A Decomposition of Total Variation Depth for Understanding Functional Outliers

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

A Decomposition of Total Variation Depth for Understanding Functional Outliers

Huang Huang & Ying Sun

To cite this article: Huang Huang & Ying Sun (2019): A Decomposition of Total Variation Depth for Understanding Functional Outliers, Technometrics, DOI: 10.1080/00401706.2019.1574241

To link to this article: https://doi.org/10.1080/00401706.2019.1574241

Accepted author version posted online: 11 Feb 2019.

A Decomposition of Total Variation Depth for Understanding Functional Outliers

Huang Huang[1] and Ying Sun[2]

[1]Statistical and Applied Mathematical Sciences Institute, Durham, North Carolina, 27709, USA. E-mail: hhuang@samsi.info

[2]CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. E-mail: ying.sun@kaust.edu.sa

***Abstract*–There has been extensive work on data depth-based methods for robust multivariate data analysis. Recent developments have moved to infinite-dimensional objects such as functional data. In this work, we propose a notion of depth, the total variation depth, for functional data, which has many desirable features and is well suited for outlier detection. The proposed depth is in the form of an integral of a univariate depth function. We show that the novel formation of the total variation depth leads to useful decomposition associated with shape and magnitude outlyingness of functional data. Compared to magnitude outliers, shape outliers are often masked among the rest of samples and more difficult to identify. We then further develop an effective procedure and visualization tools for detecting both types of outliers, while naturally accounting for the correlation in functional data. The outlier detection performance is investigated through simulations under various outlier models. Finally, the proposed methodology is demonstrated using real datasets of curves, images, and video frames.**

Some key words: data depth, functional data, total variation, outlier detection, shape outliers

Short title: Total Variation Depth

1 Introduction

Functional data, realizations of a one-dimensional stochastic process, in the form of functions, are observed and collected with increasing frequency across research fields, including meteorology, neuroscience, environmental science and engineering. Functional data analysis (FDA) considers the continuity of functions,

and various parametric and nonparametric methods can be found in Ferraty and Vieu (2006) and Ramsay et al. (2009). In recent years, extensive developments have extended typical techniques of FDA to the analysis of more complicated functional objectives. Besides model-based methods, exploratory data analysis (EDA) has been extended to functional data as well.

Sun and Genton (2011) proposed the functional boxplot as an informative visualization tool for functional data, and Genton et al. (2014) extended it to image data. Similar to the classical boxplot, if we simply extend the univariate ranking to the functional setting, the features of functional data cannot be captured. Data depth is a widely used concept in multivariate and functional data ranking. The general requirement for a data depth notion is the so-called "center-outwards" ordering, which means the natural center of the functional data should have the largest depth value and the depth decreases as the data approach outwards. Liu et al. (1999) reviewed many popular depth notions for multivariate data. Some examples in the nonparametric framework include the half-space depth (Tukey 1975) as well as its random version the random Tukey depth (Cuesta-Albertos and Nieto-Reyes 2008), the simplicial depth (Liu 1990), the projection depth (Zuo and Serfling 2000), and the spatial depth (Serfling 2002). For univariate functional data depth, there are also many existing notions, such as the (modified) band depth (López-Pintado and Romo 2009), the integrated data depth (Fraiman and Muniz 2001), the *h*-mode depth and the random projection depth (Cuevas et al. 2007), the half-region depth (López-Pintado and Romo 2011), and the extremal depth (Narisetty and Nair 2016). Some functional data depth are extended from their multivariate version, such as the functional Tukey depth (Dutta et al. 2011), the functional spatial depth (Chakraborty and Chaudhuri 2014a,b) as well as its kernel-based version, the kernelized functional spatial depth (Sguera et al. 2014). Nowadays, the study of multivariate functional data also arouses extensive interests due to its practical applications. Berrendero et al. (2011) studied the daily temperature functions on different surfaces, and Sangalli et al. (2009) as well as Pigoli and

Sangalli (2012) analyzed multivariate functional medical data. To provide a valid depth of multivariate functional data, Ieva and Paganoni (2013) derived depth measures for multivariate functional data from averaging univariate functional data depth, and Claeskens et al. (2014) gave a generalization to multivariate functional data depths by averaging a multivariate depth function over the time points with a weight function. Particularly, López-Pintado et al. (2014) proposed and studied the simplicial band depth for multivariate functional data, which is an extension of the univariate functional band depth.

Data depth-based methods provide many attractive tools in solving problems related to classification, clustering, and outlier detection. For example, Jörnsten (2004), Ghosh and Chaudhuri (2005), and Dutta and Ghosh (2012) used various depth notions to classify or cluster multivariate data. López-Pintado and Romo (2006) and Cuevas et al. (2007) extended the classification problem to functional data. Outlier detection is another challenging problem, especially for functional data, because there is no clear definition of functional outliers. Roughly speaking, functional outliers can be categorized into two types: magnitude and shape outliers. Magnitude outliers have very deviated values in some dimensions. For shape outliers, there is no universal definition because too many types of shape outliers exist. Generally speaking, shape outliers refer to realizations with any source that leads to a different pattern, ranging from oscillation with a different frequency or phase to a sudden sharp jump. Such outliers are called shape outliers because their outlyingness is not necessarily in magnitude. When talking about shape outliers, we do not imply shape outliers and magnitude outliers are mutually exclusive. Shape outliers can also be magnitude outliers if there is a magnitude deviation. Therefore, when we refer to shape outliers, they can be either pure shape outliers without any magnitude deviation or shape outliers with partial magnitude outlyingness that leads to certain changes in the overall pattern. When the outlyingness is only in magnitude, we call them pure magnitude outliers. Dang and Serfling (2010) introduced nonparametric multivariate outlier identifiers based on various

multivariate depths. For functional data, Febrero et al. (2008) used a cutoff, which is determined by a bootstrap, for the functional data depth to detect outliers. Hyndman and Shang (2010) proposed using the first two robust principal components to construct a bagplot (Rousseeuw et al. 1999), or a highest density region plot (Hyndman 1996) to detect outliers. However, the first two principle components often do not adequately describe the variability in functional data. In the functional boxplot proposed by Sun and Genton (2011), outliers were detected by the 1.5 times the 50% central region rule, which means that any departure of an observation from the fence, which is the inflation of the 50% central region by 1.5 times, makes it an outlier. The 50% central region is the envelope of the first 50% of curves that have the largest depth values, and the factor 1.5 can be adjusted according to the distribution (Sun and Genton 2012). Good performance has been shown for various types of outliers, especially for magnitude outliers. However, for shape outliers, two issues arise: outliers may appear among the first 50% of curves with the largest depth values, and outliers that are masked in the fence can never be detected even though their depth values are small. Recently, Arribas-Gil and Romo (2014) proposed the outliergram, where the relationship between modified band depth (López-Pintado and Romo 2009) and modified epigraph index (López-Pintado and Romo 2011) is studied to identify shape outliers. Hubert et al. (2015) proposed functional bagdistance and skewness-adjusted projection depth to detect various outliers for multivariate functional data. Sguera et al. (2016) used the proposed kernelized functional spatial depth to detect outliers and applied it to identifying abnormal emission levels of nitrogen oxides. Dai and Genton (2019) proposed functional directional outlyingness that accounts for the direction of outlyingness and used it to detect different types of multivariate functional data outliers.

In this paper, we introduce the total variation depth, which is essentially in the form of the integrated depth (Fraiman and Muniz 2001) but capable to detect both magnitude and shape outliers due to our novel formation of the depth notion. Specifically, the characterization of the pointwise total variation depth

leads to a useful decomposition, and the associated components in the decomposition allow us to propose shape similarity and modified shape similarity, which are powerful quantities to distinguish shape outliers. This work is motivated by the drawback of functional data depths that merely look at the centrality pointwisely and do not account for time order or temporal correlations, such as the integrated depth and modified band depth. In contrast, the decomposition of the total variation depth considers the correlation of adjacent dimensions and can be used for outlier detection. In our proposed outlier detection procedure, we show that combined with the functional boxplot, we are able to detect both magnitude and shape outliers. We also develop informative visualization tools for easy interpretation of outlyingness. The outlier detection performance has been examined through simulation studies, where we show the results of the comparison between our proposed methods and many popular existing outlier detection methods. Those existing methods include the functional boxplot using the modified band depth (MBD) and 1.5 times of the 50% central region rule proposed by Sun and Genton (2011), outlier detection performance by replacing MBD with the proposed extremal depth (ED) in the functional boxplot by Narisetty and Nair (2016), the proposed outliergram (OG) for shape outlier detection by Arribas-Gil and Romo (2014), the functional outlier map (FOM) based on the proposed adjusted outlyingness (AO) by Hubert et al. (2015), the extended FOM based on the proposed directional outlyingness (DO) by Rousseeuw et al. (2018). We also introduce the visualization tool and illustrate it through three kinds of real-data applications: curves, images and video frames. The R code implementing the total variation depth, the modified shape similarity, and our proposed outlier detection approach can be found in https://github.com/hhuang90/TVD.

