Statistical depth in abstract metric spaces

Gery Geenens 1 · Alicia Nieto-Reyes 2 · Giacomo Francisci 2, 3

Received: 12 January 2022 / Accepted: 31 January 2023 / Published online: 23 February 2023 © The Author(s) 2023

Abstract
The concept of depth has proved very important for multivariate and functional data analysis, as it essentially acts as a surrogate for the notion of ranking of observations which is absent in more than one dimension. Motivated by the rapid development of technology, in particular the advent of ‘Big Data’, we extend here that concept to general metric spaces, propose a natural depth measure and explore its properties as a statistical depth function. Working in a general metric space allows the depth to be tailored to the data at hand and to the ultimate goal of the analysis, a very desirable property given the polymorphic nature of modern data sets. This flexibility is thoroughly illustrated by several real data analyses.

Keywords Statistical depth · Metric space · Functional Data Analysis · Lens depth · Symbolic data analysis

1 Introduction
Huge parts of statistical theory, especially its nonparametric side, heavily rely on the notion of ranks, see for instance Gibbons and Chakraborti (2010). However, ranks are not well defined in a multivariate framework as there exists no natural ordering in more than one dimension. This fact motivated Tukey (1975) to introduce the notion of statistical depth as a surrogate for ‘multivariate ranks’. Concretely, a depth is a measure of how central (or how outlying) a given point is with respect to a multivariate probability distribution. Zuo and Serfling (2000), following some earlier considerations in Liu (1990), formulated the properties that a valid depth measure should satisfy. Since then, depth-based procedures have proved very important tools for robust multivariate statistical analyses, e.g. see Liu et al. (1999), Li and Liu (2004, 2008) or Zuo (2021). Serfling (2006) and Mosler (2013) offer excellent short reviews of the ideas surrounding the concept of depth, while Hallin et al. (2021) recently shed new light on the problem of ‘multivariate ranks’.

The early 21st century has also seen such technological progress in recording devices and memory capacity, that any spatio-temporal phenomenon can now be recorded essentially in continuous time or space, giving rise to ‘functional’ random objects. As a result, a solid theory for Functional Data Analysis (FDA) has been developed as well, allowing the extension of most of the classical problems of statistical inference from the multivariate context to the inherently infinite-dimensional functional case. In particular, functional versions of statistical depth have been investigated (Fraiman and Muniz 2001; Cuevas et al. 2007; López-Pintado and Romo 2009; Dutta et al. 2011; López-Pintado and Romo 2011; Sguera et al. 2014; Chakraborty and Chaudhuri 2014; Hlubinka et al. 2015; Nieto-Reyes and Battey 2021; Nieto-Reyes et al. 2021). It is worth noting that an infinite-dimensional environment implies specific theoretical and practical challenges, making the extension from ‘multivariate’ to ‘functional’ a non-trivial one (Nieto-Reyes and Battey 2016).

In this paper, we carry on with this gradual extension process by defining the statistical depth for complex random objects living in abstract metric spaces. Again, this extension is motivated by the rapid development of technology. Indeed, this is the ‘Big Data’ era, in which digital data is recorded everywhere, all the time. The information that this huge amount of data contain may enable next-generation
scientific breakthroughs, drive business forward or hold gov-
ernments accountable. However, this is conditional on the
existence of a statistical toolbox suitable for such Big Data,
the profusion and nature of which inducing commensurate
challenges. Indeed those data consist of objects as various as
high-dimensional/infinite-dimensional vectors, matrices or
functions representing images, shapes, movies, texts, hand-
writing or speech (to cite a few); and live streaming series
thereof—this is often summarised as ‘3V’ (Volume, Variety
and Velocity).

Mainstream statistical techniques often fall short for
analysing such complex mathematical objects. Yet, it remains
true that any statistical analysis requires a sense of how close
two instances of the object of interest are to one another. It
is then only natural to assume that they live in a space
where distances can be defined—that is, in a certain
metric space (Snášel et al. 2017). This motivates the need for
a statistical depth defined in an abstract metric space, as
recently acknowledged by Dai and Lopez-Pintado (2022),
who extended Tukey’s halfspace depth to such general set-
ing. Our proposal of a ‘metric depth’ continues that line of
research.

The idea that the concept of multivariate statistical depth
could be extended to general non-Euclidean settings can
be traced back to Carrizosa (1996, section 3.1). Later, Li
et al. (2011) were considering a depth-based procedure
for analysing abundance data, which are typically high-
dimensional discrete data with many observed 0’s. Because
of that particular structure, the classical Euclidean distance
is not optimal for quantifying (dis)similarities between ob-
servations, and analysts in the field usually prefer more specific
metrics such as the Bray-Curtis distance1 (Bray and Curtis
1957). In consequence, inspired by earlier works by Maa et al.
(1996) and Bartoszynski et al. (1997), Li et al. (2011) devised
a depth measure which could allow the proximity between
observations to be quantified by a specific, user-chosen dis-
tance/dissimilarity measure.

This flexibility appears even more desirable when dealing
with the polymorphous objects commonly found in mod-
ern data sets, as described above. For instance, functional
objects are much richer than just infinite-dimensional vec-
tors, and they can be compared on many different grounds:
general appearance, short- or long-range variation, oscillat-
ing behaviour, etc.; which makes the choice of the ‘proximity
measure’ between two such objects a very crucial one (Fer-
raty and Vieu 2006, Chapter 3). On a more theoretical basis,
an appropriate choice of such ‘proximity measure’ some-
times allows one to get around issues caused by the ‘Curse
of Dimensionality’ (Geenens 2011a).

1 The Bray-Curtis ‘distance’ does not satisfy triangle inequality, hence
it is rather a semi-distance.

Quantifying (dis)similarities between non-numeric objects
is even more subject to discretionary choice. As an example,
for comparing pieces of texts, the literature in text mining,
linguistics and natural language processing proposed numerous
metrics such as the Levenshtein distance, the Hamming
distance, the Jaccard index or the Dice coefficient—each tar-
getting different dimensions of words, sentences or texts,
such as similarity in spelling or similarity in meaning (Wang
and Dong 2020). It is, therefore, paramount to have access
to statistical procedures which allow a free choice of metric,
and may be tailored to the kind of data at hand and to the
ultimate purpose of the analysis.

Indeed, our proposed ‘metric depth’ (μD), defined in
Sect. 2, enables such flexible analyses. Its main properties
are explored in Sect. 3 and an empirical version (computable
from a sample) is described in Sect. 4. Section 5 illustrates its
capabilities on several real data sets, including an application
in ‘text mining’ (Sect. 5.5). Section 6 concludes.

2 Statistical depth in metric spaces: definition

Assume that the random object of interest, say \( X \), lives in a
certain space \( M \) which can be equipped with a distance \( d \). To
avoid dispensable technical complications, it will be assumed
throughout that \( (M,d) \) is a complete and separable metric
space. Let \( \mathcal{A} \) be the \( \sigma \)-algebra on \( M \) generated by the open
d-metric balls and \( \mathcal{P} \) be the space of all probability measures
defined on the Borel sets of \( \mathcal{A} \).2 This makes \( (M,\mathcal{A},\mathcal{P}) \) a
proper probability space for any \( P \in \mathcal{P} \). In particular, it will
be assumed that the distribution of \( X \) belongs to \( \mathcal{P} \). Note that
the cartesian product space \( (M \times M, \mathcal{A} \times \mathcal{A}, \mathcal{P} \times \mathcal{P}) \)
is then also a valid probability space (Parthasarathy 1967, Theorem
I.1.10). We denote:

\[
\mathbb{P}(S(X_1, X_2)) = \iint_{M \times M} S(x_1, x_2) \, d(P \times P)(x_1, x_2)
\]

for any measurable statement \( S : M \times M \to [0,1] \)—the
statement returns the value 1 if it is true, and 0 otherwise. So,
\( \mathbb{P}(S(X_1, X_2)) \) returns the probability that \( S \) is true if \( X_1, X_2 \)
are two independent replications of \( X \), whose distribution is
\( P \).

