On topological data analysis for structural dynamics: an introduction to persistent homology

T. Gowdridge, N. Dervilis, K. Worden

Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield
Mappin Street,
Sheffield, S1 3JD,
UK
tgowdridge1@sheffield.ac.uk

September 25, 2022

Abstract

Topological methods can provide a way of proposing new metrics and methods of scrutinising data, that otherwise may be overlooked. In this work, a method of quantifying the shape of data, via a topic called *topological data analysis* will be introduced. The main tool within *topological data analysis* (TDA) is *persistent homology*. Persistent homology is a method of quantifying the shape of data over a range of length scales. The required background and a method of computing persistent homology is briefly discussed in this work.

Ideas from topological data analysis are then used for nonlinear dynamics to analyse some common attractors, by calculating their embedding dimension, and then to assess their general topologies. A method will also be proposed, that uses topological data analysis to determine the optimal delay for a time-delay embedding. TDA will also be applied to a Z24 Bridge case study in structural health monitoring, where it will be used to scrutinise different data partitions, classified by the conditions at which the data were collected. A metric, from topological data analysis, is used to compare data between the partitions. The results presented demonstrate that the presence of damage alters the manifold shape more significantly than the effects present from temperature.

1 Introduction

Topological methods are very rarely used in structural dynamics generally, although considering the structure and topology of observed data may be useful. Topological methods can provide new metrics and methods of scrutinising data; the most rudimentary and powerful of which, *persistence homology*, will be discussed and used extensively in this paper. By applying topological methods, an understanding of the topological structure of data can be used to formulate arguments and develop an understanding of system parameters on a manifold shape. This statement is especially true when working with higher-dimensional data sets, where an intuitive understanding of a manifold is not easy to visualise. By using topological methods, an understanding can be quantifiably analysed. When considering engineering data, data-sets are often embedded in higher-dimensional space, and topological information is often not well explored, leaving out potentially important and insightful information.

Topological data analysis (TDA), is a recently-developed and fast-growing field that has found its way into many areas of science and engineering. The general idea of TDA applies concepts from algebraic topology to data sets. The primary focus of TDA is to determine the shape of the manifold in which sampled data are embedded. This process is achieved by identifying $2D$ holes, $3D$ cavities and higher-dimensional analogues within the data structures. From these sampled data, an approximation to the topological structure can be calculated by use of *simplicial complexes*, which are higher-dimensional analogues of graphs. From the simplicial complexes, the persistent homology can be calculated, this is then used to understand the topological structure of the data. The persistent homology is invariant for each data set, and can be used to identify the data set; much like a fingerprint.

After the simplicial complexes have been constructed for some point data, these can be manipulated using ideas from algebraic topology, in order to determine algebraic groups that will capture information about the shape and structure of the simplicial complex; these groups are topological invariants. For any simplicial complex, algebraic topology can deduce a property called

the *homology*, which encodes information about its number of $k-$dimensional holes. A generalisation of homology will be discussed here with respect to pioneering work by Edelsbrunner [1, 2], where the homology can be considered over a range of simplicial complexes; this is called the *persistent homology*, aptly named, as this uncovers how the homology persists over a range of scales.

The field of algebraic topology is well studied, but the application of the ideas to discrete point clouds has contributed to a boom in computational topology. Many standard packages exist to analyse data with regards to topological methods: GUDHI [3] and Ripser [4] being the most influential throughout this piece of work.

The layout of the paper is as follows: Section 2 will be devoted to persistent homology and its significance, and will provide an intuitive understanding. Some use cases of persistent homology will also be discussed. Section 3 will introduce and analyse the topology of some common attractors from the field of nonlinear dynamics. Section 4 will then go on to look at the classic Z24 Bridge structural health monitoring (SHM) case study, and explore the role of topology as a metric between partitions of the data set.

2 Topological Data Analysis

Only the strict mathematical formulations will be introduced here. For more information regarding the mathematics of group theory, and algebraic topology, the authors would advise referring to [5–8], and [1, 2, 9–15] for more on TDA. For some other interesting applications in economics and genomics, the reader can refer to [9, 16, 17].

2.1 Manifolds

Manifolds are continuous surfaces from which the data are assumed to be sampled. By understanding the topology of the sampled data points, it is the aim of TDA to extract topological information about the underlying manifold from the data. The manifold shape is unknown prior to analysis, and persistent homology will identify features within the manifold over a range of length scales. Thereby, understanding the shape of the sampled data, it is the conjecture of TDA that the shape of the manifold is also understood.

Formally, a manifold is a space that is locally homeomorphic to some $n-$dimensional Euclidean space, $\mathbb{R}^n$. In this work, manifolds are primarily thought of as the underlying space from which the generated or collected data are sampled. Throughout this work, there will be a reoccurring idea that a change in the parameters of a dynamic system may change the shape of the associated manifold, therefore the topology can potentially be used to identify a change in the system.

2.2 Simplicial Complex

Simplicial complexes will be used as a way of attributing quantifiable shape to the data; they can be thought of as higher-dimensional analogues of graphs, giving a way of encoding connections between vertices. In TDA, the vertices of the simplicial complexes are the observed data points. Simplicial complexes can be analysed to output the homology of the data, and following this, the persistent homology. The homology and persistent homology are key topological invariants that can be used to describe the structure of the data.

A simplicial complex is a structure made up of fundamental building blocks called *simplices*, the first four types of which are shown in Fig. 1. Each vertex in the simplex is fully connected to all the other vertices and the space enclosed by the vertices is part of that simplex. For instance, Δ^2 encloses a two-dimensional area, Δ^3 encloses a three-dimensional volume. This sequence can be generalised for Δ^k enclosing a $k-$dimensional space between $(k+1)$ fully connected vertices.
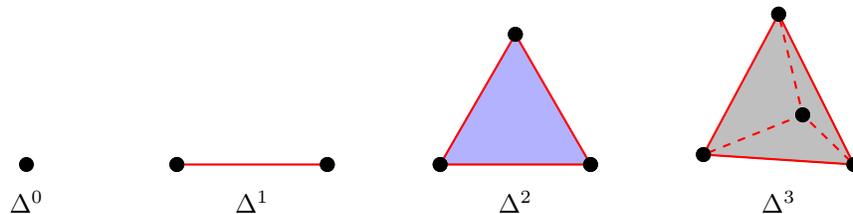

Figure 1: The first four simplices.

A *filtration* of a simplicial complex, K, is a nested sequence of sub-complexes, such that, $\emptyset = K^0 \subset K^1 \subset ... \subset K^m = K$, where $K^{i+1} = K^i \cup \Delta^i$ where Δ^i is a simplex of K [11], and $\emptyset$ is the empty set, i.e, there are no points inside.

There are many ways to construct a simplicial complex from point data. For simplification, only one method will be discussed within this paper, the *Vietoris-Rips* (VR) complex [18]. The VR complex can be constructed for point data present in a *metric space*.

A *metric* ∂_X is defined on a set X, which maps two elements from X into the positive real numbers. The mapping gives the associated distance between the two elements. For a set to be deemed a metric space, it must satisfy the following axioms:

1. for any $x, y \in X$ then $\partial_X(x, y) = \partial_X(y, x)$;

2. for any $x, y \in X$ then $\partial_X(x, y) \geq 0$, and $= 0$ iff $x = y$;

3. for any $x, y, z \in X$ then $\partial_X(x, z) \leq \partial_X(x, y) + \partial_X(y, z)$.

If these axioms are satisfied, (X, ∂_X) forms a metric space, where ∂_X is the metric on X.

An *open ball* is defined on a metric space, (X, ∂_X). For a point $a \in X$ and $\varepsilon > 0$, the subset of X consisting of all the points $x \in X$ such that $\partial_X(a, x) < \varepsilon$ is referred to as the open ball of radius ε at a.

For the VR complex $\mathrm{VR}_\varepsilon(X, \partial_X)$, let (X, ∂_X) be a finite metric space and $\varepsilon \in \mathbb{R}_{>0}$ be a fixed value that determines the scale of the VR complex [19], where X is the set of vertices and ∂_X is the metric on X. A VR complex is defined by the following condition

$$\{x_0, ..., x_k\} \in \mathrm{VR}_\varepsilon(X, \partial_X) \Leftrightarrow \|x_i - x_j\| \leq \varepsilon, \ \forall i, j \in \{0, ..., k\}. \tag{1}$$

As ε goes from 0 to $+\infty$, a nested sequence of complexes defines the *Vietoris-Rips filtration* [11].
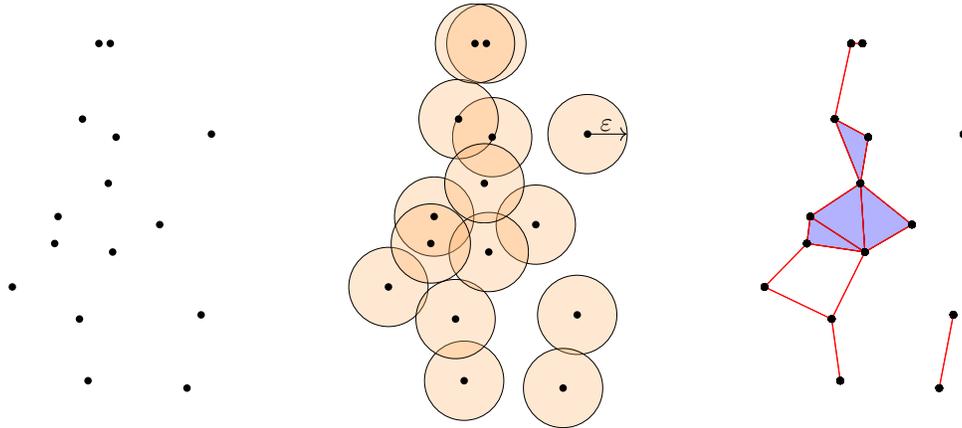

Figure 2: The process of constructing a VR complex.

This process is depicted in Fig. 2 for some randomly-sampled data and an arbitrary value of ε, which can be seen as the radius of the balls. The existence of a simplex is determined by how the balls intersect between the vertices. For a VR complex, a simplex between some set of vertices is formed if the Euclidean distance between all those vertices is less than ε.

2.3 Persistent Homology

Before delving into persistent homology, the homology of a topological space needs to be defined. Many great introductions and summaries of homology already exist; the more intrigued reader may want to consult [6, 7, 9–11, 20]. In very quick terms, the homology is calculated by computing successive boundary operations on a *chain complex* of a topological space. Homology is essentially the quantification of the number of voids present in a topological space, or in the case of TDA, a simplicial complex. The homology groups, $H_k(X)$ are invariants for the data set X, where k refers to the dimension of the homology group. Generally, the kth homology group encodes information about the number of $k-$dimensional holes in the data. Under the rules of topology, discontinuities (voids) cannot be created or destroyed under continuous maps (homeomorphisms). Therefore, a

simplicial complex can be categorised by the properties underpinned by the homology. The homology can be used to categorise and compare between simplicial complexes, and by extension, data sets.

From the homology, the *Betti numbers* are defined as the *rank* of the homology groups. If the Betti numbers for two topological spaces are different, these spaces are not topologically identical, meaning a continuous bijective map between the spaces does not exist. The zeroth Betti number, β_0, is the rank of the zeroth homology group, $H_0(X)$, and refers to the number of connected sets in the topological space. The first Betti number, β_1, is the rank of the first homology group, $H_1(X)$, and refers to the number of non-contractible holes present in the space. The second Betti number, β_2, refers to the number of enclosed volumes in the topological space. This analogy carries on further for higher dimensions.