2 Methodology

2.1 Total variation depth

Let X be a real-valued stochastic process on $\mathcal{T}$ with distribution F_X, where $\mathcal{T}$ is an interval in $\mathbb{R}$. We propose the total variation depth for a given function f with respect to (w.r.t.) F_X. We use f to denote a function, and $f(t)$ to denote the functional value at a given t. First, we define the pointwise total variation depth for a given t. Let $R_f(t) = 1\{X(t) \leq f(t)\}$, where 1 is the indicator function. It is easy to see that $p_f(t) = \mathbb{E}\{R_f(t)\} = \mathbb{P}\{X(t) \leq f(t)\}$ is associated with the relative position of $f(t)$ w.r.t. $X(t)$. For instance, if $f^*(t)$ is the true median, then $p_{f^*}(t) = 1/2$. For a given function $f(t)$ at each fixed t, we introduce the pointwise total variation depth of $f(t)$ by $D_f(t) = \mathrm{Var}\{R_f(t)\} = p_f(t)\{1 - p_f(t)\}$, which is almost the same as to the univariate simplicial depth (the univariate simplicial depth is given by $\mathbb{P}\{X(t) \leq f(t)\}\mathbb{P}\{X(t) \geq f(t)\} \times 2n/(n-1)$ for n functional data observations), however, with a different characterization. For a fixed t, we know $D_f(t)$ is maximized at the center when $p_f(t) = 1/2$, or simply the univariate median. Next, we define the functional total variation depth (TVD) for the given function $f(t)$ on $\mathcal{T}$ by

$$\mathrm{TVD}(f) = \int_{\mathcal{T}} w(t) D_f(t) dt,$$

where $w(t)$ is a weight function defined on $\mathcal{T}$. Although the proposed depth is in the form of the integrated depth (Fraiman and Muniz 2001), with the choice of a notion similar to simplicial depth as the univariate depth function, we will see later in Section 2.2 that the novel formulation of $D_f(t)$ as the variance of $R_f(t)$ leads to useful decomposition associated with shape and magnitude properties of the functions. There are many ways to choose the weight function, and we now provide two examples for the choice of $w(t)$. If we let $w(t)$ be a constant as $w(t) \equiv 1/|\mathcal{T}|$, then it can be shown that the TVD behaves similarly to the modified band depth (López-Pintado and Romo 2009). Another example is to let $w(t)$ be proportional to the vertical variability at different time points by Claeskens et al. (2014). The motivation is to put more emphasis on the order of curves in the regions with large amplitude variability, which may lead to a better separation

in the resulting sample depth values. We have considered both and notice that these two choices of weight functions perform similarly in the outlier detection procedure we proposed. Considering what Nagy et al. (2016) pointed out for the potentially inconsistent estimation of the integrated depth with weights depending on the marginal distribution, we then choose the constant weight to present our results in this paper, and focus on the novel formulation of the TVD and its useful decomposition for outlier detection.

2.2 Properties of the total variation depth

Zuo and Serfling (2000) studied the key properties of a valid multivariate data depth, and Claeskens et al. (2014) extended them to a functional setting, including affine invariance, maximality at the center, monotonicity relative to the deepest point, and vanishing at infinity. Denote the total variation depth of f w.r.t. F_X by $\mathrm{TVD}(f, F_X)$. Having a form of the integrated depth (Fraiman and Muniz 2001), the TVD enjoys the following properties immediately:

- **Affine invariance**. $\mathrm{TVD}(f, F_X) = \mathrm{TVD}(af + g, F_{aX+g})$ for any $a \in \mathbb{R} \setminus \{0\}$ and any function f and g on $\mathcal{T}$.
- **Maximality at the center**. If at each time point $t \in \mathcal{T}$, the distribution $F_X(t)$ has a uniquely defined center $f^*(t)$, then $\mathrm{TVD}(f^*, F_X) = \sup_{f \in C(\mathcal{T})} \mathrm{TVD}(f, F_X)$, where $C(\mathcal{T})$ denotes the set of all continuous functions on $\mathcal{T}$.
- **Monotonicity relative to the deepest point**. If f^* is the function such that $\mathrm{TVD}(f^*, F_X) = \sup_{f \in C(\mathcal{T})} \mathrm{TVD}(f, F_X)$, then $\mathrm{TVD}(f, F_X) \leq \mathrm{TVD}(f^* + \alpha(f - f^*), F_X)$, for any f on $\mathcal{T}$ and $\alpha \in [0,1]$.
- **Vanishing at infinity**. $\mathrm{TVD}(f, F_X) \to 0$, as $|f(t)| \to \infty$ for almost all time points t in $\mathcal{T}$.

The consistency of the sample integrated depth was first investigated by Fraiman and Muniz (2001). López-Pintado and Romo (2009) analyzed the consistency as

well as monotonicity and vanishing at infinity properties for their proposed modified band depth, which also falls into this integral form. Later, Nagy et al. (2016) further studied the consistency properties of the integrated depth with predetermined weight function, and pointed out the challenges to prove the consistency when the weight function depends on the distribution. Following these arguments, one can show that the sample TVD with the constant weight is a consistent estimator of the TVD.

Next, we show in Theorem 1 that the proposed pointwise total variation depth can be decomposed into the pointwise magnitude similarity and pointwise shape similarity, which has important practical implications for outlier detection problems. The result in Theorem 1 follows from the law of the total variance.

Theorem 1 (**Decomposition of TVD**). Let *s*, *t* be two time points where $s = t - \Delta$ provided $s, t \in \mathcal{T}$. The pointwise total variation depth has the following decomposition:

$$D_f(t) = \mathrm{Var}\{R_f(t)\} = \mathrm{Var}[\mathbb{E}\{R_f(t) \mid R_f(s)\}] + \mathbb{E}[\mathrm{Var}\{R_f(t) \mid R_f(s)\}].$$

The decomposition implies that the total variance of $R_f(t)$ can be decomposed into $\mathrm{Var}[\mathbb{E}\{R_f(t) \mid R_f(s)\}]$, the portion of the variability of $R_f(t)$ that is explained by $R_f(s)$, and $\mathbb{E}[\mathrm{Var}\{R_f(t) \mid R_f(s)\}]$, the portion of the variability of $R_f(t)$ that is independent of $R_f(s)$. We call $\mathrm{Var}[\mathbb{E}\{R_f(t) \mid R_f(s)\}]$ the shape component and $\mathbb{E}[\mathrm{Var}\{R_f(t) \mid R_f(s)\}]$ the magnitude component. If the shape component of a sample curve dominates the total variation depth, it indicates that the curve has a similar shape compared to the majority. Notice that the shape of the function *f* is characterized by considering a span of time Δ, where $1/\Delta$ is the sampling frequency. In practice, $(t - \Delta, t)$ denote two consecutive discrete time points.

To have a better insight into the two types of variability in the decomposition of $D_f(t)$, we use toy examples in Figure 1 to illustrate. There are 11 curves in total

including the outlier in each case in Figure 1. The time interval is chosen to be $[0.1,1]$ with $\Delta = 0.1$. For the magnitude outlier shown in the left panel in Figure 1, $D_f(t)$ is a constant 0.083 for all the time points. By calculations, the shape component $\mathrm{Var}[\mathbb{E}\{R_f(t) \mid R_f(s)\}]$ is exactly the same as $D_f(t) = 0.083$, and magnitude component $\mathbb{E}[\mathrm{Var}\{R_f(t) \mid R_f(s)\}]$ is 0. Even though the pointwise total variation depth of the magnitude outlier $D_f(t)$ is much smaller than that of the median, which is a constant 0.248 for all the time points, the shape component forms up the entire variability for $D_f(t)$. In contrast, for the shape outlier in the middle panel in Figure 1, the pointwise total variation depth $D_f(t)$ is a constant 0.217, which is not small compared to that of the median in the middle panel (also a constant 0.248). However, the shape components are 0.043 and 0.050 iteratively over time, which means the proportion of shape component in the entire variability is minute. As stated in the introduction in Section 1, the magnitude and shape outliers are not mutually exclusive. We show an example that is shape outlier with partial magnitude outlyingness in the right panel in Figure 1. The pointwise total variation depth $D_f(t)$ is a constant 0.248 over time $[0.1,0.5]$ and a constant 0.083 over time $[0.6,1]$, where we see $D_f(t)$ is small in the second half of the period due to its magnitude outlyingness. The shape components $\mathrm{Var}[\mathbb{E}\{R_f(t) \mid R_f(s)\}]$ are equivalent to $D_f(t)$ for all the time points in $[0.1,1]$ except time point 0.6 because there is a shape change at time point 0.6, where the shape component is 0.007, comprising a small portion of the corresponding entire variability $D_f(t) = 0.083$.

Now we give the definition of the shape similarity for a given function using the ratio of the shape component to the total variation depth.

Definition 1. For any function f on $\mathcal{T}$, the shape similarity (SS) of f given a time span Δ w.r.t. the distribution F_X is defined as

$$SS(f;\Delta) = \int_{\mathcal{T}} v(t;\Delta) S_f(t;\Delta) dt,$$

where $v(t;\Delta)$ is a weight function, and $S_f(t;\Delta)$ is given by

$$S_f(t;\Delta) = \begin{cases} \mathrm{Var}\left(\mathbb{E}[R_f(t) \mid R_f(t-\Delta)]\right) / D_f(t) & , \quad D_f(t) \neq 0 \\ 1 & , \quad D_f(t) = 0 \end{cases}.$$

Now we discuss the choice of the weight function $v(t;\Delta)$. Notice that at a given t, $R_f(t) = 1\{X(t) \leq f(t)\}$ is an indicator function. If $X(t) \leq f(t), R_f(t) = 1$ no matter how large the value of $f(t)$ is, which means $S_f(t;\Delta)$ is rank-based and hence does not depend on the difference in magnitude. To account for the magnitude change of $f(t)$, we choose the weight function $v(t;\Delta)$ to be the normalized changes in $f(t)$ on $\mathcal{T}$. More precisely, $v(t;\Delta)$ is given by $v(t;\Delta) = |f(t) - f(t-\Delta)| / \int_{\mathcal{T}} |f(t) - f(t-\Delta)|$. Then, more weights are assigned to the time intervals where $f(t)$ has larger changes in magnitude. This choice of $v(t;\Delta)$ is especially useful in practice when sample curves have magnitude changes within short time intervals. From the definition of $S_f(t;\Delta)$, it is easy to see that $0 \leq S_f(t;\Delta) \leq 1$. If $S_f(t;\Delta) = 0$ for all t, then $SS(f) = 0$. One example in this case is that $R_f(t)$ is independent of $R_f(t-\Delta)$ w.r.t F_X for all t, and then for any f, $\mathrm{Var}[\mathbb{E}\{R_f(t) \mid R_f(t-\Delta)\}] = \mathrm{Var}[\mathbb{E}\{R_f(t)\}] = 0$. Thus, $X(t)$ is an independent process. It implies that no functions share similar patterns. If $S_f(t;\Delta) = 1$ for all t, then $SS(f;\Delta) = 1$. One example in this case is that $X(t)$ is a perfectly correlated process, and then $\mathrm{Var}[\mathbb{E}\{R_f(t) \mid R_f(t-\Delta)\}] = \mathrm{Var}\{R_f(t)\}$ indicating $S_f(t;\Delta) = 1$. Thus, all the functions have the same shape.