Then we give the following definition:

**Definition 2.1** The ‘metric depth’ (‘\( \mu D \)) of the point \( X \) in
the metric space \( (M,d) \) with respect to the probability measure
\( P \in \mathcal{P} \) is defined as:

\[
\mu D(X, P) = \mathbb{P}(d(X_1, X_2) > \max(d(X_1, X), d(X_2, X)))
\]

2 In a separable metric space, the Borel \( \sigma \)-algebra is generated by the
open balls.
For each fixed $\chi \in \mathcal{M}$, the set $\{ (\chi_1, \chi_2) \in \mathcal{M} \times \mathcal{M} : d(\chi_1, \chi_2) > \max \{ d(\chi_1, \chi), d(\chi_2, \chi) \} \}$ belongs to the $\sigma$-algebra $\mathcal{A} \times \mathcal{A}$, with $\mathcal{A}$ defined above, making the probability statement $\mathbb{P}$ in (2.1) a well-defined one for any $P \in \mathcal{P}$.

The interpretation of (2.1) in terms of depth is clear: a point $\chi \in \mathcal{M}$ is deep with respect to the distribution $P$ if it is likely to find it ‘between’ two objects $\chi_1$ and $\chi_2$ in $\mathcal{M}$ randomly generated from $P$. ‘Between’ here means that the side joining $\chi_1$ and $\chi_2$ is the longest in a ‘triangle’ of $\mathcal{M}$ with vertices $\chi_1, \chi_2$ and $\chi$, or, in other words, that $\chi$ belongs to the intersection of the two open $d$-balls $B_d(d(\chi_1, \chi_2))$ and $B_d(d(\chi_2, d(\chi_1, \chi_2)))$, where $B_d(d(\chi_1, d(\chi_1, \chi_2)))$ is the ball with center $\chi_1$ and radius $d(\chi_1, \chi_2)$. In this sense, (2.1) is an extension of the vectorial ‘lens depth’ (Liu and Modares 2011), as they both coincide when $(\mathcal{M}, d) = (\mathbb{R}^p, d_2)$, where $d_2$ is the Euclidean distance over $\mathbb{R}^p$, $p \geq 1$. If we define

$$L_d(\chi_1, \chi_2) := B_d(\chi_1, d(\chi_1, \chi_2)) \cap B_d(\chi_2, d(\chi_1, \chi_2)).$$

the ‘lens’ defined by $\chi_1$ and $\chi_2$ in $(\mathcal{M}, d)$, then $\mu D(\chi, P) = \mathbb{P}(L_d(\chi_1, \chi_2) \ni \chi)$. This is the probability that a random set contains a certain element $\chi$, and interesting parallels can be drawn with the theory of random sets, in particular Choquet capacities and related ideas (Molchanov 2005, Chapter 1). Note that, independently of this work, Cholaquidis et al. (2022) recently explored the extension of the ‘lens depth’ to general metric spaces as well while Cholaquidis et al. (2021) studied the associated level sets. Their focus and the content of their papers are, however, much different to what is investigated here.

Finally, it is interesting to verify that, in the particular case where $\mathcal{M} = \mathbb{R}$ and $d$ is the usual distance $d(\chi, \xi) = |\chi - \xi|$, the metric depth $\mu D$ exactly coincides with the simplicial depth $D_S$ of Liu (1990) (defined using open simplices):

$$\mu D(\chi, P) = \mathbb{P}(|\chi_1 - \chi_2| > \max(|\chi_1 - \chi|, |\chi_2 - \chi|)) = \mathbb{P}(\chi \in \min(\chi_1, \chi_2), \max((\chi_1, \chi_2))) = D_S(\chi, P).$$

Note that the same occurs for the ‘lens depth’—in $\mathbb{R}$, where the notion of rank is unequivocally defined, it is expected that all reasonable depth measures are in agreement, indeed.

### 3 Main properties

The fact that the distance $d$ is left free really makes the metric depth $\mu D$ a very flexible tool, as any meaningful $d$ equipping $\mathcal{M}$ can be used in (2.1) without altering the theoretical properties which we explore below.

In addition, we note that no-where in the developments, it is used explicitly the fact that $d(\chi, \xi) = 0 \iff \chi = \xi$ for any two $\chi, \xi \in \mathcal{M}$ (identity of indiscernibles). A proximity measure which satisfies all the properties of a distance (non-negativity, symmetry and triangle inequality) but not ‘identity of indiscernibles’ is called a pseudo-distance. Hence, the metric depth (2.1) can be used in conjunction with a pseudo-distance, while keeping its essential features. We can, for instance, assess the proximity between two objects by comparing the coefficients of their leading terms when expanded in certain bases, such as a spline basis in the case of functional data when smoothing the original data is necessary (Ramsay and Silverman 2005, Chapter 3). Other examples are given in Sect. 5.

#### 3.1 Elasticity invariance

(P1) Let $\varphi : \mathcal{M} \rightarrow \mathcal{M}$ be an ‘elastic’ map in the sense that for any $\chi, \xi, \chi', \xi' \in \mathcal{M}$, $d(\chi, \xi) < d(\chi', \xi') \iff d(\varphi(\chi), \varphi(\xi)) < d(\varphi(\chi'), \varphi(\xi'))$. Then, $\mu D(\varphi(\chi), P) = \mu D(\chi, P)$, where $P$ is the push-forward distribution of the image through $\varphi$ of a random object of $\mathcal{M}$ having distribution $P$.

This follows from the fact that $d(\varphi(\chi_1), \varphi(\chi_2)) > \max \{ d(\varphi(\chi_1), \varphi(\chi)), d(\varphi(\chi_2), \varphi(\chi)) \} \iff d(\chi_1, \chi_2) > \max \{ d(\chi_1, \chi), d(\chi_2, \chi) \}$ for such a map $\varphi$. These obviously include any isometry, such as $d(\chi, \xi) = d(\varphi(\chi), \varphi(\xi))$, or other dilation-type transformations such that $d(\chi, \xi) = a_\varphi d(\varphi(\chi), \varphi(\xi))$, for some positive scalar constant $a_\varphi$, but not only. Clearly, (P1) establishes $\mu D$ as a purely topological concept. On another note, (P1) may be thought of as an extension of property P1 in Zuo and Serfling (2000, p. 463)—that a depth measure in $\mathbb{R}^d$ ‘should not depend on the underlying coordinate system or, in particular, on the scales of the underlying measurements’.

#### 3.2 Vanishing at infinity

Assume that $(\mathcal{M}, d)$ is an unbounded metric space, i.e., $\sup_{\chi, \xi \in \mathcal{M}} d(\chi, \xi) = \infty$. Then:

(P2) For any $P \in \mathcal{P}$ and $\chi \in \mathcal{M}$, $\lim_{R \to \infty} \sup_{\xi \in B_d(\chi, R)} \mu D(\xi, P) = 0$.

This follows from Proposition 1(a) in Cholaquidis et al. (2022). It is obviously the analogue to Zuo and Serfling (2000)’s P4: ‘The depth of a point $x$ should approach 0 as $|x|$ approaches infinity’.

