Topological Data Analyses of Time Series Using Witness Complexes

by

N. F. Sanderson

B.A., University of California, Davis, 2012

M.A., University of Colorado, Boulder, 2015

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

2018

This thesis entitled:
Topological Data Analyses of Time Series Using Witness Complexes
written by N. F. Sanderson
has been approved for the Department of Mathematics

_____

Dr. Carla Farsi

_____

Dr. James D. Meiss

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Sanderson, N. F. (Ph.D., Mathematics)

Topological Data Analyses of Time Series Using Witness Complexes

Thesis directed by Dr. Carla Farsi

Real-time regime shift detection between chaotic dynamical systems via time series analysis demands quick and correct, theoretically guaranteed methods. Often the best implemented techniques in the field are well-motivated heuristics and even interpretable statistics are scarce. Topological data analysis can contribute to the canon of traditional methods for analyzing nonlinear time series but is not computationally cheap. We introduce a topological membership test for sliding windows of time series data that uses a sparse simplicial complex - the witness complex - to model the data and assess its performance across a range of model parameters affecting computational efficiency. We then explore how the topology of witness complexes changes across this range of model parameters. We next define a simplicial complex whose construction incorporates the temporal information available with time series data. We experimentally show that this construction results in filtrations with fewer simplices and improved topological signature. We apply our techniques to synthetic time series data including numerical solutions of classical low dimensional chaotic systems Lorenz and Rössler systems of ODEs as well as regimes of the higher dimensional Brunel neuronal network model and experimental live voltage recordings of musical instruments.

Dedication

To my parents.

Acknowledgements

I would like to thank my advisors for their years of guidance and patience. I would not be where I am today without their deep-rooted interest in developing scientific thinkers for future generations to come, and I am grateful to have had to the opportunity to learn from their motivating questions and expertise.

I would also like to thank Vanessa Robins for their mentorship. Their care and intentionality in research serve as an inspiration to me. I would also like to thank the Australian National University for hosting me in June 2017.

I would like to thank Kathryn Hess for inviting me to visit their team at the École polytechnique fédérale de Lausanne and the Blue Brain Project in October 2017. Their impressive classification results of network dynamics encouraged me to test my methods on time series measurements from this data set, which greatly enhanced this thesis.

I would like to thank Abby Thompson for suggesting graduate school.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

Important real-world phenomena evolve in time according to complicated relationships between a collection of variables. The safety of a nuclear reactor is dependent on the heat of the system generated from numerous atomic reactions being successfully contained, the stability of a bridge requires damping of the multitude of vibrational frequencies governed by weight load and windspeed, cooperation of millions of users ensures the security of the internet, and even the dynamics of the human brain and human heart are dictated by geometric and chemical-based interactions between billions of cells. Modeling these phenomena using equations of evolution allows one to obtain a qualitative description of such time-varying systems. The field of dynamical systems aims to classify systems according to their behavior over time and uses properties of this behavior to make predictions.

Regime-shift detection is the process of determining when changes in the dynamics occur. If a system is operating normally, a shift in dynamics can become a fault in the system. The bridge can collapse; the nuclear reactor can melt. Detecting such a shift can serve as a warning of larger changes to come.

Time series analysis aims to study dynamics under the restriction of having only a single measurement of the complex system. Whether this signal is the temperature of the nuclear reactor, the predominant frequency of the vibrating bridge, the number of users on a certain website or the number of heartbeats per minute, the constraints from having one measurement have been met with creative approaches for extracting impressive amounts of information about the complicated

underlying system. The goal of this dissertation is to combine topological data analysis (TDA) and time series analysis to address the challenges and utilize the unique structure of temporally ordered streams of data.

Topology is loosely the study of shape properties resulting from nearby connections. This perspective brings under scrutiny the notion of components, paths, cycles and holes. Most topological data analysis (TDA) methods model physical or abstract objects without a sense of temporal development. Modern research on TDA involving time series and dynamical systems data centers around periodic data or distinguishing between periodic versus chaotic dynamical systems. Topological analyses of dynamical systems rarely begin with time series and those beginning with time series do not often look for higher dimensional features to describe the dynamics.

Time and cost are critical in any practical data analysis. This is especially true when the intention is to rapidly detect changes in signals monitoring important human endeavors. As a step toward an online topology-based regime-shift detection algorithm, this research is driven by the task of efficiently classifying chaotic dynamical systems from time series via computational topology. The thesis of this dissertation is that a sparse model - the witness complex - can contribute to successful topological methods of time series analysis either with care in selecting model parameters or upon modifying the model.

In Chapter 4, we present a topological membership test for machine learning on sliding windows of time series data and investigate its strengths and limitations. The variability in topological representation of the same data set based on model parameters leads us to introduce heuristics for parameter selection in Chapter 5 that encourage the model to capture the global topology with minimal local topology, or noise, added to the signal. In Chapter 6, we define a new topological model that reduces the cost of obtaining a correct and consistent topological signature of chaotic dynamical systems from time series data. We explain our methods with experimental periodic time series data and test the methods on synthetic time series from classical low-dimensional chaotic dynamical systems. We demonstrate the limitations of our methods on non-stationary synthetic time series data from high-dimensional dynamics on a neuronal network. These data sets are presented

in detail in Chapter 3.

In the next Chapter 2 we review historical motivations for investigating chaos and techniques of studying dynamical systems and time series. We present the method of Taken's delay coordinate embedding in both theory and practice as it plays a key role in our topological approach for analyzing time series. We then recall constructions from TDA and present the statistics we implement as validation of the success of our novel topological techniques for time series analysis.

Chapter 2

Background

2.1 Historical Motivations for Studying Chaos

In the late 1800s, Poincaré wrestled with the gravitational 3 body problem. This interest in the movement of the stars led to his discovery of chaos. Working with differential equations to describe flows in phase space, Poincaré was concerned with both qualitative description and prediction of time-varying systems. He set out much of the groundwork of algebraic topology in 1895, namely homology, cohomology and the fundamental group, in his Analysis Situs, which aimed at answering "What makes a sphere a sphere?" and generally classifying topological spaces. A few of his mathematical contributions include return maps, stable and unstable manifolds, and periodic orbits [51].

In 1927, Birkhoff introduced the concepts of wandering and non-wandering sets for flows on manifolds, further exploring the nuanced ideas of attracting and invariant sets of dynamical systems [4] . In 1930s, the Morse-Hedlund papers introduced symbol spaces and shift maps as a means of studying geodesic flows, and from this developed symbolic dynamics [37]. An early goal of symbolic dynamics became to classify, up to conjugacy, the shift maps of a shift space of finite type. Meanwhile, the question of structural stability of flows under perturbation remained a focus of the global analysis community, including Kolmogorov and Arnold [31], [3] . In 1968 and 1970, respectively, Sinai and Bowen introduced Markov partitions to discretize the phase space of one-dimensional maps, shifting the some of the attention in symbolic dynamics back to dynamics from the fields of coding and information theory to which it had drifted [58], [6].

Central to the modern study of dynamical systems is the work of Stephen Smale who in the 1960s on the beaches of Rio continued the deep study of topology, proving sphere eversion, the Poincaré Conjecture for dimensions greater than or equal to five, and the h-cobordism theorem. In 1967, he published the important survey paper "Differentiable Dynamical Systems" that lays down the "what-is-what" with diffeomorphisms, flows and conjugacy[59]. Motivated by the Van der Pol oscillator, Smale defined the horseshoe map which, by squishing, stretching and folding, creates infinitely many periodic orbits amongst chaotic trajectories. Along with R.F. Williams, Smale also presented the solenoid as an expanding attractor of a hyperbolic dynamical system, a fruitful classification of dynamical systems upon which much successful mathematics has been based [70]. Around this time, the meteorologist Edward Lorenz had begun studying nonlinear weather models and in 1963 discovered a family of strange attractors, basically inventing chaos and making popular the idea of "the butterfly effect" [35]. Exemplifying the ideas of Poincaré in practice, Hnon then introduced an even simpler fractal attractor in a lower dimension[21].

These surprising examples of peculiar attractors that arose from the original intersection of pure thought and visual computer experimentation continue to motivate the creation of novel mathematics that strive to capture the essence of such dynamical systems to this day.

2.2 Dynamical Systems

A dynamical system consists of three pieces of data, (φ_t, I, X): "an **evolution rule** that defines a trajectory as a function of a single parameter (**time**) on a set of **states** (**phase space**)" [41]. The parameter I serves as an indexing set and has at minimum the structure of a monoid. The evolution rule φ_t is a family of maps $\{\varphi_t : X \to X\}_{t\in I}$ that must satisfy $\varphi_t \circ \varphi_s = \varphi_{s+t}$ and $\varphi_0 = identity$. That is, the monoid structure must be passed to the composition structure of the family of maps.

In the case of continuous dynamical systems, the indexing set is $\mathbb{R}$, the phase space is a smooth manifold $\mathcal{M}$, and the evolution rule is sometimes called a **flow**. Yet when we only have knowledge of the system in a single direction (i.e. its past or its future), our indexing set I is the

monoid $\mathbb{R}^-$ of $\mathbb{R}^+$ with operation addition. A discrete indexing set like $\mathbb{Z}$ produces a discontinuous evolution rule known as a **map**. The rules governing flows and maps are different. For example, chaos can arise in a minimum of three dimensions from a flow but can occur in two dimensions from a map. A dynamical system is called **deterministic** if the evolution rule is a function that sends each state to a unique state. This is in contrast to random dynamical systems where the evolution rule acts on a probability distribution over the set of states. A common technique for studying non-deterministic dynamical systems is to couple a noise model either additively or multiplicatively to a deterministic evolution rule to produce a stochastic differential equation.

Our focus in this thesis is on classifying chaotic deterministic continuous dynamical systems given discrete sequential measurements known as a time series. Classifying continuous deterministic dynamical systems can be tricky, and various measures of similarity and difference have been developed over the years. One method is to use topology and geometry to distinguish the attractors of dynamical systems. We explore and improve upon modern TDA techniques to this aim. This hot topic has a rich history. To contextualize and motivate our experimental analyses with theory, we recall the following definitions [41]:

- a set $\Lambda \subseteq X$ is **invariant under** φ if $\varphi_t(\Lambda) = \Lambda$ for all $t \in I$.
- the collection of all limit points of the forward trajectory, $\varphi_t(x)$ for $t > 0$, of a point x is called the **omega-limit set of x** and is denoted $\omega(x)$
- a compact set N such that $\varphi_t(N) \subset interior(N)$ for $t > 0$ is called a **trapping region**
- a set Λ is called an **attracting set** if there exists compact trapping region N containing Λ such that $\Lambda = \bigcap_{t>0} \varphi_t(N)$

Def. An **attractor** of a dynamical system (φ_t, I, X) is an attracting set Λ such that there exists some point $x \in X$ such that $\Lambda = \omega(x)$.