In Definition 1, smaller values of the shape similarity are associated with larger shape outlyingness. However, for outlying pairs $(f(t-\Delta), f(t))$ with a small value of the shape component $\mathrm{Var}\left(\mathbb{E}[R_f(t) \mid R_f(t-\Delta)]\right)$, if $D_f(t)$ in the denominator is too small, $S_f(t;\Delta)$ may not be small enough. To better reflect the shape

outlyingness via the shape similarity, we shift $(f(t-\Delta), f(t))$ to the center, such that $(\tilde{f}(t-\Delta), \tilde{f}(t)) = (f(t-\Delta), f(t)) - \delta_t$, where $\delta_t = f(t) - \text{median}\{X(t)\}$, and then define the modified shape similarity using the shifted pairs as follows:

Definition 2. For any function f on $\mathcal{T}$, the modified shape similarity (MSS) of f given a fixed time span Δ w.r.t. the distribution F_X is defined as

$$\text{MSS}(f;\Delta) = \int_{\mathcal{T}} v(t;\Delta) S_{\tilde{f}}(t;\Delta) dt,$$

where for any pair $(t-\Delta, t)$, $\tilde{f}$ is given by

$$\tilde{f}(s;\Delta) = \begin{cases} \text{median}\{X(s)\} & , \quad s = t \\ f(s) - f(s+\Delta) + \text{median}\{X(s+\Delta)\} & , \quad s = t-\Delta \end{cases},$$

and $S_{\tilde{f}}(t;\Delta) = \text{Var}\left(\mathbb{E}[R_{\tilde{f}}(t) \mid R_{\tilde{f}}(t-\Delta)]\right) / D_{\tilde{f}}(t).$

Similar to the shape similarity, the modified shape similarity characterizes shape outlyingness of f via each pair of function values, but it is a stronger indicator by shifting each pair to the center when calculating $S_{\tilde{f}}(t;\Delta)$. Arribas-Gil and Romo (2014) used a similar idea for shape outlier detection, whereas the entire extreme curve at all time points is shifted to the center by the same amplitude.

It is noteworthy that one charm of TVD is that the way we formulate it can lead to the decomposition allowing us to compute MSS to cope with shape outlyingness. Once we get MSS, TVD with a constant weight works the same way as MBD.

In real applications, we observe functional data at discrete time points and only have samples from the true distribution. We replace the cumulative density function F_X by its empirical estimate and obtain the sample total variation depth and the sample modified shape similarity. All the estimations use sample proportions to approximate probabilities, and the details are provided in the Appendix.

2.3 Outlier detection rule and visualization

To obtain robust inferences, outlier detection is often necessary; however, the procedure is challenging for functional data, because the characterization of functional data is in infinite dimensions and appropriate outlier detection rules are needed. There are many proposed ways to detect functional outliers, some of which have been discussed in Section 1, and here, we propose a new outlier detection rule by making good use of the decomposition property of TVD. Although TVD is invariant under time permutation and thus can deal with magnitude outliers only, SS or MSS will change if we permute all curves over a series of time t, which is the reason why they can reflect the shape of a curve and be used to detect shape outliers. Suppose we observe n sample curves, then the outlier detection procedure is summarized as follows:

1. Estimate the total variation depth and modified shape similarity for each curve as described in Section 2.1 and Section 2.2.
2. Draw a classical boxplot for the n values of the modified shape similarity and detect outliers by the $F \times \mathrm{IQR}$ empirical rule, where the factor F can be adjusted by users, depending on the distributions. For example, Hubert and Vandervieren (2008) discussed the adjustment for skewed distributions and provide a more robust rule for outliers if the distribution is not symmetric. In the simulation study and applications, we choose $F = 3$ following the conservative outlier detection rule in a classical boxplot. Curves with modified shape similarity values below the lower fence in the boxplot are identified as shape outliers.
3. Remove detected shape outliers and draw a functional boxplot using the total variation depth to detect all the magnitude outliers by the 1.5 times of the 50% (w.r.t. the number of original observation before removing shape outliers) central region rule in the functional boxplot (Sun and Genton 2011). The factor 1.5 can be adjusted using bootstrap methods (Sun and Genton 2012).

It is noteworthy that using the functional boxplot along with the boxplot of the modified shape similarity enables us to detect both magnitude outliers and shape outliers even with small oscillations. We call the boxplot of the modified shape similarity the shape outlyingness plot because the modified shape similarity is a good indicator of shape outlyingness. It is especially useful in detecting shape outliers without a significant magnitude deviation.

In summary, the outlier detection procedure consists of two steps. In the first step, the shape outlyingness plot detects all shape outliers. The detected shape outliers in the first step can be either pure shape outliers or shape outliers with partial magnitude outlyingness as long as the shape outlyingness is big enough to be detected by MSS. In the second step, all the detected shape outliers are removed, and then the functional boxplot with TVD is applied to the rest of the data to detect magnitude outliers.

Next, we propose a set of informative visualization tools following the outlier detection procedure. Figure 2 (a) shows examples of simulated datasets containing different types of outliers. The model details will be introduced in Section 3.1. The grey curves are the non-outlying observations and the black curve is the functional median with the largest total variation depth value. The highlighted orange curves are the shape outliers detected by the shape outlyingness plot, as shown in Figure 2 (b), where the points in orange correspond to shape outliers. The red curve is a magnitude outlier detected by the functional boxplot constructed in Step 3 of the outlier detection procedure after removing the shape outliers. Finally, the functional boxplot in Figure 2 (c) displays the median, the 50% central region, the maximal and minimal envelopes, and the magnitude outlier.

3 Simulation Study

3.1 Outlier models

In this section, we choose a simulation design that some of the outlier models are originated from López-Pintado and Romo (2009), Sun and Genton (2011), and Narisetty and Nair (2016), but we also introduce some new outlier models. We study seven models in total, where Model 1 is the base model with no outliers compared to Model 2 – 6, and Model 7 concerns another kind of contamination with a different base model. The contamination ratio for outliers in Model 2 – 7 is $\rho = 10\%$. For all the models, $\mathcal{T}$ is set to be $\mathcal{T} = [0,1]$.

Model 1: $X_i(t) = 4t + e_i(t)$, for $i = 1, \ldots, n$, where $e(t)$ is a zero-mean Gaussian process with covariance function $c(s,t) = \exp\{-|s-t|\}$, for $s, t \in \mathcal{T}$.

Model 2: $X_i(t) = 4t + e_i(t) + 6c_i\sigma_i$, for $i = 1, \ldots, n$ where c_i ~Bernoulli(ρ) and σ_i takes values of 1 and –1 with probability 1/2, respectively, and $e(t)$ has the same definition as before.

Model 3: $X_i(t) = 4t + e_i(t) + 6c_i\sigma_i$, if $t \geq T_i$, and $X_i(t) = 4t + e_i(t)$, if $t < T_i$, for $i = 1, \ldots, n$, where $T_i \sim \text{Unif}([0,1])$, and $e_i(t)$, c_i and σ_i have the same definition as before.

Model 4: $X_i(t) = 4t + e_i(t) + 6c_i\sigma_i$, if $T_i \leq t \leq T_i + l$, and $X_i(t) = 4t + e_i(t)$ otherwise, for $i = 1, \ldots, n$, where $l = 0.08, T_i \sim \text{Unif}([0, 1-l])$, and $e_i(t)$, c_i and σ_i have the same definition as before.

Model 5: $X_i(t) = 4t + (1-c_i)e_i(t) + c_i\tilde{e}_i(t)$, for $i = 1, \ldots, n$, where c_i and $e_i(t)$ have the same definition as before, and $\tilde{e}_i(t)$ is another zero-mean Gaussian process with a different covariance function $\tilde{c}(s,t) = 6\exp\{-|s-t|^{0.1}\}$.

Model 6: $X_i(t) = 4t + e_i(t) + c_i\left(0.5\sin(40\pi t)\right)$, for $i = 1, \ldots, n$, where c_i and $e_i(t)$ have the same definition as before.

Model 7: $X_i(t) = 2\sin(15\pi x + 2c_i) + e_i(t),$ and the corresponding base model in Model 7 is $X_i(t) = 2\sin(15\pi x) + e_i(t)$, for $i = 1, \ldots, n$, where c_i and $e_i(t)$ have the same definition as before.

Then, Model 2 contains symmetric pure magnitude outliers with a vertical shift by adding a comparatively large positive or negative constant; Model 3 introduces the magnitude shift starting from a random time point, which leads to shape outlyingness in those outliers; Model 4 generates shape outliers with partial magnitude outlyingness, which have peaks lasting for a short time period; Model 5 has outliers showing different shapes as well as magnitude deviations caused by the larger variance in the covariance model; Model 6 considers pure shape outliers by adding an oscillating function, where the oscillation is frequent in time but close to the majority in magnitude; and Model 7 introduces shape outliers by changing the phase in the sine function in the base model, and in the meantime, the phase shift makes the outliers deviate from the majority in magnitude in some periods. The generated outliers from Model 2 – 7 are illustrated in Figure 3. For the following simulation studies, we generate $n = 100$ curves taking values on an equally spaced grid of $[0,1]$ with $p = 50$ time points for each model.

3.2 Central region

In this section, we study how the 50% central region is affected by different choices of depth notions in the given outlier models. The 50% central region is the envelope of the first 50% of curves that have the largest depth values. A good 50% central region should not contain or be affected by outliers. Ideally, it should be the smallest region that contains the deepest 50% functional data realizations ordered by a functional data depth. In our TVD and MSS method (TVD+MSS), the 50% central region is constructed by 50% of the deepest curves after removing the detected shape outliers by MSS as described in Section 2.3. We compare the central region constructed by our approach with those by the MBD (López-Pintado and Romo 2009) only and the ED (Narisetty and

Nair 2016). In Figure 4, we show the constructed central regions by different depth notions in different models. For Model 2, all the central regions look good because all the magnitude outliers have smaller depth values by the three notions and do not fall into the first 50% of curves. For Model 3, we see that the central region using MBD only is contaminated by the outliers. ED and our approach give good central regions. The reason why MBD fails is that it only accounts for averaged magnitude outlyingness, and for outliers that only deviate in a short time period, MBD may assign large depth values, so that these outliers fall into the first half of the deepest curves. By definition, TVD also considers the averaged magnitude outlyingness as what MBD does. However, by removing the detected shape outliers first, the constructed central region is robust and does not get affected by the outliers. It should be noteworthy that these shape outliers do show partial magnitude outlyingness, but not enough to be detected by MBD. ED focuses on the extremity, and assigns small depth values to those outliers even though the extremity may only occur in a short period. Therefore, the resulting central region is expected to exclude the outliers. The outliers in Model 4 are similar to those in Model 3 but with an even further shorter period for the outliers deviating from the center. That is why MBD has worse performance in providing the central regions whereas ED and our approach give the non-contaminated central regions. For Models 5 - 7, all the outliers do not fall into the first 50% of curves that have the largest depth values by the three notions so that all the central regions look good. Overall, the performance of our proposed method shows appealing property in constructing the 50% central region.

