\left(|\sigma(P_u)-\sigma(Q_u)|+|\mu(P_u)-\mu(Q_u)|\right)$$

where P_u denotes the law of $\langle u,\mathbf{X}\rangle$, $\mathbf{X}\sim P$, and μ and σ are univariate location and scale functions such that $\mu(aZ+b)=a\mu(Z)+b$ and $\sigma(aZ+b)=|a|\sigma(Z)$ for any $a,b\in\mathbb{R}$. Clearly, by applying Cauchy-Schwarz and linearity of μ, we have

$$\sup_{\|u\|=1}|\mu(P_u)-\mu(Q_u)|\leq \mathbb{E}[\|\mathbf{X}-\mathbf{Y}\|].$$

Additionally, when $\sigma(Z):=\mu(|Z-\mu(Z)|)$ represents the absolute mean deviation,

$$\sup_{\|u\|=1}|\sigma(P_u)-\sigma(Q_u)|\leq \sup_{\|u\|=1}2\mu(|P_u-Q_u|)\leq 2\mathbb{E}[\|\mathbf{X}-\mathbf{Y}\|]].$$

Consequently, for some constant $C>0$, it follows from Jensen's inequality for concave functions

$$\|D_{\text{Proj},P}-D_{\text{Proj},Q}\|_\infty\leq C\mathbb{E}[\|\mathbf{X}-\mathbf{Y}\|_2]\leq CW_2(P,Q),$$

and D_{Proj} meets (**A2**) with $p=2$. The result is still valid when σ is the median absolute deviation as generally used in the definition of D_{Proj}.

3. Illustrations and simulations

Supervised k-class classification is a fundamental task in machine learning and statistics where the goal is to categorize data points into one of the k predefined classes or categories. It involves using labeled training data to train a model that can accurately assign new, unlabeled data points to one of the k classes. This topic is of substantial interest in the fields of machine learning, pattern recognition and is widely used in various applications, including spam email detection, medical diagnosis, sentiment analysis, and fraud detection. The field

of k-class classification encompasses a range of algorithms and techniques, including linear classifiers, nearest neighbor methods (such as the classical k-NN), support vector machines, and deep learning models. The performance of these models is evaluated using metrics such as accuracy, precision, recall, F1-score... In our study, we confine our discussion to the simplest case $k = 2$ of two-class classification. In this context, we consider two independent sets of data samples, denoted as $X_1^{(0)}, X_2^{(0)}, \ldots, X_n^{(0)}$ and $X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)}$, drawn from two different distributions (or populations) labeled as P_0 and P_1 and characterized by the random variables $X^{(0)}$ and $X^{(1)}$, respectively. Hence, as described above, the goal of classification procedures is to allocate a newly observed data to either P_0 or P_1, leveraging the information acquired from the training samples. More precisely, the data are made of pairs (X_i, Y_i), where X_i represents the input variables (upon which the classification is based), and Y_i is the associated label (usually equals 0 or 1), describing the membership of the corresponding observation to either P_0 or P_1. In this context, considering a new observation X with unknown associated label, the objective is to predict the value of this label. Using the functional depth we proposed (see Section 2.1), we develop here a method to classify functional data.

The topic of functional classification has been widely studied over the years. For example we can refer to [4, 8, 16, 19, 28, 33, 49]. On another side depth have already been used to propose classification methods. However, most of the time, only the case of low-dimensional data is considered: we can cite for instance, the work of [19] who integrates linear discrimination principles with aspects of data depth classifiers, or [41] solving multi-class classification. Other maximum multivariate depth classifiers were studied in the literature; the reader may refer to [20, 15, 34] and references therein. Finally, we can mention a few propositions to perform functional classification with depths ([43, 10, 52, 47, 11]).

Using the works of [11] and [43] serving as benchmarks, we show strong performances of our maximum PCD_J-based classifier.

3.1. **Description of our classification method.** Recall that when provided with a training dataset where observations are accurately categorized as either P_0 or P_1, the goal is to classify new test samples as either P_0 or P_1 based on their relative positions within the respective training datasets. In our simulation below, we consider the maximum depth classifier as originally introduced by [20] based on our depth definition PCD_J (definiton 2.2). Then, we make use of the framework of [11] as a benchmark of comparison. In contrast to parametric and semiparametric classification methodologies, maximum depth classifiers diverge from the assumption of a specific parametric structure for the decision boundary or any particular probability distribution governing the datasets. Instead, their classification procedure hinges on assigning an observation to the set with respect to which it attains maximum depth value. Formally, given $n_1, n_2, \ldots, n_k$ observations from k competing classes, the maximum PCD classifier of a new observation x, is expressed as:

$$d^J(x) := \underset{q=1,\ldots,k}{\arg\max}\, \mathrm{PCD}_J(x, P_{n_q}),$$

where P_{n_q} is the empirical measure of the qth class. We recall that

$$\mathrm{PCD}^J(x, P_n) = \frac{1}{1 + \left(\frac{\|x - x_n^J\|^2}{\sum_{i=1}^{\infty} \lambda_j^n}\right)^{\beta}} D_J\left(\mathbf{x}_n^{(J)}, P_{n,\mathbf{X}_n^{(J)}}\right),$$

with J the number of FPC's as tuning parameter, and β can be an additional parameter for optimization. However, for the sake of computational simplicity, our study is limited to

optimization in J only, as will be illustrated later on.

Subsequently, in all examined classification schemes, our approach involves the following steps, with the technical details rooted in the framework of [11], incorporating some adaptations specific to our context:

- **Discretization:** all simulated curves are discretized along 51 evenly spaced grid points that span the interval $[0, 1]$. Obviously, in the case of a real dataset, these time points are inherently confined to a specified compact interval.
- **Number of runs:** the simulation results are established on 100 independent runs. Conversely, with real observations, the procedure entails permuting a partition of the datasets into a 90 % training set and a 10 % test set, followed by cross-validation.
- **Training and test data:** in each run, we generate 200 training curves for each model. Specifically, 100 are selected from P_0, and another 100 are chosen from P_1. Additionally, a test sample comprising 100 observations is created, with 50 originating from P_0 and 50 from P_1. In the case of real datasets, these proportions adhere to the previously mentioned configuration (90 % train / 10 % test).
- **Maximum $\mathbf{PCD_J}$ - depth classifier criteria:** Given a vector of number of projections (or pc's) $\vec{J} = (0, 1, \ldots, Jmax)$ and a given multivariate depth (in the definition of PCD_J), we undertake a cross-validation procedure based on the 100 training data for each single class P_0 and P_1 (dividing the 100 curves into 20 batches of 5 curves where we consider 95 curves as train subset and 5 curves as test subset, and we swap them over the 20 batches), in the aim to choose the best J^* within $\vec{J}$, in the sense that it gives the best accuracy within $\vec{J}$ (it is a mean accuracy based the 20 swapings).
- **Evaluation:** At each run ($k = 1, \ldots, 100$), the value of J^* is updated, for which we undertake again the classification procedure on the original train and test sets. Finally, we compute the proportion of correctly classified observations within the test sample.
- **Table interpretation:** the values contained within the table cells provide fundamental descriptive statistics, including common measures such as the mean, median, quantiles, and more. These statistics offer insights into the distribution of correctly classified proportions across the 100 (or given) runs.

3.2. **Simulation study with functional data.** Here we use the two settings proposed by [11]:

- **Model 1:** the population denoted as P_0 comprises trajectories of the underlying process $X(t) := m_0(t) + e(t)$, where $m_0(t) = 30(1-t)t^{1.2}$ and $e(t)$ is a centered Gaussian process with covariance function $\mathrm{cov}(X(s), (X(t)) = 0.2\exp(-|s-t|/0.3)$. The process associated with population P_1 is essentially identical to the process $X(t)$, except for variations in its mean function. It is defined by $\mathrm{Y}(t) = m_1(t) + e(t)$, with $m_1(t) = 30(1-t)^{1.2}t$.
- **Model 2:** the population P_0 consists of trajectories of the process $X(t) = m_0(t) + e(t)$, where this time, $m_0(t) = 30(1-t)t^2 + 0.5|\sin(20\pi t)|$ and $e(t)$ is a centered Gaussian process with $\mathrm{cov}(X(s), X(t)) = 0.2\exp(-|s-t|/0.3)$. The process from P_1 is made of spline approximations (with 8 knots) of trajectories from the previous process from P_0.

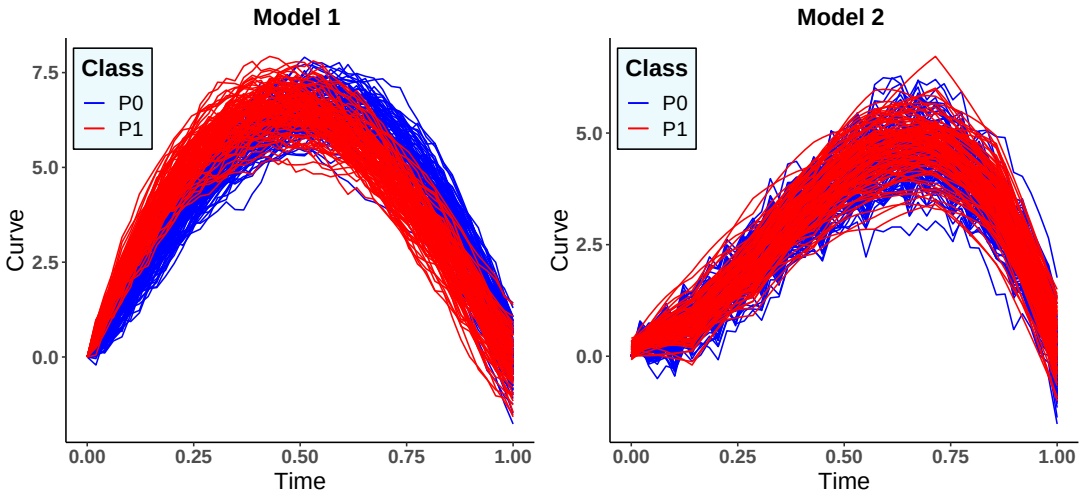

FIGURE 1. Some trajectories from Model 1 and Model 2.

As visually represented in Figure 1 (left panel), Model 1 can be observed as depicting a relatively consistent scenario characterized by rather smooth functions. In contrast, Model 2 (right panel) introduces a more irregular context where the mean function m_0 exhibits oscillations, deviating from the smooth function $t \mapsto 30(1-t)t^2$. The primary objective of this setup is to assess the efficiency of detecting trajectories from P_0 or P_1 within each model. We recall that the classification procedure is carried out with the methodology outlined in Section 3.1 above.

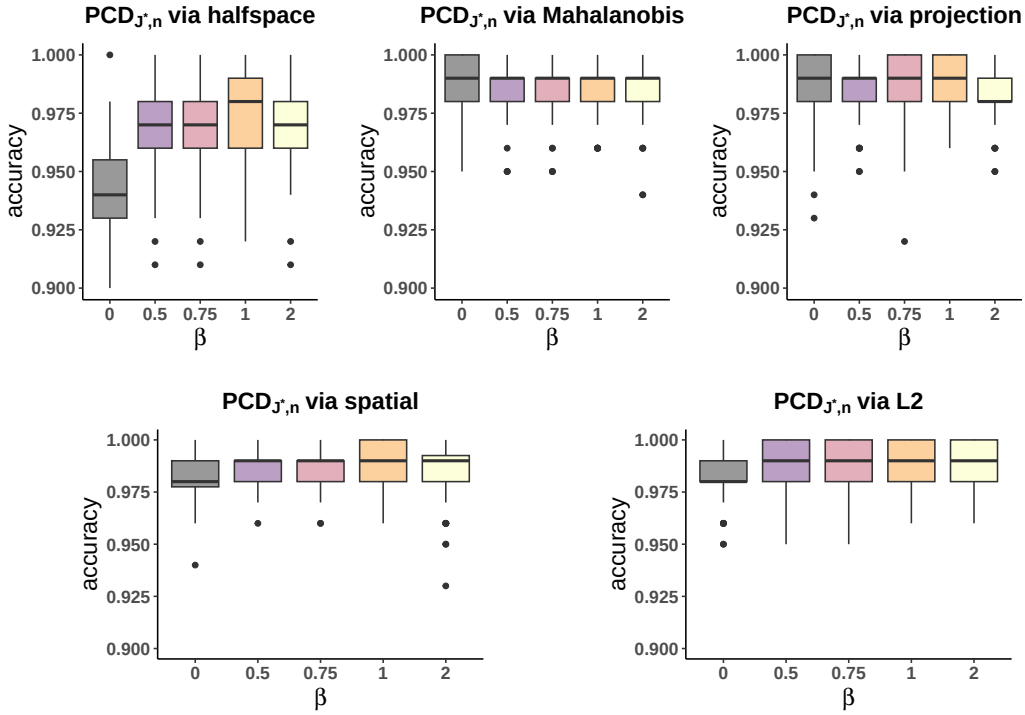

FIGURE 2. Boxplots of the correct classification proportions under PCD_{J^*} for several β values and Model 1.