In certain cases, what is meant by "the topology of the attractor" is clear. For example, limit cycles have the topology of a circle; quasiperiodic dynamics with flows φ_t that have a Fourier

series-like expansion $\varphi_t = \varphi(t) = \sum_{m\in\mathbb{Z}^d} a_m e^{im\cdot\omega t}$ with incommensurate frequency vectors ω have attractors with the topology of tori. To describe strange attractors arising from chaotic dynamical systems, we must first define chaos. To be chaotic requires satisfying the following two properties:

- A flow φ_t is **topologically transitive on an invariant set** Λ if for every pair of nonempty open sets $U, V \subset \Lambda$ there is a $t > 0$ such that $\varphi_t(U) \cap V \neq \emptyset$.
- A flow φ_t is said to exhibit **sensitive dependence to initial conditions on an invariant set** Λ if there exists a fixed $r > 0$ such that for each $x \in \Lambda$ and any $\epsilon > 0$, there is a $y \in B_\epsilon(x)$ such that $|\varphi_t(x) - \varphi_t(y)| > r$ for some $t \geq 0$.

Def. A flow φ_t is **chaotic** on a compact invariant set Λ if φ_t is transitive and exhibits sensitive dependence to initial conditions on Λ.

The set of initial conditions of a dynamical system which approach an attractor is called its **basin of attraction**.

2.2.1 Lorenz & Rössler Equations

Early claims of chaos were often merely conjectures. Famously introduced by meteorologist Edward Lorenz in 1963 [35] to model atmospheric convection, the Lorenz attractor was not confirmed chaotic until 2002 in a computer-aided proof by Tucker [63]. The Lorenz system is a three dimensional system of nonlinear ordinary differential equations in dynamic variables $x(t), y(t), z(t) \in \mathbb{R}$ and constant parameters $\sigma, \rho, \beta \in \mathbb{R}$,

$$\dot{x} = \sigma(y - x), \dot{y} = x(\rho - z) - y, \dot{z} = xy - \beta z. \tag{2.1}$$

For $\sigma = 10, \rho = 28$ and $\beta = \frac{8}{3}$, solutions to Lorenz system for generic initial conditions within the basin of attraction are chaotic. Computer-generated plots of numerically integrated solutions to these equations showed confined yet unpredictable behavior. This led to the belief that the analytical solutions of the Lorenz system were in fact chaotic. Below in Figure 2.1 is a computer-

generated plot of a solution to the Lorenz equations with initial condition in the basin of attraction of the strange Lorenz attractor. Notice the striking features, like the two large wings of the "Lorenz butterfly" and the single holes in the centers of each of the wings. As well, note the subtle features, like the Cantor-like banding extending radially outward from the holes in the center of the butterfly wings.

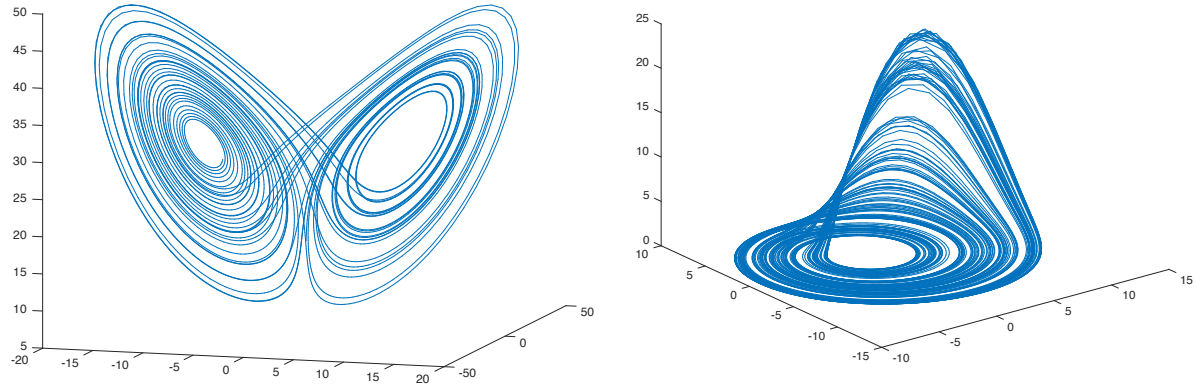


Figure 2.1: Typical trajectories of chaotic dynamical system spend time approaching a strange attractor according to its invariant measure: (a) Lorenz attractor (b) Rössler attractor.

Another classical example of a chaotic dynamical system introduced and proved as such in 1976 [52] is the Rössler system of equations created by medical doctor Otto Rössler that were later found useful for modeling chemical kinetics [53]:

$$\dot{x} = -y - z, \dot{y} = x + ay, \dot{z} = b + z(x - c) \tag{2.2}$$

for constant parameters $a, b, c \in \mathbb{R}$. With only one nonlinear term, the Rössler system is considered less complex than the Lorenz system. Chaos is observed for generic initial conditions when $a = 0.2, b = 0.2$ and $c = 5.7$. The behavior of a typical trajectory is to rotate about the equilibrium near the xy-plane, occasionally going up and down in the z-dimension as if over a bridge due to a far away equilibrium in the phase space that nonetheless affects typical trajectories of the Rössler system. An interesting feature is that the height of this bridge varies with each visit to this region

of the phase space. Figure 2.1 exhibits the Cantor-like banding of a trajectory approaching the Rössler attractor.

2.2.2 Lyapunov Exponents

One way to quantify the sensitive dependence on initial conditions necessary to prove chaos is by **Lyapunov exponents**, which report the exponential growth constant of the distance between the forward trajectories of two nearby points as a function of time. The Jacobian $D\varphi_t|_{x_*}$ associated to a flow $\varphi : X \to X$ at $x_* \in X$ defines a set of Lyapunov exponents $\{h_i(x_*)\}_{i=1}^k$ as the logarithm of the eigenvalues of the limiting matrix solution

$$\Lambda(x_*) = \lim_{t\to\infty} \frac{1}{t}(\Phi(t, x_*)^T \Phi(t, x_*))^{\frac{1}{2t}} \quad (2.3)$$

where $\Phi(t, x_*)$ is the fundamental matrix solution to $\dot{\Phi}(t, x_*) = D\varphi|_{\varphi_t(x_*)}\Phi(t, x_*)$ with $\Phi(0, x_*) = 0$ [34]. Oseledec's multiplicative ergodic theorem says that if the orbit of $x_* \in X$ under the flow φ generates an ergodic probability measure, this limit of matrix multiplication given by Λ exists and is the same for almost every x_* with respect to this ergodic measure [49]. We will revisit this notion of ergodicity momentarily, noting here that it allows us to capture descriptive information about dynamical systems from a generic single sequence of observations, rather than having to study the system behavior for all possible initial conditions. If we order the $\{h_i\}$ in decreasing order and consider

$$h_1 \geq h_2 \geq \ldots h_j \geq 0 \geq h_{j+1} \geq \ldots \geq h_k, \quad (2.4)$$

the associated eigenvectors corresponding to the first j Lyapunov exponents determine local directions in which stretching occurs while those corresponding to the last $k - j$ Lyapunov exponents determine locally contracting directions. Every autonomous flow has a zero Lyapunov exponent in the direction of the flow. Having the **maximal Lyapunov exponent** $h_1 > 0$ for a compact attractor is an indication of chaos.

In practice one uses computer algorithms like finite-time Lyapunov exponents or finite-size Lyapunov exponents to approximate these theoretical definitions [61]. The Lyapunov exponents for the Lorenz system are approximately (0.9056, 0, -14.5723), while the Rossler system has Lyapunov exponents approximately (0.0714, 0, -5.3943) [60]. How the fundamental matrices stretch and rotate a unit sphere to provide principle axes of various radii is in part captured by the set of Lyapunov exponents. What the Lyapunov exponents miss is how the rotational aspect of the sequence of matrices evolves. Rotations, leading to folding rather than stretching, are another geometric property one can measure of a dynamical system. A finite-time curvature algorithm recently introduced by Erik Bollt can computationally capture these features of chaotic dynamical systems [61].

2.2.3 Fractal Dimensions

Chaotic dynamical system also exhibit self-similarity which can be measured geometrically through the idea of dimension. One definition of dimension called the **Lyapunov (Kaplan-Yorke) dimension** is defined in [20] using the Lyapunov exponents as

$$D_{KY} \equiv k + \frac{\sum_{i=1}^{k} h_i}{|h_{i+1}|}. \tag{2.5}$$

where k is the largest integer for which $\sum_{i=1}^{k} h_i > 0$. For instance, the Lorenz and Rössler system have such $k = 2$ and the Kaplan-Yorke dimensions are 2.06215 and 2.0132, respectively [60]. This non-integer dimension shows fractal structure, here revealing the foliation of the wings of the Lorenz butterfly and the Rössler bridge that require just a little more than two dimensions to exist. Another notion of dimension is the **box-counting (capacity) dimension**.

Def. The **box-counting dimension** of a set S in a metric space is the absolute value of the limit as $\epsilon \to 0$ of exponential growth constant of the number of boxes $N(\epsilon)$ needed to cover a

set of points with a grid of boxes of length ϵ,

$$D_{\text{box}} \equiv -\lim_{\epsilon \to 0} \frac{\ln(N(\epsilon))}{\ln(\epsilon)}, \tag{2.6}$$

when this limit exists. When the limit does not exist, one defines the **upper box-counting dimension** and **lower box-counting dimension** as the lim inf and lim sup, respectively. This definition has the advantage of being easy to compute. Weighting the regions of the phase space that are visited more frequently leads to the one-parameter family of generalized dimensions D_q introduced by Reyni [20]. This approach to classifying dynamical systems exploits their measure-theoretic properties. The further structure of a measure is also useful for developing the ergodic theory of dynamical systems necessary for extrapolating information about the evolution of the entire state space from sequential observations from a single initial condition. Measurements reported by a time series are ideally of this form, making ergodic theorems crucial for analyzing dynamical systems from time series.

2.2.4 Ergodicity

A **measure preserving transformation**, $(X, \beta, \mu, \varphi_t)$ is a dynamical system with the additional data of a σ-algebra β of measurable sets of X for some measure μ such that φ_t is measurable and $\mu(\varphi_t(E)) = \mu(E)$ for all $E \in \beta$. Such a measure μ is an **invariant measure** of the dynamical system and a measurable set $E \in \beta$ is **invariant** if $\varphi_t^{-1}(E) = E$. When $\mu(X) = 1$ we have a **probability preserving transformation**.