3.3 Outlier detection

In this section, we use our outlier detection rule described in Section 2.3 to detect outliers. We compare the outlier detection performance between our proposed methods and existing outlier detection methods that include the functional boxplot of the MBD by Sun and Genton (2011), the functional boxplot of the ED by Narisetty and Nair (2016), the OG by Arribas-Gil and Romo (2014), the FOM

based on the AO by Hubert et al. (2015), the extended FOM based on the DO by Rousseeuw et al. (2018). In addition to these existing outlier detection methods that analyze the functional data itself, an alternative way is to consider the derivative of the functional data since the derivative may reflect some shape outlyingness for shape outliers that we are interested in. For comparisons to our methods, we also include several methods from this idea. One is to use the functional boxplot based on MBD of the original functional data and the derivative separately to detect outliers. The other one is to consider the original functional data and the derivative as a bivariate functional data, and then use the modified simplicial band depth (MSBD) proposed by López-Pintado et al. (2014) to detect outliers. We apply two weights for calculating MSBD: the constant weight and the same weight $v(t;\Delta)$ as ours in Definition 1. All the methods including ours that are used in outlier detection are summarized below. To have a better insight into the performance of our proposed methods, we also present the results when only MSS is used to detect the shape outliers. A detailed description of each method that is used in the outlier detection in this section is given below.

- **TVD+MSS**. Shape outliers are detected by the boxplot of the modified shape similarity (MSS) first, and then the total variation depth (TVD) is used in the functional boxplot to detect magnitude outliers.
- **MBD**. The modified band depth (MBD) is used in the functional boxplot to detect outliers.
- **ED**. The extremal depth (ED) is used in the functional boxplot to detect outliers.
- **OG+MBD**. Shape outliers are detected by the outliergram (OG) first, and then the modified band depth (MBD) is used in the functional boxplot to detect magnitude outliers.
- **FOM(AO)**. The functional outlier map (FOM) based on the adjusted outlyingness (AO) is used to detect outliers.
- **FOM(DO)**. The functional outlier map (FOM) based on the directional outlyingness (DO) is used to detect outliers.

- **MSBD1**. The modified simplicial band depth (MSBD) calculated by the constant weight is used in the functional boxplot to detect outliers.
- **MSBD2**. The modified simplicial band depth (MSBD) calculated by the weight $v(t;\Delta)$ as ours in Definition 1 is used in the functional boxplot to detect outliers.
- **MBD+MBD**. The functional boxplot based on the modified band depth (MBD) of the functional data is used to detect magnitude outliers and the functional boxplot based on MBD of the derivative functional data is used to detect shape outliers.
- **MSS-only**. Only shape outliers are detected by the boxplot of the modified shape similarity (MSS).

In the experiment, we assess the performance by the true positive rate (TPR), which is the ratio of the number of correctly detected outliers to the number of true outliers, and the false positive rate (FPR), which is the ratio of the number of wrongly detected outliers to the number of true non-outliers. A larger TPR means more outliers are correctly detected, and a smaller FPR means fewer non-outliers are falsely detected as outliers.

The results of 500 experiments are shown in Table 1 for each model and each method, with means and standard deviations (in parentheses) shown as percentages. We see that our proposed outlier detection method has an excellent performance for all models, especially for Models 3 - 7, where shape outliers exist, and MSS-only is successful in detecting nearly all the shape outliers. Even for Model 4, which ED favors, our detection method successfully detects all of the outliers with a TPR of 100%, and it outperforms the functional boxplot with ED, for which the TPR is 86.62%. For Model 6, where shape outliers only have small oscillations, all the other methods fail with very low TPRs, but our method still gains high accuracy. For shape outliers with partial magnitude outlyingness as in Models 3, 4, 5 and 7, the outliergram correctly detects most of the outliers, while our method has similar or higher TPRs, but much lower FPRs.

Generally speaking, the functional outlier map with directional outlyingness performs better than that with adjusted outlyingness, and we see it succeeds in detecting outliers that lay outside the center for a long time period in Model 3. But for outliers only peaking shortly as in Model 4 and other types of shape outlyingness as in Model 5 - 7, the detection performance is not satisfactory. The natural idea of making use of the derivatives indeed works. From the comparisons, we find that for the models we have considered, detecting outliers by using the original functional data and its derivative separately is generally better than treating them as bivariate functional data and using modified simplicial band depth. The performance of the method MBD+MBD has good TPRs in most models except for Model 6 where an oscillating noise exists. For outliers in Model 2, all the methods perform well except for the MSS-only because these outliers are pure magnitude outliers and do not have a different shape. For Model 1, where no outliers exist, all the methods except the outliergram retain a low FPR. In fact, a relatively high FPR of the outliergram is observed for all the models, indicating that the outliergram tends to falsely detect too many outliers. Overall, we see our proposed outlier detection procedure outperforms many popular existing methods and succeeds in detecting both the magnitude and shape outliers.

The mean running time of the 500 experiments for each outlier detection method in this simulation study is shown in Table 2 with the corresponding standard deviation in parentheses. The experiments are conducted on a 3.1 GHz Intel Core i7 processor. We see MBD is significantly faster than all other methods because of the fast algorithm developed by Sun et al. (2012). The MBD-based method MBD+MBD comes the next since it only requires computing the MBD twice with the original curve and derivative in addition to the small overhead for calculating the derivative. OG+MBD and MSBD are observed much slower than others. Our proposed approach is in the middle regarding computation time. Even though the computation time is not our major concern, it may be an issue if

the functional dataset to be analyzed is too large, and in this case, MBD-based methods may be favored to provide results more quickly.

4 Applications

In this section, we apply our proposed outlier detection procedure and visualization tools to three applications: the sea surface temperature in Niño zones, the sea surface height in the Red Sea, and the surveillance video of a meeting room. The three applications cover examples of datasets consisting of curves, images, and video frames.

4.1 Sea surface temperature

The El Niño southern oscillation (ENSO), irregular cycles of warm and cold temperatures in the eastern tropical Pacific Ocean, has a global impact on climate and weather patterns, including temperature, rainfall, and wind pressure. There are several indices to measure ENSO, one of which is the sea surface temperature in the Niño regions. The warm phase of ENSO is referred as El Niño, and the cold phase is called La Niña. For example, the years 1982–1983, 1997–1998, and 2015–2016, were reported as strong El Niño years, because the sea surface temperature anomalies were significantly larger for several consecutive seasons.

In this application, we use our method on the monthly sea surface temperature in one of the Niño zones ($0-10^{\circ}$ South and $90-80^{\circ}$ West) from July 1982 to June 2016. In Figure 5 (a), each curve represents a yearly sea surface temperature from July to next June. There are 35 curves in total. We see in Figure 5 (b) that there are no shape outliers, and two magnitude outliers have been detected, which are the years 1982–1983 and 1997–1998. The outliers coincide with the strong El Niño events. The upper envelop shown as the upper blue line in Figure 5 (c), partly comprises July and August of 1983 (shown by the green line in Figure 5 (a)), which is the extension of the El Niño year 1982–1983, and September to December of 2015 (shown by the purple line in Figure 5 (a)),

agreeing with the El Niño year 2015–2016. The median is the year 1989–1990 shown as the black line in Figure 5 (a), when no obvious El Niño event occurred.

4.2 Sea surface height

Sea surface height plays an important role in understanding ocean currents. Oceanographers run simulations with different initial conditions to generate ensembles. To understand the variability in these model runs, we apply our method to 50 ensemble runs of simulated sea surface heights of the Red Sea on January 1, 2016, where each observation is an image with 25,808 valid values. We first order the observations along the axis of the Red Sea in a zigzag fashion and obtain the one-dimensional functional data as shown in Figure 6. Then we apply our outlier detection method to the constructed functional data and get two detected shape outliers, one of which is shown in Figure 8 (b). The associated MSS values are illustrated in Figure 7. After removing the two shape outliers, the total variation depth is used in the functional boxplot to detect magnitude outliers and we find seven with one shown in Figure 8 (c). To have a better insight into the most representative sample of the sea surface height, we show the median in Figure 8 (a). We see that compared to the median, which is treated as the most representative realization among the 50 ensemble runs, the shape outlier has a different pattern in the northern part of the Red Sea. From the median as shown in Figure 8 (a), there is a connected area that has extremely lower height in the northern part of the Red Sea, but in the shape outlier as shown in Figure 8 (b), the area of extremely lower height tends to be isolated. For the detected magnitude outlier as shown in Figure 8 (c), the highest and lowest heights have a larger deviation from zero, as indicated by the dark red in the center part of the Red Sea and dark blue in the northern part of the Red Sea.

The outlier detection results may be different if a different way of ordering is used to transform the image data into the one-dimensional data. We have tried different ways other than the zigzag fashion, such as ordering by row, or column and found the outlier detection results are slightly different. The difference is

small because the ordering only affects the way to define the adjacent indices. Since the proposed MSS is essentially calculated from two adjacent indices, the shape outlier detection will not be affected very much if we use a reasonable ordering that makes adjacent indices close to each other in the two-dimensional space. In contrast, the ordering does not affect TVD and has no influence on detecting magnitude outliers. Therefore, in this paper, we chose the zigzag fashion of ordering, which makes the two adjacent indices reflect their smallest distance. However, extending the application of MSS to image data by properly defining adjacent indices in two-dimensional space requires future research.

