stats

MDPI

Article

A Geometric Perspective on Functional Outlier Detection

Moritz Herrmann * and Fabian Scheipl

Department of Statistics, Ludwig-Maximilians-University, Ludwigstr. 33, 80539 Munich, Germany; fabian.scheipl@stat.uni-muenchen.de

* Correspondence: moritz.herrmann@stat.uni-muenchen.de; Tel.: +49-89-2180-3198

Abstract: We consider functional outlier detection from a geometric perspective, specifically: for functional datasets drawn from a functional manifold, which is defined by the data's modes of variation in shape, translation, and phase. Based on this manifold, we developed a conceptualization of functional outlier detection that is more widely applicable and realistic than previously proposed taxonomies. Our theoretical and experimental analyses demonstrated several important advantages of this perspective: it considerably improves theoretical understanding and allows describing and analyzing complex functional outlier scenarios consistently and in full generality, by differentiating between structurally anomalous outlier data that are off-manifold and distributionally outlying data that are on-manifold, but at its margins. This improves the practical feasibility of functional outlier detection: we show that simple manifold-learning methods can be used to reliably infer and visualize the geometric structure of functional datasets. We also show that standard outlier-detection methods requiring tabular data inputs can be applied to functional data very successfully by simply using their vector-valued representations learned from manifold learning methods as the input features. Our experiments on synthetic and real datasets demonstrated that this approach leads to outlier detection performances at least on par with existing functional-data-specific methods in a large variety of settings, without the highly specialized, complex methodology and narrow domain of application these methods often entail.

Keywords: functional data analysis; outlier detection; manifold learning; dimension reduction; multidimensional scaling; local outlier factors

check for updates

Citation: Herrmann, M.; Scheipl, F. A Geometric Perspective on Functional Outlier Detection. *Stats* **2021**, *4*, 971–1011. https://doi.org/10.3390/stats4040057

Academic Editor: Manuel Oviedo de la Fuente

Received: 14 September 2021
Accepted: 12 November 2021
Published: 24 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Problem Setting and Proposal

Outlier detection for functional data is a challenging problem due to the complex and information-rich units of observations, which can be "outlying" or unusual in many different ways. Functional outliers are often categorized into magnitude and shape outliers [1,2], whereas Hubert et al. [3] differentiated between isolated and persistent outliers, the latter were further subdivided into shift, amplitude, and shape outliers. However, neither of these taxonomies yield precise, explicit, fully general definitions, which makes it difficult to theoretically describe, analyze, and compare functional outliers. Magnitude outliers, for example, have been defined as functional observations "outlying in some part or across the whole design domain" [1] (p. 1), or as "curves lying outside the range of the vast majority of the data" [2] (p. 2), whereas Hubert et al. [3] (p. 3) defined isolated outliers as observations that "exhibit outlying behavior during a very short time interval", in contrast to persistent outliers, which "are outlying on a large part of the domain".

To cut through the confusion, we propose a geometric perspective on functional outlier detection based on the well-known "manifold hypothesis" [4,5]. This refers to the assumption that ostensibly complex, high-dimensional data lie on a much simpler, lower-dimensional manifold embedded in the observation space and that this manifold's structure can be learned and then represented in a low-dimensional space, often simply called embedding space. We argue that such a perspective both clarifies and generalizes the

concept of functional outliers, without the need for any strong assumptions or prior knowledge about the underlying data-generating process or its outliers. In terms of theoretical development, the approach allows us to consistently formalize and systematically analyze functional outlier detection in full generality. We also demonstrate that procedures based on this perspective simplify and improve functional outlier detection in practice: this suggests a principled, yet flexible approach for applying well-established, highly performant standard outlier-detection methods such as local outlier factors (LOF) [6] to functional data, based on embedding coordinates obtained via manifold learning or dimension-reduction methods. Our experiments show that doing so performs at least on par with existing functional-data-specific outlier-detection methods, without the methodological complexity and limited applicability that methods specific to functional data often entail. Moreover, such lower-dimensional representations serve as an easily accessible visualization and exploration tool that helps uncover complex and subtle data structures that cannot be sufficiently reflected by one-dimensional outlier scores or labels, nor captured by many of the previously proposed 2D diagnostic visualizations for functional outliers.

1.2. Background and Related Work

Functional data analysis (FDA) [7] focuses on data where the units of observation are realizations of stochastic processes over compact domains. In many cases, the intrinsic dimensionality of functional data (FD) is much lower than the observed. First, while FD are infinite-dimensional in theory, they are high-dimensional in practice: functional observations are usually recorded on fine and dense grids of argument values. Second, the dominant drivers of the differences among functional observations are often comparatively low-dimensional, so that just a few modes of variation capture most of the structured variability in the data.

However, FD usually contain shape and translation, as well as phase variation, i.e., both "vertical" and "horizontal" variability. These different kinds of variability contribute to the difficulty of precisely defining and differentiating the various forms of functional outliers and developing methods that can "catch them all", making outlier detection a highly investigated research topic in FDA. For example, Arribas-Gil and Romo [2] argued that the proposed outlier taxonomy of Hubert et al. [3] can be made more precise in terms of expectation functions $f(t)$ and $g(t)$, with $f(t)$ a "common" process; see Figure 1.

$$\begin{array}{l} \textit{Functional outliers} \left\{ \begin{array}{l} \textit{Isolated outliers}: \ g(t)=f(t)+\mathbb{1}\{t\in S\}h(t),\ S\subset U \\ \textit{Persistent outliers} \left\{ \begin{array}{l} \textit{Magnitude outliers}: \ g(t)=\alpha f(t) \\ \textit{Shape outliers}: \ g \textit{ not related to } f \\ \textit{Shift outliers}: \ g(t)=\alpha+f(t),\ g(t)=f(t+\alpha),\ g(t)=f(h(t)) \end{array} \right. \end{array} \right. \\ \textit{Functional outliers} \left\{ \begin{array}{l} \textit{Magnitude outliers} \left\{ \begin{array}{l} \textit{Isolated outliers} \\ \textit{Persistent outliers} \end{array} \right. \\ \textit{Shape outliers} \left\{ \begin{array}{l} \textit{Isolated outliers} \\ \textit{Persistent outliers} \end{array} \right. \end{array} \right. \end{array}$$

Figure 1. Functional outlier taxonomies. Bottom: standard taxonomy. Top: the taxonomy as introduced by Hubert et al. Reprinted by permission from Springer Nature: Springer, Statistical Methods & Applications, Discussion of "Multivariate functional outlier detection", Arribas-Gil Ana, Romo Juan, Copyright 2015.

Despite these attempts, some fundamental issues remain unsolved. The proposed taxonomies do not provide precise definitions, and some of the definitions are contradictory to some extent. Finally, many outlier scenarios for realistic data-generating processes are not covered by the described taxonomies at all. As Arribas-Gil and Romo [2] themselves pointed out that settings with phase-varying data (i.e., "horizontal" variability through elastic deformations of the functions' domains) are not sufficiently reflected, as functions

deviating in terms of phase may be considered as shape outliers in cases where there are only a few such functions, but not in settings where all functions display such variation.

In addition, the taxonomy in Figure 1 provides a reasonable conceptual framework only if the nonoutlying data from the "common" data-generating process is characterized adequately just by its global mean function. This cannot be assumed for many real datasets, which often contain highly variable sets of functions, which display several modes of phase, shape, and/or amplitude variation simultaneously and/or come from multiple classes with class-specific means and higher moments (see Figure 5).

Published research focuses mostly on the development of outlier detection methods specifically for functional data, and a multitude of methods based on a variety of different concepts such as functional data depths [8,9], functional PCA [10], functional isolation forests [11], robust functional archetypoids [12], or functional outlier metrics such as directional outlyingness [13,14], often narrowly focused on detecting specific kinds of functional outliers, have been put forth. Dai et al. [1] proposed a transformation-based approach to functional outlier detection and claimed that sequentially transforming shape outliers, which "are much more challenging to handle", into magnitude outliers makes them easier to detect with established methods [1] (p. 2). The approach allows defining functional outliers more precisely in terms of the transformations being used, such as normalizing or centering functions or taking their derivatives, but practitioners still need to be able to come up with appropriate transformations for the data at hand first.

Recently, Xie et al. [15] introduced a decomposition of functional observations into amplitude, phase, and shift components, based on which specific types of outliers can be identified in a more general geometric framework without necessarily requiring functional data to be of comparatively low rank. Similar in spirit to our proposal, Hyndman and Shang [16] used kernel density estimation and half-space depth contours of two-dimensional robustified FPCA scores to construct functional boxplot equivalents and detect outliers, and Ali et al. [17] used data representations in two dimensions obtained from manifold methods for outlier detection and clustering, but the focus of both was on practicalities without considering the theoretical implications and general applicability of embedding-based approaches, nor did they consider the necessity of higher-dimensional representations. While Hyndman and Shang's HDR boxplots were based on a similar combination of methods as our approach, they did not consider their geometrical foundations and, thus, did not make use of their full potential, firstly by considering only the two largest PCs and secondly by dichotomizing observations into outliers and inliers instead of providing continuous scores of outlyingness. Yu et al. [18] developed a test statistic for outlier detection based on the observed maxima of scaled PC score vectors, i.e., outlyingness defined in terms of a single mode of variation. However, this NHST framework for outlier detection needs to assume both that the common data have a *single* consistent mean function and that all deviations from this mean function are i.i.d. realizations of a mean-zero Gaussian process. Both of these assumptions seem highly restrictive to us and are likely to be untenable in many real-world applications.

The remainder of the paper is structured as follows: We provide a theoretical formalization and discussion of our geometric approach in Section 2. Based on these theoretical considerations, Section 3 presents extensive experiments. Section 3.1 covers a detailed qualitative analysis of real-world data, while Section 3.2 provides quantitative experiments and systematic comparisons to previously proposed methods on complex synthetic outlier scenarios. We conclude with a discussion in Section 4.

2. Functional Outlier Detection as a Manifold-Learning Problem

In this section, we first define two forms of functional outliers from a geometric view point: off- and on-manifold outliers. We then illustrate how this perspective contains and extends existing outlier taxonomies and how it can be used to formalize a large variety of additional scenarios for functional data with outliers.

2.1. The Two Notions of Functional Outliers: Off- and On-Manifold

Our approach to functional outlier detection rests on the manifold assumption, i.e., the assumption that observed high-dimensional data are intrinsically low-dimensional. Specifically, we put forth that observed functional data $x(t) \in \mathcal{F}$, where $\mathcal{F}$ is a function space, arise as the result of a mapping $\phi : \Theta \to \mathcal{F}$ from a (low-dimensional) parameter space $\Theta \subset \mathbb{R}^{d_2}$ to $\mathcal{F}$, i.e., $x(t) = \phi(\theta)$. Conceptually, a d_2-dimensional parameter vector $\theta \in \Theta$ represents a specific combination of values for the modes of variation in the observed functional data, such as level or phase shifts, amplitude variability, class labels, and so on. These parameter vectors are drawn from a probability distribution P over $\mathbb{R}^{d_2}$: $\theta_i \sim P \ \forall \ \theta_i \in \Theta$, with $\Theta = \{\theta : f_P(\theta) > 0\}$ and f_P the density to P. Mapping this parameter space to the function space creates a functional manifold $\mathcal{M}_{\Theta,\phi}$ defined by ϕ and Θ: $\mathcal{M}_{\Theta,\phi} = \{x(t) : x(t) = \phi(\theta) \in \mathcal{F}, \theta \in \Theta\} \subset \mathcal{F}$, and an example is depicted in Figure 2. For $\mathcal{F} = L^2$ with data from a single functional manifold that is isomorphic to some Euclidean subspace, Chen and Müller [19] developed the notions of a manifold mean and modes of variation. Similarly, Dimelgio et al. [20] developed a robust algorithm for template curve estimation for connected smooth submanifolds of $\mathbb{R}^d$.

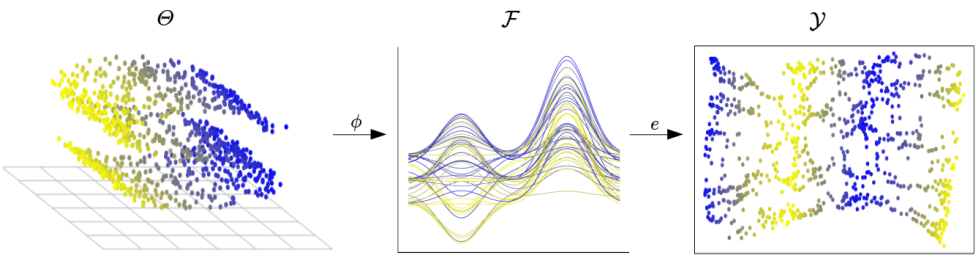


Figure 2. Functional data from a manifold-learning perspective. Image source: Herrmann and Scheipl [21]; use permitted under the Creative Commons Attribution License CC BY-SA 4.0.

Unlike these single-manifold settings, our conceptualization of outlier detection is based on two functional manifolds. That is, we assume a dataset $X = \{x_1(t), \ldots, x_n(t)\}$ with n functional observations coming from two separate functional manifolds $\mathcal{M}_c = \mathcal{M}_{\Theta_c,\phi_c}$ and $\mathcal{M}_a = \mathcal{M}_{\Theta_a,\phi_a}$, with $\mathcal{M}_j \subset \mathcal{F}$, $j \in \{c, a\}$ and $X \subset \{\mathcal{M}_c \cup \mathcal{M}_a\}$, with $\mathcal{M}_c$ representing the "common" data-generating process and $\mathcal{M}_a$ containing anomalous data. Moreover, for the purpose of outlier detection and in contrast to the settings with a single manifold described in the referenced literature, we are less concerned with precisely approximating the intrinsic geometry of each manifold. Instead, it is crucial to consider the manifolds $\mathcal{M}_c$ and $\mathcal{M}_a$ as submanifolds of $\mathcal{F}$, since we require not just a notion of distance between objects on a single manifold, but also a notion of distance between objects on different manifolds using the metric in $\mathcal{F}$. Note that function spaces such as $\mathcal{C}$ or L^2, which are commonly assumed in FDA [22], are naturally endowed with such a metric structure. Both $\mathcal{C}(D)$ and all $L^p(D)$ spaces over compact domain D are Banach spaces for $p \geq 1$ and, thus, also metric spaces [23].

Finally, we assume that we can learn from the data an embedding function $e : \mathcal{F} \to \mathcal{Y}$ that maps observed functions to a d_1-dimensional vector representation $y \in \mathcal{Y} \subset \mathbb{R}^{d_1}$ with $e(x(t)) = y$, which preserves at least the topological structure of $\mathcal{F}$, i.e., if $\mathcal{M}_c$ and $\mathcal{M}_a$ are unconnected components of $\mathcal{F}$, their images under e are also unconnected in $\mathcal{Y}$ and ideally yield a close approximation of the ambient geometry of $\mathcal{F}$.

Definition 1. *Off- and on-manifold outliers in functional data.*

Without loss of generality, let $r = \frac{|\{x_i(t):x_i(t)\in\mathcal{M}_a\}|}{|\{x_i(t):x_i(t)\in\mathcal{M}_c\}|} \lll 1$ be the outlier ratio, i.e., most observations are assumed to stem from $\mathcal{M}_c$. Furthermore, let Θ_c and Θ_a follow the distributions P_c and P_a, respectively. Let $\Omega^*_{\alpha,P}$ be an α-minimum volume set of P for some $\alpha \in (0,1)$, where $\Omega^*_{\alpha,P}$ is defined as a set minimizing the quantile function

$V(\alpha) = \inf_{C \in \mathcal{C}}\{\text{Leb}(C) : P(C) \geq \alpha\}, 0 < \alpha < 1\}$ for i.i.d. random variables in $\mathbb{R}^d$ with distribution P, $\mathcal{C}$ a class of measurable subsets in $\mathbb{R}^d$, and Lebesgue measure Leb [24], i.e., $\Omega^*_{\alpha,P}$ is the smallest region containing a probability mass of at least α.

A functional observation $x_i(t) \in X$ is then:

- An off-manifold outlier if $x_i(t) \in \mathcal{M}_a$ and $x_i(t) \notin \mathcal{M}_c$;
- An on-manifold outlier if $x_i(t) \in \mathcal{M}_c$ and $\theta_i \notin \Omega^*_{\alpha,P_c}$.

To paraphrase, we assume that there is a single "common" process generating the bulk of observations on $\mathcal{M}_c$ and an "anomalous" process defining structurally different observations on $\mathcal{M}_a$. We follow the standard notion of outlier detection in this, which assumes that there are two data-generating processes [1,25,26]. Note that this does not necessarily imply that off-manifold outliers are similar to each other in any way: P_a could be very widely dispersed and/or $\mathcal{M}_a$ could consist of multiple unconnected components representing different kinds of anomalous data. The essential assumption here is that the process from which most of the observations are generated yields structurally relatively similar data. This is reflected by the notion of the two manifolds $\mathcal{M}_c$ and $\mathcal{M}_a$ and the ratio r. We consider settings with $r \in [0, 0.1]$ as suitable for outlier detection. By definition, the number of on-manifold outliers, i.e., distributional outliers on $\mathcal{M}_c$ as opposed to the structural outliers on $\mathcal{M}_a$, only depends on the α-level for Ω^*_{α,P_c}.

Note that outlyingness in functional data is often defined only in terms of shape or magnitude, but the concept ought to be conceived much more generally. The most important aspect from a practical perspective is that any kind of structural difference will be reliably reflected in low-dimensional representations that can be learned via manifold methods, as we show in Section 3. These methods yield embedding coordinates $y \in \mathcal{Y}$ that capture the structure of data and their outliers.

2.2. Methods

To illustrate some of the implications of our general perspective on functional outlier detection and showcase its practical utility, we mostly use metric multidimensional scaling (MDS) [27] for dimension reduction and local outlier factors (LOF) [6] for outlier scoring in the following. Note, however, that the proposed approach is not at all limited to these specific methods, and many other combinations of outlier detection methods applied to lower-dimensional embeddings from manifold-learning methods are possible. However, MDS and LOF have some important favorable properties: First of all, both methods are well understood and widely used and tend to work reliably without extensive tuning since they do not have many hyperparameters. Specifically, LOF only requires a single parameter `minPts`, which specifies the number of nearest neighbors used to define the local neighborhoods of the observations, and MDS only requires specification of the embedding dimension.

More importantly, our geometric approach rests on the assumption that functional outlier detection can be based on some notion of distance or dissimilarity between functional observations, i.e., that abnormal or outlying observations are separated from the bulk of the data in some ambient (function) space. As MDS optimizes for an embedding, which preserves all pairwise distances as closely as possible (i.e., tries to project the data isometrically), it also retains a notion of the distance between unconnected manifolds in the ambient space. This property of the embedding coordinates retaining the ambient space geometry as much as possible is crucial for outlier detection. This also suggests that manifold-learning methods such as ISOMAP [28], t-SNE [29], or UMAP [30], which do not optimize for the preservation of ambient space geometry via isometric embeddings by default, may require much more careful tuning in order to be used in this way. Our experiments support this theoretical consideration, as can be see in Figure 11. For LOF, this implies that larger values for `minPts` are to be preferred here, since such LOF scores take into account more of the global ambient space geometry of the data instead of only the local neighborhood structure. In Section 3, we show that `minPts` $= 0.75n$, with n the

number of functional observations in a dataset, seems to be a reliable and useful default for the range of datasets we consider.

Two additional aspects need to be pointed out here. First, throughout this paper, we compute most distances using the L_2 metric. This yields MDS coordinates that are equivalent to standard functional PCA scores (up to rotation). The proposed approach, however, is not restricted to L_2 distances. Combining MDS with distances other than L_2 yields embedding solutions that are no longer equivalent to PCA scores, and suitable alternative distance measures may yield better results in particular settings. We illustrate this aspect using the L_{10} metric and two phase-specific distance measures in Section 3.3, which we apply to simulated data with isolated outliers and a real dataset of outlines of neolithic arrowheads, respectively. Similarly, using alternative manifold-learning methods could be beneficial in specific settings, as long as they are able to represent not just the local neighborhood structure or on-manifold geometry, but also the global ambient space geometry.

Second, even though the LOF could also be applied directly to the dissimilarity matrix of a functional dataset without an intermediate embedding step, most anomaly-scoring methods cannot be applied directly to such distance matrices and require tabular data inputs. By using embeddings that accurately reflect the (outlier) structure of a functional dataset, any anomaly-scoring method requiring tabular data inputs can be applied to functional data as well. In this work, we apply LOF on MDS coordinates to evaluate whether functional data embeddings can faithfully retain the outlier structure. Furthermore, embedding the data before running outlier-detection methods often provides large additional value in terms of visualization and exploration, as the ECG data analysis in Section 3.1 shows.

2.3. Examples of Functional Outlier Scenarios

We can now give precise formalizations of different functional outlier scenarios and investigate the corresponding low-dimensional representations. In this section, we first show that the geometrical approach is able to describe existing taxonomies (see Figure 1) more consistently and precisely. We then illustrate its ability to formalize a much broader general class of outlier detection scenarios and discuss the choice of the distance metric and the dimensionality of the embedding.