Now suppose that, $\forall \chi \in \mathcal{M}$,

$$\mathbb{P}(d(\chi_1, \chi_2) = \max \{ d(\chi_1, \chi), d(\chi_2, \chi) \}) = 0. \quad (3.1)$$

This kind of continuity condition guarantees that, with probability 1, a given $\chi \in \mathcal{M}$ will not lie exactly on the boundary of a random lens such as (2.2)—in fact, for $\mathcal{M} = \mathbb{R}$, this
condition exactly amounts to the continuity of the distribution $P$. Then, we can prove the following properties $(P_3)$ and $(P_4)$.

### 3.3 Continuity in $\chi$

$(P_3)$ For any $P \in \mathcal{P}$ such that (3.1) holds, $\forall \chi \in \mathcal{M}$ and $\forall \epsilon > 0$, there exists $\delta > 0$ such that

$$
\sup_{\xi : d(\chi, \xi) < \delta} |\mu D(\xi, P) - \mu D(\chi, P)| < \epsilon.
$$

Indeed, for any $\chi \in \mathcal{M}$, take $\xi \in \mathcal{M}$ such that $d(\chi, \xi) < \delta$, for some $\delta > 0$. Then, by the triangle inequality, for any $X_1, X_2 \in \mathcal{M}$,

$$
\max \{d(X_1, \chi), d(X_2, \chi)\} - \delta < \max \{d(X_1, \xi), d(X_2, \xi)\} < \max \{d(X_1, \chi), d(X_2, \chi)\} + \delta.
$$

Hence, $\mu D(\xi, P) = \mathbb{P}(d(X_1, X_2) > \max \{d(X_1, \xi), d(X_2, \xi)\})$ is such that

$$
\mathbb{P}(d(X_1, X_2) > \max \{d(X_1, \chi), d(X_2, \chi)\}) + \delta
\leq \mu D(\xi, P) \leq \mathbb{P}(d(X_1, X_2) > \max \{d(X_1, \chi), d(X_2, \chi)\}) - \delta.
$$

Now, see that $\Psi(x) = \mathbb{P}(\max \{d(X_1, \chi), d(X_2, \chi)\} - d(X_1, X_2) \leq x)$ is a cumulative distribution function assumed to be continuous at $x = 0$ by (3.1). This means that, for any $\epsilon > 0$, we can find a $\delta > 0$ such that $|\Psi(\delta) - \Psi(0)| < \epsilon$. As $\Psi(0) = \mu D(\chi, P)$, the claim follows.

A similar result appears in Chalouquidis et al. (2022, Proposition 2 in Supplementary Material) under a condition slightly stronger than (3.1).

### 3.4 Continuity in $P$

$(P_4)$ For any $P \in \mathcal{P}$ such that (3.1) holds, $\forall \chi \in \mathcal{M}$ and $\forall \epsilon > 0$, there exists $\delta > 0$ such that $|\mu D(\chi, Q) - \mu D(\chi, P)| < \epsilon - P$-almost surely for all $Q \in \mathcal{P}$ with $d_P(P, Q) < \delta P$-almost surely, where $d_P$ metricises the topology of weak convergence on $\mathcal{P}$.

This follows directly from classical results on convergence of probability measures on separable metric spaces—e.g., Dudley (2002, Theorem 11.1.1)—as $\mu D(\chi, P)$ and $\mu D(\chi, Q)$ in (2.1) are simple probability statements on elements of $\mathcal{A} \times \mathcal{A}$. Note that (3.1) guarantees that the ‘lens’, (2.2) is a continuity set in the sense of Dudley (2002, section 11.1).

### 3.5 Further comments

Zuo and Serfling (2000) listed two more desirable properties for a depth measure on $\mathbb{R}^d$: ‘Maximality at centre’ and ‘Monotonicity relative to deepest point’ (their properties P2 and P3). Similar features are difficult to investigate here for $\mu D$ without giving a stronger structure to $(\mathcal{M}, d)$, such as some sort of convexity, or $d$ to satisfy a parallelogram inequality, for example. As illustration, Zuo and Serfling (2000)’s P2 ‘Maximality at centre’ requires the depth to be maximum at a uniquely defined ‘centre’ with respect to some notion of symmetry. Without assuming a stronger structure $(\mathcal{M}, d)$, even the very definition of symmetry in $\mathcal{M}$ is unclear. As our aim here is to stay as flexible as possible with the proposed metric depth, we do not investigate further in that direction. Those properties of $\mu D$ may (or may not) be established on specific applications when $\mathcal{M}$ and $d$ are precisely defined, though.

On a side note, even if Liu and Modarres (2011) supposedly showed (their Theorem 6) that their Euclidean ‘lens depth’ in $\mathbb{R}^d$—of which (2.1) can be thought of as an extension—satisfies ‘Maximality at centre’ for centrally symmetric distributions, their proof appears wrong as pointed out in Kleindessner and von Luxburg (2017). Yet, Kleindessner and von Luxburg (2017) conceded that they believe that the statement is true. In Appendix, we give three counter-examples, establishing that the statement is actually not true: Liu and Modarres (2011)’s ‘lens depth’ does not generally satisfy neither ‘Maximality at centre’ nor ‘Monotonicity relative to deepest point’ (Zuo and Serfling 2000’s P2 and P3) for centrally symmetric distributions on $\mathbb{R}^d$.

A last important point is the following. Suppose that the balls $B_d(\cdot, \cdot)$ are convex in $(\mathcal{M}, d)$. Then it can easily be checked that, for any non-degenerate distribution $P \in \mathcal{P}$ (i.e., not a unit point mass at some $\chi \in \mathcal{M}$), $\mu D$ cannot be degenerated in the sense that $\mu D(\chi, P) \equiv 0$ for all $\chi \in \mathcal{M}$. Indeed, by convexity, the intersection $\bigcap_{\chi_1, \chi_2 \in \mathcal{M}} B_d(\chi_1, d(\chi_1, \chi_2)) \cap B_d(\chi_2, d(\chi_1, \chi_2))$ is non-empty as soon as $\chi_1 \neq \chi_2$, so, there always exists some $\chi \in \mathcal{M}$ which gets a positive depth by (2.1). It is known that some instances of statistical depth admit such a degenerate behaviour. For instance, that is the case of López-Pintado and Romo (2009, 2011)’s band and half-region depths for a wide class of distributions on common functional spaces (Chakraborty and Chaudhuri 2014, Theorems 3 and 4).

---

3 Kleindessner and von Luxburg (2017) noticed that analogous proofs for the spherical depth (Elmore et al. 2006) and the $\beta$-skeleton depth (Yang and Modarres 2017) are mistaken as well. Furthermore, we have found that the proof of a similar property for the band depth given in López-Pintado and Romo (2009, Theorem 1(2)) is likewise erroneous.
4 Empirical metric depth

4.1 Definition and main statistical properties

Assume now that we have a random sample \{\chi_i; i = 1, \ldots, n\} drawn from the unknown distribution \(P\) on \(M\). Then the depth of some point \(\chi \in M\) with respect to \(P\) must actually be estimated. The empirical analogue of (2.1) is naturally \(\mu D(\chi, \hat{P}_n)\), where \(\hat{P}_n\) is the empirical measure associated to \(\{\chi_i; i = 1, \ldots, n\}\), i.e., the collection of 1/n-weighted point masses at \(\chi_1, \ldots, \chi_n\). This yields

\[
\mu D(\chi, \hat{P}_n) = \frac{1}{\binom{n}{2}} \sum_{i < j} \mathbb{I}_{d(\chi_i, \chi_j) > \max\{d(\chi_i, \chi), d(\chi_j, \chi)\}}. \tag{4.1}
\]

Obviously, \(\hat{P}_n \xrightarrow{P\text{-a.s.}} P\), which guarantees under (3.1) the strong pointwise consistency of the estimator \(\mu D(\chi, \hat{P}_n)\), that is

\[
\mu D(\chi, \hat{P}_n) \xrightarrow{P\text{-a.s.}} \mu D(\chi, P), \tag{4.2}
\]

for all \(\chi \in M\). This easily follows from Property (P4). Under stronger assumptions, Cholaquidis et al. (2021, Theorem 4) established the uniform strong consistency of \(\mu D\) on compact subsets of \(M\); viz.

\[
\sup_{\chi \in \Phi} |\mu D(\chi, \hat{P}_n) - \mu D(\chi, P)| \xrightarrow{\text{a.s.}} 0, \quad n \to \infty, \tag{4.3}
\]

for any compact set \(\Phi \subset M\).