4.3 Surveillance video

Video datasets consist of a sequence of video frames, each of which can be treated as one functional observation. Rousseeuw et al. (2018) illustrated their outlier detection method using a beach surveillance video. We explore a video of a meeting room, filmed by Li et al. (2004).

In this video, the meeting room is empty and no one was inside at the beginning. The curtain was moving with the wind. At a certain time, one person came into the meeting room, stayed for a while, and left. That person entered the room three times in total in the entire dataset. We applied our outlier detection approach to detect outliers, expecting to find the frames where either the curtain is moving too far away, or the person is in the meeting room.

This video was filmed for 94 seconds and consists of 2,964 frames. We first equally sub-sampled 1,280 out of 20,480 pixels in each video frame to capture the main features of the image. Then, the selected pixels are ordered by row in a zigzag fashion to obtain the one-dimensional data vector. The constructed functional data that consist of 2,964 curves can be seen in Figure 9.

After applying our outlier detection method, we get 306 shape outliers and 123 magnitude outliers. An illustration of the MSS can be seen in Figure 10. It is

noteworthy that the shape outliers and magnitude outliers are not mutually exclusive, and that shape outliers can also have partial magnitude outlyingness. It turns out that all the 306 detected shape outliers (either pure shape outliers or shape outliers with partial magnitude outlyingness) by MSS are those frames with the person inside the room. Figure 11 (b) is one example of the detected shape outliers. After removing all the detected shape outliers and applying the functional boxplot using TVD, the 123 detected magnitude outliers contain both the frames with the presence of a person and the frames with the curtain moving away. Figure 11 (c) is one example of the detected magnitude outliers, where we see the curtain moved far away from the still position. We also show the median, which indicates the most representative frame of the entire dataset in Figure 11 (a), where we see the still position of the curtain and nobody in the meeting room.

5 Discussion

In this paper, we proposed a depth notion, the total variation depth, for functional data. We illustrated that this notion has many attractive properties, and in particular, its decomposition is highlighted for the strength in detecting shape outliers of the most challenging type. We proposed the modified shape similarity defined via the decomposition as an indicator of the shape outlyingness and constructed the shape outlyingness plot to detect shape outliers. Together with the functional boxplot, our proposed outlier detection procedure can detect both magnitude and shape outliers. Good performances of outlier detection were demonstrated through simulation studies with various types of outlier models. Moreover, we also developed a set of visualization tools for functional data, where we display the original observations with informative summary statistics, such as the median and the central region, and we highlight the detected magnitude and shape outliers, along with the corresponding shape outlyingness plot and the functional boxplot.

Another advantage of the proposed total variation depth is that it can be easily extended to multivariate functional data by properly defining the indicator function $R_f(t) = 1\{X(t) \leq f(t)\}$. For example, for a multivariate function $\mathbf{f}(t)$ that takes values in $\mathbb{R}^p$, let $R_{\mathbf{f}}(t) = 1\{\mathbf{f}(t) \in \mathbf{S}(t)\}$, where $\mathbf{S}(t)$ is a properly defined set for the multivariate stochastic process $\mathbf{X}$ at time t. One naive choice would be $\mathbf{S}(t) = \{\mathbf{X}(t) : (\mathbf{X}(t) \leqslant (\mathbf{f}(t)\}$. Another possible extension is that when we compute the MSS in practice, we use the previous time point as the conditional set in $S_f(t;\Delta)$ in a Markovian manner. To characterize the shape similarity in a more generalized way, we can choose two or more time points ahead of the underlying time point. However, further research is needed for these extensions with a careful investigation.

Appendix

Estimation of the total variation depth and shape similarity

Recall that the pointwise total variation depth is defined as $D_f(t) = p_f(t)\{1 - p_f(t)\}$, where $p_f(t) = \mathbb{P}\{X(t) \leq f(t)\}$. Given n observations at m time points $X_j(t_i),\ j = 1,\ldots,n,\ i = 1,\ldots,m$, we estimate $D_f(t_i)$ by $\hat{D}_f(t_i) = \hat{p}_f(t_i)\{1 - \hat{p}_f(t_i)\}$, where $\hat{p}_f(t_i) = \#\{j : X_j(t_i) \leq f(t_i)\}/n$, the proportion of $X_j(t_i)$'s that are below $f(t_i)$. To estimate the shape similarity, which is defined by $S_f(t;\Delta) = \mathrm{Var}\big(\mathbb{E}[R_f(t) \mid R_f(t-\Delta)]\big)/D_f(t)$ for $D_f(t) \neq 0$, namely, $p_f(t_i) \neq 1$, we show how to estimate $\mathrm{Var}\big(\mathbb{E}[R_f(t) \mid R_f(t-\Delta)]\big)$. The population estimation is done on the discrete time points, and at each $t_i, i > 1$, we let $\Delta = t_i - t_{i-1}$. Since

$$\mathbb{E}\{R_f(t_i) \mid R_f(t_{i-1})\} = \mathbb{P}\{R_f(t_i) = 1 \mid R_f(t_{i-1})\} = \mathbb{P}\{X(t_i) \leq f(t_i) \mid R_f(t_{i-1})\},$$

then

$$\begin{aligned}\mathrm{Var}[\mathbb{E}\{R_f(t_i) \mid R_f(t_{i-1})\}] &= \mathrm{Var}[\mathbb{P}\{X(t) \le f(t_i) \mid R_f(t_{i-1})\}] \\ &= \mathbb{E}[\mathbb{P}^2\{X(t_i) \le f(t_i) \mid R_f(t_{i-1})\}] - \mathbb{E}^2[\mathbb{P}\{X(t_i) \le f(t_i) \mid R_f(t_{i-1})\}].\end{aligned}$$

We estimate the second term $\mathbb{E}^2[\mathbb{P}\{X(t_i) \le f(t_i) \mid R_f(t_{i-1})\}]$ by $\hat{p}_f^2(t_i)$, and we show the first term can be simplified as follows:

$$\begin{aligned}\mathbb{E}[\mathbb{P}^2\{X(t_i) \le f(t_i) \mid R_f(t_{i-1})\}] &= \mathbb{P}\{R_f(t_{i-1}) = 0\}\mathbb{P}^2\{X(t_i) \le f(t_i) \mid R_f(t_{i-1}) = 0\} \\ &\quad + \mathbb{P}\{R_f(t_{i-1}) = 1\}\mathbb{P}^2\{X(t_i) \le f(t_i) \mid R_f(t_{i-1}) = 1\} \\ &= \mathbb{P}\{X(t_{i-1}) \le f(t_{i-1})\}\mathbb{P}^2\{X(t_i) \le f(t_i) \mid X(t_{i-1}) \le f(t_{i-1})\} \\ &\quad + \mathbb{P}\{X(t_{i-1}) > f(t_{i-1})\}\mathbb{P}^2\{X(t_i) \le f(t_i) \mid X(t_{i-1}) > f(t_{i-1})\}.\end{aligned}$$

Let

$$\hat{p}_f(t_i, t_{i-1}^-) = \#\{j : X_j(t_i) \le f(t_i), X_j(t_{i-1}) \le f(t_{i-1})\}/n,$$

and

$$\hat{p}_f(t_i, t_{i-1}^+) = \#\{j : X_j(t_i) \le f(t_i), X_j(t_{i-1}) > f(t_{i-1})\}/n,$$

we then estimate $\mathbb{E}[\mathbb{P}^2\{X(t_i) \le f(t_i) \mid R_f(t_{i-1})\}]$ by

$$\hat{p}_f^2(t_i, t_{i-1}^-)/\hat{p}_f(t_{i-1}) + \hat{p}_f^2(t_i, t_{i-1}^+)/(1 - \hat{p}_f(t_{i-1})),$$

when $p_f(t_{i-1}) \neq 1$. For $\hat{p}(t_{i-1}) = 1$, the second part vanishes. Then, the estimator for $S_f(t_i; t_{i-1})$ is

$$\hat{S}_f(t_i;t_{i-1}) = \begin{cases} 1 & , \quad \hat{p}(t_i) = 1, \\ \dfrac{\hat{p}_f^{\,2}(t_i,t_{i-1}^{-}) \,/\, \hat{p}_f(t_{i-1}) - \hat{p}_f^{\,2}(t_i)}{\hat{p}_f(t_i)\{1-\hat{p}_f(t_i)\}} & , \quad \hat{p}(t_i) \neq 1,\ \hat{p}(t_{i-1}) = 1, \\ \dfrac{\left\{\dfrac{\hat{p}_f^{\,2}(t_i,t_{i-1}^{-})}{\hat{p}_f(t_{i-1})} + \dfrac{\hat{p}_f^{\,2}(t_i,t_{i-1}^{+})}{1-\hat{p}_f(t_{i-1})}\right\} - \hat{p}_f^{\,2}(t_i)}{\hat{p}_f(t_i)\{1-\hat{p}_f(t_i)\}} & , \quad \hat{p}(t_i) \neq 1,\ \hat{p}(t_{i-1}) \neq 1. \end{cases}$$

For the weight function $w(t)$, which is the normalized standard deviation of the random function at time t, we use the sample standard deviation for estimation. At time t_i, the empirical estimator of weight function $w(t_i)$ is given by

$$\hat{w}(t_i) = \frac{\mathrm{sd}(t_i)}{\sum_{r=1}^{m} \mathrm{sd}(t_r)},$$

$$\mathrm{sd}(t_i) = \left[\sum_{j=1}^{n} \left\{ X_j(t_i) - \frac{\sum_{k=1}^{n} X_k(t_i)}{n} \right\}^2 / (n-1)\right]^{1/2}.$$

where Similarly for the weight function $v(t)$ in the shape similarity, which is the normalized difference in function values, the empirical estimator at time t_i is

$$\hat{v}(t_i) = \frac{|f(t_i) - f(t_{i-1})|}{\sum_{r=2}^{m} |f(t_r) - f(t_{r-1})|}.$$

Finally, the estimators of the TVD and SS are given by

$$\mathrm{TVD}(f) = \sum_{i=1}^{m} \hat{w}(t_i)\hat{D}_f(t_i),$$

$$\mathrm{SS}(f) = \sum_{i=2}^{m} \hat{v}(t_i)\hat{S}_f(t_i;t_{i-1}).$$

References

Arribas-Gil, A. and J. Romo (2014). Shape outlier detection and visualization for functional data: The outliergram. *Biostatistics 15* (4), 603–619.