2.3.1. Outlier Scenarios Based on Existing Taxonomies

Structure induced by shape: In the taxonomy depicted in Figure 1, top, the common data-generating process is defined by the expectation function $f(t)$. This can be formalized in our geometrical terms as follows: the set of functions defined by the "common process" $f(t)$ defines a functional manifold (in terms of shape), i.e., the structural component is represented by the expectation function of the common process. That means we can define $\mathcal{M}_c = \{x(t) : x(t) = \theta f(t) = \phi(\theta, t)\}$ or $\mathcal{M}_c = \{x(t) : x(t) = f(t) + \theta = \phi(\theta, t), \theta \in \mathbb{R}\}$. More generally, we can also model this jointly with $\mathcal{M}_c = \{x(t) : \theta_1 f(t) + \theta_2 = \phi(\theta, t), \theta = (\theta_1, \theta_2)' \in \mathbb{R}^2\}$. In each case, magnitude and (vertical) shift outliers as defined in the taxonomy correspond to on-manifold outliers in the geometrical approach, as such observations are elements of $\mathcal{M}_c$. Isolated and shape outliers, on the other hand, are by definition off-manifold outliers, as long as "g is not related to f" is specified as $g \neq \theta f \, \forall \, \theta \in \mathbb{R}$. For example, if we define $\mathcal{M}_a = \{x(t) : x(t) = \theta g(t)\}$, it follows that $\mathcal{M}_c \cap \mathcal{M}_a = \emptyset$. The same applies to isolated outliers, because $g(t) = f(t) + I_U(t)h(t) \neq \theta_1 f(t) + \theta_2$.

Figure 3 shows an example of such an outlier scenario taken from [8]. Following their notation, the two manifolds can be defined as $\mathcal{M}_c = \{x(t) | x(t) = b + 0.05t + \cos(20\pi t), b \in \mathbb{R}\}$ and $\mathcal{M}_a = \{x(t) | x(t) = a + 0.05t + \sin(\pi t^2), a \in \mathbb{R}\}$ with $t \in [0, 1]$ and $a \sim N(\mu = 5, \sigma = 4)$, $b \sim N(\mu = 5, \sigma = 3)$. Note that the off-manifold outliers lie within the mass of data in the visual representation of the curves, whereas in the low-dimensional embedding, they are clearly separable.

However, we argue that the way shape outliers are defined in Figure 1 is too restrictive, as many isolated outliers clearly differ in shape from the main data, but are not captured by the given definition if the shape is considered in terms of "g not related to f". In contrast, the geometrical perspective with its concepts of off- and on-manifold outliers reflects that consistently. Another issue with the considered taxonomy concerns horizontal shift outliers $f(t+\alpha)$ or $f(h(t))$. Aribas-Gil and Romo [2] specifically tackled that aspect in their discussion. They distinguished between situations where "all the curves present horizontal variation" (Case I), which is the no-outlier scenario for them, and situations where only a few phase-varying observations are present (Case II), which constitutes an outlier scenario. Again, the geometric perspective allows reflecting that consistently. In Appendix A, we make these two notions explicit by defining manifolds accordingly.

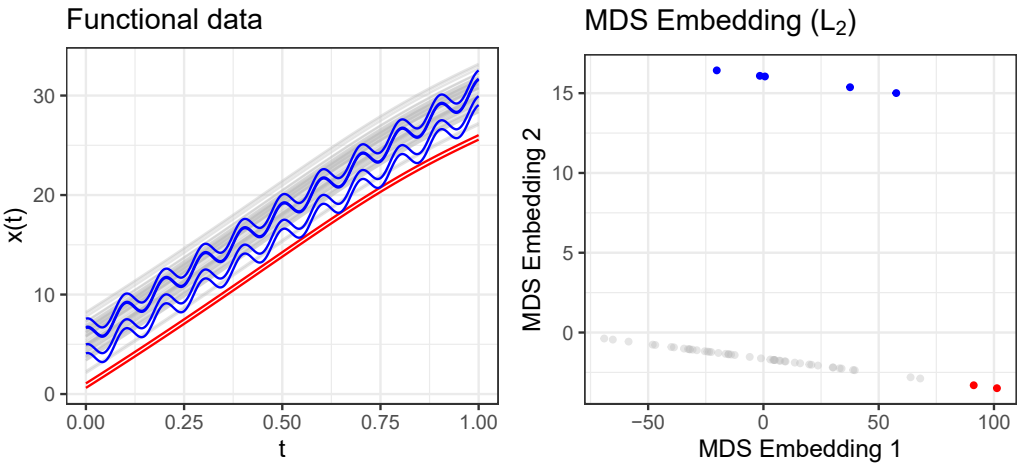


Figure 3. Functional outlier scenario ($n = 54, r = 0.09$) with shape variation inducing structural differences. Off-manifold outliers colored in blue; two on-manifold outliers colored in red.

2.3.2. General Functional Outlier Scenarios

As already noted, the concept of structural difference we propose is much more general. It is straightforward to conceptualize other outlier scenarios with an induced structure beyond shape. Consider the following theoretical example: take a parameter manifold $\Theta \subset [0,\infty] \times [0,\infty] \times [0,\infty] \times [0,\infty]$ and an induced functional manifold $\mathcal{M} = \{f(t); t \in [0,1] : f(t) = \theta_1 + \theta_2 t^{\theta_3} + I(t \in [\theta_4 \pm 0.1])\}$. Each dimension of the parameter space controls a different characteristic of the functional manifold: θ_1 the level, θ_2 the magnitude, θ_3 the shape, and θ_4 the presence of an isolated peak around $t = \theta_4$. One can now define a "common" data-generating process, i.e., a manifold $\mathcal{M}_c$, by holding some of the dimensions of Θ fixed and only varying the rest, either independently or not. On the other hand, one can define an "anomalous" data-generating process, i.e., a structurally different manifold $\mathcal{M}_a$, by letting those fixed in $\mathcal{M}_c$ vary, or simply setting them to values unequal to those used for $\mathcal{M}_c$, or by using different dependencies between parameters than for $\mathcal{M}_c$, e.g., if $\theta_1 = \theta_2$ for $\mathcal{M}_c$, let $\theta_1 = -\theta_2$ for $\mathcal{M}_a$. This implies that one can define data-generating processes so that any functional characteristic (level, magnitude, shape, "peaks", and their combinations) can be on-manifold or off-manifold outliers, depending on how the "common" data manifold $\mathcal{M}_c$ is defined.

Figure 4 shows a setting in which $\mathcal{M}_c$ is defined purely in terms of complex shape variation, while $\mathcal{M}_a$ contains vertically shifted versions of elements in $\mathcal{M}_c$: Let $\mathcal{M}_c$ be the functional manifold of Beta densities $f_B(t;\theta_1,\theta_2)$ with shape parameters $\theta_1, \theta_2 \in [1,2]$, and let $\mathcal{M}_a$ be the functional manifold of Beta densities with shape parameters $\theta_1, \theta_2 \in [1,2]$ shifted vertically by some scalar quantity $\theta_3 \in [0, 0.5]$, that is $\mathcal{M}_c = \{f(t); t \in [0,1] : f(t) = f_B(t;\theta_1,\theta_2)\}$ with $\Theta_c = [1,2]^2$ and $\mathcal{M}_a = \{f(t); t \in [0,1] : f(t) = f_B(t;\theta_1,\theta_2) + \theta_3\}$ with $\Theta_a = \Theta_c \times [0, 0.5]$.

As can be seen in Figure 4, both manifolds contain substantial shape variation that is identically structured, but those from $\mathcal{M}_a$ are also shifted upwards by small amounts. Note

that many shifted observations lie within the main bulk of the data on large parts of the domain. In the 2D embeddings based on unnormalized L_1-Wasserstein distances [31] (also know as the "Earth mover's distance", top right) and 3D embeddings based on standard L_2 distances (bottom right), we see that this structure is captured with high accuracy, even though it is hardly visible in the functional data, with most anomalous observations clearly separated from the common manifold data, whose embeddings are concentrated on a narrow subregion of the embedding space. An observation on $\mathcal{M}_a$ that is very close to $\mathcal{M}_c$, lying well within the main bulk of functional observations, also appears very close to $\mathcal{M}_c$ in both embeddings. This example shows that the two functional manifolds do not need to be completely disjoint, nor yield visually distinct observations for our approach to yield useful results. It also shows that the choice of an appropriate dissimilarity metric for the data can make a difference: a 2D embedding is sufficient for the more suitable Wasserstein distance, which is designed for (unnormalized) densities (top right panel), while a 3D embedding is necessary for representing the relevant aspects of the data geometry if the embedding is based on the standard L_2 metric (lower right panels). For a comparison with currently available outlier visualization methods for this example, see Figure A4 in Appendix D.

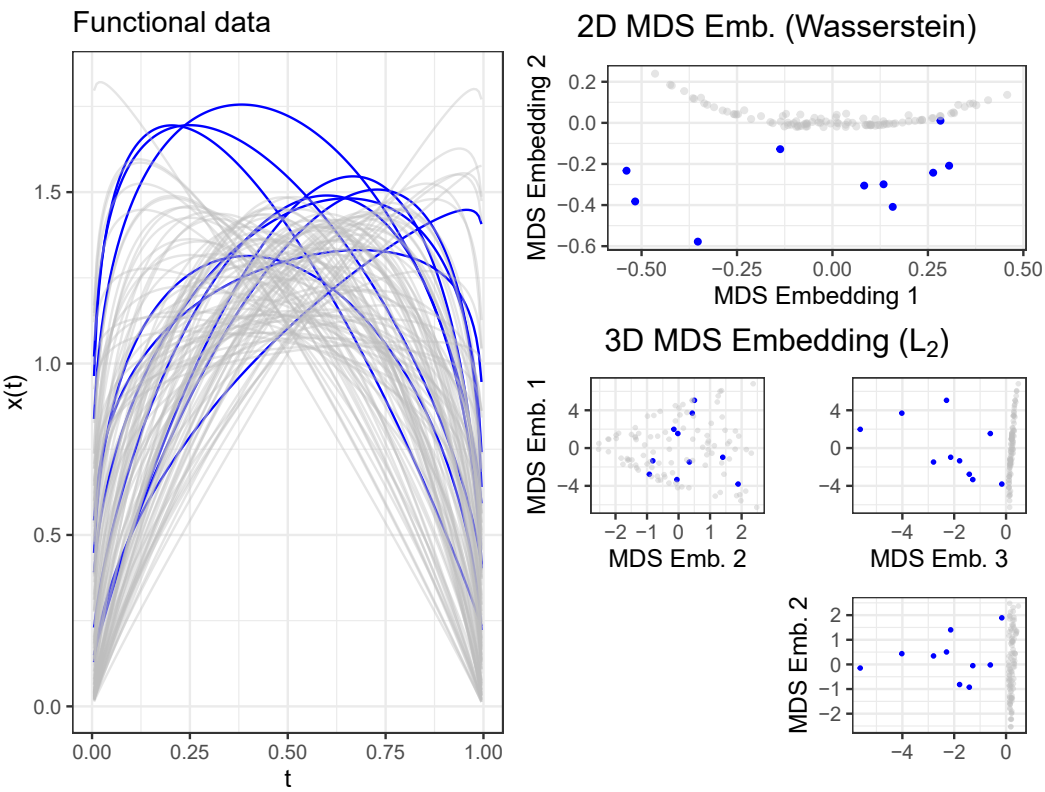


Figure 4. Functional outlier scenario ($n = 100, r = 0.1$) with vertical shifts inducing structural differences. MDS embeddings based on unnormalized L_1-Wasserstein distances and L_2 (Euclidean) distances on the right.

In summary, we propose that the manifold perspective allows defining and representing a very broad range of functional outlier scenarios and data-generating processes. We argue that these properties make the geometrical approach very compelling for functional data, because it is flexible, conceptualizes outliers on a much more general level (for example, structural differences not in terms of shape) than before, and allows theoretically assessing a given setting.

Beyond its theoretical utility of providing a general notion of functional outliers, it has crucial practical implications: outlier characteristics of functional data, in particular structural differences, can be represented and analyzed using low-dimensional representa-

tion provided by manifold-learning methods, regardless of which functional properties define the "common" data manifold and which properties are expressed in structurally different observations. From a practical perspective, on-manifold outliers will appear "connected", whereas off-manifold outliers will appear separated in the embedding, and the clearer these structural differences are, the clearer the separation in the embedding will be. Note that this implies that shape outliers, which pose particular challenges to many previously proposed methods, will often be particularly easily detectable. Moreover, all methods for outlier detection that have been developed for tabular data inputs can be (indirectly) applied to functional data as well based on this framework, simply by using the embedding coordinates as feature inputs. The embedding space $\mathcal{Y}$ is typically a low-dimensional Euclidean space in which conventional outlier detection works well, and the essential geometrical structure encoded in the pairwise functional distance matrix is conserved in these lower-dimensional embeddings. In the next section, we illustrate this practical utility in detail by extensive quantitative and qualitative analyses.

3. Experiments

To illustrate the practical relevance of the outlined geometrical approach, we first qualitatively investigate real datasets. In the second part of this section, we quantitatively investigate the anomaly detection performance of several detection methods based on synthetic data.

3.1. Qualitative Analysis of Real Data

We start with an in-depth analysis of the ECG200 data [32,33], a functional dataset with a complex structure: it seems to contain subgroups with phase and amplitude variation and different mean functions. As a result, the dataset appears visually complex (Figure 5, left). Without the color coding, it would be challenging to identify the three subgroups (as in the lower left plot in Figure 6). Moreover, there are five left-shifted observations (apparent at $t \in [10, 25]$) and a single (partly) vertically shifted outlier (apparent at $t \in [50, 75]$), clearly detectable by the naked eye.

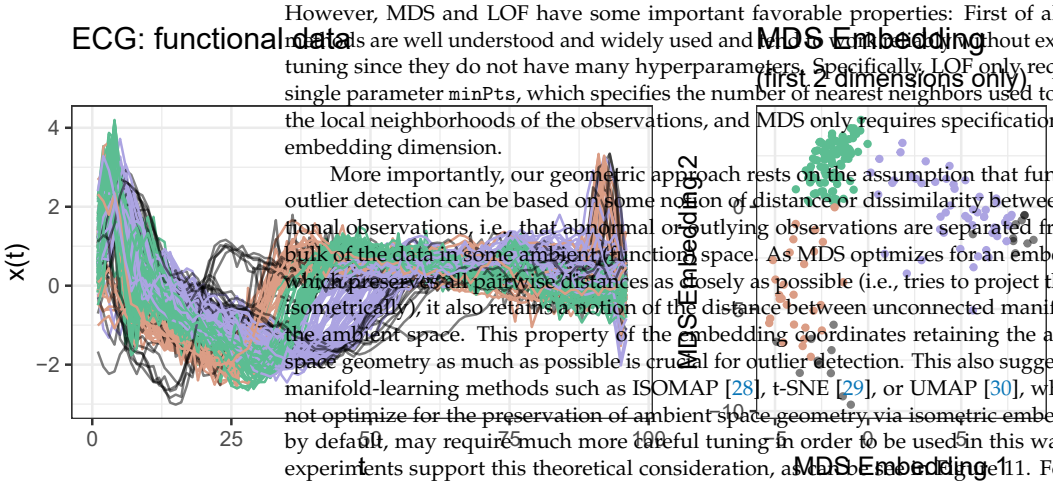


Figure 5. ECG curves and first two embedding dimensions (of five). Colors highlight subgroups apparent in the embeddings. Potential outliers with our embedding LOF scores (`minPts` $= 0.75n$) with the top decile shown in black.

Much of the general structure (and the anomaly structure in particular) becomes evident in a 5D MDS embedding. To begin with, in the first two embedding dimensions, depicted on the right-hand side of Figure 5, three subgroups are easily recognizable. The color coding in Figure 5 is based on this visualization. It makes apparent that the substructures correspond to two smaller, horizontally shifted subgroups of curves (orange: left-shifted, purple: right-shifted) and a central subgroup encompassing the majority of the observations (green). In addition, we computed LOF scores on the 5D embedding

$\mathcal{M}(\alpha) = \inf\{\mathrm{Vol}(C) : P(C) \geq \alpha\}$, $0 \leq \alpha \leq 1$ [24]. Let X be a random variable in $\mathbb{R}^d$ with distribution P, $\mathcal{C}$ a class of measurable subsets in $\mathbb{R}^d$, and Lebesgue measure Leb [24], i.e., $\Omega^*_{\alpha,P}$ is the smallest region containing a probability mass of at least α.

A functional observation $x_i(t) \in \mathcal{X}$ is then:

- An off-manifold outlier if $x_i(t) \in \mathcal{M}_a$ and $x_i(t) \notin \mathcal{M}_c$;
- An on-manifold outlier if $x_i(t) \in \mathcal{M}_c$ and $x_i \notin \Omega^*_{\alpha,P_c}$.

To paraphrase, we assume that there is a single "common" process generating the bulk of observations on $\mathcal{M}_c$ and an "anomalous" process defining structurally different observations in $\mathcal{M}_a$. We follow the standard outlier detection framework, which assumes that there are two data-generating processes [1,25,26]. Note that this does not necessarily imply that off-manifold outliers are similar to each other in any way; P_a could be very widely dispersed and/or $\mathcal{M}_a$ could consist of multiple unconnected components representing different kinds of anomalous data. The essential assumption here is that the process from which most of the observations are generated yields structurally relatively similar data. This is reflected by the notion of the two manifolds $\mathcal{M}_c$ and $\mathcal{M}_a$ and the ratio r. We consider settings with $r \in [0, 0.1]$ as suitable for outlier detection. By definition, the number of off-manifold outliers, i.e., distributional outliers on $\mathcal{M}_a$ as opposed to the structural outliers on $\mathcal{M}_c$, only depends on the α in Ω^*_{α,P_c}.

Note that outliers in functional data can be defined only in terms of shape or magnitude, but the concept ought to be conceived much more generally. The most important aspect from a practical perspective is that any kind of structural difference will be reliably reflected in low-dimensional representations that can be learned via manifold methods, as we show in Section 3. These methods yield embedding coordinates $y \in \mathcal{Y}$ that capture the structure of the data.

2.2. Methods

To illustrate some of the implications of our general perspective on functional outlier detection and showcase its practical utility, we mostly use multivariate multidimensional scaling (MDS) [27] for dimension reduction and local outlier factors (LOF) [4] for outlier scoring in the following. Note however, that the proposed approach is not at all limited to these specific methods, and many other combinations of outlier detection methods applied to lower-dimensional embeddings from manifold-learning methods are possible. However, MDS and LOF have some important favorable properties: First of all, both methods are well understood and widely used and MDS is computationally without extensive tuning since they do not have many hyperparameters. Specifically, LOF only requires a single parameter `minPts`, which specifies the number of nearest neighbors used to define the local neighborhoods of the observations, and MDS only requires specification of the embedding dimension.

More importantly, our geometric approach rests on the assumption that functional outlier detection can be based on some notion of distance or dissimilarity between functional observations, i.e., that abnormal or outlying observations are separated from the bulk of the data in some ambient function space. As MDS optimizes for an embedding, which preserves all pairwise distances as closely as possible (i.e., tries to project the data isometrically), it also retains a notion of the distance between unconnected manifolds in the ambient space. This property of the embedding coordinates retaining the ambient space geometry as much as possible is crucial for outlier detection. This also suggests that manifold-learning methods such as ISOMAP [28], t-SNE [29], or UMAP [30], which do not optimize for the preservation of ambient space geometry via isometric embeddings by default, may require much more careful tuning in order to be used in this way. Our experiments support this theoretical consideration, as we show in Figure 11. For LOF, this implies that larger values for `minPts` are to be preferred here, since such LOF scores take into account more of the global ambient space geometry rather than just a fixed number of nearest neighbors. In Section 3, LOF scores (`minPts` $= 0.75n$), with n the

coordinates. The observations with LOF scores in the top decile are shown in black in Figure 5. This set contains all the clearly outlying observations.

More importantly, note that these observations are clearly separated from the rest in the 5D embedding shown in Figure 6: the five clearly left-shifted observations in the fourth embedding dimension and the single vertically shifted observation in the subspace spanned by the first and third embedding dimension. The figure shows a scatterplot matrix of all five embedding dimensions with observations color-coded according to the 5D-embedding LOF scores. The clearly left-shifted outliers obtain the highest LOF scores due their isolation in the subspaces including the fourth embedding dimension. Note, moreover, that other observations with higher LOF scores appear in peripheral regions of the different subspaces, but they are not as clearly separable as the six observations described before. Regarding Figure 7A, which shows the 20 most outlying curves according to LOF scores, this can be explained by the fact that these other observations stem from one of the two shifted subgroups and can thus be seen as on-manifold outliers, whereas

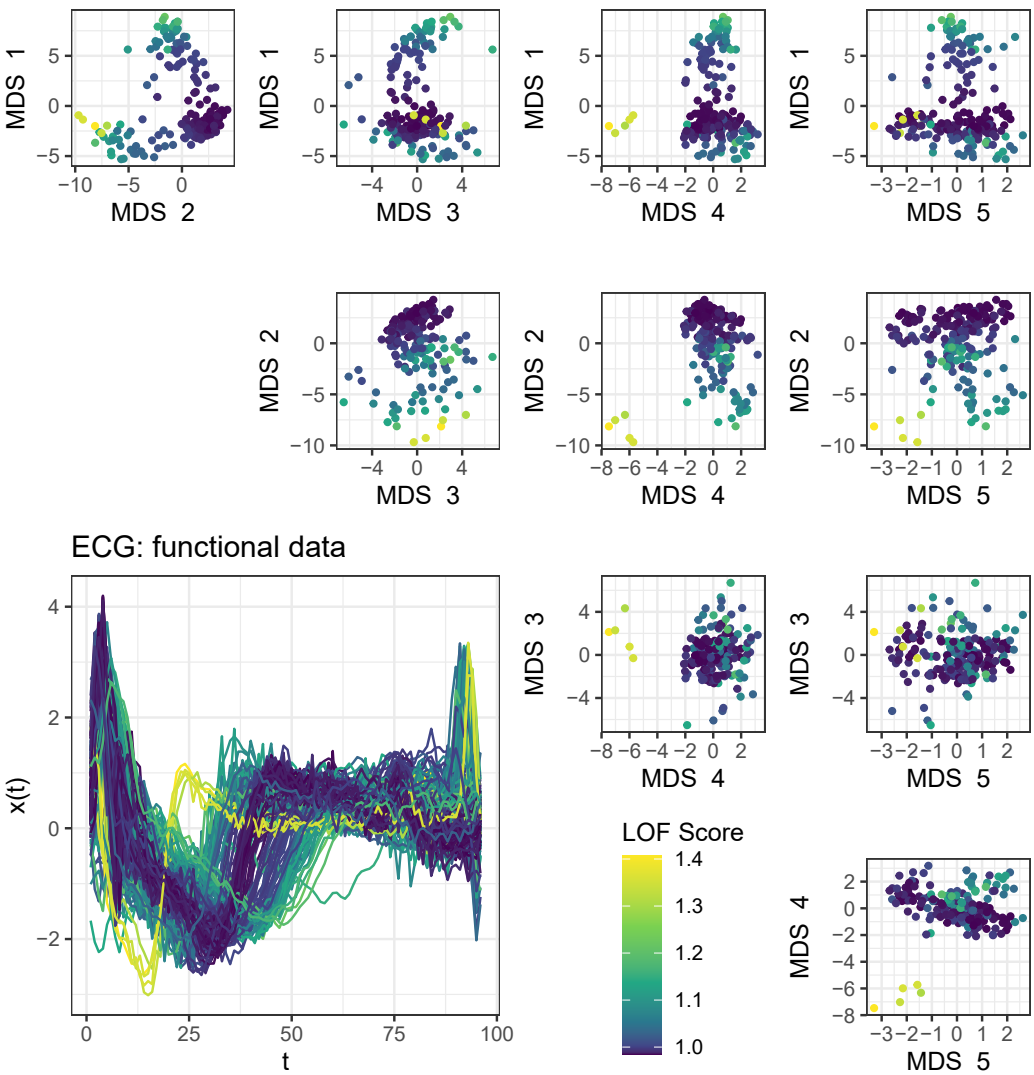


Figure 6. ECG data: scatterplot matrix of all 5 MDS embedding dimensions and curves; lighter colors for the higher LOF scores of 5D embeddings.

