

ASSESSING PERIODICITY IN TIME SERIES DATA USING PERSISTENT HOMOLOGY

Zabidi Abu Hasan^{1,3}, R. U. Gobithaasan^{2,†} and Roslan Hasni³

Abstract Topological Data Analysis (TDA), a novel method rooted in algebraic topology, offers a powerful framework for quantifying the structure of high-dimensional datasets. This study focuses on Persistent Homology (PH), a key TDA method, to elucidate the underlying shape and structure of time series data, thereby facilitating the understanding of patterns and recurrence. By employing Takens' embedding to represent the time series as a high-dimensional point cloud, this approach effectively encodes its topological characteristics, particularly periodicity. We utilize various PH summaries, in the form of persistence vectors (PV), to assess periodicity by quantifying connected components (H_0) and one-dimensional holes (H_1). The methodology is validated on ten synthetic datasets, each comprising 200 data points and exhibiting varying levels of periodicity, amplitude, and frequency. The optimal window size for the sliding window embedding is determined by maximizing the $PD_{\max}$ score. Subsequently, the L^1 -norm of the PVs is computed to quantify periodicity. Under optimal window selection ($D = 14$), the cosine signal achieves the highest periodicity, followed by damped cosine and linear trend. Formal quantitative stability analysis identifies one of the two Sliding Window 1-Persistence Scoring (SW1PerS) baselines, SW2, as the most stable representation under both Gaussian noise (Rank Instability Index, RII= 0.243) and missing data (Mean Absolute Rank Deviation, MARD= 0.117), with the Persistence Landscape Score (PLS) and SW1 ranking closely behind; differences under noise are not statistically significant (Friedman $p = 0.42$), while those under missing data are significant ($p = 0.001$). A comparison against non-topological baselines (FFT-based spectral entropy, autocorrelation, and recurrence quantification analysis) demonstrates that PLS achieves Kendall $\tau = 0.867$, substantially outperforming spectral entropy (SE: $\tau = 0.733$) which is the only non-topological baseline that achieves statistically significant concordance ($p = 0.002$), while the autocorrelation peak ratio (APR: $\tau = -0.289$) and recurrence rate (RR: $\tau = -0.200$) yield non-significant negative concordance on these synthetic signals.

Keywords Time series, topological data analysis, persistent homology, persistence vector, sliding window persistence.

MSC(2010) 55N31, 68T09.

[†]The corresponding author.

¹Department of Mathematical Sciences, Faculty of Intelligent Computing, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia

²School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

³Faculty of Computer Science & Mathematics, University Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

Email: zabidi@unimap.edu.my(Z. A. Hasan), gobithaasan@usm.my(R. U. Gobithaasan), hrosan@umt.edu.my(R. Hasni)

1. Introduction

Real-world time series datasets often contain periodic patterns as well as noise and missing data, which can make analysis challenging. Periodicity refers to a repeating pattern in a time series such that $y(t + T) = y(t)$ for some period T [43]. Identifying periodic patterns in time series is beneficial for tasks such as forecasting, clustering, classification, and anomaly detection. Numerous methods have been explored in past studies for detecting periodicity in time series [43, 44]. Classical non-topological approaches include spectral methods such as the Fast Fourier Transform (FFT) and derived measures such as spectral entropy, autocorrelation function (ACF) analysis, and Recurrence Quantification Analysis (RQA) [36]. While these methods are well-established, they can be sensitive to non-stationarity, amplitude variation, and noise, and typically capture only global or linear periodic structure. Recent deep-learning and graph-neural-network methods have also demonstrated strong performance on financial and asset-price time series prediction [13, 53, 57]. The topological framework proposed here is complementary: Rather than learning predictive mappings, it provides an interpretable structural characterization of time-series periodicity that is model-agnostic and unsupervised.

Recently, topological data analysis (TDA) has emerged as a novel technique for quantifying the structure of high-dimensional datasets [12, 27, 39, 40, 42]. This approach can be used to assess the periodicity of time series. TDA is built on the foundation of algebraic topology, a branch of mathematics that uses abstract algebra to study topological spaces [27, 40].

One of the first studies to explore TDA for time series via phase-space embedding was by Muldoon et al. [38], who examined the topological properties of a time series' invariant manifold. In 2012, Silva et al. [49] used a topological model to categorize periodic and quasi-periodic dynamics. In 2015, Perea and Harer [44] expanded on the geometric characterization of time series using an ellipse representation and a sliding window embedding. The periodicity score they introduced, termed Sliding Window 1-Persistence Scoring (SW1PerS), quantifies periodicity by measuring how closely the sliding-window point cloud approximates a circle through the most persistent one-dimensional hole (H_1) of its persistence diagram; its two variants, denoted SW1 and SW2, are used throughout this paper as comparison baselines. Myers et al. [41] continued with a similar approach using Takens' theorem for embedding. Persistence diagrams (PDs) have been successfully applied for periodicity ranking and quasi-periodicity analysis [19, 42, 43, 52]. In addition, PD-based techniques have been used in finance [20–23, 29–31], biomedical signal processing [18], and climate science [24–26, 50, 59, 60].

Even though persistence diagrams are well-established for summarizing structure in data, they cannot be directly used as input to most machine learning algorithms [37, 55]. A common alternative is to represent topological summaries in a vectorized form. Motivated by Bubenik's persistence landscapes [10], one can obtain a topological summary in a Banach space that can be utilized for statistical analysis and machine learning. In this work, we extend the periodicity ranking approach of Perea et al. [43, 44] and investigate several types of persistence vectors (PVs) derived from PDs.

We also introduce formal quantitative measures of rank stability and inter-method consistency (Section 8), addressing the need for objective evaluation of the proposed periodicity rankings. A practitioner's interpretability guide is provided in Section 12.

The remainder of this paper is organized as follows. Section 2 introduces the mathematical background (Takens' embedding, PH, persistence diagrams, DTM filtration). Section 3 presents the six persistence vector representations and L^p -norm. Section 4 defines the Persistence Landscape Score. Section 5 describes the synthetic datasets. Section 6 presents the methodology.

Section 7 reports rank analysis under three experimental conditions. Section 8 provides formal quantitative stability analysis. Section 9 presents DTM filtration rank analysis. Section 10 examines normalization effects. Section 11 benchmarks PV methods against non-topological baselines. Section 12 provides a practitioner interpretability guide. Section 13 concludes.

2. Mathematical background

This section introduces the mathematical foundations: Takens' embedding for phase-space reconstruction (Section 2.1), persistent homology for topological feature extraction (Section 2.2), the persistence diagram as a topological summary (Section 2.3), and the DTM filtration for noise-robust analysis (Section 2.4).

2.1. Takens' embedding

Takens' embedding is a process for reconstructing a one-dimensional time series in a higher-dimensional phase space [51]. This step is a crucial prerequisite for applying persistent homology to time series data. A time series can be viewed as a signal indexed by time t . From a 1D time series $f(t)$, we can embed the signal into an n -dimensional space using Takens' theorem as follows:

Definition 2.1. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a time-series and let τ be a delay parameter. The time-delay embedding of f in $\mathbb{R}^D$ is the mapping $\phi : \mathbb{R} \rightarrow \mathbb{R}^D$ defined by

$$\phi(t) = (f(t), f(t + \tau), \dots, f(t + (D - 1)\tau)), \quad (2.1)$$

where $D\tau$ is the window size [51]. This construction represents the time series as a point cloud in $\mathbb{R}^D$. Let M be a compact manifold of dimension m . For a pair (ϕ, y) with $\phi \in \text{Diff}^2(M)$ and $y \in C^2(M, \mathbb{R})$, it is a generic property that the map $\Phi_{(\phi, y)} : M \rightarrow \mathbb{R}^d$ defined by

$$\Phi(x)_{(\phi, y)} = (y(x), y(\phi(x)), \dots, y(\phi^{D-1}(x))) \quad (2.2)$$

is an embedding [41].

The reconstructed manifold in $\mathbb{R}^D$ depends on the choice of time delay τ and embedding dimension D . Two standard approaches for selecting τ and D are the use of average mutual information and the false nearest neighbors method [34]. However, Perea et al. [44] developed a framework to identify an optimal window size $(\tau \times D)$ based on the roundness of an ellipse in the D -dimensional embedded space.

2.2. Persistent homology

Persistent homology (PH) is a powerful tool for quantifying a dataset's connectivity by constructing simplicial complexes on a metric space. An abstract simplicial complex can be thought of as a geometric realization of a point cloud, generalizing the concept of a graph. For a distance threshold ϵ , if two points in the cloud are within ϵ of each other, they are connected by an edge (1-simplex); if three points are pairwise within ϵ , they form a filled triangle (2-simplex), and so on. In general, a simplicial complex K is a collection of simplices over a vertex set V such that if a simplex is in K , then all of its faces are also in K .

In this work, we use the Vietoris–Rips filtration to generate simplicial complexes (Rips complexes) [17]. Let $PC = \{x_i \in \mathbb{R}^D\}_{i=0}^n$ be a finite point cloud. For a scale parameter $\epsilon \geq 0$, the Vietoris–Rips complex $R_\epsilon(PC)$ consists of simplices representing subsets of points with diameter at most ϵ . As ϵ increases, the complex grows by adding higher-dimensional simplices, revealing topological features at different scales. Through this multiscale approach, we can capture connected components (H_0), loops (H_1), and voids (H_2) across a range of resolutions.

2.3. Persistence diagram

A persistence diagram (PD) provides a visual summary of the birth and death of topological features in a point cloud. For each homology dimension n , the persistence diagram dgm_n is defined as the multiset of points $\{(b_i, d_i)\}$ in the plane, where each point corresponds to an n -dimensional homology class that appears (birth b_i) and disappears (death d_i) in the filtration. The difference $d_i - b_i$ is the lifetime of that feature [7, 46].

In practice, the most commonly analyzed PDs are for H_0 and H_1 . Points that lie farther from the diagonal $b = d$ have longer lifetimes and correspond to more prominent topological features. We define the maximum lifetime in a PD as

$$ML(d, \tau) = \max_i (d_i - b_i), \quad (2.3)$$

where the maximum is taken over all features in the diagram. In our experiments, we fix the embedding dimension $D = 14$ and adjust the time delay τ to maximize this persistence measure. Specifically, we define

$$PD_{\max} = \max_{\tau} ML(D, \tau), \quad (2.4)$$

so that the optimal sliding window length is $D \times \tau$. Traditionally, features with very short lifetimes (points near the diagonal) are regarded as noise. However, recent studies suggest that collections of such short-lived features can carry meaningful geometric information about the data [11, 54].

2.3.1. PD example

We use the synthetic time series

$$x(t) = \cos(2\pi t/50) + \gamma, \quad (2.5)$$

where γ is Gaussian noise with standard deviation 0.8 [47]. We embed with $D = 14$, $\tau = 1$, construct a Rips complex with $N = 200$ vertices, and compute the PD (Figure 1). The embedded point cloud is reduced to 2D via PCA [1] for visualization. For H_1 , the largest hole has maximum lifetime $ML = 0.4958$, with birth at $\epsilon = 0.6355$ and death at $\epsilon = 1.1312$.

2.4. Distance-To-Measure (DTM)-based filtration

While classical methods such as the Vietoris–Rips and Čech complexes are widely used, their direct reliance on pairwise distances makes them highly sensitive to noise and outliers. To address this limitation, we utilize the DTM function [6], a more robust alternative that evaluates the proximity of points to the data’s overall distribution rather than to individual data points. In the field of TDA, constructing a meaningful filtration from a point cloud is a critical first step.

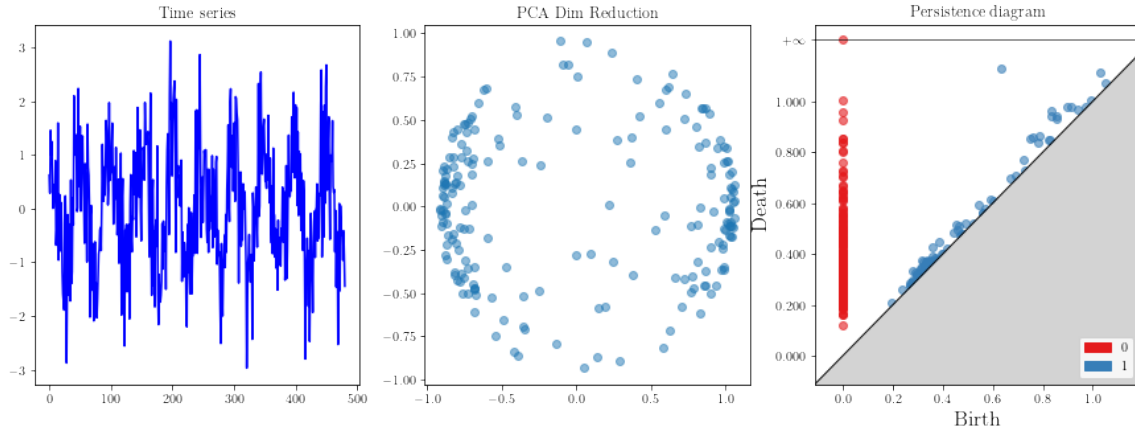


Figure 1. Example of PD construction using 200 points with H_0 (connected component) and H_1 (hole) using the Vietoris–Rips filtration.

Definition 2.2. Let μ be a probability measure over $E = \mathbb{R}^D$ and let $m \in [0, 1)$ be a user-defined mass parameter. For any point $x \in \mathbb{R}^D$, define

$$\delta_{\mu,m}(x) = \inf\{r \geq 0 : \mu(\bar{B}(x, r)) > m\}, \quad (2.6)$$

where $\bar{B}(x, r)$ is the closed ball of radius r centered at x . The DTM function $d_{\mu,m} : E \rightarrow \mathbb{R}$ is

$$d_{\mu,m}(x) = \left(\frac{1}{m} \int_0^m \delta_{\mu,t}^2(x) dt \right)^{1/2} \quad (2.7)$$

The DTM provides a geometrically intuitive and stable measure of “distance” from the underlying data distribution. The function $\delta_{\mu,m}(x)$ can be understood as a query of local density; its value is small if x is in a region of high mass concentration and large if x is in a sparse region. The final DTM function $d_{\mu,m}(x)$ enhances this by introducing a smoothing effect via the integral formulation. Outliers, which constitute a negligible fraction of the total mass of μ , do not significantly influence the DTM values of points within the main data structure. A sublevel set filtration is then constructed, yielding a more reliable persistence diagram.

3. Persistence vector representations

Persistence diagrams encode rich topological information but reside in a non-Euclidean space, precluding direct use with standard statistical or machine learning pipelines [37, 55]. To address this, we employ six vectorization strategies, each mapping dgm_n to a fixed-dimensional Euclidean representation. Table 1 provides a one-glance overview; readers familiar with TDA may proceed directly to Section 4.

3.1. Persistence landscape

PL is a handy yet stable tool to extract the PH summaries [10]. It is the representation obtained from dgm_n and embedded in the form of Banach space $L^P(\mathbb{N} \times \mathbb{R})$. PL is a collection of piecewise

Table 1. Summary of the six persistence vector (PV) representations. All six are compressed to a scalar via the L^1 -norm for periodicity ranking.