Berrendero, J. R., A. Justel, and M. Svarc (2011). Principal components for multivariate functional data. *Computational Statistics and Data Analysis 55* (9), 2619–2634.

Chakraborty, A. and P. Chaudhuri (2014a). On data depth in infinite dimensional spaces. *Annals of the Institute of Statistical Mathematics 66* (2), 303–324.

Chakraborty, A. and P. Chaudhuri (2014b). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The Annals of Statistics 42* (3), 1203–1231.

Claeskens, G., M. Hubert, L. Slaets, and K. Vakili (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association 109* (505), 411–423.

Cuesta-Albertos, J. A. and A. Nieto-Reyes (2008). The random tukey depth. *Computational Statistics and Data Analysis 52* (11), 4979–4988.

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

Dai, W. and M. G. Genton (2019). Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis 131*, 50–65.

Dang, X. and R. Serfling (2010). Nonparametric depth-based multivariate outlier identifiers, and masking robustness properties. *Journal of Statistical Planning and Inference 140* (1), 198–213.

Dutta, S. and A. K. Ghosh (2012). On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics 64*, 657–676.

Dutta, S., A. K. Ghosh, and P. Chaudhuri (2011). Some intriguing properties of tukey's half-space depth. *Bernoulli 17* (4), 1420–1434.

Febrero, M., P. Galeano, and W. González-Manteiga (2008). Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics 19* (4), 331–345.

Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis - Theory and Practice*. Springer Science & Business Media.

Fraiman, R. and G. Muniz (2001). Trimmed means for functional data. *Test 10* (2), 419–440.

Genton, M. G., C. Johnson, K. Potter, G. Stenchikov, and Y. Sun (2014). Surface boxplots. *Stat 3* (1), 1–11.

Ghosh, A. K. and P. Chaudhuri (2005). On maximum depth and related classifiers. *Scandinavian Journal of Statistics 32* (2), 327–350.

Hubert, M., P. J. Rousseeuw, and P. Segaert (2015). Multivariate functional outlier detection. *Statistical Methods and Applications 24* (2), 177–202.

Hubert, M. and E. Vandervieren (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis 52* (12), 5186–5201.

Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician 50* (2), 120–126.

Hyndman, R. J. and H. L. Shang (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics 19* (1), 29–45.

Ieva, F. and A. M. Paganoni (2013). Depth measures for multivariate functional data. *Communications in Statistics - Theory and Methods 42* (7), 1265–1276.

Jörnsten, R. (2004). Clustering and classification based on the L1 data depth. *Journal of Multivariate Analysis 90* (1), 67–89.

Li, L., W. Huang, I. Y. H. Gu, and Q. Tian (2004). Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing 13* (11), 1459–1472.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics 18* (1), 405–414.

Liu, R. Y., J. M. Parelius, and K. Singh (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Annals of Statistics 27* (3), 783–858.

López-Pintado, S. and J. Romo (2006). Depth-based classification for functional data. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science. Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications 72*, 103–120.

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

López-Pintado, S. and J. Romo (2011). A half-region depth for functional data. *Computational Statistics and Data Analysis 55* (4), 1679–1695.

López-Pintado, S., Y. Sun, J. K. Lin, and M. G. Genton (2014). Simplicial band depth for multivariate functional data. *Advances in Data Analysis and Classification 8* (3), 321–338.

Nagy, S., I. Gijbels, M. Omelka, and D. Hlubinka (2016). Integrated depth for functional data: statistical properties and consistency. *ESAIM: Probability and Statistics 20*, 95–130.

Narisetty, N. N. and V. N. Nair (2016). Extremal depth for functional data and applications. *Journal of the American Statistical Association 111* (516), 1705–1714.

Pigoli, D. and L. M. Sangalli (2012). Wavelets in functional data analysis: Estimation of multidimensional curves and their derivatives. *Computational Statistics and Data Analysis 56* (6), 1482–1498.

Ramsay, J., G. Hooker, and S. Graves (2009). *Functional Data Analysis*. Wiley Online Library.

Rousseeuw, P. J., J. Raymaekers, and M. Hubert (2018). A measure of directional outlyingness with applications to image data and video. *Journal of Computational and Graphical Statistics 27* (2), 345–359.

Rousseeuw, P. J., I. Ruts, and J. W. Tukey (1999). The bagplot: a bivariate boxplot. *The American Statistician 53* (4), 382–387.

Sangalli, L. M., P. Secchi, S. Vantini, and A. Veneziani (2009). A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. *Journal of the American Statistical Association 104* (485), 37–48.

Serfling, R. (2002). A depth function and a scale curve based on spatial quantiles. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L1-Norm and Related Methods*, pp. 25–38. Basel: Birkhäuser Basel.

Sguera, C., P. Galeano, and R. Lillo (2014). Spatial depth-based classification for functional data. *Test 23* (4), 725–750.

Sguera, C., P. Galeano, and R. E. Lillo (2016). Functional outlier detection by a local depth with application to NOx levels. *Stochastic Environmental Research and Risk Assessment 30* (4), 1115–1130.

Sun, Y. and M. G. Genton (2011). Functional boxplots. *Journal of Computational and Graphical Statistics 20* (2), 316–334.

Sun, Y. and M. G. Genton (2012). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics 23* (1), 54–64.

Sun, Y., M. G. Genton, and D. W. Nychka (2012). Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? *Stat 1* (1), 68–74.

Tukey, J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians 2*, 523–531.

Zuo, Y. and R. Serfling (2000). General notions of statistical depth function. *Annals of Statistics 28* (2), 461–482.

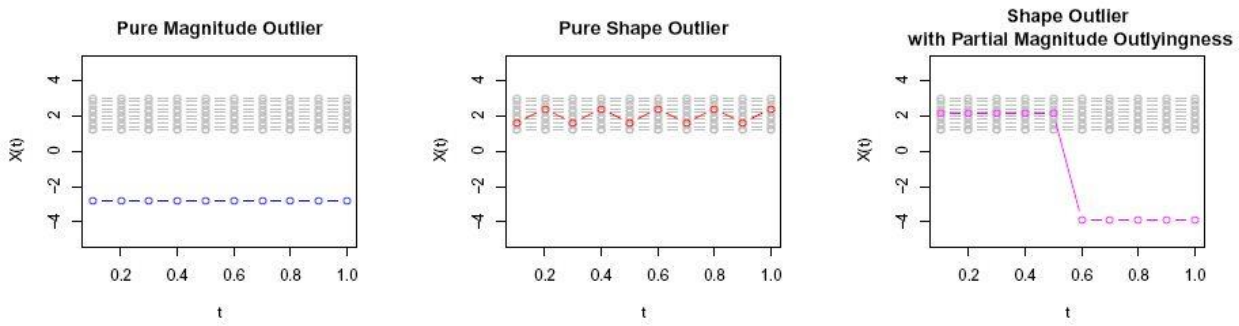

Fig. 1 Toy examples for the pure magnitude outlier (blue), the pure shape outlier (red), and the shape outlier with partial magnitude outlyingness (magenta). Grey curves are the non-outliers in each case.

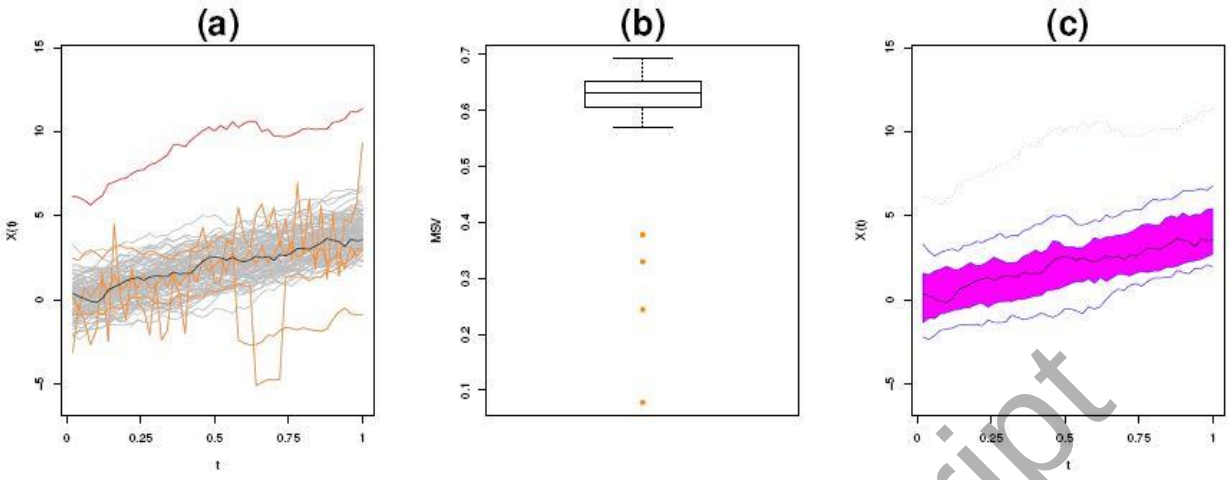

Fig. 2 Visualization tools: (a) Observations with a detected magnitude outlier (red) and shape outliers (orange). (b) Shape outlyingness plot with points in orange corresponding to shape outliers. (c) Functional boxplot using the total variation depth after removing detected shape outliers.

