Technometrics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

PERCEPT: A New Online Change-Point Detection Method using Topological Data Analysis

Xiaojun Zheng, Simon Mak, Liyan Xie & Yao Xie

To cite this article: Xiaojun Zheng, Simon Mak, Liyan Xie & Yao Xie (2023) PERCEPT: A New Online Change-Point Detection Method using Topological Data Analysis, Technometrics, 65:2, 162-178, DOI: 10.1080/00401706.2022.2124312

To link to this article: https://doi.org/10.1080/00401706.2022.2124312

View supplementary material

Published online: 31 Oct 2022.

Submit your article to this journal

Article views: 304

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=utch20

TECHNOMETRICS
2023, VOL. 65, NO. 2, 162–178
https://doi.org/10.1080/00401706.2022.2124312

Taylor & Francis
Taylor & Francis Group

Check for updates

PERCEPT: A New Online Change-Point Detection Method using Topological Data Analysis

Xiaojun Zheng[a], Simon Mak[a], Liyan Xie[b], and Yao Xie[c]

[a]Department of Statistical Science, Duke University, Durham, NC; [b]School of Data Science, The Chinese University of Hong Kong, Shenzhen, China; [c]H. Milton Stewart School of Industrial and Systems Engineering (ISyE), Georgia Institute of Technology, Atlanta, GA

ABSTRACT

Topological data analysis (TDA) provides a set of data analysis tools for extracting embedded topological structures from complex high-dimensional datasets. In recent years, TDA has been a rapidly growing field which has found success in a wide range of applications, including signal processing, neuroscience and network analysis. In these applications, the online detection of changes is of crucial importance, but this can be highly challenging since such changes often occur in low-dimensional embeddings within high-dimensional data streams. We thus propose a new method, called PERsistence diagram-based ChangE-PoinT detection (PERCEPT), which leverages the learned topological structure from TDA to sequentially detect changes. PERCEPT follows two key steps: it first learns the embedded topology as a point cloud via persistence diagrams, then applies a nonparametric monitoring approach for detecting changes in the resulting point cloud distributions. This yields a nonparametric, topology-aware framework which can efficiently detect online geometric changes. We investigate the effectiveness of PERCEPT over existing methods in a suite of numerical experiments where the data streams have an embedded topological structure. We then demonstrate the usefulness of PERCEPT in two applications on solar flare monitoring and human gesture detection.

ARTICLE HISTORY
Received February 2022
Accepted August 2022

KEYWORDS
Change-point detection; Human gesture detection; Online monitoring; Persistent homology; Solar flare monitoring; Topological data analysis

1. Introduction

Topological Data Analysis (TDA) is a thriving field at the intersection of statistics, machine learning, and algebraic topology, which has gained traction in recent years. TDA methods provide a set of tools for studying the shapes of complex high-dimensional datasets, by extracting their underlying low-dimensional geometric structures. This field of study has found success in a wide range of applications, including signal processing (Perea and Harer 2015), computational biology (Cang et al. 2015), time series analysis (Seversky et al. 2016), and neuroscience (Sizemore et al. 2019).

Despite promising developments in recent years, there has been little work on integrating topological structure for sequential change-point detection, which is a fundamental problem in many of the aforementioned applications. Change-point detection here refers to the detection of a possible change in signal distribution over time. Traditional change-point detection methods largely focus on likelihood ratio tests, which presume that observations are independently and identically distributed, both before and after the change (Siegmund 1985; Poor and Hadjiliadis 2008; Basseville and Nikiforov 1993; Tartakovsky et al. 2015). When the pre- and post-change distributions are known, one can apply the cumulative sum (CUSUM) detection rule, which has been proved to be optimal (Lorden 1971; Moustakides 1986). In practice, the post-change distribution is typically unknown, in which case one can sequentially estimate it and construct a generalized likelihood ratio test (Lai 1998).

However, a key limitation with these traditional methods is that they may perform poorly for *high-dimensional* data with *complex* low-dimensional structure. This is because the distributions in such setting would typically be difficult to approximate well using parametric families, and the required density ratio would thus be difficult to estimate as well. One solution in this high-dimensional setting is to first extract the underlying low-dimensional structure from data, and then construct detection statistics based on the extracted information. CUSUM-type detection algorithms were proposed in Xie et al. (2020) and Jiao et al. (2018) to detect changes characterized by an unknown subspace structure in the covariance matrix. These methods work well when the true low-dimensional embedding is precisely the embedding used in the detection procedure, that is, a linear subspace. In complex problems where the true embedding is nonlinear, however, this model misspecification may result in considerable deterioration in detection performance (Molloy and Ford 2017). There is thus a need for a high-dimensional change detection method which can integrate a broader yet realistic framework for modeling low-dimensional structure.

In recent years, the rise of TDA methods suggests that, for many problems, the underlying data may have embedded *topological* structure. Indeed, in such problems, the extracted topology from TDA often captures intuitive features from

CONTACT Simon Mak *sm769@duke.edu* Department of Statistical Science, Duke University, Durham, NC.
© 2022 American Statistical Association and the American Society for Quality

high-dimensional data which are interpretable for the practitioner. For example, in computer vision, topological features can represent segmented regions of a 3D shape (Beksi and Papanikolopoulos 2016). Similarly, for time series, periodic signals can be captured by topological features via Taken's Embedding (Seversky et al. 2016). However, the integration of such structure for change detection has largely remained unexplored. A recent work (Islambekov et al. 2019) proposed a TDA detection approach for time series data, by converting this to a sequence of so-called Betti numbers (Edelsbrunner and Harer 2008), which capture the number of k-dimensional holes on a topological surface. However, Betti numbers are known to provide a weak summary of topology (Ghrist 2008) and thus may be insensitive to certain topological changes. Another recent work (Ofori-Boateng et al. 2021) used a richer topological summary called a *persistence diagram* (Edelsbrunner and Harer 2008), which uniquely captures the topological features of the data at different spatial resolutions (more on this in Section 2.1). With persistence diagrams computed at each point in time, the method then makes use of the Wasserstein distance between diagrams from adjacent times as the detection test statistic. However, there are two limitations with this approach: (i) it relies on the Wasserstein distance, which may not fully capture the topological differences between two persistence diagrams (more on this later), and (ii) its test statistic relies on information from only the immediate past, which can greatly reduce detection efficiency.

To address this, we propose a new *nonparametric*, *topology-aware* framework called PERsistence diagram-based ChangE-PoinT detection (PERCEPT). As in Ofori-Boateng et al. (2021), PERCEPT leverages the extracted persistence diagrams (which capture topological features of the data at each time) for change detection. However, instead of using the Wasserstein distance of diagrams from adjacent times, PERCEPT extends a recent nonparametric change detection method (Xie and Xie 2021) to detect changes directly on the diagram point clouds. This yields two advantages: (i) it offers a distribution-based approach which amplifies changes in topological features; and (ii) its test statistic makes use of data within a past time window, which addresses information loss. We demonstrate the effectiveness of PERCEPT over existing methods in a suite of simulation experiments and in applications to solar flare monitoring and human gesture detection.

The rest of the article is organized as follows. Section 2 provides preliminaries on persistent homology and motivations. Section 3 outlines the PERCEPT methodology. Section 4 demonstrates the effectiveness of this method in numerical experiments. Section 5 applies the method to solar flare monitoring and human gesture detection.

2. Preliminaries and Motivation

We first provide a brief overview of TDA, then discuss two baseline methods, the Hotelling's T^2 statistic and the Wasserstein distance approach in Ofori-Boateng et al. (2021), for high-dimensional change-point detection. We then motivate the proposed PERCEPT method via our motivating solar flare detection application.

2.1. TDA Preliminaries

A primary tool in TDA is *persistent homology*, which extracts topological features (e.g., connected components, holes, and their higher-dimensional analogues) from point cloud data. In what follows, we provide a brief overview of persistent homology, following Ghrist (2008) and Edelsbrunner and Harer (2008). For a given point cloud dataset, persistent homology provides a representation of this as a *simplicial complex*, defined as a set of vertices, edges, triangles, and their higher-dimensional counterparts. A common simplicial complex is the *Rips complex*, which depends on a single scale parameter ϵ. At a given radius $\epsilon > 0$, the Rips complex contains all edges between any two points whose distance is at most ϵ, and contains triangular faces for any three points whose pairwise distances are at most ϵ. Figure 1 (adapted from Han et al. 2018) illustrates this for a toy dataset. Clearly, a single ϵ cannot capture all geometric structures embedded in the data. Thus, a sequence of scale parameters is used to build a sequence (or *filtration*) of simplicial complexes. This filtration provides a means for extracting topological structure from the data: zero-dimensional holes (or connected components), one-dimensional holes, and their higher-dimensional analogues.

Under this framework, suppose a topological feature appears in the filtration at some radius ϵ and disappears at a larger radius $\epsilon' > \epsilon$. The pair (ϵ, ϵ') gives the *persistence* of the feature, with ϵ and ϵ' being its *birth* and *death*, respectively. A prominent topological feature in the point cloud data would have long persistence, whereas a small or noisy topological feature would have short persistence. The persistence information from all topological features can be captured by an (untilted) *persistence diagram* (PD), a collection of points in $\mathbb{R}^2$ where each feature is represented by a point (ϵ, ϵ'), with ϵ is its birth time and ϵ' is its death time. Figure 1 illustrates the persistent homology pipeline from point cloud data to a persistence diagram. We will distinguish this untilted PD from its tilted variant later.

We use a simple example to illustrate this pipeline from topological structures to a PD. Figure 2 (left) shows the point cloud data generated from two disjoint circles with radii 1, and Figure 2 (right) shows the corresponding PD of this data. For the zero-dimensional homology, which captures connected components (black points in Figure 2 right), we observe many points with small persistence, but one point which persists for a long time. This last point with large persistence suggests the data has two connected components, which is indeed true. For the one-dimensional homology, which captures one-dimensional holes (red points in Figure 2 right), we observe two red points with large persistence, which reflects the two holes within the circles. The choice of zero-dimensional or one-dimensional homology for change detection will depend on prior knowledge of the data, and will be discussed later in numerical experiments.

From a dimension reduction perspective, one might wonder how the low-dimensional (namely, two-dimensional) nature of persistence diagrams can serve as an effective feature extraction tool. One reason is that, when such topology is present in data, the embedded geometric structure is typically low-dimensional enough to be captured on the two-dimensional persistence diagrams, even when the data is high-dimensional.

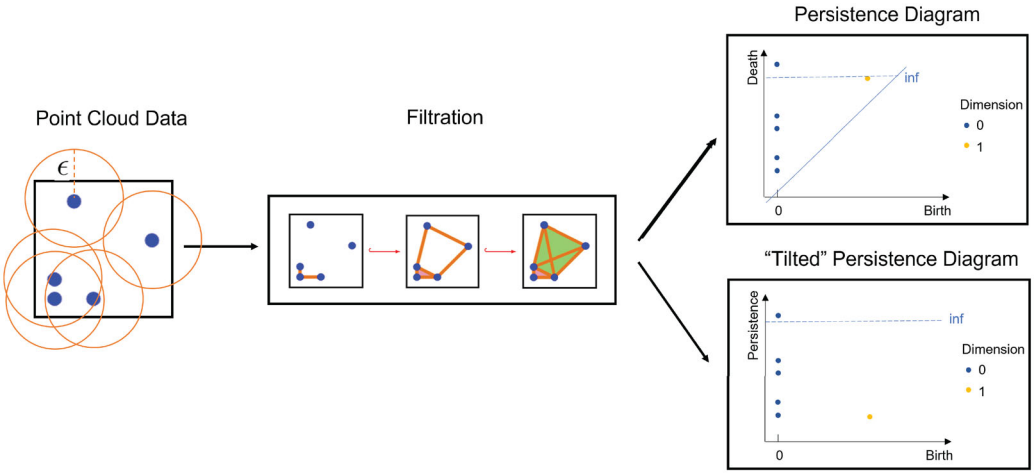

Figure 1. An illustration of the persistent homology pipeline, from point cloud data to a filtration of simplicial complexes to a persistence diagram. The Rips complex with radius ϵ in the left plot corresponds to the second simplicial complex in the filtration.