Finally, the obvious \(U\)-statistics structure of (4.1) allows us to easily deduce, through an appropriate Central Limit Theorem, the asymptotic normality of \(\mu D(\chi, \hat{P}_n)\), as a straightforward consequence of (Arcones and Giné 1993, Theorem 4.10) and (Giné 1996, Proposition 10); see also (Cholaquidis et al. 2021,Theorem 6). This result can be used for inference, for instance to build a confidence region for the ‘true’ median element, i.e. the deepest element with respect to the population distribution \(P\) (Serfling and Wijesuriya 2017). Note that this median need not be unique, and careful treatment is necessary in that case. Indeed, as discussed above, the depth is not guaranteed to be monotonically decreasing when moving away from the deepest point, making sets of deepest points not necessarily convex. In such cases, we may make use of the level sets of maximum depths, following results of Cholaquidis et al. (2021).

4.2 Computational aspects

The practical computation of the empirical metric depth (4.1) involves obtaining all pairwise distances between the elements of the set \(\{\chi\} \cup \{\chi_i; i = 1, \ldots, n\}\); and checking \(d(\chi_i, \chi_j) > \max\{d(\chi_i, \chi), d(\chi_j, \chi)\}\) for \(1 \leq i < j \leq n\).

A simple (if not naive) implementation of this would be carried out in maximum \(O(n^2)\) flops. Thus, the computational cost for computing the metric depth is in general lower than the simplicial depth, which requires \(O(n^{p+1})\) operations in \(\mathbb{R}^p\), and Tukey depth, which involves infinite unidimensional projections. The above observation assumes that the dimension of the objects of interest is fixed; e.g., functional data are represented by vector of a certain length which does not vary with \(n\). Now, even in such cases, the objects of interest may be complex and have dimension close to or larger than the sample size \(n\)—this may have to be taken into account when computing the distance \(d\). For example, an \(L_r\)-distance, \(1 \leq r \leq \infty\), between functional data represented by a vector of size \(N\), is computed in \(O(N)\) operations. If \(N\) is comparable to \(n\), the total computational cost would effectively be \(O(Nn^2)\).

5 Data examples

In this section, we illustrate the usefulness of the proposed metric depth \(\mu D\) on 5 real data sets: two one-dimensional functional datasets (Sects. 5.1 and 5.2), a bidimensional functional dataset (Sect. 5.3), a symbolic data set (Sect. 5.4) and a non-numeric data set (text) (Sect. 5.5).

5.1 Canadian weather data

The Canadian temperature data set is a classical functional data set available from the R package \texttt{fda}. The data give the daily temperature records of 35 Canadian weather stations over a year (365 days, day 1 is 1st of January) averaged over 1960 to 1994, see Fig. 1. First, the depth of the 35 curves with respect to the sample has been computed from the empirical functional metric depth (4.1) with \(d\) being the usual \(L_2\) distance between two square-integrable functions, i.e. \(d^2(\chi, \xi) = \int (\chi(t) - \xi(t))^2 \, dt\). The 5 deepest and
least deep curves are shown in Fig. 2. The suggested depth measure identifies the Sherbrooke (deepest curve), Thunder Bay, Fredericton, Quebec and Calgary stations as the most representative of a median Canadian weather, in terms of temperature. On the other hand, the most outlying curves are seen to be Resolute (least deep curve), Victoria, Vancouver, Inuvik and Iqaluit. It is visually obvious that those curves are much different to the others: Resolute, Inuvik and Iqaluit are Arctic stations, with much colder temperatures across the year than the other stations, while Vancouver and Victoria lie on the south Pacific coast of Canada and enjoy much milder winters. We can appreciate that Vancouver and Victoria are ‘shape outliers’, whereas the Arctic stations are ‘location outliers’. These are equally easily flagged by the metric depth $\mu D$—this has to be stressed, as some functional depths have been shown to be able to identify one type of outlier but not the other, or vice-versa (Serfling and Wijesuriya 2017).

Of course, the daily average temperature curves are particularly noisy, which could heavily affect the $L_2$-distances computed between pairs of curves, hence the whole calculation of the depths. One can deal with the roughness of those curves in different manners: first, one could use smoothed versions of the initial curves, for instance the monthly average temperatures as in Serfling and Wijesuriya (2017); second, one could use for $d$ a distance less affected by such noise than the $L_2$ one, for instance the supremum ($L_\infty$) distance; finally one can expand the different curves in a certain basis and focus only on the first terms when assessing the proximity between them. We achieved that by expanding each curve in the empirical Principal Components basis (Hall 2011) and keeping only the first two principal scores: the curves re-constructed from those two components only are indeed smooth approximations to the initial, rough curves. So, each curve is now represented by a point in the 2-dimensional space of the first two Principal Components, and the proximity between two curves quantified by the $L_2$-distance between the corresponding two points. In effect, this defines a pseudo-distance between the initial curves (see Ferraty and Vieu 2006, section 3.4.1). The depths assigned to each station according to these 4 methods are shown in Table 1. The four depth measures are in very good agreement, essentially identifying the same central and outlying curves. This shows that the depth measure $\mu D$ (2.1) and its empirical version (4.1) are quite robust to any reasonable choice of $d$.

5.2 Lip movement data

Malfait and Ramsay (2003) studied the relationship between lip movement and time of activation of different face muscles, see also (Ramsay and Silverman 2002, Chapter 10) and Gervini (2008). The study involved a subject saying the word ‘bob’ 32 times and the movement of their lower lip was recorded each time. Those trajectories are shown in Fig. 3, and all share the same pattern: a first peak corresponding to the first /b/, then a plateau corresponding to the /o/ and finally a second peak for the second /b/. These functions being very smooth (actually, they are smoothed versions of raw data not publicly available), it seems natural to use again the classical $L_2$ distance for assessing their relative proximity. Hence, the respective depth of each curve with respect to the sample was obtained by (4.1) with $d^2(\chi, \xi) = \int (\chi(t) - \xi(t))^2 \, dt$. The 5 deepest and 5 least deep curves are shown in the top row of Fig. 4. In particular, this depth identifies as outliers the three curves showing a second peak at a much later time than for the rest of the curves, which were already hived off by Gervini (2008). The remaining two outlying curves show two peaks of lower amplitude than the others, with a second peak occurring earlier than the bunch.