This discussion now raises the question: which length scale ε is representative of the topology of the data? When constructing the VR complexes, for the same data set, different values of epsilon, will result in different values for the Betti numbers. The hyper-parameter ε determines the Betti numbers for that specific instance of some point cloud data. Additionally, when the feature present within the data is at a length scale less than ε this feature will not be expressed, as ε will span the feature. This reasoning means that only topological properties that are described at a length scale greater than ε can be captured. A problem arises here, as usually the feature scale is not known prior to analysis, and a manifold may have many multi-scale features.

The answer to this problem is to vary ε and see how the Betti numbers evolve and *persist*. When varying ε, a *filtered simplicial complex* is formed. This can be thought of as a chain or sequence of simplices; with n disconnected points at the beginning and a fully connected $(n-1)$- simplex at the end. The next simplicial complex in the *filtration* is the previous simplicial complex plus the next simplex to be formed as ε increases. Fig. 3 shows some simplicial complexes in this process. A filtered simplicial complex is an ordering of simplices to show how they evolve as a distance scale is increased. From this, one can form an idea of how the topology evolves over the filtration.

Obtaining the homology for a single value of ε provides very limited information; this notion is almost redundant, because of potential varying feature length scales in the manifold. For this reason, it is vital to consider what homological features persist as ε is varied. The goal of persistent homology is to track the homology classes as ε is varied. The process of varying ε does not bias any disk size, as all are being considered. This process will give an initial value, $\varepsilon_{\min}$, where a specific homological feature comes to life and $\varepsilon_{\max}$, where the feature is no longer considered for that simplicial complex. This range of values $[\varepsilon_{\min}, \varepsilon_{\max}]$ is called the *persistence interval* for that homological feature. Each persistence interval is attributed a Betti number. Following this, the set of all persistence intervals is descriptive for that manifold, giving information about in which dimension a hole exists in the data and over what range of values it persists.

The persistence intervals obtained can be represented in two ways: *barcodes* or *persistence diagrams*. In a barcode representation, the $x-$axis refers to the value of ε. As ε increases, the barcode shows which features persist. The set of intervals are plotted with each interval beginning at its $\varepsilon_{\min}$ and ending at its $\varepsilon_{\max}$. The colour of the interval on the barcode refers to the Betti number, β_k [15]. The value of the $y-$axis can simply be thought of as an indexing of the intervals in the barcode.
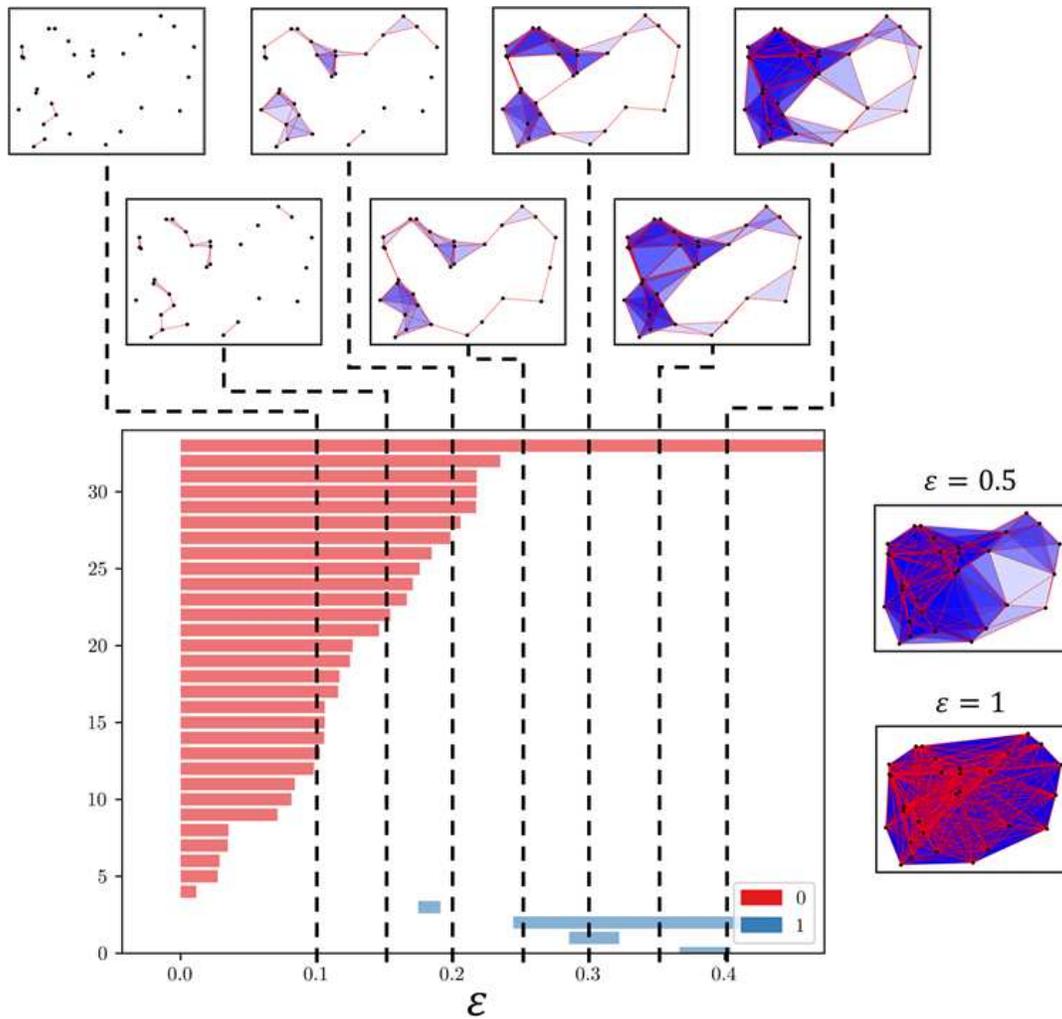
Figure 3: A persistence barcode with realisations showing which simplicial complexes are present at values of ε of 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, and 1.00.

An example of a barcode can be seen in Fig. 3, with vertical dotted lines showing the intersections with the intervals, showing which features are present for the corresponding simplicial complex in the filtration. The red bars refer to the persistence intervals for β_0, similarly the blue bars are for β_1. A bar is formed when that feature begins and ends when that feature dies. The number of intersections of the persistence intervals with the slices (denoted by the dashed black line) at each ε refers to the Betti number for each feature. For instance, for $\varepsilon = 0.3$, the simplicial complex consists of one connected component (as there is only one intersection with the red persistence intervals) and there are two holes present (denoted by two intersections with the blue persistence intervals). For $\varepsilon = 0.50, 1.00$, these are shown besides the barcode, as there is no change in the barcode for these values. For $\varepsilon > 0.45$, all of the holes have been spanned, and there is only one connected component. The vertices only become more and more connected as ε increases, and this has little topological interest. The length of the interval represents for how long the feature persists. The longer the feature persists, the higher the probability that this feature is characteristic of the manifold. Shorter intervals are generally regarded as topological noise.
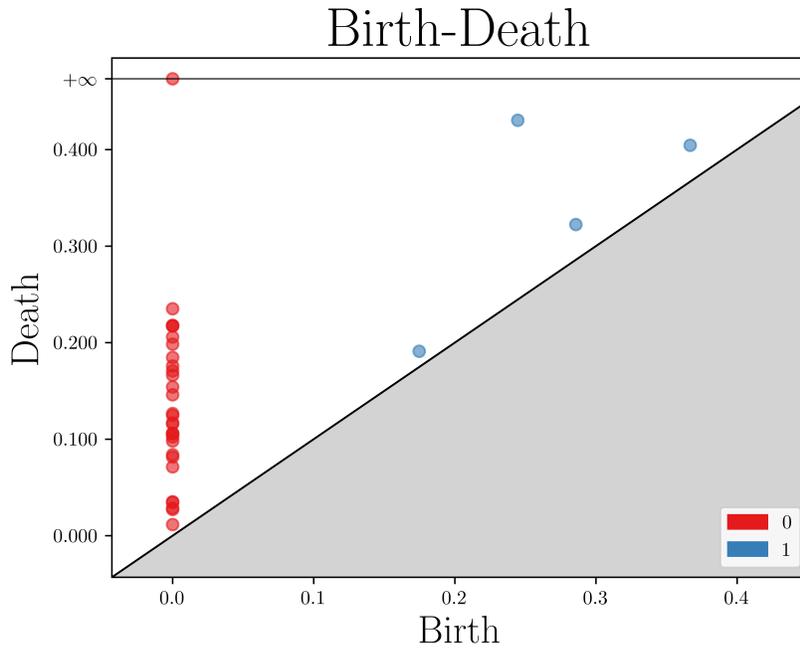
Figure 4: Birth-death diagram for the same random data present in Fig. 3.

The other method of visualising the set of persistence intervals is the persistence diagram, or *birth-death* diagram. In this representation, $\varepsilon_{\min}$ is plotted on the $x-$axis and $\varepsilon_{\max}$ is plotted on the $y-$axis, with each interval represented by the point $(\varepsilon_{\min}, \varepsilon_{\max})$. Intuitively, there is a line defined by $y = x$, below which points will not be plotted. This fact means that the grey region in Fig. 3 will never contain any points. The line $y = x$ has the interpretation that the feature must first exist before it can die. Reading these diagrams is as intuitive as reading the barcodes; the vertical height of the point from the line $y = x$ is analogous to the length of the interval, that is, the further a point is from the line $y = x$, the more the feature is persistent. An example of a birth-death diagram can be seen in Fig. 4.

On both the barcode and birth-death diagrams, it can be seen that one feature persists to ∞; this is because there will always be a fully-connected simplex that persists to infinity. There will be a value of $\varepsilon_{\mathrm{fc}}$ that results in a fully-connected simplex, where every vertex is connected to every other vertex. For values $\varepsilon > \varepsilon_{\mathrm{fc}}$ the simplex will remain fully connected, and therefore this will continue to infinity. The removal of this infinite interval is called the *reduced homology*. The reduced homology is required for use in many calculations.

The space of barcodes actually forms a metric space [9]; the distance between the barcodes is a measure of similarity of two barcodes. As the persistent intervals are invariant for a manifold, the data manifolds can be represented by their persistent homology. This notion of a metric space allows one to compare the similarity of manifolds. Metrics between barcodes are well established and the one used in this paper is the $p-$*Wasserstein distance* [21].

Definition 2.1. Given two barcodes B_1 and B_2. For $p > 0$, the $p-$*Wasserstein distance*, ∂_{W_p}, is given by,

$$\partial_{W_p}(B_1, B_2) = \left(\inf \sum_{Z \in B_1} d_\infty(Z, \phi(Z))^p \right)^{\frac{1}{p}}$$

where ϕ is a matching between B_1 and B_2, d_∞ is the supremum metric, and Z is a persistence interval in B_1 [9].

The Wasserstein distance is used to measure the similarity between two persistence barcodes, by computing the sum over their edge lengths. The q-Wasserstein distance is defined as the minimal value achieved by a perfect matching between the points of the two diagrams, where the value of a matching is defined as the qth root of the sum of all edge lengths to the power q [21].