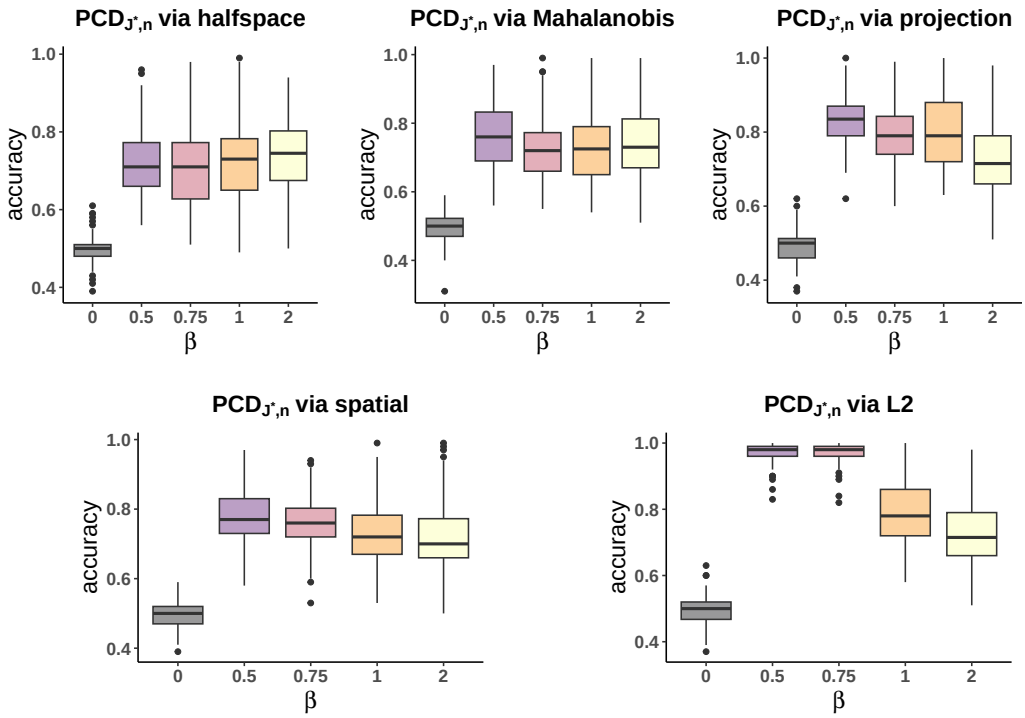

FIGURE 3. Boxplots of the correct classification proportions under PCD_{J^*} for several β values and Model 2.

Boxplots in Figures 2 (representing Model 1) and 3 (representing Model 2) illustrate the accuracy variations across selected values of β in both models. Take note that in this study, opting for $\beta = 1$ generally results in slightly better outcomes in Model 1, emphasizing that the choice of β does not affect accuracy results. Conversely, the opposite trend is observed for Model 2 and $\beta = 1/2$ seems a better choice. Therefore, in addition to optimizing J^*, the inclusion of the parameter β for further optimization in some β^* may be considered, however, this may significantly increase computation time.

The performances of our Monte Carlo procedure are succinctly summarized in Tables 3 and 4.

D_{J^*} depth	Min	First Quartile	Median	Mean	Third quartile	Max
halfspace	0.88	0.96	0.98	0.9737	0.99	1
Mahalanobis	0.96	0.98	0.99	0.9853	0.99	1
projection	0.9	0.98	0.99	0.9891	1	1
spatial	0.96	0.98	0.99	0.9894	1	1
L2	0.96	0.98	0.99	0.9879	1	1
k-NN	0.98	1	1	0.9977	1	1
RP	0.97	0.99	1	0.9961	1	1
RP2	0.93	0.98	0.99	0.9849	0.9925	1
RPD	0.94	0.97	0.98	0.981	0.99	1

TABLE 3. Descriptive statistics for the correct classification proportions under PCD_{J^*} with $\beta = 1$ for Model 1. The second bloc in the table displays the reference results within the same framework, based on the work of [11].

D_{J^*} depth	Min	First Quartile	Median	Mean	Third quartile	Max
halfspace	0.56	0.66	0.71	0.7204	0.7725	0.96
Mahalanobis	0.56	0.69	0.76	0.7645	0.8325	0.97
projection	0.62	0.79	0.835	0.8309	0.87	1
spatial	0.58	0.73	0.77	0.7826	0.83	0.97
L2	0.83	0.96	0.98	0.9694	0.99	1
k-NN	0.82	0.87	0.89	0.8912	0.91	0.97
RP	0.63	0.78	0.8	0.8042	0.84	0.93
RP2	0.95	0.9975	1	0.9962	1	1
RPD	0.9	0.98	0.99	0.9884	1	1

TABLE 4. Descriptive statistics for the correct classification proportions under PCD_{J^*} with $\beta = 1/2$ for Model 2. The second bloc in the table displays the reference results within the same framework, based on the work of [11].

Under Model 1, one may observe from Table 3 that the k-Nearest Neighbors (k-NN) method demonstrates superior efficiency, with the performance of the Random Projection (RP) method being relatively comparable. In contrast, when still examining Model 1, our method based on PCD_J-depth exhibits comparable performance, slightly outperforming RP2 and RPD. In particular, it demonstrates significant enhancements in the minimum value, as seen when employing parameters such as $\beta = 1$ and projection, spatial, Mahalanobis, or L2-depths.

In contrast, under Model 2, there is notably higher variability, but the methods based on derivatives, specifically RP2 and RPD, clearly outperform the remaining classifiers (refer to Table 4). Consequently, when anticipating differences in the smoothness of the functions between P_0 and P_1, there is a compelling argument for employing the double projection methods.

We highlight the persistent preference for $J = 5$, sporadically $J = 4$, rarely $J = 3$ and the lack of $J = 0, 1, 2$ across all models throughout the procedure highlights the inherent capability of PCD_J to dynamically ascertain an optimal dimensionality for classification purposes. This adaptive attribute is actualized through the process of cross-validation performed on the parameter J, as expounded upon earlier.

Last but not least, we mention that the Karhunen-Loève decomposition can be extended to more general Hilbert spaces, such as Sobolev spaces, where the principal components will as such take into account the functions and their derivatives. This is indeed particularly relevant in scenarios where the smoothness of functions holds significance. Hence, PCD_J depth can be adapted in order to take into account regularity of functions, but this is beyond the scope of this paper. Therefore, the PCD_J depth measure can be modified to accommodate the regularity of functions, although this is beyond the scope of the present paper.

3.3. **Application on real data.** For the purpose of comparing performances, we align our analysis with specific real data, as investigated in [43], which, in the next two subsections, will consist of the Adelaide electricity demand [1] and the Tecator dataset [2] (loaded from the `ddalpha` package in `R`).

Following the identical classification methodology detailed in Section 3.2, specifically requiring optimization of the number of principal components J, we draw out accuracy outcomes for both real datasets and compare them to those derived in [43].

[1] Available at https://rdrr.io/cran/fds/man/Electricitydemand.html

[2] Available at https://search.r-project.org/CRAN/refmans/ddalpha/html/dataf.tecator.html

3.3.1. *Adelaide electricity demand.* The Adelaide electricity demand [39], encompasses the daily electricity consumption in megawatts (MW) recorded in Adelaide between 1997 and 2007. This dataset comprises 508 curves for each weekday, derived from half-hourly measurements (thus, along 48 time points), which underwent B-splines smoothing.

For the classification focus, two business days—Monday and Tuesday—and the two weekend days—Saturday and Sunday—were selected. Four leave-one-out classification schemes were implemented, coupling each business day with each weekend day. The four considered datasets can be visualized in Figure 4. Although there is a significant overlap in the electricity demand curves as one may notice in figure 4, distinctive variations in shape serve to differentiate observations between business days and weekends. Naturally, business days are characterized by a shorter duration of daily low demand and a more pronounced surge in demand preceding office hours, in contrast to the two weekend days. Indeed, weekends demonstrate a lower daily mean, usually achieved one to two hours later.

For the Adelaide dataset, once again, it happened that $\beta = 1/2$ serves as a suitable parameter for achieving the highest accuracies. Table 5 presents the leave-one-out classification accuracy rates for the four described models, employing $\beta = 1/2$, as well as the reference rates of [43]. The parameter J is determined through leave-one-out cross-validation, following the same procedure led in Section 3.1.

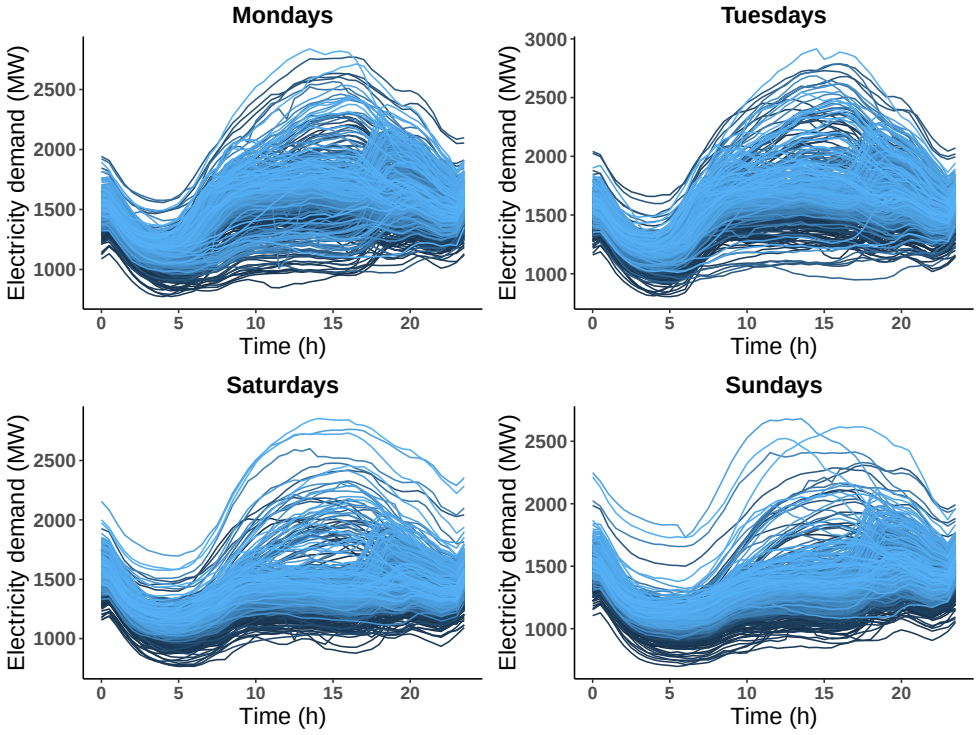

FIGURE 4. Adelaide electricity demand curves over Mondays, Tuesdays, Saturdays and Sundays between years 1997 and 2007.

	halfspace	projection	MHD	spatial	L2	FD_3^k	PC	PLS	k-NN
Monday-Saturday	0.8799	0.8622	0.8878	0.8868	0.9104	0.907	0.904	0.755	0.931
Monday-Sunday	0.8986	0.9045	0.933	0.9252	0.939	0.902	0.945	0.828	0.94
Tuesday-Saturday	0.874	0.8976	0.9242	0.9291	0.9488	0.923	0.92	0.778	0.948
Tuesday-Sunday	0.937	0.9576	0.9842	0.9862	0.9882	0.966	0.988	0.864	0.977

TABLE 5. Accuracy rates for the Adelaide-Electricity leave-one out classification problem described above, based on maximum depth classification using PCD_{J^*} and $\beta = 1/2$. Reference accuracy rates based on the work of [43] are represented in the last four columns.

Upon comparing our findings with the established reference results, notice that PCD_J surprisingly consistently demonstrates strong performance w.r.t other methodologies across all the examined models. For model 1 (Monday-Saturday), the k-NN classifier stands out as the most effective, even though, our L2-depth based PCD_J demonstrates a competitive proximity to the performance of k-NN in this particular model. It is intriguing to observe that L2-depth based PCD_J, outperforms the reference depths FD_J^k, $J = 1, 2, 3$ [43] for all four models. Remarkably, L2-based PCD_J shows alignment with the top-performing classifiers (PC with Model 2 and 4, k-NN with Model 3). Notice further, that MHD and spatial depths exhibit analogous performances. However, it is important to mention that PC-classifier is acknowledged as an optimal classifier within specific assumptions about the underlying generating process, as outlined in [13].

On another note, the consistent selection of $J = 6$, occasionally $J = 5$, and the absence of $J = 0, 1, 2, 3, 4$ for each model through the procedure, emphasizes PCD_J's capacity to dynamically determine an optimal dimension for classification. This adaptability is achieved through cross-validation on J as mentioned earlier.

In terms of computational complexity, PCD_J is akin to the computational cost of the included multivariate depth added to the computational cost of the K-L decomposition in its definition. The complexity of this multivariate depth, in turn, may exhibit either relatively low or high complexity, depending on J and n. Specifically, the J-dimensional halfspace depth is recognized for its higher complexity, scaling at an order of $O(n^{J-1}\log(n))$. Numerous advancements have been made in developing both approximate and exact algorithms for efficiently computing halfspace depth (for comprehensive information, refer to works such as [56, 24, 37, 48]). Whereas, computing Mahalanobis, spatial or L2 depth is, for instance, significantly faster.

3.3.2. *Tecator dataset.* The Tecator spectrometric dataset, documented by [6], consists of the light wavelength absorbance patterns of 215 finely chopped meat samples measured by the Tecator Infratec spectrometer. Each meat sample's absorbance was recorded at 100 equidistant discretization points across the wavelength spectrum ranging from 850 to 1050 nanometers.