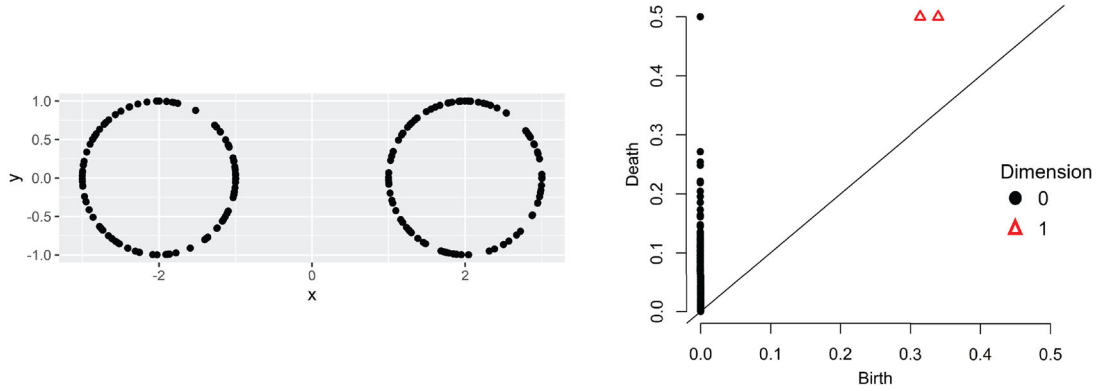

Figure 2. (Left) Point cloud data generated from two disjoint circles, and (right) its corresponding persistence diagram for zero-dimensional and one-dimensional holes.

This *targeted* extraction of *topological* features thus allows for a reduced dimension of the feature space over other dimension reduction approaches. See Chazal (2016) for further discussion, with high-dimensional applications in signal processing (Perea and Harer 2015) and neuroscience (Sizemore et al. 2019).

One potential limitation of PDs for change detection, however, is its inability to distinguish translation shifts in data. One can show that for any point cloud data, its PD representation remains the same after any translation shifts. This is not necessarily a drawback for persistent homology, which aims to extract *geometric* structure rather than translation shifts. For problems both geometric changes and translation shifts are important to detect, we propose later a simple modification which addresses this in Section 3.2.

2.2. Existing Baseline Methods

A classic baseline change detection approach for multivariate data is the Hotelling's T^2 statistic (Hotelling 1947), which makes use of the mean and covariance structure of data. Given data $\boldsymbol{x}_t \in \mathbb{R}^p$ at time t, $t = 1, \ldots, T$, the Hotelling's T^2 statistic is defined as

$$(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\intercal \Sigma_0^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0), \tag{1}$$

where $\bar{\boldsymbol{x}}$ is the sample mean vector, $\boldsymbol{\mu}_0$ and Σ_0 is the nominal mean vector and covariance matrix, respectively (this is typically given or estimated from reference data). The vanilla Hotelling's T^2 statistic is calculated using *only* data at the current time t, with all historical data discarded. To compute the test statistic more efficiently, it can be coupled with the CUSUM scheme (Page 1954), which makes use of a cumulative sum of the statistic over time. The resulting detection statistic S_t^H is then given by:

$$S_t^H = (S_{t-1}^H)^+ + (\bar{\boldsymbol{\mu}}_{t-w,t} - \hat{\boldsymbol{\mu}}_0)^\intercal \widehat{\Sigma}_0^{-1} (\bar{\boldsymbol{\mu}}_{t-w,t} - \hat{\boldsymbol{\mu}}_0) - d^H, \quad S_0^H = 0, \tag{2}$$

where $(x)^+ = \max(x, 0)$, $\bar{\boldsymbol{\mu}}_{t-w,t}$ denotes the sample average of the data vectors $\{\boldsymbol{x}_{t-w}, \ldots, \boldsymbol{x}_t\}$, and $\hat{\boldsymbol{\mu}}_0$ and $\widehat{\Sigma}_0$ are the *pre-change* mean vector and covariance matrix estimated from historical data. Here, d^H is a drift parameter that can also be estimated using historical data. When the data is known to be concentrated on a linear subspace, one can adapt the Hotelling's T^2 test by first performing Principal Component Analysis (PCA), then using the resulting principal components as data within Equation (2). With this, a change-point is then declared at time t when the statistic S_t^H first exceeds a prespecified threshold b.

The second baseline method, the Wasserstein distance approach in Ofori-Boateng et al. (2021), integrates topology in the following way. The Wasserstein distance (of order 1) between two distributions P and Q on sample space Ω is defined as

$$\mathcal{W}_1(P, Q) := \min_{\gamma \in \Gamma(P,Q)} \left\{ \mathbb{E}_{(\omega,\omega') \sim \gamma} \left[c(\omega, \omega') \right] \right\}.$$

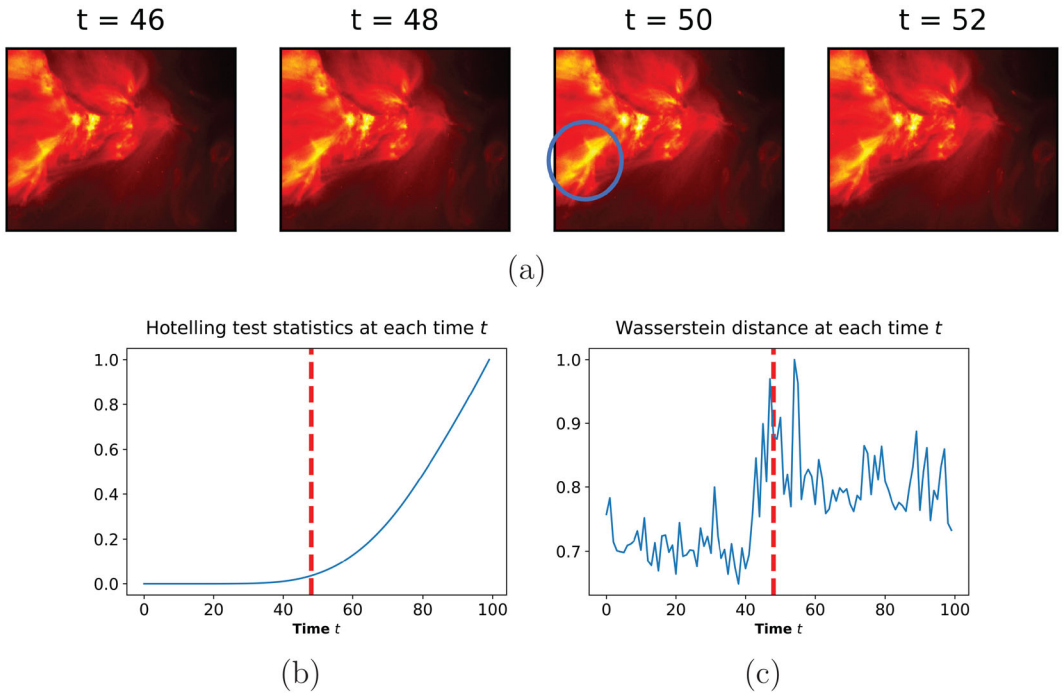

Figure 3. (a) Snapshots of the solar flare around the true change-point $t^* = 49$. (b,c) Hotelling's T^2 and Wasserstein distance statistics at each time t; the true change-point t^* is marked in red.

Here, $c(\cdot,\cdot)$ is a metric on Ω (we use the Euclidean norm), and $\Gamma(P,Q)$ denotes the collection of all (Borel) probability measures on $\Omega \times \Omega$ with marginal distributions P and Q. With this, the Wasserstein distance method is straight-forward. First, at each time t, the persistence diagram $\mathcal{D}_t \in \mathbb{R}^2$ is computed from data $\boldsymbol{x}_t$. Next, the Wasserstein distances $S_t^W = \mathcal{W}_1(\mathcal{D}_t, \mathcal{D}_{t+1})$ are computed between the PDs from adjacent time frames. One then declares a change-point at time t when the statistic S_t^W exceeds a pre-specified threshold b.

We investigate these two baseline methods for a motivating solar flare monitoring application (more on this in Section 5). The data consists of image snapshots (232×292 pixels) of a solar flare, captured by the Solar Dynamics Observatory[1] at NASA. Figure 3(a) shows several snapshots before and after the true change-point at $t^* = 49$, where the flare bursts become more pronounced and bright. However, such a change can be quite subtle visually, thus, making monitoring a challenging task (Xie et al. 2012). We applied the above two baseline methods, with the Hotelling's T^2 applied by first vectorizing the image data, then applying PCA to extract the top 15 principal components. Figure 3(b) and (c) show the detection statistics as a function of time t. We see that both methods experience a large detection delay from the true change-point at $t^* = 49$. For the Hotelling's T^2, the test statistic increases slowly after the change-point, which suggests it does not capture well the changed image features. Similarly, the Wasserstein distance statistic appears highly unstable and unable to detect the change-point. Given the limitations of the two baseline methods, we introduce next a new nonparametric, topology-aware method which provides a new framework for integrating low-dimensional geometric information for online monitoring.

[1] See *https://sdo.gsfc.nasa.gov/assets/docs/SDO_Guide.pdf*.

3. Persistence Diagram-Based Change Detection

We now introduce the proposed PERsistence diagram-based ChangE-PoinT (PERCEPT) monitoring method. We first describe the histogram binning for PDs, then show how the extracted histograms can be used for change detection. We then present useful insights for PERCEPT, and discuss methodological developments for bin and weight optimization. Figure 4 visualizes the workflow for PERCEPT; we will describe each step in detail below.

3.1. Persistence Histogram Binning

The first step in PERCEPT is to construct the so-called *persistence histograms*, a novel histogram representation which captures persistence information from the computed PDs. This histogram binning serves two purposes: it provides a robust way for filtering noise in the PD data, and allows us to leverage recent developments in histogram-based change detection methods within PERCEPT (shown next). To facilitate this, we assume all PDs are given in their *tilted* representation, where a feature is represented by a point $(\epsilon, \epsilon' - \epsilon)$, with ϵ is its birth time and $\epsilon' - \epsilon$ is its persistence time. Figure 1 provides an illustration.

Assume, as before, that the PDs $\mathcal{D}_t \in \mathbb{R}^2$ are computed for the data $\boldsymbol{x}_t$ at each time $t = 1, \ldots, T$. Further assume that the birth range for the PDs (i.e., the x-axis on $\mathcal{D}_t$) is partitioned into the L bins: $[0, b_1), [b_1, b_2), \ldots, [b_{L-1}, b_L)$, where b_l is the right break-point for the lth bin. With this, we can now bin the point cloud data $\mathcal{D}_t$. Let $f_{t,l}$ be the sum of persistences for points in $\mathcal{D}_t$ within the lth bin, that is:

$$f_{t,l} = \sum_{(u,v)\in\mathcal{D}_t, u\in[b_{l-1},b_l)} v, \tag{3}$$

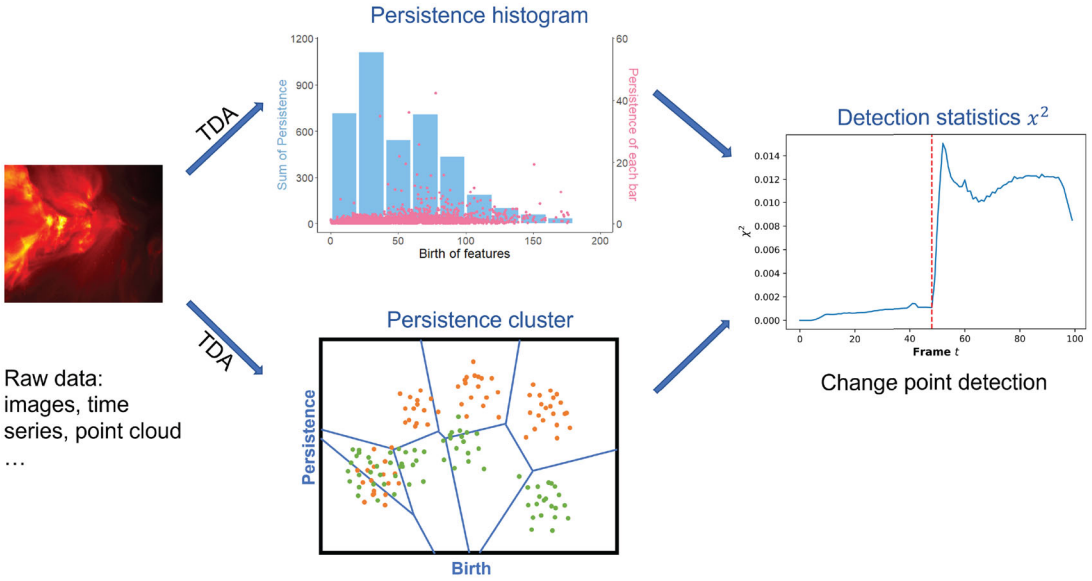

Figure 4. Visualizing the general workflow for PERCEPT. The data can take a variety of forms, including images, time series or point cloud data. Persistence homology is used to extract useful geometric features from data to persistence diagrams (Section 2.1). These diagrams are then binned using either persistence histograms or persistence clusters (Sections 3.1 and 3.4, respectively), then used within an ℓ_2 test statistic for monitoring (Section 3.2).