2.4 Calculating Fractal Dimension

The first application of persistent homology in this paper is on calculating *fractal dimension*, which is a measure of how much space a set occupies. To gauge the idea of the fractal dimension intuitively, consider that a single point is zero-dimensional, a

line spanning two points is one-dimension, and the area enclosed by three points is two-dimensional, and so on. Now, suppose a curve, consisting of purely one-dimensional elements has infinite length and is bounded by a finite region. This curve may be believed to occupy more space than something that is only considered one-dimensional, but it also does not completely fill this arbitrarily-defined space. A logical conclusion is that the dimension of this curve is somewhere in *between* one and two, depending on how efficiently it occupies this area.

The first notable work on quantifying fractal dimension was by Hausdorff [22], where a measure of roughness was conceived; this was later expanded on by Falconer, and generalised for point clustering in [23]. Mandelbrot, popularised the idea of fractal dimension in the notable work [24].

The two most common methods of quantifying fractal dimension are the *Hausdorff* and *box-counting* dimensions. The box-counting method exploits the dimensional relations to scalability and the difference in occupied space over different scales. The box-counting method is very popular because of its ease of computation. The method of using the persistent intervals, which is described later, more closely represents the box-counting method.

To calculate the box-counting dimension, assume a space X. The smallest number of sets of diameter ϵ that can cover X, is referred to as $N(\epsilon)$. The scale ϵ determines $N(\epsilon)$, and the dimension of X determines the rate of change of $N(\epsilon)$. As the scale is altered, it is expected that,

$$N(\epsilon) \simeq c\epsilon^{-d_B} \tag{2}$$

for positive constants d_B and c. It is said that X has a box dimension of d_B. To solve for the dimension of the set, this equation can be rearranged to give,

$$d_B = \lim_{\epsilon \to 0} \frac{\log(N(\epsilon))}{-\log(\epsilon)}, \tag{3}$$

with the constant term, c, disappearing in the limit [23].

Inherently, fractals are infinitely-complex objects that are self-similar, often over an infinite range of scales. This view gives a problem when working with finite point-data sets, as only finite information can be captured in a finite point-data set. When zooming in excessively, points will become more sparse and the approximations will inevitably become less exact. In reality for finite sets, self similarity may only be visible over a small number of scales. This process can be seen later, in Fig. 9, where the self similarity is clear over two length scales, then thereafter points become sparse.

2.5   Fractal Dimension - Persistent Homology

Persistent homology can be used to calculate fractal dimension, given some point data sampled from a fractal. This idea first came from that of the *minimal spanning tree* (MST) [25, 26]. The MST method was shown as a viable method to calculate the fractal dimension of a set [27]. Numerous works have showed that the MST method is equivalent to using the zeroth homology group for calculating the fractal dimension [27–30]. Unlike using the MST method, the persistent homology can be generalised for higher-order homology groups, therefore returning more information regarding the topology of the fractal shapes.

Often when analysing the persistent data, the smaller intervals are discarded as noise, as these mean that this specific feature only persists for a short while. In the case of calculating fractal dimension, there is information contained in all of the persistent intervals, as the idea here is to deduce the fractal dimension of X_n points as $n \to \infty$ and seeing how the lengths of the persistence intervals in each homology group vary. The shorter intervals provide a good measure of how the local geometry is present in the finite random sample.

To calculate the fractal dimension from the persistent homology, the α-weighted sum for $\alpha > 0$, of the persistent intervals, in a given dimension should be computed, as follows,

$$E_\alpha^i(x_1, ..., x_n) = \sum_{I \in PH_i(x_1, ..., x_n)} |I|^\alpha, \tag{4}$$

$(x_1, ..., x_n)$ are n points sampled from a manifold, most interestingly one that exhibits fractal behaviour. $PH_i(x_1, ..., x_n)$ is the ith dimension persistent homology group. $|I|$ is the length of the interval from PH_i. The alpha-weighted sum of the persistence intervals tracks the rate at which the topological noise decays, and similarly to the box-counting method, this can be linked to the fractal dimension [31]. The value of alpha will give a larger weighting to longer intervals.

Definition 2.2. Let X be a bounded subset of a metric space and μ a measure on X. For all $i \in \mathbb{N}$ and a value $\alpha > 0$, which is used to give a weighting to larger persistence intervals in the calculation [31], the persistent homology dimension can be defined as,

$$\dim_{PH_i^\alpha}(\mu) = \frac{\alpha}{1-\beta}, \tag{5}$$

where,

$$\beta = \lim_{n \to \infty} \sup \frac{\log(\mathbb{E}(E_\alpha^i(x_1, ..., x_n)))}{\log(n)} \tag{6}$$

Here, $\sup$ refers to the largest value in the set. The operator $\mathbb{E}$, is used as the expected value of a random variable. Finally, $E_\alpha^i(x_1, ..., x_n)$ is the alpha-weighted cumulative sum over the n points. This result means that the embedding dimension $d = \dim_{PH_i^\alpha}(\mu)$, of a manifold can be calculated if $E_\alpha^i(x_1, ..., x_n)$ scales with $n^{\frac{d-\alpha}{d}}$ [29, 31].

For the case when $i = 0$, the persistence intervals are equivalent to the lengths of the edges in the MST. From the MST, there is a set of n vertices and a set of $n-1$ edges, where each edge spans two vertices. This work is built on ideas from Kruskal [25] and Prim [26] in the 1950s. The edges in the MST are equivalent to the persistence intervals in the reduced zeroth homology group.

The MST approach formulates the problem in terms of graph theory, with V being the set of *vertices* and E being the *edge-set*, connecting the vertices. Two vertices are referred to as *connected*, if a *path* connects them. A *path* is the successive joining of adjacent edges from one vertex to another, i.e, one can walk from one vertex to the other without leaving the path. If these edges form a closed loop, this is called a *circuit*. If a graph does not contain a closed circuit, it is called a *tree*.


(a) VR Complex   (b) MST
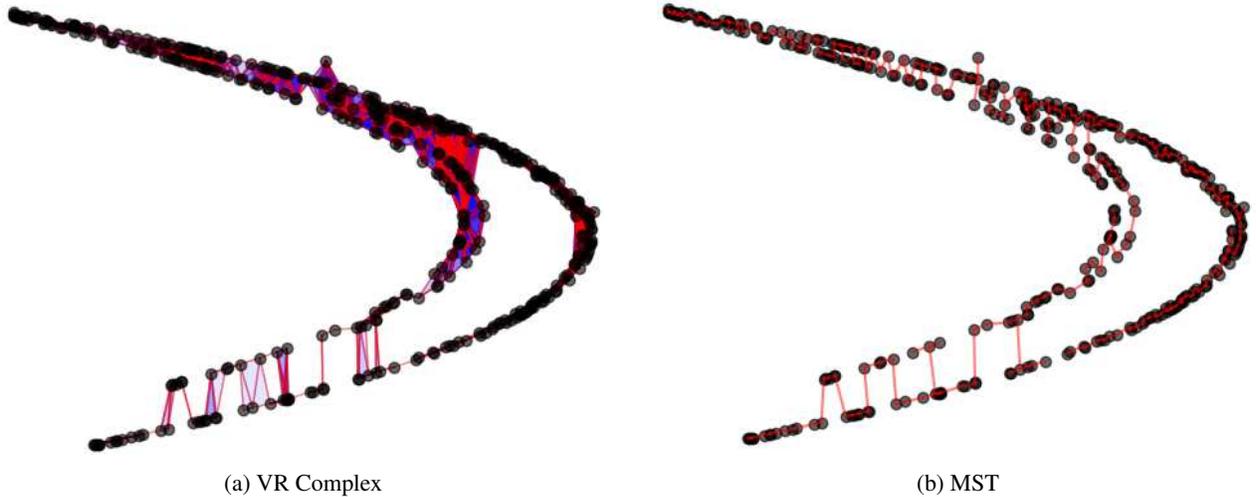
Figure 5: Hénon attractor constructed with the two methods of calculating fractal dimension.

A spanning tree has an attribute called its *length*. The length of a spanning tree is the sum of all the edges in the tree. The MST is the spanning tree that spans the points most efficiently, by minimising the length of the edges. The algorithms used to calculate MSTs are very well optimised, meaning that competitive results for calculating fractal dimensions can be obtained from their usage [31]. One can use the lengths of the edges in an MST to approximate the dimension of the manifold from which the vertices are sampled. The asymptotic behaviour of Eq. (4) is studied to calculate the fractal dimension.

This property is useful, as algorithms for calculating MSTs are much faster than calculating persistent intervals. However, the MST is only equivalent to the zeroth homology group. If more information is required for higher-dimensional homological features, the slower, but more informative persistence algorithms must be used. Fig. 5 shows the VR complex and the MST for an attractor.

When dealing with higher-order persistent homology, $i \geq 1$, things become more tricky. With PH_0, it is known that there are going to be $n-1$ edges for n points. Whereas, for higher-order persistent homology, this is not the case. There exist metric spaces where the number of persistent intervals grows faster than linearly in the number of points. To get around this problem, a limit for the upper bound can be proven. In the case of the VR complex, Schweinhart proved $|PH_1| = O(n)$ for the points $(x_1, ..., x_n)$ [29].

8

3 Attractors

Strange attractors arise from nonlinear dynamical systems. A dynamical system is one which is modelled (and evolves) in a phase space embedded in $\mathbb{R}^n$, where the geometric object is embedded in Euclidean space and parameterised by time. Attractors represent the state to which a dissipative dynamical system will eventually converge, seemingly regardless of initial conditions. Attractors are common-place for study as they describe the asymptotic behaviour of many dynamical systems [32–34].

3.1 Lorenz Attractor

The Lorenz attractor is defined by the differential equations,

$$\dot{x} = -\sigma \tag{7}$$
$$\dot{y} = \rho u - v - uw \tag{8}$$
$$\dot{z} = -\beta w + uv \tag{9}$$

The attractor was first discovered by Lorenz, when studying non-periodic turbulent flows [32]. It displays an interesting topology, with two holes being present in the manifold. It also shows a Cantor-set like behaviour over its cross sections [35] for certain parameters.


(a) Dense Lorenz Attractor.

(b) log-log plot to determine fractal dimension, with gradient (red).
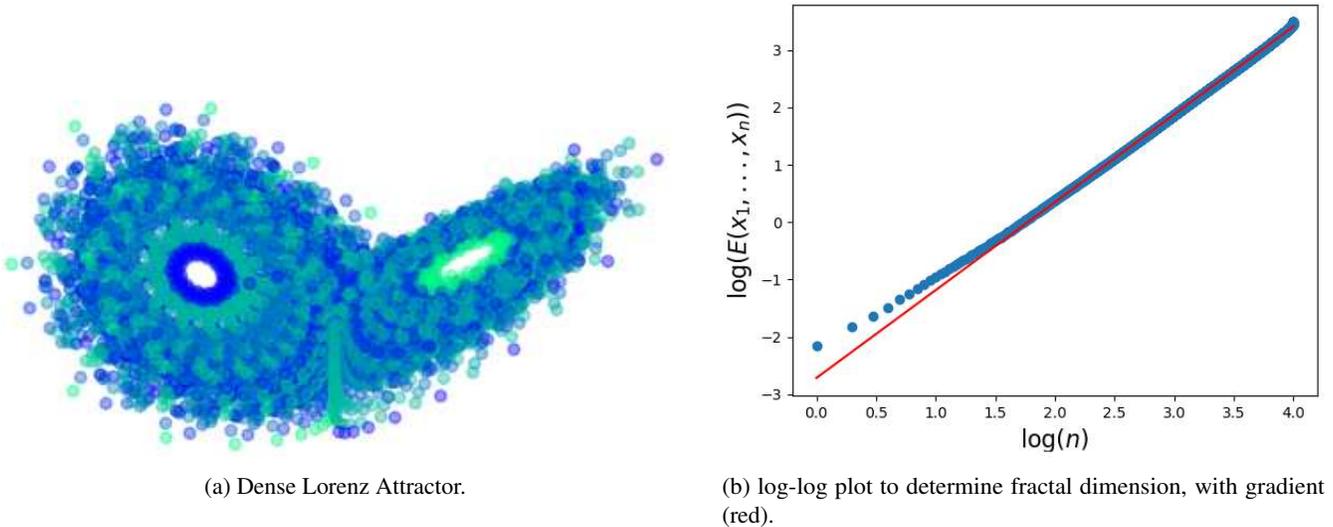
Figure 6: Lorenz Attractor.