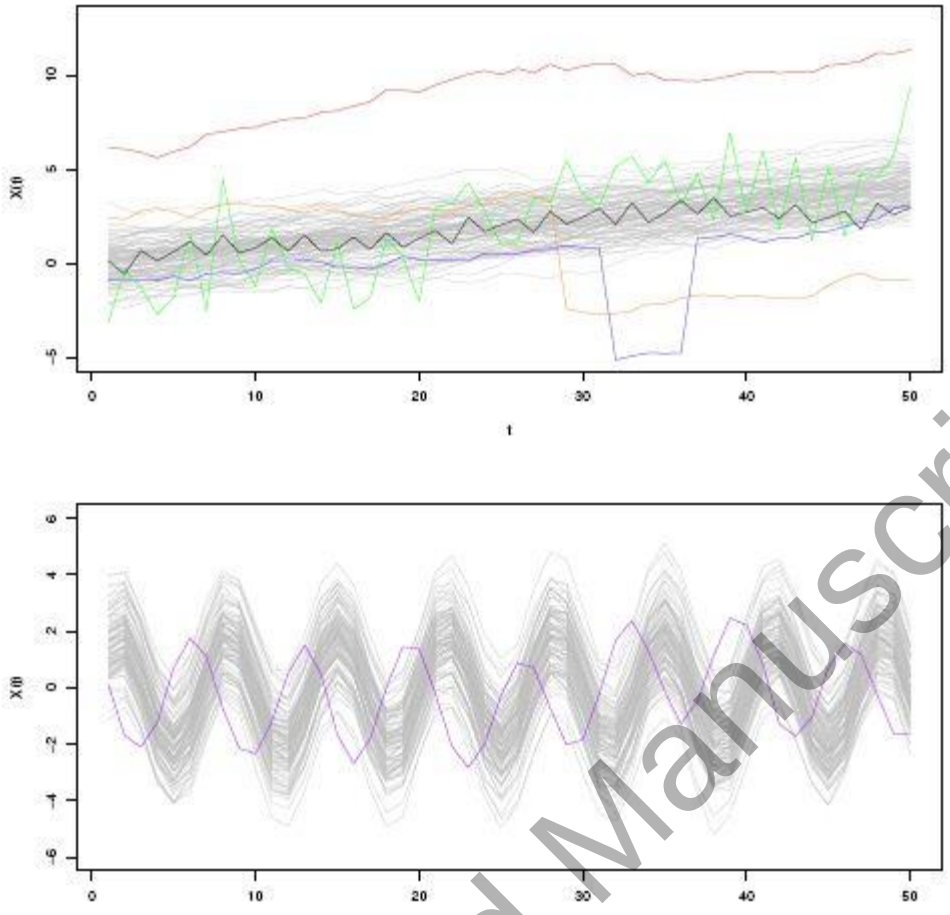

Fig. 3 Top panel: 100 curves (grey) from Model 1 with one of different outliers from Model 2 (red), Model 3 (orange), Model 4 (blue), Model 5 (green) and Model 6 (black), respectively. Bottom panel: 100 curves (grey) contaminated by one outlier from Model 7 (purple).

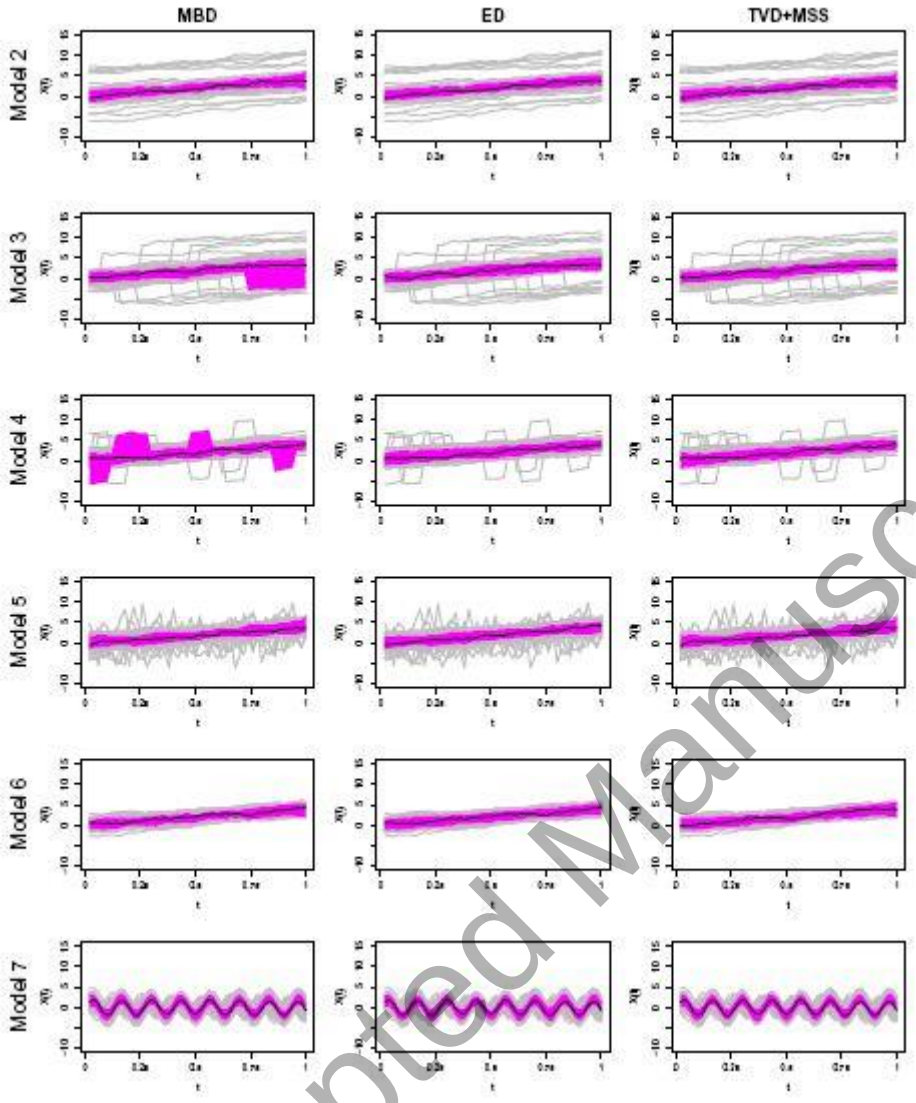

Fig. 4 The central regions constructed by MBD (left), ED (middle), and TVD+MSS (right) for Models 2 - 7. The central region is displayed by the magenta polygon. The solid black line in the middle is the median curve, which has the largest depth value. All the grey lines in the background are the dataset curves.

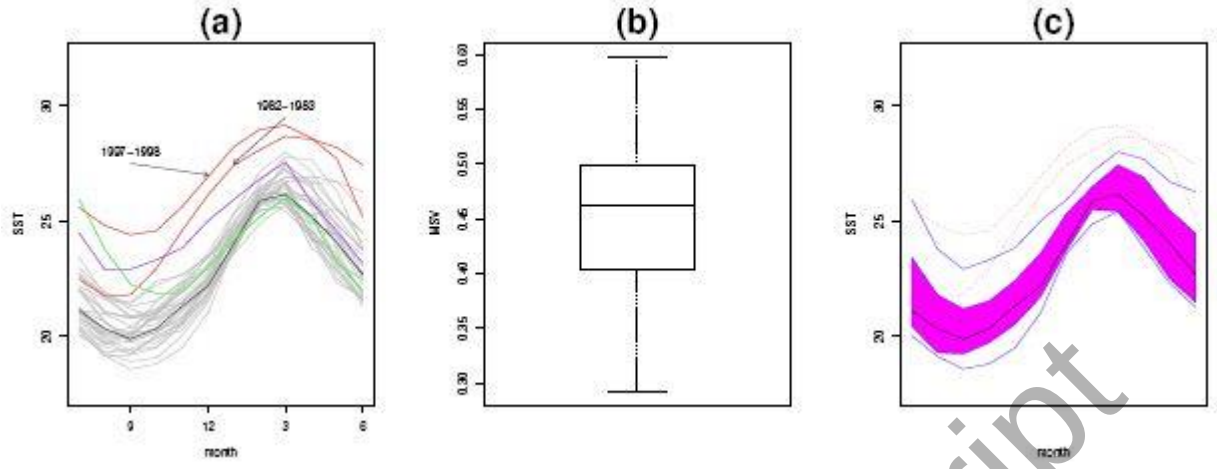

Fig. 5 The sea surface temperatures in a Niño zone from July 1982 to June 2016. (a) The original data. (b) The boxplot of the modified shape similarity. (c) The functional boxplot with the total variation depth.

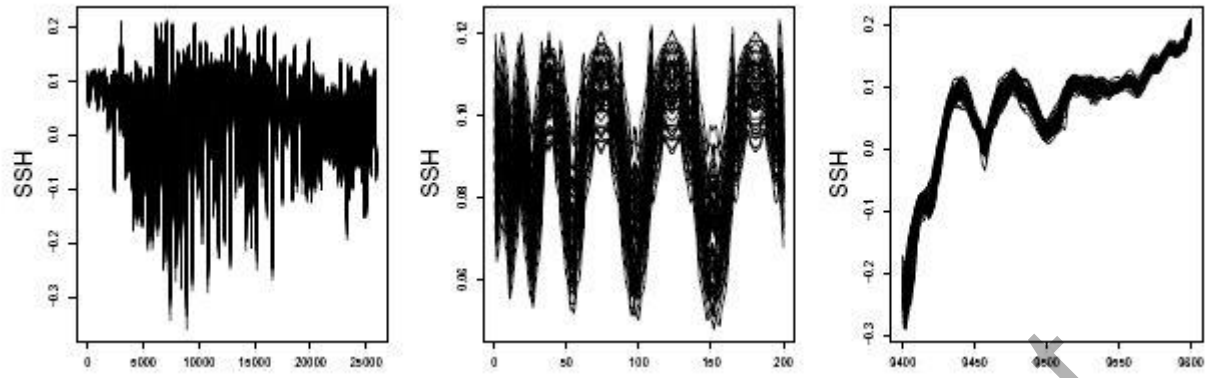

Fig. 6 The 50 functional curves that are transformed from the sea surface height images (left), and the curves that are zoomed in different areas (middle and right).