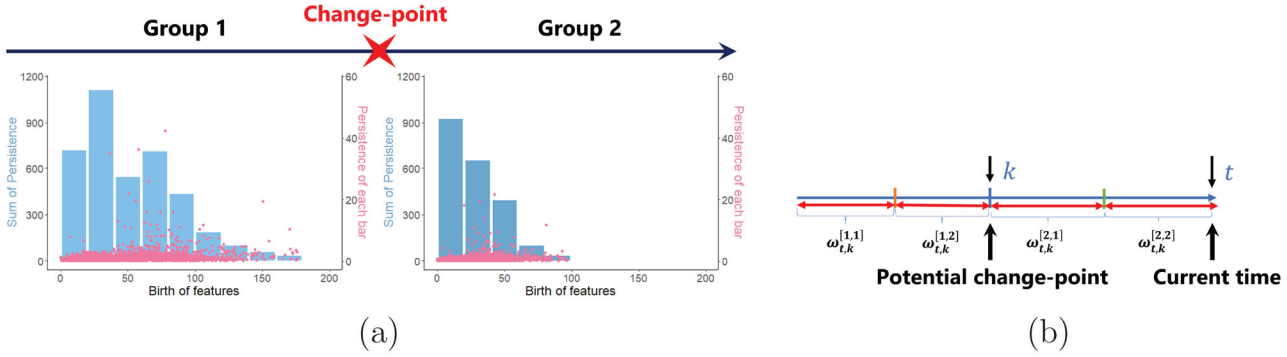

Figure 5. (a) Persistence histograms for a pre-change and post-change solar flare image. (b) Visualizing the intervals used for the weighted ℓ_2 detection statistic.

where $v = \epsilon' - \epsilon$ is the persistence time of point (u, v), and let $\omega_{t,l} = f_{t,l} / \sum_{l'=1}^{L} f_{t,l'}$ be its corresponding proportion. The histogram frequencies and distribution of the PD $\mathcal{D}_t$ can thus be represented by the vectors $\boldsymbol{f}_t = (f_{t,1}, \ldots, f_{t,L})$ and $\boldsymbol{\omega}_t = (\omega_{t,1}, \ldots, \omega_{t,L})$, respectively. We will call $\boldsymbol{\omega}_t$ the *persistence histogram* of the PD $\mathcal{D}_t$. Figure 5(a) visualizes this binning procedure. We will discuss the choice of L in Section 3.5.

In the absence of prior information on pre- and post-change data, the persistence histogram breakpoints $b_1, \ldots, b_L$ can be placed in such a way that the resulting histogram bins have equal widths. The same bins are then used throughout the monitoring procedure. When there is prior information on only the pre-change data, the breakpoints $b_1, \ldots, b_L$ can be chosen such that there is (roughly) an equal sum of persistences within each histogram bin. When prior information are available on both pre- or post-change data, one can employ more elaborate binning approaches; see Section 3.4 for further discussion.

As an illustration, consider again the solar flare images from Figure 3. Using the lower star filtration (Bendich et al. (2011), discussed later in Section 5.1), we can obtain persistence diagrams for each flare image. Figure 5(a) shows the binned persistence histogram (with $L = 10$ bins) of a PD computed from a pre-change flare image, and Figure 5(b) shows the persistence histogram for a "post-change" image using the same bins. Given a significant change in topological structure, the corresponding pre- and post-change persistence histograms (which capture topology information) should be sufficiently different to capture this change. We will leverage this next to formulate our test statistic.

3.2. Test Statistic

The second step in PERCEPT is to use the extracted persistence histograms to construct a monitoring test statistic. The idea is as follows. Suppose the bins $[0, b_1), [b_1, b_2), \ldots, [b_{L-1}, b_L)$ are fixed. Then, at each time t, one can treat the observed persistence histograms $\boldsymbol{f}_t$ as *data* sampled from an underlying discrete distribution with L levels. Let p_{pre} denote this so-called *persistence distribution* (with L levels) prior to the change, and p_{post} be the persistence distribution after the change. The goal of detecting topological changes can then be thought of as testing for differences between the persistence distributions p_{pre} and p_{post}. We thus desire a test which investigates the following hypotheses:

$$H_0 : p_{\text{pre}} = p_{\text{post}}, \quad H_A : p_{\text{pre}} \neq p_{\text{post}}. \tag{4}$$

To test (4), we extend a recent nonparametric test in Xie and Xie (2021) for detecting changes on discrete distributions. At the current time t, we search for all possible change points at time $k < t$ within a fixed-sized window. To investigate whether time k is a change point, we will consider four consecutive time intervals (see Figure 5(b)): the first two intervals are immediately before time k, and the last two intervals are immediately after k. All four intervals have the same time length of $\lfloor (t-k)/2 \rfloor$. We call the first group of intervals (representing potential *pre*-change times) as "Group 1," and the latter group (representing potential *post*-change times) as "Group 2."

Suppose the histogram breakpoints $b_1, \ldots, b_L$ are given. Consider now the persistence histogram proportions within the two time intervals in Group 1 and the two intervals in Group 2; we denote this as $\boldsymbol{\omega}_{t,k}^{[1,1]}$, $\boldsymbol{\omega}_{t,k}^{[1,2]}$, $\boldsymbol{\omega}_{t,k}^{[2,1]}$, and $\boldsymbol{\omega}_{t,k}^{[2,2]}$, respectively. This nonparametric weighted-ℓ_2 statistic can then be defined as

$$\chi_{t,k} = (\boldsymbol{\omega}_{t,k}^{[1,1]} - \boldsymbol{\omega}_{t,k}^{[2,1]})^{\intercal} \Sigma (\boldsymbol{\omega}_{t,k}^{[1,2]} - \boldsymbol{\omega}_{t,k}^{[2,2]}). \tag{5}$$

Here, $\Sigma = \text{diag}\{\sigma_1 \ldots, \sigma_L\}$ is the diagonal weight matrix, where $\sigma_l \geq 0, l = 1, \ldots, L$. Note that, if time k were indeed a change point, the resulting test statistic $\chi_{t,k}$ would likely be large, thus, providing evidence for a change. In the absence of prior information on pre- and post-change data, an intuitive choice of Σ is the identity matrix, which assumes equal weights over persistence histogram bins. However, when some prior data are available on either the pre- or post-change PDs, the weights in Σ can be optimized to better highlight distributional differences. Section 3.5 provides further discussion on this.

Using the statistic (5), an online detection procedure is then given by the stopping time:

$$\mathcal{T} = \inf\{t : \chi_t^{\max} \geq b\}, \quad \chi_t^{\max} = \max_{0 \leq k \leq t} \chi_{k,t}, \tag{6}$$

where b is a pre-specified threshold parameter. Here, $\mathcal{T}$ is the time at which the procedure raises an alarm indicating a change-point has occurred before time t. In other words, when the test statistic $\chi_t^{\max}$ exceeds a certain threshold, a change is declared indicating the persistence histograms before and after time k are sufficiently different.

The threshold b in (6) can be set by controlling the false alarm rate to a desired pre-specified level, or, equivalently, by controlling the average run length (ARL, more on this later) to be above a pre-specified level. For the above weighted-l_2 statistic, theoretical approximations are available for the ARL as a function of threshold b (Xie and Xie 2021):

$$\mathbb{E}_\infty[\mathcal{T}] = \frac{1}{2} \frac{b^{-1} e^{b^2/(2\sigma_p^2)} [2\pi\sigma_p^2]^{1/2}}{\int_{[4b^2/(m_0\sigma_p^2)]^{1/2}}^{[4b^2/(m_1\sigma_p^2)]^{1/2}} y\nu^2(y)dy} (1 + o(1)).$$

Here, $\sigma_p^2 = 4[\sum_{l=1}^{L} \sigma_i^2 p_i^2 (1-p_i)^2 + \sum_{i \neq j} \sigma_i \sigma_j p_i^2 p_j^2]$, $p = p_{\text{pre}}$ is the pre-change distribution, m_0 and m_1 are two known constants, and $\nu(\cdot)$ is a special function that is easy to calculate. Further details can be found in Xie and Xie (2021). With this, one can set the threshold b to bound the ARL $\mathbb{E}_\infty[\mathcal{T}]$, which then controls the false alarm rate. Furthermore, for computational efficiency, one may adopt a window-limited procedure which considers only $k \in [t - m_1, t - m_0]$ in (6), instead of searching over all possible change-points $k < t$. Typically, m_1 is set to be larger than the desired detection delay, and m_0 is set to ensure the reliable estimation of the persistence histogram distribution. In later experiments, we set $m_0 = 20$ and $m_1 = 80$, which seems to provide good performance.

As discussed earlier, one potential limitation of PDs for change detection is that it might not capture translation shifts of the data. For problems where both geometric changes and translation shifts are of interest, an easy fix is as follows. One can perform the proposed PERCEPT procedure (6) (which can detect geometric changes) in parallel with standard procedures which are adept at mean-shift detection, such as Shewhart control charts (Shewhart 1925), exponential weighting mean average charts (Roberts 1959), CUSUM test (Page 1954; Lau et al. 2018), etc. The joint monitoring of these two statistics will enable the joint tracking of both geometric changes and mean shifts.

3.3. Connection between EDD and Topology

We now provide a useful connection between the expected detection delay (EDD) and topology, which sheds light on how PERCEPT may be useful for topological change detection. Consider two fundamental metrics in online change-point detection: EDD and Average Run Length (ARL). Let $\mathbb{E}_\infty$ denote the expectation under the probability measure when there is no change-point, and let $\mathbb{E}_0$ denote the expectation under the probability measure when the change happens at time 0. For a given stopping time $\mathcal{T}$ of a monitoring procedure, its ARL is defined as $\mathbb{E}_\infty[\mathcal{T}]$, the expected run length to false alarm when there is no change, and the EDD is defined as $\mathbb{E}_0[\mathcal{T}]$, the number of samples needed to detect the change. Theoretically, the EDD is known to be linearly related to the log(ARL) (Tartakovsky et al. 2015).

Next, we introduce the bottleneck distance, a standard metric for topological distance (Ghrist 2008). Suppose, for two point cloud datasets with different topologies, one computes its (untilted) PDs $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. The extracted topological features in these PDs can then be compared via a *matching* η. This matching is performed in two steps: (i) it pairs each point in the first PD $\mathcal{D}_1$ with a point in the second PD $\mathcal{D}_2$ or a point on the diagonal line, and (ii) it pairs each point in $\mathcal{D}_2$ with a point in $\mathcal{D}_1$ or a point on the diagonal. The *bottleneck distance* (Ghrist 2008) between the PDs $\mathcal{D}_1$ and $\mathcal{D}_2$ is then defined as

$$d_B(\mathcal{D}_1, \mathcal{D}_2) = \inf_{\eta} \sup_{\boldsymbol{y} \in \mathcal{D}_1} ||\boldsymbol{y} - \eta(\boldsymbol{y})||_\infty.$$

Here, the supremum is taken over all matched points in $\mathcal{D}_1$, and the infimum is taken over all possible matchings η. The bottleneck distance can also be viewed as the Wasserstein distance of order ∞. Clearly, a larger bottleneck distance indicates the extracted features from the first PD are quite different from that for the second PD (and vice versa). This then suggests the topology for the first dataset is markedly different from that for the second dataset. This link between the bottleneck distance and topological differences is formalized by the Stability Theorem (Cohen-Steiner et al. 2007), a key theorem in TDA.

With this, the EDD of the proposed PERCEPT method can then be linked to the bottleneck distance of the topologies of the pre- and post-change data. Recall that PERCEPT makes use of the nonparametric l_2-statistics on the underlying persistence diagrams. It is known (Xie and Xie 2021) that the EDD for such a procedure can be upper bounded by

$$\text{EDD} \leq \frac{2b}{(\min_i \Sigma_{ii})||p_{\text{pre}} - p_{\text{post}}||_2^2}, \tag{7}$$

where b is the prespecified detection threshold and Σ_{ii} is the ith diagonal entry of Σ. In other words, the larger the ℓ_2-difference between the pre- and post-change persistence distributions p_{pre} and p_{post}, the smaller the EDD for PERCEPT. We can then show (technical details in Appendix 6) that, under certain asymptotic approximations, $||p_{\text{pre}} - p_{\text{post}}||_2^2$ can be lower bounded by the bottleneck distance between a pre-change PD $\mathcal{D}_{\text{pre}}$ and post-change PD $\mathcal{D}_{\text{post}}$, which quantifies the change in topology. This, together with (7), suggests that, for PERCEPT, *the greater (or smaller) the topological difference is between pre- and post-change data, the smaller (or greater) its expected detection delay tends to be*, which is as desired.

3.4. Persistence Cluster Binning

The persistence histogram binning approach in Section 3.1 can be viewed as partitioning the persistence space into vertical *rectangular* regions, which are then used to bin the PDs for monitoring. This may have two limitations. First, since persistences are summed within each bin, the procedure can distinguish topological features with different birth times, but not features with similar birth times but different persistences. Second, the restriction of partitions to be vertical and rectangular may hamper the identification of regions of greatest change between the pre- and post-change distributions. When training data are available on either the pre- or post-change PDs, we present an alternate novel *persistence clustering* approach which can yield a more informed partition of the persistence space.