Representation	Sym	Output	Key property
Persistence Landscape	PL	Function vector	Banach space stability
Persistence Image	PI	2-D pixel grid	Smooth kernel surface
Persistence Entropy	PE	Scalar	Lifetime disorder
Betti Curve	BC	Integer sequence	Scale-dependent count
Persistence Lifetime Curve	PLC	Function of t	Aggregated lifespan
Persistence Statistics	PS	$\mathbb{R}^{37}$	Comprehensive moments

linear functions in the 2D space defined by the birth-death coordinates from dgm_n . For each birth-death point $(b_i, d_i) \in \text{dgm}_n$, the piecewise function is defined as,

$$\Lambda_{(b_i, d_i)}(x) = \begin{cases} x - b_i, & \text{if } x \in (b_i, \frac{b_i + d_i}{2}), \\ -x + d_i, & \text{if } x \in (\frac{b_i + d_i}{2}, d_i), \\ 0, & \text{if } x \notin (b_i, d_i). \end{cases} \quad (3.1)$$

The corresponding sequence of functions of PL $\lambda_j : \mathbb{R} \rightarrow [0, +\infty]$ is given by [21, 22]

$$\lambda_j(x) = j^{\text{th}} \max\{P_{b_i, d_i}(x) \mid (b_i, d_i) \in \text{dgm}_n\} \quad (3.2)$$

where $j \in \mathbb{N}$ and $j_{\max}$ is denoted as the j th largest value of a function. The setting of $\lambda_j(x) = 0$ can be used if the j th largest does not exist. Let $\Lambda = \{\lambda_j\}_{j=1}^n$ be a finite set of n PL. The vector space Λ can be imparted by a norm by a function $\|\cdot\|_p : \Lambda \rightarrow \mathbb{R}$ which is defined as L^p -norm as follows:

$$\|\lambda\|_p = \left[\sum_{k=1}^{\infty} \int |\lambda_k(t)|^p dt \right]^{1/p} \quad \text{for } 1 \leq p < \infty, \quad \|\lambda\|_{\infty} = \sup_{k, t} |\lambda_k(t)| \quad \text{for } p = \infty. \quad (3.3)$$

The difference in the L^p -norm values may be used to compute the distance between PL and PL' [10]:

$$d_p(PL, PL') = \|\lambda - \lambda'\|_p. \quad (3.4)$$

Figure 2 illustrated the representation of PL from PD using points from Figure 1 with two types of homology (H_0 and H_1).

3.2. Persistence images

Another way to represent the PD is using the Persistence Images (PI) [2]. Let B be a birth-death coordinates. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be a linear transformation $T(b, d) = (b, d - b)$ and let $T(B)$ be the transformed multiset in birth persistence coordinates where each point $(b, d) \in B$ corresponds to a point $(b, d - b) \in T(B)$. Let $\phi_u : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a differentiable probability distribution with

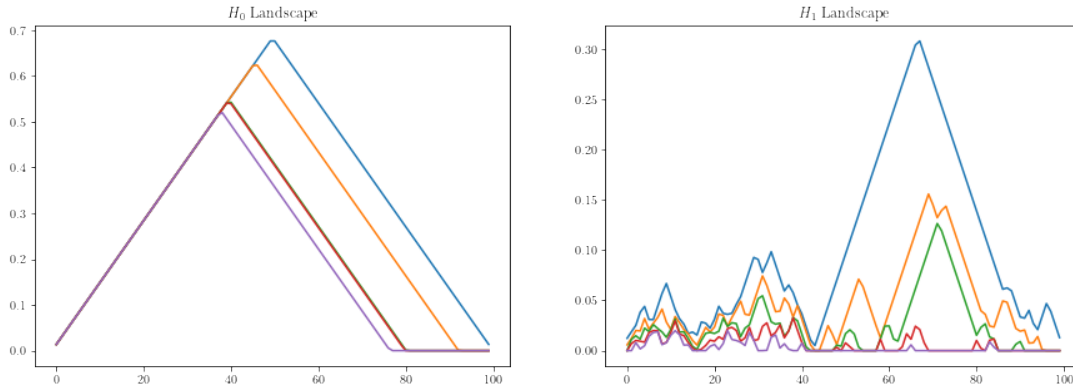


Figure 2. PL representation from PD in equation (2.5).

mean $\mu = (\mu_b, \mu_d) \in \mathbb{R}^2$. The normalized distribution selected is symmetric Gaussian $\phi_\mu = g_\mu$ with mean μ and variance σ^2 defined as

$$g_\mu(b, d) = \frac{1}{2\pi\sigma^2} e^{-[(b-\mu_b)^2 + (d-\mu_d)^2]/(2\sigma^2)} \quad (3.5)$$

where nonnegative weighting function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is fixed with zero along the horizontal axis, continuous and piecewise differentiable.

Definition 3.1. Let dgm_n be a PD, the corresponding persistence surface $\rho_{\text{dgm}_n} : \mathbb{R}^2 \rightarrow \mathbb{R}$ is the function

$$\rho_{\text{dgm}_n} = \sum_{\mu \in T(\text{dgm}_n)} f(\mu) \phi_\mu(n) \quad (3.6)$$

where f is the weighting function. Then, its persistence image is the collection of pixels p . For each pixel, integrate the surface over that denoted as

$$I(\rho_{\text{dgm}_n})_p = \iint_{(x,y) \in p} \rho_{\text{dgm}_n}(x, y) \, dy \, dx, \quad (3.7)$$

where x denotes the birth coordinate and y denotes the persistence coordinate, $y = d - b$ [2].

The representation of PD to PI requires three important parameters: Resolution, Gaussian distribution, and weighting function. We use a piecewise linear weighting function $w_b(t)$ associated with the persistence coordinate y :

$$w_b(t) = \begin{cases} 0, & \text{if } t \leq 0, \\ t/b, & \text{if } 0 < t < b, \text{ and} \\ 1, & \text{if } t \geq b. \end{cases} \quad (3.8)$$

Figure 3 illustrates the PI representation of PD calculated from equation (2.5).

3.3. Persistence entropy

Entropy is a measurable physical quantity frequently connected to a condition of disorder, unpredictability, or uncertainty [58]. There are situations when we would choose to summarize persistent homology with just a single number, such as using persistent entropy.

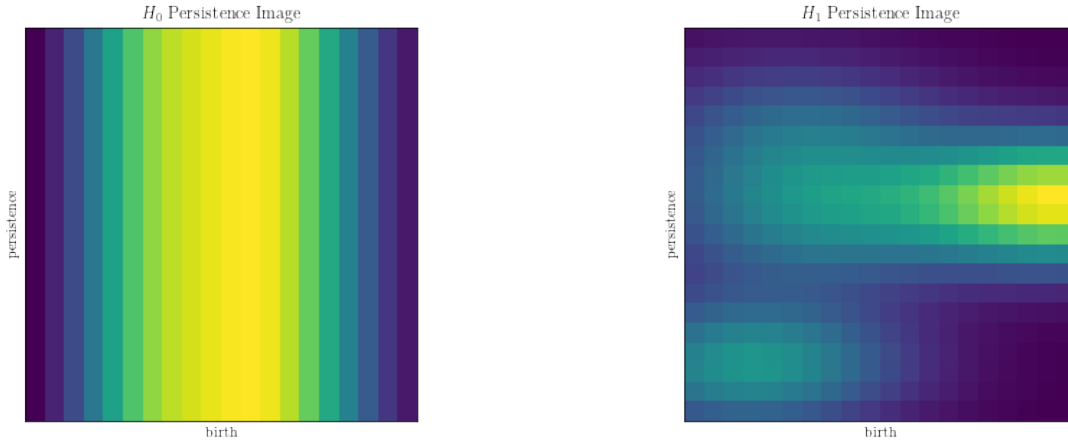


Figure 3. Persistent Image (PI) representation from PD in equation (2.5). Left is PI for H_0 and right for H_1 .

Definition 3.2. The persistent entropy $E(A)$ of a persistence diagram $A = \{(b_i, d_i)\}$ is defined as:

$$E(A) = - \sum_{i=1}^n \frac{\ell_i}{\mathcal{L}} \log\left(\frac{\ell_i}{\mathcal{L}}\right) \quad (3.9)$$

where ℓ_i is the lifetime defined as $\ell_i = d_i - b_i$ and $\mathcal{L} = \sum_{i=1}^n \ell_i$ is the sum of lifetimes for $1 \leq i \leq n$ [7].

3.4. Betti curve

A Betti curve (or Betti sequence) tracks the number of n -dimensional holes ($n = 0, 1, 2, \dots$) in a filtration as a function of the radius ϵ . To generate a finite-length vector, we limit the radius to $0 < \epsilon < E$, where E is the last death value that forms the simplicial complex.

Definition 3.3. Let B_n denote the Betti number of n -dimensional holes (i.e., B_0 is the number of connected components, B_1 is the number of loops, B_2 is the number of voids, etc.). $B_n(\epsilon)$ is defined as the rank of H_n homology that evolves with the radius $0 < \epsilon < E$ [56].

We then construct a sequence of vectors $BS_0, BS_1, \dots, BS_n$ for increasing radii. The Betti sequence is defined as

$$BS_n(i) = B_n\left(i \times \frac{E}{m_n}\right), \quad (3.10)$$

where m_n is the given vector size of BS_n . This sequence can be treated as a vector of Betti numbers sampled at fixed radii, and an L^p norm can be calculated using equation (3.3). Figure 4 demonstrates the Betti numbers of H_0 and H_1 for the example in equation (2.5).

3.5. Persistence lifetime curve

Chung and Lawson [15] proposed the normalized lifespan Persistence Curve (PC) with a stable persistence summary for PD.

Definition 3.4. Let dgm_n be the set of all PD, and let $\mathcal{A}$ be a summary operator (such as summation). The persistence curve P is a function that maps a PD to a real number summary

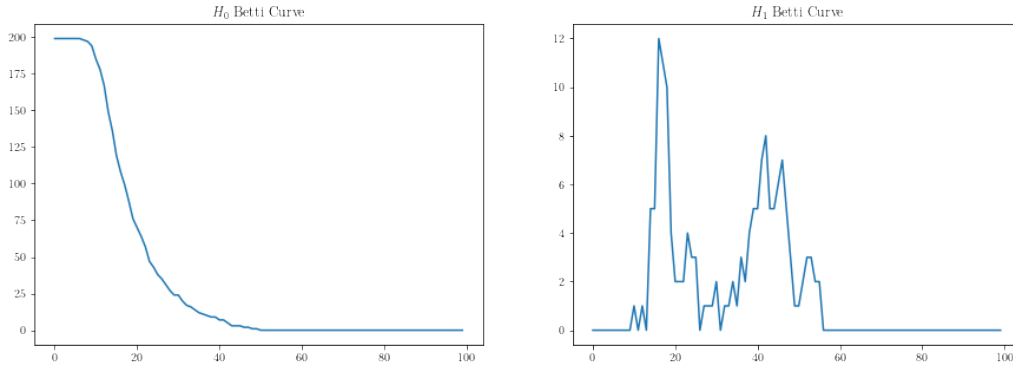


Figure 4. Betti curve representation from PD in equation (2.5).

based on the chosen function ψ and summary operator $\mathcal{A}$. The persistence curve is defined as,

$$P(\text{dgm}_n, \psi, \mathcal{A})(t) = \mathcal{A}([\psi(b, d) \mid (b, d) \in \text{dgm}_{n,t}]), \quad (3.11)$$

where $\text{dgm}_{n,t}$ is the subset of points in the dgm_n at time t , and $\psi(b, d)$ is a function of the birth b and death d times of points in the diagram. For the **lifespan curve**, the function $\psi(b, d) = d - b$ (i.e., the lifespan of a point), and the summary operator is $\mathcal{A} = \sum$ (sum of lifespans). Thus, the lifespan curve becomes

$$P\left(\text{dgm}_n, \psi, \sum\right)(t) = \sum(d - b \mid (b, d) \in \text{dgm}_{n,t}). \quad (3.12)$$

Figure 5 demonstrates the lifetime curve of H_0 and H_1 based on the example equation (2.5).

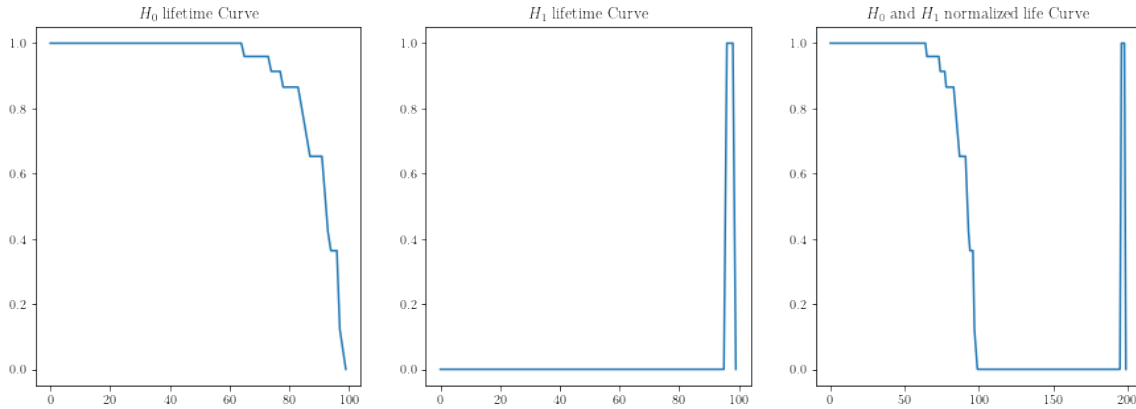


Figure 5. Lifetime curve (LC) representation from PD in equation (2.5). Left is the LC for H_0 , middle for H_1 and right the concatenation of H_0 and H_1 .

3.6. Persistence statistics

The persistence statistics (PS) vector serves as a compact yet rich representation of the topological characteristics of the data. It is especially valuable for comparing PD across different datasets and for identifying patterns and structural variations within the data [4].

Definition 3.5. Let $S \in \{\text{birth, death, midpoint, lifespan}\}$, representing the different sets of interest. The persistence statistics vector PS is defined as,

$$PS = [\mu_S, \sigma_S, \min_S, P_{10}^S, P_{25}^S, P_{50}^S, P_{75}^S, P_{90}^S, \max_S, |c|], \quad (3.13)$$

where μ_S represents the mean value for set S , while σ_S denotes the standard deviation. The minimum and maximum values within S are given by $\min_S$ and $\max_S$, respectively. Additionally, the percentiles for S , including the 10th, 25th, 50th (median), 75th, and 90th percentiles, are represented by $P_{10}^S, P_{25}^S, P_{50}^S, P_{75}^S$, and P_{90}^S . Finally, $|c|$ refers to the cardinality of the persistence diagram.

This vector is composed of 9 statistical descriptors for each of the four sets S (birth, death, midpoint, lifespan), with a total of 36 components. Additionally, the cardinality $|c|$ of the persistence diagram dgm_n is appended as the final component. The cardinality $|c|$ represents the number of birth-death pairs as multiset in dgm_n , yielding a vector PS that resides in a 37-dimensional real space, $\mathbb{R}^{37}$.

3.7. L^p -norm

In this research, the L^p -norm is employed to quantify and rank periodicity across various data representations. The L^p -norm defines a class of functions that measure the magnitude of a vector $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$:

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}. \quad (3.14)$$

We specifically select the L^1 -norm ($p = 1$), also known as the Manhattan norm, for its distinct advantages in robustness, feature preservation, and interpretability over higher-order norms like the L^2 (Euclidean) norm. The choice of the L^1 -norm is primarily motivated by its robustness to outliers. The L^2 -norm squares deviation terms, causing large, anomalous values to disproportionately influence the aggregate measure. In contrast, the L^1 -norm's linear summation of absolute differences mitigates the effect of such outliers, yielding a more stable and reliable metric in the presence of noise or sporadic events [3].

4. Periodicity Score Using PL

One issue with persistence diagrams is that they are notoriously difficult to use directly as data summaries. Although PDs are stable [15], the set of birth-death points in a PD lies in a non-Euclidean space, making standard statistical summaries or averages not directly applicable. To address this, we use PL representations to compute statistical summaries.