We contrast these findings with the results of directional outlyingness [14,34], which performs very well (see Section 3.2) on simple synthetic datasets. Figure 7 shows the ECG curves color-coded by the variation of directional outlyingness (B), the 20 most outlying curves by the variation of directional outlyingness (C), and the observations labeled as

outliers by directional outlyingness respectively by the MS-plot (D). First of all, it can be seen that many observations yield a high variation of directional outlyingness, and observations in the right-shifted subgroup obtain most of the highest values. In fact, among the twenty observations with the highest variation of directional outlyingness, only one is from the left-shifted group, and thirteen are from the right-shifted group. Moreover, applying directional outlyingness to this dataset results in 72 observations being labeled as outliers, which is about 36% of all observations. We would argue that it is questionable whether 36% of all observations should be labeled as outliers.

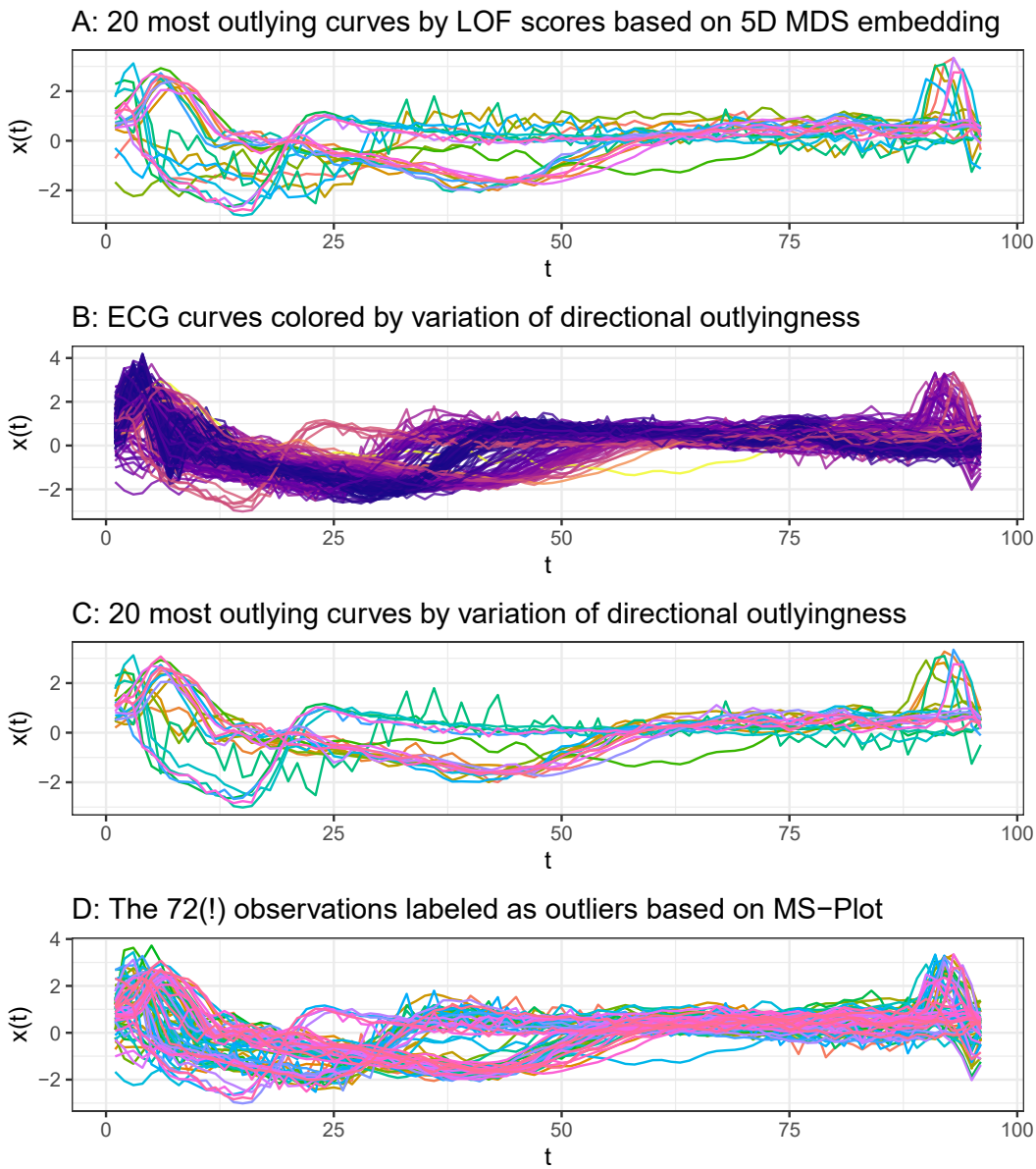


Figure 7. ECG data: LOF on MDS embeddings in contrast to directional outlyingness.

In this regard, the ECG data serve as an example that illustrates the advantages of the geometric approach. First of all, it yields readily available visualizations, which reveal much more of the inherent structure of a dataset than just its anomaly structure. This is specifically important for data with a complex structure (i.e., subgroups or multiple modes and large variability). Moreover, it allows applying well-established and powerful outlier scoring methods such as LOF to functional data. This exemplifies that the approach not only improves theoretical understanding and consideration as outlined in the previous section, it also has large practical utility in complex real data settings in which previously proposed methods may not provide useful answers.

In the ECG example, we saw that a 5D embedding yielded reasonable results and sufficiently reflected many aspects of the data. In particular, the extremely left-shifted observations became clearly separable in the fourth embedding dimension. In Appendix E, we analyze a synthetic dataset in the same way as the ECG data, which yields similar findings. Moreover, note that the Spearman rank correlation between LOF scores computed on the 5D embedding and LOF scores computed directly on the ECG data distances is 0.99. This shows that the outlier structure retained in the 5D embedding is highly consistent with the outlier structure in the high-dimensional observation space, an important aspect with respect to anomaly-scoring methods requiring (low-dimensional) tabular inputs.

Finally, note that even fewer than five embedding dimensions may suffice to reflect much of the inherent structure. Consider the examples depicted in Figure 8, which shows the functional observations and the first two embedding dimensions of a corresponding 5D MDS embedding of another four real datasets. The Octane data consist of spectra from 60 gasoline samples [35,36], the Spanish weather data of annual temperature curves of 73 weather stations [37], the Tecator data of spectrometric curves of meat samples [37,38], and the Wine data of spectrometric curves of wine samples [32,39]. As before, the observations are colored according to LOF scores based on the 5D embedding. In addition, the 12 observations with highest LOF scores are depicted as triangles. These datasets are much simpler than the ECG data, and the first two embedding dimensions already reflect the (outlier) structure fairly accurately: observations with high LOF scores appear separated in the first two embedding dimensions, and more general substructures are revealed as well. The substructure of the weather data is rather obvious already regarding the functional observations, for example, the observations with less variability in terms of temperature, all of which obtained high LOF scores. The substructure of the wine data—for example, the small cluster in the lower part of the embedding—is much harder to detect based on visualizations of the curves alone.

Appendix B summarizes a more detailed analysis of the sensitivity of the approach to the choice of the dimensionality of the embedding. We conclude that sensitivity seems to be fairly low. For all five real datasets we considered, the rank order of LOF scores is very similar or even identical whether based on two-, five-, or even twenty-dimensional embeddings (cf. Table A1).

Following Mead [40], we quantified the goodness of fit (GOF) for a d_1-dimensional MDS embedding as: $GOF(d_1) = \frac{\sum_{i=1}^{d_1} \max(0,\lambda_i)}{\sum_{j=1}^{n} \max(0,\lambda_j)}$, where λ_k are the eigenvalues (sorted in decreasing order) of the kth eigenvectors of the centered distance matrix. For all of the considered real datasets, a 5D embedding achieved a goodness of fit over 0.8, the four less-complex examples even over 0.95 (see Figure A2). As a rule of thumb, the embedding dimension does not seem crucial as long as the goodness of fit (GOF) of the embedding is over 0.8 for L_2 distances. This rule of thumb also yielded compelling quantitative performance results, as shown in Section 3.2.

Figures 6 and 8 show visualizations that combine MDS embeddings with LOF outlier scores. To put them into context, we compare them to existing visualization techniques in this section. For the sake of clarity, only the results are summarized here. The figures for the various alternative methods can be found in Appendix D. Figure A5 shows the results for the MBD-MEI "Outliergram" by Aribas-Gil and Romo [41] (implementation: [42]) for shape outlier detection and the magnitude–shape plot method of Dai and Genton [34]. Figures A6 and A7 show the results for the translation–phase–amplitude boxplots by Xie et al. [15] and the elastic depth boxplot for shape outlier detection by Harris et al. [9]. Finally, Figures A8–A13 show the corresponding functional and bivariate HDR boxplots by Hyndman and Shang [16] (implementation: [43]). Considering the MBD-MEI outliergram and the magnitude–shape plots, both of these visualization methods mostly fail to identify shift outliers (by design, in the case of the outliergram). The outliergram tends to mislabel very central observations as outliers in datasets with little shape variability (e.g., the supposed "shape outliers" detected by MBD-MEI in the central region of the Tecator data) and fails to detect even egregious shape outliers in datasets with high variability (e.g., not

a single MBD-MEI outlier in ECG 200), as well as shape outliers that are also outlying in their level (e.g., the three shape outliers identified by `msplot` in the upper region of the Tecator data). Note that some central functions of the Spanish weather data, which are labeled as outliers by the magnitude–shape plot (and partly by the outliergram) are also reflected in the 2D embedding in Figure 8.

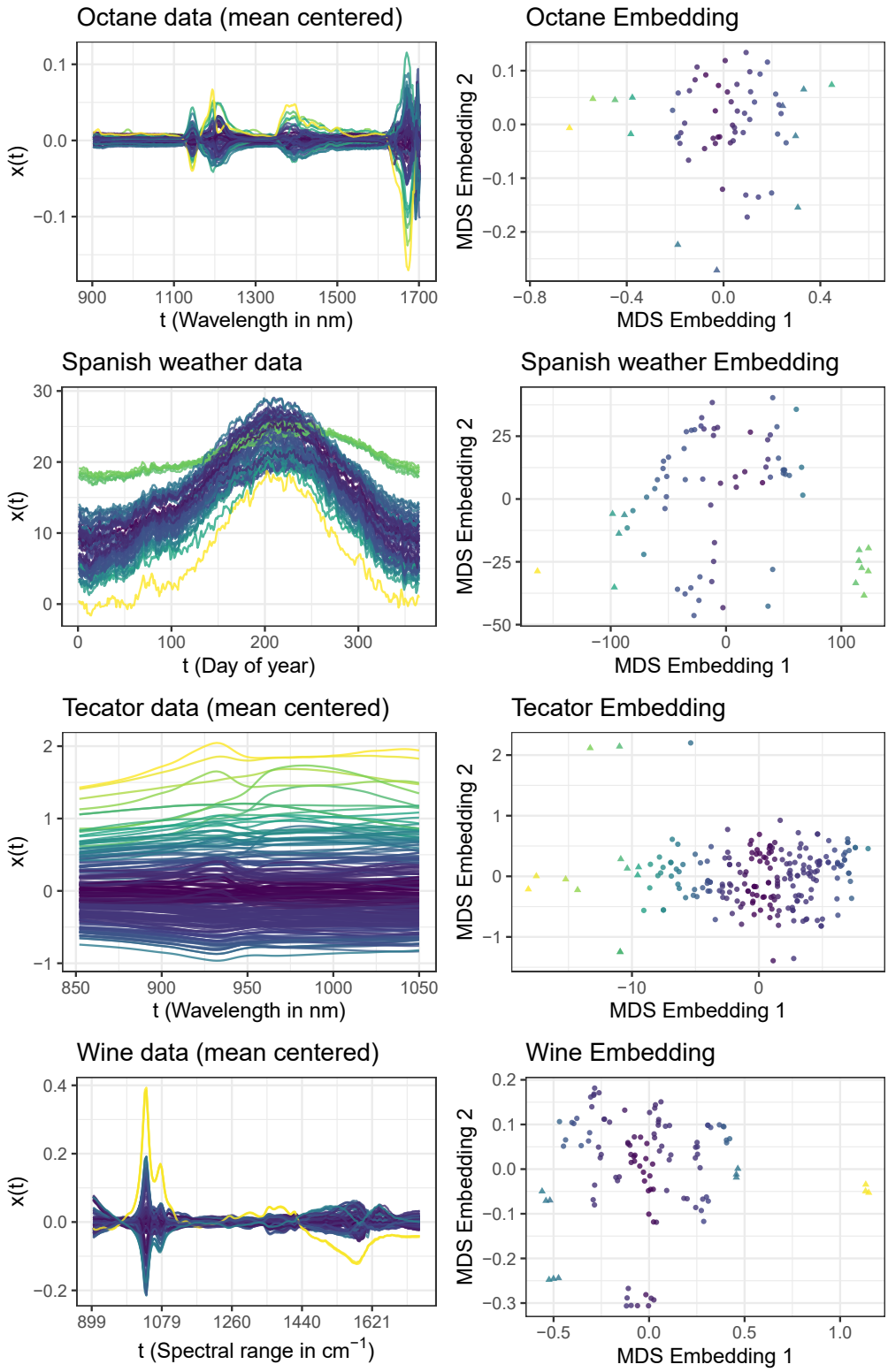


Figure 8. Further examples of real functional data colored by LOF score. The 12 most outlying observations depicted as triangles in the embedding.

They are fairly numerous relative to the overall sample size and are very similar to each other. As such, they form a clearly defined separate cluster within the data, which can be seen in the middle bottom part of the embedding. The translation–phase–amplitude boxplots mostly fail to detect outliers in data with high variability: no outliers at all are detected for the Spanish weather data despite their visually apparent anomalies, and only a single translation outlier is detected for the ECG data. Moreover, the implementation of the approach seems to break down for data with very little variation, and it was not possible to compute the phase boxplot for the Wine data, a dataset with almost no variability in terms of phase.

The results of the elastic depth boxplots do not seem to be consistent over all considered datasets. The results appear reasonable for the Octane, the Wine, and, in part, for the ECG data, where both amplitude and phase outliers are detected. However, in the ECG data, mostly observations from the left-shifted subgroup are detected as phase outliers and only two from the right-shifted subgroup. The results for the Spanish weather and the Tecator data are even less convincing. Among the Tecator data, the method labels 41 curves, i.e., 19% of all observations, as outliers, while it does not discover a single outlier in the Spanish weather data. Note, however, that the elastic depth boxplots are more robust than the translation–phase–amplitude boxplots. While the latter method only detected a single translation outlier and was not able to compute the phase boxplot for the Wine data at all, the elastic depth boxplots detect several amplitude outliers and simply do not yield phase outliers.

Finally, HDR boxplots based on PC projections of the data yield mostly similar results as the L_2-distance-based MDS embeddings. However, we would argue that dichotomizing the observations into inliers and outliers by a fixed outlier threshold makes the visualizations much less suited as an exploratory tool. Consider, for example, the Spanish weather data. The small cluster of observations with rather constant temperature ($\sim$17–25°) does not fall into the outlier region according to the dichotomization threshold, and so, they are also not shown individually in the functional HDR boxplots. Whether they are considered to be outliers or rather a subgroup surely depends on the observer, but we would argue that an outlier visualization method should emphasize and not hide such structures. Our approach of colors according to continuous scores does that very well, reflecting at the same time both the general and the outlier structure. More importantly, the outlier structure of the ECG dataset is not captured in the embedding used by the HDR boxplots. As outlined, more than two embedding dimensions are necessary to fully reflect the outlier structure of this dataset, and the density estimators underlying the HDR boxplot will break down fairly rapidly as the number of embedding dimensions increases. As such, the available implementation is limited to only using the first two PC scores for the embedding, regardless of the actual rank of the underlying data.

3.2. Quantitative Analysis of Synthetic Data

In this section, we investigate the outlier detection performance quantitatively, based on synthetic datasets for which the true (outlier) structure is known.

3.2.1. Methods

In addition to applying LOF to 5D embeddings and directly to the functional data, we investigate the performance of four "functional data"-specific outlier-detection methods: directional outlyingness (DO) [14,34], total variational depth (TV) [44], elastic depth (ED_amp, ED_pha) [9], and the approach based on translation, phase, and amplitude boxplots (AP_BOX) presented by Xie et al. [15]. For the first two methods, we use implementations provided by the package `fdaoutlier` [45] and use the variation of directional outlyingness as returned by the function `dir_out` as outlier scores for DO and the total variation depths as returned by the function `total_variation_depth` for TV. For the latter two methods, we use implementations provided by Harris et al. [9]. Outlier scores for these methods are based on elastic depths as computed by the function `depth.R1` from

the package `elasticdepth` [9] and time-warped functions as computed by the function `time_warping` from the package `fdasrvf` [46]. Note that the elastic depth approach does not produce a single outlier score per observation, but scores amplitude and phase outliers separately. Both amplitude (ED_amp) and phase (ED_pha) scores are shown in Figure 9.

3.2.2. Data-Generating Processes

The methods are applied to data from four different data-generating processes (DGPs), the first two of which are based on the simulation models introduced by Ojo et al. [25] and provided in the corresponding R package `fdaoutlier` [45]. We also provide the results of additional experiments based on the original DGPs from the package `fdaoutlier` in Appendix C. However, we consider most of these DGPs as too simple for a realistic assessment, as most methods achieve almost perfect performance on them, and we use more complex DGPs here. In both DGPs 1 and 2, the inliers from `simulation_model1` from the package `fdaoutlier` serve as $\mathcal{M}_c$, i.e., the common data-generating process. This results in simple functional observations with a positive linear trend. In addition, `simulation_model1` generates simple shift outliers. Additionally, our DGP 1 also includes shape outliers stemming from `simulation_model8`, which serves as $\mathcal{M}_a$. In contrast, DGP 2 contains shape outliers from all of the other DGPs in `fdaoutlier`, which means $\mathcal{M}_a$ contains observations from several different data-generating processes.

For DGPs 3 and 4, we define $\mathcal{M}_c$ by generating a random, wiggly template function over $[0, 1]$ for each dataset, generated from a B-spline basis with 15 or 25 basis functions, respectively, with i.i.d. $\mathcal{N}(0, 1)$ spline coefficients. Functions in $\mathcal{M}_c$ are generated as elastically deformed versions of this template, with random warping functions drawn from the ECDFs of Beta(a, b) distributions with $a, b \sim U[4, 6]$ (DGP 3) or $a, b \sim U[3, 8]$ (DGP 4). Functions in $\mathcal{M}_a$ are also generated as elastically deformed versions of this template, with Beta ECDF random warping functions with $a, b \sim U[3, 4]$ for DGP 3 and with 50:50 Beta mixture ECDF random warping functions with $a, b \sim 0.5U[3, 8] : 0.5U[0.1, 3]$ (DGP 4). Finally, white noise with $\sigma = 0.1, 0.15$, respectively, for DGPs 3 and 4 is added to all resulting functions. Appendix F shows visualizations of example datasets drawn from these DGPs.

3.2.3. Performance Assessment

From these four DGPs, we sampled data $B = 500$ times with three different outlier ratios $r \in \{0.1, 0.05, 0.01\}$. Based on the outlier scores, we computed the area under the ROC curve (AUC) and Mathew's correlation coefficient (MCC) as the performance measures and report the results over all 500 replications. Note that, for $r \in \{0.1, 0.05\}$, the number of sampled observations was $n = 100$, whereas for $r = 0.01$, we sampled $n = 1000$ observations. Since computing the elastic depths and time-warped functions requires more than an hour for a single dataset with 1000 observations, we only included them for the settings with 100 observations.