Consider first the case where training data are available for both pre- and post-change regimes. The idea is to find a clustering of these point clouds, so that the corresponding partition of the persistence space can discriminate well the pre- and post-change distributions. We construct these *persistence clusters* as follows. First, we perform k-means clustering (Lloyd 1982) on the pre-change PD point clouds, which returns C_{pre} cluster centers $\mathcal{C}_{\text{pre}} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{C_{\text{pre}}}\}$. Then we do the same on the post-change PDs to obtain C_{post} cluster centers $\mathcal{C}_{\text{post}} = \{\boldsymbol{c}'_1, \ldots, \boldsymbol{c}'_{C_{\text{post}}}\}$. Using the combined centers $\mathcal{C} = \mathcal{C}_{\text{pre}} \cup \mathcal{C}_{\text{post}}$, we then form a *Voronoi diagram* using centers $\mathcal{C}$ (Aurenhammer 1991), that is, a partition of $\mathbb{R}^2$ to its closest point in $\mathcal{C}$. Figure 6 visualizes this persistence clustering procedure. The number of clusters can be determined by the elbow method in k-means clustering (Ketchen and Shook 1996).

Guided by prior knowledge, the modeler may choose to use alternate clustering approaches in lieu of k-means clustering, such as BIRCH (Zhang et al. 1996), DBSCAN (Ester et al. 1996), and spectral clustering (Shi and Malik 2000). In particular, distribution-based clustering methods, for example, DBSCAN or other recent methods (Krishna et al. 2019), may be able to capture distributional features on the PD space, thus, leading to clusters which can better discriminate pre- and post-change distributions. We make use of the standard k-means clustering for persistence binning in later experiments, which turns out to perform quite well, but aim to explore more sophisticated clustering approaches as future research.

There may, of course, be cases where only pre-change (and not post-change) data are available, that is, when the modeler has prior information on data topology before a change but not after. In such cases, one may perform k-means clustering (or other clustering approaches) on the pre-change PDs for persistence binning.

With these clusters, we then employ the same weighted-ℓ_2 test statistic in (5). The only difference is that, instead of taking $\boldsymbol{\omega}_{t,k}^{[1,1]}$, $\boldsymbol{\omega}_{t,k}^{[1,2]}$, $\boldsymbol{\omega}_{t,k}^{[2,1]}$, and $\boldsymbol{\omega}_{t,k}^{[2,2]}$ as persistence histogram proportions, we take these as the proportions over persistence clusters. As Figure 6(b) shows, these persistence clusters may yield improved discrimination between the pre- and post-change distributions compared to the earlier persistence histograms, particularly when there is a large number of points (or features) captured in the persistence diagrams. When there is only a small number of features, however, we would recommend the use of the persistence histogram approach instead, since there may be insufficient data to fit the more complex persistence clusters. We will demonstrate this later in numerical experiments.

3.5. Weight Optimization

Similarly, given available training data on both the pre-change and post-change PDs, the weights $\Sigma = \text{diag}\{\sigma_1, \ldots, \sigma_L\}$ in the test statistic (5) can also be carefully specified to amplify differences between the pre- and post-change distributions. For this, we adopt the weight optimization approach in Xie and Xie (2021), which aims to find weights Σ to maximize the worst-case weighted ℓ_2 distance. This can be formulated as

$$\max_{\sigma \geq 0, g(\sigma) \leq 1} f(\sigma), \quad g(\sigma) := \max_{p_{\text{pre}} \in \Delta} \sum_i \sigma_i^2 p_{\text{pre},i}^2,$$
$$f(\sigma) := \min_{p_{\text{pre}}, p_{\text{post}}} \left\{ \sum_i \sigma_i (p_{\text{pre},i} - p_{\text{post},i})^2 : p_{\text{pre}}, p_{\text{post}} \in \Delta, \right.$$
$$\left. ||p_{\text{pre}} - p_{\text{post}}||_2 \geq \rho \right\}. \tag{8}$$

Here, $p_{\text{pre},i}$ and $p_{\text{post},i}$ denote the ith entry in the vector p_{pre} and p_{post}, respectively. The minimization in $f(\sigma)$ is taken over all possible pre- and post-change distributions p_{pre} and p_{post} (within the probability simplex Δ) that are ρ-separable for a given $\rho > 0$. The choice of ρ depends on the scale of change that one wants to detect; in our experiments later, ρ is set to be 0.1. For persistence histograms, the optimal number of bins L can be optimized simultaneously by selecting L which yields the highest $f(\sigma)$ in (8). In practice, the training pre- and post-change data are required for estimating the persistence distributions p_{pre} and p_{post}. Further details on this optimization can be found in Xie and Xie (2021). These optimized weights can then be used within the test statistic (5) for change detection.

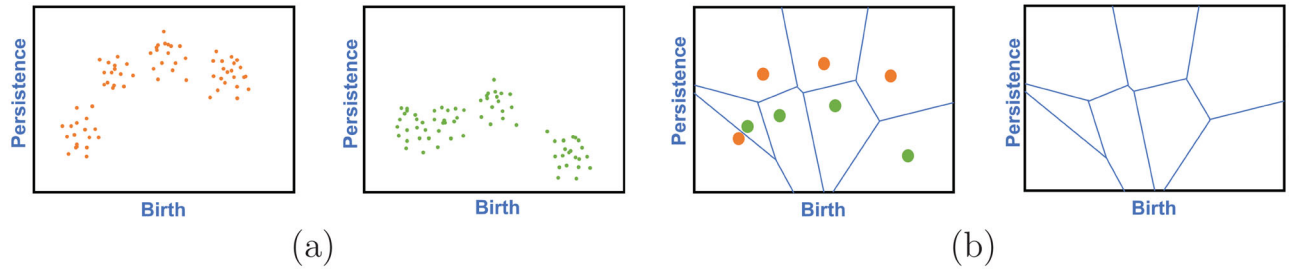

Figure 6. (a) A sample pre-change (orange) and post-change (green) PD. (b) Pre- and post-change cluster centers and its corresponding persistence cluster via Voronoi diagrams.

In implementation, we have found that a small modification of (8) can yield noticeably improved performance. Note that when $p_{\text{pre},i}/p_{\text{post},i}$ or $p_{\text{post},i}/p_{\text{pre},i}$ is exceedingly larger from 1, and $|p_{\text{pre},i} - p_{\text{post},i}|$ is small for some i, it may be difficult to pick out bin i as an important bin using the above formulation. However, such a bin i distinguishes the pre- and post-change distributions because their relative difference is large. Thus, we see that this optimization is distribution-dependent; when p_{pre} and p_{post} are extremely different from the uniform distribution, the above weights may not be ideal for distribution discrimination. In practice, we suggest using the *relative* difference of p_{pre} and p_{post} instead of the *absolute* difference in the above formulation, that is, minimizing $\sum_i \sigma_i\{(p_{\text{pre},i} - p_{\text{post},i})/p_{\text{post},i}\}^2$ instead of $\sum_i \sigma_i(p_{\text{pre},i} - p_{\text{post},i})^2$ for $f(\sigma)$ in (8).

As before, there may be scenarios where only pre-change data are available. In such a setting, one could use the weights which minimize the best-case weighted ℓ_2 distance over just the pre-change distributions. Specifically, we can divide the pre-change data into two groups, with binned distributions denoted as p_{pre_1} and p_{pre_2}. We then replace $p_{\text{pre},i}$ and $p_{\text{post},i}$ with $p_{\text{pre}_1,i}$ and $p_{\text{pre}_2,i}$, take the maximization in $f(\sigma)$, and choose weights which minimize $f(\sigma)$ in (8). With such a choice, the detection statistics are expected to be small prior to the change-point, and increase after the change-point.

4. Simulation Experiments

We now explore the performance of PERCEPT in a suite of simulation studies where the data is generated with topological structure. We investigate several change scenarios, including topology changes, noise changes, and its scalability in higher dimensions.

Three baseline methods are used here for comparison. The first is the aforementioned parametric Hotelling's T^2 test, using the 15 extracted principal components from PCA on the original data. The second is the Wasserstein distance method proposed in Ofori-Boateng et al. (2021), which makes use of the Wasserstein distance between PDs in adjacent times. Details on both methods can be found in Section 2.2. The third method is the maximum mean discrepancy (MMD) test (Gretton et al. 2012; Li et al. 2015), a widely used nonparametric change detection method. Given a class of functions $\mathcal{F}$ and two distributions p and q, the MMD distance between p and q is defined as $\text{MMD}_{\mathcal{F}}(p,q) = \sup_{f\in\mathcal{F}}(\mathbb{E}_{x\sim p}[f(x)] - \mathbb{E}_{y\sim q}[f(y)])$. When $\mathcal{F}$ is a reproducing kernel Hilbert space (RKHS) associated with kernel function $K(\cdot,\cdot)$, this MMD statistic can be written as

$$S_t^M = \frac{1}{n_{\text{pre}}^2}\sum_{i,i'=1}^{n_{\text{pre}}} K(\boldsymbol{x}_i,\boldsymbol{x}_{i'}) + \frac{1}{n_{\text{post}}^2}\sum_{j,j'=1}^{n_{\text{post}}} K(\boldsymbol{x}_j,\boldsymbol{x}_{j'}) - \frac{2}{n_{\text{pre}}n_{\text{post}}}\sum_{i=1}^{n_{\text{pre}}}\sum_{j=1}^{n_{\text{post}}} K(\boldsymbol{x}_i,\boldsymbol{x}_j). \quad (9)$$

In our implementation, we used the standard Gaussian radius basis function (RBF) kernel $K(\cdot,\cdot)$, where the kernel bandwidth is chosen using the so-called "median trick" (Bernhard et al. 2018), that is, set to be the median of the pairwise distances between data points.

The simulation set-up is as follows. We simulate data $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_T$ with $T = 400$, and the change-point is set at time $t^* = 200$. In other words, $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{200}$ are generated from the pre-change distribution, and $\boldsymbol{x}_{201},\ldots,\boldsymbol{x}_{400}$ are generated from the post-change distribution. In our simulations, this point cloud data is generated with topological structure from two simple geometric shapes, the unit sphere and the ellipsoid (in varying dimensions). A quick inspection of the PDs shows the presence of many persistence features, thus, we decided to use the persistence clustering approach across all simulation studies. The goal is to have PERCEPT learn this topological structure from data, then leverage such structure to quickly identify the change-point. Here, we make use of the one-dimensional homology for PERCEPT (see Section 2.1), since both the circle and ellipsoid have a hole structure.

4.1. Shape Change

We first consider the case of geometric shape changes, where the pre-change data is sampled with noise from the unit two-dimensional circle, and the post-change data is sampled with noise from a two-dimensional ellipse. Two noise settings are considered for this experiment: $N(0, 0.05)$ and $N(0, 0.10)$. Figure 7 shows the detection statistics from PERCEPT, Hotelling's T^2 and MMD under the two noise settings. We see that both PERCEPT and MMD are able to quickly detect the change: both monitoring statistics peak up immediately after the change-point at $t = 200$. The Hotelling's T^2 statistic, however, shows a relatively larger delay, which is unsurprising since such a method is adept at detecting mean and covariance shifts, but not shape changes. This suggests that, by integrating topological structure, PERCEPT can yield improved performance over the Hotelling's T^2. For the Wasserstein distance method, the statistic experiences a large spike at exactly $t = 200$ for the noise setting $N(0, 0.05)$, but is unable to detect the change at the larger noise