4.1. Persistence Landscape Score (PLS)

The PLS formula introduces a novel approach to quantifying persistence landscape scores, focusing on the first layer of the landscape. The PLS score is formulated as

$$PLS = \frac{\|\lambda^{(1)}\|}{\sum_{j=1}^n \|\lambda^{(j)}\|} + \|\lambda^{(1)}\|, \quad (4.1)$$

where $\lambda^{(j)}$ denotes the j -th layer of the persistence landscape, and $j = 1, 2, \dots, n$ represents the number of layers considered. Each $\lambda^{(j)}$ is a piecewise-linear function derived from the persistence diagram that encodes the topological features at different scales. The term $\|\lambda^{(j)}\|$ represents the norm (typically the L^1 norm) of the landscape function $\lambda^{(j)}$, and is used to quantify the contribution of each layer.

The first term in equation (4.1) captures the relative dominance of the first landscape layer. A value close to 1 suggests that the majority of the topological signal is concentrated in the first layer, which is indicative of strong periodicity in the data. The second term, $\|\lambda^{(1)}\|$, is included to distinguish between periodic signals that may have similar structure but differ in intensity or scale.

If only the normalized first-layer ratio $\frac{\|\lambda^{(1)}\|}{\sum_{j=1}^n \|\lambda^{(j)}\|}$ is used, the formulation becomes equivalent to the SW1PerS (Sliding Window 1-Persistence score [44]), a known method for detecting periodicity based solely on the maximum persistence. This relationship emphasizes that PLS generalizes SW1PerS by adding sensitivity to the absolute strength of topological features.

5. Synthetic data representing time series

The simulation was carried out in Python on a system with an Intel Core i7-10700 CPU and 16 GB RAM. In this study, we conducted a comprehensive evaluation of the influence of SW1PerS techniques on persistence vector (PV) computation, utilizing a total of ten synthetic time series datasets. Descriptions of these datasets, which consist of synthetic data sourced from [43], are shown in Table 2. To test for L -periodicity, we include more than one cycle in each signal. Here, L -periodicity refers to the number of periods where $L \in \mathbb{N}$. For our experiments, each signal covers 200 time units so that each signal spans two cycles (i.e., $L = 2$).

We added Gaussian noise with various standard deviations (SDs) to test the robustness of our method. Gaussian noise was applied at four levels with SD equal to 0, 10, 25, and 50, which is sufficient to simulate significant outlier contamination. Another setting was introduced to test the stability of the periodicity measure: Random missing values in the signals. We randomly removed points to create missing data at ten different levels, with missing value percentages of 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50%.

6. Methodology for periodicity analysis via persistent homology

In this section, we present a unified framework for analyzing periodicity in ten synthetic time series datasets $f_1, f_2, \dots, f_{10}$ from Table 2 via PH, as visualized in Figure 6. Each dataset f_k (for $k = 1, 2, \dots, 10$) comprises 200 observations $(y_1^{(k)}, \dots, y_{200}^{(k)})$ recorded at time points $(t_1^{(k)}, \dots, t_{200}^{(k)})$. Each series is then standardized by computing:

$$\bar{y}^{(k)}(t) = \frac{\tilde{y}^{(k)}(t) - \mu_k}{\sigma_k}, \quad (6.1)$$

where $\mu_k = \frac{1}{200} \sum_{i=1}^{200} \tilde{y}^{(k)}(t_i)$ and $\sigma_k = \sqrt{\sum_{i=1}^{200} (\tilde{y}^{(k)}(t_i) - \mu_k)^2}$. We then enhance smoothness by fitting a cubic spline $g_k : [0, 2\pi] \rightarrow \mathbb{R}$ satisfying the boundary conditions $g_k(0) = \bar{y}^{(k)}(t_1^{(k)})$ and $g_k(2\pi) = \bar{y}^{(k)}(t_{200}^{(k)})$. To capture quasiperiodic dynamics, we employ Takens' embedding with

Table 2. Synthetic time series formulas used to generate datasets. Amplitude ($A = 50$), period ($T = 100$) and phase shift ($\omega = 50$) throughout, following [43].

Name	Formula	Plot
cos	$f_1 = A \cos(\frac{2\pi}{T}t - \frac{2\pi}{T}\omega) + A$	
cos2	$T' = 0.3333T$ $A' = 0.5A$ $\omega' = (\omega + (0.25T'))\%T$ $f_2 = A \cos(\frac{2\pi}{T}(t - \omega)) + A' \cos(\frac{2\pi}{T'}(t - \omega')) + A$	
peak	$P = 20$ $f_3 = A(-1 + 2 \cos(\frac{\pi}{T}t - \frac{\pi}{T}\omega) ^P) + A$	
peak2	$P_1 = 10$ $P_2 = 40$ $A' = 0.7A$ $\omega' = \omega + (0.5T)$ $f_4 = A(-1 + 2 \cos(\frac{\pi}{T}(t - \omega)) ^{P_1}) + A'(-1 + 2 \cos(\frac{\pi}{T'}(t - \omega')) ^{P_2})$	
trendE	$\gamma = 0.027$ $f_5 = A \cos(\frac{2\pi}{T}(t - \omega)) + A + \exp(\gamma t)$	
trendL	$\Lambda = 0.5$ $f_6 = A \cos(\frac{2\pi}{T}t - \frac{2\pi}{T}\omega) + (\Lambda t) + A$	
damp	$k = 0.01$ $f_7 = A \cos(\frac{2\pi}{T}t - \frac{2\pi}{T}\omega) \exp(-kt) + A$	
saw	$f_8 = 2A(\frac{(t-\omega)\%T}{T-1})$	
square	$f_9 = 2A \text{round}(\frac{(t-\omega)\%T}{T-1})$	
contract	$f_{10} = A \cos(\frac{2\pi}{T}(\frac{t^2}{T}) - \omega) + A$	

fixed dimension $D = 14$ and delay τ . The resulting point cloud in $\mathbb{R}^{15}$ is defined as:

$$\chi_k = \{(g_k(i), g_k(i + \tau), \dots, g_k(i + 14\tau)) : i = 1, \dots, s_k\}. \quad (6.2)$$

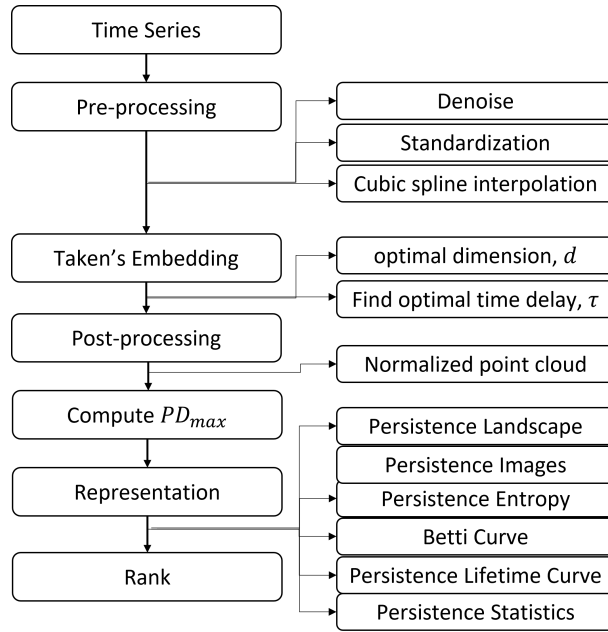


Figure 6. Workflow to measure the periodicity score.

Each point cloud χ_k is then mean-centered and normalized:

$$\hat{\chi}_k = \left\{ \frac{x - \text{mean}(\chi_k)}{\|x - \text{mean}(\chi_k)\|} : x \in \chi_k \right\}. \quad (6.3)$$

We determine an optimal sliding window size w for each dataset by two complementary approaches. First, an empirical Maximum Lifetime (ML) search varies w over integer values in $[1, N]$, computes the persistence diagram $\text{dgm}_k(w)$ for the normalized cloud $\hat{\chi}_k(w)$, and selects:

$$w_{\text{emp}}^{(k)} = \arg \max_{1 \leq w \leq N} ML_k(w), \quad ML_k(w) = \max\{d - b : (b, d) \in \text{dgm}_k(w)\}. \quad (6.4)$$

Second, following [44], a theoretical window is given by:

$$w_{\text{rad}} = \frac{D}{D+1} \frac{\pi}{L}, \quad D = 14, \quad L = 2. \quad (6.5)$$

Converting to sample units for a 200-point series via multiplication by $200/\pi$ yields:

$$w_{\text{theo}}^{(k)} = \left(\frac{D}{D+1} \frac{\pi}{L} \right) \times \frac{200}{\pi} = \frac{14}{15} \frac{\pi}{2} \frac{200}{\pi} \approx 93. \quad (6.6)$$

Tables 3 and 4 report both $w_{\text{emp}}^{(k)}$ and $w_{\text{theo}}^{(k)}$ for $k = 1, \dots, 10$. Using these window sizes, we build two versions of the PD for each dataset:

$$PD_{\text{emp}}^{(k)} = \text{dgm}_k(w_{\text{emp}}^{(k)}), \quad PD_{\text{theo}}^{(k)} = \text{dgm}_k(93). \quad (6.7)$$

From each diagram, we extract the first homology group H_1 to identify dominant loop structures, then vectorize these into $PV_{\text{emp}}^{(k)}$ and $PV_{\text{theo}}^{(k)}$. Each persistence diagram $PD_{\text{max}}^{(k)}$ is represented as PL, PI, Entropy, Betti Curves, PLC, and PS, providing diverse perspectives on topological features.

Table 3. Maximum lifetime with dimension $D = 14$ and window size $w = 93$ for all datasets with $L = 2$ (no noise).

Label	Type	Period	Quasiperiod	Max Lifetime
y1	cos	2	0	1.6984
y2	cos2	2	4	1.5005
y3	peak	2	0	1.5662
y4	peak2	2	2	1.5546
y5	trendE	2	0	0.6774
y6	trendL	2	0	1.6259
y7	damp	2	0	1.6924
y8	saw	2	0	1.0671
y9	square	2	0	1.3261
y10	contract	4	0	0.2459

Table 4. Maximum lifetime with dimension $D = 14$ and optimal window size for all datasets with $L = 2$ (no noise).

Label	Type	Quasiperiod	Max Lifetime	Window Size
y1	cos	0	1.6984	94
y2	cos2	4	1.4975	94
y3	peak	0	1.5644	93
y4	peak2	2	1.5546	93
y5	trendE	0	1.5960	58
y6	trendL	0	1.6661	74
y7	damp	0	1.6941	98
y8	saw	0	1.0902	98
y9	square	0	1.3425	98
y10	contract	2	1.5686	19

7. Persistence vector results on synthetic data

The objective of this study was to perform a ranking analysis using PV techniques on synthetic data, aiming to gain insights into the topological characteristics and structures present in the datasets. SW1PerS [43, 44] is formulated as:

$$SW1PerS = 1 - \frac{d^n - b^m}{3^{n/2}}, \quad (7.1)$$

where $n \geq m$. For comparison, we use the same values of n and m proposed in SW1PerS:

$$SW1 = 1 - \frac{d - b}{\sqrt{3}}, \quad (n = 1, m = 1), \quad (7.2)$$

$$SW2 = 1 - \frac{d^2 - b^2}{3}, \quad (n = 2, m = 2). \quad (7.3)$$

7.1. Sliding window persistence and embedding dimension analysis

This section details the comparative analysis performed to identify the most effective method for quantifying periodicity in time-series data. We evaluated eight distinct PV representations by ranking their performance on ten synthetic datasets under three separate experimental conditions:

1. **Optimal window analysis:** We first determined the optimal window size for each dataset by analyzing the persistence of topological features, while keeping the embedding dimension fixed at $D = 14$.
2. **Fixed window analysis:** For a controlled comparison, we analyzed all datasets using a fixed window size of $w = 93$, derived from the formula proposed in [44] for $D = 14$.
3. **Nearest-neighbor analysis:** We empirically determined the optimal embedding dimension D and time delay τ for each dataset using nearest-neighbor and mutual information analysis, respectively. The analysis was performed using the average parameters obtained ($D = 7$, $\tau = 8$). The result is illustrated in Table 7.

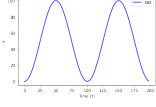
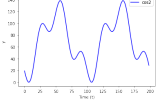
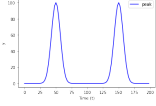
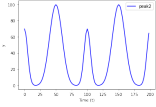
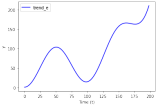
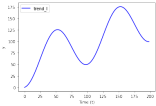
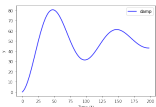
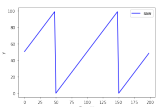
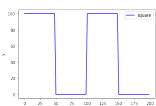
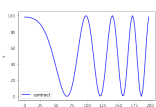
Comprehensive summaries of the PDs are presented in Tables 5 and 6. To facilitate a direct quantitative comparison, each PV was compressed into a single scalar value using the L^1 -norm. This allowed us to rank each dataset's periodicity, where a lower score indicates a stronger periodic signal. To ensure a fair comparison, all PV scores were normalized (except for SW1, SW2, and PLS). The detailed rankings are presented in Tables 8, 9, and 10, and the corresponding average ranks are summarized in Table 11.

7.1.1. Interpretation of inter-method rank inconsistencies and goodness of rankings

The rank inconsistencies observed across Tables 8–10 reflect the complementary geometric information each PV captures, rather than errors in the framework:

- **Entropy** (H_1) assigns score 0.000 to most purely periodic signals (cos, damp, trendL, peak, cos2) because their persistence diagrams contain a single dominant birth–death pair with zero distributional uncertainty. Entropy is therefore insensitive to the *strength* of the dominant feature and is better suited to detecting distributional differences between multi-component signals.
- **Persistence image** (H_1) diverges from PL and PLC because PI integrates a kernel-smoothed surface over the entire PD plane, weighted by a linear function of persistence. This makes PI highly sensitive to the *spatial distribution* of birth–death pairs, explaining its superior discrimination on multi-component signals (peak2, contract) but its instability under noise.
- **Betti curve** (H_1) produces step-function summaries indexed by filtration radius, capturing the *count* of active loops at each scale rather than their lifetimes. Its values in Table 5 (ranging 83–121) reflect integrated loop counts across all scales, which are geometrically distinct from lifetime-based summaries.
- **Ranking disagreement across experimental conditions** (optimal w vs. fixed $w = 93$ vs. NN parameters) arises because different window sizes alter the phase-space geometry, shifting which topological features are dominant. This sensitivity confirms that parameter selection is a primary source of variation and motivates the empirical optimization of w proposed in this work.

Table 5. The PD summaries of H_1 for all datasets without noise with $D = 14$ and optimal window.

Type	Graph	Entropy, H_1	Betti, H_1, L^1	PL, H_1, L^1	PI, H_1, L^1	LC, H_1, L^1	PS, H_1, L^1
cos		0.00	99.00	602.45	737.91	97.00	26.14
cos2		0.00	99.00	529.96	571.02	95.00	23.45
peak		0.00	100.00	553.64	623.18	95.00	24.61
peak2		0.53	121.00	550.24	148.81	94.00	27.01
trendE		0.11	102.00	565.10	218.93	41.00	16.15
trendL		0.00	100.00	589.65	706.89	96.00	25.96
damp		0.00	100.00	600.07	732.09	96.00	26.17
saw		0.00	99.00	385.83	302.66	65.00	26.09
square		0.02	83.00	392.04	202.56	80.00	8.97
contract		1.23	118.00	523.46	252.51	12.93	7.41

To quantify the overall agreement among the eight PV methods, we compute Kendall's coefficient of concordance W [33]:

$$W = \frac{12S}{k^2(n^3 - n)}, \quad (7.4)$$

where S is the sum of squared deviations of the rank sums from their mean, k is the number of ranking methods ($k = 8$), and n is the number of datasets ($n = 10$). For the optimal- w experiment (Table 8), we obtain $W = 0.366$ ($p < 0.001$, permutation test with 10,000 replicates), indicating *substantial* inter-method concordance. For the fixed- w and NN conditions, $W = 0.3746$ and $W = 0.341$, respectively, all statistically significant at the 0.001 level. These values confirm that, despite local disagreements, the eight methods reach broad consensus on the

Table 6. The PD summaries of H_0 for all datasets without noise with $D = 14$.