Def. A measure preserving transformation is **ergodic** if every invariant measurable set or its complement has measure 0.

The ergodicity of the Lorenz attractor was proved by Tucker in 2002 when establishing its chaos through the existence of an invariant measure [63] . Note that such ergodicity does not hold for the Lorenz system of equations as defined on all of $\mathbb{R}^3$, only for the dynamics when restricted to the Lorenz attractor itself.

Additive ergodic theorems often say that integrating a measurement function over a trajectory of some dynamical system gives **in the time limit** the same result as integrating that measurement function according to the invariant measure of the dynamical system. The specifics of the type of function may change, but the existence and equality of these limits provides the ability to glean information from a single time series about an entire dynamical system. This "time average equals space average" property has been crucial for scientific advancement. The first published theorem of this nature was Birkoff's in December 1931 for L^1 functions achieving point-wise convergence almost everywhere [5].

Thm. (Birkhoff's Pointwise Ergodic Theorem): Let (X, β, μ, φ) be an ergodic probability measure preserving transformation and let $f \in L^1(\beta, \mu)$ be an integrable function. Then for μ-almost every $\omega \in \beta$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(\varphi_N(\omega)) \to \int_X f d\mu. \tag{2.7}$$

Months later in 1932, von Neumann published a mean version for L^2 functions and convergence in the operator norm [68].

Both of these are sufficient to guarantee the "time average equals the space average" phenomena the notion of ergodicity seeks to describe, yet they predate the modern study of dynamical systems which was born with the advent of computational experimentation in the 1960s and 1970s. When the study of dynamical systems began incorporating local analyses of the dynamics involving matrices in the tangent bundle, new theorems were needed to ensure the global application of such descriptions. To show that Lyapunov exponents are well-defined requires the multiplicative version of the ergodic theorem provided by Oseledec in 1968 [49], [13]. We recall the following definition for a cocycle of an dynamical system (X, φ, I) where X is an n-dimensional manifold.

- A **cocyle** is a map $A : (X, I) \to \mathbb{R}^{n^2}$ that satisfies $A(x, 0) = Id$ and $A(x, s+t) = A(\varphi_t(x), s) \cdot A(x, t)$ $\forall s, t \in I$.

Thm. (Oseledec's Multiplicative Ergodic Theorem) Let A be a cocyle over an ergodic

probability preserving transformation φ_t such that $log||A|| + log||A^{-1}|| \in L^1(\beta, \mu)$. Then there exist real numbers $\lambda_1 > \lambda_2 > \ldots > \lambda_k$ and φ_t-invariant subbundles $0 \subsetneq A^{\leq \lambda_k} \subsetneq \ldots \subsetneq A^{\leq \lambda_1} = A$ of A defined for almost every $x \in X$ such that for any vector $v \in A_x^{\leq \lambda_i} \backslash A_x^{\leq \lambda_{i+1}}$, we have

$$\lim_{t \to \infty} \frac{1}{t} log||A_x^{(t)} v|| \to \lambda_i. \tag{2.8}$$

To study the flow associated to continuous deterministic dynamical systems the cocycle A is often taken to be the Jacobian $D\varphi_s|_{x_*}$ for some choice of discrete time step s[69]. The existence of these subbundles invariant under rate of expansion guarantees that the Lyapunov exponents as computed from the fundamental matrix are the same regardless of the point on the attractor at which this fundamental matrix is based.

Coarse-graining the invariant measures on state spaces or attractors themselves and normalizing in various ways brings some ideas of probability and information theory back into the realm of dynamical systems.

2.2.5 Entropy

One of the central ideas of information theory is entropy. Introduced in 1948 by Claude Shannon to answer the question of how many binary questions one would need to answer to determine the state of a system, **information entropy** is colloquially the average amount of surprise in a system. For example, if you have a heavily weighted coin that flips **"heads"** $9/10^{th}$s of the time and **"tails"** only $1/10^{th}$ of the time, you would expect the coin to flip **"heads"** and be surprised very infrequently and unsurprised most of the time. However, if you have a fair coin that flips **"heads"** $1/2$ of the time and **"tails"** $1/2$ of the time, you would not have any reason to expect a certain outcome so the outcome of each coin flip surprises you. Having a uniform probability distribution over the set of states results in a maximal entropy system and having more states results in a higher maximal entropy. For instance, a fair coin flip has one bit of information entropy 1 while a fair die has $\log_2(6) \approx 2.58$ bits of entropy. More formally, for a discrete random variable

X with events x_j occurring with probability p_j, the **information (Shannon) entropy** is

$$H(X) = -\sum_j p_j \log_2(p_j). \tag{2.9}$$

The use of the logarithm in this definition allows independent events to affect the entropy additively. That is, flipping two fair coins produces 2 bits of entropy and rolling a pair of dice has about 5.16 bits of entropy.

Given a partition $Q = \{Q_1, \ldots, Q_k\}$ of the state space X of a measure-preserving dynamical system (X, β, μ, φ), the **entropy of the partition** Q is defined by

$$H(Q) = -\sum_{i=1}^{k} \mu(Q_i) \log_2 \left(\mu(Q_i)\right). \tag{2.10}$$

The **measure-theoretic entropy of a dynamical system** (X, β, μ, φ) then is defined as growth rate of the entropy of the refinement of the partition under pre-images of the flow

$$h_\mu(\varphi, Q) = \lim_{T\to\infty} \frac{1}{T} H\left(\bigvee_{t=0}^{T} \varphi^{-t}(Q)\right). \tag{2.11}$$

where the refinement of the flow given by $\bigvee_{t=0}^{k} \varphi^{-t}(Q) = \{A \cap B | A \in \bigvee_{t=0}^{k-1} \varphi^{-t}(Q), B \in \varphi^{-k}(Q)\}$ is defined iteratively. From this, the **Kolmogorov-Sinai (measure-theoretic) entropy** of a dynamical system is defined as the supremum of the entropy of the dynamical system with respect to a partition Q of X over all partitions Q. Like their topological and geometric counterparts, information and measure-theoretic quantities like entropy can help classify dynamical systems as chaotic as well as distinguish between different chaotic dynamical systems. A Kolmogorov-Sinai entropy $h_\mu > 0$ indicates a system is chaotic. The Rössler attractor has measure-theoretic entropy approximately $h_\mu = 1.18$ [15].

But what if all that we have is a single measurement from the dynamical system? What can still be deduced? Compressing the information from the original dynamical system by taking a single measurement and producing a time series is a transformation of the dynamical

system. Some of this compressed information can often be extracted by reconstructing the time series as a point cloud in higher dimensional space using Takens method of delay coordinate embedding, another transformation of the original dynamical system. Entropy is not invariant under many kinds of transformations. A feature of geometric properties like the Lyapunov exponent and box-counting dimension is that they are invariant under certain transformation of the phase space of the dynamical system. This is key for analyses of dynamical systems obtained from discrete scalar valued observations of the system also known as time series.

2.3 Time Series Analysis

Time series are important as measurements from scientists, economists, health practitioners, internet explorers, analysts of business, politics, social networks, the cosmos and beyond. Measurements are also fundamental in mathematics. Sometimes it is only possible to collect a single perhaps noisy measurement from a complicated system. This measurement might not directly report a driving force of the system, but important information about the system can still be extracted. A sequence of single measurements is called a time series. A time series is formally a particular outcome $\{x_t\}_{t\in T}$ of a stochastic process. Recall that a **real-valued random variable** $X : \Omega \to \mathbb{R}$ is a measurable function from a probability space (Ω, β, μ) to $\mathbb{R}$ with the Lebesgue measure.

Def. A **stochastic process** is a family of random variables $\{X_t\}_{t\in T}$ indexed by a time parameter and defined on a common probability space $\{\Omega, \beta, \mu\}$ for all $t \in T$.

2.3.1 Nonlinearity

Linear time series analysis aims to model a time series by a linear combination of predictors that can then be used for description and prediction. **Auto-regressive moving average** (ARMA)

models are **linear stochastic models** that come in the form

$$x_t = a_0 + \sum_{i=1}^{M_{AR}} a_i x_{t-i} + \sum_{j=0}^{M_{MA}} b_j \eta_{t-j} \tag{2.12}$$

where η_t are random Gaussian noise that add stochasticity to the otherwise deterministic linear process.

Predictability has often been considered a sign of determinism, and unpredictable systems are often classified as stochastic, or random. But linear predictors, like best fitting data to a line, are built to detect linear correlation among system variables as a function of time. Nonlinear structures often go undetected. Linear systems of ordinary differential equations have a solutions that are linear superpositions of exponentially damped or growing harmonic oscillations. Nonlinear systems of ordinary differential equations can produce more complicated behavior. In particular, drifts in system parameter values do not change the behavior of linear dynamical systems but can dramatically alter the dynamics of a nonlinear system, as studied in the field of bifurcation theory [32]. This suggests a need for specifically nonlinear time series analysis techniques.

A reason for the widespread use of linear statistics is that the **estimates**, or quantities calculated from sample data, can be rigorously related to the population descriptor, or **statistic**, given some assumptions are satisfied; this is not the case for many nonlinear estimates. Given a sequence of measurements $\{s_n\}_{n=1}^N$, one can compute linear estimates like the mean and variance,

$$\langle s \rangle = \frac{1}{N} \sum_{n=1}^{N} s_n \ , \ \sigma^2 = \frac{1}{N-1} \sum_{n=1}^{N} (s_n - \langle s \rangle)^2. \tag{2.13}$$

Both of these estimates aggregate information across the time series. To get information about the time evolution of the system, one can look at the **autocorrelation at time lag τ**

$$c_\tau = \frac{1}{\sigma^2} \langle (s_n - \langle s \rangle)(s_{n-\tau} - \langle s \rangle) \rangle = \frac{\langle s_n s_{n-\tau} \rangle - \langle s \rangle^2}{\sigma^2}. \tag{2.14}$$

This fluxuates periodically in τ for a periodic signal, achieving a minimum at $\frac{1}{4}$ the period. However,

the autocorrelation function behaves similarly for deterministic chaotic dynamical systems as well as for random processes, decaying exponentially, making this an unideal statistic for distinguishing between systems.

Rather than looking at the time series itself, one often performs a statistical analysis of the frequency spectrum of the Fourier transform, $\{\tilde{s}_k\}$, of the time series, $\{s_n\}$, to reveal the ratios of the major periodicities $\frac{N\Delta t}{k}$ present in the system

$$\tilde{s}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} s_n e^{2\pi i k n/N}. \tag{2.15}$$

This power density spectrum is the Fourier transform of the autocorrelation function. Since it only picks up linear correlation, or superpositions of simple sinusoidal signals present in a time series, it is not advisable for nonlinear data.