For the system parameters $\rho = 28$, $\sigma = 10$, and $\beta = \frac{8}{3}$, the fractal dimension has previously been calculated by Viswanath [35] to be 2.063 . To obtain a structure with the fine details, showing the Cantor-like fractal cross sections, a large number of points are required for the embedding. 10000 points were calculated here for a dense embedding of the Lorenz attractor. Using the MST method of calculating the fractal dimension [31, 36, 37], of the embedded attractor shown in Fig. 6, an approximation to the Hausdorff dimension of 2.0826 ± 0.03603 was calculated. This value is calculated by determining the linear gradient of the log-log plot of the cumulative sum of the MST edges, which are analogous to the persistent intervals in PH_0.

Now, in order to capture the global topology of the system, fewer points are required to calculate the persistent homology. In this example, a more sparse Lorenz attractor is sampled with only 1000 points, shown in Fig. 7. These samples were taken over the same range as the previous case but now there is a greater time step between each point. This procedure ensures a relatively dispersed distribution of points over the Lorenz attractor. By taking a dispersed distribution of points on the manifold, the global topology of the manifold should become more clear.

(a) Lorenz Attractor.　　(b) Persistence Homology for the Lorenz attractor.
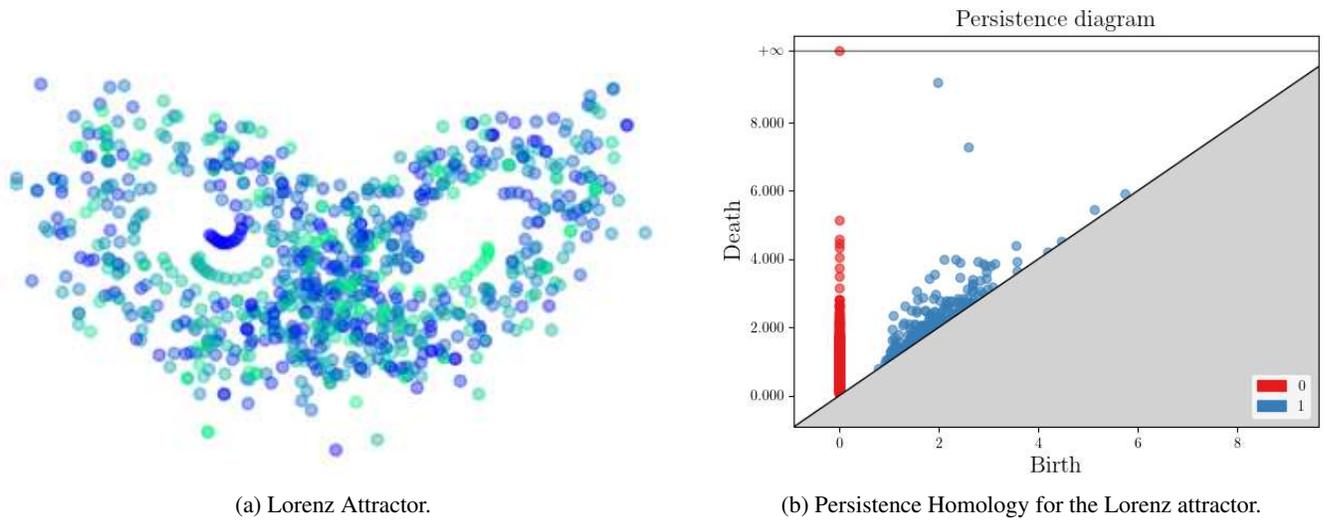
Figure 7: Lorenz Attractor.

The birth-death diagram can be seen on the right hand side of Fig. 7. There are no large differences in β_0, represented by the red points; this strongly indicates that the Lorenz attractor consists of only one connected component. For the case of the first homology group, there is a large amount of noise present close to the line $y = x$. Most interestingly from this plot, two features persist long enough to be deemed properties of the manifold. These points are $(2.59, 7.29)$ and $(1.97, 9.17)$. The holes represented by these points can be visibly seen on the left hand side of Fig. 7 as the holes present within the manifold. The smaller of the two intervals represents the void in the left-hand side of the plot.

3.2 Hénon Attractor

The Hénon attractor is a two-dimensional quadratic map with a constant Jacobian; it was first conceived as a simplified discrete map of the Lorenz system. As the map is discrete, it has become a common object for study in dynamical systems, as computations for generating a large number of points are fast [34]. The Hénon attractor used in this work is defined by,

$$x_{n+1} = 1 - ax_n^2 + y_n \tag{10}$$
$$y_{n+1} = bx_n \tag{11}$$

In this case, the well studied parameters of $a = 1.4$, $b = 0.3$, and an initial point of $(0.1, 0.3)$ were used, as these give a convergent solution. 2000 iterations were taken to give a finite approximation of the orbit of this specific Hénon attractor, with the initial conditions listed. Using the MST method of calculating the fractal dimension [31, 36, 37], a value of 1.2558 ± 0.04476 was obtained. A previously-calculated value from work by Grassberger [38], calculated the dimension to be 1.26.

(a) Hénon Attractor generated with 2000 points.

(b) log-log plot to determine fractal dimension, with gradient (red).
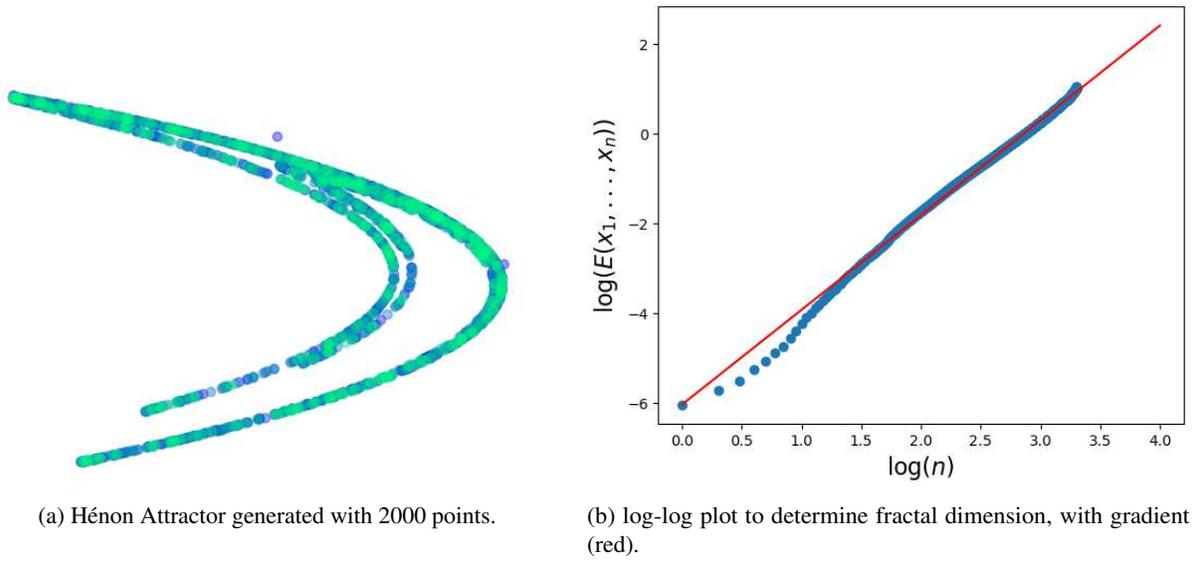
Figure 8: Hénon Attractor.

The Hénon attractor is well known to exhibit a self-similar, Cantor-like, behaviour over its cross sections [34]. The longitudinal structure of the Hénon attractor is simple, with each curve appearing to be a 1D manifold. The traversal structure is the interesting part; this can been seen to be self similar in Fig. 9.

(a) $x \in [0.5, 0.75]$, $y \in [0.15, 0.21]$.

(b) $x \in [0.62, 0.64]$, $y \in [0.185, 0.191]$.

(c) $x \in [0.63, 0.6325]$, $y \in [0.1889, 0.1895]$.
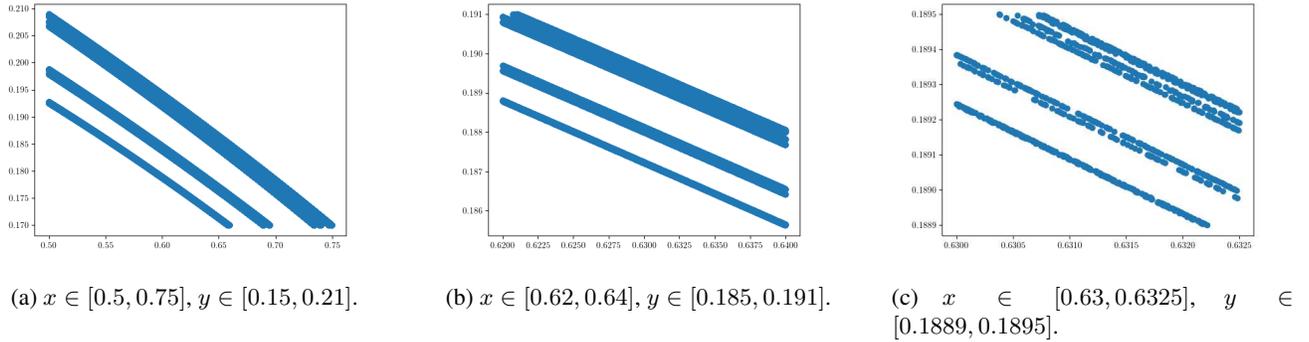
Figure 9: Hénon points sampled from three different scales, showing the Cantor cross section.

The dimensions of (a), (b), and (c) in Fig. 9 are 1.233, 1.245, 1.526; this shows that, as the data become more sparse, this fractal dimension calculation method becomes less accurate as more data points are missing at the smaller length scales.