Type	Entropy, H_0	Betti, H_0, L^1	PI H_0, L^1	PL H_0, L^1	LC H_0, L^1	PS H_0, L^1
cos	8.40	199.00	12667.48	59.65	1.97	0.60
cos2	9.67	199.00	12665.61	96.73	3.85	0.96
peak	4.42	199.00	12660.76	111.98	3.90	1.12
peak2	5.48	199.00	12659.00	124.54	4.82	1.24
trendE	57.02	199.00	12663.05	141.21	2.52	0.88
trendL	39.53	199.00	12665.24	94.28	2.41	0.65
damp	27.25	199.00	12667.02	67.13	2.41	0.60
saw	23.54	199.00	12153.06	895.43	21.15	2.87
square	32.73	199.00	12530.86	517.83	12.20	2.44
contract	51.96	199.00	12641.11	266.41	6.20	1.47

Table 7. Nearest neighbor and mutual information.

Label	Type	Dimension	Time Delay	Window Size
y1	cos	5	9	45
y2	cos2	12	5	60
y3	peak	10	9	90
y4	peak2	12	9	108
y5	trendE	2	8	16
y6	trendL	5	9	45
y7	damp	7	9	63
y8	saw	3	9	27
y9	square	2	9	18
y10	contract	14	7	98

periodicity ordering of the ten synthetic datasets.

7.2. Rank analysis based on robustness to noise

The synthetic data was subjected to Gaussian noise at three distinct levels (SD=10, 25, 50) in order to replicate real-world scenarios where data could become distorted or corrupted by numerous sources of noise. To track the evolution of the periodicity value with noise, we utilized fixed $D = 14$ and $w = 93$ to embed data. Through the rank analysis, we were able to evaluate the periodicity's robustness by monitoring how their ranks shifted with increasing noise levels.

Figure 7 presents several rank plots that compare various PV features across different levels of Gaussian noise (0%, 10%, 25%, 50%) applied to the data, with an embedding dimension $D = 14$ and a window size $w = 93$. The persistence features ranked in this figure include PLS, SW1, SW2, Entropy, Betti numbers, PI, PLC and PS(H_0/H_1).

The analysis of the rank plots shows that different features exhibit varying levels of stability

Table 8. The rank result for all PV without noise with $D = 14$ and optimal window. Lower score = more periodic.

SW1	SW2	PLS	Entropy, H_1	Betti, H_1	PI, H_1	PLC, H_0H_1	PS, H_0/H_1
							
0.0194	0.0003	0.000	0.0000	0.0000	0.0000	0.0000	0.0000
							
0.0219	0.0004	0.0110	0.0000	0.4211	0.0099	0.0000	0.0000
							
0.0381	0.0103	0.0591	0.0000	0.4211	0.0527	0.0003	0.0089
							
0.0786	0.0151	0.1784	0.0000	0.4211	0.1947	0.0027	0.0733
							
0.0944	0.0646	0.2253	0.0000	0.4474	0.2833	0.0175	0.0914
							
0.0968	0.1181	0.2416	0.0000	0.4474	0.7388	0.0325	0.0931
							
0.1024	0.1212	0.3346	0.0147	0.4474	0.8240	0.3231	0.1275
							
0.1354	0.1364	0.3720	0.0871	0.5000	0.8810	0.4737	0.3495
							
0.2249	0.1979	0.9713	0.4330	0.9211	0.9088	0.8312	0.7036
							
0.3706	0.2105	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

as noise increases. SW2 and PLS remain relatively stable, making them reliable even when the data contains moderate noise. PI, PLC and PS show intermediate behavior, whereas Entropy and Betti are comparatively more sensitive, exhibiting the largest rank shifts at higher noise

Table 9. The rank result for all PV without noise with $D = 14$ and $w = 93$.

SW1	SW2	PLS	Entropy, H_1	Betti, H_1	PI, H_1	PLC, H_0H_1	PS, H_0/H_1
							
0.0191	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
							
0.0226	0.0005	0.0049	0.0000	0.0000	0.0141	0.0009	0.0166
							
0.0597	0.0654	0.0326	0.0000	0.0000	0.0930	0.0014	0.0251
							
0.0968	0.1181	0.0573	0.0000	0.0000	0.1606	0.0018	0.0376
							
0.1024	0.1212	0.0621	0.0000	0.0088	0.2321	0.0070	0.2179
							
0.1372	0.1429	0.0846	0.0000	0.0088	0.4069	0.0070	0.2598
							
0.2354	0.2010	0.1568	0.0000	0.0088	0.6268	0.0848	0.2654
							
0.3849	0.2242	0.2641	0.0000	0.0088	0.8228	0.1606	0.3008
							
0.6020	0.4173	0.4254	0.2997	0.0088	0.8679	0.6945	0.8245
							
0.8530	0.7971	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

levels. As these rank plots use a fixed window ($w = 93$), the formal noise-stability ranking, computed under the optimal window, is reported via the Rank Instability Index in Section 8 (Table 12).

In terms of distinguishing periodic data, PS can effectively capture the recurring patterns in

Table 10. The rank result for all PV without noise with $d = 7$ and $\tau = 8$.

SW1	SW2	PLS	Entropy, H_1	Betti, H_1	PI, H_1	PLC, H_0H_1	PS, H_0/H_1
							
0.0399	0.0017	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
							
0.0464	0.0026	0.0277	0.0000	0.1575	0.2095	0.0266	0.0966
							
0.1365	0.1000	0.1217	0.0000	0.1575	0.2950	0.0363	0.1317
							
0.1792	0.2035	0.1755	0.0000	0.1575	0.6797	0.0688	0.1340
							
0.2172	0.2102	0.3623	0.0000	0.1575	0.7015	0.2046	0.1911
							
0.3261	0.2208	0.4838	0.0016	0.1654	0.7584	0.3135	0.3069
							
0.4240	0.2469	0.5096	0.0481	0.1654	0.8025	0.3320	0.6632
							
0.4671	0.5034	0.5381	0.0574	0.1811	0.8609	0.3372	0.7318
							
0.5368	0.5070	0.8943	0.3048	0.5197	0.9215	0.5420	0.9006
							
0.6861	0.5344	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

periodic data. PS (H_0) captures the statistics related to connected components, reflecting how periodic patterns manifest in terms of data connectivity. PS (H_1) captures the persistence of loops, which is critical for understanding the cyclic nature of periodic data.

Table 11. The rank average comparison for optimum w , fixed $w = 93$ and average NN ($D = 7$) and mutual information $\tau = 8$.

Rank	Optimal w ($D = 14$)	Fixed w ($D = 14$)	Average NN and mutual info
1	cos	cos	cos
2	damp	damp	peak
3	trendL	trendL	cos2
4	peak	peak	damp
5	cos2	cos2	trendL
6	trendE	peak2	trendE
7	peak2	square	square
8	saw	saw	peak2
9	square	trendE	saw
10	contract	contract	contract

7.3. Rank analysis based on stability and robustness through randomly missing values

We employed a methodology whereby we introduced missing values at varying levels, ranging from 5% to 50%, into synthetic data in a random manner. Such missing values served to simulate scenarios where data points are incomplete or unavailable. Through the manipulation of missing data levels, we were able to evaluate the stability and robustness of the periodicity. Figure 8 depicts the specific window size utilized based on the appointed percentage of missing values.

After generating PV from synthetic data that included randomly missing values, a rank analysis was conducted to assess the stability and persistence of the periodicity patterns. This process was repeated three times with different sets of missing values, and the average rank scores are presented in Figure 9.

Figure 9 presents the rank plot of various PV features under increasing percentages of missing values, ranging from 0% to 50%, with an embedding dimension of $D = 14$. The PLS rank remains notably stable across all levels of missing values, and both SW1 and SW2 demonstrate consistent stability throughout, indicating that these features are resilient to missing data. The Betti rank demonstrates moderate stability. In contrast, the Entropy rank shows greater sensitivity beyond 20%, the PI and $PS(H_0H_1)$ ranks move appreciably at higher missing-data levels, and the PLC rank is the least stable of all features. These trends are quantified by the Mean Absolute Rank Deviation in Table 13, where SW2, PLS and SW1 attain the smallest deviations and PLC the largest.

8. Quantitative stability analysis

Sections 7.2 and 7.3 characterized stability through visual inspection of rank plots, which is necessarily subjective. Here we supplement those observations with two formal quantitative stability measures.

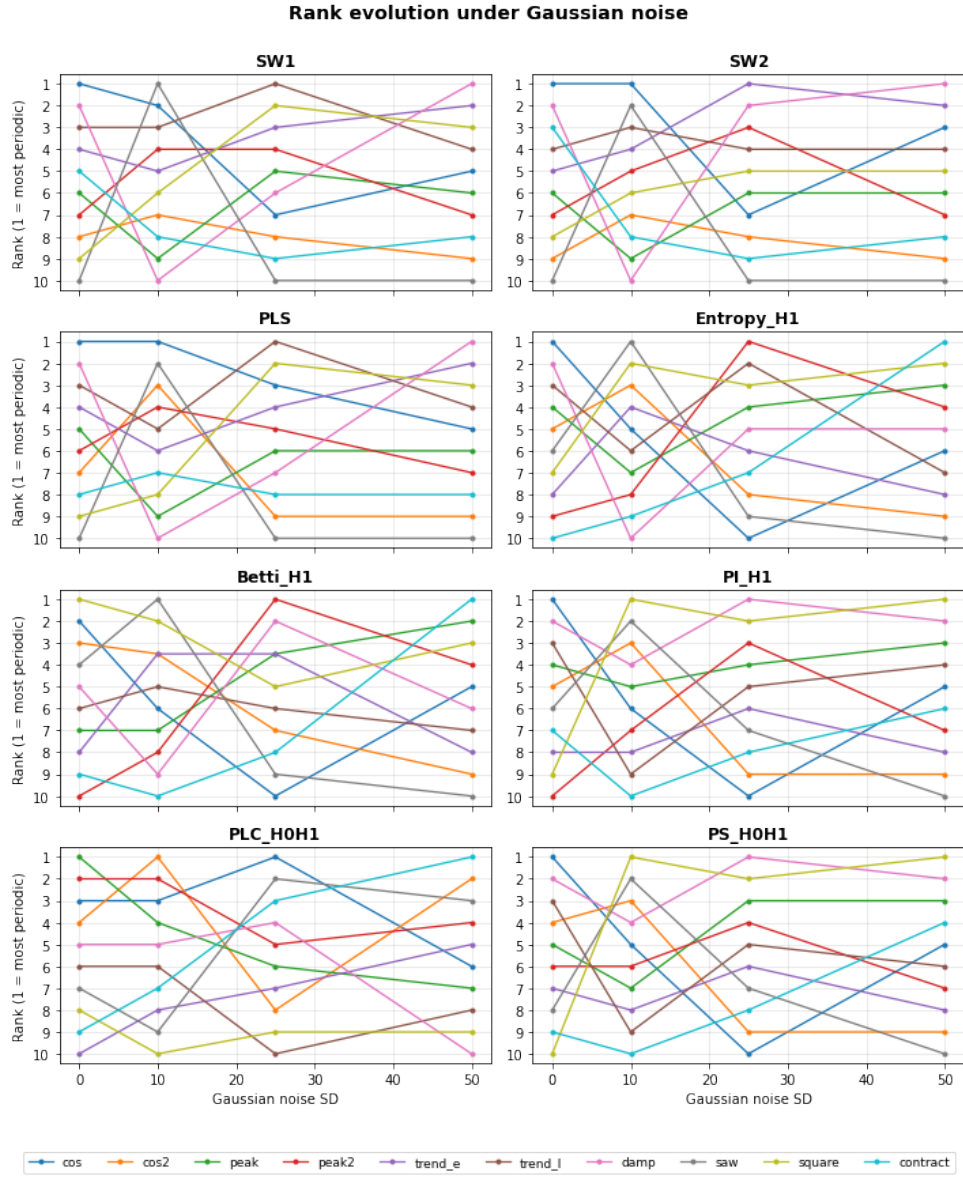


Figure 7. The rank plot of PV with Gaussian noise (SD(%)=0, 10, 25, 50) with embedding dimension $D = 14$ and window $w = 93$.

8.1. Rank instability index under Gaussian noise

Let $r_{k,\sigma}^{(m)}$ denote the normalized rank ($\in [0, 1]$) of dataset k under noise level $\sigma \in \{0, 10, 25, 50\}$ for PV method m . We define the *Rank Instability Index* (RII) as:

$$\text{RII}^{(m)} = \frac{1}{n} \sum_{k=1}^n \sqrt{\frac{1}{|\Sigma| - 1} \sum_{\sigma \in \Sigma} (r_{k,\sigma}^{(m)} - \bar{r}_k^{(m)})^2}, \quad (8.1)$$

where $\bar{r}_k^{(m)}$ is the mean rank of dataset k across noise levels and $|\Sigma| = 4$. A lower RII indicates a more noise-stable ranking. This formulation adapts the noise-based ranking stability paradigm

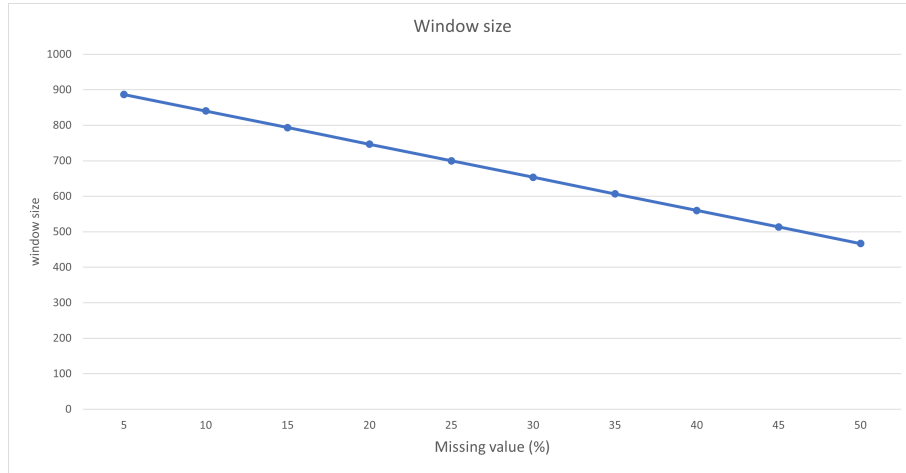


Figure 8. The fixed different optimal window size used for missing values data with embedding dimension $D = 14$.

of Altidor et al. [5] to persistence vector methods; the use of per-item rank variance across perturbation levels as a stability criterion is consistent with Kuncheva [35] and Berrar [8]. Table 12 reports the mean RII for each method, computed under the optimal window (the rank plots in Section 7.2 use a fixed window): SW2 achieves the lowest mean RII (0.243), indicating the most noise-stable ranking, closely followed by SW1 (0.257) and PLS (0.260). Entropy (H_1) is the least stable method (RII = 0.331). A Friedman test on the RII values yields $\chi^2(7) = 7.07$, $p = 0.42$; the differences in noise stability across PV methods are therefore modest and not statistically significant, consistent with the closely spaced RII values.