2.3.2 Non-Stationarity

Many linear methods of time series analysis assume stationarity. Stationarity in general means that all parameters relevant to the system's dynamics are fixed and constant during the period of observation [20]. The concept of **weak stationarity** requires that only up to second order statistics, like the mean and standard deviation, be constant. A **sliding window analysis** is often invoked to determine whether a sequence of samples $\{s_t\}_{t=0}^{N}$ comes from a stationary or non-stationary stochastic process. Typically an estimate of a statistic is computed over windows $W_i = \{s_t\}_{t=i\cdot k}^{(i+1)\cdot k-1}$ for $i \in \mathbb{N}$ and k equal to the **slizing window size**. A change in the estimate signals non-stationarity, or a change in the dynamics. This is called **change-point** or **regime-shift detection**.

A major technique for nonlinear time series analysis is to begin by reconstructing the discrete time series as a point cloud in a higher dimensional ambient space and to study properties of the dynamics from properties of the reconstruction. A simple, yet powerful, idea is to use successive delays of the time series itself as the further coordinates. We turn to this subject now, mentioning

briefly some alternative methods for nonlinear time series analysis that do not rely on geometry or phase space reconstruction include using neural networks and well-chosen sets of more complicated basis of functions, such as radial basis functions, to model the nonlinear process.

2.4 Delay Coordinate Reconstruction

Poincaré thought of being limited to observations of continuous dynamical systems confined to a subset of the phase space led to his construction of the return map, a discrete subsequence of the continuous time indexing set and a reduced state space comprised of the subset. Assume a continuous dynamical system $(M, \varphi, \mathbb{R})$.

Def. A **local Poincaré section** for a neighborhood U of a periodic point $p \in M$ is a hypersurface S of codimension 1 in M such that for all $x \in U$ there exists a $t \in I$ such that $\varphi_t(x) \in S$ and $\varphi_s(x) \notin S$ for all st and the resulting map $\rho : U \to S$ such that $x \in U$ maps to $\varphi_t(x) \in S$ is a diffeomorphism.

Def. The **Poincaré return map** for a neighborhood U of a periodic point $p \in M$ and a Poincaré section S is $\rho : U \to S$ such that $x \in U$ maps to $\varphi_t(x) \in S$.

In recording the location and times of a single trajectory revisiting a well-chosen subset of the phase space, the return map is a model reduction from which information about the original continuous dynamics can still be extrapolated. Delay reconstruction is a similar idea from about a century later in the opposite direction: given a time series, we can deduce information about the underlying continuous dynamics by mapping the sequence of measurements into a higher dimension.

The idea of delay coordinate reconstruction was put forth in [46] by Packard, Crutchfield, Farmer and Shaw in 1980. Working together at the University of California Santa Cruz, they experimentally studied a complicated fluid flow by taking measurements from a leaky faucet. Their claim was that, using successive delays of a discrete sequence of measurements from a continuous dynamical system, they could capture key aspects of the underlying dynamics from the vectorized data points in this m-dimensional reconstructed state space.

Def. The method of **delay coordinate reconstruction** takes as input a time series and

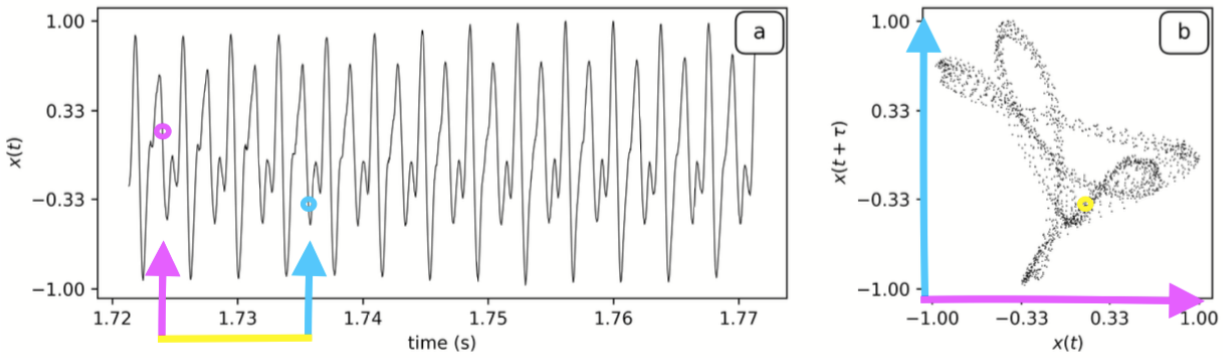


Figure 2.2: The method of delay reconstruction takes m points from a time series that are evenly spaced apart by τ and considers that as a single point in an m-dimensional space. We demonstrate this process for $m = 2$ where the delay amount τ is represented by the yellow bar between the two arrows pointing at the time series.

produces as output a sequence of vectors,

$$\{x(t)\}_{t=0}^{N} \mapsto \{x(t), x(t+\tau), \ldots, x(t+m\cdot\tau)\}_{t=0}^{N-m\cdot\tau}. \tag{2.16}$$

Here τ is known as the time-delay. We will say more below about the important issues of selecting τ, m and N. Published in 1981, Takens theorem asserts that the reconstructed set of points from the time series can be diffeomorphic to attractor of the underlying dynamical system being measured [47]. The catch is that this diffeomorphism is only guaranteed under some technical assumptions of genericity. These requirements are based on the dimension of the manifold and later in the work of Sauer et al [62] the box-counting dimension of the attractor and the major periodicities of the dynamical system, typically unknown features of a dynamical system from a time series measurement prior to analysis. This schism between practice and theory demands heuristics to determine these parameters and human intervention to implement the Takens method appropriately. While this can be expensive to execute, geometric features like the Lyapunov exponents and box-counting dimension are two properties of dynamical systems that can be extracted from time series measurements via Takens method of delay coordinate reconstruction, making it an attractive approach for analyzing nonlinear time series data.

2.4.1 Delay Coordinate Embedding In Theory

We restate the main theorems in Takens original 1981 paper [47], highlighting the assumptions on the dynamical system and genericity properties for which the results hold along the way.

Takens Theorem 1 Let M be a compact manifold of dimension m. For pairs $(\varphi, y), \varphi : M \to M$ and $y : M \to \mathbb{R}$, a C^2 diffeomorphism and C^2 function respectively, it is a generic property that $\Phi : M \to \mathbb{R}^{2m+1}$ given by $\Phi_{(\varphi,y)}(x) = (y(x), y(\varphi(x)), \ldots, y(\varphi^{2m}(x)))$ is an embedding.

Here Takens genericity condition is that any points $x \in M$ with period $k \leq 2m + 1$ are required to have distinct eigenvalues of $D\varphi^k|_x$ not equal to 1, no two fixed points can have the same function value under y, and $Dy|_x, Dy\varphi|_x, \ldots Dy\varphi^{2m}|_x$ must span $T_x^*(M)$ at fixed points x.

Takens Theorem 4 Let M be a compact manifold, X a vector field on M with flow φ_t and p a point in M. Consider the one-parameter family of diffeomorphisms $\varphi_\alpha : M \to M$ given by restricting the continuous indexing set of φ_t to integer multiples of α. Then for a residual set of α, the omega-limit sets $\omega_{\varphi_t}(p)$ and $\omega_{\varphi_\alpha}(p)$ are equal.

A residual set is the complement of a countable union of nowhere dense sets [14].

Takens Corollary 5 Let M be a compact manifold of dimension m. Consider (X, y, p, α), a vector field on M associated to flow φ_t, a function from M to the real numbers, a point on M, and a real number, respectively. For generic (X, y) and α, there is a smooth embedding of M into $\mathbb{R}^{2m+1}$ mapping the omega limit set $\omega(p)$ bijectively onto the set of limit points of

$$\{(y(\varphi_{k\cdot\alpha}(p), y(\varphi_{(k+1)\cdot\alpha}(p), \ldots, y(\varphi_{(k+2m)\cdot\alpha}(p)))\}_{k=0}^{\infty}. \tag{2.17}$$

This last corollary is what is needed to apply Takens theorem to time series analysis. Extensions to this theory provided by Davies [25] and Muldoon [38] allow for the theory to still hold in the face of noisy data when the underlying system is no longer a continuous deterministic dynamical system but a stochastic process. Yet in application technicalities still abound.

2.4.2 Delay Coordinate Reconstruction In Practice

When working with experimental data one rarely begins knowing the required information to satisfy Takens theorem and guarantee a diffeomorphism between the reconstructed data and the underlying attractor. The following are classical heuristics used for determining an appropriate reconstruction dimension and delay.

2.4.2.1 False Nearest Neighbors

The **false nearest neighbors algorithm** due to [39]is a statistical method for determining the efficacy of an m-dimensional reconstruction. In too low of a reconstruction dimension, projection can cause problems with achieving a embedding. Usually one defines a neighborhood around each data point based on the standard deviation as a regularization technique to discount outliers. One then determines clusters of similar data points within that neighborhood. These clusters seek to identify the points in the neighborhood that remain near each each other when the reconstruction dimension in increased. If increasing the reconstruction dimension unfolds a projection, some points in the neighborhood of the projection should become farther apart with the addition of this new dimension. To capture this, one computes the ratio of the distances between a point and its nearest neighbor when the time series is embedded in dimension $m+1$ versus m. This ratio is then compared to a threshold r. If this ratio is larger, [39] call this a **false nearest neighbor**. Averaging the count of false nearest neighbors over the data set and normalizing by the size of the neighborhood one gets the statistic

$$X_{fnn}(r) = \frac{\sum_{n=1}^{N-m-1} \Theta\left(\frac{|\mathbf{s}_n^{(m+1)} - s_{k(n)}^{(m+1)}|}{|\mathbf{s}_n^{(m)}|} - r\right) \Theta\left(\frac{\sigma}{r} - |\mathbf{s}_n^{(m)} - \mathbf{s}_{k(n)}^{(m)}|\right)}{\sum_{n=1}^{N-m-1} \Theta\left(\frac{\sigma}{r} - |\mathbf{s}_n^{(m)} - \mathbf{s}_{k(n)}^{(m)}|\right)}, \tag{2.18}$$

where Θ is the Heaviside step function.

This algorithm takes $\mathcal{O}(k^m)$ computations where m is the reconstruction dimension and k is the number of nearest neighbors we are averaging over at a single data point.

2.4.2.2 First Minimum of Time-Delayed Mutual Information

The **first minimum of the autocorrelation function** discussed in [20] can be used to establish a delay τ that creates maximally uncorrelated embedding dimensions, but this only accounts for linear relationships between the dimensions. For nonlinear time series, [2] suggest using the **first minimum of the time delayed mutual information**. Consider a partition of $\mathbb{R}$ into intervals of length ϵ with P_i denoting the i^{th} element of the partition.