3.3 Rössler Attractor

This attractor was first formulated in [33, 39], it is obtained by solving the differential equations,

$$\dot{x} = -y - z \tag{12}$$
$$\dot{y} = x + ay \tag{13}$$
$$\dot{z} = b + z(x - c) \tag{14}$$
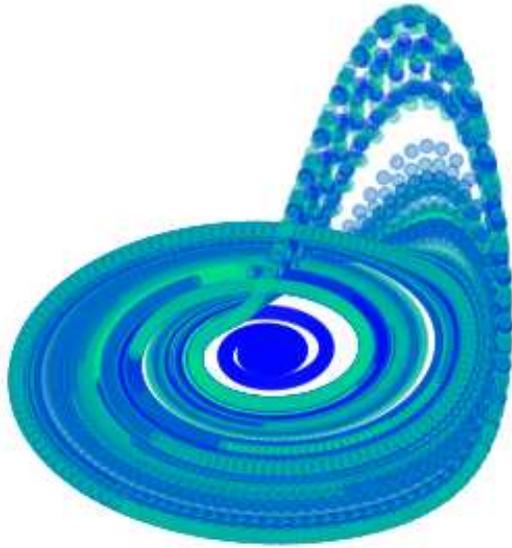
(a) Rössler Attractor.
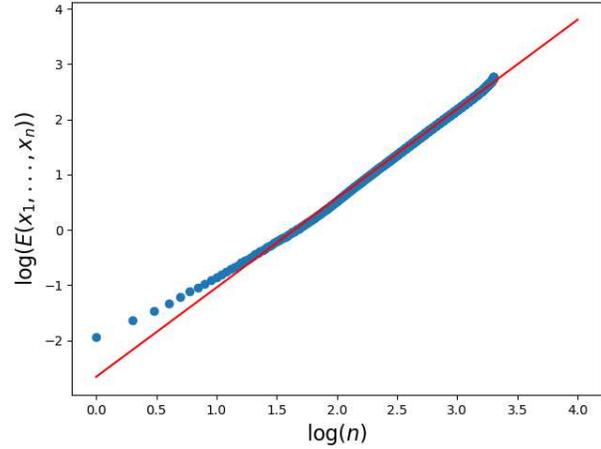
(b) log-log plot to determine fractal dimension, with gradient (red).

Figure 10: Rössler Attractor with 20000 points, used to calculate the embedding dimension.

The values for the initial conditions and system parameters were $a = 0.2$, $b = 0.2$, $c = 5.7$, with the initial point, $p_0 = (0, 0, 0)$. With 20000 points sampled from the Rössler attractor, the dimension estimate gives a value of 2.025 ± 0.0246 [31, 36, 37]. Kuznetsov calculated the dimension of the Rössler attractor, for these parameters to be 2.0160 [40].
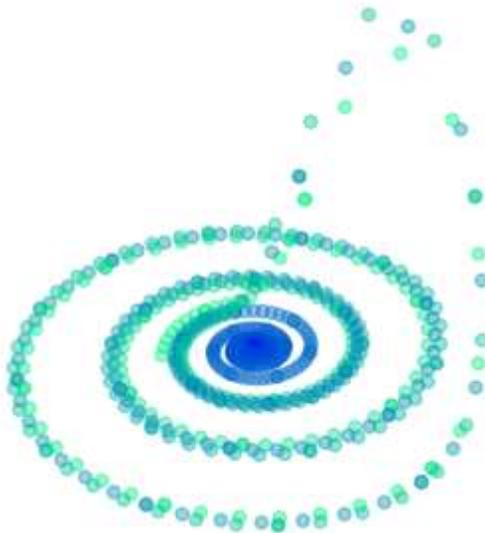
(a) Rössler Attractor.
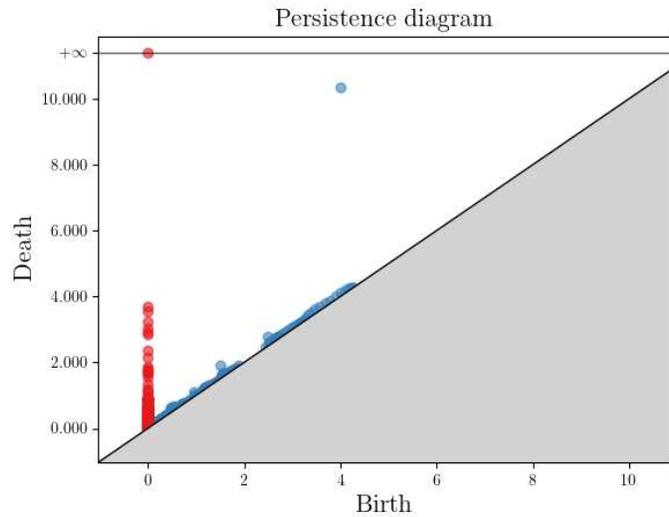
(b) Persistence Homology of the Rössler attractor.

Figure 11: Rössler Attractor, with 700 points, used to analyse the topology.

From looking at the persistence barcode in Fig. 11, it can be seen that all the red dots, corresponding to β_0 are close and show no clear split between these values. This observation strongly implies that the Rössler attractor is a fully-connected manifold. The range in which these occur is most likely because of the data points along the 'flick' of the Rössler attractor being more sparse; if this were a continuous embedding, these features would not be present.

For the case of the first-dimensional Betti numbers, represented by the blue dots in Fig. 11, there is a lot of topological noise

close to the line $y = x$. Interestingly, there is a point far from the line $y = x$, which is representative of the topology of the Rössler attractor. This point has the coordinates $(4.00, 10.3)$, and is representative of the hole formed over the 'flick' of the attractor. These persistence data will now act as the benchmark for the Rössler attractor, and multiple embeddings will be compared to this. The aim here is to minimise the Wasserstein distance (WD) between the reconstructed object and the original attractor. The smaller the WD, the closer the topologies.

3.3.1 Reconstruction from Time Delay Embedding

In this case, the topologies of a reconstructed phase space, formed from a 1D time series from the attractor, and the original point cloud are compared. Persistent homology will be used to determine the optimal embedding parameter from the 1D time series.

Time-delay embeddings were first introduced in [41] as a method of inducing geometry from a one-dimensional time series. This work was further expanded on by Takens [42], proving that the topology can be perfectly reconstructed for chaotic attractors. Skraba et al [43], then showed that persistent homology could be used with the time-delay embedding to yield useful topological results.

Given a time-varying series, $f : t \rightarrow \mathbb{R}$, the time-delay embedding can be stacked d times, each with a delay α, to give a new embedding $\phi : t \rightarrow \mathbb{R}^d$ where the new embedding is represented by,

$$\phi(t) = (f(t), f(t + \alpha), ..., f(t + (d-1)\alpha)). \tag{15}$$

From this embedding, the time series now has a topology induced. This embedding is a reconstruction of the topology from a one-dimensional time series into any desired dimension. Takens' theorem shows that under certain circumstances, this reconstructed attractor is homeomorphic to the original attractor [42].

The delay embedding highlights periodic recurrent features in the time series. Recurrent behaviour will be highlighted in the time-delay embedding as a loop. Persistent homology can then be used to quantify the size of these loops.

If the delay is too small, there will not be sufficient information to form meaningful topology, and the reconstruction will be too similar to a straight diagonal line. Therefore, by maximising the size of the holes present in the attractor by using the persistent homology, an optimal time-delay embedding can be determined. Conversely, using a too-large delay will result in nonsense, as the gap between the readings will be too large and will not show a local change over the manifold, resulting in a deformed attractor.

By computing a range of delay embeddings, the persistent homology of these can be calculated and then the WD can be used as a metric of topological similarity, to give the optimal delay embedding. Fig. 12 shows the WD between the original Rössler attractor and attractors formed from a delay embedding over a range of delays. By taking the delay for the minimal WD, this implies that the topology of the original attractor and the reconstructed delay embedding have the most similar topology. Therefore, the delay that minimises the WD is the one that gives the optimal delay embedding.
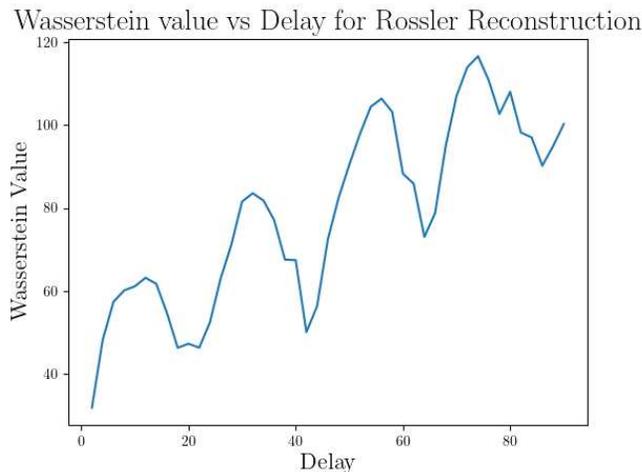

Figure 12: WD vs time delay embedding for Rössler.

13

There is a clear periodic trend emerging in Fig. 12; this trend occurs as the time-delay embedding comes in and out of phase with the original attractor. The general trend on Fig. 12 shows an increase in the WD as the delay is increased. This trend is because of artefacts induced in the reconstruction, as successive points are sampled over different periods from the attractor.


(a) $\alpha = 1$.
(b) $\alpha = 57$.
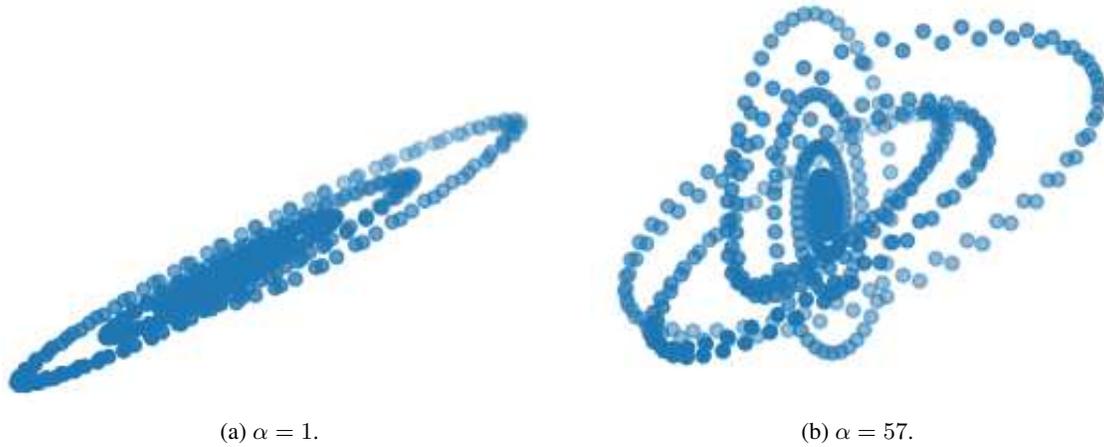
Figure 13: Examples of Bad Rössler Reconstructions.

From Fig. 12 it can be seen that a delay equal to one gives the smallest WD; in this case, the reconstructed topology is too similar to the diagonal set, and the topology is relatively uninteresting. If one was to assume an delay embedding of $\alpha = 0$, this would result in the straight line $x = y = z$, which will clearly have an uninteresting topology. Therefore, in this analysis, only values after the first peak ($\alpha = 13$) will be considered, as these ensure an interesting topology is being formed from the embedding. On the other hand, using a too-large value for the delay gives a deformed manifold. Two bad examples of delay embeddings, $\alpha = 1$ and $\alpha = 57$, can be seen in Fig. 13.