In the context of a binary classification problem, the meat samples have been categorized into two groups based on their fat content. Indeed, the goal is to differentiate between meat samples with low fat content (typically less than or equal to 20 % giving rise to 138 observations) and those with high fat content (usually exceeding 20 %, consisting of 77 remaining observations). Figure 5 depicts the previous classical binary grouping.

This dataset is a recurrent focus in the functional data analysis literature, often aiming to classify meat samples into different groups based on their fat content as mentioned above, posing challenges due to overlappings and similar smoothness properties in the spectrometric curves (the reader may refer to [16]). Nonetheless, distinctions in curvature are remarkable

among the observations (for instance, based on Figure 5, the blue curves display two slight peaks compared to the red curves).

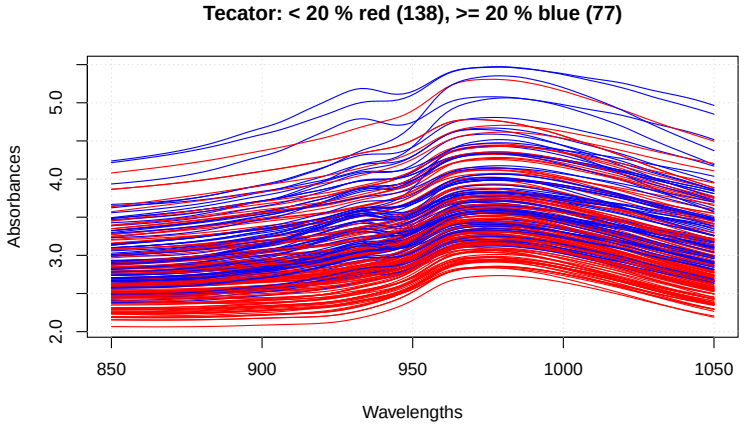

FIGURE 5. Tecator light wavelength absorbance data of 215 curve samples of meat. Curves with less (resp. more) than 20 % of fat content are in red (resp. blue).

$\boldsymbol{\beta}$	halfspace	projection	MHD	spatial	L2	FD_3^k	PC	PLS	k-NN
$1/2$	0.8573	0.8849	0.95	0.9681	0.9399	0.907	0.907	0.963	0.81
1	0.8799	0.8972	0.9537	0.9681	0.9435	0.907	0.907	0.963	0.81

TABLE 6. Accuracy rates for the Tecator classification problem described above, based on maximum depth classification using PCD_{J^*}. Reference accuracy rates based on the work of [43] are represented in the last four columns.

As displayed in Table 6, PCD_J consistently delivers robust accuracy results, whatever $\beta = 1/2$ or $\beta = 1$. The spatial-based PCD_J is surpassing all other classifiers presented. Within the Tecator dataset, the persistent selection of $J = 5$ by the blue curves and $J = 4$ by the red curves, from the set $J = 0, 1, 2, 3, 4, 5, 6$, during the classification process, points out again the adaptability of PCD_J to this specific dataset. Notably, the necessity for an additional projection for the blue functions enables the capture of their more curved shapes, contrasting with the less wavy red functions.

3.3.3. *Functional CCTE based on Nice weather and power demand.* As it is of keen interest to us, the present section is dedicated to a real world environmental application. The dataset, which was downloaded from *RTE France*[3] and provided on demand by *Meteo France*[4] consists of daily power consumptions (in MWh) and temperatures (in °C) for the region of *Nice* (France) over a five-year period (2018-2022). This gives rise to $n_2 := 1825$ observed curves along 24 timepoints corresponding indeed to curves observed during a day/24hours. We mention that the dataset of temperatures was provided to us since 1993. However, power consumption data is available starting by the end of the year 2017, which is why we consider the period 2018-2022 for our study in what comes next.

To be more precise, the considered dataset is made of $n_2 = 1825$ observations of the form $\{(\mathrm{Y}_i, \mathbf{X}_i)\}_{1\leq i\leq n_2} \in (\mathbb{R} \times \mathbb{L}^2[0,1])^{n_2}$, where $\mathbf{X}_i := (X_i(t))_{t=1,2,\ldots,24}$ is the 24-hours curve of

[3] https://www.rte-france.com/

[4] https://meteofrance.com/

temperature on day i, $1 \leq i \leq n_2$ and $\mathrm{Y}_i \in \mathbb{R}$ is the average electricity consumption on day i. We may underline the fact that Y_i was computed using a 96-timepoints electricity statement corresponding to electricity consumption measured each 15 minutes of the day i.

The goal of the present section is to study an empirical functional version of the multivariate Covariate-Conditional-Tail-Expectation (CCTE) of [1]. The empirical CCTE is a multivariate depth based risk measure defined by:

$$\widehat{\mathrm{CCTE}}^{n_1,n_2}_{\mathrm{PCD}_J,\alpha}(\mathrm{Y},\mathbf{X}) := \frac{\sum_{i=1}^{n_2} \mathrm{Y}_i \mathbb{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1,J}(\alpha)}}{\sum_{i=1}^{n_2} \mathbb{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1,J}(\alpha)}}, \tag{3.1}$$

with the convention $0/0 = 0$. In the above, $S_{n_2} := \{(\mathrm{Y}_i, \mathbf{X}_i)\}_{i=1,..,n_2}$ are i.i.d and $\mathcal{L}_{n_1,J}(\alpha) := \mathcal{L}_{D_{J,n_1}}(\alpha)$ is the empirical functional level set associated to PCD_{J,n_1} and computed from a n_1-sample $\tilde{S}_{n_1} := (\tilde{\mathbf{X}}_i)_{i=1,..,n_1}$ independent from S_{n_2} $(n_1 + n_2 = n)$:

$$\mathcal{L}_{n_1,J}(\alpha) := \left\{ x \in \mathbb{L}^2[0,1] : \mathrm{PCD}_{J,n_1}(x) \leq \alpha \right\}.$$

Recall that D_J is the multivariate depth function involved in the definition of PCD-depth.

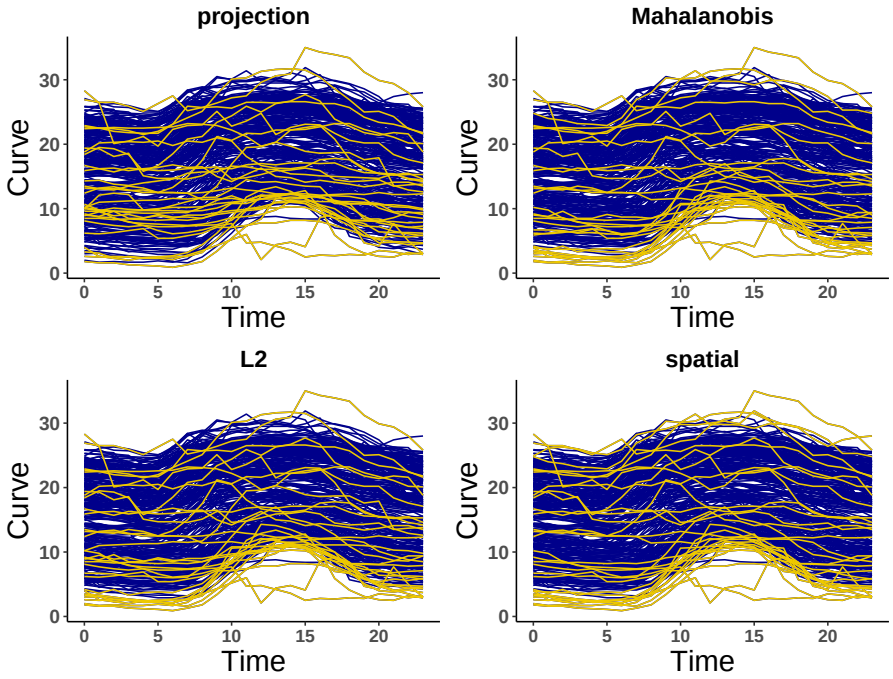

FIGURE 6. Functional 10 % - PCD_J level sets $\mathcal{L}_{n_1,J}(0.1)$ (golden curves) w.r.t some randomly chosen curves (blue curves) among the large period of 1993-2022. In this setting, $J = 2$ and $\beta = 1$.

As a reminder, the concept of CCTE was construed as a tool quantifying the cost associated with a specific rare event characterized as 'high-risk'. In our case, for each temperature curve $\mathbf{X}_i$, we associate a cost variable Y_i, representing the power consumption (in MWh) on day i. It's worth noting that the initial definition of depth-CCTE was formulated within a multivariate context based on a multivariate depth [1], and as such, the depth level set existed in $\mathbb{R}^J$. In the current iteration of functional CCTE, the PCD-level set $\mathcal{L}_{n_1,J}(\alpha)$ is then a functional

alternative of the multivariate depth level set. Thus, $\mathcal{L}_{n_1,J}(\alpha)$ is regarded as the region of risk pertaining to temperature curves with elevated risk levels.

This category of problems has been extensively investigated for many years within the realm of short and long-term load forecasting. Typically, the objective is to predict the electric daily load one to five days in advance, relying on weather-related input variables such as temperature and humidity for one or two preceding days (Dudek et al., 2016; Oreshkin et al., 2021) [14], [45].

In the aforementioned setting, the empirical CCTE at level α can be explained as the average power consumption when temperature curves deviate significantly from the norm (e.g., when considering low values of α). This perspective holds value for electric utility companies like EDF, a prominent French multinational electric utility company. For instance, if there are anticipations of severe heatwaves or cold spells in the coming decade, EDF could use this information to forecast future electricity investments over the specified period.

Subsequently, we explore the relationship between $\widehat{\text{CCTE}}_\alpha$ and α, considering four distinct multivariate depths (in the definition of PCD). These depths aim to encapsulate the general 'pattern' of the functional risk factors dataset, which are the daily temperature curves $\mathbf{X}_i$ for the years 2018-2022.

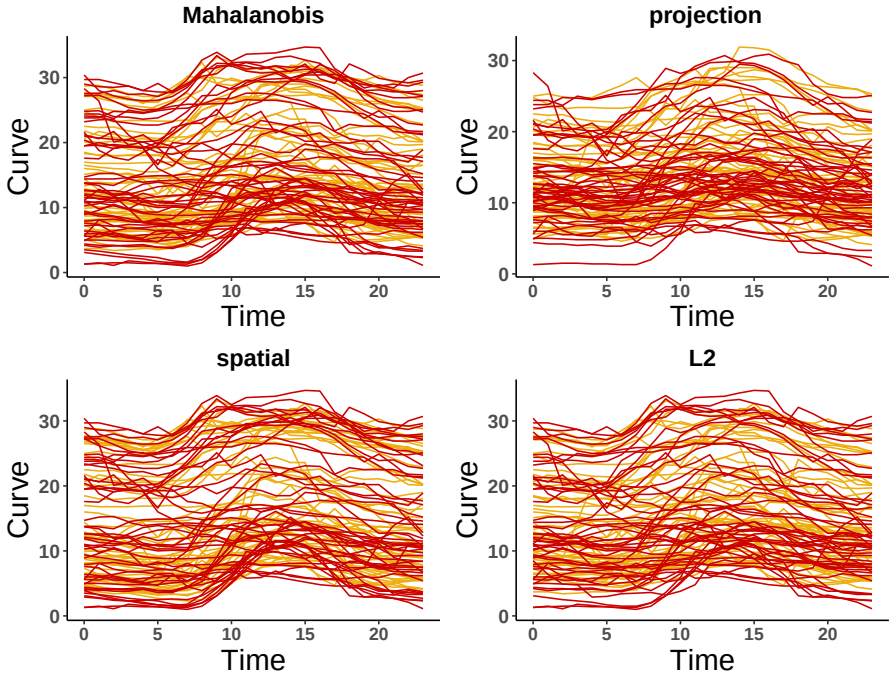

FIGURE 7. Functional 10% - PCD$_J$ level sets of the least deep temperature curves, $J = 2$ and $\beta = 1$, for two overlaying periods: the period of 2000-2004 (red curves) and 2018-2022 (golden curves).

Based on Figure 7 above, it is noteworthy that the periods spanning from 2000 to 2004 and from 2018 to 2022, exhibit similar patterns in the 'extreme' functional temperature level sets $\mathcal{L}_{n_1,J}$. In these figures, it appears that the PCD$_J$-level sets of the 10% least deep curves largely overlap. Analogous findings were noted for time intervals such as 2008-2012 and 2018-2022 (see Figure 11 in appendix C). Nonetheless, time periods such as 1993-1997 and 2018-2022 clearly exhibit a shift which is one discernible aspect of climate change (see Figure 10 in appendix C). As a result, it is reasonable to rather use the temperature curve data from 2000 to 2017 (and some 20 % curves of 2018-2022) for the purpose of estimating the PCD level set, while

focusing our primary analysis on the dataset from 2018 to 2022 to compute the mean CCTE.