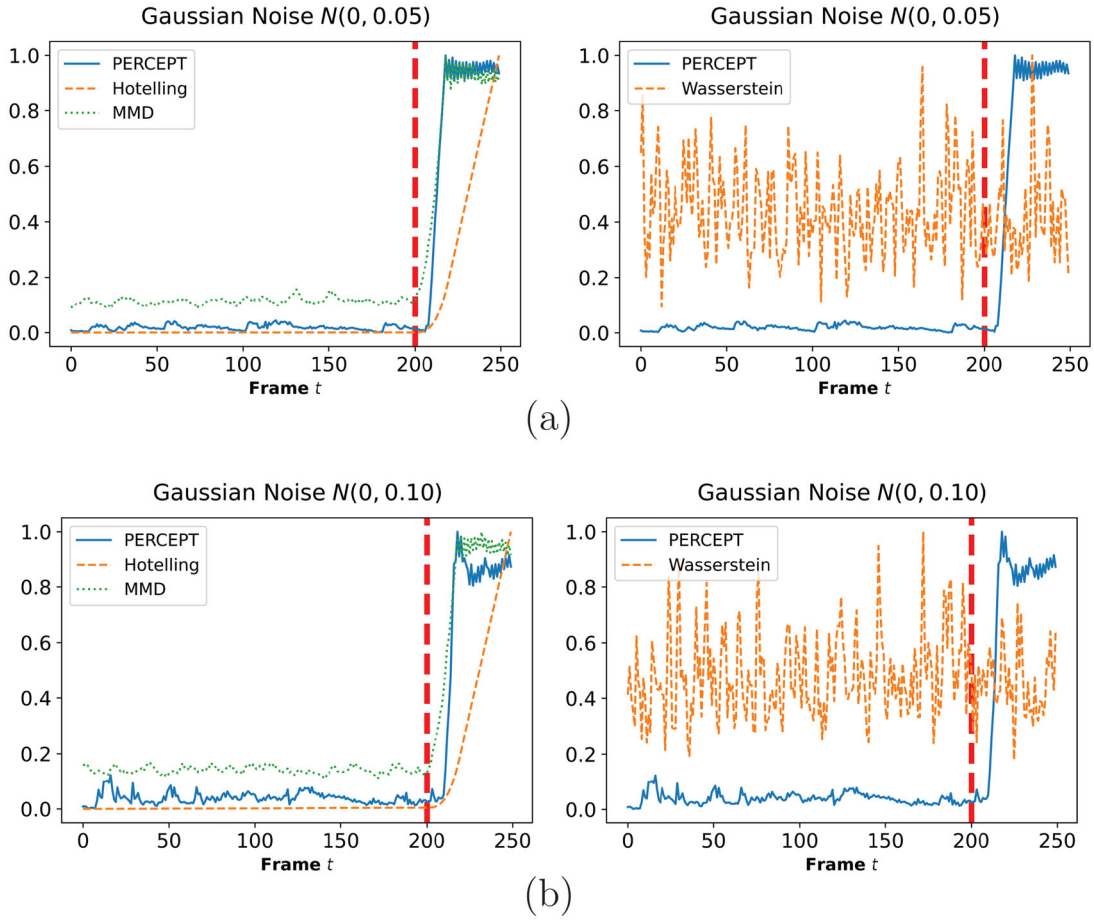

Figure 7. (a) Test statistics using PERCEPT, Hotelling's T^2, MMD and Wasserstein distance at each time t, for the setting of Gaussian noise $N(0, 0.05)$. The vertical red dashed lines indicate the true change-point $t^* = 200$. (b) The same comparison for the setting of Gaussian noise $N(0, 0.10)$.

level $N(0, 0.10)$. This affirms the earlier observation that the Wasserstein distance is a weaker metric which may not be able to distinguish local topological differences.

4.2. Noise Change

Next, we consider the case of noise changes, where the pre- and post-change data are generated from the same unit two-dimensional circle (or ellipse), but with noise levels $N(0, 0.05)$ and $N(0, 0.10)$, respectively. Figure 8 shows the test statistics from PERCEPT and the Hotelling's T^2 for the circle and ellipse. We see that PERCEPT can quickly detect the underlying change: its pre-change test statistic is quite stable, and it peaks up quickly after the change-point at $t = 200$. The Wasserstein distance approach also seems to perform quite well here, although its pre-change statistic is noticeably more unstable, which may lead to frequent false alarms (i.e., lower ARL). On the other hand, the Hotelling's T^2 statistic increases noticeably more slowly after the change, which suggests a larger detection delay; this indicates that Hotelling's T^2 can indeed be slow at detecting small variance changes. For the MMD statistic, we see that while it peaks up after the change-point, its pre-change statistic is quite unstable and volatile, which again leads to frequent false alarms. This is not unexpected, since the MMD approach does not leverage the underlying topological structure for change detection, which may result in inefficient monitoring performance.

4.3. Increasing Dimensionality

We now investigate how well these methods perform on data generated on higher-dimensional geometric structures, namely, the three-dimensional and four-dimensional unit spheres and ellipsoids. We consider here the same noise change as in Section 4.2, and we compared the performance of PERCEPT with the classic parametric Hotelling's T^2 test, MMD statistics and Wasserstein distance. Figure 9(a) shows the detection statistics for the four-dimensional sphere, and the detection statistics for the three-dimensional sphere can be found in the Appendices. We see that the proposed PERCEPT method consistently outperforms existing methods: its pre-change statistics are stable, and its post-change statistics peak up quickly after the change. Comparatively, the increase in the Hotelling-T^2 statistic is more subdued after the change (which results in greater detection delay), and the MMD statistic pre-change is noticeably more unstable (which results in increased false alarms). A likely reason for the poorer performance of the MMD statistic is that it does not integrate topological structure, which when present, can improve monitoring performance. The Wasserstein distance approach is again unable to detect this change, for similar reasons as before. Furthermore, comparing with the lower-dimensional setting in Figure 8, PERCEPT appears to yield greater improvements over existing methods, which is unsurprising since it leverages the underlying low-dimensional topological structure in the high-dimensional data. Figure 9(b)

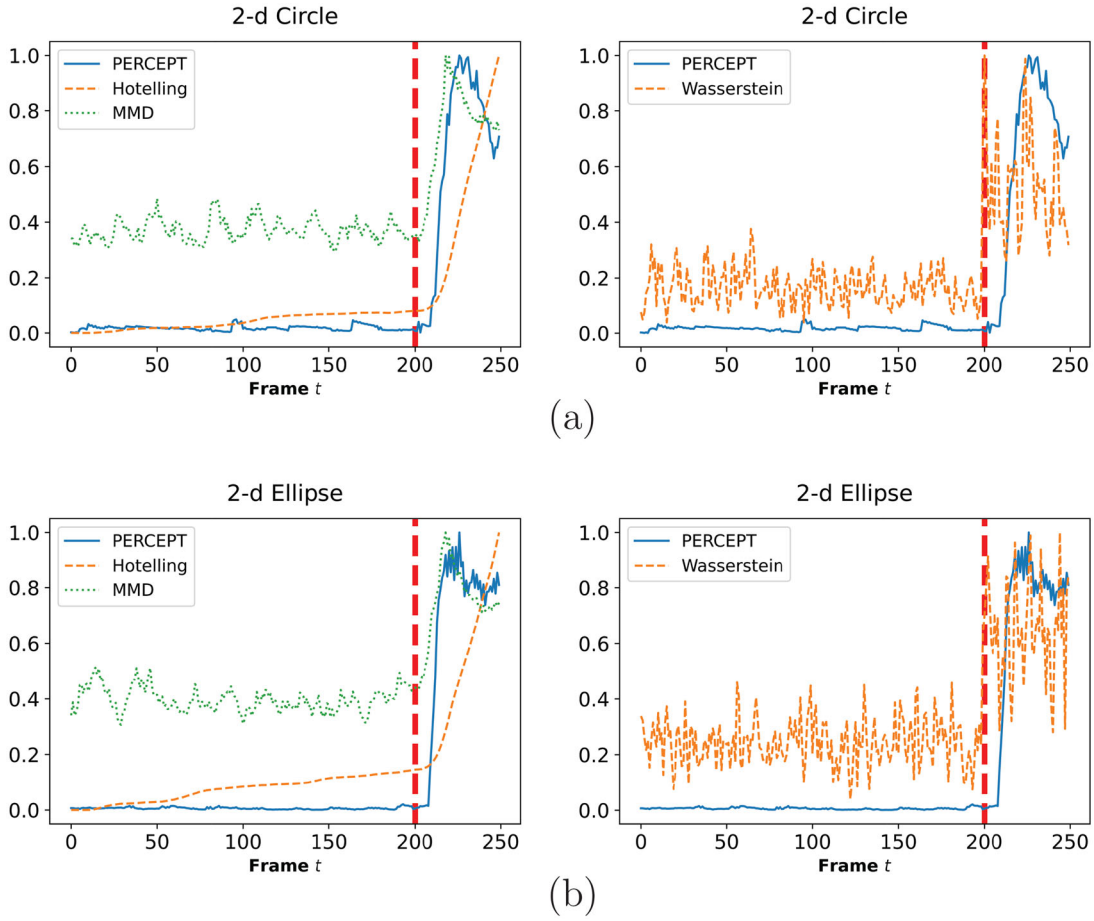

Figure 8. (a) The test statistics using PERCEPT, Hotelling's T^2, MMD and Wasserstein distance at each time t for the two-dimensional circle, where the noise changes from $N(0, 0.05)$ to $N(0, 0.10)$. Again, the vertical red dashed lines indicate the true change-point at $t^* = 200$.

shows the detection statistics for the four-dimensional ellipsoid, and the detection statistics for the three-dimensional ellipsoid can be found in the Appendices. Again, we see that PERCEPT yields improved performance over the existing benchmarks, despite having slightly more unstable pre-change statistics. These experiments suggest that, by learning and integrating low-dimensional topological structure, PERCEPT can efficiently detect changes when such structures are present in high-dimensional data.

4.4. EDD versus ARL

Finally, we compare the performance of these methods via the two metrics introduced in Section 3.3: the Average Run Length (ARL) and the Expected Detection Delay (EDD). Recall that ARL measures the expected run length to a false alarm when there is no change, and EDD measures how quickly the change is detected. A method with large ARL and small EDD is desired, since this means fewer false alarms and a smaller detection delay. We approximate the ARL and EDD of the compared methods using different thresholds b, under the two-dimensional circle experiment with Gaussian noise $N(0, 0.05)$ for pre-change regime and Gaussian noise $N(0, \sigma^2)$ for post-change regime, where $\sigma^2 = 0.09$, 0.10, and 0.11. We adopted the experiments from Xie and Xie (2021), and details can be found in Appendix 6.

Figure 10 plots the EDD versus log-ARL curves for the two-dimensional circle experiment. There are two interesting observations. First, for all noise levels and ARL levels, we see that PERCEPT yields much lower EDD compared to the Hotelling's T^2 procedure, which suggests that it indeed yields improved detection performance by integrating topological structure. Second, the proposed method is more robust to noise perturbations. As the noise level decreases, the EDD for the Hotelling's T^2 becomes noticeably larger for fixed ARL levels. Again, this is not surprising, since this change is difficult to detect without first identifying the underlying topological structure. In contrast, the EDD for PERCEPT is more stable and nearly remains the same as the noise level decreases, which shows the robustness of PERCEPT. The relationship between ARL and EDD for the four-dimensional case is not shown here since it is expensive to compute, but similar conclusions are expected from earlier results.

5. Applications

5.1. Solar Flare Change Detection

We now demonstrate the effectiveness of PERCEPT in the earlier motivating solar flare detection problem. A solar flare is an intense emission of radiation on the Sun's atmosphere, and the

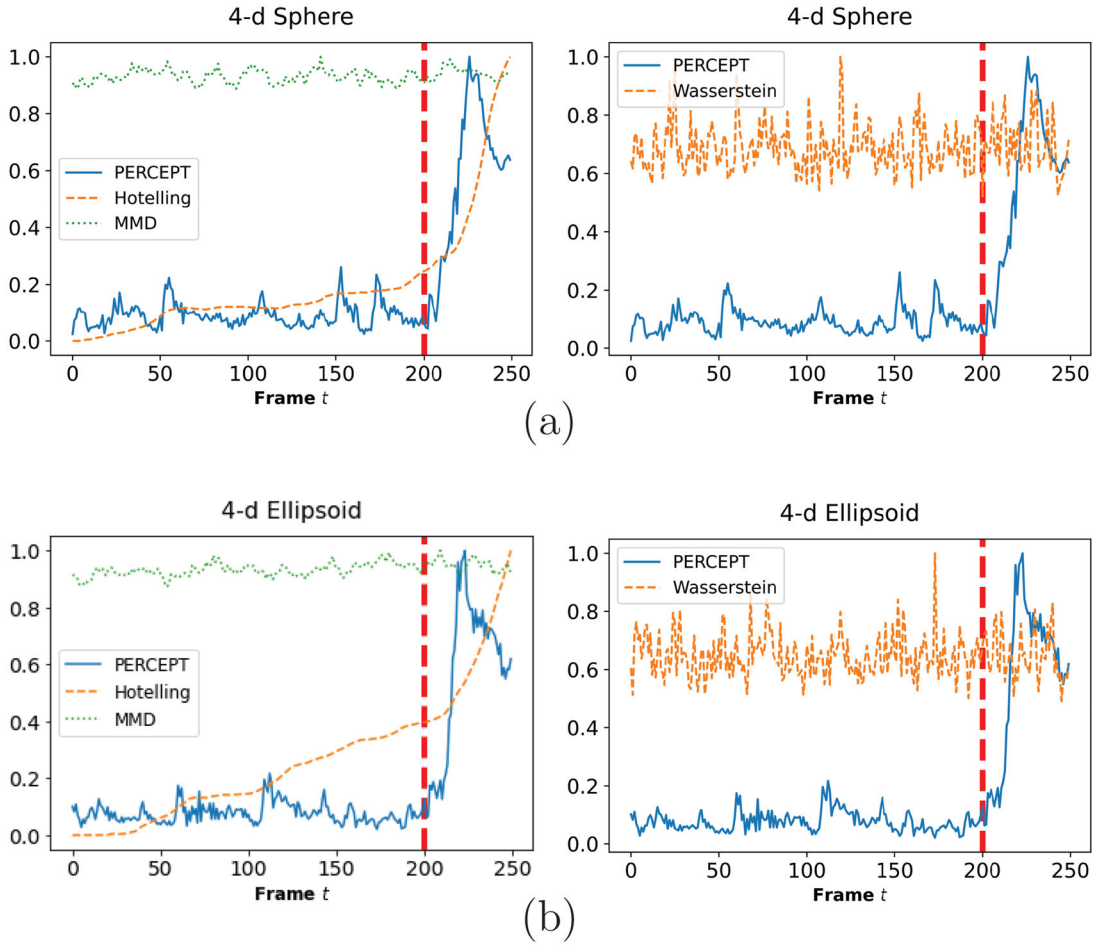

Figure 9. (a) The test statistics using PERCEPT, Hotelling's T^2, MMD and Wasserstein distance at each time t for the four-dimensional sphere, where the noise changes from $N(0, 0.05)$ to $N(0, 0.10)$. Again, the vertical red dashed lines indicate the true change-point at $t^* = 200$. (b). Same for four-dimensional ellipsoids.