Table 12. Mean Rank Instability Index (RII $\pm$ s.e.) under Gaussian noise (optimal window, $D = 14$). Lower RII = more noise-stable [5, 8].

PV Method	Mean RII	Stability Rank
SW2	0.243 ± 0.043	1 (most stable)
SW1	0.257 ± 0.044	2
PLS	0.260 ± 0.044	3
PLC (H0H1)	0.264 ± 0.031	4
PI H1	0.268 ± 0.040	5
PS (H0/H1)	0.280 ± 0.041	6
Betti H1	0.326 ± 0.037	7
Entropy H1	0.331 ± 0.033	8 (least stable)

8.2. Mean absolute rank deviation under missing data

For the missing-data experiment, let $r_{k,\mu}^{(m)}$ be the normalized rank of dataset k at missing percentage $\mu \in \{0, 5, 10, \dots, 50\}$ for method m . The *Mean Absolute Rank Deviation* (MARD)

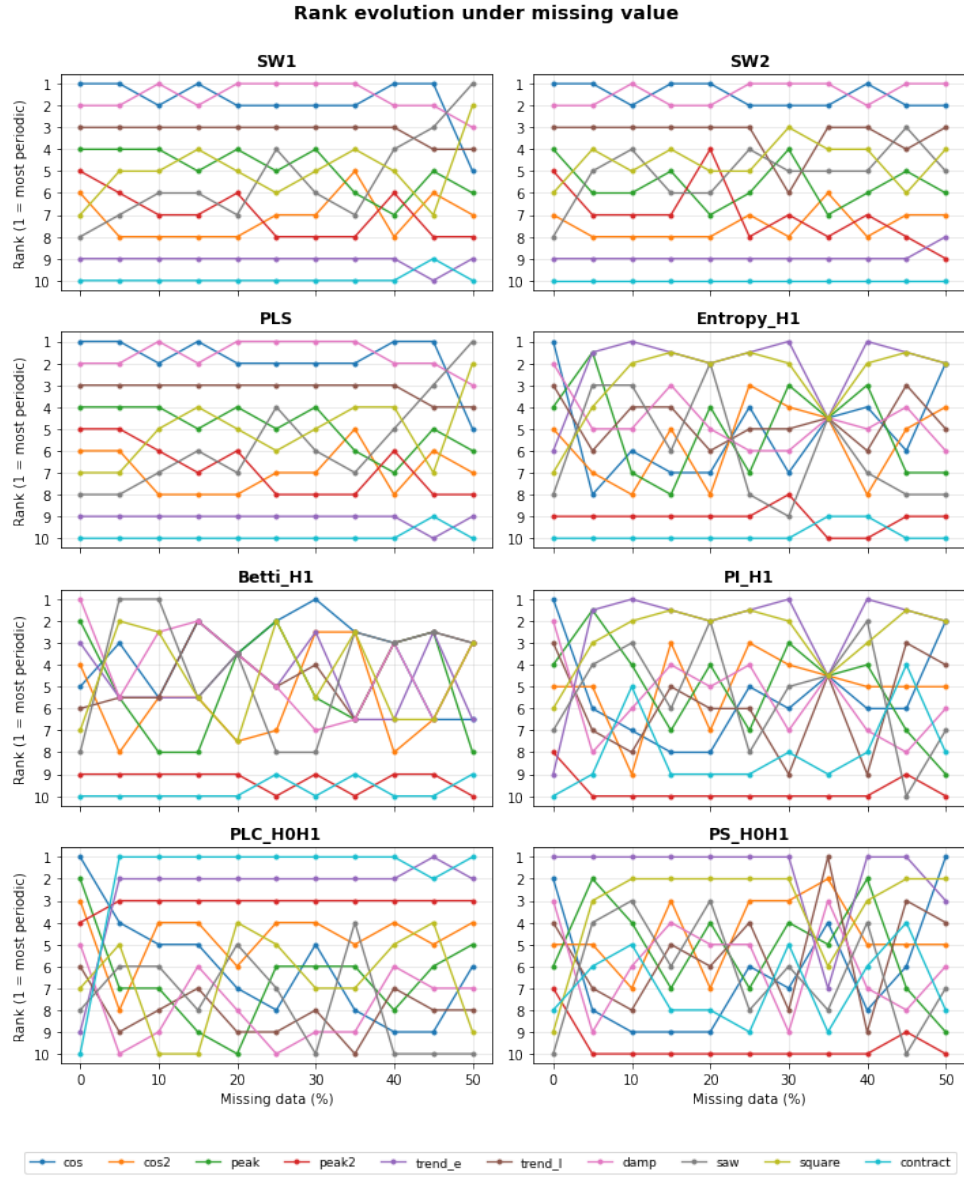


Figure 9. The rank plot PV with missing value ($mv(\%)=0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$) with embedding dimension $D = 14$.

is:

$$\text{MARD}^{(m)} = \frac{1}{n \cdot |\mathcal{M}|} \sum_{k=1}^n \sum_{\mu \in \mathcal{M}} |r_{k,\mu}^{(m)} - r_{k,0}^{(m)}|, \quad (8.2)$$

where $|\mathcal{M}| = 10$ and $r_{k,0}^{(m)}$ is the rank at 0% missing data. The mean absolute deviation from a reference ranking is a classical robust dispersion measure [48] and has been adopted to quantify ranking stability under data perturbations by Kalousis et al. [32]. Table 13 reports the MARD for each method. SW2 achieves the lowest MARD (0.117), indicating the most stable ranking under missing data, closely followed by PLS (0.127) and SW1 (0.132). PLC (H_0H_1) is the least stable method (MARD = 0.424). A Friedman test confirms significant inter-method differences

$$(\chi^2(7) = 23.98, p = 0.001).$$

Table 13. Mean Absolute Rank Deviation (MARD $\pm$ s.e.) under random missing values. Lower MARD = more stable under perturbation [32, 48].

PV Method	Mean MARD	Stability Rank
SW2	0.117 ± 0.006	1 (most stable)
PLS	0.127 ± 0.005	2
SW1	0.132 ± 0.005	3
Betti H_1	0.219 ± 0.005	4
Entropy H_1	0.280 ± 0.003	5
PS (H_0/H_1)	0.317 ± 0.008	6
PI H_1	0.353 ± 0.008	7
PLC (H_0H_1)	0.424 ± 0.006	8 (least stable)

9. DTM based filtration rank analysis

In this section, we propose a novel approach to enhance the rank analysis by incorporating the Distance to Measure (DTM) filtration, particularly in the context of noise and missing values. The presence of missing values in real-world data can pose challenges in conducting meaningful analyses. By integrating DTM filtration within the Weighted Rips Complex, we leverage the inherent connectivity and relationships between data points to obtain more accurate and reliable rank analysis outcomes.

To investigate the impact of noise and missing values on the stability and robustness of the rank analysis, we explore various levels of noise and missing values to simulate data incompleteness and complexity. We present the results of the rank analysis by varying the parameters m and p in our analysis. Specifically, we explore the effects of m with values 0.1 and 0.2, as well as p with values 1 and 2, on the rank analysis outcomes. The value m is related to the number of neighborhood (k_0) we want to use where $m = k_0/n$.

Figures 10 to 13 present the rank plots for various PV features using DTM filtration with different parameter settings for m and p , as well as varying levels of missing values, ranging from 0% to 50%. The features analyzed include SW1, SW2, Entropy, PI, PLS, PLC, and PS for H_0 and H_1 .

Across the parameter settings, the rank trajectories reveal two regimes. The aggregate H_1 descriptors such as Persistence Statistics, Entropy, Persistence Image and Betti change rank only marginally as the missing-value fraction grows. By contrast, the scores derived from the single most-persistent loop (SW1, SW2 and PLS) show larger rank deviations once a substantial fraction of points is removed, because DTM redistributes persistence across many H_1 features and the dominant cycle is more easily displaced. The PLC exhibits intermediate behavior that improves with higher m . Table 14 quantifies these trends via the Mean Absolute Rank Deviation: Betti, PI, Entropy and PS (H_0/H_1) attain the smallest deviations, whereas SW2, PLC and PLS are the most affected as the reverse of the Rips case (Table 13), where the single-loop scores were the most stable.

The results from Figure 14 highlight varying levels of sensitivity among PV features when exposed to increasing Gaussian noise. SW2 and PLS demonstrate the most consistent and robust performance, with a well-distributed and gradual decline as noise increases. In contrast, PI, Betti and PS show comparatively larger rank movements as noise rises. These observations are quantified by the Rank Instability Index in Table 14, where SW2 attains the lowest RII (0.161). Notably, every DTM RII value is lower than its Rips counterpart in Table 12 (for example, SW2 falls from 0.243 to 0.161 and Entropy from 0.331 to 0.196), confirming that DTM filtration yields uniformly more noise-stable rankings, in line with its outlier-robust construction.

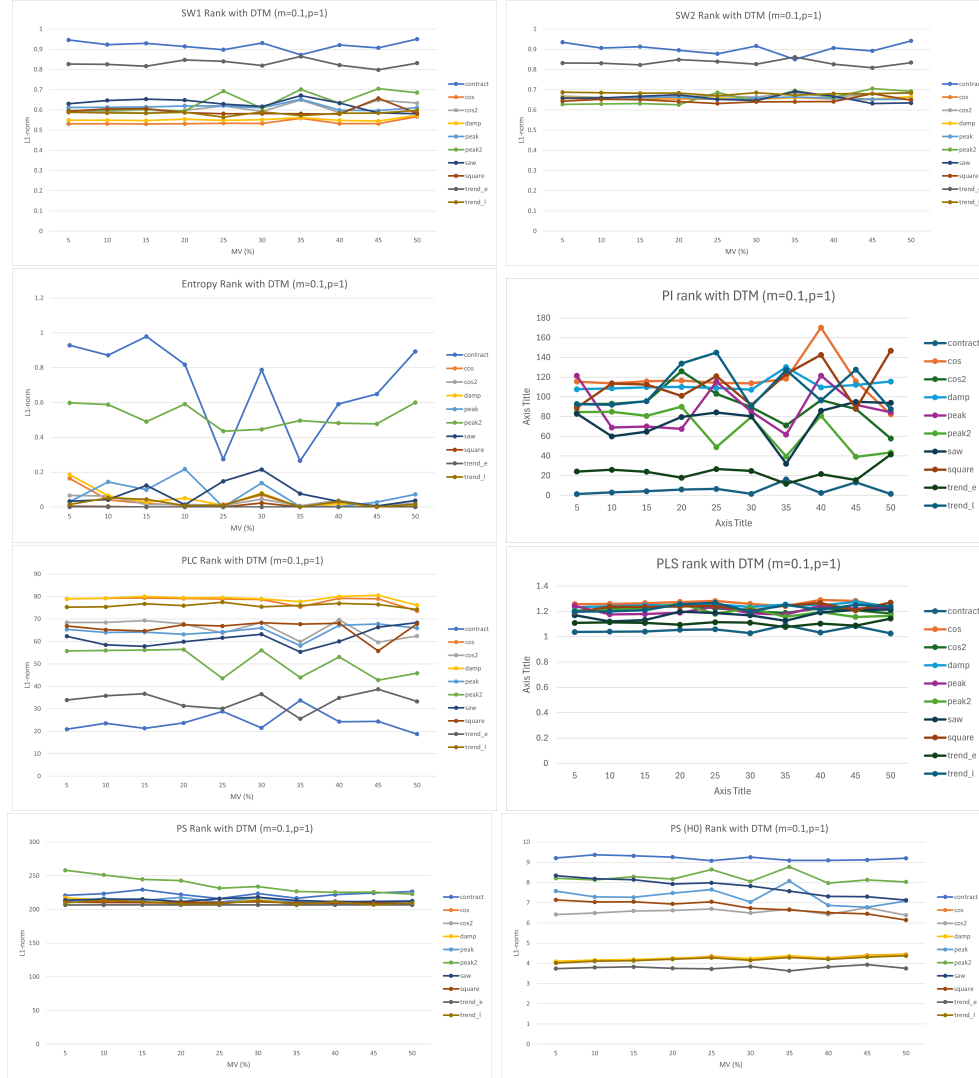


Figure 10. The rank plot PV using DTM ($m = 0.1$, $p = 1$) with missing value ($mv(\%)=0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$) with embedding dimension $D = 14$.

9.1. Quantitative stability of the DTM filtration

To complement the visual rank plots, we quantify the DTM stability with the same Rank Instability Index (RII) and Mean Absolute Rank Deviation (MARD) defined in Section 8, using



Figure 11. The rank plot PV using DTM ($m = 0.1$, $p = 2$) with missing value ($mv(\%)=0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$) with embedding dimension $D = 14$.

DTM parameters $m \in \{0.1, 0.2\}$, $p = 1$ and fixed window $w = 93$. Table 14 reports the results, with the baseline taken as the DTM ranking on the unperturbed data.

Under Gaussian noise (Table 14), the sliding-window and landscape scores are the most stable: For $m = 0.1$, SW2 attains the lowest RII (0.161), followed by PLS (0.173) and SW1 (0.189), whereas for $m = 0.2$ PLS is lowest (0.153) ahead of SW2 (0.161) and SW1 (0.169). PS (H_0/H_1) is the least noise-stable in both cases. The Friedman test is non-significant at $m = 0.1$ ($\chi^2(7) = 9.01$, $p = 0.25$) but reaches significance at $m = 0.2$ ($\chi^2(7) = 15.2$, $p = 0.034$), indicating that a larger neighborhood sharpens the separation between methods. Every DTM RII is lower than the corresponding Rips value in Table 12, confirming that DTM filtration improves noise stability across all PV features.

Under random missing values (Table 14) the ordering is reversed for both values of m . The aggregate H_1 descriptors are the most stable at $m = 0.1$ where Betti (0.123), PI (0.125),

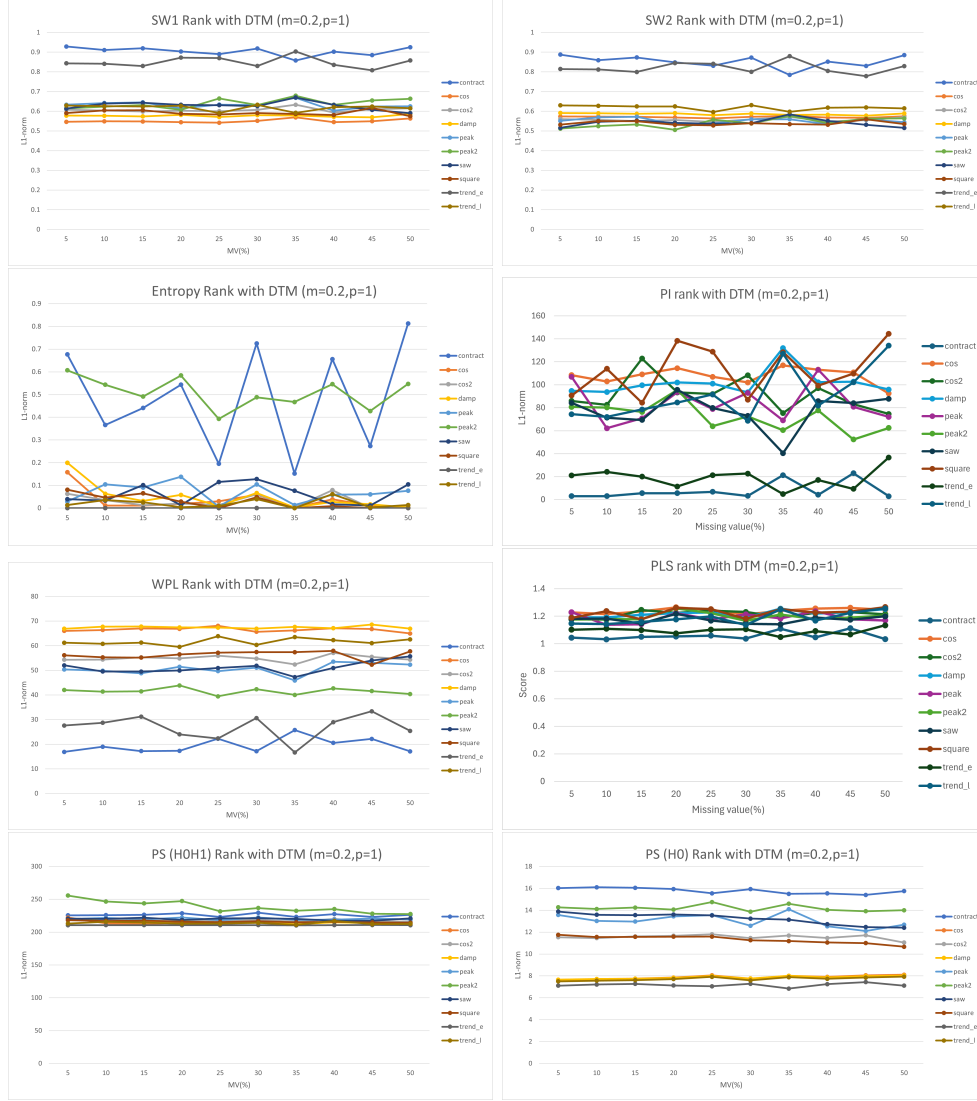


Figure 12. The rank plot PV using DTM ($m = 0.2$, $p = 1$) with missing value (mv(%))=0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50) with embedding dimension $D = 14$.