In our analysis, we partitioned the overall datasets at hand in a way s.t. the first set $\tilde{S}_{n_1} := \left\{(\tilde{\mathbf{X}}_i)\right\}_{i=1,..,n_1}$, consists in $n_1 = 2000$ randomly chosen temperature curves between 2000 and 2017. Recall that the latter is used to estimate the functional PCD-level set and comprises only temperature curves $(\tilde{\mathbf{X}}_i)_i$, since only the temperature curves $\tilde{\mathbf{X}}_i's$ are required for level set estimation. Meanwhile, the second subset $S_{n_2} := \{(\mathrm{Y}_i, \mathbf{X}_i)\}_{i=1,..,n_2}$, $n_2 = 1825$ of the period 2018-2022, is employed for the mean CCTE estimation. Nevertheless, we reclaim 20 % of the n_2-sample and exclusively select the temperature curves from it, which are subsequently incorporated into the level-set defined by the n_1-sample. Consequently, the final sizes of the samples are adjusted to $n_1 = 2365$ and $n_2 = 1460$. In other words, $\mathcal{L}_{n_1,J}(\alpha) = \mathcal{L}_{2365,2}(\alpha)$ in the previous definition of functional CCTE.

We mention that, beforehand, we excluded instances with missing electricity records (NA electricity data points) and subsequently calculated the mean power consumption. Therefore, this computation was sometimes performed using fewer than 96 data points, specifically for the days with incomplete data. On the other hand, from the level-set sample, we already removed the curves corresponding to days having several missing temperature measures.

All the temperature curves in each datasets $\tilde{S}_{n_1}$ and S_{n_2}, were smoothed using B-splines with degree four. The selection of the B-spline degree plays a critical role, as it determines the degree of approximation that provides the closest representation of the actual curve (this is depicted in figure 9, Appendix C).

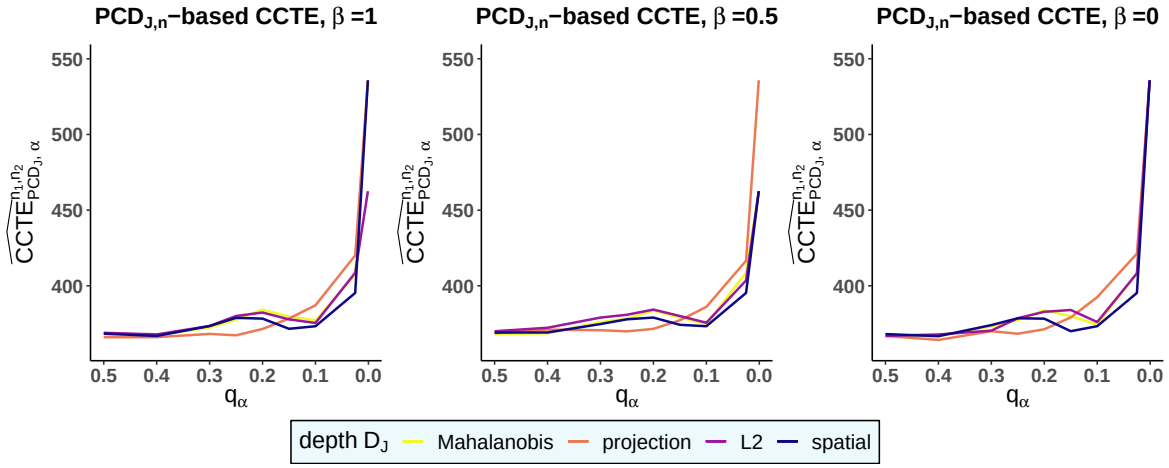

FIGURE 8. Functional PCD$_J$-based CCTE, $J = 2$, corresponding to 98.8% of explained variance. The x-axis must be read from left to right according to the decreasing order of $q_\alpha = 50\%, \ldots, 0.1\%$.

In Figure 8 above, we draw the behavior of the $\widehat{\mathrm{CCTE}}_\alpha$ (3.1) as a function of the empirical α-quantiles q_α. Here, q_α takes values in the set $\{50\%, \ldots, 1\%, 0.1\%\}$ and represents the proportion of the least deep observations $\mathbf{X}_i's$ in S_{n_2} that fall within the level set $\mathcal{L}_{n_1,J}$, resulting in matching PCD depth values α allowing the computation of $\widehat{\mathrm{CCTE}}_\alpha$ (3.1) a. We consider three configurations where $\beta = 1, 0.5, 0$, depicting an overall similar behavior of $\widehat{\mathrm{CCTE}}_\alpha$. In our study, we also draw the PCD$_J$-based CCTE for different classical multivariate depths (Figure 3.1).

As expected, similar to the multivariate setting, whatever the value of β, the functional estimator $\widehat{\text{CCTE}}_\alpha$ exhibits an upward trend as q_α decreases, for all four chosen depths, as illustrated in Figure 3.1 above. This behavior implies that as we move towards less central and hence "riskier" curves (in the PCD_J sense), the associated cost becomes increasingly significant. In simpler terms, low values of q_α correspond to temperature curves $\mathbf{X}_i$ that are notably 'extreme' (either very low or very high temperatures, or simply unusual daily temperatures, see Figures 6, leading naturally to higher power consumption.

4. Conclusion and perspectives

In this paper, we have introduced a new functional depth measure, PCD_J, constructed by including finite projections via the J principal components. This essentially reduces the definition to a finite-dimensional depth measure. We have established theoretical results to demonstrate the consistency of its respective estimator under mild conditions. To showcase our investigation, we have delved into binary classification problems—some simulated and others based on real data—previously explored using existing functional depths, providing a basis for comparison with our PCD_J classifier. Gratifyingly, our proposed classifier displayed superior performances when compared to the reference methods, demonstrating adaptability in selecting the number of principal components J according to the available data. Thus, there is potential to enhance classification performances by optimizing in β through cross-validation and similarly optimizing the choice of the J-variate depth; however, this would lead to a substantial increase in computational cost.

On the other hand, we have observed limitations in the applicability of our PCD_J classifier to functional data exhibiting significant regularity. Hence, a future challenge lies in refining the PCD_J depth measure to fit such functions. This entails exploring the K-L decomposition within general Hilbert spaces, representing an engaging avenue for future research.

References

[1] S. Armaut, R. Diel, and T. Laloë. Depth level set estimation and associated risk measures. *Electronic Journal of Statistics*, 16(2):6584–6630, 2022.

[2] Michal Benko, Wolfgang Härdle, and Alois Kneip. Common functional principal components. *The Annals of Statistics*, pages 1–34, 2009.

[3] Jan Beran and Haiyan Liu. Estimation of eigenvalues, eigenvectors and scores in fda models with dependent errors. *Journal of Multivariate Analysis*, 147:218–233, 2016.

[4] Gérard Biau, Florentina Bunea, and Marten H Wegkamp. Functional classification in hilbert spaces. *IEEE Transactions on Information Theory*, 51(6):2163–2172, 2005.

[5] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[6] Claus Borggaard and Hans Henrik Thodberg. Optimal minimal neural interpretation of spectra. *Analytical chemistry*, 64(5):545–551, 1992.

[7] Ignacio Cascos and Miguel López-Díaz. On the uniform consistency of the zonoid depth. *Journal of Multivariate Analysis*, 143:394–397, 2016.

[8] Frédéric Cérou and Arnaud Guyader. Nearest neighbor classification in infinite dimension. *ESAIM: Probability and Statistics*, 10:340–355, 2006.

[9] Gerda Claeskens, Mia Hubert, Leen Slaets, and Kaveh Vakili. Multivariate functional halfspace depth. *Journal of the American Statistical Association*, 109(505):411–423, 2014.

[10] Juan A Cuesta-Albertos and Alicia Nieto-Reyes. Functional classification and the random tukey depth. practical issues. In *Combining soft computing and statistical methods in data analysis*, pages 123–130. Springer, 2010.

[11] A. Cuevas, M. Febrero, and R. Fraiman. Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22(3):481–496, 2007.

[12] Jacques Dauxois, Alain Pousse, and Yves Romain. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of multivariate analysis*, 12(1):136–154, 1982.

[13] Aurore Delaigle and Peter Hall. Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74(2):267–286, 2012.

[14] Grzegorz Dudek. Pattern-based local linear regression models for short-term load forecasting. *Electric power systems research*, 130:139–147, 2016.

[15] Subhajit Dutta and Anil K Ghosh. On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics*, 64:657–676, 2012.

[16] Frédéric Ferraty and Philippe Vieu. *Nonparametric functional data analysis: theory and practice*, volume 76. Springer, 2006.

[17] Bernhard Flury. *Common principal components & related multivariate models*. John Wiley & Sons, Inc., 1988.

[18] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability theory and related fields*, 162(3-4):707–738, 2015.

[19] Anil K Ghosh and Probal Chaudhuri. On data depth and distribution-free discriminant analysis using separating surfaces. *Bernoulli*, 11(1):1–27, 2005.

[20] Anil K Ghosh and Probal Chaudhuri. On maximum depth and related classifiers. *Scandinavian Journal of Statistics*, 32(2):327–350, 2005.

[21] I. Gijbels and S. Nagy. On a general definition of depth for functional data. *Statistical Science*, 32(4):630–639, 2017.

[22] Peter Hall and Mohammad Hosseini-Nasab. Theory for high-order bounds in functional principal components analysis. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 146, pages 225–256. Cambridge University Press, 2009.

[23] Peter Hall, Hans-Georg Müller, and Jane-Ling Wang. Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34(3):1493 – 1517, 2006.

[24] Marc Hallin, Davy Paindaveine, Miroslav Šiman, Ying Wei, Robert Serfling, Yijun Zuo, Linglong Kong, and Ivan Mizera. Multivariate quantiles and multiple-output regression quantiles: From l1 optimization to halfspace depth [with discussion and rejoinder]. *The Annals of Statistics*, pages 635–703, 2010.

[25] F.R. Hampel, E.M. Ronchetti, P. Rousseeuw, and W.A. Stahel. *Robust statistics: the approach based on influence functions*. Wiley-Interscience; New York, 1986.

[26] Lajos Horvath and Piotr Kokoszka. *Inference for functional data with applications*, volume 200. Springer Science & Business Media, 2012.

[27] Rob J Hyndman and Han Lin Shang. Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45, 2010.

[28] Gareth M James and Trevor J Hastie. Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(3):533–550, 2001.

[29] Gareth M James, Trevor J Hastie, and Catherine A Sugar. Principal component models for sparse functional data. *Biometrika*, 87(3):587–602, 2000.

[30] MC Jones and John A Rice. Displaying the important features of large collections of similar curves. *The American Statistician*, 46(2):140–145, 1992.

[31] Kari Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung: akademische Abhandlung*. Sana, 1947.

[32] Gleb Koshevoy and Karl Mosler. Zonoid trimming for multivariate distributions. *The Annals of Statistics*, 25(5):1998–2017, 1997.

[33] Thomas Laloë. A k-nearest neighbor approach for functional regression. *Statistics & Probability Letters*, 78(10):1189–1193, 2008.

[34] Tatjana Lange, Karl Mosler, and Pavlo Mozharovskyi. Fast nonparametric classification based on data depth. *Statistical Papers*, 55:49–69, 2014.

[35] Regina Y Liu, Jesse M Parelius, and Kesar Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh). *The annals of statistics*, 27(3):783–858, 1999.

[36] R.Y. Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, pages 405–414, 1990.

[37] Xiaohui Liu, Karl Mosler, and Pavlo Mozharovskyi. Fast computation of tukey trimmed regions and median in dimension p$>$ 2. *Journal of Computational and Graphical Statistics*, 28(3):682–697, 2019.

[38] S. López-Pintado and J. Romo. On the concept of depth for functional data. *Journal of the American statistical Association*, 104(486):718–734, 2009.

[39] Luciana Magnano, John W Boland, and Robin John Hyndman. Generation of synthetic sequences of half-hourly temperature. *Environmetrics: The official journal of the International Environmetrics Society*, 19(8):818–835, 2008.

[40] Ivan Mizera and Milos Volauf. Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *Journal of Multivariate Analysis*, 83(2):365–388, 2002.

[41] Karl Mosler and Richard Hoberg. Data analysis and classification with the zonoid depth. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:49, 2006.

[42] Karl Mosler and Pavlo Mozharovskyi. Choosing among notions of multivariate depth statistics. *Statistical Science*, 37(3):348–368, 2022.

[43] S. Nagy, S. Helander, G. Van Bever, L. Viitasaari, and P. Ilmonen. Flexible integrated functional depths. *Bernoulli*, 27(1):673–701, 2021.

[44] A. Nieto-Reyes and H. Battey. A topologically valid definition of depth for functional data. *Statistical Science*, 31(1):61–79, 2016.

[45] Boris N Oreshkin, Grzegorz Dudek, Paweł Pełka, and Ekaterina Turkina. N-beats neural network for mid-term electricity load forecasting. *Applied Energy*, 293:116918, 2021.

[46] Justin Petrovich and Matthew Reimherr. Asymptotic properties of principal component projections with repeated eigenvalues. *Statistics & Probability Letters*, 130:42–48, 2017.

[47] Sara Pintado and Juan Romo. Depth-based classification for functional data. *?*, 72, 11 2005.

[48] Oleksii Pokotylo, Pavlo Mozharovskyi, and Rainer Dyckerhoff. Depth and depth-based classification with r-package ddalpha. *arXiv preprint arXiv:1608.04109*, 2016.

[49] James O Ramsay and Bernard W Silverman. *Applied functional data analysis: methods and case studies*. Springer, 2002.