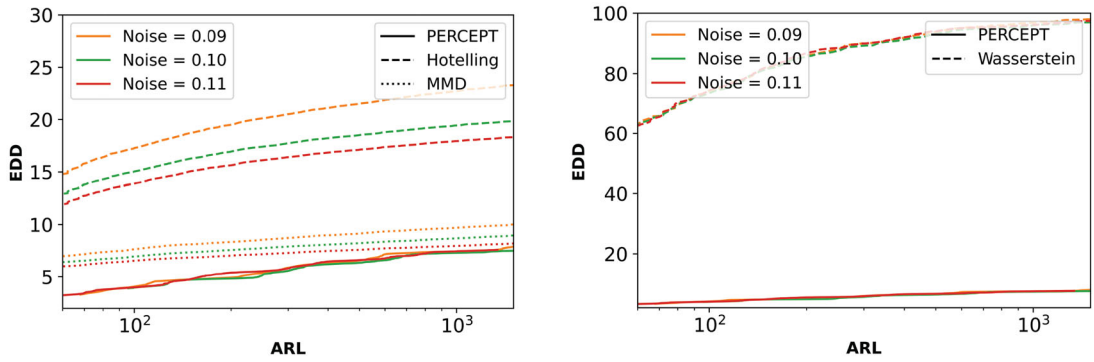

Figure 10. EDD–ARL curves for two-dimensional circle experiment with noise change, using varying post-change noise levels for PERCEPT, Hotelling's T^2, MMD, and the Wasserstein distance method.

monitoring of such changes via satellite imaging is critical for predicting geomagnetic storms (Xie et al. 2012). This detection can be highly challenging, however, as noted in Section 2.2, due to the high dimensionality of images and the subtlety of such a change. Recent work in image processing shows that a wide range of image features can be captured via topology (Bendich et al. 2011), which suggests our topology-aware approach may provide a solution. The data used here are $T = 100$ image snapshots taken by the SDO at NASA, where the true change-point is at $t^* = 49$. Further details can be found in Section 2.2.

To apply PERCEPT, we need to first map the image data to a filtration, or a sequence of simplicial complexes (see Section 2.1 for details). We will make use of the so-called *lower star filtration*, which has shown success in capturing useful topological features in image data (Bendich et al. 2011). Let $f(x)$ be a mapping from each pixel location x to its intensity value, and for a given ϵ, define the *sublevel set* of f as $X(\epsilon) = \{x | f(x) \leq \epsilon\}$. One can then form a simplicial complex from $X(\epsilon)$ by first considering image pixels as vertices on a grid, then triangulating this grid by placing an edge between two points that are horizontally, vertically, or diagonally adjacent, and a triangular face for

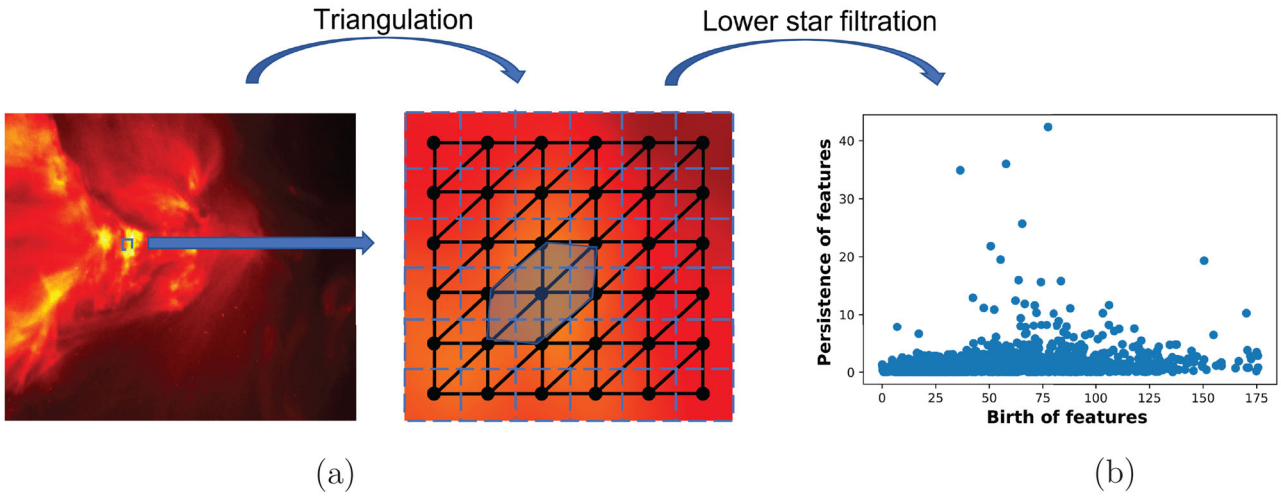

Figure 11. (a) Freudenthal triangulation of the solar flare image. (b) "Tilted" persistence diagram of the solar flare image in (a).

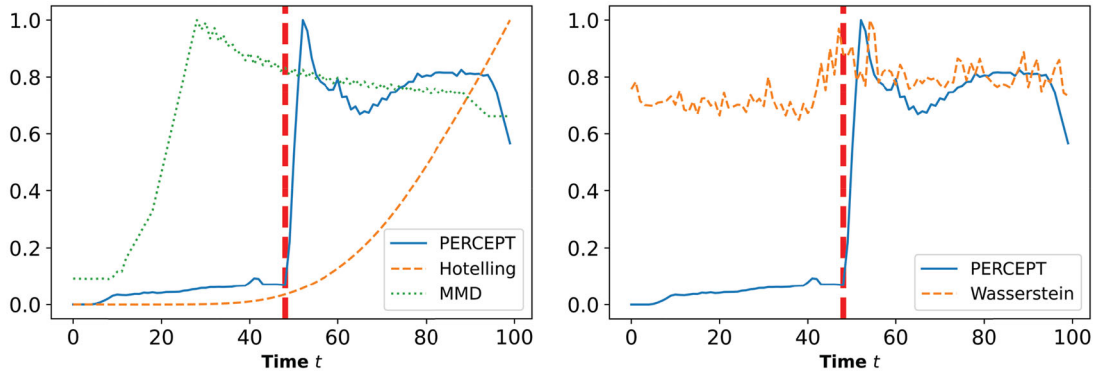

Figure 12. Detection statistics for each method in the solar flare monitoring application. Vertical red dashed lines indicate the true change-point at $t^* = 49$.

any three adjacent points forming a triangle. This is known as the Freudenthal triangulation (Freudenthal 1942). Figure 11(a) visualizes this mapping from image to simplicial complex. For a sequence $0 < \epsilon_1 < \epsilon_2 < \cdots < \epsilon_m$, the *sublevel set filtration* of the image is defined as $X_i = X(\epsilon_i)$, $i = 1, \ldots, m$, which forms a sequence of nested simplicial complexes. This filtration can be used to construct a PD (Figure 11(b)) as described in Section 2.1. One can view this filtration pipeline as a way to extract important image topological features as a PD point cloud. Here, the zero-dimensional homology is used for PERCEPT, since different solar flares can be characterized by connected components in the lower star filtration.

With the mapped PD in hand for each image, we can then proceed with the detection framework outlined in Section 3. Since there is a wide range of solar flares, we assume here that only pre-change training data are available, and we use the persistence binning approach in Section 3.1 and equal weights over persistence histogram bins. A quick inspection of the computed PDs show only a small amount of points (or features), thus, we decided to use persistence histograms instead of persistence clusters (see Section 3.4). Here, the histogram breakpoints $b_1, \ldots, b_L$ are chosen such that there is (roughly) an equal sum of persistences within each histogram bin in the first solar flare image. The Hotelling's T^2 is performed using the 30 extracted principal components from PCA on the image data, and the MMD test is performed on the image data directly, with the RBF kernel, and the bandwidth is chosen using the "median trick" as described in Section 4. The Wasserstein distance approach is performed on the mapped PDs from the lower star filtration.

Figure 12 shows the proposed detection statistic $\chi_t^{\max}$ using $L = 10$ histogram bins, the Hotelling's T^2, the MMD test statistics and the Wasserstein distance statistic. For PERCEPT, we note a sudden increase in the test statistic after the true change-point $t^* = 49$ (vertical red dashed line), which suggests the proposed method is able to pick out the underlying topological change in the images with little detection delay. Comparatively, for the Hotelling's T^2, the increase in its test statistic is much more subdued and gradual, which indicates a much larger detection delay. The MMD approach yields poor performance here: its pre-change statistics are highly volatile and unstable, and its post-change statistics experience a decrease after the change-point. The Wasserstein distance approach is not able to detect the change, since its test statistic does not increase significantly at the change-point. This is in line with results from earlier simulations, and an in-depth discussion of plausible reasons is provided there. By learning and integrating low-dimensional topological structure within a nonparametric change detection framework, PERCEPT can yield significant improvements over existing methods when such structure is indeed present for image data.

Despite the required mapping from images to PDs via the lower star filtration, PERCEPT appears to be quite efficient. This

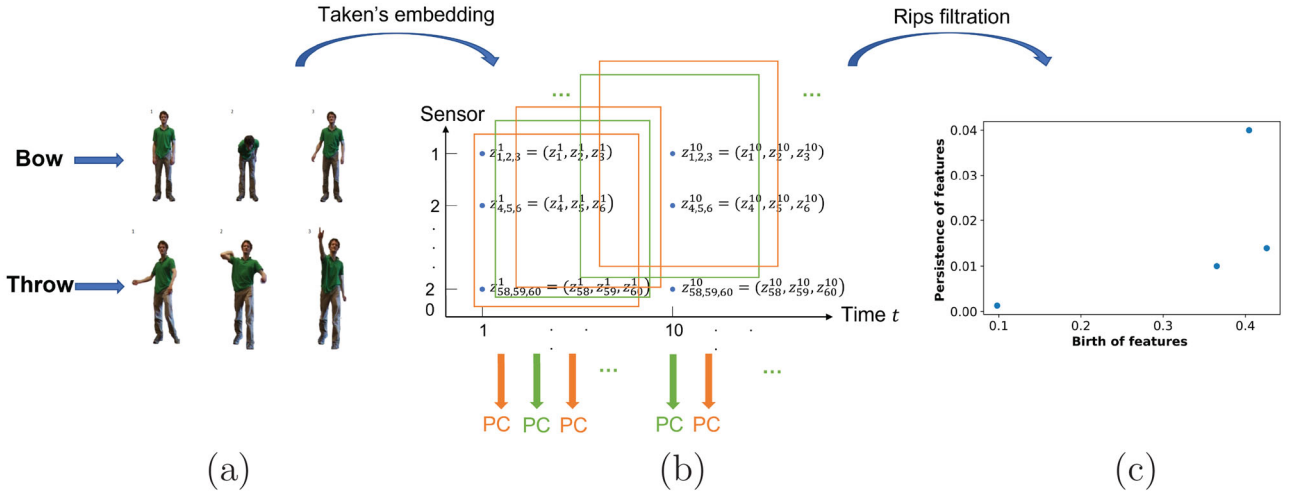

Figure 13. (a) Visualizing the two gestures: taking a bow (top) and throwing an object (bottom). (b) Embedding of a multivariate time series to point cloud (PC) data via Taken's embedding, to a persistence diagram via persistence homology. (c) "Tilted" persistence diagram of the embedded point cloud at $t = 510$.

Table 1. Total computation time (in seconds) for the compared methods on the solar flare application.