3.2.4. Results

We note that LOF applied directly to functional data distances yielded very similar results as LOF applied to their 5D embeddings. This agrees with our findings in the qualitative analyses. In the following, we simply refer to the geometrical approach and do not distinguish between the LOF based on MDS embeddings and the LOF applied directly to the functional distance matrix. Figure 9 shows that the proposed geometrical approach is highly competitive with existing functional-data-specific outlier-detection methods. It yields better results than TV for all of the four DGPs and performs at least on par with DO. In comparison to DO, it performs better on DGP 1 and DGP 3, on par on DGP 4, and worse on DGP 2. Note that DO struggles to detect simple shift outliers: among these methods, it performs worst on the first DGP. Similar conclusions can be reported for other settings, where it performs even worse if there are only shift outliers (cf. Figures A3 and A15). Moreover, while the approaches based on elastic depth proposed by Harris et al. (ED_amp

and ED_pha) and the approach proposed by Xie et al. (AP_BOX) perform well on DGP 2, they are outperformed by DO in this setting, and on DGPs 1, 3, and 4, they clearly perform the worst. Thus, these two methods yield the worst performances overall.

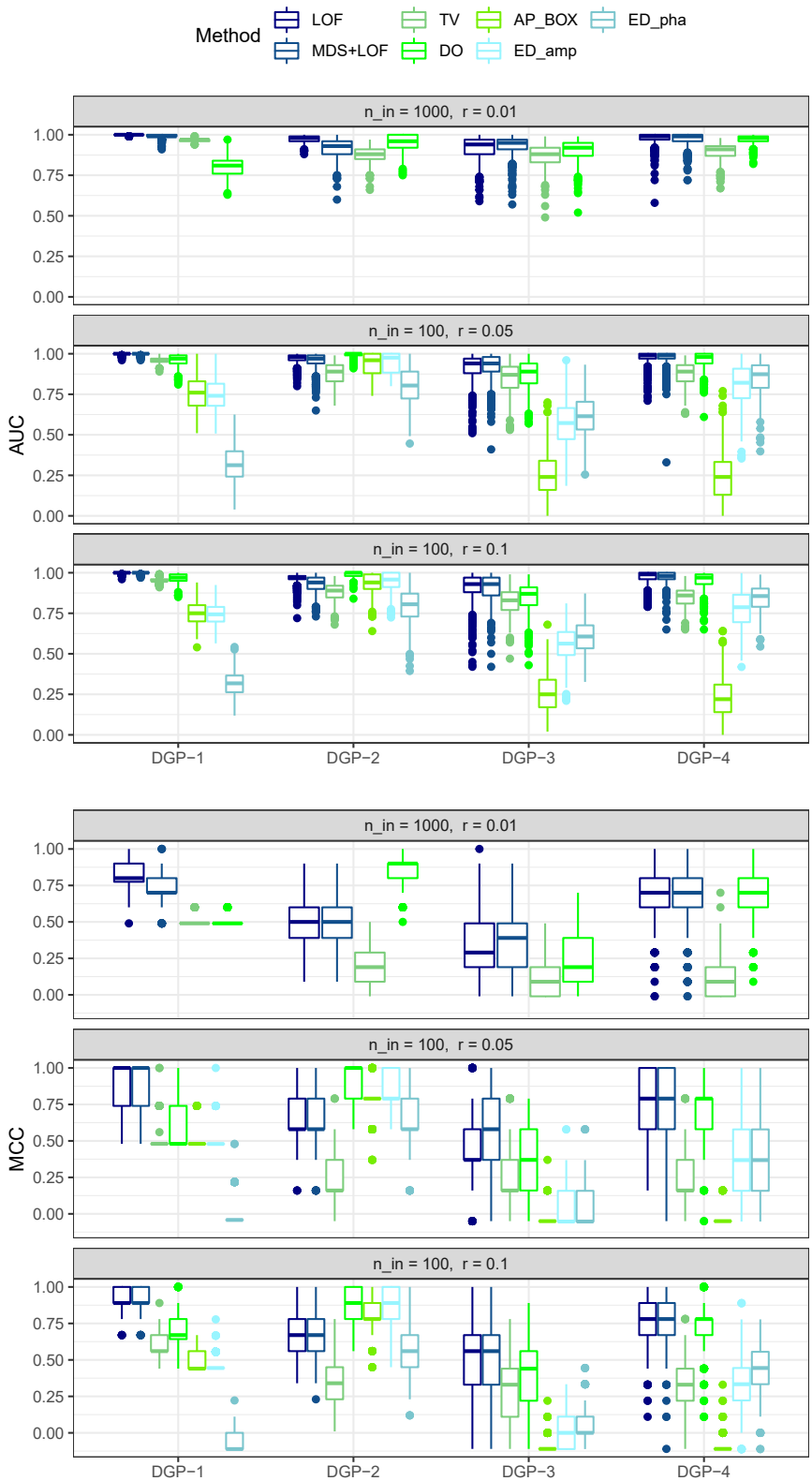


Figure 9. Distribution of the AUC and MCC over the 500 replications for the different data-generating processes (DGPs), outlier-detection methods, and outlier ratios r.

Note that the insights we gain on synthetic data are confirmed by all of the real data applications we investigate in Section 3.1. In addition to the experiments conducted here, we applied the considered methods and their accompanying visualization approaches to these five real datasets. The results of the previously proposed visualizations are presented in detail in Appendix D, Figures A5–A7. In contrast to the proposed geometrical approach, none of them yields satisfactory results consistently for all of the considered datasets. For example, the outliergram, as well as the approach based on translation, phase and amplitude boxplots and the elastic depth approach fail to identify any outliers in some of these datasets, while the magnitude–shape plot, for example, labels an entire third of all observations in the ECG data as outliers (as already outlined in Section 3.1).

In summary, based on the conducted experiments, the proposed geometrical approach yields very compelling results: On synthetic data, it leads to outlier scoring performances at least on par with specialized functional-outlier-detection methods even in its simplest version (MDS with L_2 distances and LOF). Moreover, in contrast to the other methods, it yields consistently useful and sensible results on all of the considered real datasets, while providing more intuitive and more easily interpretable visualizations. Going further, our approach can be adapted to specific settings simply by choosing metrics other than L_2. As the next section shows, this can improve the outlier-detection performance considerably.

3.3. General Dissimilarity Measures and Manifold Methods

So far, we have computed MDS embeddings mostly based on L_2 distances. In the following, we show that the approach is more general. The geometric structure of a dataset is captured in the matrix of pairwise distances among observations. Different metrics emphasize different aspects of differences in the data and can thus lead to different geometries. MDS based on L_2 distances yielded compelling results in many of the examples considered above, but other distances are likely to lead to better performance in certain settings. To illustrate the effect, we consider two additional settings—one simulated and one on real data—in the following. The results are displayed in Figure 10.

The simulated setting is based on isolated outliers, i.e., observations that deviated from functions in $\mathcal{M}_c$ only on small parts of their domain. In such settings, higher-order L_p metrics lead to better results, since such metrics amplify the contribution of small segments with large differences to the total distance. We use as an example data generated from `simulation_model2` from the package `fdaoutlier`. Figure 10A shows the AUC values of LOF scores on MDS embeddings based on L_2 and L_{10} distances. Again, 500 datasets were generated form the model over different outlier ratios. In contrast to L_2-based MDS, using L_{10} distances yielded almost perfect detection. In embeddings based on L_{10}, isolated outliers are clearly separable in the first two or three embedding dimensions.

As a second example, we consider the *ArrowHead* dataset [47,48], which contains outlines of three different types of neolithic arrowheads (see Appendix G for visualizations of the dataset). Using the 78 structurally similar observations from class "Avonlea" as our data on $\mathcal{M}_c$ and sampling outliers from the 126 structurally similar observations from the other two classes, we can compute AUC values based on the given class labels. We generate 500 datasets for each outlier ratio $r \in \{0.05, 0.1\}$. Since there are only 78 observations in the class "Avonlea", we do not use $r = 0.01$ for this example. Embeddings are computed using three different dissimilarity measures: the standard L_2 metric, the unnormalized L_1-Wasserstein metric [31], and the dynamic time warping (DTW) distance [49]. Note that the DTW distance does not define a proper metric [50].

Figure 10B shows that small performance improvements can be achieved in this case if one uses dissimilarity measures that are more appropriate for the comparison of shapes, but not as much as in the isolated outlier example. Note that even though the DTW distance is not a proper metric, it improves the outlier-scoring performance in this example. This indicates that, from a practical perspective, general dissimilarity measures can be sufficient for our approach to work. This opens up further possibilities, as there are many general dissimilarity measures for functional data, for example the semimetrics introduced by

Fuchs et al. [51]. Overall, these examples illustrate the generality of the approach: using suitable dissimilarity measures can make the respective structural differences more easily distinguishable.

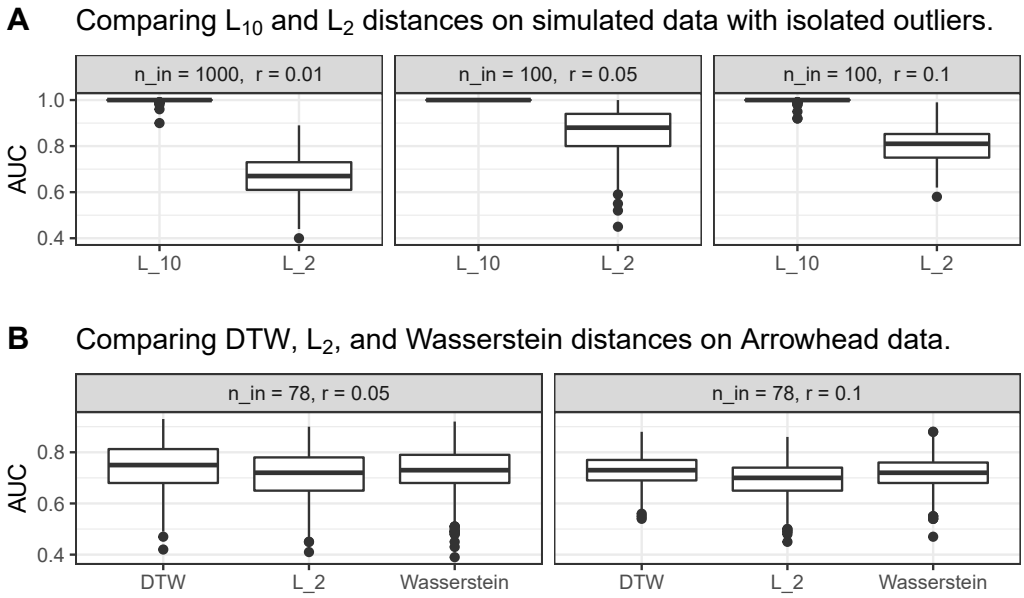


Figure 10. Comparing the effects of different distance measures. Depicted are the distributions of the AUC over 500 replications for the LOF based on MDS embeddings computed with the respective distance measures for different outlier ratios r. (**A**) Comparing the L_{10} and L_2 metrics on a dataset with isolated outliers generated via Simulation Model 2 from the package fdaoutlier. (**B**) Comparing the DTW, L_2, and unnormalized L_1-Wasserstein distance measures on the real dataset ArrowHead. Note: the DTW distance is not a metric.

More complex embedding methods, on the other hand, do not necessarily lead to better or even comparable results as MDS. Figure 11 shows the distribution of the AUC for embedding methods ISOMAP and UMAP. Both methods require a parameter that controls the neighborhood size used to construct a nearest neighbor graph from which the manifold structure of the data is inferred. The larger this value, the more of the global structure is retained. For both methods, embeddings were computed for very small and very large neighborhood sizes of five and ninety.

The results show that neither method performs better than MDS; UMAP even performs considerably worse. Note that ISOMAP is equivalent to MDS based on the geodesic distances derived from the nearest neighbor graph, and the larger the neighborhood size the more similar to direct pairwise distances these geodesic distances become. This is also reflected in the results, as ISOMAP-90 performs better than ISOMAP-5 on average. For DGP-2, ISOMAP-90 slightly outperforms MDS, indicating that more complex manifold methods could improve the results somewhat in specific settings.

In general, however, these findings confirm the theoretical considerations sketched in Section 2.2. Embedding methods that preserve the geometry of the space $\mathcal{F}$ of which $\mathcal{M}_c$ and $\mathcal{M}_a$ are submanifolds, i.e., the ambient space geometry, are more suited for outlier detection than methods that focus on approximating the intrinsic geometry of the manifold(s). Thus, more sophisticated embedding methods, which often focus on approximating the intrinsic geometry, should not be applied lightly and certainly require careful parameter selection in order to be applicable for outlier detection. Since hyperparameter tuning for unsupervised methods remains an unsolved problem, this is unlikely to be achieved in real-world applications. In particular, consider that both UMAP and t-SNE [29] have been found to be—in general—oblivious to local density, which means that clusters of different density in the observation space tend to become clusters of more equal density in the embedding space [52]. Although there may exist a parameter setting where this effect is

reduced (note that there are now density-preserving versions of t-SNE and and UMAP [52]), we are skeptical that outliers can be faithfully represented in such an embedding given the difficulties of hyperparameter tuning in unsupervised settings. Moreover, these methods are not designed to preserve important aspects of the outlier structure. For example, UMAP is subject to a local connectivity constraint, which ensures that every observation is at least connected to its nearest neighbor (in more technical terms: that a vertex in the fuzzy graph approximating the manifold is connected by at least one edge with an edge weight equal to one [30]), which makes it unlikely that UMAP can be tuned so that it is able to sensibly embed off-manifold outliers, which should, by definition, not be connected to the common data manifold. The poor performance of UMAP embeddings in our experiments confirms these concerns.

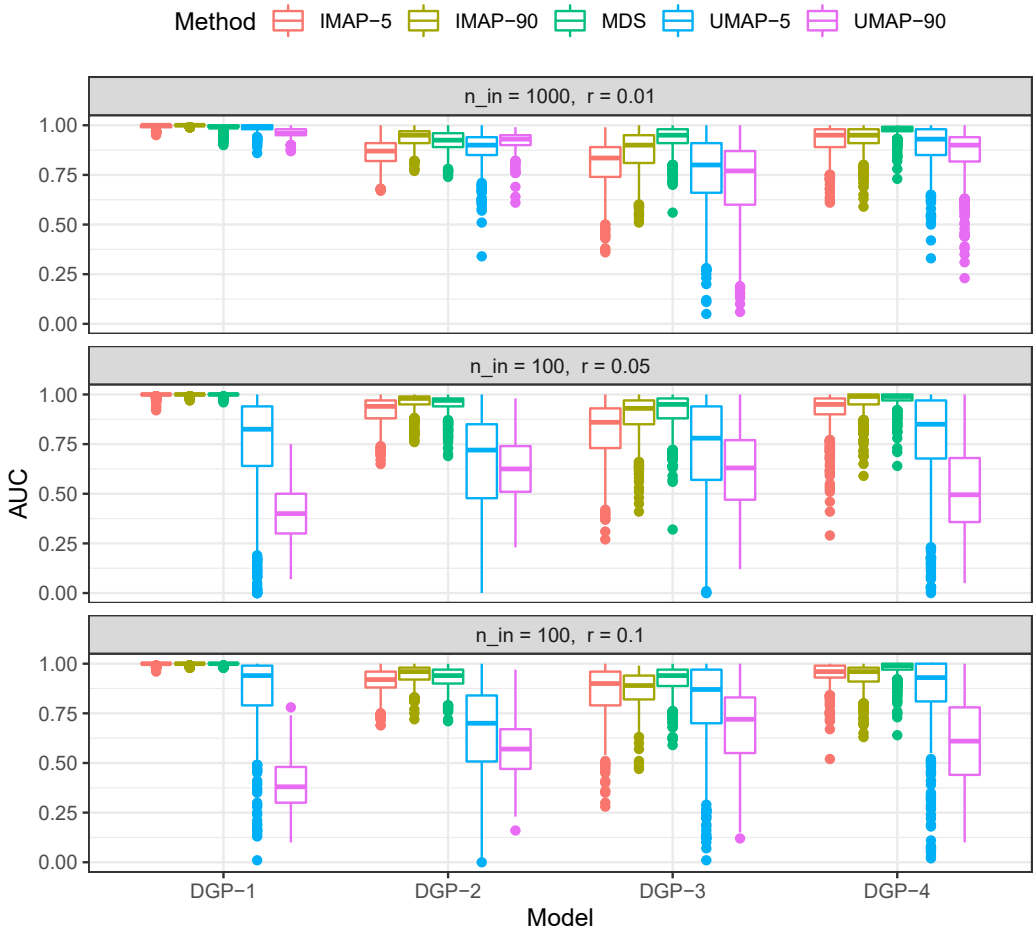


Figure 11. Comparing UMAP and ISOMAP to MDS. UMAP and ISOMAP embeddings were computed for two different locality parameter values: 5 and 90. The distribution of the AUC over 500 replications of the four DGPs for different outlier ratios r. The AUC computed on LOF scores based on 5D embeddings.

4. Discussion

Based on a geometrical perspective of functional outlier detection, we defined two general types of functional outliers: off- and on-manifold outliers. Our investigation showed that this perspective clarifies the theoretical concepts and improves practical results. From a theoretical perspective, it allows formalizing functional outlier scenarios in precise and consistent terms, beyond differences in terms of either shape, level, or magnitude. This simplifies reasoning about specific outlier settings and provides a fully general theoretical conceptualization of the problem.

From an applied perspective, we formulated two important consequences. First of all, as was demonstrated with a comprehensive analysis of a complex, real dataset of

ECG curves, the geometrical approach allows for easily accessible and highly informative visualizations. These are obtained by means of low-dimensional embeddings reflecting the inherent structure of a functional dataset in much detail. Such visualizations provide more accurate and complete pictures of the (outlier) structure of functional data. In particular, off-manifold outliers reliably appear as clearly separated (groups of) points in the low-dimensional embeddings.

Second, the proposed approach makes it possible to apply highly developed and performant standard outlier-detection methods to functional data, since the geometric structure of the data is captured and reflected in their pairwise distance matrices. Outlier detection and scoring methods that can be applied to distance matrices can therefore directly be used for functional data as well. Furthermore, detection methods requiring tabular inputs can also be applied simply by using the embedding coordinates obtained with embedding methods as proxy data for the original functions. Our experiments using LOF scores showed that the two approaches yielded very similar results. This simultaneously simplifies and improves functional outlier detection: It simplifies since functional data analysis becomes more accessible to a broader audience with general outlier-detection methods that are widely used in other areas and that do not require an understanding of complex methodological details of functional data methods. It improves the state-of-the-art since many functional outlier methods can only detect specific kinds of functional outliers by design or fail in more complex realistic data that are widely dispersed or that contain multiple nonoutlying subgroups, such as the ECG data. Moreover, note that our proposal is not limited to univariate functional data. Extending it to multivariate functions is completely straightforward, as long as a suitable dissimilarity measure is available to compute pairwise distances.

In this paper, most embeddings were obtained using MDS based on L_2 distances. This implies a close similarity to functional bagplots and highest-density region (HDR) boxplots [16], which are based on the first two robust principal component scores. However, this similarity only applies if our geometrical approach is implemented with 2D MDS embeddings based on L_2 distances. As outlined, our proposal is neither limited to the L_2 metric as a distance measure nor to MDS as an embedding method or just two embedding dimensions. Other metrics and (higher-dimensional) embedding methods can be used as well, and our results indicate that an alternative distance measure can further improve the performance in specific settings, sometimes considerably. In particular, even nonmetric dissimilarity measures may be applicable as our results based on DTW distances indicate. On the other hand, the results also show that more sophisticated embedding methods such as ISOMAP and UMAP cannot be used as straightforwardly as MDS. Such methods, which do not take into account the ambient space geometry by default, at least require very careful parameter selection.

In terms of practical applicability, the $O(n^3)$ time complexity and $O(n^2)$ storage complexity of standard MDS may prove problematic for large data, but generalizations such as Landmark MDS [53], Pivot MDS [54], or multilevel MDS exploiting GPU performance [55] scale much better with the number of available observations.

Finally, we would argue that existing functional outlier detection approaches mostly lack the principled geometrical underpinning and conceptualization presented here. As outlined, we argue that such a conceptualization is necessary to make functional outlier detection tractable in full generality. Specifically, consider that existing methods typically limit themselves to creating a 1D or 2D representation of each curve (e.g., MBD-MEI, MO-VO, functional bagplots, HDR plots), often based on preconceived notions of the characteristics of functional outliers. Our investigations and experiments suggested that this is often not sufficient for real-world functional outlier detection: there is no valid reason to limit our representations to two dimensions with modern outlier-detection methods, and the geometrical perspective often strongly suggests otherwise in the case of complex functional data. Even more importantly, it is much more flexible to learn maximally informative low-dimensional representations directly from data instead of starting with

rigid notions of which characteristics to look at and to ignore the rest. The latter is likely to lead to results not capturing the entire (outlier) structure of a given dataset, which is essential in real-world unsupervised settings and exploratory analyses.

Based on the theoretical considerations and the empirical results outlined above, we conclude that the proposed approach is well suited for both the theoretical conceptualization and the practical implementation of functional outlier detection. In particular, the choice of embedding method should consider whether it is able to preserve the extrinsic geometry of the function space, and simple MDS embeddings based on functional distances provide a very strong baseline for that. On the basis of this work, we intend to further investigate the implications of the geometrical perspective, such as the effects other dissimilarity measures, embedding, and outlier-detection methods, in future research. We are also investigating the use of mass volume curves [56] for hyperparameter tuning in functional outlier detection. Such a criterion will permit analysts to optimize the combination of the functional distance metric, embedding dimensionality, and outlier-scoring method parameters. In the absence of quantitative criteria for optimizing these settings, our recommendations are to (1) use the standard L_2 metric as the default, which proved to be a very strong baseline in our experiments for a wide variety of data settings and outlier types, (2) make use of substantive knowledge about the data at hand, either from an initial exploratory data analysis or expertise about the data-generating process, in order to choose metrics that are sensitive to the relevant kinds of structural deviations, and (3) supplement and verify the results with results based on alternative metrics, since our proposal has a low computational cost for typical functional dataset sizes.