[50] JO Ramsay and BW Silverman. Principal components analysis for functional data. *Functional data analysis*, pages 147–172, 2005.

[51] Robert Serfling. Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:1, 2006.

[52] Carlo Sguera, Pedro Galeano, and Rosa Lillo. Spatial depth-based classification for functional data. *Test*, 23:725–750, 2014.

[53] BW Silverman and JO Ramsay. Applied functional data analysis: methods and case studies. *Journal*, pages 1–18, 2002.

[54] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, 100(470):577–590, 2005.

[55] Y. Zuo and X. He. On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *The Annals of Statistics*, 34(6):2879–2896, 2006.

[56] Y. Zuo and R. Serfling. General notions of statistical depth function. *Annals of statistics*, pages 461–482, 2000.

Appendix A. Proof of main theorems 2.6 and 2.7

To establish our convergence results, we introduce a new functional to separately handle the projection component. Let us define for $x \in \mathbb{L}^2$, for any family $v = (v_j)_{1\le j\le J}$ in $\mathbb{L}^2$ and any $P_X \in \mathcal{P}_2(\mathbb{L}^2[0,1])$,

$$F(x,v,P) := \frac{1}{1+\left(\frac{\|x-x^v-(m_X-m_X^v)\|^2}{\sum_{j=1}^{\infty}\lambda_j^{P_X}}\right)^{\beta}} D_J\left(x^{(v)}-m_X^{(v)}, P_{X^{(v)}-m_X^{(v)}}\right)$$

where

$$x^v := \sum_{j=1}^{J}\langle x, v_j\rangle\, v_j, \quad x^{(v)} := (\langle x, v_j\rangle)_{1\le j\le J}$$

$$m_X = \int_{\mathbb{L}^2[0;1]} xP(dx), \quad m_X^v := \sum_{i=1}^{J}\langle m_X, v_j\rangle\, v_j, \quad \text{and} \quad m_X^{(v)} := (\langle m_X, v_j\rangle)_{1\le j\le J}$$

It is then obvious that

$$\mathrm{PCD}^J(x,P) = F(x,(e_j)_{1\le j\le J},P) \quad \text{and} \quad \mathrm{PCD}^J(x,P_n) = F(x,(e_j^n)_{1\le j\le J},P_n),$$

where $(e_j)_{1\le j}$ is any K-L orthonormal basis associated to P and $(e_j^n)_{1\le j\le J}$ any K-L orthonormal basis associated to P_n.

We establish some regularity properties of the functional F which will be used to prove our theorems.

Lemma A.1. *Consider the set $\mathcal{O}_J(\mathbb{L}^2[0,1])$ of the orthonormal families of $\mathbb{L}^2[0,1]$ with J elements.*

*(1) Suppose that depth D_J satisfies **P-1** and **P-6**, then F is continuous as a function from $(\mathbb{L}^2[0,1])^{J+1}\times\mathcal{P}_2(\mathbb{L}^2[0,1])$ to $\mathbb{R}$ where the space $(\mathbb{L}^2)^{J+1}\times\mathcal{P}_2(\mathbb{L}^2[0,1])$ is equipped with the distance*

$$d((x,v,P),(\tilde{x},\tilde{v},\tilde{P})) := \|x-\tilde{x}\| + \sum_{j=1}^{J}\|v_j-\tilde{v_j}\| + d_{\mathscr{P}}(P,Q),$$

*$d_{\mathscr{P}}$ being the distance appearing in **P-6**.*

*(2) Suppose that the depth D_J satisfies Properties **P-1**, **P-4** and **P-6U**. Fix some probability measure $P \in \mathcal{P}_2(\mathbb{L}^2[0,1])$ and some orthnormal family $e \in \mathcal{O}_J(\mathbb{L}^2[0,1])$. Then we have for any $(\tilde{e},\tilde{P}) \in \mathcal{P}_2(\mathbb{L}^2[0,1])\times\mathcal{O}_J(\mathbb{L}^2[0,1])$ converging to (e,P),*

$$\lim_{(\tilde{e},\tilde{P})\to(e,P)}\ \sup_{x\in\mathbb{L}^2[0,1]}\left|F(x,\tilde{v},\tilde{P}) - F(x,v,P)\right| = 0.$$

*(3) Suppose that depth D_J satisfies **P-1** and Assumptions **A1** and **A2** for some $p \ge 1$ and denote by $\bar{\beta} = \beta \wedge \frac{1}{2}$.*

There is a continuous function $M : \mathcal{P}_2(\mathbb{L}^2[0,1]) \to]0;\infty[$ s.t. for any $P_X \in \mathcal{P}_2(\mathbb{L}^2[0,1])$,

$$\forall x \in \mathbb{L}^2[0,1],\ \forall e \in \mathcal{O}_J(\mathbb{L}^2[0,1]),\ F(x,e,P_X) \le \frac{M(P_X)}{1+\|x\|^{2\bar{\beta}}}.$$

For any $P_X \in \mathcal{P}_p(\mathbb{L}^2[0,1])$, there is a neighborhood $V_{P_X} \subset \mathcal{P}_p(\mathbb{L}^2[0,1])$ of P_X and a constant C_{P_X} depending only on D_J and P_X such that,

$$\forall x \in \mathbb{L}^2[0,1],\ \forall e, \tilde{e} \in \mathcal{O}_J(\mathbb{L}^2[0,1]),\ \forall P_{\tilde{X}} \in V_{P_X},$$
$$|F(x,e,P_X) - F(x,\tilde{e},P_{\tilde{X}})| \leq$$
$$C_{P_X}\Bigg(\|m_X - m_{\tilde{X}}\|^{2\bar{\beta}} + \Big|\big(\sum_j \lambda_j(P_X)\big)^{1/2} - \big(\sum_j \lambda_j(P_{\tilde{X}})\big)^{1/2}\Big|^{2\bar{\beta}} + W_p(P_{X^{(e)}}, P_{\tilde{X}^{(e)}})$$
$$+ \big(\|x\|+1\big)^{2\bar{\beta}}\Big(\sum_{j=1}^J \|e_j - \tilde{e}_j\|\Big)^{2\bar{\beta}} + \big(\|x\|+1\big)\sum_{j=1}^J \|e_j - \tilde{e}_j\|\Bigg)$$

Proof.
(1) Recall that, when Assumptions **P-1** and **P-6** are fulfilled by D_J, the depth D_J is continuous with respect to both variables x and P. As the J first eigenvalues and the functions $(x,v) \to x^v$ and $(x,v) \to x^{(v)}$ are also continuous, this gives the continuity of the functional F.
(2) Remark first that, as D_J satisfies **P-1**, the expression of F can be simplified:

$$F(x,e,P) := \frac{1}{1 + \left(\frac{\|x - x^e - (m_X - m_X^e)\|^2}{\sum_{j=1}^J \lambda_j^{P_X}}\right)^\beta} D_J\left(x^{(e)}, P_{X^{(e)}}\right)$$

Fix some $x \in \mathbb{L}^2$ and two orthonormal families e and $\tilde{e}$ in $\mathcal{O}_J(\mathbb{L}^2[0,1])$. The difference $x^e - x^{\tilde{e}}$ can be written:

$$x^e - x^{\tilde{e}} = \sum_{j=1}^J \langle x, e_j - \tilde{e}_j\rangle\, e_j + \langle x, \tilde{e}_j\rangle\,(e_j - \tilde{e}_j).$$

Using triangular inequality and Cauchy-Schwarz inequality and the fact that e and $\tilde{e}$ are two orthonormal families, we get:

$$\|x^e - x^{\tilde{e}}\| \leq 2\|x\| \sum_{j=1}^J \|e_j - \tilde{e}_j\| \tag{A.1}$$

Same kind of arguments give the bound

$$\|x^{(e)} - x^{(\tilde{e})}\| \leq \|x\| \sqrt{\sum_{j=1}^J \|e_j - \tilde{e}_j\|^2} \leq \|x\| \sum_{j=1}^J \|e_j - \tilde{e}_j\| \tag{A.2}$$

Fix a probability measure P_X in $\mathcal{P}_2(\mathbb{L}^2[0,1])$, an orthonormal family $e = (e_j)_{1\leq j\leq J} \in \mathcal{O}(\mathbb{L}^2[0;1])$, and some $\epsilon > 0$. Property **P-4** in the depth definition implies that there is a $M_1 > 0$ such that for any $x \in \mathbb{L}^2[0;1]$ if $\|x^{(e)}\| \geq M_1$,

$$D_J(x^{(e)}, P_{X^{(e)}}) \leq \epsilon \tag{A.3}$$

and, as $\|x^e\| = \|x^{(e)}\|$, there exists some value M_2 depending on e and P such that (A.3) is satisfied if $\|x^e\| \geq M_2$. Up to increasing the value of M_2, we have also, if $\|x - x^e\| \geq M_2$,

$$\frac{1}{1 + \left(\frac{\|x - x^e - (m_X - m_X^e)\|^2}{\sum_{j=1}^\infty \lambda_j}\right)^\beta} \leq \epsilon / \|D_J(\cdot, P_{X^{(e)}})\|_\infty.$$

So that if $x \in \mathbb{L}^2[0;1]$ is such that $\|x\| \geq \sqrt{2}M_2$ then $\|x^e\| \geq M_2$ or $\|x - x^e\| \geq M_2$ and thus $F(x,e,P) \leq \epsilon$. Moreover, Property **P-6U** shows that for $(\tilde{e}, \tilde{P})$ close enough to (e,P),

$$\sup_{\|x\| \geq 2M_2} F(x, \tilde{e}, \tilde{P}) \leq 2\epsilon. \tag{A.4}$$

We now bound the two differences

$$|F(x,e,P_X)-F(x,e,P_{\tilde{X}})| \text{ and } |F(x,e,P_{\tilde{X}})-F(x,\tilde{e},P_{\tilde{X}})|$$

which will directly give the result. For the sake of clarity, we introduce the following notations:

$$a=\|x-x^e-(m_X-m_X^e)\|,\ \tilde{a}=\|x-x^e-(m_{\tilde{X}}-m_{\tilde{X}}^e)\|$$
$$b=\sqrt{\sum_j \lambda_j(P_X)} \text{ and } \tilde{b}=\sqrt{\sum_j \lambda_j(P_{\tilde{X}})}\,.$$

So that for any $x\in\mathbb{L}^2[0;1]$,

$$\begin{aligned}
&|F(x,e,P_X)-F(x,e,P_{\tilde{X}})|\\
\leq&\frac{1}{1+\left(\tilde{a}/\tilde{b}\right)^{2\beta}}\left|D_J(x^{(e)},P_{X^{(e)}})-D_J(x^{(e)},P_{\tilde{X}^{(e)}})\right|+\left|\frac{1}{1+(a/b)^{2\beta}}-\frac{1}{1+\left(\tilde{a}/\tilde{b}\right)^{2\beta}}\right|D_J(x^{(e)},P_{X^{(e)}})\\
\leq&\sup_{y\in\mathbb{R}^J}\left|D_J(y,P_{X^{(e)}})-D_J(y,P_{\tilde{X}^{(e)}})\right|+\left|\frac{1}{1+(a/b)^{2\beta}}-\frac{1}{1+\left(\tilde{a}/\tilde{b}\right)^{2\beta}}\right|\sup_{y\in\mathbb{R}^J}D_J(y,P_{X^{(e)}}) \quad \text{(A.5)}
\end{aligned}$$

The application of the β-power onto the terms outlined above serves to differentiate between two scenarios: the first, where $2\beta\leq 1$, and the second, where $2\beta>1$.