PERCEPT	Hotelling	MMD	Wasserstein
32.3 sec	2.1 sec	3.4 sec	341.8 sec

NOTE: Computation is performed on a standard desktop computer using an AMD Ryzen 5 2600 processor with 6 cores and 12 threads.

is due in part to the availability of well-maintained packages for TDA algorithms. Table 1 summarizes the computation time for each method in this application, where persistence homology (for PERCEPT) is performed using the Python package `Ripser` (Tralie et al. 2018). Here, the lower star filtration on 100 images takes around 30 sec on a standard desktop computer, which is fast given these images are quite large (232×292 pixels). The Wasserstein distance approach takes quite long due to optimal matching. With this filtration computed, the detection statistic $\chi_t^{\max}$ can then be evaluated with minimal additional computation, thus, enabling efficient detection.

5.2. Human Gesture Change Detection

Next, we investigate the performance of PERCEPT in a human gesture detection application. The detection of human body gestures is an important task in computer vision (Turaga et al. 2008), and has immediate applications in visual surveillance and sign language interpretation (Oh et al. 2011). One challenge for this detection lies in the high-dimensional time series data, and the low-dimensional embedding of human body gestures within such data. Recent developments in time series analysis suggest that many time series features can be captured via TDA (further details below), and thus PERCEPT may be promising for this task. To explore this, we will use the human gestures dataset from the Microsoft Research Cambridge-12 Kinect (Fothergill et al. 2012), which consists of time series observations of human skeletal body part movements, collected from 20 sensors on 30 people performing 12 distinct gestures. We study in particular the transition of gestures from a "bow" to a "throw" sequence, as shown in Figure 13(a).

To apply PERCEPT, we need to first transform the multidimensional time series to point cloud data, on which the usual TDA pipeline (see Section 2.1) can then be performed. A popular transformation is Taken's embedding (Gidea and Katz 2017), which is widely used in TDA. Suppose we observe the d-dimensional time series $\{z_k(t), t \geq 0\}, k = 1, \ldots, d$ (here, $d = 20 \times 3 = 60$, since each of the 20 body sensors return a three-dimensional coordinate). For each time t, define the point $\boldsymbol{z}_t = (z_1(t), \ldots, z_d(t))$ as the cross-section of the multivariate time series. With a given window size w, Taken's embedding returns the point cloud representation $\mathcal{Z}_t = \{\boldsymbol{z}_i, \ldots, \boldsymbol{z}_{i+w-1}\}$ at each time t. After this transformation, one can then perform the standard TDA filtration in Section 2.1 to convert the point cloud $\mathcal{Z}_t$ to a persistence diagram $\mathcal{D}_t$. It can be shown (Gidea and Katz 2017) that this embedding captures key periodicity and dynamic system information on the multivariate time series. Figure 13(b) provides a visualization of Taken's embedding. Here, we set the window size w to be approximately the periodicity of the gestures, which is based on prior knowledge. We further make use of one-dimensional homology for PERCEPT, since it is known that periodicity of time series can be represented as one-dimensional holes via Taken's embedding (Gidea and Katz 2017).

With this, we can then apply PERCEPT for detecting gesture changes. A quick inspection shows that there are limited points in the computed PDs $\mathcal{D}_t$. Thus, we use the persistence histograms in Section 3.1 rather than the persistence clusters in Section 3.4. We then used 60 frames from "bow" and "throw" sequences to choose the number of histogram bins and optimize weights. The Hotelling's T^2 is performed using the 30 extracted principal components from PCA on the embedded point cloud data $\mathcal{Z}_t$, and the MMD test is performed on the embedded point cloud data $\mathcal{Z}_t$ directly with the RBF kernel, and the bandwidth is chosen using the "median trick" as described in Section 4. The Wasserstein distance approach is performed on the mapped PDs from the Rips filtration.

Figure 14 shows the proposed detection statistic $\chi_t^{\max}$ using $L = 2$ histogram bins, the Hotelling's T^2, the MMD test statistics, and the Wasserstein distance over time, with the vertical line indicating the true change-point (i.e., the transition from

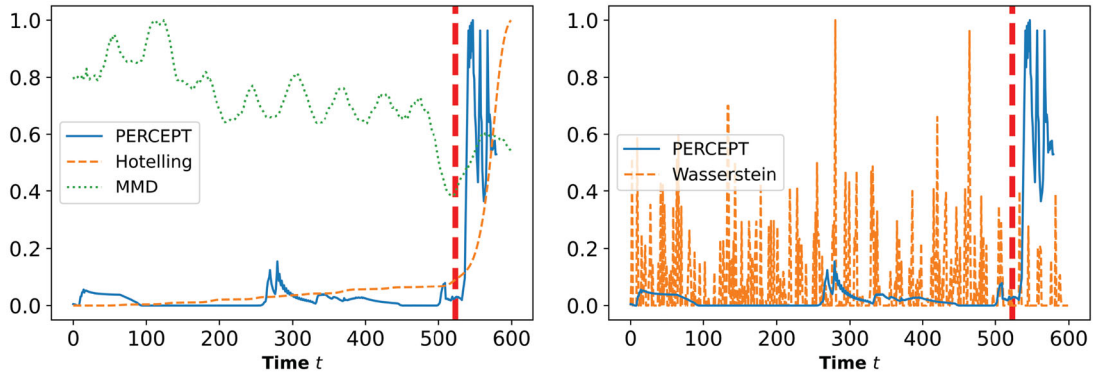

Figure 14. Detection statistics for each method in the human gesture change detection application. Vertical dashed lines indicate the true change-point from "bow" to "throw."

"bow" to "throw"). For PERCEPT, we see the test statistic is relatively stable pre-change and experiences a large jump immediately after the change. This suggests our method has a large ARL and small detection delay, which is as desired. The Hotelling's T^2 again experiences a much more subtle and gradual increase after the change, which indicates a large detection delay. As before, the MMD method performs poorly: its pre-change statistic is highly unstable pre-change and experiences a notable decrease post-change, which suggests it is unable to detect the change at all. The Wasserstein distance method is again not able to detect the change-point. This demonstrates the advantages of a topology-aware nonparametric change detection framework: when such a low-dimensional structure exists and can be leveraged, one can achieve efficient monitoring performance even with high-dimensional data.

6. Conclusion

We have proposed a novel topology-aware, nonparametric monitoring framework called the Persistence Diagram-based Change-Point Detection (PERCEPT) method, which leverages tools from topological data analysis for change detection. The idea is to first extract the topological structure of the data via persistence diagrams, then leverage a nonparametric, histogram-based change detection approach on these diagrams to sequentially detect topological changes. A suite of simulation experiments and two applications show that, when the underlying change is topological in nature, the proposed PERCEPT approach yields noticeably improved detection performance over existing approaches.

There are several interesting future directions for refining PERCEPT for broader applications. First, we are exploring a more localized detection approach, which can pinpoint and monitor local changes (e.g., local translation/rotation shifts) in topology. Second, there has been recent work on two-parameter persistence (Wright and Zheng 2020), particularly on its robustness in extracting topological structure in the presence of noise; integrating such ideas within PERCEPT would allow for a more robust topology-aware monitoring approach. Third, we aim to explore a broader range of clustering approaches for persistence binning, including DBSCAN. Finally, we will investigate broader uses of PERCEPT in applications for which TDA has found recent success, particularly neuroscience (Wang et al. 2020, 2022) and complex physical systems (Mak et al. 2018).

Appendices

Appendix A: Experiments for Approximating the ARL and EDD

Experiment 1: To compute the ARL, we generate n sequences of pre-change samples of length m. We create a random pool with M pre-change samples to reduce the computational cost. For each sequence, m samples are drawn randomly from the pool with replacement, and we repeat this process n times. Let $\mathcal{T}$ be the stopping time of the detection procedure, and if there is no change-point in the sequence of length m (all samples are drawn from the pre-change distribution), from the discussion in Xie and Xie (2021) we have:

$$P(\mathcal{T} > m) = P(\max_{0\le t\le m} \chi_t < b) \approx \exp\{-m/\lambda\},$$

where λ is the estimation of the ARL. For any given threshold b, we could get the estimation of the ARL based on the percentage of sequences whose maximum online statistics is below the threshold, among all n sequences. More specifically, we approximate the ARL as $m/(-\ln \hat{p})$, where $\hat{p}$ is the percentage.

Experiment 2: To compute the EDD, we generate n sequences of pre-change samples of length m' followed by post-change samples of length m. The pre-change samples are only used as historical data, in order to construct the pre-change histograms $\boldsymbol{\omega}_{t,k}^{[1,1]}$, $\boldsymbol{\omega}_{t,k}^{[1,2]}$ as shown in Figure 5(b). The online test statistics are calculated only from the onset of post-change samples to obtain the detection delay. We consider the same list of threshold b as we used in computing the ARL, and find the average detection delay over n sequences.

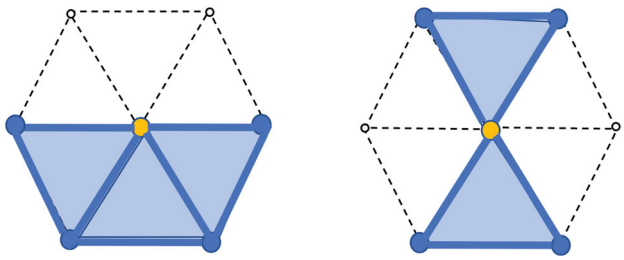

Figure B.1. Visualizing a regular point (left) and a saddle point (right) in the lower star filtration.

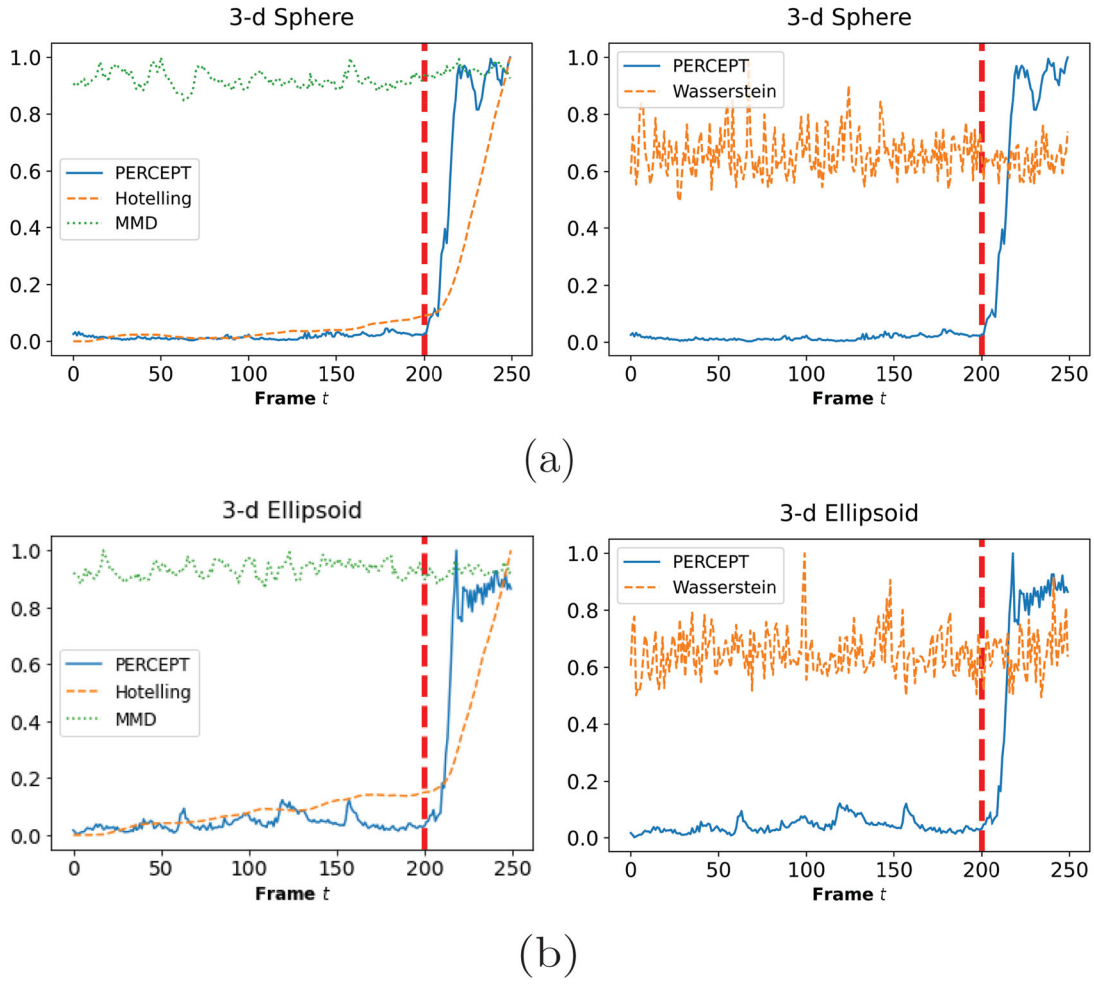

Figure D.1. (a) The test statistics using PERCEPT, Hotelling's T^2, MMD and Wasserstein distance at each time t for the three-dimensional sphere, where the noise changes from $N(0, 0.05)$ to $N(0, 0.10)$. Again, the vertical red dashed lines indicate the true change-point at $t^* = 200$. (b) Same for three-dimensional ellipsoids.