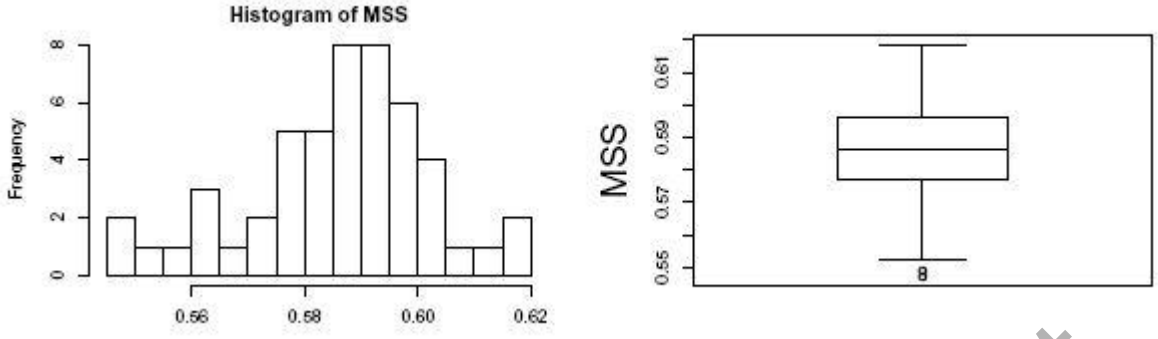

Fig. 7 The histogram (left) and boxplot (right) of MSS associated with the 500 functional curves for the sea surface height.

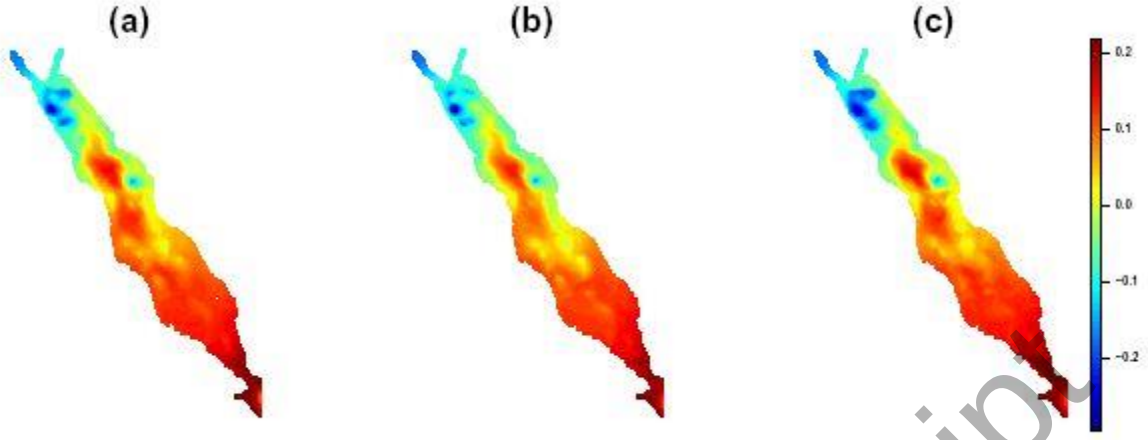

Fig. 8 Model runs of the sea surface height of the Red Sea on January 1, 2016. (a) The median. (b) One of the shape outliers. (c) One of the magnitude outliers.

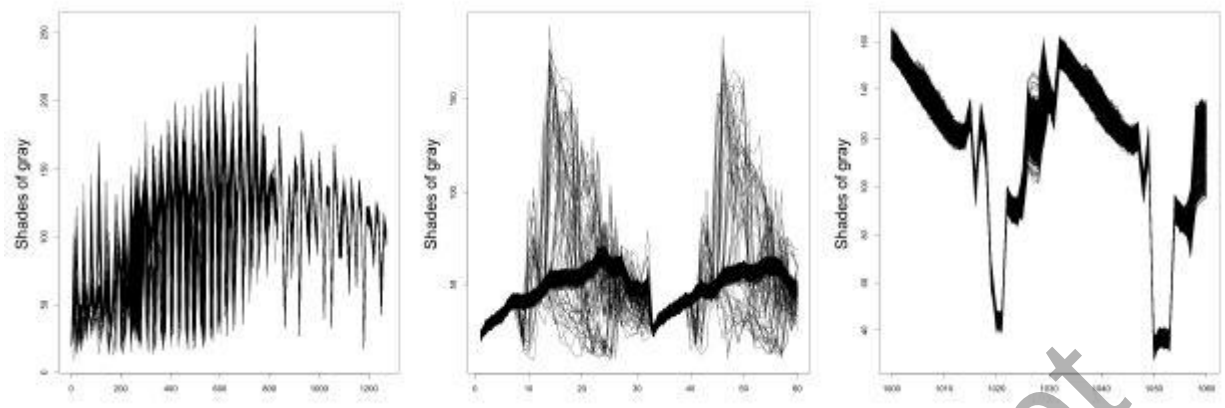

Fig. 9 The 2,964 functional curves that are transformed from the video frames (left), and the curves that are zoomed in different areas (middle and right).

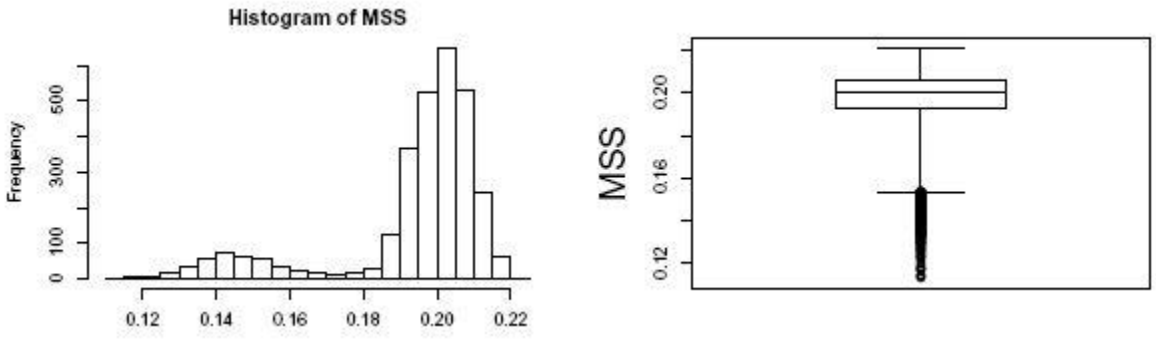

Fig. 10 The histogram (left) and boxplot (right) of MSS associated with the 2,964 functional curves for the video frames.

Fig. 11 The video frames. (a) The median. (b) One of the shape outliers. (c) One of the magnitude outliers.

Table 1 Results of outlier detection using different methods for different models. TPR is true positive rate, and FPR is false positive rate. The values are the means shown as percentages in the 500 experiments for each case, and the values in parentheses are the corresponding standard deviations.

Method	Model 1	Model 2		Model 3		Model 4	
	FPR	TPR	FPR	TPR	FPR	TPR	FPR
TVD+MSS	0.07(0.27)	98.88(3.39)	0.04(0.22)	**98.08(4.68)**	0.25(0.55)	**100.00(0.00)**	0.04(0.22)
MBD	0.07(0.27)	98.87(3.35)	0.04(0.21)	82.19(14.87)	0.05(0.25)	47.86(20.01)	0.06(0.27)
ED	0.03(0.18)	98.50(4.16)	0.02(0.18)	**86.15(13.73)**	0.02(0.14)	**86.62(12.27)**	0.02(0.16)
OG+MBD	5.38(2.40)	98.87(3.35)	4.90(2.50)	**89.09(11.69)**	3.30(2.07)	**75.85(18.14)**	3.22(2.05)
FOM(AO)	0.38(0.75)	96.87(10.57)	0.06(0.26)	69.36(22.89)	0.08(0.36)	24.09(24.31)	0.10(0.37)
FOM(DO)	0.84(1.12)	99.83(1.34)	0.19(0.54)	**79.71(18.11)**	0.25(0.64)	43.66(29.73)	0.24(0.58)
MSBD1	0.06(0.25)	98.14(4.72)	0.02(0.13)	77.39(17.10)	0.02(0.15)	45.79(20.48)	0.04(0.20)
MSBD2	0.04(0.21)	97.02(5.87)	0.01(0.08)	**84.23(13.72)**	0.01(0.10)	**80.18(14.79)**	0.01(0.11)
MBD+MBD	0.07(0.27)	98.87(3.35)	0.04(0.21)	**92.15(9.74)**	0.05(0.25)	**81.10(14.23)**	0.06(0.27)
MSS-only	0.00(0.05)	0.00(0.00)	0.00(0.00)	97.79(5.23)	0.00(0.00)	100.00(0.00)	0.00(0.00)

Method	Model 5	Model 6	Model 7

	TPR	FPR	TPR	FPR	TPR	FPR
TVD+MSS	**100.00(0.00)**	0.03(0.19)	**99.84(1.99)**	0.03(0.19)	**100.00(0.00)**	0.03(0.19)
MBD	84.27(14.14)	0.03(0.20)	0.06(0.74)	0.04(0.22)	44.18(24.34)	0.03(0.18)
ED	82.10(14.02)	0.01(0.12)	0.03(0.55)	0.01(0.12)	34.53(22.36)	0.01(0.11)
OG+MBD	**99.99(0.26)**	2.13(1.73)	6.74(8.61)	4.82(2.23)	**99.89(2.53)**	2.16(1.77)
FOM(AO)	22.22(21.57)	0.07(0.29)	0.36(2.09)	0.32(0.65)	1.59(7.45)	0.08(0.32)
FOM(DO)	36.05(25.43)	0.19(0.52)	0.91(3.21)	0.86(1.12)	9.68(21.08)	0.22(0.50)
MSBD1	81.92(14.62)	0.04(0.22)	0.08(0.87)	0.03(0.18)	39.87(23.24)	0.03(0.17)
MSBD2	80.8(15.07)	0.03(0.20)	0.01(0.32)	0.02(0.15)	39.08(23.03)	0.02(0.15)
MBD+MBD	**100.00(0.00)**	0.03(0.20)	7.04(9.91)	0.04(0.22)	**100.00(0.00)**	0.03(0.18)
MSS-only	100.00(0.00)	0.00(0.00)	99.84(1.99)	0.00(0.00)	100.00(0.00)	0.00(0.00)

Table 2 Running time of outlier detection using different methods for different models. The mean running time of the 500 experiments is shown, and the values in parentheses are the corresponding standard deviations.

Method	Running Time (seconds)	Method	Running Time (seconds)
TVD+MSS	0.089(0.010)	FOM(DO)	0.067(0.012)
MBD	0.005(0.002)	MSBD1	0.692(0.037)
ED	0.045(0.008)	MSBD2	0.694(0.038)
OG+MBD	1.157(0.062)	MBD+MBD	0.010(0.002)
FOM(AO)	0.069(0.008)	MSS-only	0.087(0.008)