Def. The **time-delayed mutual information** of a time series $\{s_n\}$ is a function of the delay parameter τ given by

$$I_\epsilon(\tau) = \sum_{i,j} p_{i,j}(\tau) \ln p_{i,j}(\tau) - 2 \sum_i p_i \ln p_i \tag{2.19}$$

where p_i is the probability that $s_t \in P_i$ and $p_{i,j}(\tau)$ is the joint probability that $s_t \in P_i$ and $s_{t+\tau} \in P_j$. The first minimum of the time-delayed mutual information identifies a time scale at which the time series is least predictable given past information. Choosing the first minimum helps ensure that this lack of predictability is due to the multiple state space variables governing the dynamics being measured rather than the randomness introduced by taking a delay too large compared to the time scale of the system. Mutual information can also be thought of as reporting the decrease in entropy of the stochastic process $\{X_t\}$ given knowledge of the stochastic process $\{X_{t-\tau}\}$, that is

$$I_\epsilon(\tau) = H(X_t) - H(X_t|X_{t-\tau}). \tag{2.20}$$

Pecora and Moniz [33] introduce a statistically verifiable method for selecting the reconstruction dimension m and delay τ in tandem that incorporates the topological concept of continuity. In their algorithm, the delay parameter τ is actually a delay vector $\{\tau_k\}_{k=1}^{m-1}$ which allows the delay to vary for each of the m coordinates.

Garland recently demonstrated that **for predictive purposes** incomplete reconstructions that do not satisfy the theoretical demands of Takens theory can actually outperform proper delay

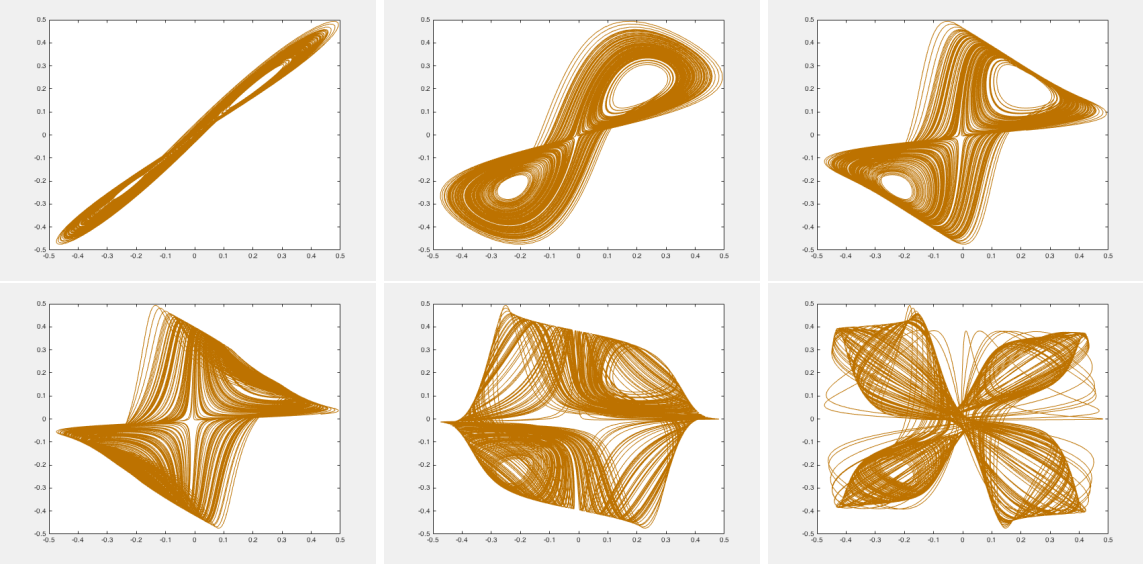


Figure 2.3: A wide range of representations of the Lorenz attractor can result when varying the delay parameter in the process of delay coordinate reconstruction of time series data.

coordinate embeddings [22]. We are intrigued by this philosophy of creating lemons to make lemonade, and exploiting the cheap algorithm implementations that result. Our need to play fast-n-loose with requirement is ultimately driven by the task of online regime-shift detection - classifying dynamical systems in real time from a stream time series data - where the size and speed of the analysis is of great importance.

This brings us to the primary investigation of this thesis: **How can we exploit the unique structure of time series data to create new TDA methods better at extracting topological and geometric properties of discrete representations of attractors obtained via Takens delay reconstruction on measurements from continuous dynamical systems?**

2.5 Topological Data Analysis

Classic work spanning the 1980s through the early 2000s by Birman, Muldoon and Mindlin recognized the need for coarse-grained approximations to provide topological signatures of chaotic dynamical systems and time series [29], [42], [10]. Technical developments in the field of computational topology by Robins, Edelsbrunner and Carlsson led to an explosion of specialized interests

and application in both academia and industry [54], [19], [16].

Time series data obtained from measurements of dynamical systems have been studied using persistent homology in [27] and [18]. Recent work by Perea et al [27] uses the 1-dimensional persistent homology of delay reconstructions to quantify periodicity. Research by Neville et al [18] uses the 0-dimensional persistent homology of filtrations of sub-level sets to classify chaotic dynamical systems. Persistent (co)homology has been applied to study dynamical systems to distinguish between periodic limit cycles and fixed points [24] to identify circle structures [66] and to detect chaotic from non-chaotic regimes [57], [30]. Most of these analyses use a Vietoris-Rips complex construction, a simplification of the Čech construction.

2.5.1 Topological Modeling of Data

A creative aspect of data analysis is developing a model from which to glean information about the data. Simplicial complexes are advantageous models because they represent the data using smaller pieces whose relationship to each other can be represented by matrices in a computer. Many analyses assume the data lies in a metric space and this provides a geometric scale-parameter from which to model data topologically.

Common geometric scale-based constructions of simplicial complexes from point cloud data that result in a filtration include the Čech complex, Vietoris-Rips complex, and alpha complex. The construction of the Čech complex looks for non-trivial higher order intersections of balls of radius ϵ centered at the n data points, a task that requires $\mathcal{O}(2^n)$ computations for a set with n data points. While the Vietoris-Rips complex simplifies this task by looking only for pairwise intersections of balls of radius ϵ to produce the 1-skeleton and then constructing the clique complex from that, this clique construction is less restrictive so there are more simplices in the Vietoris-Rips complex. Both the Čech and Vietoris-Rips complexes can contain simplices of far greater dimension than the dimension of the ambient space of the data set, up to dimension $n-1$. Partitioning the ambient space into cells demarcating the regions nearest to each data point, the intersections formed by regions in the ambient space equidistant to two or more data points create the cell complex structure of the

Voronoi diagram. The alpha complex reduces the dimension of the simplicial complex constructed by restricting the ball of radius ϵ centered at each data point to its intersection with that data points Voronoi cell. The alpha complex can also be thought of a subset of the Delaunay complex, or the nerve of the Voronoi diagram, which takes $\mathcal{O}(n \log n)$ to compute in dimension two or less. However, computing the Voronoi diagram is $\mathcal{O}(n^{\frac{m}{2}})$ in dimension $m \geq 3$ [44]. Current research is to adapt computations for the two-dimensional construction to the torus and hyperbolic plane and make three-dimensional computations tractable for large data sets [40].

Recall a presheaf with coefficients in Abelian groups is a contravariant functor from the category of open sets of a topological space X with morphisms the inclusions to the category of Abelian groups,

$$\mathcal{F} : \mathcal{O}_X \to \text{Ab}. \tag{2.21}$$

The Čech cohomology of a particular open cover $\mathcal{U}$ of X can be define by taking as the n-cochains maps from the intersections of n open sets in the $\mathcal{U}$ into Ab and as the coboundary operator from the k-cochains to $k+1$-cochains the alternating sum of the subgroups associated to the n-cochains obtained by restriction. The Čech complex construction in topological data analysis described above takes the nerve of the particular cover $\mathcal{U} = \{B(x_i, \epsilon)\}_{i=1}^n$ consisting of open balls of radius ϵ centered at each of the $\{x_i\}_{i=1}^n$ data points. From the Borsuk-Weil Theorem, this nerve is homotopic to the cover $\mathcal{U}$ [48],. Typically the constant presheaf that sends all open sets to the same Abelian group is used to compute Čech cohomology of an open cover $\mathcal{U}$ of a data set, as this corresponds to the simplicial homology of the nerve. Yet modeling data using less trivial sheaves or cosheaves and computing the sheaf cohomology or cosheaf homology can have interesting applications [17], [9], [55].

When a construction of a simplicial complex is dependent on a scale parameter ϵ that produces a filtration, or nested sequence of simplicial complexes, $\mathcal{F} = F_{\epsilon_0} \subset \ldots \subset F_{\epsilon_{\max}}$ as a function of that parameter, one can compute the persistent homology as the image of the induced map on homology

of the inclusions in the filtration,

$$\mathcal{PH}_k^{ij}(\mathcal{F}) = Im(\iota_*^{(i,(j-i))} H(F_i)). \tag{2.22}$$

Our aim is to do fast classification of dynamical systems from a sliding window topological data analysis of time series. To that end, we exploit the witness complex construction that begins by reducing the number of vertices in the construction from n to the size of a selected subset of the data points called **landmarks**.

2.5.2 Witness Complexes

The **strict witness complex** originally introduced in [65] was a stagnant construction. Let X be a data set in some metric space with cardinality $n, \{\ell_0, \ldots, \ell_{N-1}\} = L \subset X$ a size N subset of the data points called **landmarks**, and A the $n \times N$ distance matrix. We discuss the matter of landmark selection in Chapter 5.

Def. An edge is in the **strict witness complex** $W_\infty(A)$ if there is a data point $x \in X$ whose two nearest landmarks are the vertices of that edge. Similarly, a k-simplex is in $W_\infty(A)$ if there is a data point $x \in X$ whose $k+1$ nearest landmarks are the $k+1$ vertices spanning that k-simplex. Akin to the Vietoris-Rips complex and the Čech complex, the **lazy witness complex** is the clique complex on the edge set of the strict witness complex.

One can create a filtration of such complexes by dimension or alternative criteria **a posteriori**. Yet this is in a sense different than constructing a nested sequence of complexes based on a data-set specific parameter ϵ. The witness complex we utilize and adapt aligns with the second definition provided by Carlsson and de Silva.

Def. A simplex is in the **nested family of witness complexes for** $\nu \in \mathbb{Z}_{\geq 0}$ $W(A; \epsilon, \nu)$ if there is an $x \in X$ such that the maximum distance between x the vertices of that simplex is less than ϵ more than the distance from x to its ν^{th} closest landmark.