(a) $\alpha = 19$.
(b) Persistence Homology.
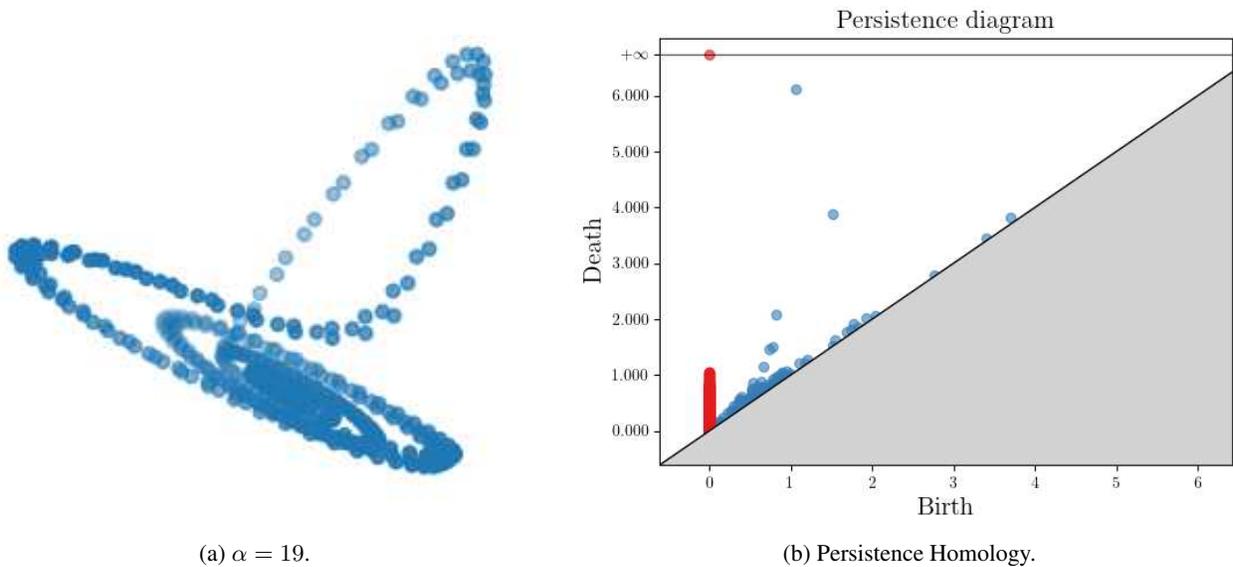
Figure 14: Optimal Reconstructed Rössler Attractor.

For $\alpha = 19$, not only the topology of the attractor is most accurately being captured, but from Fig. 14 (a), there is also a strong likeness to the original geometry. This is not likely to be the case for all other reconstructions, as Takens' only proved this for the topology [42].
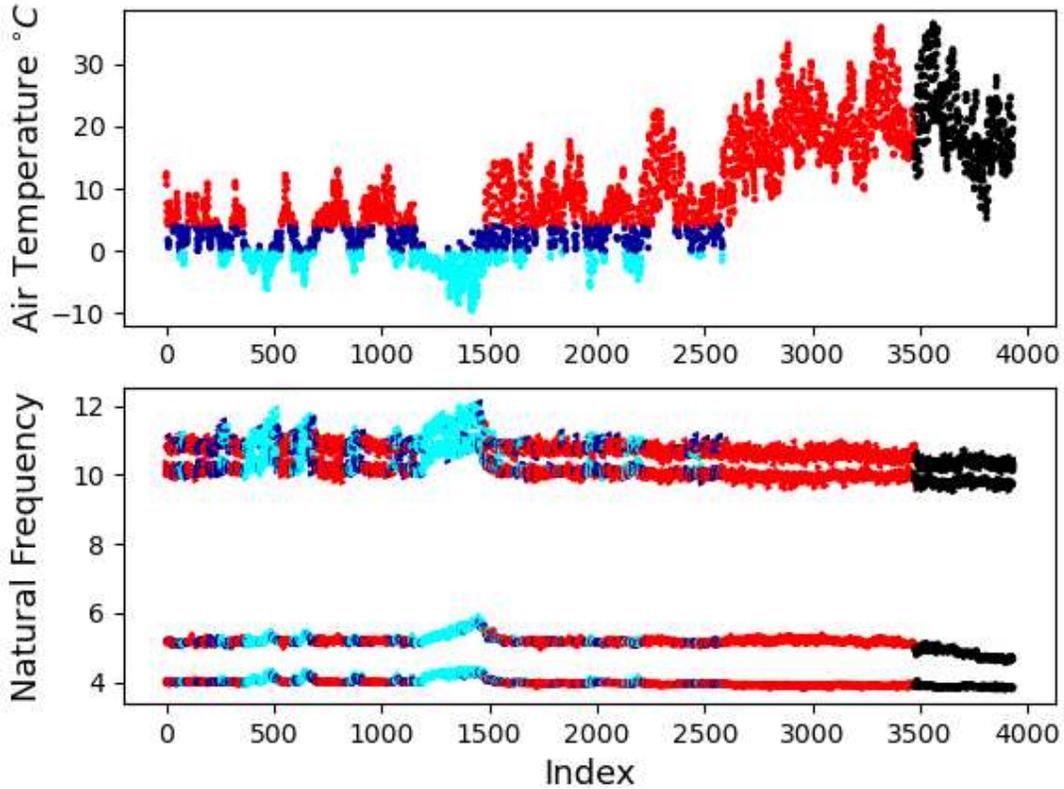
Figure 15: Natural frequency classes according to their temperature.

4 A Case Study of Interest for SHM: the Z24 Bridge

4.1 Introduction to the problem

The Z24 bridge was a structure connecting the two regions, Koppigen and Utzenstorf, in Switzerland. Data were collected over a whole year, with a sensor network placed over the bridge to collect modal parameters. Sensors were also used to measure the air temperature, soil temperature and humidity. Because of the extreme conditions of the Swiss weather, air temperature was recorded as low as $-9°C$ and as high as $36°C$. As a result, the temperature effects are clearly visible on the calculated natural frequencies [44]. Shortly before the destruction of the Z24 bridge, there was controlled damage introduced to the system, which is also visible in the natural frequencies after a certain point in time.

The change in temperature is the biggest contributing factor to the change in the natural frequency ($\approx 30\%$). The change in temperature has a greater impact than introducing damage ($\approx 7\%$). For this reason, the change in the magnitude of the natural frequencies offers little insight into the presence of damage. The main problem here is to separate the damage case from the temperature effects.

The data set can be broken down into four categories, according to the air temperature at the time of the measurements, and whether damage was present. Fig. 15 shows the temperature readings and the first four calculated natural frequencies; the corresponding colours refer to:

- Light blue making up the freezing data set. This is any value with a temperature reading $T < 0°C$.

- Dark blue, making up the cold data set. This is any value in the temperature range $0°C \leq T < 4°C$.

- Red, making up the warm data set. This is any value with a temperature reading $4°C \leq T$.

- Black, making up the damage-state data set. This is any reading taken after an index of 3475, irrespective of the temperature.

At every measurement instance, the natural frequencies $\{\omega_i\}$, can be calculated; where ω_i is the resonance frequency for the ith mode. The first set of the n natural frequencies can be represented as a point in $\mathbb{R}^n$, where the ith axis is for the value of ω_i.

As the Z24 data set has been sampled over different environmental and operational variations (EOVs), it is expected that the resonance frequencies will vary, depending on the environmental variables on the day. The bigger the change in the EOVs, the greater the change in ω_i. As the temperature changes, it is expected that material properties will be affected, therefore resulting in a change in the natural frequency.

In the case of the Z24 Bridge, by sampling over the time frame of a year, there will be slight changes in humidity, air temperature and soil temperature, that mean that each natural frequency will be slightly different from any other day. As the first four natural frequencies have previously been extracted for each reading [44], each point can be plotted in R^4. Plotting these points will trace out a manifold shape that is paramaterised by all of these EOVs. The data points are then assumed to lie on a manifold, representative of this specific bridge. TDA can then be used to form an understanding of what is expected for the shape of the natural frequency manifold for the Z24 data set.

Since the data have been partitioned into freezing, cold, warm, and damage, TDA can be used to compare the relative shapes of these manifolds. A significant change in the shape of the manifold could potentially be an indicator of the presence of damage.

4.2 Wasserstein Distance of Z24 partitions

There will be three case studies presented here, all on the Z24 data set. Each case study will include a new aspect of analysis to make the analysis more concrete.

1. The first case study will compare the manifolds embedded in four dimensional space, where each of the axes corresponds to a natural frequency ω_1, ω_2, ω_3 and ω_4. This case will also show how normalising by the number of points when using the WD makes a more robust metric for comparing data sets of different sizes.

2. The second case study will remove ω_2, and plot the manifold embedded in $\mathbb{R}^3$. Previous analysis of the Z24 bridge, has shown that a lot of the nonlinear features are present in the second natural frequency. This case study is useful, as now that the natural frequency data are embedded in three dimensions, the data can be plotted and visualised, whilst showing that topological arguments are still valid in lower-dimensional shadows.

3. The third case study shows the robustness of the manifolds to a linear dimension-reduction algorithm. In this case, the four-dimensional embedding of the data will be compressed down to two and three dimensions [45]. The reduced-dimension manifolds are then analysed with TDA.

4.2.1 Raw Natural Frequency Case

For the first case, a full walk-through of the calculation procedure will be displayed. For the succeeding cases, this will be omitted to limit repetition. This case will take the first four natural frequencies obtained from the Z24 data set [44]. The natural frequencies will be represented by a point in four-dimensional space. It is believed that the introduction of damage will change the shape of the manifold in a more substantial way than the change in temperature. TDA can be used here to compare between the different shapes of the partitioned manifolds.

The persistent homology of the manifolds are calculated accordingly. The partitions are not all of the same size. The warm data set is the largest data set, and this will be randomly split in half to form two data sets, the original warm data set will remain in the analysis. The second can be used to verify the results; the two new random subsets will have a very similar topological structure, as they are all sampled from the same manifold. There will be slight differences because of topological noise formed in the persistence intervals from missing points in the smaller samples. These effects should be negligible when over the true global structure of the manifold, there are enough points to adequately describe the topology. For smaller partitions, the absence of many missing points will affect the topology in a significant way.

	Freezing	*Cold*	*Warm*	*Damage*	*Warm1*	*Warm2*
Freezing	0.00	9.39	22.92	10.62	12.62	12.50
Cold	9.39	0.00	21.47	5.35	8.78	8.50
Warm	22.92	21.46	0.00	23.44	14.26	14.51
Damage	10.62	5.35	23.44	0.00	10.09	9.59
Warm1	12.62	8.78	14.26	10.09	0.00	1.80
Warm2	12.50	8.50	14.51	9.59	1.80	0.00

Table 1: WD values over each partition.

Table 1 shows the WDs over the different partitions of the data set. These values are relatively uninteresting, as the WD values are a factor of the number of points present in the data. This effect can be seen clearly between the warm and the warm subsets, as all the WD values for the full warm data set are roughly twice the size of the two random subsets. This effect shows that the number of points present in the point cloud is linked to the size of the WD value.

As a more informative measure, the WD values can be summed along the rows. Summing along the rows, gives an understanding of how different a data set is from all the others. The larger the value, the more different it is. As well as this, one can normalise by the number of points present in the data set; this gives a normalised value, independent of size of the data set. This measure then acts as a metric to discriminate between manifold shapes for manifolds with a varying number of points present.


(a) Raw.