Now, Malfait and Ramsay (2003), in their original study, were more interested in the acceleration of the lip during
Table 1: Canadian weather data—metric depth measures for 4 different (pseudo-)distances $d$:

- $\mu D_2$: $L_2$ distance;
- $\mu D_\infty$: Supremum ($L_\infty$) distance;
- $\mu D_{2m}^2$: $L_2$ distance on the average monthly temperature curves;
- $\mu D_{PCA}$: $L_2$ distance in the plane of the first two principal components.

| Station     | $\mu D_2$ | $\mu D_\infty$ | $\mu D_{2m}^2$ | $\mu D_{PCA}$ | Station     | $\mu D_2$ | $\mu D_\infty$ | $\mu D_{2m}^2$ | $\mu D_{PCA}$ |
|-------------|-----------|-----------------|-----------------|----------------|-------------|-----------|-----------------|-----------------|----------------|
| Sherbrooke  | 0.514     | 0.474           | 0.516           | 0.506          | Thunder B   | 0.511     | 0.513           | 0.504           | 0.496          |
| Fredericton | 0.504     | 0.479           | 0.504           | 0.491          | Quebec      | 0.504     | 0.513           | 0.506           | 0.497          |
| Calgary     | 0.492     | 0.375           | 0.492           | 0.503          | Bagottville | 0.482     | 0.457           | 0.484           | 0.476          |
| Edmonton    | 0.474     | 0.459           | 0.472           | 0.494          | Arvida      | 0.469     | 0.476           | 0.472           | 0.476          |
| Regina      | 0.429     | 0.435           | 0.427           | 0.420          | Charlottvl  | 0.425     | 0.435           | 0.424           | 0.435          |
| Pr.George   | 0.422     | 0.380           | 0.425           | 0.469          | Ottawa      | 0.410     | 0.479           | 0.408           | 0.403          |
| Winnipeg    | 0.403     | 0.418           | 0.405           | 0.398          | Pr.Albert   | 0.400     | 0.383           | 0.400           | 0.390          |
| Montreal    | 0.383     | 0.457           | 0.380           | 0.375          | Halifax     | 0.380     | 0.348           | 0.378           | 0.368          |
| Whitehorse  | 0.371     | 0.378           | 0.373           | 0.395          | The Pas     | 0.360     | 0.358           | 0.358           | 0.348          |
| Sydney      | 0.324     | 0.284           | 0.333           | 0.348          | Uranium C.  | 0.316     | 0.257           | 0.319           | 0.311          |
| Toronto     | 0.313     | 0.363           | 0.309           | 0.309          | Schefferville| 0.294     | 0.318           | 0.292           | 0.291          |
| St.Johns    | 0.271     | 0.237           | 0.272           | 0.257          | London      | 0.267     | 0.324           | 0.267           | 0.261          |
| Yellowknife | 0.227     | 0.207           | 0.224           | 0.200          | Yarmouth    | 0.208     | 0.195           | 0.208           | 0.205          |
| Dawson      | 0.207     | 0.138           | 0.208           | 0.242          | Churchill   | 0.160     | 0.242           | 0.160           | 0.160          |
| Kamloops    | 0.150     | 0.202           | 0.148           | 0.148          | Pr.Rupert   | 0.106     | 0.108           | 0.106           | 0.109          |
| Iqaluit     | 0.096     | 0.134           | 0.096           | 0.096          | Inuvik      | 0.066     | 0.064           | 0.066           | 0.066          |
| Vancouver   | 0.045     | 0.050           | 0.045           | 0.040          | Victoria    | 0.017     | 0.002           | 0.017           | 0.018          |
| Resolute    | 0         | 0               | 0               | 0              |             | 0         | 0               | 0               | 0              |

Fig. 3: Lip movement data

The process rather than on the lip motion itself. The study aimed at explaining time of activation of face muscles, and the acceleration reflects the force applied to tissue by muscle contraction. Hence, in this application, it may be worth contrasting the lip trajectories in terms of their corresponding accelerations, that is, comparing the second derivatives of the position curves. The $L_2$ distance between the second derivatives of the curves is naturally a pseudo-distance between the initial curves (Ferraty and Vieu 2006, Section 3.4.3), which can be used in (4.1). The 5 deepest curves, according to (4.1) based on the ‘acceleration’ pseudo-distance, are shown in the middle row of Fig. 4 and differ from those in the first row of Fig. 4. Naturally, the focus here is no more on the exact position of the curves, but rather on the more fundamental underlying dynamics. For instance, the 5 deepest curves show a first peak of distinctly different heights, but in terms of their second derivatives, they are in fact quite similar and representative of the sample (bottom row of Fig. 4), and that is what matters in Malfait and Ramsay (2003)’s study. As argued in Sect. 1, the flexibility of $\mu D$ (2.1) in terms of the choice of $d$ allows the analyst to tailor the depth measure to the given factors and the goal of the analysis.

5.3 Handwriting data

The ‘handwriting’ data set consists of twenty replications of the printing of the three letters ‘fda’ by a single individual. The position of the tip of the pen has been sampled 200 times per second. The data, available in the R package fda, have already been pre-processed so that the printed characters are scaled and oriented appropriately, see Fig. 5.

These data are essentially bivariate functional data. Indeed, each instance $\chi$ of the word ‘fda’ arises through the simultaneous realisation of two components $(\chi_X(t), \chi_Y(t))$, where $\chi_X(t)$ and $\chi_Y(t)$ give the position along the horizontal axis and the vertical axis, respectively, of the pen at time $t$. This is illustrated for one instance of ‘fda’ in Fig. 6. Hence, an appropriate functional metric space here could be $(M, d)$ with $M = L_2(T) \times L_2(T), T = [0, 2.3]$ (the time interval on which the position of the pen was recorded) and $d$ being the Euclidean distance on $L_2(T) \times L_2(T)$ whose square is defined by
Fig. 4   Top and middle row: Five deepest curves (left; the darkest curves are the deepest) and five least deep curves (right; the lightest curves are the least deep) according to (4.1) with (i) $d(\chi, \xi) = \|\chi - \xi\|_2$, the $L_2$ distance between the curves (top row) and (ii) $d(\chi, \xi) = \|\chi'' - \xi''\|_2$, the $L_2$ distance between the second derivatives of the curves (middle row). Bottom row: Five deepest acceleration curves (left; the darkest curves are the deepest) and five least deep acceleration curves (right; the lightest curves are the least deep)

\[
d^2(\chi, \xi) = \|\chi_X - \xi_X\|_2^2 + \|\chi_Y - \xi_Y\|_2^2 = \int_T (\chi_X(t) - \xi_X(t))^2 \, dt + \int_T (\chi_Y(t) - \xi_Y(t))^2 \, dt. \tag{5.1}
\]

This distance can be used directly in (4.1) to identify the 5 deepest and 5 least deep instances of ‘fda’, see Fig. 7. The bivariate nature of the data at hand does not cause any particular complication and the definition (2.1) need not be re-adapted to this case. Again, the so-defined depth only focuses on the ‘drawings’ fda themselves, and identifies the deepest instances. However, it was argued in the related literature that the tangential acceleration of the pen during the process was also a key element to analyse for understanding the writing dynamics, for instance for discriminating between genuine handwritings and forgeries (Geenens 2011a, b). As in Subsection 5.2, one could therefore use (4.1) with $d$ a pseudo-distance assessing the proximity between two instances of fda through their tangential acceleration curves only, if that was to be the focus of the analysis.
5.4 Age distribution in European countries

Symbolic Data Analysis (SDA) has recently grown as a popular research field in statistics (Billard and Diday 2003, 2007). Indeed, the intrinsically large ‘Big Data’ sets often need to be summarised so that the resulting summary datasets are of a manageable size, and so-called ‘symbolic data’ typically arise from such a process. No longer formatted as single values like classical data, they are meant to be ‘aggregated’ variable typically represented by lists, intervals, histograms, distributions and the like. In this section we give a closer look at a ‘distribution-valued’ symbolic data set. Specifically, we analyse the distribution of the age of the population of the 44 European countries (see Table 2).