Author Contributions: Conceptualization, M.H.; methodology, M.H. and F.S.; software, M.H.; validation, M.H. and F.S.; formal analysis, M.H. and F.S.; investigation, M.H. and F.S.; resources, M.H. and F.S.; data curation, M.H.; writing—original draft preparation, M.H.; writing—review and editing, M.H. and F.S.; visualization, M.H. and F.S.; supervision, F.S.; project administration, F.S.; funding acquisition, F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The authors of this work take full responsibility for its content.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All R code and data to fully reproduce the results are freely available on GitHub: https://github.com/HerrMo/fda-geo-out.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LOF	Local outlier factor
FD(A)	Functional data (analysis)
(F)PCA	(Functional) principle component analysis
HDR	High-density region
NHST	Null hypothesis significance testing
ECG	Electrocardiogram
MDS	Multidimensional scaling
DTW	Dynamic time warping
MS-plot	Magnitude–shape plot
GOF	Goodness of fit
DO	Directional outlyingness
TV	Total variational depth
ED	Elastic depth
DGP	Data-generating process
ECDF	Empirical cumulative distribution function
AUC	Area under the ROC curve

MBD	Modified band depth
MEI	Modified epigraph index
MO	Mean directional outlyingness
VO	Variability of directional outlyingness

Appendix A. Formalizing Phase Variation Scenarios

Appendix A.1. Phase Variation: Case I

The manifold $\mathcal{M} = \{x(t) : x(t) = \theta_1\varphi(t-\theta_2), \theta = (\theta_1, \theta_2)' \in \Theta\}$, with $\varphi(.)$ the standard Gaussian pdf and $\Theta = [0.1, 2] \times [-2, 2]$, defines a functional data setting with independent amplitude and phase variation. Since there is a single manifold only, there are no structural novelties. Figure A1, top, depicts the functional observations on the left and a 2D embedding obtained with MDS on the right. Note that all of the curves are subject to amplitude and phase variation to a varying extent; however, there are no clearly "outlying" or "outstanding" observations in terms of either amplitude or phase. This is reflected in the corresponding embedding, which does not show any clearly separated observations in the embedding space, indicating that there are no structurally different observations. The situation in the second case of phase-varying data, however, is different.

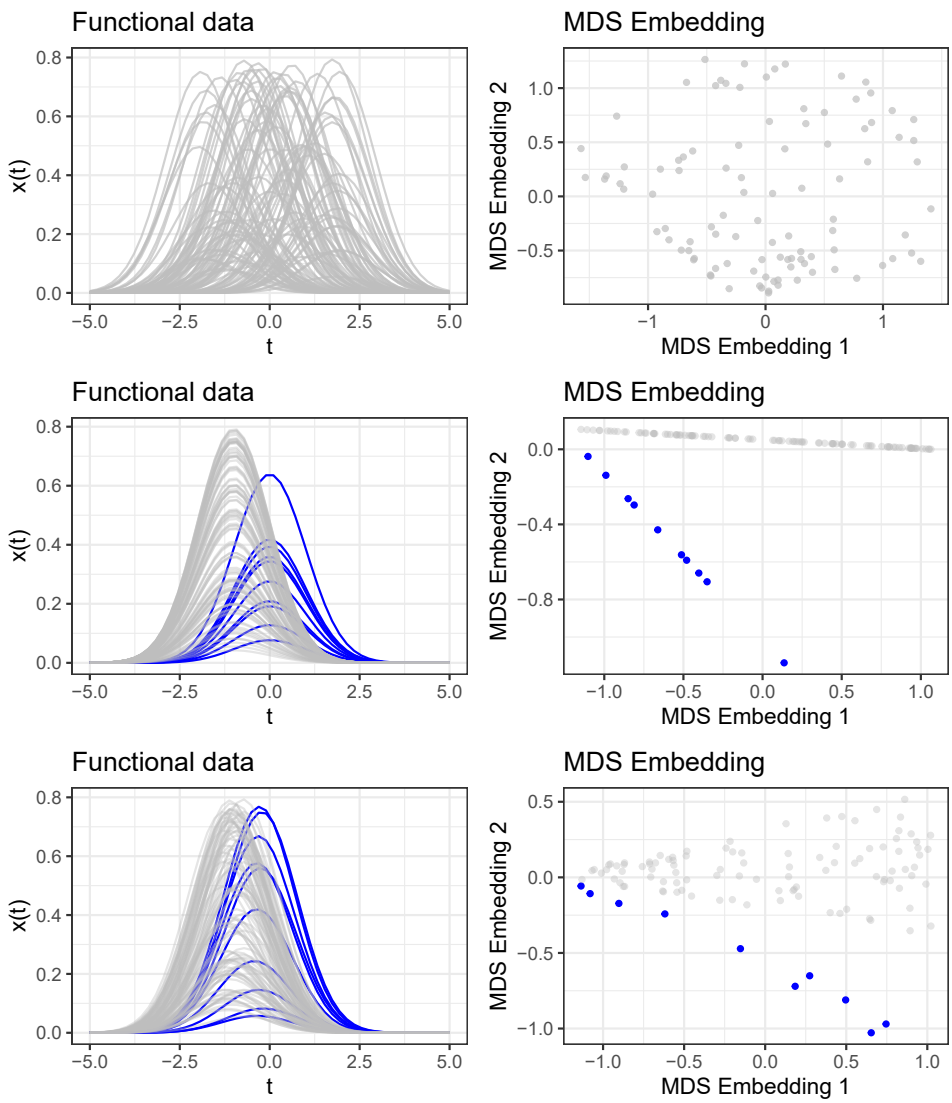


Figure A1. Functional data with phase variation and different levels of structural difference. **Top**: scenario with no off-manifold outliers. **Middle**: scenario with clear off-manifold outliers. **Bottom**: intermediate scenario.

Appendix A.2. Phase Variation: Case II

The two manifolds $\mathcal{M}_c = \{x(t) : x(t) = \theta\varphi(t+1), \theta \in \Theta\}$ and $\mathcal{M}_a = \{x(t) : x(t) = \theta\varphi(t), \theta \in \Theta\}$, with $\Theta = [0.1, 2]$ describe a similar scenario as before; however, there are two structurally different manifolds induced by the shift in the argument of φ. In contrast to the first case, there are on-manifold and off-manifold outliers. Figure A1, middle, depicts the functional observations and the corresponding embedding. Clearly, in this example, a few (blue) curves, the ones from $\mathcal{M}_a$, show a horizontal shift compared to the normal data, and consequently, those few curves appear horizontally "outlying". Within the main data manifold, only on-manifold outliers in terms of amplitude exist. These aspects are reflected in the corresponding embedding: the low-dimensional representations of the blue curves are clearly separated from those of the main data in grey.

Of course, such clear settings—in particular, phase-varying functional data with fixed and distinct phase parameters—will seldom be observed in practice. A more realistic example is given by $\mathcal{M}_c = \{x(t) : x(t) = \theta_1\varphi(t-\theta_2), (\theta_1, \theta_2)' \in \Theta_c\}$ and $\mathcal{M}_a = \{x(t) : x(t) = \theta_1\varphi(t-\theta_2), (\theta_1, \theta_2)' \in \Theta_a\}$, with $\Theta_c = [0.1, 2] \times [-1.3, -0.7]$ and $\Theta_a = [0.1, 2] \times [-0.5, 0.1]$. Here, we have again two structurally different manifolds. This is more realistic, since the "phase parameters" θ_2 are not fixed, but are subject to random fluctuations. In addition, the structural difference induced by the phase parameters is much smaller. Considering Figure A1, bottom, again, this is reflected in the embedding: there are two separable structures; however, the differences are not as clear as in the second example above.

The three examples together show that the less similar the processes are and/or the less variability there is within the phase parameters defining the manifolds, the clearer structural differences induced by horizontal variation become visible in the embeddings.

Appendix B. Sensitivity Analysis

The differences in complexity among the ECG and the other four real datasets become apparent in Figure A2 as well, which shows how the goodness of fit (GOF) of the embeddings is affected by their dimensionality. For the L_2 metric, a goodness of fit over 0.9 is achieved with two to three embedding dimensions for the less complex datasets. Moreover, all of them reach a saturation point at five dimensions. This is in contrast to the ECG data, where the first five embedding dimensions lead to a goodness of fit of 0.8. Moreover, the ranking induced by LOF scores is very robust to the number of embedding dimensions. As Table A1 shows, the rank correlations between LOF scores based on five and LOF scores based on twenty embedding dimensions are very high for all datasets.

Table A1. Spearman correlation between LOF scores based on embeddings of different dimensionality for the 5 considered real datasets and metrics $L_{0.5}$, L_1, ..., L_{10}, and unnormalized L_1-Wasserstein. MDS embeddings with 5 dimensions are compared to embeddings with 2 (2 vs. 5) and 20 (5 vs. 20) dimensions.

	$L_{0.5}$		L_1		L_2		L_3		L_4		L_5	
	2 vs. 5	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**
ECG	0.96	0.97	0.98	0.97	0.97	0.99	0.94	0.99	0.94	0.98	0.90	0.97
Octane	0.94	0.99	0.96	0.98	0.97	0.99	0.98	0.99	0.98	0.99	0.96	0.98
Weather	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Tecator	0.97	0.99	0.96	0.99	0.99	1.00	0.99	1.00	0.99	1.00	1.00	1.00
Wine	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00

	L_6		L_7		L_8		L_9		L_{10}		**Wasserstein**	
	2 vs. 5	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**	**2 vs. 5**	**5 vs. 20**
ECG	0.89	0.96	0.87	0.96	0.86	0.95	0.86	0.95	0.85	0.94	0.98	0.97
Octane	0.96	0.98	0.95	0.99	0.96	0.98	0.94	0.97	0.94	0.97	0.95	0.96
Weather	1.00	1.00	0.99	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00
Tecator	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	0.99	1.00	0.96	0.99
Wine	0.99	1.00	0.98	1.00	0.98	1.00	0.98	0.99	0.98	0.99	0.99	1.00

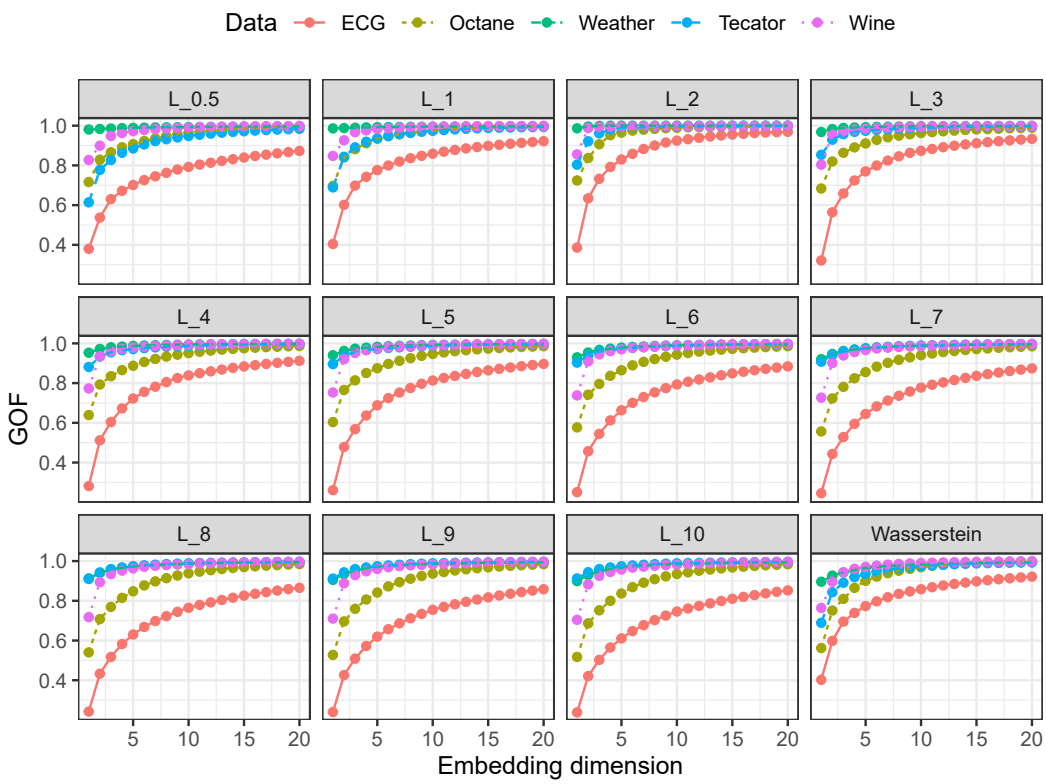


Figure A2. Goodness of fit (GOF) of different embedding dimensions for the five considered real datasets and $L_{0.5}, L_1, ..., L_{10}$, and unnormalized L_1-Wasserstein metrics.

Appendix C. Quantitative Results on the `fdaoutlier` Package DGPs

The simulation models presented by Ojo et al. [25] cover different outlier scenarios: vertical shifts (Model 1), isolated outliers (Model 2), partial magnitude outliers (Model 3), phase outliers (Model 4), various kinds of shape outliers (Models 5–8), and amplitude outliers (Model 9). A detailed description can be found in the vignette (https://cran.r-project.org/web/packages/fdaoutlier/vignettes/simulation_models.html, accessed on 15 November 2021) accompanying their R package. In the following, the proposed geometrical approach is compared to directional outlyingness (DO) and total variational depth (TV) using the AUC as a performance measure.

As Figure A3 shows, (almost) perfect performance is achieved by at least two methods for Models 1, 3, 4, 8, and 9; DO shows almost perfect performance for all models except Model 1. For Models 2, 5, 6, and 7, the methods based on the geometric approaches do not perform equally well (as does TV). However, as outlined in Section 3.3, perfect performance can be achieved for Model 2 by using L_{10} distances instead of L_2 distances.

Furthermore, for Models 5, 6, and 7, it has to be taken into account that the AUC values only reflect the detection of "true outliers", which can now—given the geometric perspective—be specified more precisely as off-manifold outliers (observations from $\mathcal{M}_a$). However, this does not take into account possible on-manifold outliers. Due to their distributional nature, by chance, some on-manifold outliers (observations on $\mathcal{M}_a$) can be "more outlying" than some of the off-manifold outliers and thus correctly obtain higher LOF scores. However, such cases are not correctly reflected in the performance assessment approach, as—in contrast to off-manifold outliers—such on-manifold outliers are not labeled as "true outliers". The observed lower performance in terms of the AUC thus can simply mean that there are on-manifold outliers obtaining relatively high LOF scores. In particular, this also does not imply that off-manifold outliers fail to be separated in a subspace of the embedding, as will be outlined in Appendix E in more detail, nor that perfect AUC performance cannot be obtained via the geometric approaches for these settings. If the geometric approach is applied to the derivatives instead (depicted in

Figure A3 as "deriv"), almost perfect performances can be achieved. Obviously, functions of the same shape (i.e., all observations from $\mathcal{M}_c$) are very similar on the level of derivatives regardless of how strongly dispersed they are in terms of vertical shift.

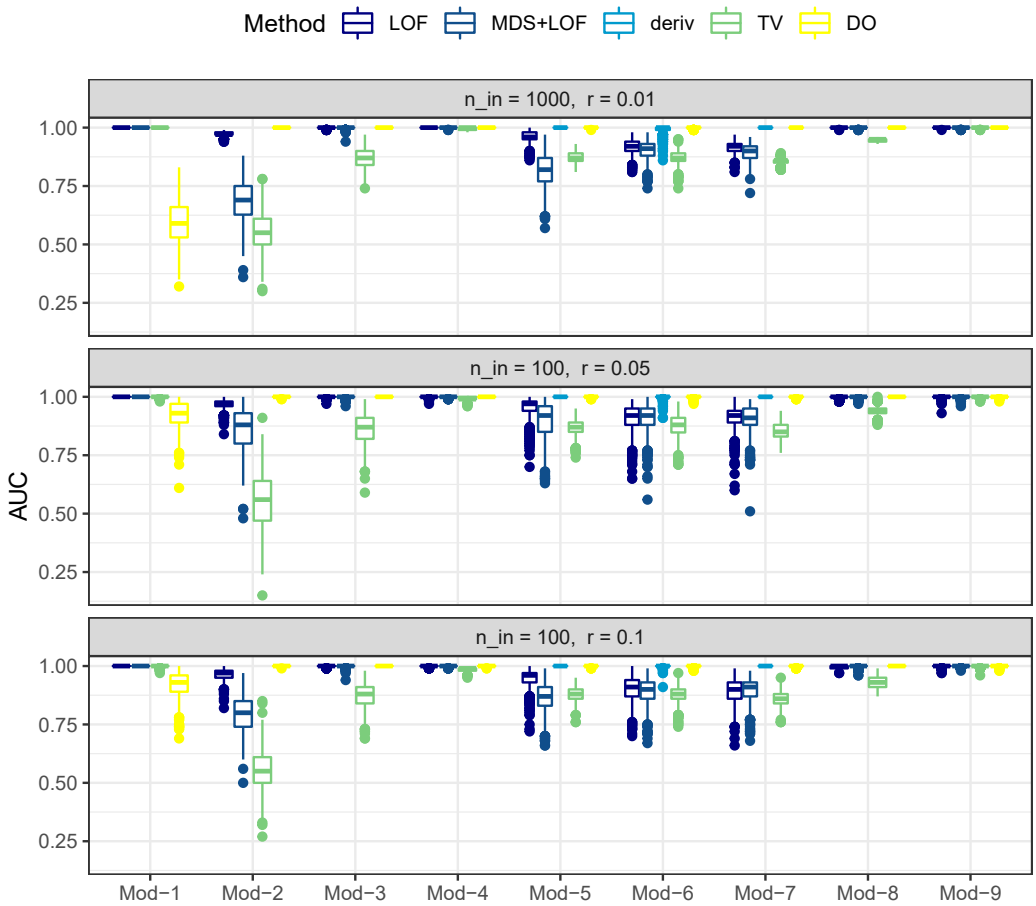


Figure A3. Distribution of the AUC over the 500 replications for the different outlier-detection methods, simulation models (Mod) from the package fdaoutier, and outlier ratios r.

Appendix D. Visualization Methods: `roahd::outliergram`, `fdaoutlier::msplot`, Translation–Phase–Amplitude Boxplots, Elastic Depth Boxplots, and HDR Boxplots

Figure A4 shows the results for the synthetic data example of Figure 4 with ten true outliers, where the MS plot yields six false positives and only three true positives, while the Outliergram fails to detect even a single outlier. The elastic depth boxplots labels twenty-six observations as outliers, only two of which are among the shifted observations. Moreover, note that observations labeled phase outliers are also labeled amplitude outliers at the same time. In contrast, the translation–phase–amplitude boxplots correctly detect the 10 shifted observations as translation outliers; however, 15 other observations are also labeled outliers. Note that some observations obtain multiple labels, for example, all phase outliers are also labeled as amplitude outliers. The HDR boxplots yield six false positives and no true positive (see Figure A13). In summary, neither of the methods are capable of correctly capturing the outlier structure of this dataset, in contrast to the proposed geometrical approach.

Figure A5 shows results for the MBD-MEI "Outliergram" by Aribas-Gil and Romo [41] (implementation: [42]) for shape outlier detection, and the magnitude–shape plot method of Dai and Genton [34] for the example datasets shown in Figures 5 and 8. Figures A6 and A7 show the results for the translation–phase–amplitude boxplots by Xie et al. [15] and the elastic depth boxplot for shape outlier detection by Harris et al. [9] for these datasets. Finally, Figures A8–A13 show the results of the HDR boxplots by Hyndman and Shang [16] (implementation: [43]). For a detailed discussion, see Section 3.1.

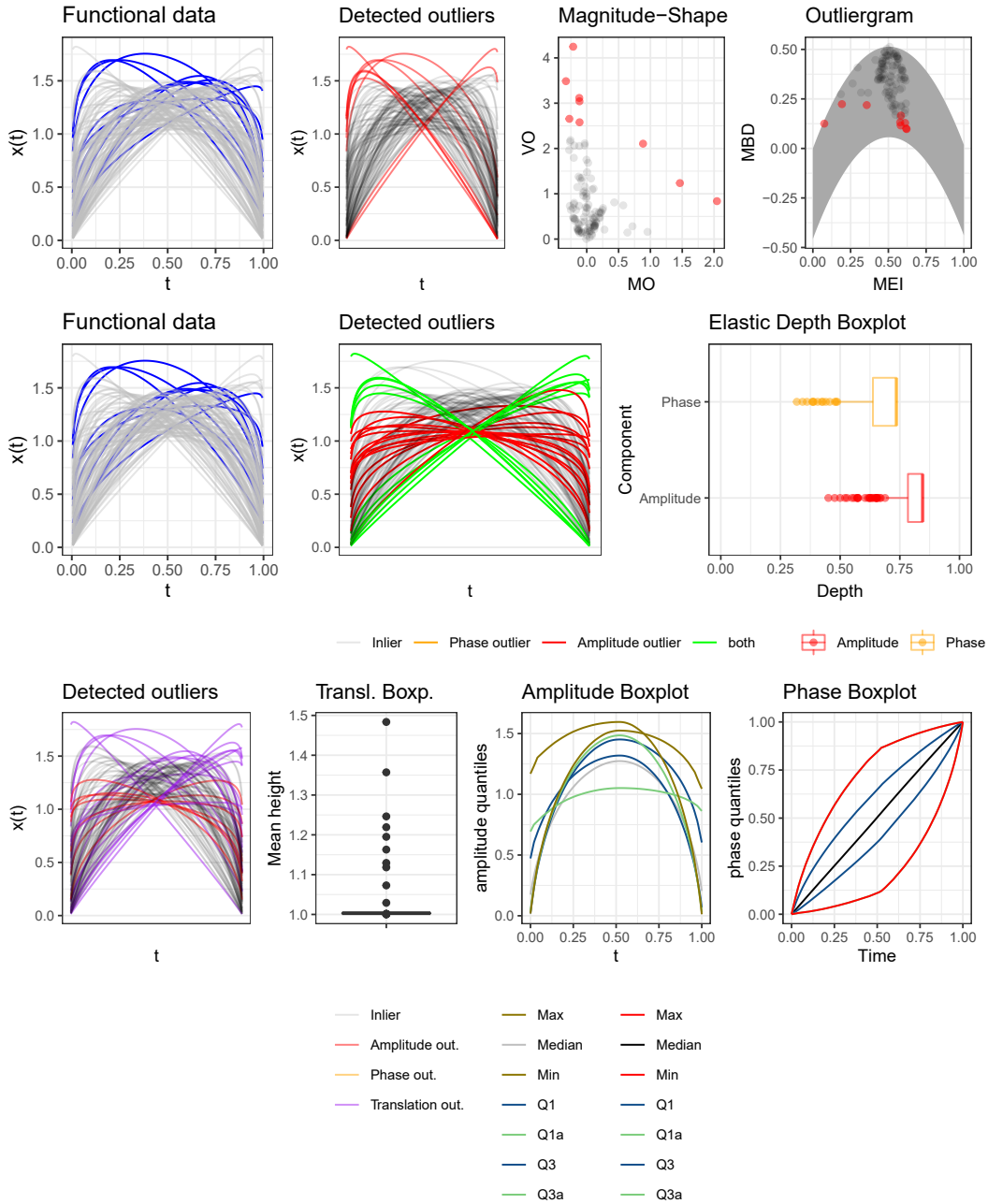


Figure A4. First column of first two rows: data with true outliers in blue; subsequent columns: data with detected outliers in color. First row: magnitude–shape plot of mean directional outlyingness (MO) versus variability of directional outlyingness (VO) and outliergram of the modified epigraph index (MEI) versus modified band depth (MBD) with the inlier region in grey. Second row: Elastic depth boxplots. Third row: translation–phase–amplitude boxplots. For the results of the HDR boxplots on the data, see Figure A8.