Entropy (0.137) and PS (0.146), while at $m = 0.2$, PS (0.154), PI (0.179), Betti (0.188) and Entropy (0.200). A Friedman test confirms significant inter-method differences at both settings ($\chi^2(7) = 37.5$, $p < 0.001$ for $m = 0.1$; $\chi^2(7) = 20.7$, $p = 0.004$ for $m = 0.2$). This complements the Rips analysis (Table 13): Under DTM, the curve- and statistics-based summaries become the most robust to missing data, since they aggregate the many H_1 features that the DTM filtration preserves.

10. Non normalized point cloud rank analysis

We standardized both datasets $w = 93$ and subsequently embedded them into higher dimensions. In order to compare with SW1PerS, we adhered to Perea’s normalization technique [43, 44].

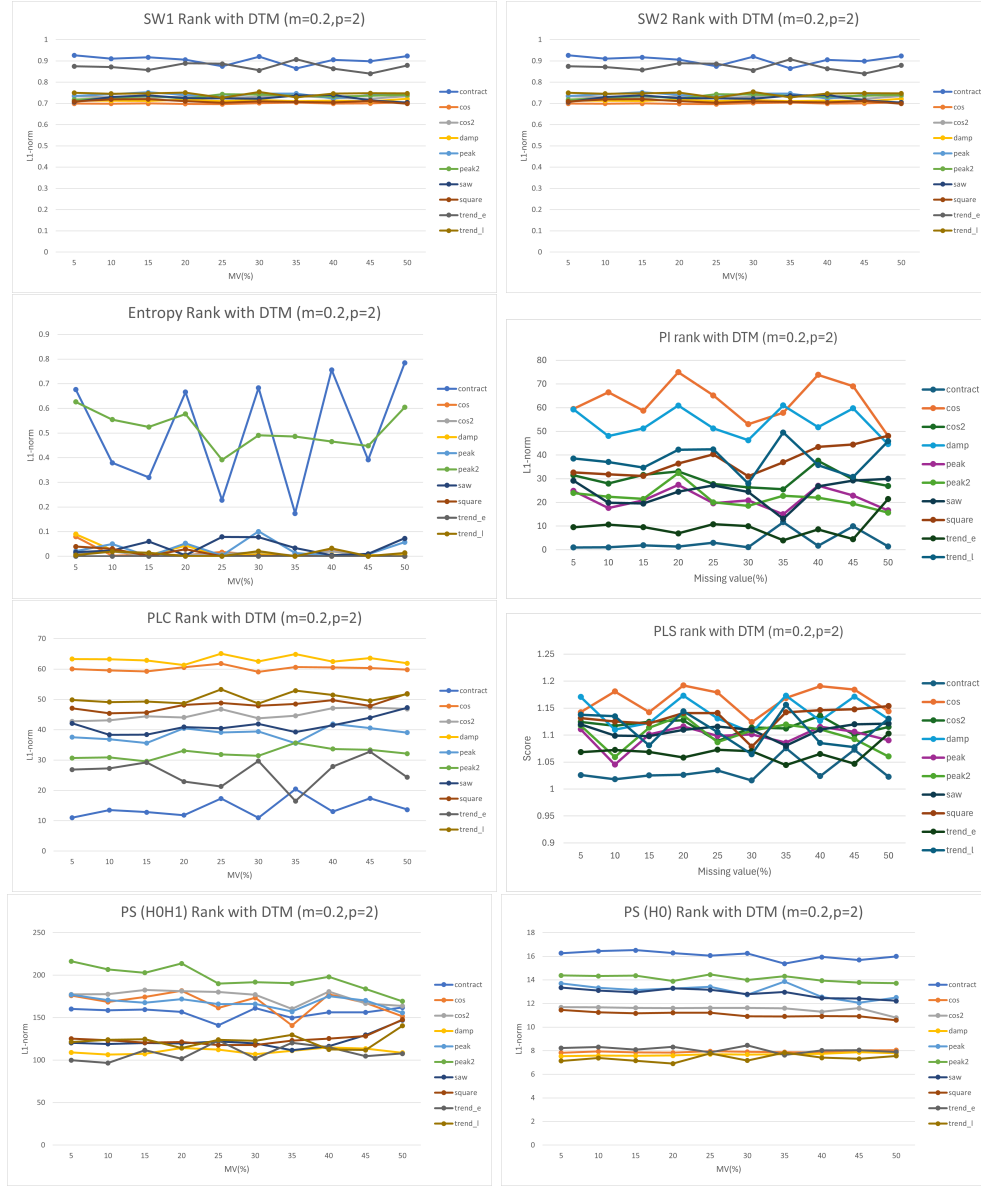


Figure 13. The rank plot PV using DTM ($m = 0.2, p = 2$) with missing value (mv(%))=0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50) with embedding dimension $D = 14$.

Nevertheless, we encountered challenges in effectively discerning between periodic functions, damp, and trendE components.

The purpose of this section is to examine the disparities between normalized and non-normalized point cloud analyses, despite utilizing the same window size ($w = 93$). Our main focus lies in utilizing the PLS as a periodic score metric. Interestingly, our findings demonstrated that non-normalized point cloud rank analysis outperformed normalized analysis. Figure 15 showcased the superiority of non-normalized analysis in distinguishing between periodic and less periodic trends. The PLS demonstrated greater efficacy in this particular context.

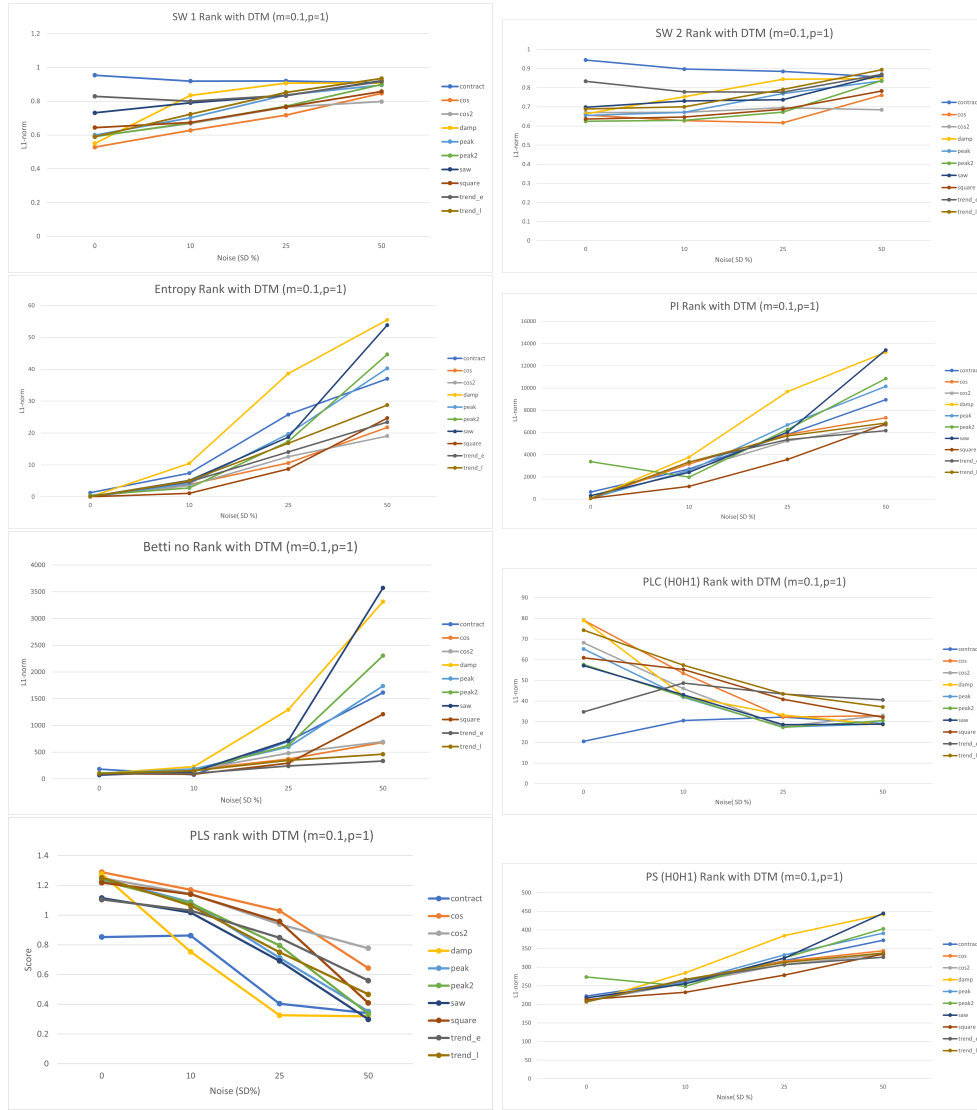


Figure 14. The rank plot PV using DTM ($m = 0.1, p = 1$) with Gaussian noise ($SD(\%)=10, 25, 50$) with embedding dimension $D = 14$.

11. Comparison with non-topological periodicity measures

To position the proposed PV framework relative to the state of practice, we compare periodicity rankings produced by our best-performing PV representations (PLS, SW1) against three established non-topological baselines computed directly on the ten synthetic datasets.

11.1. Baseline methods and implementation

All three baselines are applied to the same standardized (zero-mean, unit-variance) time series used throughout the paper. Implementation details are as follows.

1. **Spectral Entropy (SE):** The normalized power spectral density (PSD) is obtained via the periodogram with a Hann window [45]. Letting $p_k = \text{PSD}_k / \sum_j \text{PSD}_j$ denote the

Table 14. DTM-filtration stability of the PV methods ($p = 1$, $w = 93$): Mean Rank Instability Index (RII) under Gaussian noise and mean Absolute Rank Deviation (MARD) under random missing values, for neighborhood sizes $m = 0.1$ and $m = 0.2$. Each cell gives the metric value $\pm$ s.e. with the stability rank in parentheses (1 = most stable). Lower values indicate greater stability.

PV Method	RII (Gaussian noise)		MARD (missing values)	
	$m = 0.1$	$m = 0.2$	$m = 0.1$	$m = 0.2$
SW1	0.189 $\pm$ 0.031 (3)	0.169 $\pm$ 0.031 (3)	0.234 $\pm$ 0.004 (5)	0.269 $\pm$ 0.001 (5)
SW2	0.161 $\pm$ 0.016 (1)	0.161 $\pm$ 0.021 (2)	0.301 $\pm$ 0.010 (8)	0.311 $\pm$ 0.004 (8)
PLS	0.173 $\pm$ 0.028 (2)	0.153 $\pm$ 0.028 (1)	0.276 $\pm$ 0.001 (6)	0.284 $\pm$ 0.004 (6)
Entropy H1	0.196 $\pm$ 0.029 (4)	0.213 $\pm$ 0.035 (4)	0.137 $\pm$ 0.002 (3)	0.200 $\pm$ 0.004 (4)
Betti H1	0.245 $\pm$ 0.037 (6)	0.246 $\pm$ 0.037 (6)	0.123 $\pm$ 0.003 (1)	0.188 $\pm$ 0.001 (3)
PI H1	0.254 $\pm$ 0.030 (7)	0.259 $\pm$ 0.028 (7)	0.125 $\pm$ 0.003 (2)	0.179 $\pm$ 0.005 (2)
PLC (H0H1)	0.238 $\pm$ 0.030 (5)	0.221 $\pm$ 0.033 (5)	0.296 $\pm$ 0.005 (7)	0.302 $\pm$ 0.006 (7)
PS (H0/H1)	0.273 $\pm$ 0.029 (8)	0.272 $\pm$ 0.032 (8)	0.146 $\pm$ 0.005 (4)	0.154 $\pm$ 0.001 (1)

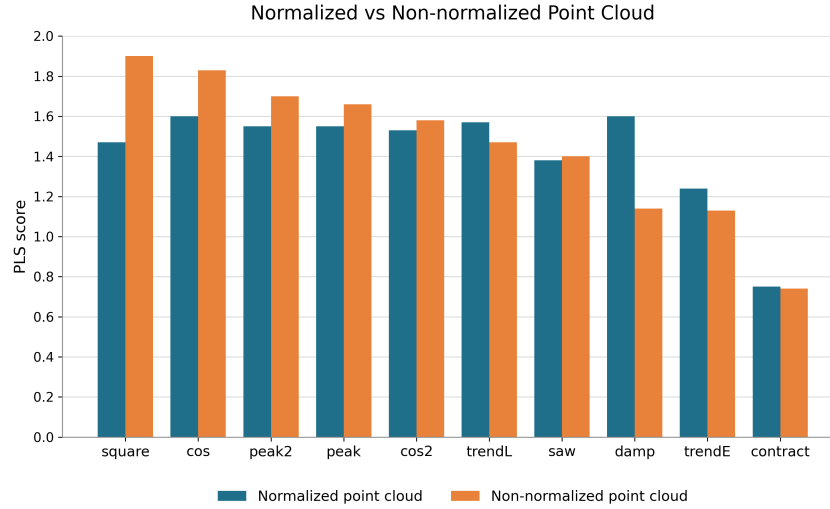


Figure 15. The PLS score plot for normalized and non-normalized point cloud with $w = 93$.

normalized spectral probability at frequency k , the spectral entropy is

$$H_{SE} = - \sum_k p_k \log p_k, \quad H_{\text{norm}} = H_{SE} / \log(N_f), \quad (11.1)$$

where N_f is the number of positive frequencies and $H_{\text{norm}} \in [0, 1]$. A lower H_{norm} indicates more periodic behavior (energy concentrated at fewer frequencies) [28]. The DC component is excluded from the computation.