Appendix B: Details on the Lower Star Filtration

When a new vertex is added in the sublevel set, the topological change depends on whether the vertex is a maximum, minimum, regular, or a saddle of the function. Figure B.1 visualizes a regular point and saddle point (in yellow), and the edges and faces in the sublevel sets (in blue). The topological features do not change after introducing a regular point, but the number of connected components decreases by one after introducing a saddle point.

Appendix C: Connection on Distributional and Bottleneck Distances

Here, we elaborate further on the connection between the difference between the pre- and post-change persistence distributions, $\|p_{\text{pre}} - p_{\text{post}}\|_2$, and its corresponding bottleneck distance $d_B(\mathcal{D}_{\text{pre}}, \mathcal{D}_{\text{post}})$. Since the bottleneck distance is defined on the sampled PDs $\mathcal{D}_{\text{pre}}$ and $\mathcal{D}_{\text{post}}$ at a given time t, we will make this connection on the *empirical* pre- and post-change persistence distributions $\hat{p}_{\text{pre}}$ and $\hat{p}_{\text{post}}$ at time t, respectively.

The following proposition provides a link between the empirical persistence distribution difference $\|\hat{p}_{\text{pre}} - \hat{p}_{\text{post}}\|_2$ and the bottleneck distance of its corresponding persistence diagrams $\mathcal{D}_{\text{pre}}$ and $\mathcal{D}_{\text{post}}$, under asymptotic conditions.

Proposition 1. Let $\mathcal{D}_{\text{pre}}$ and $\mathcal{D}_{\text{post}}$ be the PDs of samples from pre- and post-change data at time t, and let $\hat{p}_{\text{pre}}$ and $\hat{p}_{\text{post}}$ be the corresponding persistence histograms given a fixed partition of L bins at time t. Suppose all birth times are unique, and suppose the number of histogram bins L are sufficiently large such that the persistence histograms at each time has at most one point in each bin. Then $||\hat{p}_{\text{pre}} - \hat{p}_{\text{post}}||_2^2 \geq d_B^2(\mathcal{D}_{\text{pre}}, \mathcal{D}_{\text{post}})$.

In other words, as the number of histogram bins L goes to infinity, the distributional difference $||\hat{p}_{\text{pre}} - \hat{p}_{\text{post}}||_2$ can be lower bounded by the bottleneck distance $d_B(\mathcal{D}_{\text{pre}}, \mathcal{D}_{\text{post}})$. This, combined with Equation (7) of the main article, suggests that the greater the topological difference is between pre- and post-change data, the smaller its detection delay, as desired.

Appendix D: Increasing Dimensionality on Geometrix Structures

In Section 4.3, we investigate how well the methods perform on data generate on higher dimensional geometric structures under noise change, and Figure D.1 shows the detection statistics for three-dimensional sphere and ellipsoids.

Acknowledgments

The authors thank Yi Ji and Alessandro Zito for helpful comments on the article.

Data Availability Statement

Code accompanying this article can be found at *https://github.com/Xiaojzheng/PERCEPT*.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Yao Xie is partially supported by NSF DMS-2134037, CCF-1650913, DMS-1938106, and DMS-1830210. Simon Mak is partially supported by NSF CSSI Frameworks 2004571 and NSF DMS-2210729.

ORCID

Yao Xie https://orcid.org/0000-0001-6777-2951

References

Aurenhammer, F. (1991), "Voronoi Diagrams — A Survey of a Fundamental Geometric Data Structure," *ACM Computing Surveys*, 23, 345–405. [168]

Basseville, M., and Nikiforov, I. V. (1993), "*Detection of Abrupt Changes: Theory and Application*, Upper Saddle River, NJ: Prentice Hall. [162]

Beksi, W. J., and Papanikolopoulos, N. (2016), "3D Eegion Segmentation using Topological Persistence," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1079–1084. [163]

Bendich, P., Edelsbrunner, H., and Kerber, M. (2011), "Computing Robustness and Persistence for Images," *IEEE Transactions on Visualization and Computer Graphics*, 6, 1251–1260. [166,172]

Bernhard, S., Smola, A. J., and Bach, F. (2018), *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, Cambridge, MA: MIT Press. [169]

Cang, Z., Mu, L., Wu, K., Opron, K., Xia, K., and Wei, G.-W. (2015), "A Topological Approach for Protein Classification," *Molecular Based Mathematical Biology*, 3, 140–162. [162]

Chazal, F. (2016), "High-Dimensional Topological Data Analysis," in *Handbook of Discrete and Computational Geometry*, eds. C. D. Toth, J. O'Rourke, J. E. Goodman, (3rd ed.), pp. 663–683, Boca Raton, FL: CRC Press. [164]

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. (2007), "Stability of Persistence Diagrams," *Discrete & Computational Geometry*, 37, 103–120. [167]

Edelsbrunner, H., and Harer, J. (2008), "Persistent Homology – A Survey," *Contemporary Mathematics*, 453, 257–282. [163]

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996), "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 226–231. AAAI Press. [168]

Fothergill, S., Mentis, H., Kohli, P., and Nowozin, S. (2012), "Instructing People for Training Gestural Interactive Systems," *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, pp. 1737–1746. [174]

Freudenthal, H. (1942), "Simplizialzerlegungen von beschrankter flachheit," *Annals of Mathematics*, 43, 580–582. [173]

Ghrist, R. (2008), "Barcodes: The Persistent Topology of Data," *Bulletin of the American Mathematical Society*, 45, 61–75. [163,167]

Gidea, M., and Katz, Y. A. (2017), "Topological Data Analysis of Financial Time Series: Landscapes of Crashes," *Physica A: Statistical Mechanics and its Applications*, 491, 820–834. [174]

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012), "A Kernel Two-Sample Test," *Journal of Machine Learning Research*, 13, 723–773. [169]

Han, S., Okonek, T., Yadav, N., and Zheng, X. (2018), "Distributions of Matching Distances in Topological Data Analysis," *SIAM Undergraduate Research Online* 13, 57–76. [163]

Hotelling, H. (1947), *Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights*, pp. 111–184, New York: McGraw-Hill. [164]

Islambekov, U., Yuvaraj, M., and Gel, Y. (2019), "Harnessing the Power of Topological Data Analysis to Detect Change Points," *Environmetrics*, 31, e2612. [163]

Jiao, Y., Chen, Y., and Gu, Y. (2018), "Subspace Change-Point Detection: A New Model and Solution," *IEEE Journal of Selected Topics in Signal Processing*, 12, 1224–1239. [162]

Ketchen, D. J., and Shook, C. L. (1996), "The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique," *Strategic Management Journal*, 17, 441–458. [168]

Krishna, A., Mak, S., and Joseph, R. (2019), "Distributional Clustering: A Distribution-Preserving Clustering Method," arXiv preprint arXiv:1911.05940. [168]

Lai, T. L. (1998), "Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems," *IEEE Transactions on Information Theory*, 44, 2917–2929. [162]

Lau, T. S., Tay, W. P., and Veeravalli, V. V. (2018), "A Binning Approach to Quickest Change Detection with Unknown Post-Change Distribution," *IEEE Transactions on Signal Processing*, 67, 609–621. [167]

Li, S., Xie, Y., Dai, H., and Song, L. (2015), "M-statistic for Kernel Change-Point Detection," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 3366–3374. [169]

Lloyd, S. (1982), "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, 28, 129–137. [168]

Lorden, G. (1971), "Procedures for Reacting to a Change in Distribution," *Annals of Mathematical Statistics*, 42, 1897–1908. [162]

Mak, S., Sung, C.-L., Wang, X., Yeh, S.-T., Chang, Y.-H., Joseph, V. R., Yang, V., and Wu, C. F. J. (2018), "An Efficient Surrogate Model for Emulation and Physics Extraction of Large Eddy Simulations," *Journal of the American Statistical Association*, 113, 1443–1456. [175]

Molloy, T. L., and Ford, J. J. (2017), "Misspecified and Asymptotically Minimax Robust Quickest Change Detection," *IEEE Transactions on Signal Processing*, 65, 5730–5742. [162]

Moustakides, G. V. (1986), "Optimal Stopping Times for Detecting Changes in Distributions," *Annals of Statistics*, 14, 1379–1387. [162]

Ofori-Boateng, D., Dominguez, I. S., Akcora, C., Kantarcioglu, M., and Gel, Y. R. (2021), "Topological Anomaly Detection in Dynamic Multilayer Blockchain Networks," in *Machine Learning and Knowledge Discovery in Databases. Research Track*, eds. M. Ceci, J. Hollmén, L. Todorovski, C. Vens, and S. Džeroskipp, pp. 788–804, Cham: Springer. [163,164,169]

Oh, S., Hoogs, A., Perera, A., Cuntoor, N., Chen, C.-C., Lee, J. T., Mukherjee, S., Aggarwal, J. K., Lee, H., Davis, L., Swears, E., Wang, X., Ji, Q., Reddy, K., Shah, M., Vondrick, C., Pirsiavash, H., Ramanan, D., Yuen, J., Torralba, A., Song, B., Fong, A., Roy-Chowdhury, A., and Desai, M. (2011), "A Large-Scale Benchmark Dataset for Event Recognition in Surveillance Video," *IEEE Xplore*, pp. 3153–3160. [174]

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100–115. [164,167]

Perea, J. A., and Harer, J. (2015), "Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis," *Foundations of Computational Mathematics*, 15, 799–838. [162,164]

Poor, H. V., and Hadjiliadis, O. (2008), *Quickest Detection*, Cambridge: Cambridge University Press. [162]

Roberts, S. W. (1959), "Control chart Tests based on Geometric Moving Averages," *Technometrics*, 1, 239–250. [167]

Seversky, L. M., Davis, S., and Berger, M. (2016), "On Time-Series Topological Data Analysis: New Data and Opportunities," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1014–1022. [162,163]

Shewhart, W. A. (1925), "The Application of Statistics as an Aid in maintaining Quality of a Manufactured Product," *Journal of the American Statistical Association*, 20, 546–548. [167]

Shi, J., and Malik, J. (2000), "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 888–905. [168]

Siegmund, D. O. (1985), *Sequential Analysis: Tests and Confidence Intervals*. Springer Series in Statistics, New York: Springer. [162]

Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R., and Bassett, D. S. (2019), "The Importance of the Whole: Topological Data Analysis for the Network Neuroscientist," *Network Neuroscience*, 3, 656–673. [162,164]

Tartakovsky, A., Nikiforov, I., and Basseville, M. (2015), *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Ser. Monographs on Statistics and Applied Probability 136. Boca Raton, FL: Chapman & Hall/CRC Press. [162,167]

Tralie, C., Saul, N., and Bar-On, R. (2018), "Ripser.py: A Lean Persistent Homology Library for Python," *The Journal of Open Source Software*, 3, 925. [174]

Turaga, P., Chellappa, R., Subrahmanian, V. S., and Udrea, O. (2008), "Machine Recognition of Human Activities: A Survey," *IEEE Transactions on Circuits and Systems for Video Technology*, 18, 1473–1488. [174]

Wang, H., Xie, L., Cuozzo, A., Mak, S., and Xie, Y. (2020), "Uncertainty Quantification for Inferring Hawkes Networks," in *Advances in Neural Information Processing Systems*. [175]

Wang, H., Xie, L., Xie, Y., Cuozzo, A., and Mak, S. (2022), "Sequential Change-Point Detection for Mutually Exciting Point Processes over Networks," *Technometrics* (accepted), DOI: 10.1080/00401706.2022.2054862. [175]

Wright, M., and Zheng, X. (2020), "Topological Data Analysis on Simple English Wikipedia Articles," *The PUMP Journal of Undergraduate Research*, 3, 308–328. [175]

Xie, L., and Xie, Y. (2021), "Sequential Change Detection by Optimal Weighted ℓ_2 Divergence," *IEEE Journal on Selected Areas in Information Theory*, 2, 747–761. [163,167,168,171,175]

Xie, L., Xie, Y., and Moustakides, G. V. (2020), "Sequential Subspace Change Point Detection," *Sequential Analysis*, 39, 307–335. [162]

Xie, Y., Huang, J., and Willett, R. (2012), "Change-Point Detection for High-Dimensional Time Series with Missing Data," *IEEE Journal of Selected Topics in Signal Processing*, 7, 12–27. [165,172]

Zhang, T., Ramakrishnan, R., and Livny, M. (1996), "Birch: An Efficient Data Clustering Method for Very Large Databases," *SIGMOD Record*, 25, 103–114. [168]