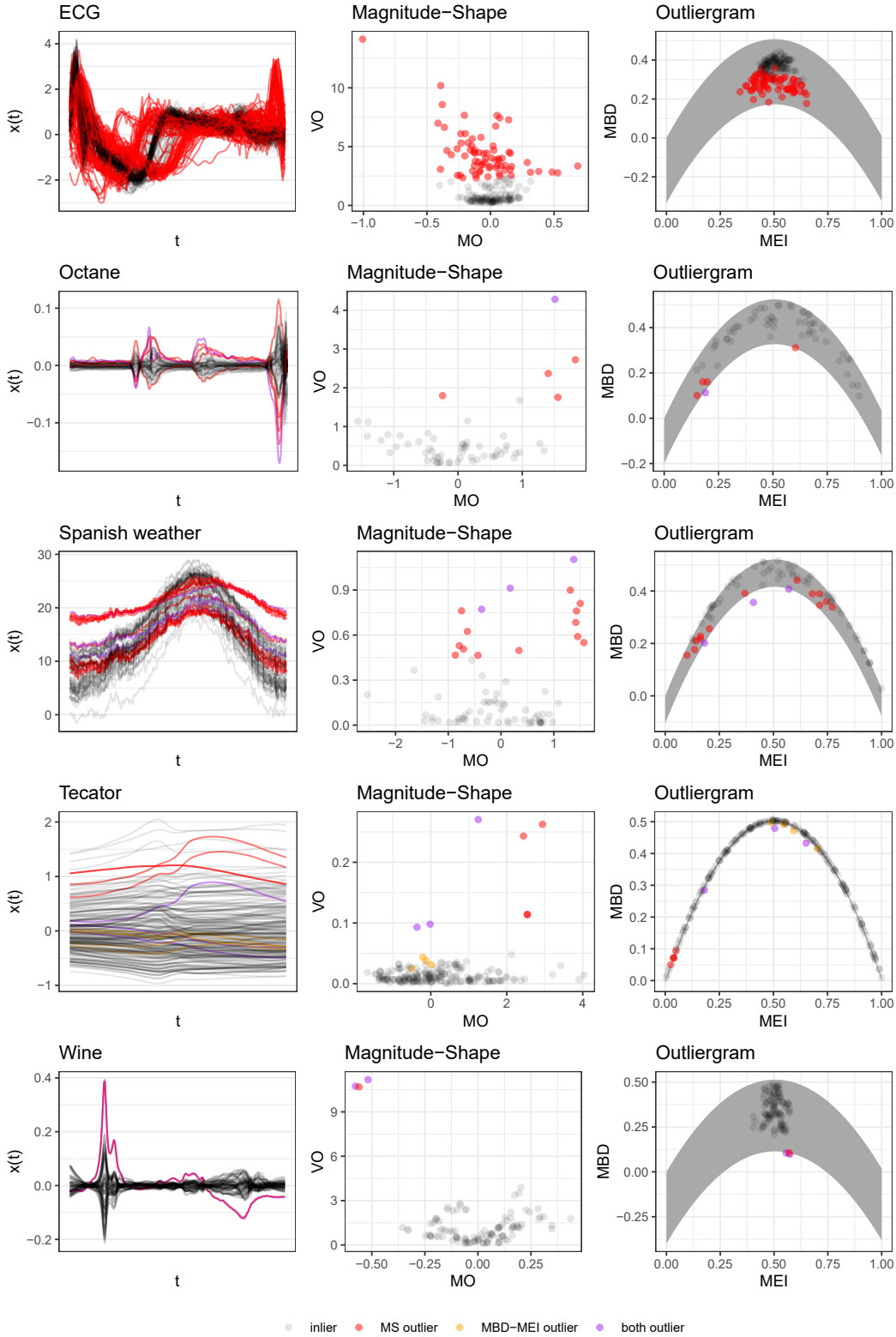


Figure A5. Left column: data; middle column: magnitude–shape plots of mean directional outlyingness (MO) versus variability of directional outlyingness (VO); right column: outliergram of the modified epigraph index (MEI) versus modified band depth (MBD) with the inlier region in grey. Curves and points are colored according to outlier status as diagnosed by `fdaoutlier::msplot` and/or `roahd::outliergram`.

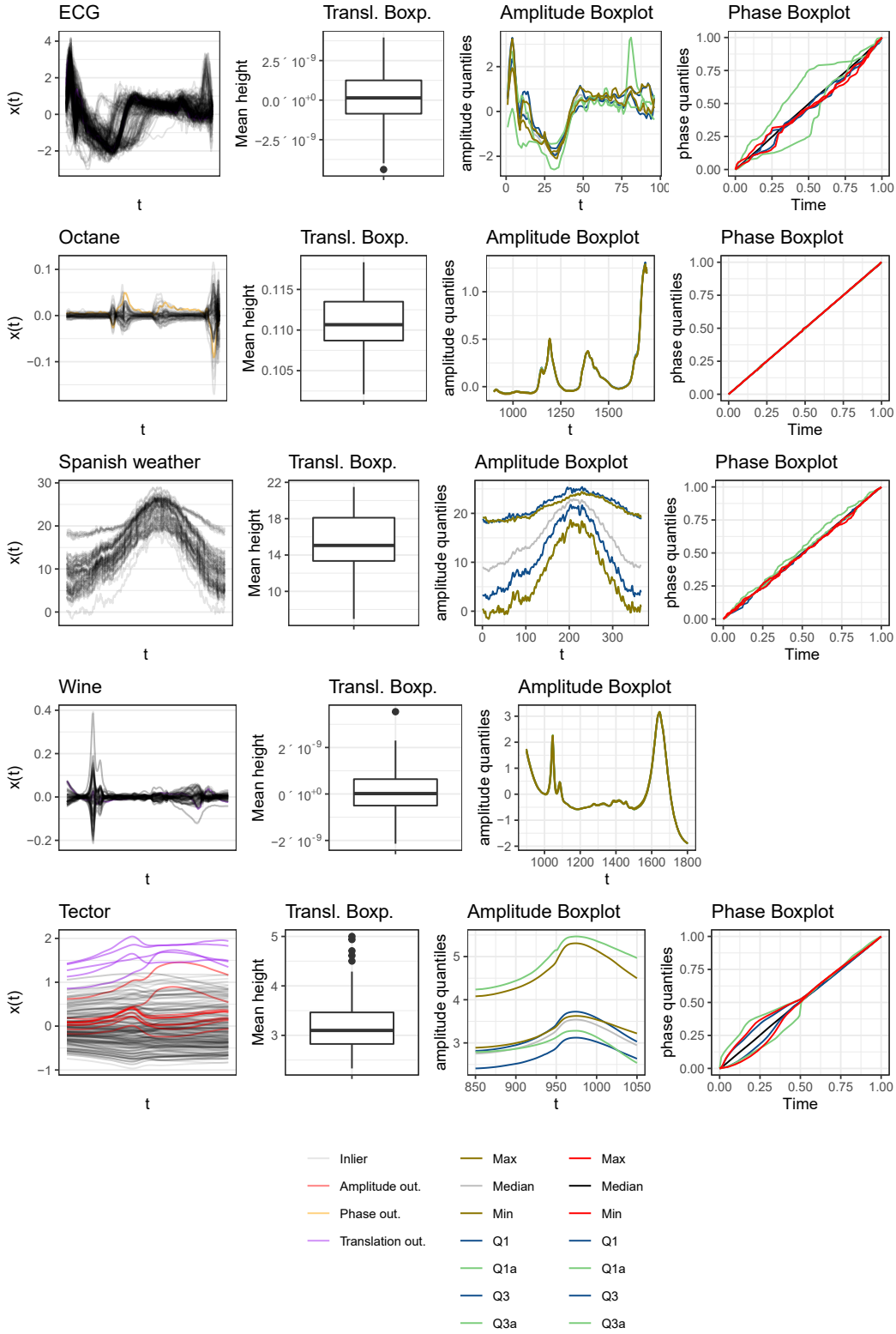


Figure A6. First column: data; second column: translation boxplots of average curve heights; third and fourth column: amplitude, respectively phase boxplots with the maximum and minimum extreme curves (Max, Min), the first and third quartile curves (Q1 and Q3), and the 0.05- and 0.95-quantile curves (Q1a, Q3a). Curves in the first column colored according to the outlier status by translational outlyingness, amplitude outlyingness, and phase outlyingness (the latter two as diagnosed by `fdasrvf::AmplitudeBoxplot` and `fdasrvf::AmplitudeBoxplot`). Note, for the Wine data, it was not possible to compute the phase boxplot.

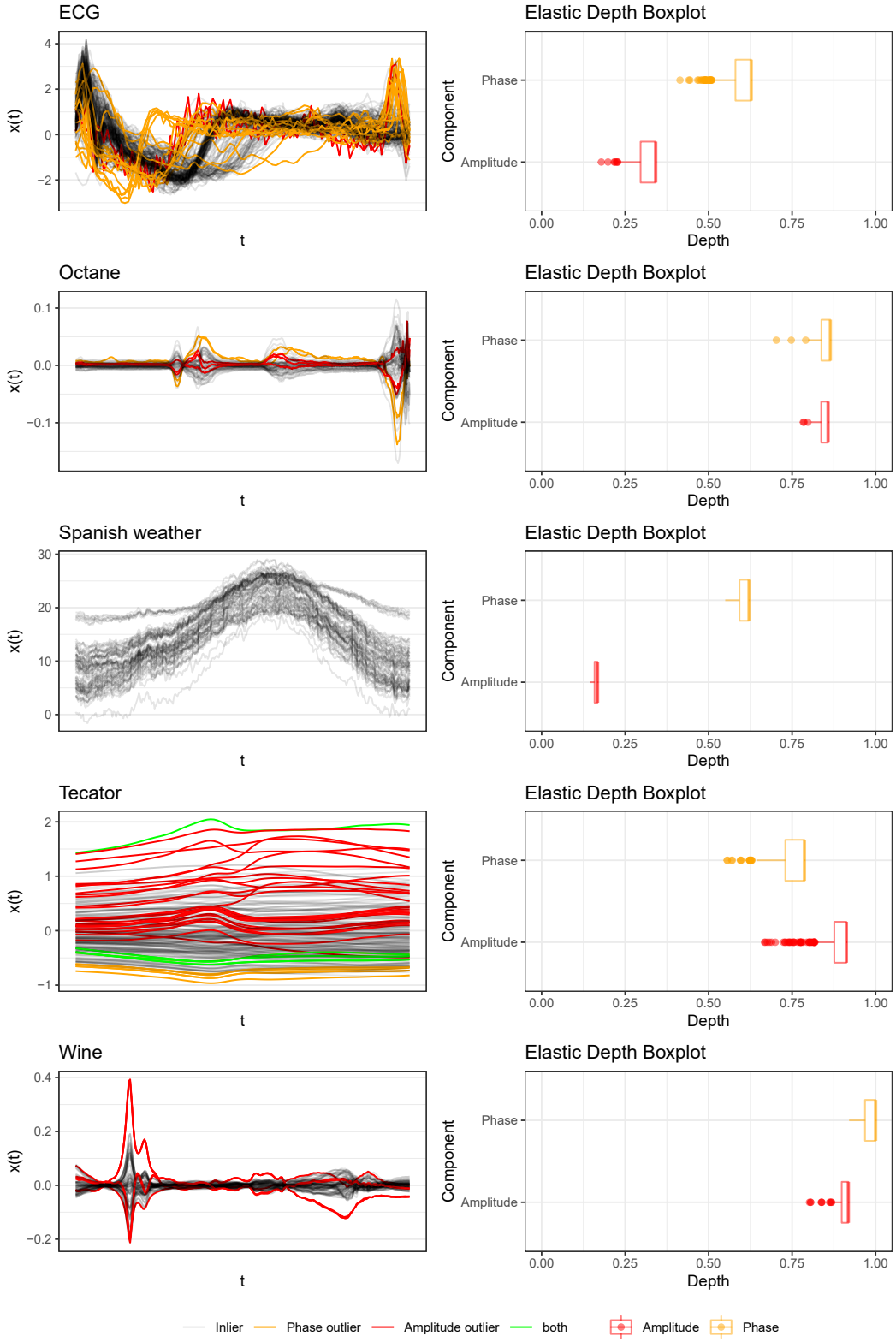
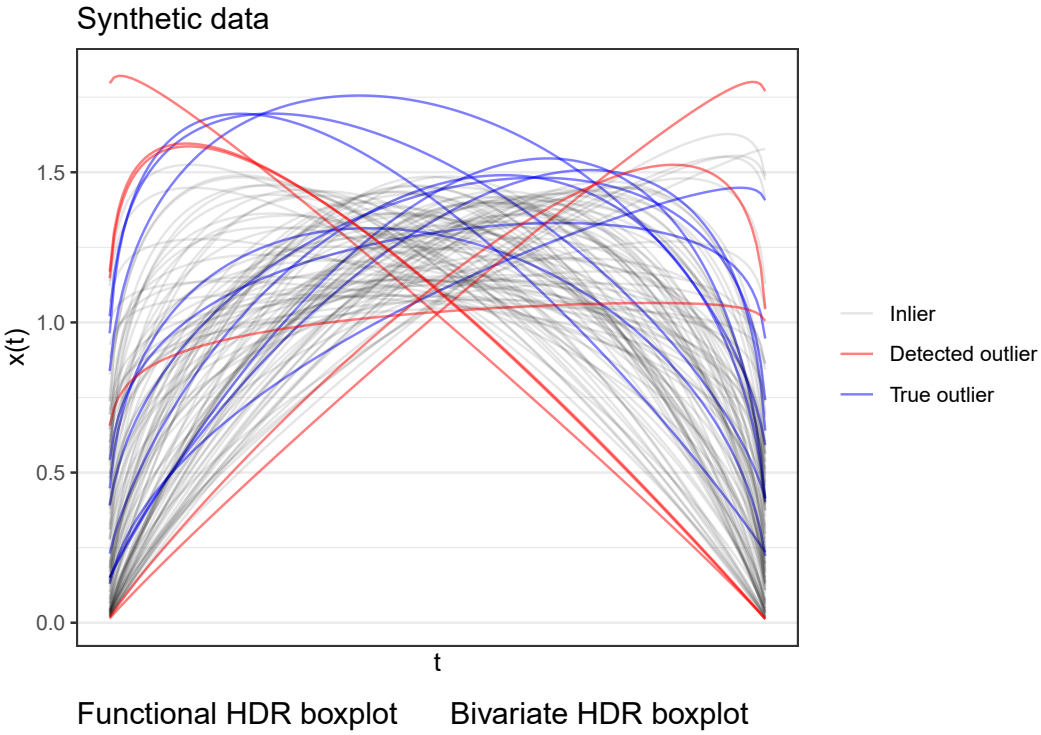


Figure A7. Left column: data; right column: elastic depth boxplots for amplitude and phase variability. Curves in the left column colored according to the outlier status by amplitude outlyingness and phase outlyingness as diagnosed by `elasticdepth::elastic_outliers`.



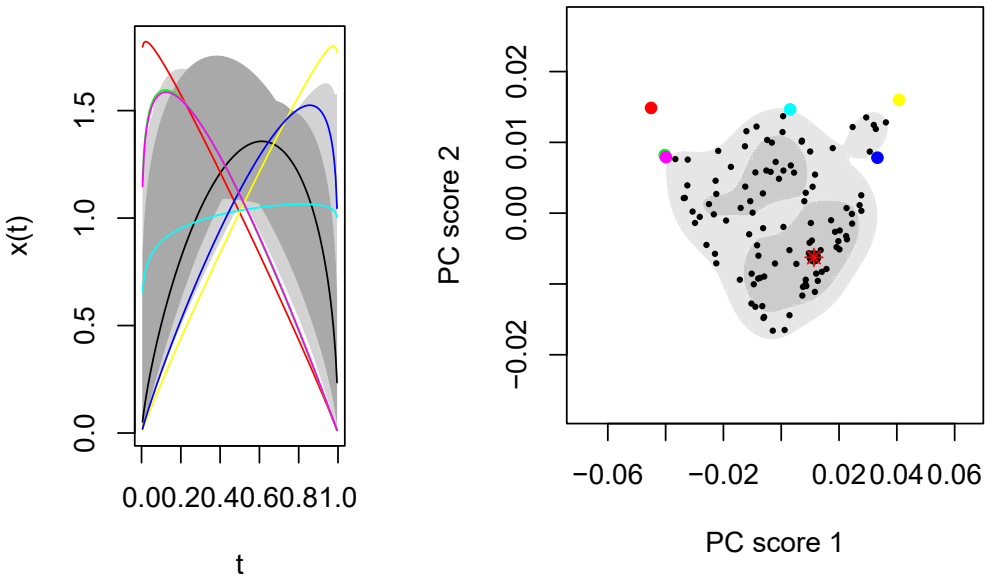


Figure A8. Upper row: synthetic data. Lower row, left column: functional HDR boxplot; right column: bivariate HDR boxplot. Colored curves/points are outliers according to a coverage probability of 0.05 for the functional HDR boxplot. HDR boxplots computed with `rainbow::fboxplot`.

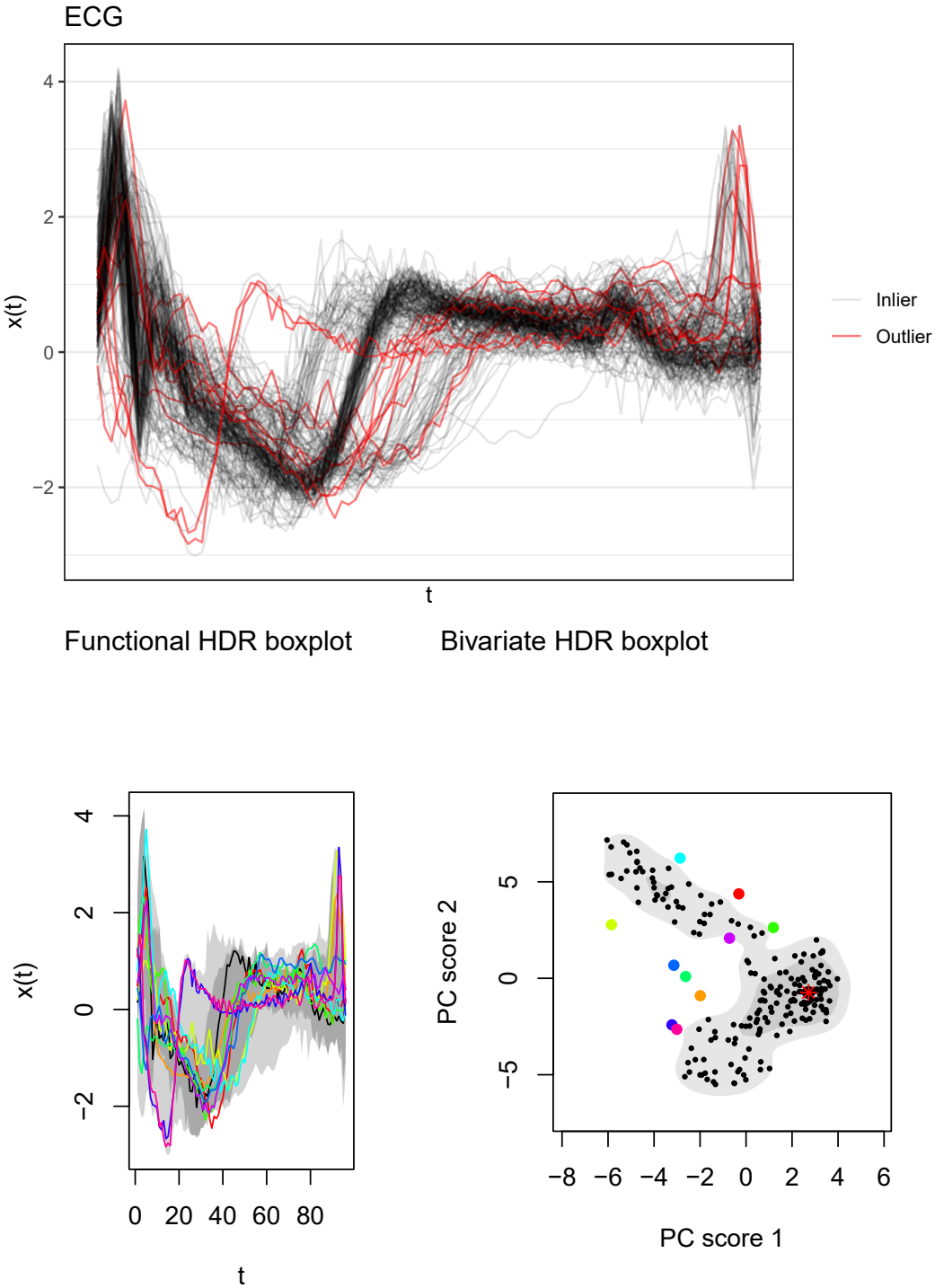


Figure A9. Upper row: ECG data. Lower row, left column: functional HDR boxplot; right column: bivariate HDR boxplot. Colored curves/points are outliers according to a coverage probability of 0.05 for the functional HDR boxplot. HDR boxplots computed with `rainbow::fboxplot`.