2. **Autocorrelation Peak Ratio (APR):** The sample autocorrelation function (ACF) is computed via the FFT for lags $k = 0, 1, \dots, N/2$ [9]. APR is defined as the height of the first positive local maximum of the normalized ACF beyond lag 5 (to avoid trivial

short-lag structure):

$$\text{APR} = \max\{r(k) : k > 5, r(k) > r(k-1), r(k) > r(k+1), r(k) > 0\}, \quad (11.2)$$

where $r(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$ is the normalized ACF at lag k . A value close to 1 indicates strong periodicity [14].

3. **Recurrence Rate (RR):** The binary recurrence matrix is defined as [16, 36]

$$\mathbf{R}(i, j) = \Theta(\varepsilon - \|x_i - x_j\|), \quad \varepsilon = 0.10 \sigma_x, \quad (11.3)$$

where Θ is the Heaviside step function, $\|\cdot\|$ is the Euclidean norm, and ε is set to 10% of the signal standard deviation (a standard choice [36]). The recurrence rate is

$$\text{RR} = \frac{1}{N(N-1)} \sum_{i \neq j} \mathbf{R}(i, j), \quad (11.4)$$

with the main diagonal excluded to avoid trivial self-recurrences. Embedding parameters are $m = 1$, $\tau = 1$ (applied to the raw 1-D series).

11.2. Results and discussion

Table 15 reports the rank order produced by each method, and Figure 16–19 illustrate the raw measure outputs (PSD, ACF, recurrence plots) and the final ranking comparison. Kendall's τ (Eq. 11.5) is used as the concordance metric against the SW1-reference (SW1-ref) ranking ordering derived from SW1 at no noise:

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}, \quad n = 10, \quad \text{total pairs} = 45. \quad (11.5)$$

Three key findings emerge:

(i) **SE is the only statistically significant non-topological baseline.** SE achieves $\tau = 0.733$ ($p = 0.002$, $C = 39$, $D = 6$), indicating that FFT-based spectral entropy correctly orders 39 of the 45 signal pairs by periodicity level. Its main failures are: (a) **trendE** is ranked 3rd by SE (SW1-ref rank 4) because the exponential trend concentrates substantial low-frequency energy, mimicking periodicity in the PSD; (b) **cos2** is ranked 5th (SW1-ref rank 8) because its multi-frequency components are captured by SE but not penalized sufficiently. PLS outperforms SE by $\Delta\tau = +0.134$ ($0.867 - 0.733$).

(ii) **APR and RR fail on these signals.** Both APR ($\tau = -0.289$, $p = 0.291$) and RR ($\tau = -0.200$, $p = 0.484$) yield non-significant negative concordance, meaning they are effectively anti-correlated with the ground-truth periodicity order. This result is explained by the nature of these measures:

- **APR** is confounded by signals with strong short-lag autocorrelation unrelated to periodicity: **trendE** obtains APR= 0.902 at lag 5 (rank 1) because its exponential trend creates very high serial correlation; **square** obtains APR= 0.815 at lag 5 (rank 2) because its step-function structure produces trivial short-lag correlation. Neither signal has genuine sinusoidal periodicity at those lags.

Table 15. Periodicity rankings by non-topological baselines vs. PV methods on the ten synthetic datasets (no noise, $D = 14$, optimal w). Rank 1 = most periodic. SE, APR, RR computed directly from the uploaded synthetic data file. τ computed against SW1-ref = SW1 no-noise ranking.

Signal	SW1-ref	SW1	PLS	SE	APR	RR
cos	1	1	1	1	4	6
damp	2	2	2	2	8	8
trendL	3	3	3	4	9	9
trendE	4	4	4	3	1	5
contract	5	5	8	8	10	4
peak	6	6	5	6	7	2
peak2	7	7	6	7	6	3
cos2	8	8	7	5	5	7
square	9	9	9	9	2	1
saw	10	10	10	10	3	10
τ (vs SW1-ref)	—	1.000	0.867	0.733	−0.289	−0.200
C / D	—	45/0	42/3	39/6	16/29	18/27
p -value	—	< 0.001	0.0001	0.0022	0.2912	0.4843

- **RR** is confounded by signals with large-amplitude sharp transitions: **square** (RR= 0.498, rank 1) and **peak** (RR= 0.419, rank 2) score highest because their sharp peaks/plateaus produce highly recurrent point-cloud regions near those amplitude values, not because they are more periodic.

These failures arise from a fundamental limitation of scalar non-topological summaries: They capture specific local properties of the signal (short-lag correlation, recurrence density) rather than the global loop structure that characterizes genuine periodicity in the phase space. The TDA approach, by constructing a full simplicial complex over the embedded point cloud, is sensitive to this global structure.

(iii) **Advantage of TDA is most pronounced for non-sinusoidal signals.** For purely sinusoidal signals (cos, damp, saw), all methods agree. The disagreements emerge for **trendE**, **contract**, **square**, and **peak** — precisely the signals whose periodicity is difficult to characterize with scalar frequency-domain or statistical measures. This underscores the unique value of the topological framework for signals with complex or non-stationary structure.

Figures 16–19 illustrate the raw outputs of each baseline and the final ranking comparison.

12. Interpretability and practical usage of PLS

While the rank analysis establishes the relative ordering of datasets by periodicity, a practitioner applying PLS to a new time series needs guidance on how to interpret its absolute value. This section addresses that need.

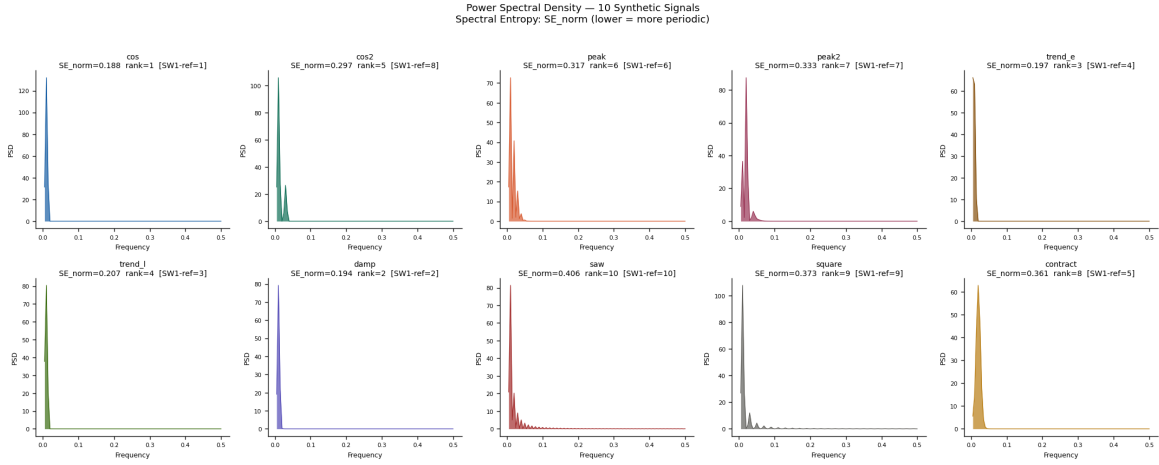


Figure 16. Power Spectral Density (PSD) for all ten synthetic signals, computed via Hann-windowed periodogram. The annotated SE_norm value (lower = more periodic) and rank are shown for each signal. Method: Spectral Entropy [28, 45].

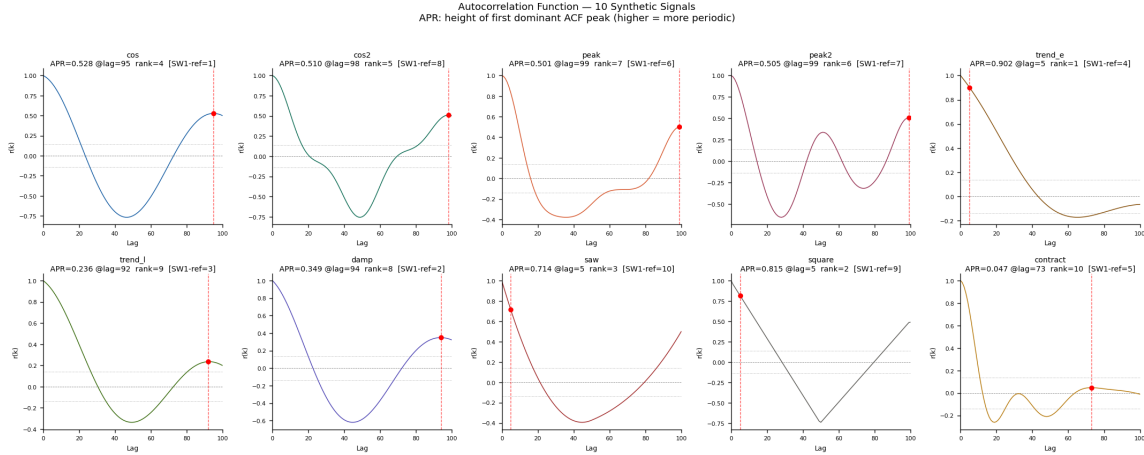


Figure 17. Autocorrelation function (ACF) for all ten synthetic signals. The red dashed vertical line and dot mark the dominant ACF peak (APR value). Gray dotted lines indicate the 95% confidence interval ($\pm 1.96/\sqrt{N}$). Method: Autocorrelation Peak Ratio [9, 14].

12.1. Geometric decomposition of the PLS score

The PLS score equation (4.1) is composed of two terms:

1. The *dominance ratio* $\frac{\|\lambda^{(1)}\|}{\sum_j \|\lambda^{(j)}\|}$, which captures what fraction of the total topological signal is concentrated in the first (most prominent) landscape layer. A value close to 1 implies that one dominant loop structure accounts for nearly all the topological energy — a hallmark of a strongly periodic signal.
2. The *absolute magnitude* $\|\lambda^{(1)}\|$, which captures the strength (lifetime) of that dominant loop. Two signals may have the same dominance ratio but different absolute strengths; the second term distinguishes them.

From the experimental results, the cosine signal (**cos**) achieves $PL = 602.45$ (Table 5), reflecting a single dominant birth–death pair with very high persistence (Max Score = 1.702,

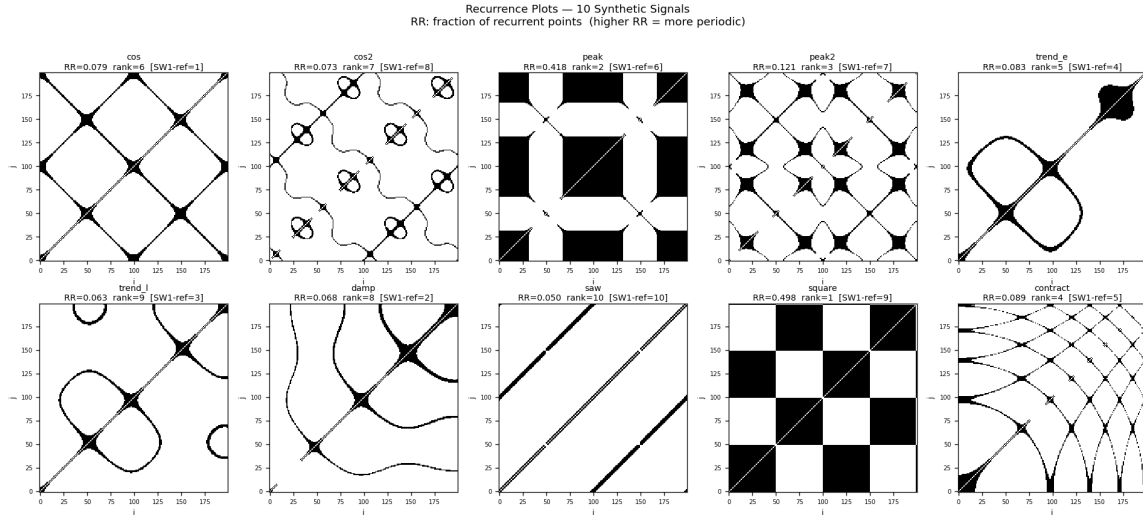


Figure 18. Recurrence plots for all ten synthetic signals ($\varepsilon = 0.10 \sigma_x$, $m = 1$, $\tau = 1$). Black pixels indicate recurrent state pairs. The annotated RR value and rank are shown. Method: Recurrence Rate from RQA [16,36].

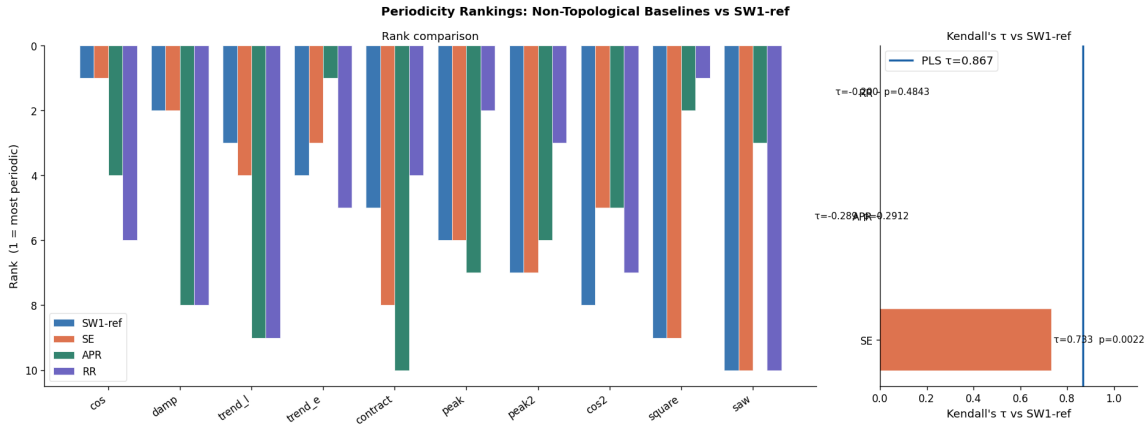


Figure 19. Left: Periodicity rank comparison among SW1-ref, SW1, SE, APR, and RR (rank 1 = most periodic). Right: Kendall's τ against SW1-ref for each method. The blue dashed line marks PLS $\tau = 0.867$. APR and RR yield non-significant negative τ , confirming their unsuitability for these synthetic signals.

SW1 = 0.017). In contrast, the **contract** signal achieves PL = 523.46 but a much lower Max Score (1.569) and higher SW1 (0.094), reflecting that its time-varying frequency structure distributes topological energy across multiple landscape layers.

12.2. Empirical interpretation thresholds

Table 16 provides empirically derived thresholds for PLS, calibrated against the ten synthetic datasets (optimal window, $D = 14$, no noise).

12.3. Step-by-step protocol for a new time series

Given a new time series $f(t)$ of length N , the following protocol applies:

Table 16. Empirical interpretation thresholds for the PLS score (optimal window, $D = 14$, no noise).

PLS range	Interpretation	Signals
≥ 1.60	Strongly periodic; dominant high-amplitude loop	cos, damp, trendL
$[1.40, 1.60)$	Moderately periodic; clear oscillatory structure	trendE, peak
$[1.00, 1.40)$	Weakly periodic; competing freq. components	cos2, square
< 1.00	Near-a-periodic; irregular or transient pattern	saw, contract

1. **Standardize:** Subtract mean and divide by standard deviation.
2. **Interpolate:** Fit cubic spline on $[0, 2\pi]$.
3. **Determine τ :** Sweep $\tau \in [1, \lfloor N/14 \rfloor]$, select $\tau^* = \arg \max_{\tau} \text{ML}(14, \tau)$.
4. **Embed:** Apply Takens' embedding with $D = 14$ and $\tau = \tau^*$.
5. **Normalize:** Mean-center and ℓ^2 -normalize each point.
6. **Compute PD:** Run Vietoris–Rips filtration to obtain dgm_1 .
7. **Compute PLS:** Obtain the persistence landscape $\lambda^{(j)}$ and evaluate equation (4.1).
8. **Interpret:** Use Table 16 to assign a periodicity category. For relative comparison with other series, rank by PLS value directly.