(b) Scaled.
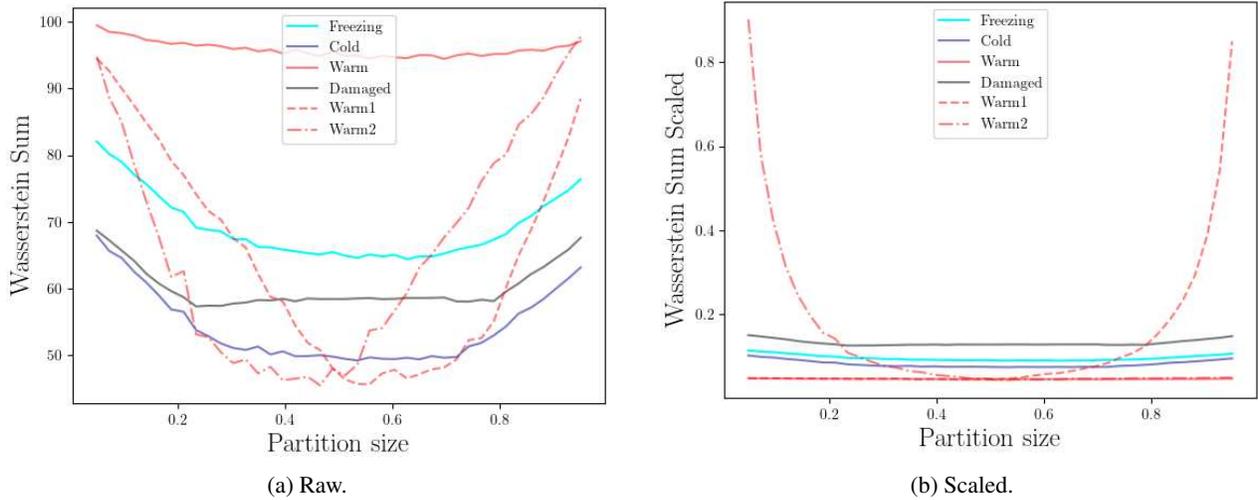
Figure 16: Size of the WD values depending on the size of the warm partition size.

Here, the change in the partition size and its effect on the WD value will be discussed, (Figure 16). The left-hand plot in Fig. 16 shows how the sum of the WD values for any data set varies as the partition size of the warm data set changes. It can be seen that the very cold, cold, warm, and damage sets are all changing proportionally, as the partition size is varied, as is expected. On the other hand, the two warm subsets vary a lot as the partition size changes. For small partition sizes, the topology of the manifold is subsampled to an extreme and a likeness between the manifolds cannot be established.

The right-hand plot of Fig. 16 shows the sum of all the WDs for a data set, and then scaled by the number of points presents in the data set. This figure shows how dividing by the number of points almost perfectly maps the warm subsets to the full warm set, in the regions where there are enough points to adequately describe the topology.

	WD Sum	*Number of Points*	*Scaled WD Sum*
Freezing	68.050	720	0.095
Cold	53.481	666	0.080
Warm	96.584	2089	0.046
Damage	59.095	457	**0.129**
Warm1	47.549	1044	0.046
Warm2	46.903	1045	0.045

Table 2: Summed and scaled WD data.

As can be seen here, the damage case is clearly the most different in terms of manifold structure, compared to the other data sets. The freezing data are the next most distinct.

For reference, if the partitions are not included in the analysis, results are obtained as in Tables 3 and 4. The damage manifold comes out as less different from the other values, as this now contains a greater weight in the analysis. When the warm subsets were included, this gave a much larger weight to this condition. Despite this, the damage manifold still comes out as the most topologically-dissimilar manifold.

	Freezing	*Cold*	*Warm*	*Damage*
Freezing	0.00	9.39	22.92	10.62
Cold	9.39	0.00	21.47	5.35
Warm	22.92	21.46	0.00	23.44
Damage	10.62	5.35	23.44	0.00

Table 3: WDs with subsets not included.

	WD Sum	*Number of Points*	*Scaled WD Sum*
Freezing	42.922	720	0.060
Cold	36.203	666	0.054
Warm	67.817	2089	0.032
Damage	39.411	457	**0.086**

Table 4: Scaled summed WDs with subsets not included.

4.2.2 3D Shadow: ω_1, ω_3, and ω_4

Previous analysis of the Z24 Bridge data has determined that ω_2 contains the most nonlinear features. The analysis presented in the previous section can be repeated here, but after eliminating ω_2 and plotting the manifold in three dimensions. This restriction acts as a visual aid as the data can now be plotted; it also presents a case where a lower-dimensional shadow is taken of the data, and the topology is preserved over this type of subsampling.
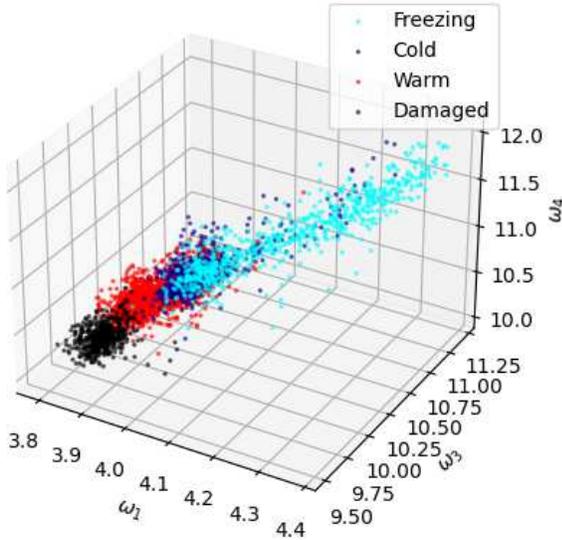
Figure 17: ω_1, ω_3, ω_4 3D plot, with the colours specifying data partitions.

	WD Sum	*Number of Points*	*Scaled WD*
Freezing	43.373	720	0.060
Cold	32.461	666	0.049
Warm	54.486	2089	0.026
Damage	41.078	457	**0.090**
Warm1	29.190	1044	0.028
Warm2	29.013	1045	0.028

Table 5: WD values with the highly nonlinear ω_2 removed.

This example shows the resilience of the TDA procedure; how a lower-dimensional shadow still contains enough information to distinguish the damage data partition as the most different.

4.2.3 Principal Component Analysis

In this section, a linear dimension-reduction algorithm, called Principal Component Analysis (PCA) [45] will be applied to the data. Naturally, when reducing the dimension of the data, some information will be lost. The $4D$ space will be reduced to the principal components present in $3D$ and $2D$; the results of both will be presented. This demonstration will show how the data's topological structure is preserved over linear transformations, and how the accuracy degrades as the dimension reduction becomes more distant from the true embedding dimension.

The first principal component can be thought of as the direction that maximises the variance of the data into a projected space. The successive principal components are the ones that maximise the variance of the data into the projected space that are also orthonormal to all the previous components.

To clarify how the topological structure is altered when taking the principal components, a case will be considered that uses some points randomly sampled from a torus. The example is shown in Fig. 18. This figure shows that as the PCA embedding dimension is reduced, topological information is being lost. When going from a $3D$ embedding to a $2D$ embedding, the volume enclosed by the torus is lost. Following this progression, when going from $2D$ to $1D$, the embedding is simply a straight line with a distribution of points weighted at the ends. These changes in the embedding dimension mean topological information in the dimensions higher than the current embedding are lost, i.e. in Fig. 18 (b) the $3D$-volume is lost, and in Fig. 18 (c), the $2D$-hole is lost. Despite this loss of information, the topology captured in the remaining principal components is still adequately represented, in Fig. 18 (b), the $2D$-hole is still visible, and in (c), the manifold still consists of one connected component.

(a) Regular torus.　　　　(b) First two principal components.　　　　(c) First principal component.
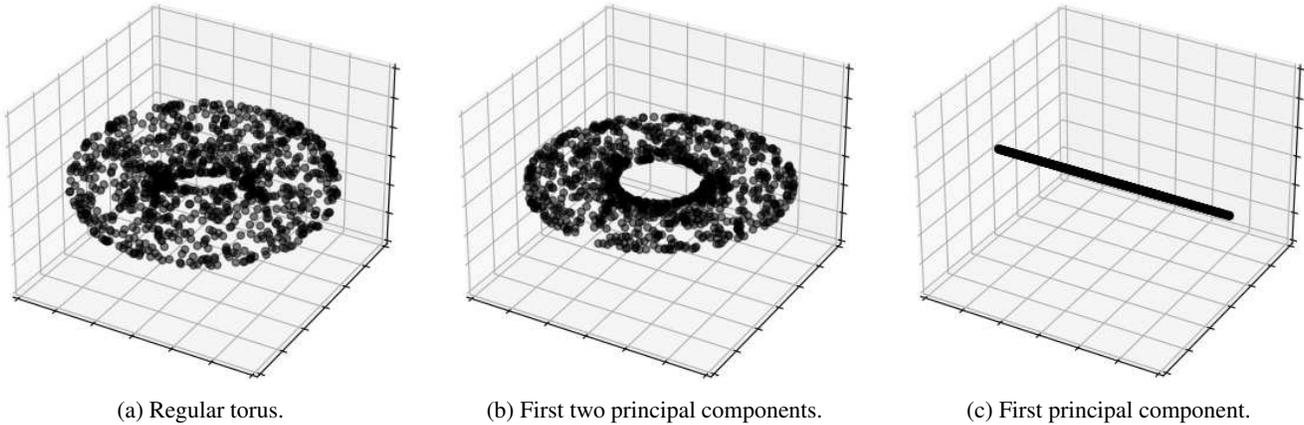
Figure 18: Torus principal components.

The first PCA example presented here on the Z24 Bridge data set, is for the dimension reduction from $4D$ space into $3D$ space. As can be seen in Fig. 19, calculating the PCA of the data has still kept the clusters in somewhat separate regions of space. Under close inspection, the warm, cold, and freezing all seem to be single clusters. Whereas, the damage cluster appears to be comprised of two clusters positioned very close to one another. This property is one that will result in the topology of the damage partition being different to the other clusters.
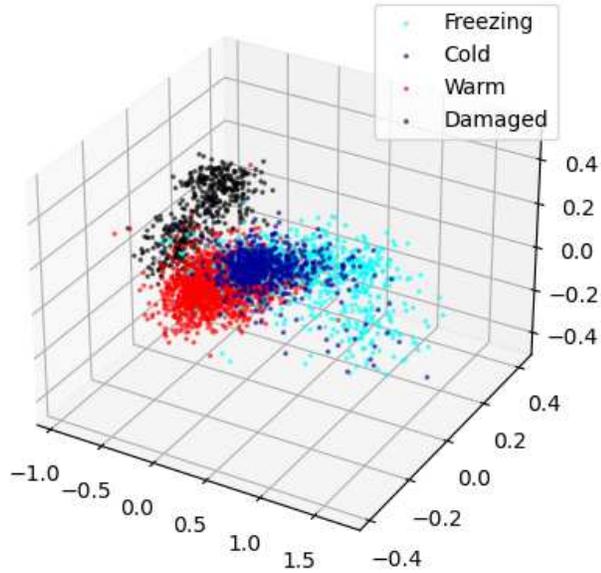
Figure 19: A visualisation of the first three principal components of Z24 natural frequency data.

	WD Sum	Number of Points	Scaled WD
Freezing	56.763	720	0.079
Cold	42.665	666	0.064
Warm	69.743	2089	0.033
Damage	44.878	457	**0.098**
Warm1	36.561	1044	0.035
Warm2	37.414	1045	0.036

Table 6: WD values for data reduced from $4D$ to $3D$.

Performing the PCA of the data is interesting, as it shows that the topology and persistent homology is conserved over the first three principal components. As can be seen in Table 6, using the principal components of the data still results in the same orderings between the different partitions.