In [71], we reintroduce the witness complex $W(A; \epsilon, 1)$ as a Dowker complex [12].

Def A **Dowker relation** between two sets Z and Y is a subset of the product space $R(Z,Y) \subset Z \times Y$.

Def A finite subset $\{z_0, \ldots, z_n\} \subset Z$ is a simplex σ in a **Dowker complex** if there is an element $y_i \in Y$ such that $\{z_0, \ldots, z_n\} \times y_i \subset R(Z,Y)$. This complex is homotopy equivalent to the complex constructed using z_j and $\{y_0, \ldots, y_k\} \subset Y$.

The **fuzzy witness complex** is then the Dowker complex the following **fuzzy witness relation** built using data from each $x \in X$ and simplices $\sigma \subset L$.

Def. The **fuzzy witness relation** is the subset of $X \times L$ given by

$$\mathcal{R}_W(X,L) = \{(x,\ell) | A(x,\ell) \leq A(x,L) + \epsilon\}. \tag{2.23}$$

This witness relation looks for shared witnesses near the boundaries of neighboring Voronoi cells centered at the landmarks. The geometric ϵ parameter increases the "fuzziness" of the Voronoi cell boundaries by extending the traditional straight-line boundary of the Voronoi cell by a hyperbola shifted a distance of $\frac{\epsilon}{2}$ from the boundary at its center, which lies on the line connecting two landmarks. When $\epsilon = 0$, the witnesses are partitioned by the Voronoi cells centered at each of the landmarks and thus only witness the landmark to whose Voronoi cell they belong. Unless some witnesses are exactly on the Voronoi cell boundary between multiple landmarks, there should be no higher dimensional simplices. As ϵ increases, the inflated Voronoi cells no longer have intersections of measure 0. The existence of witnesses in these non-convex regions of intersection determines the simplices included in the witness complex. See Fig. 2.4 for illustrations of the Čech complex and witness complex constructions.

Recall, our objective is to extract from this topological model a signature of the dynamics in a sliding window. Changes in this topological signature should indicate a regime-shift in a time series. Because we seek to implement this analysis in real-time, keeping computational costs cheap and runtime low is key to our methodology of TDA for time series analysis. Balancing correctness, consistency and cost forces us to develop a new perspective on the following questions.

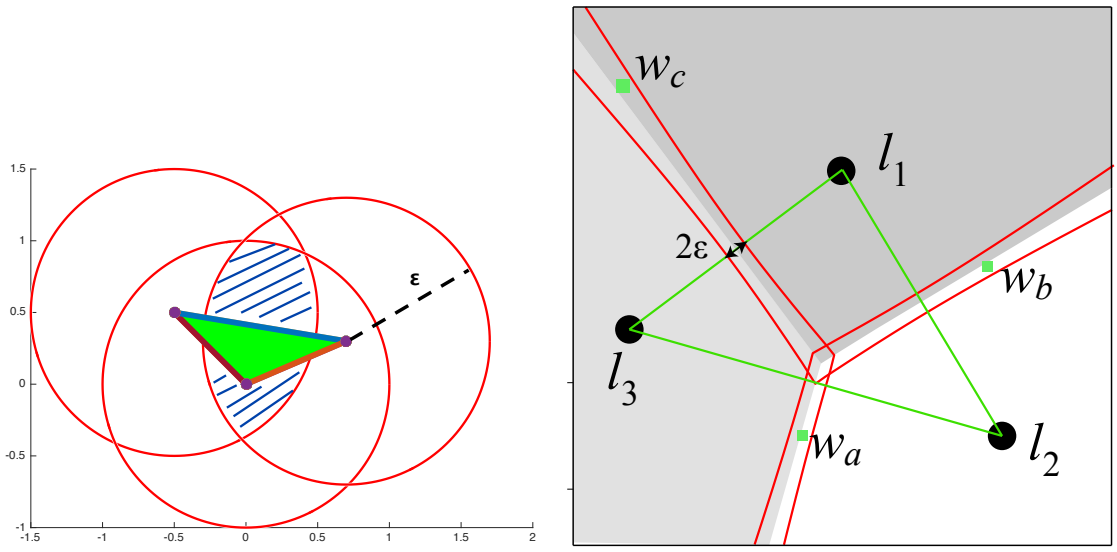


Figure 2.4: The (a) Čech and (b) witness complex require different types and amounts of computation for their construction.

How much time series data should be used for delay reconstruction? How many subsample points should be used in witness complex model of point cloud? At what scale-parameter do we terminate the filtration of witness complexes we are constructing? How do these choices affect the topological signature? We investigate these questions in Chapter 4.

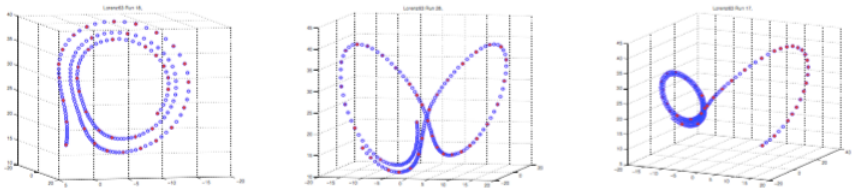
Our data sets not only lie in a metric space, they are also derived from a dynamical system. The point cloud we are modeling topologically is inherently not a stagnant object. Properties of the dynamics can inform us about the underlying topology represented by the point cloud that may be obscured by the geometry of the delay reconstruction. Assuming our discrete time series measurements are of an attractor of a C^2 continuous dynamical system, nearby points should remain nearby. Moreover, they should travel in roughly the same direction.

Are there better definitions for scale-based simplicial complexes to model delay reconstructed time series data? Can we utilize the extra information in the data structure itself, the time ordering of the points? We present the family of directionally

distorted witness complexes and statistics supporting their improvement over the standard witness complex for modeling two-dimensional delay reconstructions of attractors from chaotic dynamical systems data in Chapter 6.

2.5.3 Sliding Window Size

The input to our regime-shift detection pipeline begins as a stream of time series data. The first question we must address is how much data we would like to look at. A small amount of data will likely not cover the attractor of the dynamical system and lead to an incorrect topological signature. A large amount of data might well cover the attractor but may be too expensive to store or analyze.



(A) circle (B) butterfly (C) leash

Figure 2.5: Three 200 point trajectories of the Lorenz system show great variability in topology based on initial conditions.

The invariant measure addresses the idea of average coverage of an attractor over time regardless of initial conditions. This suggests that there is some length of time at which the entire attractor will achieve some minimal density of coverage, even if parts are more densely covered than others. Until observing this amount of data, the geometric variation of specific sampled trajectories based on initial condition can lead to an inconsistent topological signature for the dynamical system. Figure 2.5 demonstrates a few of the of topologies that can occur even when increasing the amount of data.

Figure 2.7 shows how this variability manifests in the delay reconstructions of short segments

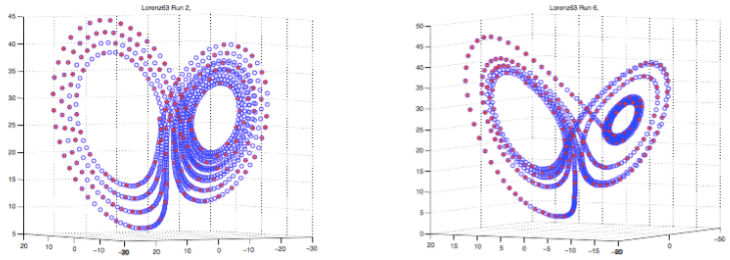


Figure 2.6: Two 1000 point trajectories of the Lorenz system reveal geometric variation of trajectories based on initial conditions.

of the time series acquired from projecting the Lorenz system of equations onto the x-coordinate.

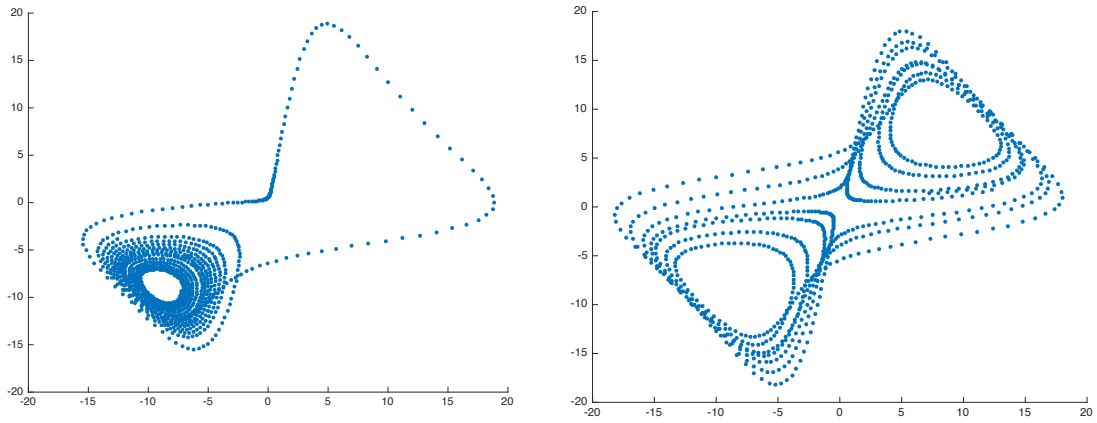


Figure 2.7: Two 1000 point delay reconstructions of the x-projection of solutions to the Lorenz system with distinct initial conditions reveal geometric variation in the topologies.

We hypothesize that when the invariant measure does not apply, not only will there be inconsistency within a data set of topological representation, there will also be extra homology in the point clouds where lower measure regions of the attractor remain unsampled. Some regions of the attractor are less dense because the flow is faster there. In these cases, the one-dimensional trajectory structure can be emphasized with a higher sampling rate. Lack of sampling due to not observing the system for a long enough time-scale produces non-uniform density traverse to the flow that form holes in the wings of the Lorenz attractor or along the bridge of the Rossler attractor.

Therefore, more data will lead to less erroneous holes of this kind. For an insufficient **sliding window size**, we anticipate the geometric differences between the radii of the boundary holes in the attractor preventing us from distinguishing these erroneous holes from the meaningful ones representing a boundary. These features are reflected in Figure 2.6. One property of the boundary holes is that each point has at least one other point on the hole whose tangent vector is opposing its own. This differs from the tangent vectors along an erroneous hole. This is the feature we aim to exploit with the metric we introduce in Chapter 6 to improve the topological signature using a smaller **sliding window size** and **number of landmarks**.