- **Case $\boldsymbol{\beta\leq 1/2}$**: In this case, for any $t,s>0$, $|t^{2\beta}-s^{2\beta}|\leq|t-s|^{2\beta}$ and some simple computations give

$$\left|\frac{1}{1+(a/b)^{2\beta}}-\frac{1}{1+\left(\tilde{a}/\tilde{b}\right)^{2\beta}}\right|\leq\frac{1}{b^{2\beta}}\left(|b-\tilde{b}|^{2\beta}+|a-\tilde{a}|^{2\beta})\right)$$

- **Case $\boldsymbol{\beta>1/2}$**: In this case, the mean value theorem used with $z\in\mathbb{R}_+\mapsto z^{2\beta}$ and some simple computations show that

$$\left|\frac{1}{1+(a/b)^{2\beta}}-\frac{1}{1+\left(\tilde{a}/\tilde{b}\right)^{2\beta}}\right|\leq\frac{2\beta\max(\tilde{b}^{2\beta},b^{2\beta},1)}{b^{2\beta}}\left(|b-\tilde{b}|+|a-\tilde{a}|\right)$$

Thus, in any case, denoting $\bar{\beta}=\beta\wedge\frac{1}{2}$, we get

$$\left|\frac{1}{1+(a/b)^{2\beta}}-\frac{1}{1+\left(\tilde{a}/\tilde{b}\right)^{2\beta}}\right|\leq\frac{2\beta\max(\tilde{b}^{2\beta},b^{2\beta},1)}{b^{2\beta}}\left(|b-\tilde{b}|^{2\bar{\beta}}+|a-\tilde{a}|^{2\bar{\beta}}\right)\,.$$

Remark now that

$$|a-\tilde{a}|\leq\|(m_{\tilde{X}}-m_{\tilde{X}^e})-(m_X-m_{X^e})\|\,,$$

therefore we obtain the overall upper bound for the first term (A.5),

$$\begin{aligned}
|F(x,e,P_X)-F(x,e,P_{\tilde{X}})|\leq&\sup_{y\in\mathbb{R}^J}\left|D_J(y,P_{X^{(e)}})-D_J(y,P_{\tilde{X}^{(e)}})\right|+ \quad \text{(A.6)}\\
&2\beta\|D_J(\cdot,P_{X^{(e)}})\|_\infty\frac{\max\left(1,\left(\sum_j\lambda_j(P_X)\right)^\beta,\left(\sum_j\lambda_j(P_{\tilde{X}})\right)^\beta\right)}{\left(\sum_j\lambda_j(P_X)\right)^\beta}\times\\
&\left(\|(m_X-m_X^e)-(m_{\tilde{X}}-m_{\tilde{X}}^e)\|^{2\bar{\beta}}+\left|\left(\sum_j\lambda_j(P_X)\right)^{1/2}-\left(\sum_j\lambda_j(P_{\tilde{X}})\right)^{1/2}\right|^{2\bar{\beta}}\right)
\end{aligned}$$

Thus, the continuity of the total variance with respect to the distribution (represented by the series of all eigenvalues), the continuity of the function $x \to x^{(e)}$ and Properties **P-4** and **P-6U** lead to

$$\lim_{P_{\tilde{X}} \to P_X} \sup_{x \in \mathbb{L}^2[0;1]} |F(x, e, P_X) - F(x, e, P_{\tilde{X}})| = 0. \tag{A.7}$$

We now give a bound when the orthonormal family e varies. Recall the notation previously introduced

$$\tilde{a} = \|x - x^e - (m_{\tilde{X}} - m^e_{\tilde{X}})\| \text{ and } \tilde{b} = \sqrt{\sum_j \lambda_j(P_{\tilde{X}})}$$

and denote moreover $\tilde{\tilde{a}} = \|x - x^{\tilde{e}} - (m_{\tilde{X}} - m_{\tilde{X}^{\tilde{e}}})\|$. We easily obtain that

$$\begin{aligned}
&|F(x, e, P_{\tilde{X}}) - F(x, \tilde{e}, P_{\tilde{X}})| \\
\leq &\left| \frac{1}{1 + \left(\tilde{a}/\tilde{b}\right)^{2\beta}} - \frac{1}{1 + \left(\tilde{\tilde{a}}/\tilde{b}\right)^{2\beta}} \right| D_J(x^{(e)}, P_{\tilde{X}^{(e)}}) + \frac{1}{1 + \left(\tilde{\tilde{a}}/\tilde{b}\right)^{2\beta}} \left| D_J(x^{(e)}, P_{\tilde{X}^{(e)}}) - D_J(x^{(\tilde{e})}, P_{\tilde{X}^{(\tilde{e})}}) \right|
\end{aligned}$$

Using Property (**P-1**), we see that

$$\begin{aligned}
\left| D_J(x^{(e)}, P_{\tilde{X}^{(e)}}) - D_J(x^{(\tilde{e})}, P_{\tilde{X}^{(\tilde{e})}}) \right| &= \left| D_J(x^{(e)}, P_{\tilde{X}^{(e)}}) - D_J(x^{(e)}, P_{\tilde{X}^{(\tilde{e})} + x^{(e)} - x^{(\tilde{e})}}) \right| \\
&\leq \sup_{y \in \mathbb{R}^J} \left| D_J(y, P_{\tilde{X}^{(e)}}) - D_J(y, P_{\tilde{X}^{(\tilde{e})} + x^{(e)} - x^{(\tilde{e})}}) \right|
\end{aligned}$$

And same kind of computations as above give

$$\left| \frac{1}{1 + \left(\tilde{a}/\tilde{b}\right)^{2\beta}} - \frac{1}{1 + \left(\tilde{\tilde{a}}/\tilde{b}\right)^{2\beta}} \right| \leq \frac{2\beta \max(\tilde{b}^{2\beta}, 1)}{\tilde{b}^{2\beta}} |\tilde{a} - \tilde{\tilde{a}}|^{2\bar{\beta}}.$$

where $\bar{\beta} = \beta \wedge \frac{1}{2}$. Remarking now that $|\tilde{a} - \tilde{\tilde{a}}| \leq \|x^e - x^{\tilde{e}}\| + \|m_{\tilde{X}^e} - m_{\tilde{X}^{\tilde{e}}}\|$, we get

$$\begin{aligned}
&|F(x, e, P_{\tilde{X}}) - F(x, \tilde{e}, P_{\tilde{X}})| \\
\leq &2\beta \max\left(1 \vee \left(\sum_j \lambda_j(P_{\tilde{X}})\right)^{-1/2}\right) \sup_{y \in \mathbb{R}^J} D_J(y, P_{\tilde{X}^{(e)}}) \left(\|x^e - x^{\tilde{e}}\| + \|m_{\tilde{X}^e} - m_{\tilde{X}^{\tilde{e}}}\|\right)^{2\bar{\beta}} \\
&+ \sup_{y \in \mathbb{R}^J} \left| D_J(y, P_{\tilde{X}^{(e)}}) - D_J(y, P_{\tilde{X}^{(\tilde{e})} + x^{(e)} - x^{(\tilde{e})}}) \right| \\
\leq &2^{2\bar{\beta}+1} \beta \max\left(1 \vee \left(\sum_j \lambda_j(P_{\tilde{X}})\right)^{-1/2}\right) \sup_{y \in \mathbb{R}^J} D_J(y, P_{\tilde{X}^{(e)}}) \left(\|x\| + \|m_{\tilde{X}}\|\right)^{2\bar{\beta}} \left(\sum_{j=1}^J \|e_j - \tilde{e}_j\|\right)^{2\bar{\beta}} \\
&+ \sup_{y \in \mathbb{R}^J} \left| D_J(y, P_{\tilde{X}^{(e)}}) - D_J(y, P_{\tilde{X}^{(\tilde{e})} + x^{(e)} - x^{(\tilde{e})}}) \right|. \qquad \text{(A.8)}
\end{aligned}$$

where inequality (A.1) was used to obtain the last line.

Now, bound (A.8) and bound (A.4) prove that

$$\limsup_{(\tilde{e}, P_{\tilde{X}}) \to (e, P_X)} \sup_{x \in \mathbb{L}^2} \left| F(x, e, P_{\tilde{X}}) - F(x, \tilde{e}, P_{\tilde{X}}) \right| = 0.$$

With (A.7), this gives the required result.

(3) Once again, we have to control the two differences:

$$|F(x, \tilde{e}, P_X) - F(x, \tilde{e}, P_{\tilde{X}})| \text{ and } |F(x, e, P_X) - F(x, \tilde{e}, P_X)|$$

which will directly give the global bound of the lemma. Therefore, the bound (A.6) and Assumptions **A1** and **A2** and the fact that the sum of the eigenvalues is a continuous function

of the distribution show that for some constant M_{P_X} depending only on D_J and P_X, for any $P_{\tilde{X}}$ in some neighborhood of P_X we have

$$|F(x,e,P_X) - F(x,e,P_{\tilde{X}})| \leq$$
$$M_{P_X}\left(\|m_X - m_{\tilde{X}}\|^{2\bar{\beta}} + \left|\left(\sum_j \lambda_j(P_X)\right)^{1/2} - \left(\sum_j \lambda_j(P_{\tilde{X}})\right)^{1/2}\right|^{2\bar{\beta}}\right) + W_p(P_{X^{(e)}}, P_{\tilde{X}^{(e)}})\Bigg)$$

For the second difference, (A.8) and Assumptions **A1** and **A2** show that for some constant M_{P_X} depending only on D_J and P_X, for any $P_{\tilde{X}}$ in some neighborhood of P_X we have

$$|F(x,e,P_{\tilde{X}}) - F(x,\tilde{e},P_{\tilde{X}})| \leq \tag{A.9}$$
$$M_{P_X}\left(\left(\|x\| + \|m_{\tilde{X}}\|\right)^{2\bar{\beta}} \sum_{j=1}^{J} \|e_j - \tilde{e}_j\|\right)^{2\bar{\beta}} + W_p(P_{\tilde{X}^{(e)}}, P_{\tilde{X}^{(\tilde{e})}+x^{(e)}-x^{(\tilde{e})}})\Bigg).$$

Using Minkowski inequality, it is easily checked that

$$W_p(P_{\tilde{X}^{(e)}}, P_{\tilde{X}^{(\tilde{e})}+x^{(e)}-x^{(\tilde{e})}}) \leq \|x^{(e)} - x^{(\tilde{e})}\| + W_p(P_{\tilde{X}^{(e)}}, P_{\tilde{X}^{(\tilde{e})}})\,.$$

And Hölder's inequality gives,

$$W_p(P_{\tilde{X}^{(e)}}, P_{\tilde{X}^{(\tilde{e})}}) \leq \mathbb{E}[\|(<\tilde{X}, e_j - \tilde{e}_j>)_{j\leq J}\|^p]^{\frac{1}{p}} \leq \mathbb{E}[\|\tilde{X}\|^p]^{\frac{1}{p}} \sqrt{\sum_{j=1}^{J} \|e_j - \tilde{e}_j\|^2}$$
$$\leq \mathbb{E}[\|\tilde{X}\|^p]^{\frac{1}{p}} \sum_{j=1}^{J} \|e_j - \tilde{e}_j\|.$$

As $\|m_{\tilde{X}}\| \leq \mathbb{E}[\|\tilde{X}\|]$, the bounds (A.1) and (A.2) and the above calculations, all give

$$|F(x,e,P_{\tilde{X}}) - F(x,\tilde{e},P_{\tilde{X}})|$$
$$\leq M_{P_X}\left(\left(\|x\| + \mathbb{E}[\|\tilde{X}\|]\right)^{2\bar{\beta}} \sum_{j=1}^{J} \|e_j - \tilde{e}_j\|\right)^{2\bar{\beta}} + \left(\|x\| + \mathbb{E}[\|\tilde{X}\|^p]^{\frac{1}{p}}\right) \sum_{j=1}^{J} \|e_j - \tilde{e}_j\|\Bigg).$$

Moreover, if $P_{\tilde{X}}$ is close enough to P_X,

$$\|x\| + \mathbb{E}[\|\tilde{X}\|] \leq (1 + \mathbb{E}[\|X\|])(\|x\| + 1)$$

and

$$\|x\| + \mathbb{E}[\|\tilde{X}\|^p]^{\frac{1}{p}} \leq (1 + \mathbb{E}[\|X\|^p]^{\frac{1}{p}})(\|x\| + 1)\,.$$

Hence, the two inequalities for both terms of interest $|F(x,e,P_X) - F(x,e,P_{\tilde{X}})|$ and $|F(x,e,P_{\tilde{X}}) - F(x,\tilde{e},P_{\tilde{X}})|$ lead to the desired bound of the third point of the lemma.

Finally, the existence of a positive continuous function M such that

$$F(x,e,P_X) \leq \frac{M(P_X)}{1 + \|x\|^{2\beta}}$$

is a direct consequence of **A1**, and of the equality $\|x^e\| = \|x^{(e)}\|$. □

With these new tools, we can now prove our convergence theorems.

Proof of Theorem 2.6 and 2.7. The proofs of both Theorem 2.6 and 2.7 rely on the control of the difference

$$\Delta_n(x) = \left|\mathrm{PCD}^J(x,P) - \mathrm{PCD}^J(x,P_n)\right|.$$

Remark first that any K-L basis associated to P_n does not necessarily converge to its theoretical counterpart, a K-L orthonormal basis associated to P when n goes to infinity as there is

no unicity of the eigen-vectors basis. Thus, the first step is to construct a consistent empirical basis from the previously fixed K-L basis associated to P, $(e_j)_{1\leq j\leq J}$. Precisely, for $n\geq 1$ we want to construct a sequence of orthonormal basis $(e_j^n)_{j\leq J}$ of the space generated by the first J empirical eigenvalues, counted with multiplicity, such that

$$\|e_j^n - e_j\| \xrightarrow[n\to\infty]{a.s.} 0, \quad j=1,\dots,J.$$

We underline that these basis will not necessarily be K-L ones, but, as explained in Proposition 2.5, it has no impact on the value of the empirical PCD depth.

For the sake of simplicity, we consider first the case $\lambda_1=\lambda_2=\ldots=\lambda_J>\lambda_{J+1}$ that is λ_1 is of full multiplicity J with associated theoretical eigenspace $L_J := E_{\lambda_1} := \mathrm{Span}(e_1,\dots,e_J)$ and $(e_j)_{1\leq j\leq J}$ is the associated K-L basis. Let $\lambda_1^n\geq\ldots\geq\lambda_J^n$ be the empirical eigenvalue(s) associated to λ_1, which may be different or not. Recall the notation

$$L_J^n := \dotplus_{j=1}^{J} E_{\lambda_j^n}$$

corresponding to the empirical eigenspaces associated to $\lambda_j^n, j=1,\dots,J$, whether these empirical eigenvalues are equal or not.