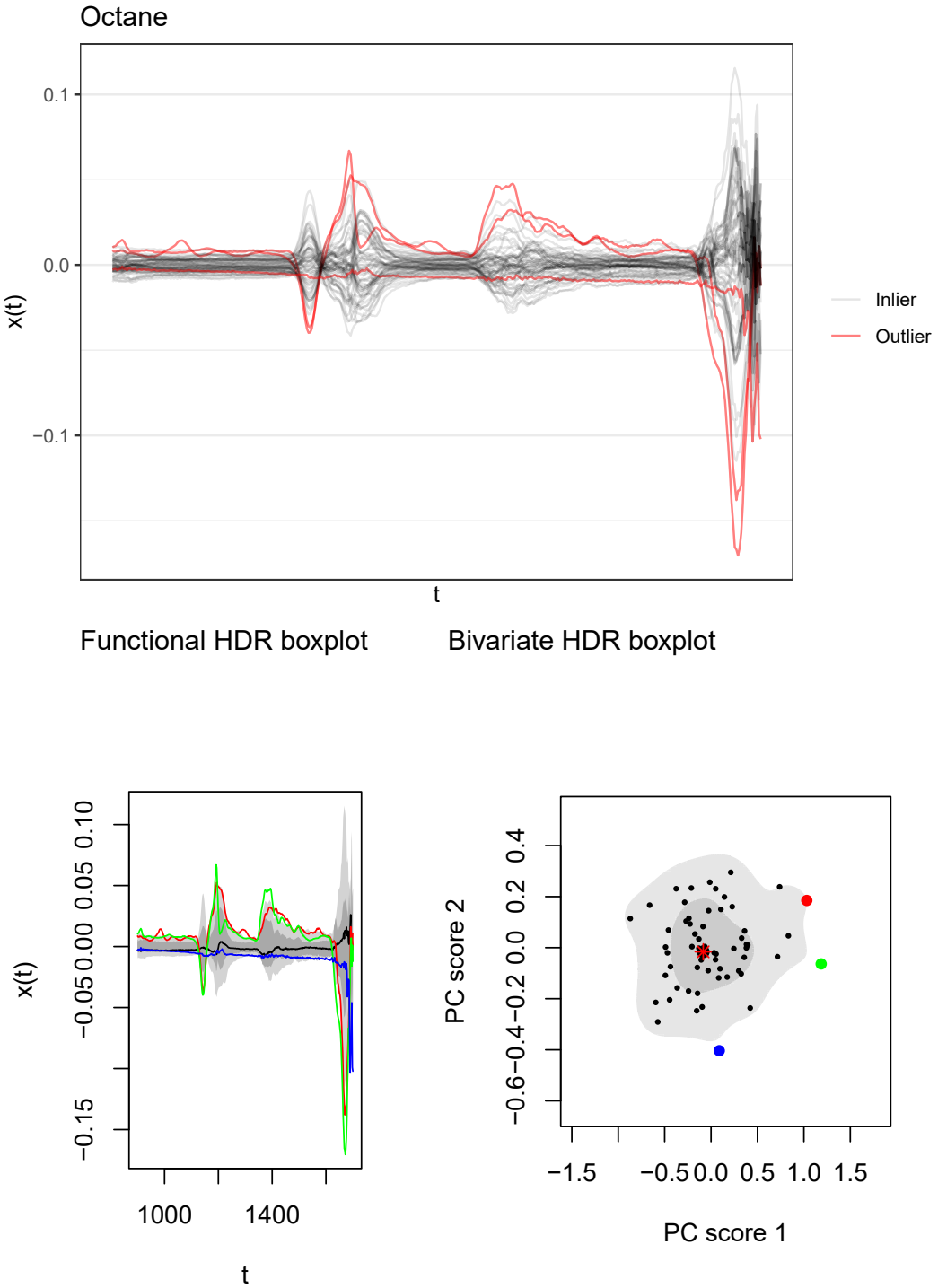


Figure A10. Upper row: Octane data. Lower row, left column: functional HDR boxplot; right column: bivariate HDR boxplot. Colored curves/points are outliers according to a coverage probability of 0.05 for the functional HDR boxplot. HDR boxplots computed with `rainbow::fboxplot`.

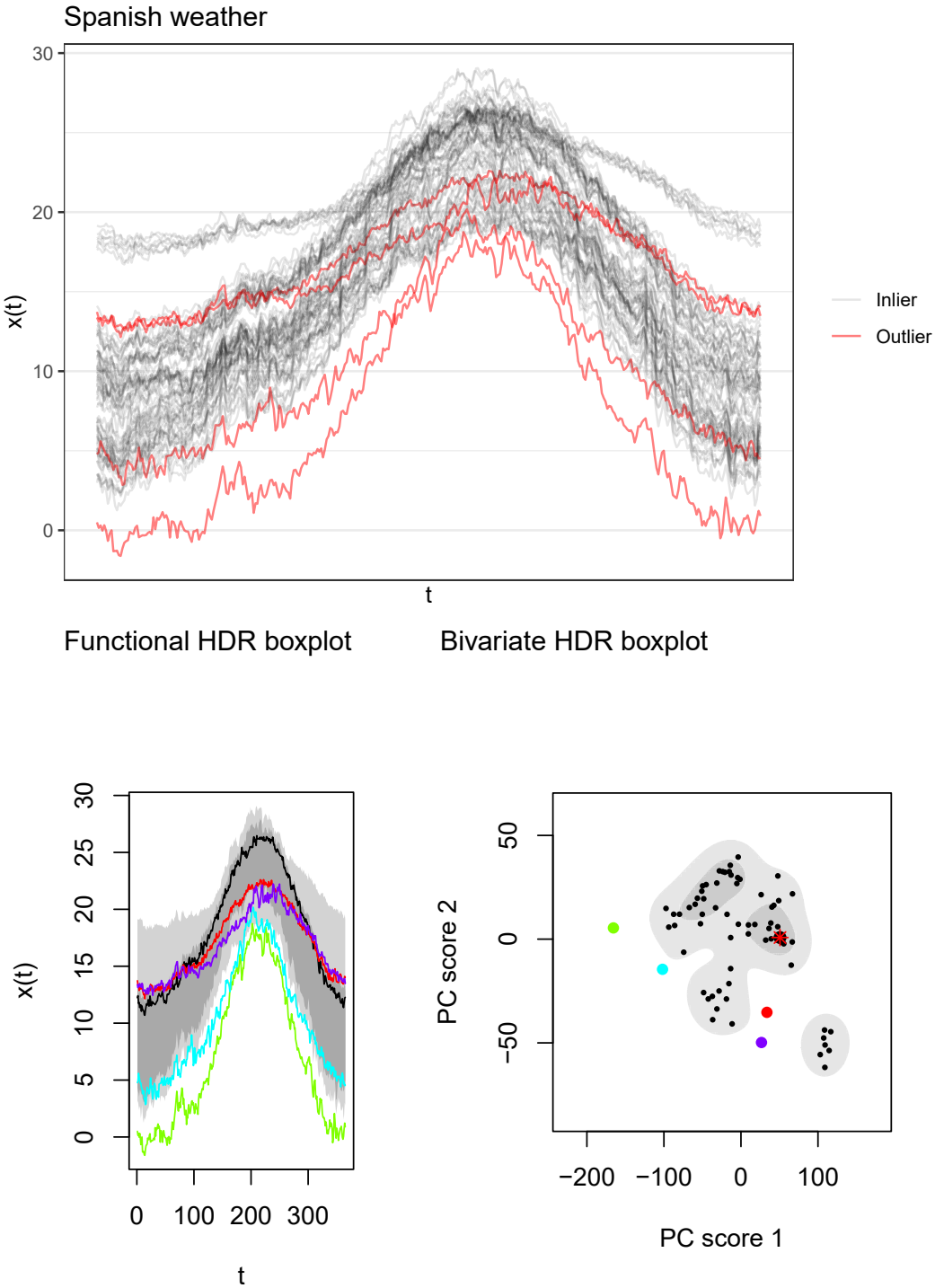


Figure A11. Upper row: Spanish weather data. Lower row, left column: functional HDR boxplot; right column: bivariate HDR boxplot. Colored curves/points are outliers according to a coverage probability of 0.05 for the functional HDR boxplot. HDR boxplots computed with `rainbow::fboxplot`.

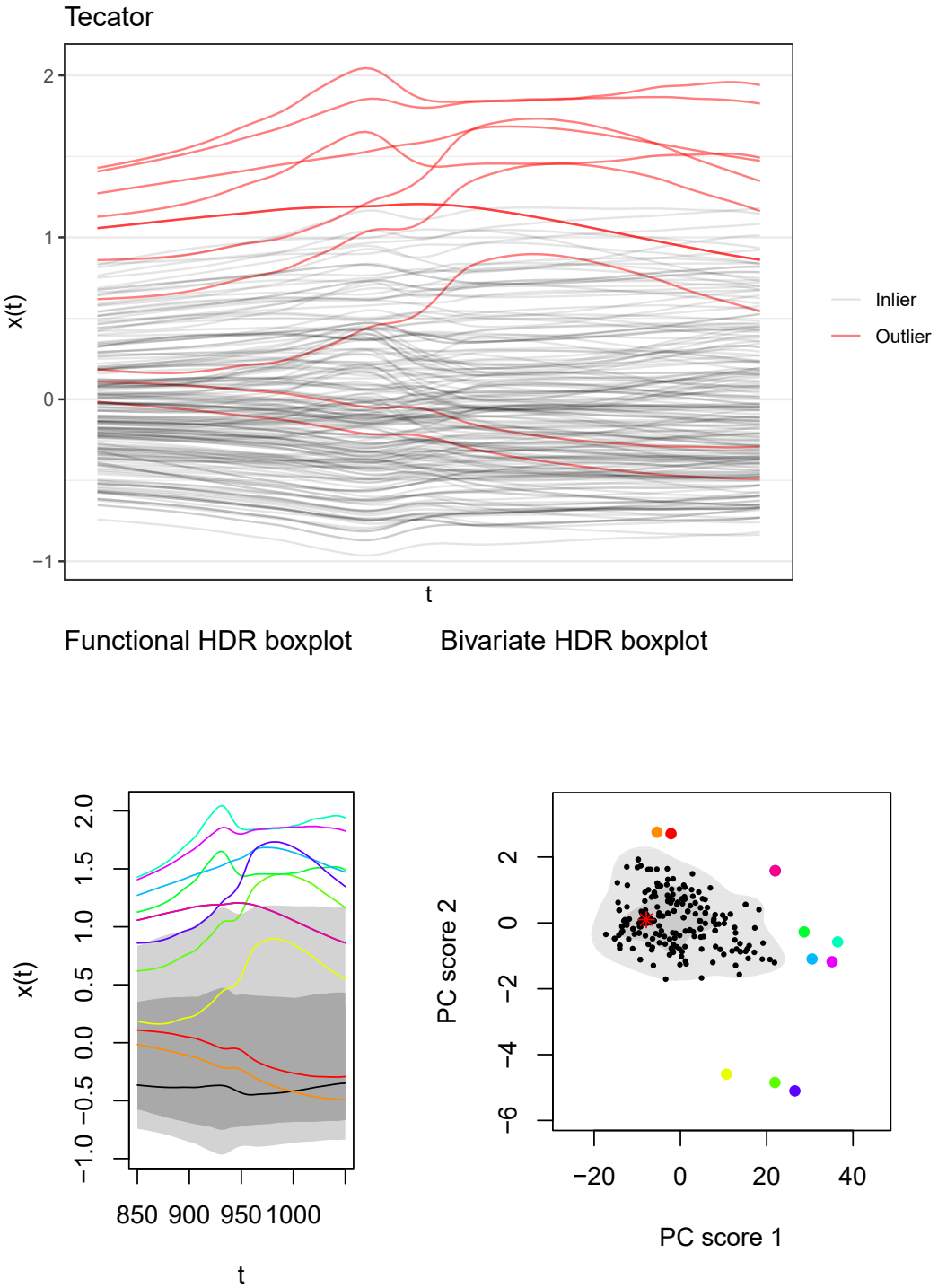


Figure A12. Upper row: Tecator data. Lower row, left column: functional HDR boxplot; right column: bivariate HDR boxplot. Colored curves/points are outliers according to a coverage probability of 0.05 for the functional HDR boxplot. HDR boxplots computed with `rainbow::fboxplot`.

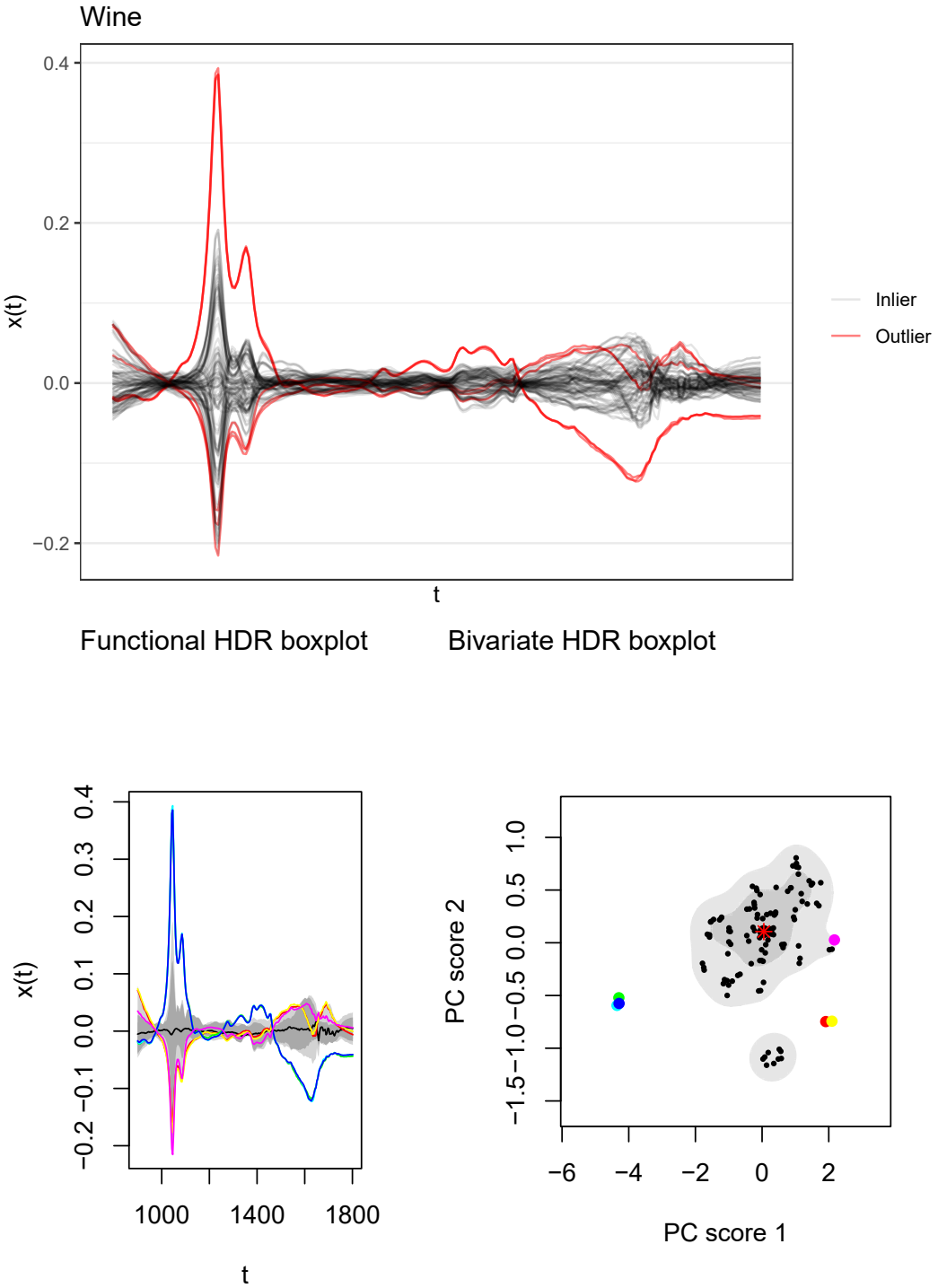


Figure A13. Upper row: Wine data. Lower row, left column: functional HDR boxplot; right column: bivariate HDR boxplot. Colored curves/points are outliers according a coverage probability of 0.05 for the functional HDR boxplot. HDR boxplots computed with `rainbow::fboxplot`.

Appendix E. In-Depth Analysis of Simulation Model 7

The analysis of the ECG data in Section 3.1 showed that embeddings can reveal much more (outlier) structure than can be represented by scores and labels. To illustrate the effects described in Appendix C, we conducted a similar qualitative analysis for an example dataset with observations sampled from Simulation Model 7; see Figure A14. The dataset consisted of 100 observations with 10 off-manifold or—in more informal terms: "true"—outliers. The functions were evaluated on 50 grid points. The analysis showed that a quantitative performance assessment alone may yield misleading results and again emphasizes the practical value of the geometric perspective and low-dimensional

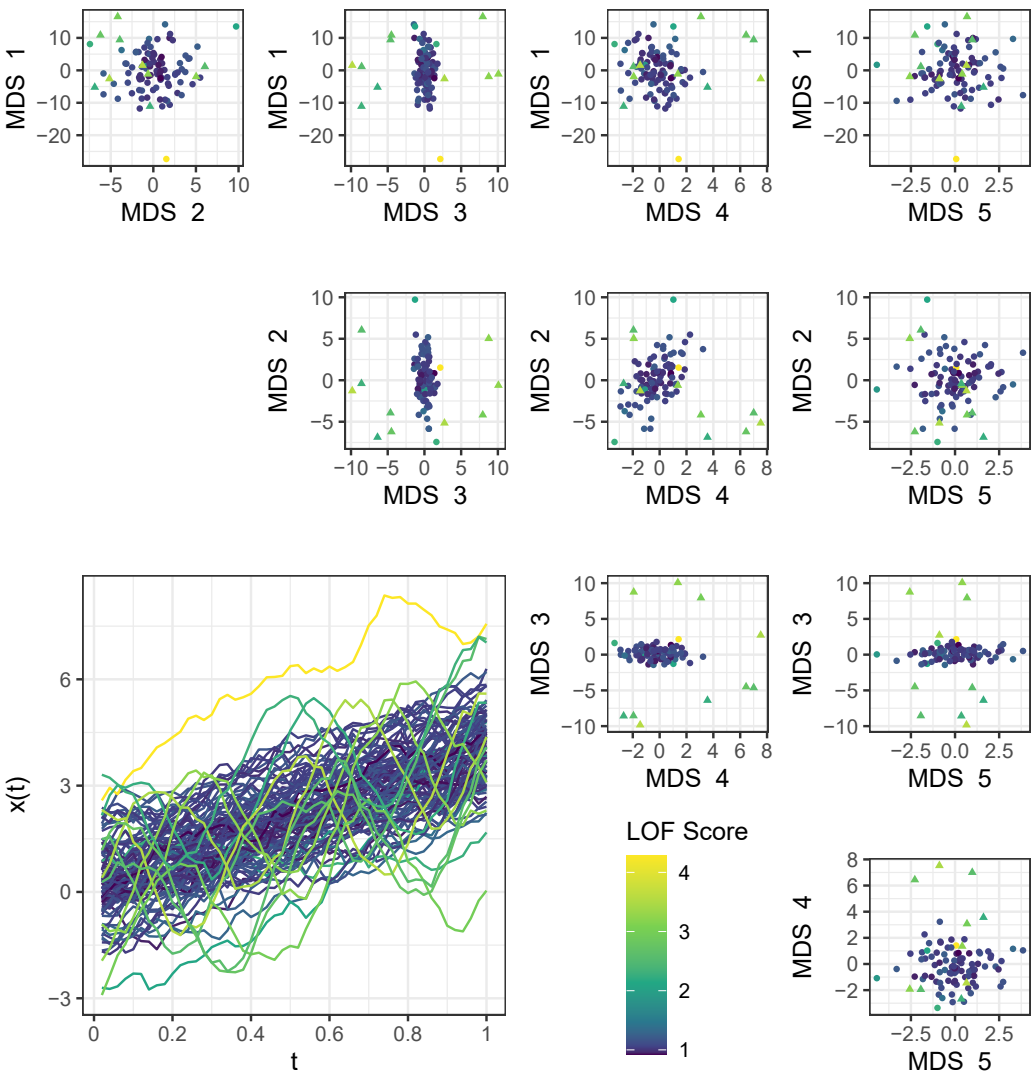


Figure A14. Model 7 data: scatterplot matrix of all 5 MDS embedding dimensions and curves; lighter colors for higher LOF score of 5D embeddings. True outliers depicted as triangles. Note that the true outliers are clearly separated from the rest of the data in embedding subspace 3 vs. 4.