The 2017 data were obtained from the US Census bureau (www.census.gov/population/international/data/). Typically, the population distribution for a given country is presented under the form of a population pyramid (that is, a histogram),
from which a proper distribution function for population age can easily be extracted (Kosmelj and Billard 2011). Hence, each country (here: ‘individual’, also called ‘concept’ in the SDA literature) is characterised by a distribution. Figure 8 displays the sample of age distributions. Here we will use the suggested metric depth \( \mu D \) to analyse which countries are most representative of the ‘European’ age distribution, and which countries can be regarded as ‘outliers’ in that respect.

The data being here distribution functions of nonnegative variables, \( M \) can be identified with a space of distribution functions supported on \( \mathbb{R}^+ \), i.e. a space of nondecreasing càdlàg functions \( F \) with \( F(0) = 0 \) and \( \lim_{t \to \infty} F(t) = 1 \), equipped with an appropriate distance. The Wasserstein distance has proved useful for a wide range of problems explicitly involving distribution functions (Rachev1984; Panaretos and Zemel 2020), hence seems a natural choice in this setting as well. For some \( r \geq 1 \), the Wasserstein distance between two distributions \( F \) and \( G \) whose \( r \)th moments exist, is defined as

\[
d_r(F, G) = \inf_{(X,Y) \sim (F, G)} \left\{ \mathbb{E} \left( |X - Y|^r \right) \right\}^{1/r},
\]

where the infimum is taken over the set of all joint bivariate distributions whose marginal distributions are \( F \) and \( G \) respectively. Properties of this distance are described in Major (1978) and Bickel and Freedman (1981). In particular, it is known that \( d_r(F, G) \) is essentially the usual \( L_r \) distance between the quantile functions \( F^{-1} \) and \( G^{-1} \) over \([0,1] \). Also, it is known that convergence in the Wasserstein distance is equivalent to convergence in distribution together with convergence of the first \( r \) moments. Hence, the distance \( d_r \) quantifies the proximity between two distributions through both their general appearance and the values of their moments. In what follows, we take \( r = 2 \), hence we consider functional data in \((M_2, d_2)\), \( M_2 \) being the space of all probability distribution functions with finite second moment.

The flexibility of (2.1) allows us to base \( \mu D \) on the Wasserstein distance so as to define a depth measure specific to distribution functions without any difficulty. The ‘Wasserstein-depths’ of the 44 countries are given in Table 2. The 5 deepest and least deep age distributions are shown in Fig. 9. The deepest distribution, hence the most representative of the age distributions in Europe, appears to be that of Switzerland, a country located at the very heart of Europe, in-between the Western and Eastern countries, and in-between the Northern countries and the Southern countries, at the meeting point between the ‘Germanic’ world (Germany, Austria) and the ‘Latin’ world (France, Italy). From that perspective, Switzerland can be regarded as really representative of a ‘median’ European country on many aspects. On the other hand, the Wasserstein-metric depth is null for Kosovo and Monaco, and indeed, the distributions for those two countries clearly lie outside the bunch of the other distributions. Monaco is a micro, mild-climate (and incidentally, tax haven) state which attracts a large amount of rich retirees from all over the continent (if not the world), hence its population is globally much older than for other countries and its age distribution is below the others. Monaco set aside, Germany and Italy show globally the oldest population of Europe. Kosovo was still recently at the heart of an armed conflict in the Balkans, which explains the low proportion of older people in that country and the position of its age distribution above all the others. To some extent, this also explains the outlying-ness of Albania’s curve. In any case, this example illustrates that one can readily define a depth measure tailored for distribution curves, which paves the way for developing rank-like procedures in Symbolic Data Analysis as well.

### 5.5 Authorship attribution by intertextual distance

Author identification on an unknown or doubtful text is one of the oldest statistical problems applied to literature. Here the capability of the proposed metric depth is illustrated within that framework. William Shakespeare and Thomas Middleton were contemporaries (late 16th-early 17th centuries), and their oeuvre are often compared. In that aim, Merriam (2003) examined 9 Middleton plays and 37 Shakespeare texts, and computed between each pair of them the so-called ‘inter-textual distance’ proposed by Labbé and Labbé (2001). Although the entities of interest are here purely non-numerical (famous literary pieces), the obtained matrix of distances allows us to outline the relative position of each text—and this is essentially all what is needed for \( \mu D \) to come into play.

As an example, Table 3 (recovered from Appendix 2 in Merriam (2003)) reports the ‘inter-textual’ distances between

---

\(^4\) It is not the purpose of this paper to describe how this index is computed or what it represents; neither do we imply that it is the panacea for the considered problem—for that matter, it has been criticised (Viprey and Ledoux 2006). Here we use it in an illustrative purpose only.
Table 2  Age distribution in European countries—metric depth for the age distributions of the 44 European countries, based on the Wasserstein distance

| Country       | Depth |
|---------------|-------|
| Switzerland   | 0.59  |
| Liechtenstein | 0.52  |
| Hungary       | 0.519 |
| Malta         | 0.508 |
| Czech Republic| 0.503 |
| Ukraine       | 0.498 |
| Netherlands   | 0.488 |
| Croatia       | 0.478 |
| Portugal      | 0.477 |
| Poland        | 0.474 |
| Belgium       | 0.451 |
| Serbia        | 0.449 |
| Denmark       | 0.434 |
| Romania       | 0.422 |
| UK            | 0.421 |
| Belarus       | 0.406 |
| Spain         | 0.406 |
| Estonia       | 0.401 |
| Bulgaria      | 0.36  |
| Montenegro    | 0.36  |
| Slovakia      | 0.349 |
| Latvia        | 0.348 |
| Sweden        | 0.322 |
| Lithuania     | 0.319 |
| Luxembourg    | 0.318 |
| Russia        | 0.292 |
| Norway        | 0.28  |
| Austria       | 0.277 |
| Bosnia-Herz   | 0.276 |
| France        | 0.27  |
| Finland       | 0.261 |
| San Marino    | 0.228 |
| Andorra       | 0.221 |
| Macedonia     | 0.201 |
| Slovenia      | 0.178 |
| Moldova       | 0.151 |
| Greece        | 0.141 |
| Iceland       | 0.14  |
| Ireland       | 0.088 |
| Italy         | 0.087 |
| Belgium       | 0.044 |
| Germany       | 0.044 |
| Kosovo        | 0     |
| Monaco        | 0     |

Fig. 9  Five deepest age distributions (left; the darkest curves are the deepest) and five least deep age distributions (right; the lightest curves are the least deep) according to (4.1) with $d$ being the Wasserstein distance $d_2$ between distributions.