It is important to underline the fact that assumption $\lambda_{J+1}<\lambda_J$ implies that $\lambda_{J+1}^n<\lambda_J^n$ for n large enough as $\lambda_{J+1}^n \xrightarrow[n\to\infty]{a.s.} \lambda_{J+1}$ and $\lambda_J^n \xrightarrow[n\to\infty]{a.s.} \lambda_J$. Hence, the empirical eigenspaces are not cut through for large n and the estimated depth does not depend on the choice of the basis (see Proposition 2.5).

Now, the idea is to project all true eigenfunctions $e_1,\dots,e_J\in L_J$ over the empirical spaces L_J^n, that is, we set

$$v_j^n := \mathrm{P}_J^n(e_j), \quad j=1,\dots,J$$

where P_J^n is the projector over L_J^n.

Lemma A.2. *Almost surely, for n large enough, the family $(v_j^n)_{j=1,\dots,J}$ is linearly independent and therefore is a basis of L_J^n.*

Proof. Let $(a_j^n)_{j\leq J}\in\mathbb{R}^J$ be scalars such that $\sum_{j=1}^J a_j^n v_j^n = 0$. This is equivalent to having

$$\mathrm{P}_J^n\Big(\sum_{j=1}^J a_j^n e_j\Big) = 0 \text{ or } (\mathrm{P}_J^n - \mathrm{P}_J)\Big(\sum_{j=1}^J a_j^n e_j\Big) + \sum_{j=1}^J a_j^n e_j = 0.$$

Now, we have

$$\|\sum_{j=1}^J a_j^n e_j\| = \|(\mathrm{P}_J^n-\mathrm{P}_J)(\sum_{j=1}^J a_j^n e_j)\| \leq \|\mathrm{P}_J^n-\mathrm{P}_J\|_{\mathrm{op}} \sum_{j=1}^J |a_j^n|.$$

According to Proposition 3 of [12], P_J^n converges almost surely to its theoretical counterpart P_J w.r.t. the Hilbert Schmidt norm and thus w.r.t. to the operator norm $\|\cdot\|_{op}$. Therefore, for n large enough,

$$\|\sum_{j=1}^J a_j^n e_j\| \leq \frac{1}{2}\|\sum_{j=1}^J a_j^n e_j\|$$

and the sum is necessarily equal to zero. The family $(e_j)_{j\leq J}$ being linearly independent, this implies that $a_j^n=0$ for all $j=1,\dots,J$ which proves the result of the lemma. □

Starting from the newly constructed basis $(v_1^n,\dots,v_J^n)$ (see Lemma A.2) we may obtain an orthonormal basis $(e_1^n,\dots,e_J^n)$ using the Gram-Schmidt process: $e_1^n = v_1^n/\|v_1^n\|$, $e_2^n = u_2^n/\|u_2^n\|$

where $u_2^n := v_2^n - \langle v_2^n, e_1^n\rangle\, e_1^n$, ... and $e_j^n = u_j^n/\|u_j^n\|$ with $u_j^n := v_j^n - \sum_{k=1}^{j-1} \left\langle v_j^n, e_k^n \right\rangle e_k^n$, $2 \le j \le J$.

Lemma A.3. *Under notations and assumptions of Theorem 2.6, we have for n large enough*

$$\|e_j^n - e_j\| \le 5^j \|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}}, \quad j = 1, \dots, J.$$

In particular, it holds

$$\|e_j^n - e_j\| \xrightarrow[n\to\infty]{a.s.} 0, \quad j = 1, 2, \dots, J.$$

Proof of Lemma A.3. Fix $1 \le j \le J$. We prove the lemma by induction on $j \le J$. It is simple to prove that

$$\|e_1^n - e_1\| \le 4\|\mathrm{P}_J^n - \mathrm{P}_J\|.$$

Suppose now that

$$\forall k < j, \quad \|e_k^n - e_k\| \le (5^k - 1)\|\mathrm{P}_J^n - \mathrm{P}_J\|.$$

Since $e_j \in L_J$, $\mathrm{P}_J(e_j) = e_j$, and recalling the definition of u_j^n, it holds

$$\begin{aligned}\|u_j^n - e_j\| &= \|v_j^n - \sum_{k=1}^{j-1} \langle \mathrm{P}_J^n(e_j), e_k^n\rangle\, e_k^n - \mathrm{P}_J(e_j)\| \\ &\le \|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}} + \sum_{k=1}^{j-1} |\,\langle \mathrm{P}_J^n(e_j), e_k^n\rangle\,|\end{aligned}$$

where the last inequality holds from having $\|e_k^n\| = 1$ by construction. Now, notice that for any $1 \le k \le j-1$, $\langle e_j, e_k\rangle = 0$, therefore

$$\begin{aligned}|\,\langle \mathrm{P}_J^n(e_j), e_k^n\rangle\,| &= |\,\langle \mathrm{P}_J^n(e_j) - \mathrm{P}_J(e_j), e_k^n\rangle + \langle e_j, e_k^n - e_k\rangle + \langle e_j, e_k\rangle\,| \\ &\le \|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}} + \|e_k^n - e_k\|.\end{aligned}$$

Particularly, with the previous calculations, it yields

$$\begin{aligned}\|u_j^n - e_j\| &\le j\|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}} + \sum_{k=1}^{j-1} \|e_k^n - e_k\| \\ &\le \|\mathrm{P}_J^n - \mathrm{P}_J\|\Big(1 + \sum_{k=1}^{j-1} 5^k\Big) = \frac{5^j - 1}{4}\|\mathrm{P}_J^n - \mathrm{P}_J\|\end{aligned}$$

Coming back to e_j^n, we get

$$\begin{aligned}\|e_j^n - e_j\| &= \frac{1}{\|u_j^n\|}\|u_j^n - e_j\|u_j^n\|\| \le \frac{1}{\|u_j^n\|}\left(\|u_j^n - e_j\| + |1 - \|u_j^n\||\right) \\ &\le \frac{2}{\|u_j^n\|}\|u_j^n - e_j\| \le \frac{5^j - 1}{2\|u_j^n\|}\|\mathrm{P}_J^n - \mathrm{P}_J\|\end{aligned}$$

Recall further that the operator norm is dominated by the Hilbert-Schmidt (H-S) norm. Thus, from Proposition 3 of [12] describing almost-sure convergence of the empirical projector P_J^n to its true counterpart P_J w.r.t. the H-S norm, we derive its convergence w.r.t. the operator norm, meaning

$$\|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}} \xrightarrow[n\to\infty]{a.s.} 0.$$

Thus, the previous computations show that $\|u_j^n - e_j\|$ goes almost surely to 0 and for large n, $\|u_j^n\| \ge |1 - \|u_j^n - e_j\|| \ge 1/2$ for n large enough. That way, all the above will imply the desired result of Lemma A.3. □

Accordingly, the proof of Lemma A.3 still applies when the eigenvalue λ_1 is not necessarily of full multiplicity. Indeed, by concatenation, the Gram-Schmidt procedure yields in the same way an overall consistent orthonormal basis $(e_j^n)_{j=1,\dots,J}$ of L_J^n satisfying the same inequality.

For the seek of clarity, we introduce now some notations using the basis $(e_j^n)_{j=1,\dots,J}$ as follows:

$$x_n^J = \sum_{j=1}^{J} \langle x, e_j^n \rangle e_j^n, \ \mathbf{x}_n^{(J)} := \langle x, e_j^n \rangle_{j=1,\dots,J} \in \mathbb{R}^J \text{ and } \mathbf{X}_n^{(J)} := \langle X, e_j^n \rangle_{j=1,\dots,J}.$$

and the two empirical distributions on $\mathbb{R}^J$ which will appear in our proof:

$$P_{n,\mathbf{X}^{(J)}} := \frac{1}{n}\sum_{i=1}^{n} \delta_{(\langle X_i, e_j \rangle)_{j=1,\dots,J}} \ P_{n,\mathbf{X}_n^{(J)}} := \frac{1}{n}\sum_{i=1}^{n} \delta_{(\langle X_i, e_j^n \rangle)_{j=1,\dots,J}}.$$

The isometry-invariance property of Proposition 2.5 leads to the following expression for $\mathrm{PCD}(x, P_n)$,

$$\forall x \in \mathbb{L}^2[0,1], \ \mathrm{PCD}(x, P_n) = \frac{1}{1 + \left(\frac{\|x - x_n^J\|^2}{\sum_{j=1}^{\infty} \lambda_j^n}\right)^{\beta}} D_J\left(\mathbf{x}_n^{(J)}, P_{n,\mathbf{X}_n^{(J)}}\right).$$

So, we write for any $x \in \mathbb{L}^2[0,1]$, using the functional F:

$$\Delta_n(x) := \left|\mathrm{PCD}^J(x, P) - \mathrm{PCD}^J(x, P_n)\right| = |F(x, e, P) - F(x, e_n, P_n)|$$

The simple convergence of $\Delta_n(x)$ is direct as F is continuous (Point (1) of Lemma A.1), P_n converges to P and e_n converges to e. The uniform convergence of $\sup_{x \in \mathbb{L}^2} \Delta_n(x)$ to 0 is, meanwhile, a direct consequence of Point (2) of Lemma A.1.

We finish with the proof of Theorem 2.7. According to Point (3) of Lemma A.1, as $P_{n,X}$ converges to P_X , there is some constant C_{P_X} depending on D_J and P_X, such that almost surely for n large enough, for any $x \in \mathbb{L}^2$

$$\begin{aligned}\Delta_n(x) \leq C_{P_X} \Bigg(& \|m_X - \sum_{i=1}^{n} X_i/n\|^{2\bar{\beta}} + \left|\left(\sum_j \lambda_j\right)^{1/2} - \left(\sum_j \lambda_j^n\right)^{1/2}\right|^{2\bar{\beta}} \Big) \\ & + \left(\|x\| + 1\right)^{2\bar{\beta}} \Big(\sum_{j=1}^{J} \|e_j - e_j^n\| \Big)^{2\bar{\beta}} + \left(\|x\| + 1\right) \sum_{j=1}^{J} \|e_j - e_j^n\| \\ & + W_p(P_{\mathbf{X}^{(J)}}, P_{n,\mathbf{X}^{(J)}}) \Bigg)\end{aligned}$$

According to Lemma A.3, for some constant depending only on J,

$$\sum_{j=1}^{J} \|e_j - e_j^n\| \leq C_J \|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}}$$

and $\lim_{n\to\infty} \|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}} = 0$. Thus, as $2\bar{\beta} \leq 1$, for n large enough,

$$\|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}} \leq \|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}}^{2\bar{\beta}}.$$

Moreover, the polynomial decrease of the functional F (hence, that of depth D_J) of Point (3) in Lemma A.1:

$$F(x, e, P_X) \leq \frac{M(P_X)}{1 + \|x\|^{2\beta}},$$

the convergence of P_n to P_X and Assumption **A2** show that, for $\|x\| > \left(\|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}}\right)^{-\bar{\beta}/(\bar{\beta}+\beta)}$ (that is when $\|x\|$ is large),

$$\Delta_n(x) \leq C_{P_X} \|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}}^{2\beta\bar{\beta}/(\bar{\beta}+\beta)}.$$

All these points taken together prove that a.s. for n large enough, up to a change in the constant C_{P_X},

$$\forall x \in \mathbb{L}^2,\ \Delta_n(x) \le C_{P_X}\Bigg(\|m_X - \sum_{i=1}^n X_i/n\|^{2\bar\beta} + \big|\big(\sum_j \lambda_j\big)^{1/2} - \big(\sum_j \lambda_j^n\big)^{1/2}\big|^{2\bar\beta}\Big)$$
$$+\big(1+\|x\|+\|x\|^{2\bar\beta}\big)\|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}}^{2\bar\beta} + W_p(P_{\mathbf{X}^{(J)}}, P_{n,\mathbf{X}^{(J)}})\Bigg) \tag{A.10}$$

and

$$\sup_{x\in\mathbb{L}^2}\Delta_n(x) \le C_{P_X}\Bigg(\|m_X - \sum_{i=1}^n X_i/n\|^{2\bar\beta} + \big|\big(\sum_j \lambda_j\big)^{1/2} - \big(\sum_j \lambda_j^n\big)^{1/2}\big|^{2\bar\beta}\Big)$$
$$+\|\mathrm{P}_J^n - \mathrm{P}_J\|_{\mathrm{op}}^{2\beta\bar\beta/(\bar\beta+\beta)} + W_p(P_{\mathbf{X}^{(J)}}, P_{n,\mathbf{X}^{(J)}})\Bigg) \tag{A.11}$$

Fix some $\epsilon > 0$. Propositions 7 in [12] shows that $n^{\frac12-\epsilon}\|\,\mathrm{P}^J - \mathrm{P}_n^J\,\|$ converges in probability to 0 when n goes to infinity, while by the central limit theorem $n^{\frac12-\epsilon}\big|(\sum_j \lambda_j) - (\sum_j \lambda_j^n)\big|$ converges in probability to 0 when n goes to infinity and the same is true for $n^{\frac12-\epsilon}\|\overline{X}_n - m_X\|$. To deal with the term $W_p(P_{\mathbf{X}^{(J)}}, P_{n,\mathbf{X}^{(J)}})$, we use concentration inequalities for the empirical measure in p-Wasserstein distance, $p \ge 1$, presented in the paper of Fournier and Guillin [18]. More precisely, using Theorem 2 in [18], under the assumptions of Theorem 2.7 and in particular the existence of a given q-th moment, with $q > 2p$, we have for any $n \ge 1$, any $t > 0$ and any $\eta \in (0, q)$,

$$\mathbb{P}\left(W_p\left(P_{\mathbf{X}^{(J)}}, P_{n,\mathbf{X}^{(J)}}\right) \ge t\right) \le a_n(t)\mathbf{1}_{t\le 1} + Cn(nt)^{-\frac{q-\eta}{p}}$$

where

$$a_n(t) = C\begin{cases} \exp(-cnt^2) & \text{if } p > J/2 \\ \exp(-cn(t/\log(2+1/t))^2) & \text{if } p = J/2 \\ \exp(-cnt^{\frac{J}{p}}) & \text{if } p \in (0, J/2). \end{cases}$$

In the above, the positive constants C and c depend only on J and $\mathrm{e}\,P_{\|\mathbf{X}^{(J)}\|^q}$. Using this result with $\eta = q/2 - p$ and

$$t_n = \delta\begin{cases} n^{-1/2+\epsilon} & \text{if } p \ge J/2 \\ n^{-p/J+\epsilon} & \text{if } p < J/2, \end{cases}$$

for any $\delta > 0$, we easily obtain that $n^{\left(\frac12\wedge\frac{p}{J}\right)-\epsilon}W_1\left(P_{\mathbf{X}^{(J)}}, P_{n,\mathbf{X}^{(J)}}\right)$ converges to 0 in probability for any ϵ.