Now for a larger dimension reduction, into $2D$ space. This result still preserves the topology, although the WD values are a lot less distinct between the different cases. This result is because more topological information is being lost, as the data is compressed down to only the first two principal components. Despite the loss of information, the ordering from the scaled WD calculations still remains the same.

Figure 20: A visualisation of the first two principal components of Z24 natural frequency data.

	WD Sum	Number of Points	Scaled WD
Freezing	28.074	720	0.039
Cold	19.816	666	0.030
Warm	25.156	2089	0.012
Damage	19.310	457	**0.042**
Warm1	16.637	1044	0.016
Warm2	16.043	1045	0.015

Table 7: WD values for data reduced from $4D$ to $2D$.

4.3 Time-Delay Embedding

In this section, the topology of the natural frequency manifold will be reconstructed from each of the $1D$ time series. For the purposes of visualisation, the Z24 time series data will be embedded in $3D$, so that an intuitive understanding of the topology can be formed.

Figures 21 - 24 show the delay embeddings of the first four natural frequencies, projected into three dimensions. The embeddings for ω_1, ω_3 and ω_4 show a similar structure at the delay value $\alpha = 75$. Whereas, ω_2 has a different topology defined around the colder temperatures. This can be inferred by the second mode having a more pronounced effect from the freezing temperature variations.
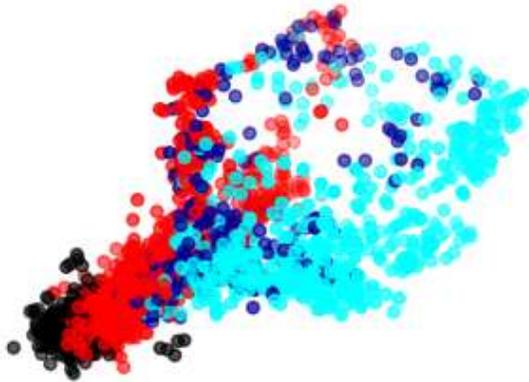

Figure 21: ω_1 Time Delay Embedding $\alpha = 75$.
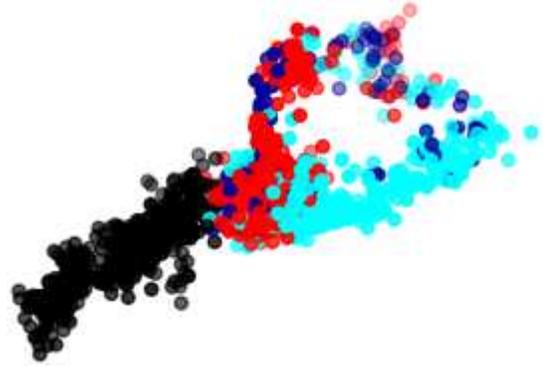
Figure 22: ω_2 Time Delay Embedding $\alpha = 75$.
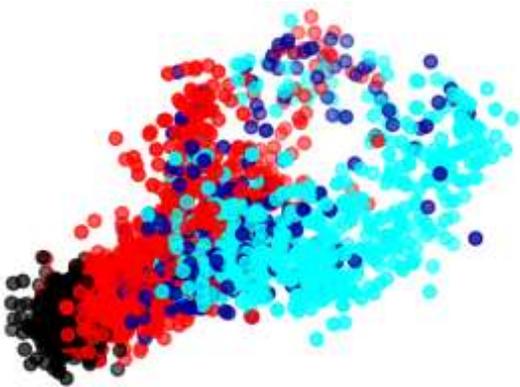
Figure 23: ω_3 Time Delay Embedding $\alpha = 75$.
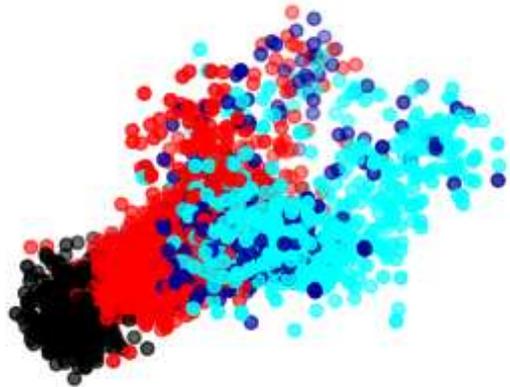
Figure 24: ω_4 Time Delay Embedding $\alpha = 75$.

When delaying the time series, this means that the delayed dimensions will have different temperature readings from the one-dimensional time series data used to form the embedding. The value for the temperature used to represent the point is the one from the initial time series data. The other dimensions will have their own respective temperatures. The classes used for the temperatures in these plots are a little less certain than in previous cases. This classification isn't too much of a problem, as the change in temperature is assumed to be a smooth change over each day, and as the value of $\alpha = 75$ is relatively small compared to the size of the data set, the time delay can be inferred as a local change.

5   Conclusion

The focus of this paper was to present an overview of TDA to the structural dynamics community, with a variety of use cases that highlight its depth and borderline limitless capabilities.

For the case of attractors, the calculation of the dimension of some common attractors was facilitated. A novel method for determining the optimal delay for constructing an attractor's topology from a $1D$ time series was also presented.

With respect to the Z24 data set, topological methods have been able to single out the damage data partition as the most topologically dissimilar. However, further analysis on topological methods for damage detection would need to be explored to understand the true limits and possibilities of TDA in SHM. An insight into the data structure provides powerful insight into the operating conditions of a machine or structure.

Future work on TDA, will look at different topological methods; aside from using the Wasserstein distance as a metric, other case studies will also be considered. A further journal paper, which extends on the ideas presented here will be submitted.

Acknowledgements

References

[1] H Edelsbrunner and J Harer. *Computational Topology: an Introduction*. American Mathematical Soc., 2010.

[2] H Edelsbrunner, D Letscher, and A Zomorodian. Topological persistence and simplification. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*. IEEE, 2000.

[3] C Maria, J D Boissonnat, M Glisse, and M Yvinec. The gudhi library: simplicial complexes and persistent homology. In *International Congress on Mathematical Software*, pages 167–174. Springer, 2014.

[4] U Bauer. Ripser: efficient computation of Vietoris-Rips persistence barcodes. *Journal of Applied and Computational Topology*, 2021.

[5] B Mendelson. *Introduction to Topology*. Courier Corporation, 1990.

[6] C Nash and S Sen. *Topology and Geometry for Physicists*. Elsevier, 1988.

[7] R Ghrist. Homological algebra and data. *The Mathematics of Data*, 25:273, 2018.

[8] M Saunders. *Homology*. Springer Science & Business Media, 2012.

[9] R Rabadán and A J Blumberg. *Topological Data Analysis for Genomics and Evolution: Topology in Biology*. Cambridge University Press, 2019.

[10] R W Ghrist. *Elementary Applied Topology*, volume 1. Createspace Seattle, 2014.

[11] J D Boissonnat, F Chazal, and M Yvinec. *Geometric and Topological Inference*. Cambridge University Press, 2018.

[12] F Chazal and B Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*, 2017.

[13] A J Zomorodian. *Topology for Computing*. Cambridge University Press, 2005.

[14] G Carlsson. Persistent homology and applied homotopy theory. *Handbook of Homotopy Theory*, 2019.

[15] R Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[16] M Gidea and Y Katz. Topological data analysis of financial time series: landscapes of crashes. *Physica A: Statistical Mechanics and its Applications*, 491:820–834, 2018.

[17] L Li, W Y Cheng, B S Glicksberg, O Gottesman, R Tamler, R Chen, E P Bottinger, and J T Dudley. Identification of Type 2 diabetes subgroups through topological analysis of patient similarity. *Science Translational Medicine*, 7, 2015.

[18] E Carlsson, G Carlsson, and V De Silva. An algebraic topological method for feature identification. *International Journal of Computational Geometry & Applications*, 16(04):291–314, 2006.

[19] E W Chambers, V De Silva, J Erickson, and R Ghrist. Vietoris-Rips complexes of planar point sets. *Discrete & Computational Geometry*, 44(1):75–90, 2010.

[20] B F Schutz. *Geometrical Methods of Mathematical Physics*. Cambridge University Press, 1980.

[21] T Lacombe, M Cuturi, and S Oudot. Large scale computation of means and clusters for persistence diagrams using optimal transport. *arXiv preprint arXiv:1805.08331*, 2018.

[22] F Hausdorff. Dimension und äußeres maß. *Mathematische Annalen*, 79(1):157–179, 1918.

[23] K Falconer. *The Geometry of Fractal Sets*. Cambridge University Press, 1986.

[24] B Mandelbrot. *The Fractal Geometry of Nature*. WH freeman New York, 1982.

[25] J Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956.

[26] R Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401, 1957.

[27] R Van de Weygaert, B Jones, and V Martinez. The minimal spanning tree as an estimator for generalized dimensions. *Physics Letters A*, 169(3):145–150, 1992.

[28] B Schweinhart. Fractal dimension and the persistent homology of random geometric complexes. *Advances in Mathematics*, 372:107291, 2020.

[29] B Schweinhart. Persistent homology and the upper box dimension. *Discrete and Computational Geometry*, 65(2):331–364, 2021.

[30] G Kozma, Z Lotker, and G Stupp. The minimal spanning tree and the upper box dimension. *Proceedings of the American Mathematical Society*, 134(4):1183–1187, 2006.

[31] J Jaquette and B Schweinhart. Fractal dimension estimation with persistent homology: a comparative study. *Communications in Nonlinear Science and Numerical Simulation*, 84:105163, 2020.

[32] E N Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20(2):130–141, 1963.

[33] O E Rössler. An equation for continuous chaos. *Physics Letters A*, 57(5):397–398, 1976.

[34] M Hénon. A two-dimensional mapping with a strange attractor. In *The Theory of Chaotic Attractors*, pages 94–102. Springer, 1976.

[35] D Viswanath. The fractal property of the Lorenz attractor. *Physica D: Nonlinear Phenomena*, 190(1-2):115–128, 2004.

[36] W B March, P Ram, and A G Gray. Fast Euclidean minimum spanning tree: algorithm, analysis, and applications. *KDD '10 Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010.

[37] R R Curtin, M Edel, M Lozhnikov, Y Mentekidis, S Ghaisas, and S Zhang. mlpack 3: a fast, flexible machine learning library. *Journal of Open Source Software*, 2018.

[38] P Grassberger and I Procaccia. Measuring the strangeness of strange attractors. In *The Theory of Chaotic Attractors*, pages 170–189. Springer, 1982.

[39] O E Rössler. An equation for hyperchaos. *Physics Letters A*, 71(2-3):155–157, 1979.

[40] NV Kuznetsov and TN Mokaev. A note on finite-time lyapunov dimension of the rossler attractor. *arXiv preprint arXiv:1807.00235*, 2018.

[41] N H Packard, J P Crutchfield, J D Farmer, and R S Shaw. Geometry from a time series. *Physical Review Letters*, 45(9):712, 1980.

[42] F Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

[43] P Skraba, Vin De S, and M Vejdemo-Johansson. Topological analysis of recurrent systems. In *NIPS 2012 Workshop on Algebraic Topology and Machine Learning, December 8th, Lake Tahoe, Nevada*, pages 1–5, 2012.

[44] B Peeters and G De Roeck. One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake Engineering and Structural Dynamics*, 30(2):149–171, 2001.

[45] S Wold, K Esbensen, and P Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):37–52, 1987.

List of Figures

List of Tables