2.5.4 Landmark Selection

In order to expedite computations we use the witness complex, which uses a subsample of the data to extract the shape of the point cloud. This subsample can be determined in a number of ways. Previously implemented methods include selecting the landmarks randomly from the witnesses or using a MaxMin algorithm to evenly space the landmarks amongst the witnesses. According to such an algorithm, an initial landmark is selected at random from the point cloud X. From that, a set L of landmarks is increased iteratively with the k^{th} landmark to be added to the previously selected $\{\ell_1, \ldots, \ell_{k-1}\}$ chosen to be $w \in X$ such that w achieves

$$\ell_k = \text{argmax}_{w \in X}\{\min_{\ell_i \in L}(d(w, \ell_i))\}. \tag{2.24}$$

Alternatively, using the temporal ordering of the data points, we can landmarks that are evenly spaced in time (EST). Given a point cloud $\{X_t\}_{t=1}^{\textbf{number of witnesses}}$ corresponding to an m-dimensional delay reconstruction of the sliding window $\{x_t\}_{t=1}^{\textbf{sliding window size}}$ with delay τ, we take landmarks according to a down sampling rate of η witnesses per landmark so that the k^{th} landmark in the landmark set L is

$$\ell_k = x_{k \cdot \eta}. \tag{2.25}$$

Such a definition should space the landmarks in the point cloud according to the invariant measure. It also avoids having to recalculate the landmark set when buffering, or adding data to a sliding window, before proceeding to slide the window. We choose to use the MaxMin landmark selection method in this research because it lends itself to estimates that depend only on the dimension of the data rather than the local density.

There also remains the big question of how many data points should be taken as landmarks. Recall the **number of landmarks** determines the number of vertices used in the witness complex. We want a cheap and correct method of classifying delay reconstructed time series data from computable topological representations. The **number of landmarks** immediately limits the number of simplices that can be created in the filtration from which the persistent homology is directly computed. It also limits the types of detectable topology. Many research endeavors involve estimating convergences in the limiting cases of infinite data and infinite time [50]. In contrast, our objective is to approach this problem from the bottom up - starting with less to see more, finding lemons to make lemonade.

2.5.5 Metrics on Persistent Homology

The n^{th} dimensional persistent homology is equivalent to the set of **persistence intervals**,

$$PI_n = \{[\epsilon_{i_B}, \epsilon_{i_D}]\}_{i=1}^{\beta_n} \tag{2.26}$$

with the i^{th} interval corresponding to the n-dimensional hole that is born at scale parameter ϵ_{i_B} and dies at scale parameter ϵ_{i_D} and β_n is the total number of n-dimensional holes to ever exist in the filtration. Depending on how persistent homology is interpreted, it can take on a variety of metrics. Persistent homology in a given dimension can be communicated through a persistence diagram where each persistence interval $[\epsilon_{i_B}, \epsilon_{i_D}]$ corresponds to a point $(\epsilon_{i_B}, \epsilon_{i_D}) \in \mathbb{R}^{2+}_{x \leq y}$. This presentation is called a **persistence diagram** and is shown in Figure 2.8(a).

Some of the first metrics introduced on the space of persistence diagrams such as the p-

Wasserstein metrics are matching-based, computationally expensive and challenging for statistical analysis [36]. We instead choose to translate the information in the persistence diagram into a cumulative functional representation of the persistent homology called the **persistent homology rank function** (PRF) introduced in [67].

$$f : \mathbb{R}^{2+}_{x\leq y} \to \mathbb{R}^{+}, \tag{2.27}$$

$$(a, b) \mapsto |\{[\epsilon_{i_B}, \epsilon_{i_D}] \in PI_n \; s.t. \; [a, b] \subseteq [\epsilon_{i_B}, \epsilon_{i_D}]\}| \tag{2.28}$$

A PRF assigns to each point $(a, b) \in \mathbb{R}^{2+}_{x\leq y}$ the number of points in the persistence diagram with $\epsilon_{i_B} \leq a$ and $\epsilon_{i_D} \geq b$. Fig. 2.5(b)) shows the corresponding PRF to Fig. 2.5(a). The colors of the points in the persistence diagram correspond to the multiplicity of homological features with that lifespan, as seen in the color bar. For instance, the yellow point at approximately $(1e^{-2}, 2e^{-2})$ in Fig. 2.5(a) represents five or more features with birth ϵ value of $1e^{-2}$ and death ϵ value of $2e^{-2}$. These features are represented in the corresponding PRF in Fig. 2.5(b) by the light green square in the lower left corner of the PRF. The value of this square is the number of homological features born before $1e^{-2}$ that die after $2e^{-2}$, which includes homological features represented by the purple point and turquoise triangle in the left side of the persistence diagram. The triangle is used to distinguish "immortal cycles", or cycles that do not terminate at the end of the filtration. The PRF shown in Fig. 2.5(b) has constant value on grid squares because of our discretization which we discuss in Chapter 4.

The span of a set of PRFs forms a subset of a Hilbert space where the distance between two PRFs f, g is

$$d_\phi(f, g)^2 = \int_0^{\epsilon_{\max}} \int_0^{y} [f(x, y) - g(x, y)]^2 \phi(x, y) dx dy, \tag{2.29}$$

the inner product between two PRFs f, g is given by

$$\langle f, g \rangle = \int_0^{\epsilon_{\max}} \int_0^y f \cdot g\phi(x, y)dxdy \tag{2.30}$$

and the norm of a PRF f is

$$||f||_{\phi L^2} = \int_0^{\epsilon_{\max}} \int_0^y f^2\phi(x, y)dxdy. \tag{2.31}$$

Here $\phi : \mathbb{R}^{2+}_y \to \mathbb{R}$ is L^1. The original motivation was to set the weight function as an exponentially decreasing function of the lifepsan to make finite the infinite distances between persistence diagrams that have "immortal cycles", cycles that do not terminate at the end of the filtration and are taken to have infinite lifespan, with slightly different birth times. We adopt the perspective that immortality is improbable and set the death time of immortal features as one more than the end of the filtration. We discuss this matter of terminating filtrations in Chapter 4. Th weighted L^2 metric on the span of a set of PRFs has the advantage that we can compute a unique point-wise mean of a set of PRFs $\{f_n\}_{n=1}^s$ with sample size s by

$$\overline{\beta}(x, y) = \frac{1}{s} \cdot \sum_{n=1}^{s} f_n(x, y). \tag{2.32}$$

We can then use this as a representation of the persistent homology of that set. From this, statistics to characterize the consistency and correctness of the topological signature can be computed. The inner product structure can also be used in further analysis, like classification via functional principal component analysis [67]. We utilize weighted L^2 metrics on PRFs as a part of both our topological classifier Chapter 4 and our heuristics for establishing lower bounds on the number of witnesses and landmarks to use for modeling dynamical systems from delay reconstructed sliding windows in Chapter 5.

An alternative way to represent persistent homology is as a family of functions called **persistence landscapes** that trace out the level sets of the PRF. This idea was introduced prior to

PRFs in [8] and relies on the monotonicity of the PRF under a change of coordinates from (**birth, death**) to (**midlife time, lifespan**). The L_∞ norm on the space of persistence landscapes is stable under perturbations of the data set.

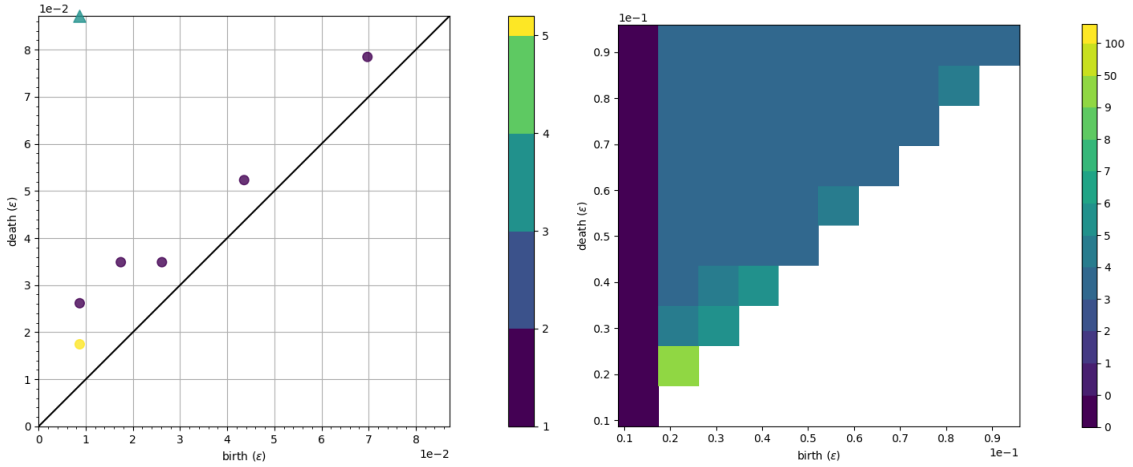


Figure 2.8: A persistence diagram and the corresponding persistent homology rank function are two ways to represent the persistent homology of a filtration.

Having these well-defined, stable and interpretable statistics is important for analyses aimed at classification, where means often serve as a class representative and variances serve as a way to threshold membership. Though common in the machine learning community, this technique has not yet been applied to topological statistics. We build upon the definitions in [67], introducing the further statistical machinery necessary and defining this topological binary classifier in Chapter 4.

2.6 Statistics for Validation Procedures

2.6.1 Receiver Operating Characteristic Curves

In Chapter 4 we introduce a topological membership test as a step towards regime shift detection from time series using TDA. For validation of the efficacy of our topological membership test, we use area under the curve (AUC) of receiver operating characteristic (ROC) curves [26].

A **Receiver Operating Charateristic** (ROC) curve plots the true positive rate (TRP), called **sensitivity**, against the false positive rate (FPR), called **specificity**, as a parameterized

curve for binary classifications where the classifier is parameterized by $z \in \mathbb{R}^+$. A good classifier has TPR growing significantly faster than FPR. Therefore, a successful classifier would have an ROC curve that slowly traces the left and top perimeter of the TPR $\times$ FPR domain as z is increased; a bad classifier would have an ROC curve that is near or below the diagonal. The information presented by the ROC curve can be summarized by the **area under the curve** (AUC). A higher AUC indicates a better classifier. A near perfect classifier has $AUC \approx 1$, a classifier that is essentially a coin flip has $AUC \approx 0.5$, and a classifier that misclassifies more than it correctly classifies has $AUC < 0.5$.

The statistical analysis we perform in Chapter 4 reveals a wide discrepancy in the stability and success of our topological binary classifier over the parameter space of the witness complex construction. We use weighted L^2 norms on PRFs to experimentally investigate how the topology detected by a filtration of witness complexes is altered by changing the number of witnesses and number of landmarks. This experimentation informs a heuristic we introduce in Chapter 5 to establish the size of the model.