First of all, note that the AUC computed for this specific dataset was 0.9, thus close to the median AUC value for LOF applied to MDS embeddings of Model 7 data, as depicted in Figure A3. Nevertheless, the "true outliers" are clearly separable in a 5D MDS embedding. As Figure A14 shows, they are clearly separable in the subspace spanned by the third and fourth embedding dimension. Note, moreover, that there is an outlying observation with an extreme shift, which also obtains a high LOF score. This observation is not labeled as a "true outlier", as it stems from $\mathcal{M}_c$. This example shows that evaluation approaches

for outlier detection methods that are based on "true outliers" may not always reflect the outlier structure adequately and may result in misleading conclusions. However, those approaches are frequently used to compare and assess different outlier-detection methods. Again, this illustrates the additional value low-dimensional embeddings have for outlier detection as such aspects become accessible.

Finally, note that the DO/MS-plots are not sensitive to vertical shift outliers as the extreme shift outlier is neither scored high based on DO nor labeled as an outlier based on the MS-plot; see Figure A15.

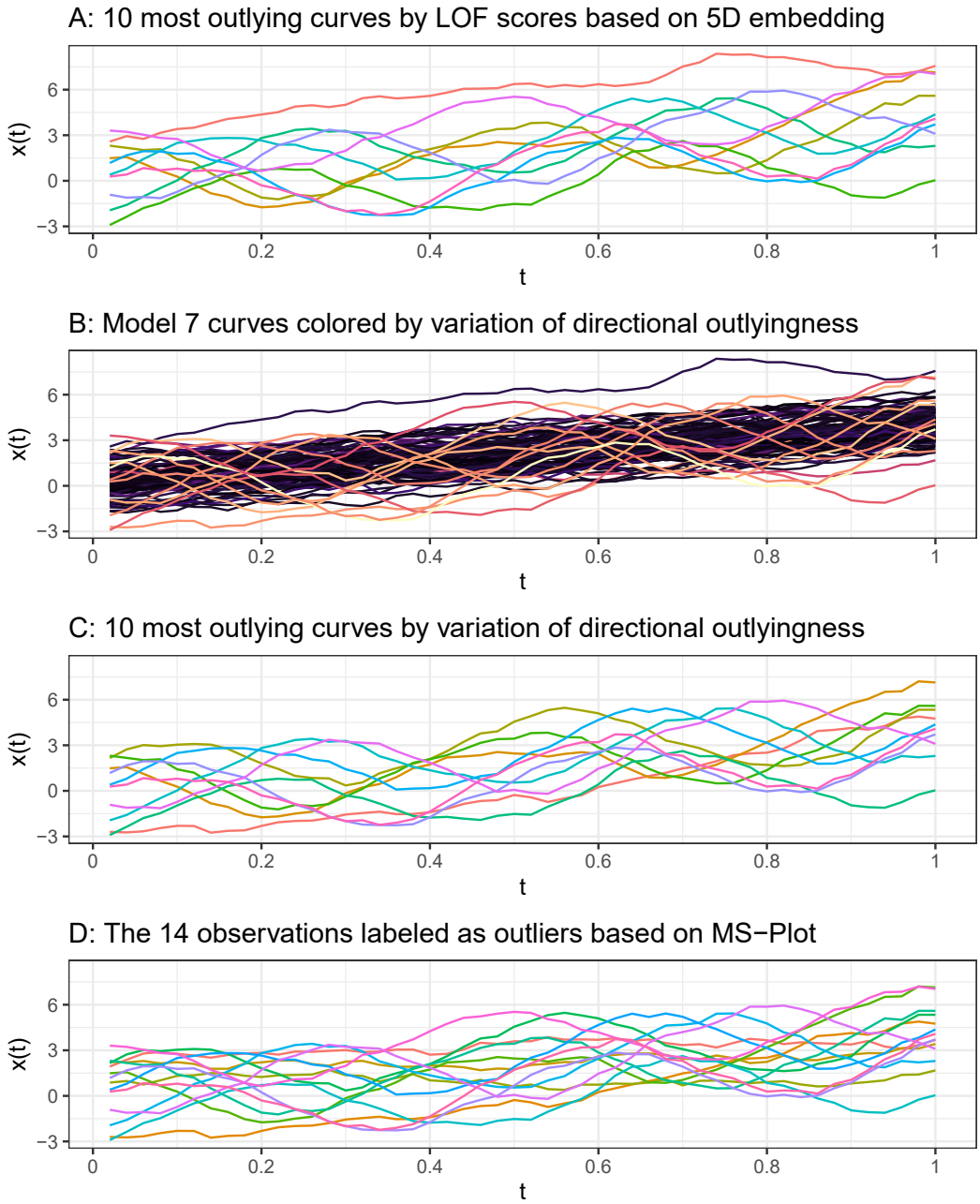


Figure A15. Model 7 data: the LOF on MDS embeddings in contrast to directional outlyingness.

Appendix F. Examples of the DGPs Used for the Quantitative Evaluation

Depicted in Figure A16 are two example datasets for each of the data-generating processes (DGPs) used in Section 3.2 for the comparison of the different outlier-detection methods.

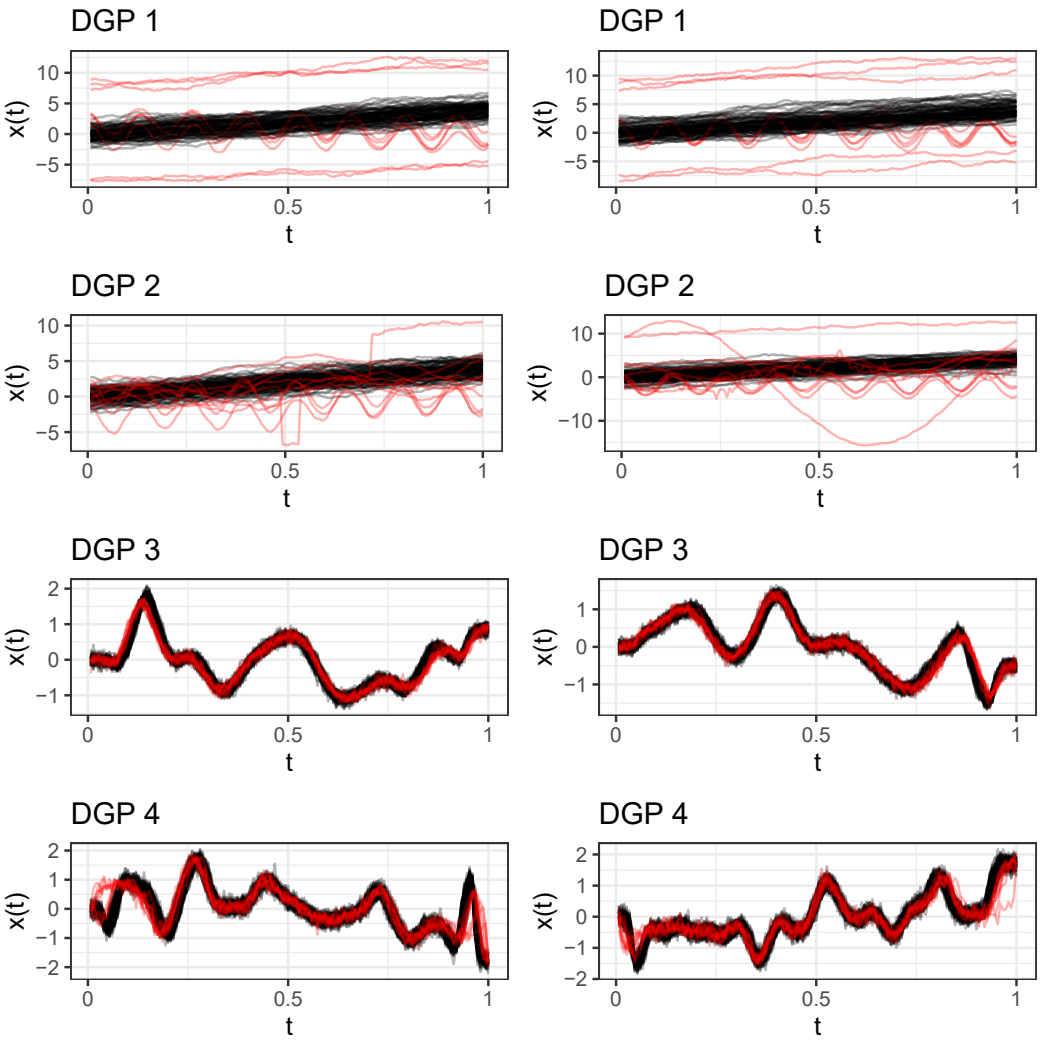


Figure A16. Example datasets for the DGPs used in the simulation study (2 each). Inliers in black; outliers in red. Outlier ratio 0.1; $n = 100$.

Appendix G. ArrowHead Data

Depicted in Figure A17 are the ArrowHead data used in Section 3.3.

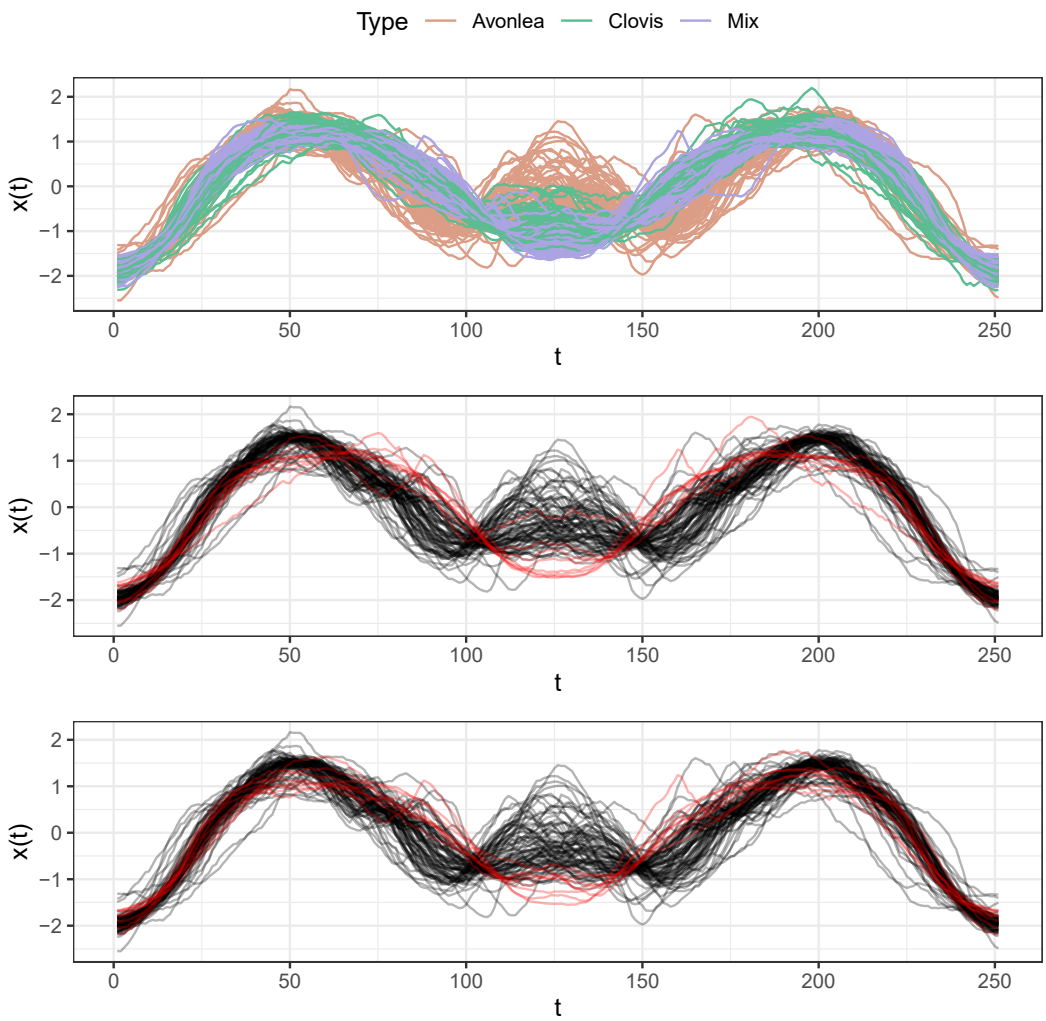


Figure A17. ArrowHead data. Top: the complete dataset. Middle and bottom: two example outlier datasets. Inliers from class "Avonlea" in black; outliers sampled from classes "Clovis" and "Mix" in red. Outlier ratio 0.1.

References

1. Dai, W.; Mrkvička, T.; Sun, Y.; Genton, M.G. Functional outlier detection and taxonomy by sequential transformations. *Comput. Stat. Data Anal.* **2020**, *149*, 106960. [CrossRef]
2. Arribas-Gil, A.; Romo, J. Discussion of "Multivariate functional outlier detection". *Stat. Methods Appl.* **2015**, *24*, 263–267. [CrossRef]
3. Hubert, M.; Rousseeuw, P.J.; Segaert, P. Multivariate functional outlier detection. *Stat. Methods Appl.* **2015**, *24*, 177–202. [CrossRef]
4. Ma, Y.; Fu, Y. *Manifold Learning Theory and Applications*; CRC Press: Boca Raton, FL, USA, 2012.
5. Lee, J.A.; Verleysen, M. *Nonlinear Dimensionality Reduction*; Springer Science & Business Media: New York, NY, USA, 2007.
6. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA, 15–18 May 2000; pp. 93–104.
7. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2005.
8. Hernández, N.; Muñoz, A. Kernel Depth Measures for Functional Data with Application to Outlier Detection. In *Artificial Neural Networks and Machine Learning–ICANN 2016*; Villa, A.E., Masulli, P., Pons Rivero, A.J., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; pp. 235–242.
9. Harris, T.; Tucker, J.D.; Li, B.; Shand, L. Elastic depths for detecting shape anomalies in functional data. *Technometrics* **2021**, *63*, 466–476. [CrossRef]
10. Sawant, P.; Billor, N.; Shin, H. Functional outlier detection with robust functional principal component analysis. *Comput. Stat.* **2012**, *27*, 83–102. [CrossRef]

11. Staerman, G.; Mozharovskyi, P.; Clémençon, S.; d'Alché Buc, F. Functional isolation forest. In Proceedings of the Eleventh Asian Conference on Machine Learning, Nagoya, Japan, 17–19 November 2019; Lee, W.S., Suzuki, T., Eds.; Volume 10, pp. 332–347.
12. Vinue, G.; Epifanio, I. Robust archetypoids for anomaly detection in big functional data. *Adv. Data Anal. Classif.* **2021**, *15*, 437–462. [CrossRef]
13. Rousseeuw, P.J.; Raymaekers, J.; Hubert, M. A measure of directional outlyingness with applications to image data and video. *J. Comput. Graph. Stat.* **2018**, *27*, 345–359. [CrossRef]
14. Dai, W.; Genton, M.G. Directional outlyingness for multivariate functional data. *Comput. Stat. Data Anal.* **2019**, *131*, 50–65. [CrossRef]
15. Xie, W.; Kurtek, S.; Bharath, K.; Sun, Y. A Geometric Approach to Visualization of Variability in Functional data. *J. Am. Stat. Assoc.* **2017**, *112*, 979–993. [CrossRef]
16. Hyndman, R.J.; Shang, H.L. Rainbow plots, bagplots, and boxplots for functional data. *J. Comput. Graph. Stat.* **2010**, *19*, 29–45. [CrossRef]
17. Ali, M.; Jones, M.W.; Xie, X.; Williams, M. TimeCluster: dimension reduction applied to temporal data for visual analytics. *Vis. Comput.* **2019**, *35*, 1013–1026. [CrossRef]
18. Yu, G.; Zou, C.; Wang, Z. Outlier Detection in Functional Observations with Applications to Profile Monitoring. *Technometrics* **2012**, *54*, 308–318. [CrossRef]
19. Chen, D.; Müller, H.G. Nonlinear manifold representations for functional data. *Ann. Stat.* **2012**, *40*, 1–29. [CrossRef]
20. Dimeglio, C.; Gallón, S.; Loubes, J.M.; Maza, E. A robust algorithm for template curve estimation based on manifold embedding. *Comput. Stat. Data Anal.* **2014**, *70*, 373–386. [CrossRef]
21. Herrmann, M.; Scheipl, F. Unsupervised Functional Data Analysis via Nonlinear Dimension Reduction. *arXiv* **2020**, arXiv:2012.11987.
22. Cuevas, A. A partial overview of the theory of statistics with functional data. *J. Stat. Plan. Inference* **2014**, *147*, 1–23. [CrossRef]
23. Malkowsky, E.; Rakočević, V. *Advanced Functional Analysis*; CRC Press: Boca Raton, FL, USA, 2019.
24. Polonik, W. Minimum volume sets and generalized quantile processes. *Stoch. Process. Their Appl.* **1997**, *69*, 1–24. [CrossRef]
25. Ojo, O.; Lillo, R.E.; Anta, A.F. Outlier Detection for Functional Data with R Package fdaoutlier. *arXiv* **2021**, arXiv:2105.05213.
26. Zimek, A.; Filzmoser, P. There and back again: Outlier detection between statistical reasoning and data mining algorithms. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1280. [CrossRef]
27. Cox, M.A.; Cox, T.F. Multidimensional scaling. In *Handbook of Data Visualization*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 315–347.
28. Tenenbaum, J.B.; Silva, V.D.; Langford, J.C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **2000**, *290*, 2319–2323. [CrossRef]
29. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
30. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2020**, arXiv:1802.03426.
31. Gangbo, W.; Li, W.; Osher, S.; Puthawala, M. Unnormalized optimal transport. *J. Comput. Phys.* **2019**, *399*, 108940. [CrossRef]
32. Bagnall, A.; Lines, J.; Bostrom, A.; Large, J.; Keogh, E. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Discov.* **2017**, *31*, 606–660. [CrossRef]
33. Olszewski, R.T. Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2001.
34. Dai, W.; Genton, M.G. Multivariate functional data visualization and outlier detection. *J. Comput. Graph. Stat.* **2018**, *27*, 923–934. [CrossRef]
35. Shang, H.L.; Hyndman, R.J. *fds: Functional Data Sets*; R Package Version 1.8; R package; 2018.
36. Kalivas, J.H. Two datasets of near infrared spectra. *Chemom. Intell. Lab. Syst.* **1997**, *37*, 255–259. [CrossRef]
37. Febrero-Bande, M.; Oviedo de la Fuente, M. Statistical Computing in Functional Data Analysis: The R Package fda.usc. *J. Stat. Softw.* **2012**, *51*, 1–28. [CrossRef]
38. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*; Springer Science & Business Media: New York, NY, USA, 2006.
39. Holland, J.; Kemsley, E.; Wilson, R. Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purees. *J. Sci. Food Agric.* **1998**, *76*, 263–269. [CrossRef]
40. Mead, A. Review of the development of multidimensional scaling methods. *J. R. Stat. Soc. Ser.* **1992**, *41*, 27–39. [CrossRef]
41. Arribas-Gil, A.; Romo, J. Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics* **2014**, *15*, 603–619. [CrossRef]
42. Ieva, F.; Paganoni, A.M.; Romo, J.; Tarabelloni, N. `roahd` Package: Robust Analysis of High Dimensional Data. *R J.* **2019**, *11*, 291–307. [CrossRef]
43. Shang, H.L.; Hyndman, R. *Rainbow: Bagplots, Boxplots and Rainbow Plots for Functional Data*; R package version 3.6; R package; 2019.
44. Huang, H.; Sun, Y. A decomposition of total variation depth for understanding functional outliers. *Technometrics* **2019**, *61*, 445–458. [CrossRef]

45. Ojo, O.T.; Lillo, R.E.; Fernandez Anta, A. *fdaoutlier: Outlier Detection Tools for Functional Data Analysis*; R package version 0.2.0.; R package; 2021.
46. Tucker, J.D. *fdasrvf: Elastic Functional Data Analysis*; R package version 1.9.7.; R package; 2021.
47. Dau, H.A.; Bagnall, A.; Kamgar, K.; Yeh, C.C.M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C.A.; Keogh, E. The UCR time series archive. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 1293–1305. [CrossRef]
48. Ye, L.; Keogh, E. Time series shapelets: A new primitive for data mining. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; pp. 947–956.
49. Rakthanmanon, T.; Campana, B.; Mueen, A.; Batista, G.; Westover, B.; Zhu, Q.; Zakaria, J.; Keogh, E. Searching and mining trillions of time series subsequences under dynamic time warping. In Proceedings of the 18th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 262–270.
50. Lemire, D. Faster retrieval with a two-pass dynamic-time-warping lower bound. *Pattern Recognit.* **2009**, *42*, 2169–2180. [CrossRef]
51. Fuchs, K.; Gertheiss, J.; Tutz, G. Nearest neighbor ensembles for functional data with interpretable feature selection. *Chemom. Intell. Lab. Syst.* **2015**, *146*, 186–197. [CrossRef]
52. Narayan, A.; Berger, B.; Cho, H. Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nat. Biotechnol.* **2021**, *39*, 765–774. [CrossRef] [PubMed]
53. De Silva, V.; Tenenbaum, J.B. Global versus local methods in nonlinear dimensionality reduction. *NIPS* **2002**, *15*, 705–712.
54. Brandes, U.; Pich, C. Eigensolver methods for progressive multidimensional scaling of large data. In *International Symposium on Graph Drawing*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 42–53.
55. Ingram, S.; Munzner, T.; Olano, M. Glimmer: Multilevel MDS on the GPU. *IEEE Trans. Vis. Comput. Graph.* **2008**, *15*, 249–261. [CrossRef] [PubMed]
56. Clémençon, S.; Thomas, A. Mass volume curves and anomaly ranking. *Electron. J. Stat.* **2018**, *12*, 2806–2872. [CrossRef]