These convergences combined to bounds (A.10) and (A.11), prove that for any $\epsilon > 0$, for any $x \in \mathbb{L}^2$,

$$n^{\left(\frac12\wedge\frac{p}{J}\wedge\beta\right)-\epsilon}\Delta_n(x)$$

converges in probability to 0 and

$$n^{\left(\frac{\beta}{1+2\beta}\wedge\frac{p}{J}\right)-\epsilon}\sup_{x\in\mathbb{L}^2}\Delta_n(x)$$

converges in probability to 0. □

Appendix B. Some depth notions

Many depth notions have been proposed in the literature, however, in this section, we review some of them in order to recall their definitions (the reader may refer to [56], [42] and references therein).

- *Mahalanobis depth.* The Mahalanobis depth is given by:

$$\mathrm{MHD}(x, P_{\mathbf{X}}) = \left(1 + d^2_{\Sigma(\mathbf{X})}(x, \mu(\mathbf{X}))\right)^{-1}, \tag{B.1}$$

  where $d^2_{\Sigma(\mathbf{X})}(x, \mu(\mathbf{X})) := \|x - \mu(\mathbf{X})\|^2_{\Sigma(\mathbf{X})} := (x - \mu(\mathbf{X}))^\top \Sigma(\mathbf{X})^{-1} (x - \mu(\mathbf{X}))$ is the *Mahalanobis distance* (1936).
- $\mathbb{L}^p$ *depth.* For $x \in \mathbb{R}^J$ and a r.v $\mathbf{X} \in \mathbb{R}^J$,

$$D_{\mathbb{L}^p}(x, P) = (1 + \mathbb{E}(\|x - \mathbf{X}\|_p))^{-1}, \tag{B.2}$$

  $1 \leq p < \infty$. The depth $D_{\mathbb{L}^p}$ measures the outlyingness of a point w.r.t the (multivariate) $\mathbb{L}^p$-distance. For $p = 2$, the depth is also mentioned as Euclidean depth.
- *Projection depth.*

$$D_{\mathrm{Proj}}(x, P_n) = \left(1 + \sup_{\|u\|=1} \frac{|\langle u, x\rangle - \mathrm{Med}(\langle u, \mathbf{X}\rangle)|}{\mathrm{MAD}(\langle u, \mathbf{X}\rangle)}\right)^{-1}, \tag{B.3}$$

  where Med(Y) denotes the median of a univariate r.v. Y, while MAD(Y) = Med(|Y - Med(Y)|) is its median absolute deviation from the median.
- *Halfspace depth.*

$$D_{\mathrm{HS}}(x, P) = \inf_{v \in \mathbb{R}^J,\ \|v\|=1} P\big(\langle v, \mathbf{X} - x\rangle \geq 0\big), \tag{B.4}$$

  where $\mathbf{X}$ denotes a random variable with distribution P.
- *Simplicial depth.*

$$D_{\mathrm{SD}}(x, P) = \mathbb{P}(x \in S[\mathbf{X}_1, \ldots, \mathbf{X}_{J+1}]), \tag{B.5}$$

  where $\mathbf{X}_1, \ldots, \mathbf{X}_{J+1}$ is a $(J+1)$-sample with distribution P and $S[\mathbf{X}_1, \ldots, \mathbf{X}_{J+1}]$ is the convex hull (or simplex based on these observations).
- *Simplicial Volume depth.*

$$D^\alpha_{\mathrm{SimpVol}}(x, P) := \left(1 + \mathbb{E}\left[\left(\frac{\Delta(S[x, \mathbf{X}_1, \ldots, \mathbf{X}_J])}{\sqrt{\det(\Sigma_{\mathbf{X}})}}\right)^\alpha\right]\right)^{-1}, \quad \alpha > 0, \tag{B.6}$$

  where Σ_X denotes the covariance matrix of $\mathbf{X} \sim P$ and Δ is the J-dimensional volume.
- *Zonoid depth[32].* For $0 < \alpha \leq 1$,

$$D^\alpha_{\mathrm{Zon}}(P) = \{\mathbb{E}[\mathbf{X}\, g(\mathbf{X})] \mid g : \mathbb{R}^J \mapsto [0, 1/\alpha] \text{ measurable and } \mathbb{E}[g(\mathbf{X})] = 1\}$$

  is the zonoid α-region of $\mathbf{X}$. For $\alpha = 0$ set $D^0_{\mathrm{Zon}} = \mathbb{R}^J$. The zonoid depth is defined as

$$D_{\mathrm{Zon}}(x, P) = \sup\{\alpha : x \in D^\alpha_{\mathrm{Zon}}(P)\}. \tag{B.7}$$

  The empirical counterpart of the zonoid depth is then given by:

$$D_{\mathrm{Zon}}(x, P_n) = \sup\left\{\alpha : \alpha\lambda_i \leq 1/n, x = \sum_{i=1}^n \lambda_i \mathbf{X}_i, \sum_{i=1}^n \lambda_i = 1, \lambda_i \geq 0 \forall i\right\}.$$

- *Spatial depth.* The spatial depth is defined as

$$D_{\mathrm{Spa}}(x, P) = 1 - \left\|\mathbb{E}\left[\frac{x - \mathbf{X}}{\|x - \mathbf{X}\|}\right]\right\|, \tag{B.8}$$

  with the covention $0/0 = 0$.

Appendix C. Real temperature dataset

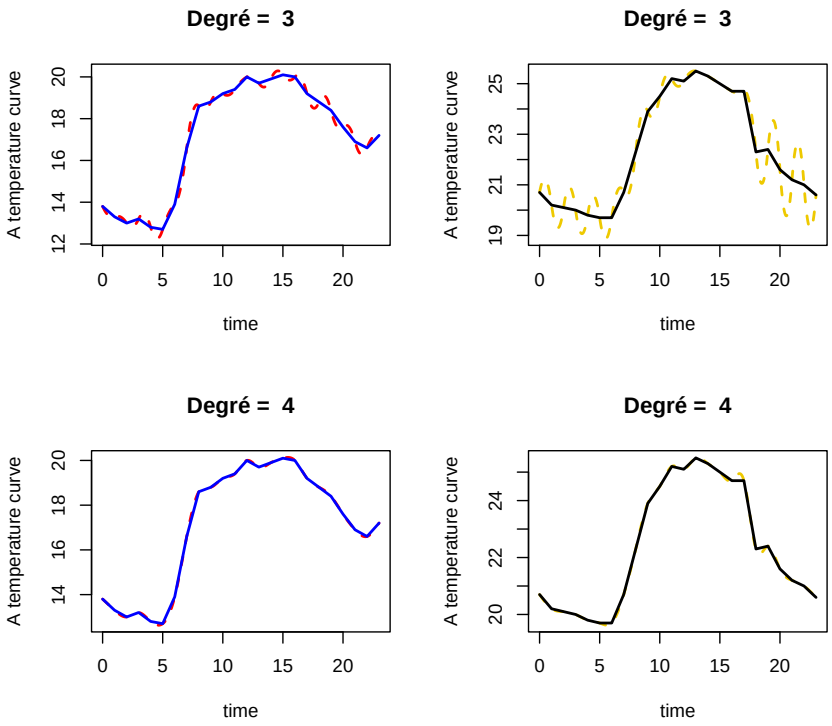

Figure 9. B-spline approximation of two temperature curves (solid blue and solid black lines) with degree 3 and 4 resp. The dashed lines corresponds to the B-spline approximation of the solid line with according degree.

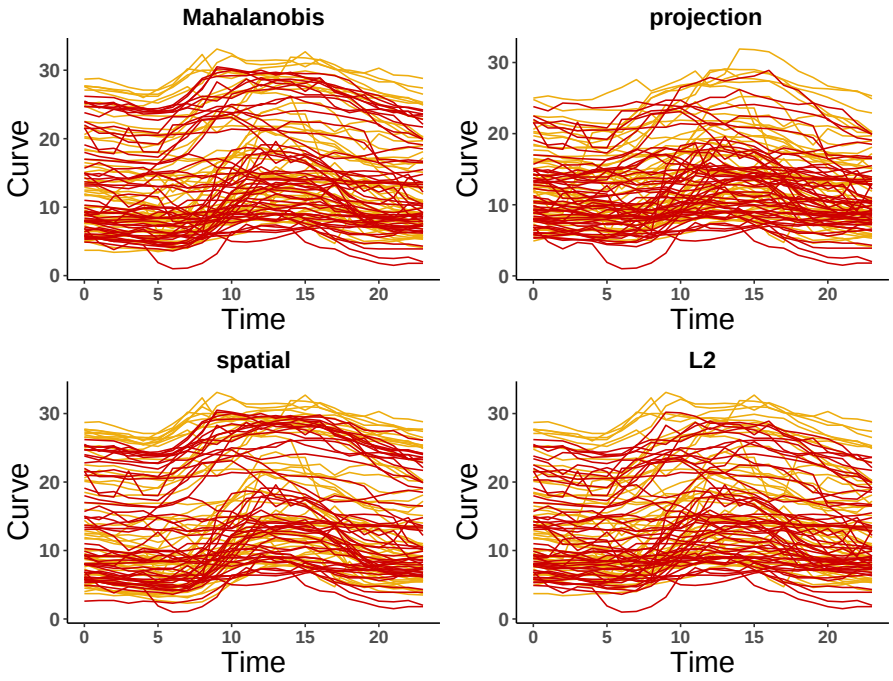

FIGURE 10. Functional 10% - PCD$_J$ level sets of the least deep temperature curves, $J = 2$ and $\beta = 1$, for two overlaying periods: the period of 1993-1997 (red curves) and 2018-2022 (golden curves).

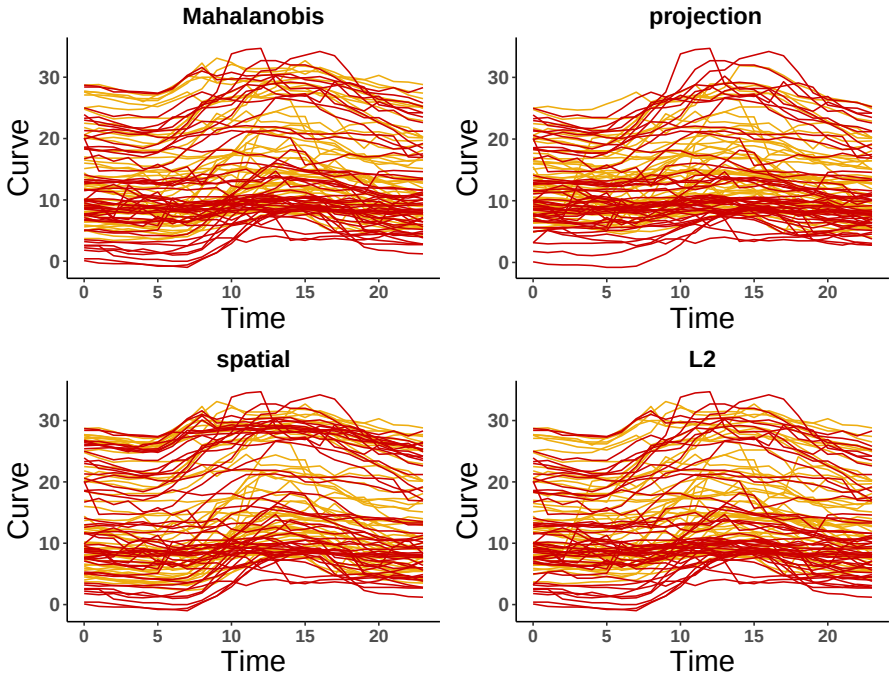

FIGURE 11. Functional 10% - PCD$_J$ level sets of the least deep temperature curves, $J = 2$ and $\beta = 1$, for two overlaying periods: the period of 2008-2012 (red curves) and 2018-2022 (golden curves).