Table 3  Matrix of ‘inter-textual’ distances between 9 essential plays by Thomas Middleton: Phn: ‘The Phoenix’; Mad: ‘A Mad World, My Masters’; Trk: ‘A Trick to Catch the Old One’; Pur: ‘The Puritan’; Alm: ‘The Almanac’; CMC: ‘A Chaste Maid in Cheapside’; Dis: ‘More Dissemblers Besides Women’; Val: ‘The Nice Valour’; WBW: ‘Women Beware Women’

|       | Phn   | Mad   | Trk   | Pur   | Alm   | CMC   | Dis   | Val   | WBW   |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Phn   | 0.000 | 0.312 | 0.301 | 0.315 | 0.335 | 0.343 | 0.319 | 0.330 | 0.322 |
| Mad   | 0.312 | 0.000 | 0.314 | 0.344 | 0.319 | 0.328 | 0.326 | 0.339 | 0.325 |
| Trk   | 0.301 | 0.314 | 0.000 | 0.310 | 0.314 | 0.326 | 0.335 | 0.338 | 0.330 |
| Pur   | 0.315 | 0.344 | 0.310 | 0.000 | 0.338 | 0.340 | 0.337 | 0.346 | 0.338 |
| Alm   | 0.335 | 0.319 | 0.314 | 0.338 | 0.000 | 0.314 | 0.313 | 0.333 | 0.344 |
| CMC   | 0.343 | 0.328 | 0.326 | 0.340 | 0.314 | 0.000 | 0.339 | 0.349 | 0.311 |
| Dis   | 0.319 | 0.326 | 0.335 | 0.337 | 0.313 | 0.339 | 0.000 | 0.318 | 0.284 |
| Val   | 0.330 | 0.339 | 0.338 | 0.346 | 0.333 | 0.349 | 0.318 | 0.000 | 0.320 |
| WBW   | 0.322 | 0.325 | 0.330 | 0.338 | 0.344 | 0.311 | 0.284 | 0.320 | 0.000 |
Table 4  37 Shakespeare’s plays (shown in chronological order)—empirical $\mu_D$ in the sample of Shakespeare’s plays only (left column) and in the sample of combined Shakespeare’s and Middleton’s works (right column)

| Play                        | $\mu_D$ - Shakespeare | $\mu_D$ - combined |
|-----------------------------|------------------------|---------------------|
| The Two Gentlemen of Verona | 0.3784                 | 0.5072              |
| The Taming of the Shrew    | 0.2192                 | 0.4396              |
| Henry VI - Part II         | 0.2508                 | 0.2097              |
| Henry VI - Part III        | 0.0526                 | 0.0425              |
| Henry VI - Part I          | 0.1532                 | 0.1227              |
| Titus Andronicus           | 0.1937                 | 0.1594              |
| Richard III                | 0.4760                 | 0.4570              |
| The Comedy of Errors       | 0.4595                 | 0.5353              |
| Love’s Labour’s Lost       | 0.4910                 | 0.4889              |
| A Midsummer Night’s Dream  | 0.3859                 | 0.3691              |
| Romeo and Juliet           | 0.5015                 | 0.5227              |
| Richard II                 | 0.1006                 | 0.0841              |
| King John                  | 0.2267                 | 0.2048              |
| The Merchant of Venice     | 0.4189                 | 0.5179              |
| Henry IV, Part I           | 0.4129                 | 0.4058              |
| The Merry Wives of Windsor | 0.3829                 | 0.4483              |
| Henry IV, Part II          | 0.3619                 | 0.4744              |
| Much Ado about Nothing     | 0.2252                 | 0.4444              |
| Henry V (prose part)       | 0.3078                 | 0.2870              |
| Henry V (verset part)      | 0.0000                 | 0.0000              |
| Julius Caesar              | 0.4294                 | 0.3990              |
| As You Like It             | 0.2763                 | 0.4792              |
| Hamlet                     | 0.4715                 | 0.4473              |
| Twelfth Night              | 0.0075                 | 0.3488              |
| Troilus and Cressida       | 0.3333                 | 0.2850              |
| Measure for Measure        | 0.3378                 | 0.5063              |
| Othello                    | 0.5030                 | 0.5585              |
| All’s Well that Ends Well  | 0.0450                 | 0.3710              |
| Timon of Athens            | 0.1141                 | 0.3971              |
| King Lear                  | 0.4640                 | 0.5478              |
| Macbeth                    | 0.3649                 | 0.3121              |
| Antony and Cleopatra       | 0.5255                 | 0.5401              |
| Coriolanus                 | 0.4655                 | 0.4309              |
| The Winter’s Tale          | 0.2462                 | 0.4261              |
| Cymbeline                   | 0.1607                 | 0.4184              |
| The Tempest                | 0.5135                 | 0.5246              |
| Henry VIII                 | 0.4099                 | 0.5024              |

the 9 essential plays of Middleton. Computing the empirical metric-depth (4.1) on each of this entry in the ‘Middleton sample’ reveals that the two deepest observations are ‘More Dissemblers Besides Women’ and ‘A Trick to Catch the Old One’ (both get a depth of 0.4167). They may, therefore, be considered as the most typical Middleton plays (as long as the ‘inter-textual’ distance is the relevant metric).

This time focusing on the 37 Shakespeare texts only, ‘Antony and Cleopatra’ is identified as Shakespeare’s most typical text; i.e., the deepest among the considered sample (depth: 0.5255)—see Table 4 (left column). The following most representative of Shakespeare plays are ‘The Tempest’ (0.5135), ‘Othello’ (0.5030) and ‘Romeo and Juliet’ (0.5015). The most outlying piece of work is the verset part of ‘Henry V’ (depth: 0), which tends to confirm a common conjecture hold by many experts on Shakespeare’s oeuvre: the verset part of ‘Henry V’ was not written by Shakespeare himself, but by Christopher Marlowe (Merriam 2002).

Now, if we computed the metric depth of the 9 Middleton’s plays in Shakespeare’s sample, all would receive depth
Fig. 10 Top row: Example A.1; Central row: Example A.2; Bottom row: Example A.3. From left to right, density function (first column), sample lens depth constructed with $n = 5,000$ sample draws from $P$ (second column), corresponding heat-map (third column) and its section along the line $x_2 = 0$ (top-right panel), $x_1 = x_2$ (central-right panel) and $x_1 = 0$ (bottom-right panel).

0—all are ‘outlying’ in Shakespeare’s oeuvre. This clearly indicates that Middleton’s work cannot be confused with Shakespeare’s, and it should be easy to assign a new piece of text to one or the other based on $\mu D$. Further, it is interesting to analyse the depth of each text in a combined sample made up both the works of Middleton and Shakespeare. In particular, some of Shakespeare’s texts which have a low depth in the ‘Shakespeare’s only’ sample, see their depth increase by large in the combined sample. This indicates that these pieces may have a strong Middleton flavour, to some extent. This hypothesis is confirmed for at least one of those plays: ‘Timon of Athens’ sees its depth increase from 0.1141 to 0.3971 if one includes Middleton’s works in the reference sample; and indeed, extensive research on the topic has provided ample evidence that Middleton wrote approximately one third of that play (Taylor 1987).

Note that computing and comparing the depth of certain observations in two different samples is the spirit of the $DD$-plot and the $DD$-classifier proposed by Li et al. (2012). These procedures can naturally be used in conjunction with the metric depth $\mu D$, enabling similar powerful depth-based analyses in abstract metric spaces.

6 Conclusion

In this paper, we have proposed a new statistical depth function, called ‘metric depth’ or just $\mu D$, defined in an abstract metric space. It is explicitly constructed on a certain distance $\delta$ that must be chosen by the analyst, which allows them to tailor the depth to the data at hand and to the ultimate goal of the analysis. This offers an unmatched flexibility about the range of problems and applications that can be addressed using the said depth measure. The usefulness of $\mu D$ has been illustrated on several real data sets, including one in the emergent field of Symbolic Data Analysis and an application in text.
mining (authorship attribution). Rejuvenating an old idea of Bartoszynski et al. (1997), its definition is very intuitive: the depth of a functional point \( x \) with respect to a distribution \( P \) is the probability to find it ‘between’ two functional objects \( A_1 \) and \( A_2 \) randomly generated from \( P \), ‘between’ meaning here that \( x \) belongs to the intersection of the two open \( d \)-balls \( B_d(A_1, d(A_1, A_2)) \) and \( B_d(A_2, d(A_1, A_2)) \). This definition is natural and enjoys many pleasant properties.