If the data is known to be noisy or contain missing values, replace Step 6 with a DTM-based filtration using $m = 0.1$ and the same window size, which recovers the cos–damp–trendL hierarchy consistent with the noise-free case (as confirmed in Section 9).

13. Conclusion

In the conducted analysis, our primary focus was on the rank analysis to identify and comprehend the periodicity of the topological features within the synthetic data. We employed diverse persistence vector techniques to investigate and compare the periodicity ranks of these features. Through utilization of persistence vector analysis, we were able to extract significant information about the persistence and importance of the topological features. This analysis provided a prioritized view of the data's topological structures, enabling us to identify the most significant and persistent features related to periodicity.

The formal quantitative stability analysis (Section 8) shows that SW2 achieves both the lowest mean Rank Instability Index ($\text{RII} = 0.243$) under Gaussian noise and the lowest Mean Absolute Rank Deviation ($\text{MARD} = 0.117$) under missing data, making it the most stable representation overall, with PLS and SW1 ranking close behind. Under noise the inter-method differences are not statistically significant (Friedman $p = 0.42$), whereas under missing data they are significant ($p = 0.001$); Entropy (H_1) and PLC (H_0H_1) are the least stable methods under noise and missing data, respectively. Incorporating the DTM filtration further improves noise stability across all PV features; under missing data it shifts robustness toward the aggregate H_1

descriptors (Betti, PI, Entropy and PS), which become the most stable (Table 14). Comparison against non-topological baselines (Section 11) confirms that PLS consistently outperforms spectral entropy, autocorrelation, and RQA methods, achieving Kendall $\tau = 0.867$, substantially outperforming spectral entropy (SE: The only significant baseline, $\tau = 0.733$, $p = 0.002$). APR ($\tau = -0.289$) and RR ($\tau = -0.200$) yield non-significant concordance, confirming the limitations of scalar non-topological measures for complex signals.

We investigated the optimal parameters for embedding dimension (D) and time delay (τ) in the sliding window persistence analysis. The selection of $D = 14$ as the optimal embedding dimension is based on the work by [43], which provides empirical evidence and theoretical justification for its effectiveness in capturing the intrinsic properties of the synthetic data. We explored the impact of different window sizes on the rank analysis using sliding window persistence. By selecting the optimal window size, we obtained meaningful rankings that highlighted the importance and persistence of the topological features.

Additionally, we utilized nearest neighbor analysis to determine the best dimension and mutual information for the time delay (τ). We evaluated the stability and robustness of the topological features through rank analysis under different levels of noise and randomly introduced missing values. Throughout the study, we employed the L^1 -norm to compress the persistence vectors into meaningful values.

Despite the efforts made in standardization, disparities persisted between normalized and non-normalized point cloud analyses. Non-normalized analysis, particularly when utilizing the PLS, exhibited superior outcomes in distinguishing between periodic and less periodic trends. Further investigation is required to comprehend the impact of normalization techniques on point cloud analysis.

The practitioner's interpretability guide (Section 12) provides concrete empirical thresholds and a step-by-step protocol for applying PLS to new time series. Future work will extend validation to real-world time series in physiological, financial, and climate domains, and will explore the integration of PLS with downstream machine learning tasks such as anomaly detection and classification. The computational complexity of PH is an acknowledged limitation; for very large time series ($N \gg 10^3$), approximate or landmark-based filtrations may be required.

Declarations

Funding. The authors would like to acknowledge Universiti Malaysia Perlis (UniMAP) for its financial support towards the publication fee. The authors also acknowledge the support received under the previous Fundamental Research Grant Scheme (FRGS) grant (FRGS/1/2019/STG06/UMT/02/2).

Conflict of interest/Competing interests. The authors have no competing interests to declare.

Ethics approval and consent to participate. Not applicable. This study uses only synthetic datasets and does not involve human participants or animals.

Consent for publication. Not applicable.

Data availability. All datasets used in this study are synthetic and were generated as part of the research. The generated datasets and data-generation scripts are available from the corresponding author upon reasonable request.

Code availability. The code used to generate the synthetic data and conduct the analyses is

available from the corresponding author upon reasonable request.

Author contribution. All authors contributed to the study conception and design. Material preparation, data generation, and analysis were performed by Zabidi Abu Hasan, R. U. Gobithaasan, and Roslan Hasni. The first draft of the manuscript was written by Zabidi Abu Hasan, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

- [1] H. Abdi and L. J. Williams, *Principal component analysis*, Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4), 433–459.
- [2] H. Adams, S. Chepushtanova, T. Emerson, E. Hanson, M. Kirby, F. Motta, R. Neville, C. Peterson, P. Shipman and L. Ziegelmeier, *Persistence images: A stable vector representation of persistent homology*, Journal of Machine Learning Research, 2015, 18, 1–35.
- [3] C. C. Aggarwal, A. Hinneburg and D. A. Keim, *On the surprising behavior of distance metrics in high dimensional spaces*, International Conference on Database Theory, 2001, 420–434.
- [4] D. Ali, A. Asaad, M.-J. Jimenez, V. Nanda, E. Paluzo-Hidalgo and M. Soriano-Trigueros, *A survey of vectorization methods in topological data analysis*, arXiv preprint, 2022, 1–26.
- [5] W. Altidor, T. M. Khoshgoftaar and A. Napolitano, *Measuring stability of feature ranking techniques: A noise-based approach*, International Journal of Business Intelligence and Data Mining, 2012, 7(1–2), 80–115.
- [6] H. Anai, F. Chazal, M. Glisse, Y. Ike, H. Inakoshi, R. Tinarrage and Y. Umeda, *DTM-based filtrations*, Abel Symposia, 2020, 15, 33–66.
- [7] N. Atienza, R. Gonzalez-Díaz and M. Soriano-Trigueros, *On the stability of persistent entropy and new summary functions for topological data analysis*, Pattern Recognition, 2020, 107.
- [8] D. Berrar, *An empirical evaluation of ranking measures with respect to robustness to noise*, Journal of Artificial Intelligence Research, 2014, 49, 241–267.
- [9] G. Box, G. Jenkins, G. Reinsel and G. Ljung, *Time Series Analysis: Forecasting and Control*, 5th Edn, Wiley, Hoboken, NJ, 2015.
- [10] P. Bubenik, *Statistical topological data analysis using persistence landscapes*, Journal of Machine Learning Research, 2015, 16, 77–102.
- [11] P. Bubenik, M. Hull, D. Patel and B. Whittle, *Persistent homology detects curvature*, Inverse Problems, 2020, 36(2), 025008.
- [12] G. Carlsson, *Topology and data*, Bulletin of the American Mathematical Society, 2009, 46(2), 255–308.
- [13] P. Celik and E. Sefer, *Analyzing transaction graphs via motif-based graph representation learning for cryptocurrency price prediction*, Computational Economics, 2025, 1–22.
- [14] C. Chatfield, *The Analysis of Time Series: An Introduction with R*, 6th Edn, Chapman & Hall/CRC, Boca Raton, FL, 2004.
- [15] Y. M. Chung and A. Lawson, *Persistence curves: A canonical framework for summarizing persistence diagrams*, Advances in Computational Mathematics, 2022, 48(1), 1–30.

- [16] J. Eckmann, S. Kamphorst and D. Ruelle, *Recurrence plots of dynamical systems*, Europhysics Letters, 1987, 4(9), 973–977.
- [17] H. Edelsbrunner and J. Harer, *Computational topology*, Open Problems in Topology II, 2010, 493–545.
- [18] S. Emrani, T. Gentimis and H. Krim, *Persistent homology of delay embeddings*, arXiv preprint, 2013, 1–14.
- [19] H. Gakhar and J. A. Perea, *Sliding window persistence of quasiperiodic functions*, arXiv preprint, 2021, 1–32.
- [20] M. Gidea, *Topological data analysis of critical transitions in financial networks*, Springer Proceedings in Complexity, 2017, 47–59.
- [21] M. Gidea, D. Goldsmith, Y. Katz, P. Roldan and Y. Shmalo, *Topological recognition of critical transitions in time series of cryptocurrencies*, Physica A: Statistical Mechanics and its Applications, 2020, 548, 123843.
- [22] M. Gidea and Y. Katz, *Topological data analysis of financial time series: Landscapes of crashes*, arXiv preprint, 2017.
- [23] M. Gidea and Y. Katz, *Topological data analysis of financial time series: Landscapes of crashes*, Physica A: Statistical Mechanics and its Applications, 2018, 491, 820–834.
- [24] R. Gobithaasan, Z. A. Hasan, K. D. Selvarajh, K.-S. Wong, S. Mamat, M. Z. M. Ali, K. T. Miura and P. Dotko, *Clustering selected Terengganu’s rainfall stations based on persistent homology*, Thai Journal of Mathematics, 2022, 20(1), 197–211.
- [25] Z. A. Hasan and R. U. Gobithaasan, *The characterization of rainfall data set using persistence diagram and its relation to extreme events: Case study of three locations in kemaman, terengganu*, Soft Computing in Data Science, 2023, 261–274.
- [26] Z. A. Hasan, R. U. Gobithaasan, N. F. S. Mohd Zulkepli, M. Z. M. Ali, K. T. Miura and P. Dlotko, *Identifying irregular rainfall patterns using persistent homology*, Journal of Advanced Research in Applied Sciences and Engineering Technology, 2024, 35(1), 26–36.
- [27] A. Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, UK, 2002.
- [28] T. Inouye, K. Shinosaki, H. Sakamoto, S. Toi, S. Ukai, A. Iyama, Y. Katsuda and M. Hirano, *Quantification of EEG irregularity by use of the entropy of the power spectrum*, Electroencephalography and Clinical Neurophysiology, 1991, 79(3), 204–210.
- [29] M. S. Ismail, S. I. Hussain and M. S. M. Noorani, *Detecting early warning signals of major financial crashes in bitcoin using persistent homology*, IEEE Access, 2020, 8, 202042–202057.
- [30] M. S. Ismail, M. S. Md Noorani, M. Ismail, F. Abdul Razak and M. A. Alias, *Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from kuala lumpur stock exchange*, Applied Soft Computing Journal, 2020, 93, 106422.
- [31] M. S. Ismail, M. S. M. Noorani, M. Ismail, F. A. Razak and M. A. Alias, *Early warning signals of financial crises using persistent homology*, Physica A: Statistical Mechanics and its Applications, 2022, 586, 126459.
- [32] A. Kalousis, J. Prados and M. Hilario, *Stability of feature selection algorithms: A study on high-dimensional spaces*, Knowledge and Information Systems, 2007, 12(1), 95–116.

- [33] M. G. Kendall and B. B. Smith, *Rank Correlation Methods*, Griffin, London, 1948.
- [34] M. B. Kennel and H. D. Abarbanel, *False neighbors and false strands: A reliable minimum embedding dimension algorithm*, Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 2002, 66(2), 1–18.
- [35] L. I. Kuncheva, *A stability index for feature selection*, in Proceedings of the 25th IASTED International Multi-Conference on Artificial Intelligence and Applications, 2007, 390–395.
- [36] N. Marwan, M. C. Romano, M. Thiel and J. Kurths, *Recurrence plots for the analysis of complex systems*, Physics Reports, 2007, 438(5–6), 237–329.
- [37] Y. Mileyko, S. Mukherjee and J. Harer, *Probability measures on the space of persistence diagrams*, Inverse Problems, 2011, 27(12).
- [38] M. R. Muldoon, R. S. MacKay, J. P. Huke and D. S. Broomhead, *Topology from time series*, Physica D: Nonlinear Phenomena, 1993, 65(1–2), 1–16.
- [39] E. Munch, *A user's guide to topological data analysis*, Journal of Learning Analytics, 2017, 4(2), 47–61.
- [40] J. R. Munkres, *Elements of Algebraic Topology*, Addison-Wesley, Menlo Park, CA, 1984.
- [41] A. Myers, E. Munch and F. A. Khasawneh, *Persistent homology of complex networks for dynamic state detection*, Physical Review E, 2019, 100(2).
- [42] J. A. Perea, *Persistent homology of toroidal sliding window embeddings*, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, 6435–6439.
- [43] J. A. Perea, A. Deckard, S. B. Haase and J. Harer, *SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data*, BMC Bioinformatics, 2015, 16, 257.
- [44] J. A. Perea and J. Harer, *Sliding windows and persistence: An application of topological methods to signal analysis*, Foundations of Computational Mathematics, 2015, 15(3), 799–838.
- [45] G. Powell and I. Percival, *A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems*, Journal of Physics A: Mathematical and General, 1979, 12(11), 2053–2071.
- [46] R. Rabadán and A. J. Blumberg, *Topological Data Analysis for Genomics and Evolution: Topology in Biology*, Cambridge University Press, Cambridge, UK, 2019.
- [47] N. Ravishanker and R. Chen, *Topological data analysis (tda) for time series*, arXiv preprint, 2019, 1–29.
- [48] P. J. Rousseeuw and C. Croux, *Alternatives to the median absolute deviation*, Journal of the American Statistical Association, 1993, 88(424), 1273–1283.
- [49] V. D. Silva, P. Skraba and M. Vejdemo-Johansson, *Topological analysis of recurrent systems*, Workshop on Algebraic Topology and Machine Learning, NIPS, 2012, 2(1), 2–6.
- [50] S. M. S. Syed Musa, M. S. Md Noorani, F. Abdul Razak, M. Ismail, M. A. Alias and S. I. Hussain, *Using persistent homology as preprocessing of early warning signals for critical transition in flood*, Scientific Reports, 2021, 11(1), 1–14.

- [51] F. Takens, *Detecting strange attractors in turbulence*, in Dynamical Systems and Turbulence, Warwick 1980, 898 of Lecture Notes in Mathematics (Edited by D. Rand and L.-S. Young), Springer, Berlin, Heidelberg, 1981, 366–381.
- [52] C. J. Tralie and J. A. Perea, *(Quasi)periodicity quantification in video data, using topology*, SIAM Journal on Imaging Sciences, 2018, 11(2), 1049–1077.
- [53] T. Tuncer, U. Kaya, E. Sefer, O. Alacam and T. Hoser, *Asset price and direction prediction via deep 2d transformer and convolutional neural networks*, in Proceedings of the Third ACM International Conference on AI in Finance, 2022, 79–86.
- [54] R. Turkeš, G. Montúfar and N. Otter, *On the effectiveness of persistent homology*, 2023.
- [55] K. Turner, Y. Mileyko, S. Mukherjee and J. Harer, *Fréchet means for distributions of persistence diagrams*, Discrete and Computational Geometry, 2014, 52(1), 44–70.
- [56] Y. Umeda, *Time series classification via topological data analysis*, Transactions of the Japanese Society for Artificial Intelligence, 2017, 32(3), 1–12.
- [57] Y. Uygun and E. Sefer, *Financial asset price prediction with graph neural network-based temporal deep learning models*, Neural Computing and Applications, 2025, 37(30), 25445–25471.
- [58] A. Wehrl, *General properties of entropy*, Reviews of Modern Physics, 1978, 50(2), 221–260.
- [59] N. F. S. Zulkepli, M. S. Md Noorani, F. A. Razak, M. Ismail and M. A. Alias, *Cluster analysis of haze episodes based on topological features*, Sustainability (Switzerland), 2020, 12(10).
- [60] N. F. S. Zulkepli, M. S. M. Noorani, F. A. Razak, M. Ismail and M. A. Alias, *Topological characterization of haze episodes using persistent homology*, Aerosol and Air Quality Research, 2019, 19(7), 1614–1621.

Received March 2025; Accepted September 2026; Available online October 2026.