Acknowledgements Gery Geenens’ research was supported by a Faculty Research Grant from the Faculty of Science, UNSW Sydney, Australia. Alicia Nieto-Reyes’ research was funded by the Spanish Ministerio de Ciencia, Innovación y Universidades Grant Number MTM2017-86061-C2-2-P.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

A Appendix

Here we give three counter-examples for illustrating that Liu and Modarres (2011)’s Euclidean ‘lens depth’ does not satisfy Zuo and Serfling (2000)’s properties P2 ‘Maximality at centre’ and P3 ‘Monotonicity relative to deepest point’ for centrally symmetric distributions—indeed, these two properties are related. For simplicity, we work in \( \mathbb{R}^2 \).

Example A.1 (Mixture of two normal distributions)
Let \( P \) be a mixture of two bivariate normal distributions with respective means \((-3, 0), (3, 0)\), identity covariance matrices and equal weights—viz., the density of \( P \) is \( f(x_1, x_2) = \frac{1}{2} \phi(x_1 - 3, x_2) + \frac{1}{2} \phi(x_1 + 3, x_2) \), for \( \phi \) the standard bivariate normal density.

Example A.2 (Bivariate normal distribution truncated to 4 squares) Let \( P \) be the distribution whose density function is
\[
f(x_1, x_2) = \phi(x_1, x_2)/\left(\int_{A} \phi(x_1, x_2) \, dx_1 \, dx_2\right) \mathbb{I}_{(x_1, x_2) \in A},
\]
where \( \phi \) is the standard bivariate normal density, \( A = A_1 \cup A_2 \cup A_3 \cup A_4 \) with \( A_1 = [-2, -1] \times [-1, 1], A_2 = [-3, -2] \times [2, 3], A_3 = [1, 2] \times [1, 2] \) and \( A_4 = [2, 3] \times [-3, -2] \).

Example A.3 (Bivariate normal distribution truncated to a frame) Let \( P \) be as in Example A.2 but with \( A_1 = [-4, -3] \times [-1, 1], A_2 = [-4, 4] \times [1, 2], A_3 = [3, 4] \times [-1, 1] \) and \( A_4 = [-4, 4] \times [-2, -1] \).

The distribution \( P \) is clearly centrally symmetric about \((0, 0)\) in each case: in Example A.1 as \( f(x_1, -x_2) = f(x_1, x_2) \), and in Examples A.2–A.3 because the standard bivariate normal distribution is centrally symmetric about \((0, 0)\) and the region \( A \) is symmetric with respect to the origin. However, Fig. 10 reveals that Liu and Modarres (2011)’s Euclidean lens depth function with respect to any of these three distributions is not maximum at the centre \((0, 0)\), nor is monotonic away from the deepest point(s)—see that the depth function admits a local maximum at \((0, 0)\) for Example A.2.

References
Arcones, M.A., Giné, E.: Limit theorems for U-processes. Ann. Probab. 21(3), 1494–1542 (1993)
Bartoszynski, R., Pearl, D.K., Lawrence, J.: A multidimensional goodness-of-fit test based on interpoint distances. J. Am. Stat. Assoc. 92, 577–586 (1997)
Bickel, P.J., Freedman, D.A.: Some asymptotic theory for the bootstrap. Ann. Stat. 9, 1196–1217 (1981)
Billard, L., Diday, E.: From the statistics of data to the statistics of knowledge: Symbolic Data Analysis. J. Am. Stat. Assoc. 98, 470–487 (2003)
Billard, L., Diday, E.: Symbolic Data Analysis: Conceptual Statistics and Data Mining, Wiley Series in Computational Statistics. Wiley (2007)
Bray, J.R., Curtis, J.T.: An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27, 325–349 (1957)
Carrizosa, E.: A characterization of half-space depth. J. Multiv. Anal. 58, 21–26 (1996)
Chakraborty, A., Chaudhuri, P.: On data depth in infinite dimensional spaces. Ann. Inst. Stat. Math. 66, 303–324 (2014)
Cholaquidis, A., Fraiman, R., Gamboa, F., Moreno, L.: Weighted lens depth: some applications to supervised classification. J. Stat. Canad (2022, to appear)
Cholaquidis, A., Fraiman, R., Moreno, L.: Level sets of depth measures in abstract spaces. Manuscript, arXiv:2011.11416 (2021)
Cuevas, A., Febrero, M., Fraiman, R.: Robust estimation and classification for functional data via projection-based depth notions. Comput. Stat. 22, 481–496 (2007)
Dai, X., Lopez-Pintado, S.: Tukey’s depth for object data. J. Am. Stat. Assoc. 1–13 (2022)
Dudley, R.M.: Real Analysis and Probability, Cambridge studies in Advanced Mathematics. Cambridge University Press (2002)
Dutta, S., Ghosh, A.K., Chaudhuri, P.: Some intriguing properties of Tukey’s half-space depth. Bernoulli 17, 1420–1434 (2011)
Elmore, R.T., Hettmansperger, T.P., Xuan, F.: Spherical data depth and a multivariate median. In: Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 72, 87 (2006)
Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis. Springer, New York (2006)
Fraiman, R., Muniz, G.: Trimmed means for functional data. TEST 10, 419–440 (2001)
Kosmelj, K., Billard, L.: Clustering of population pyramids using Mal-

Liu, Z., Modarres, R.: Lens depth and median. J. Nonparametr. Stat.

Liu, R.Y., Parelius, J.M., Singh, K.: Multivariate analysis by data depth:

Li, J., Cuesta-Albertos, J.A., Liu, R.Y.: DD-classifier: nonparametric
description of functional data. J. Am. Stat. Assoc. 105, 24–45 (2017)

Serfling, R., Wijesuriya, U.: Depth-based nonparametric description of functional data, with emphasis on use of spatial depth. Comput. Stat. Data Anal. 105, 24–45 (2017)

Sguera, C., Galeano, P., Lillo, R.E.: Spatial depth-based classification for functional data. TEST 23, 725–750 (2014)

Snášel, V., Nowaková, J., Xhafa, F., Baroliš, L.: Geometrical and topological approaches to Big Data. Futur. Gener. Comput. Syst. 67, 286–296 (2017)

Taylor, G.: The Canon and Chronology of Shakespeare’s Plays. Claren-

Paris, T., Vieu, P.: Robust principal component analysis. Statist. Comput. 9, 429–437 (1999)

Vasiliou, V., Vieu, P.: Functional principal component analysis: a survey. J. Multivar. Anal. 115, 106–125 (2013)

Vishwanathan, S.V.N., Schraudolph, N.N., Peternell, M., Bengio, Y., Kriegel, H.-P.: Graph kernel methods for semantically enriched networks and structured data. In: NIPS (2009)

Wang, J., Dong, Y.: Measurement of text similarity: a survey. Inform. Retr. 9, 1–35 (2006)

Wang, J., Dong, Y.: Measurement of text similarity: a survey. Informatica 11, 421 (2020)

Wang, J., Dong, Y.: Measurement of text similarity: a survey. Informatica 11, 421–437 (2020)

Wang, J., Dong, Y.: Measurement of text similarity: a survey. Informatica 11, 421–437 (2020)

Wang, J., Dong, Y.: Measurement of text similarity: a survey. Informatica 11, 421–437 (2020)

Yang, M., Modarres, R.: β-Skeleton depth functions and medians. Commun. Stat. Theory Methods 47, 5127–5143 (2018)

Zuo, Y.: On general notions of depth in regression. Stat. Sci. 36, 142–157 (2021)

Zuo, Y., Serfling, R.: General notions of statistical depth functions. Ann. Stat. 28, 461–482 (2000)

Publisher’s Note Springer Nature remains neutral with regard to juris-
dictional claims in published maps and institutional affiliations.