2.6.2 Hierarchical k-Clustering

For validation of the improvements on persistent homology achieved by the new witness relation for dynamical systems data we introduce in Chapter 6, we present a set of novel statistics based on the hierarchical 2-clustering of the persistence intervals as situated in the persistence diagram. We review the general method of hierarchical clustering now.

Hierarchical 2-clustering is a technique that begins with each data point in a unique cluster and iteratively assigns the data points to fewer and fewer clusters, ultimately terminating in two clusters of data points that produces the maximal distance between any possible two clusters. We use a Euclidean-based single-linkage heirarchical clustering algorithm in the scipy package of python. In general, hierarchical k-clustering looks for the minimal Hausdorff distance between all of the clusters and combine the clusters that achieve this minimal distance, repeating to successively reduce the number of clusters until there are k clusters. This is akin to reporting the last k

components when computing the β_0 persistent homology. In our case, we are considering the set of persistence intervals as data points in $\mathbb{R}^2$ with the Euclidean metric and clustering the persistent homology itself until we only have two clusters. The separation of persistence intervals in the persistence diagram in this way can serve to communicate what of the persistent homology is signal and what is noise. We take the number of points in the cluster containing the longest living homological feature as the "signal" cluster representing the number of global one-dimensional homological features, and the rest of the persistent homology as "noise" representing the local homology that is short-lived.

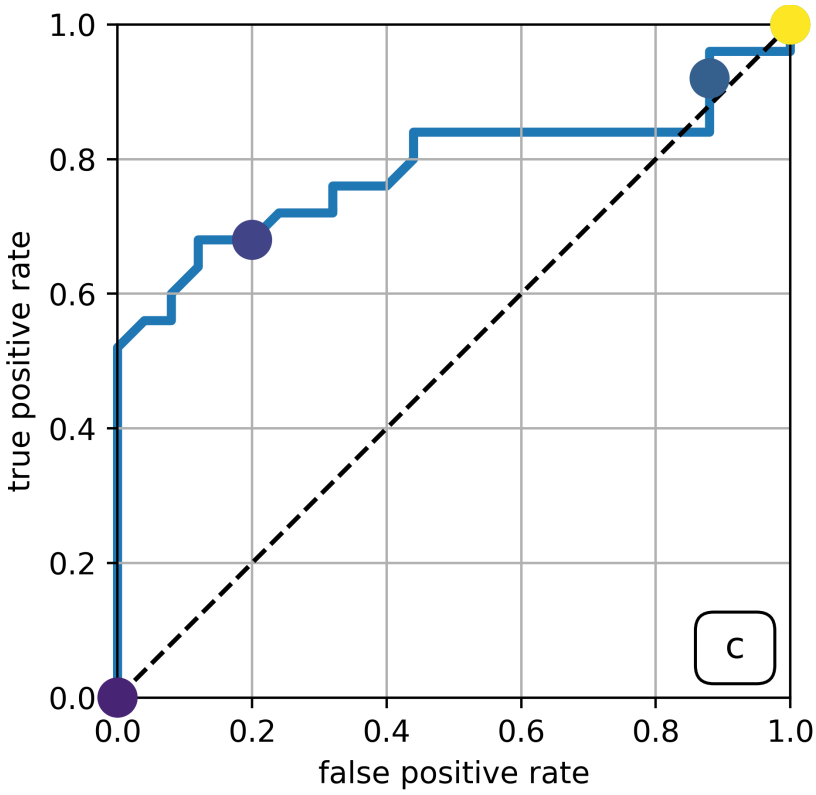


Figure 2.9: A Receiver Operating Characteristic curve reports the true positive rate and false positive rate of a classifier as a function of the classifier threshold value.

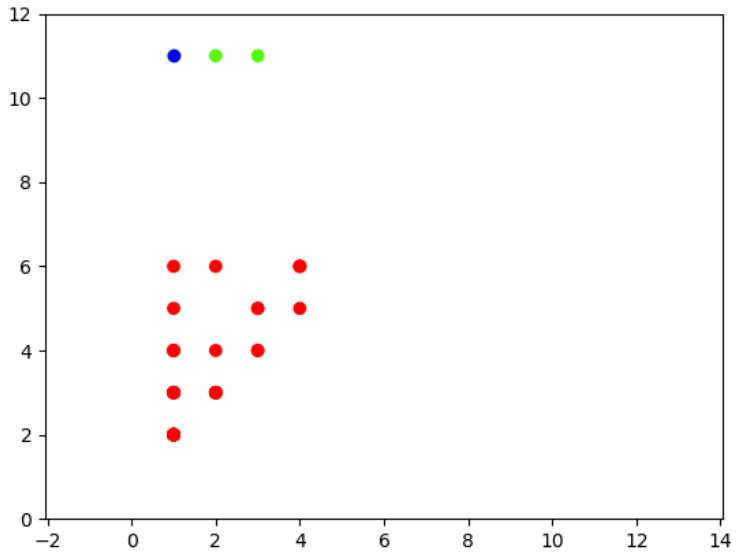


Figure 2.10: Hierarchical 2-clustering of the points in the persistence diagram separates the "signal" and "noise" in the persistent homology. The"signal" is represented by the green points in the persistence diagram and also contains the maximal lifespan feature denoted in blue. The "noise" is colored red.

Chapter 3

Datasets

3.1 Synthetic Data Sets

We establish our methodology on synthetic data generated as numerical solutions to systems of equations modeling real-world chaotic phenomena.

3.1.1 2D Delay Reconstructed Lorenz Attractor

We generate 1,000,000 data point time series from the Lorenz system in Equation 2.1 with time step 0.01 with initial condition (1,1,1) for 10,000 time units by first numerically integrating the ODEs in the full three dimensional state space using the ode45 integrator in MATLAB. We project this numerical three dimensional solution onto the x-coordinate to obtain a time series of the chaotic Lorenz data set.

Following the philosophy introduced in [23], we delay reconstruct these time series in two dimensions using the first minimum of mutual information, $\tau = 18$ time steps, as our choice of delay. This is shown in Figure 3.3.

As shown in Figure 3.3(a), for delay less than the first minimum of time-delayed mutual information, the reach of the attractor is small. Figure 3.3(b) shows a reconstruction using delay parameter equal to the first minimum of time-delayed mutual information. Now the holes in the Lorenz butterfly wings are round and the attractor is well-extended off the diagonal. As the delay increases beyond the first minimum of time-delayed mutual information, the attractor folds over itself. Figure 3.3(c) shows this projection eradicating the two main one-dimensional homological

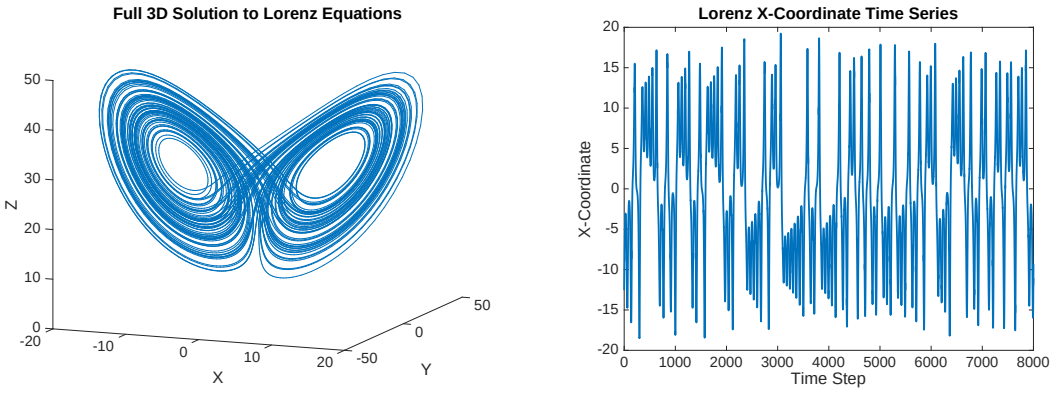


Figure 3.1: (a) A numerical 3D solution to the Lorenz system of equations approaches the strange attractor. (b) The projection of the numerical 3D solution to the Lorenz system of equations onto the x-coordinate produces a time series.

features. For delay much greater than the first minimum of time-delayed mutual information, four lobes with no one-dimensional homology emerge. This is shown in Figure 3.3(d).

3.1.2 2D Delay Reconstructed Rössler Attractor

We generate 1,000,000 data point time series from the Rössler system Equation 2.2 with time step 0.08, initial condition (1,1,1) and 80,000 time units by first numerically integrating the ODE in the full three dimensional state space using the ode45 integrator in MATLAB. We use this larger time step so that the first minimum of average time-delayed mutual information is the same for both the Lorenz and Rössler time series. We project this full 3D solution onto the x-coordinate to obtain a time series of the chaotic Rössler data set, as shown in Figure 3.4.

We delay reconstruct these time series in two dimensions using the first minimum of mutual information, $\tau = 18$ time steps, as our choice of delay. This is shown in Figure 3.6.

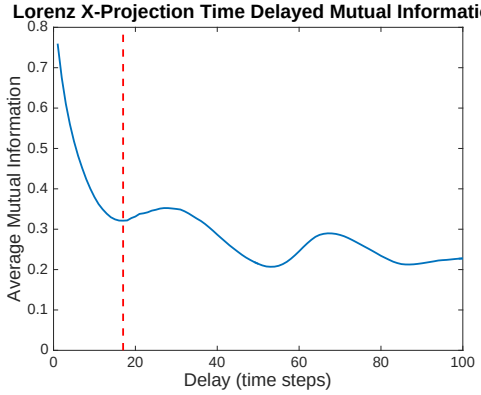


Figure 3.2: The time-delayed mutual information of the x-projection of a solution to the Lorenz system of equations is plotted as a function of the delay. The first minimum denoted with a dashed red line identifies a delay parameter for Takens delay reconstruction that results in maximally independent coordinates.

Figure 3.6(a) shows that for delays less than the first minimum of time-delayed mutual information, the attractor is still close to the diagonal which leads to small reach and short-lived persistent homology. Figure 3.6(b) shows a reconstruction using delay parameter equal to the first minimum of time-delayed mutual information. Here the hole in the center of the Rössler attractor appears circular in the delay reconstruction. As the delay increases beyond the first minimum of time-delayed mutual information, the attractor fold over itself. Figure 3.6(c) shows how this projection contains no one-dimensional homological features. As the delay is further increased past the first minimum of time-delayed mutual information, the reconstructed Rössler attractor cycles through these representations. Figure 3.6(d) shows how the attractor can re-unfold; however, the projection can distort the homology we are seeking to detect.

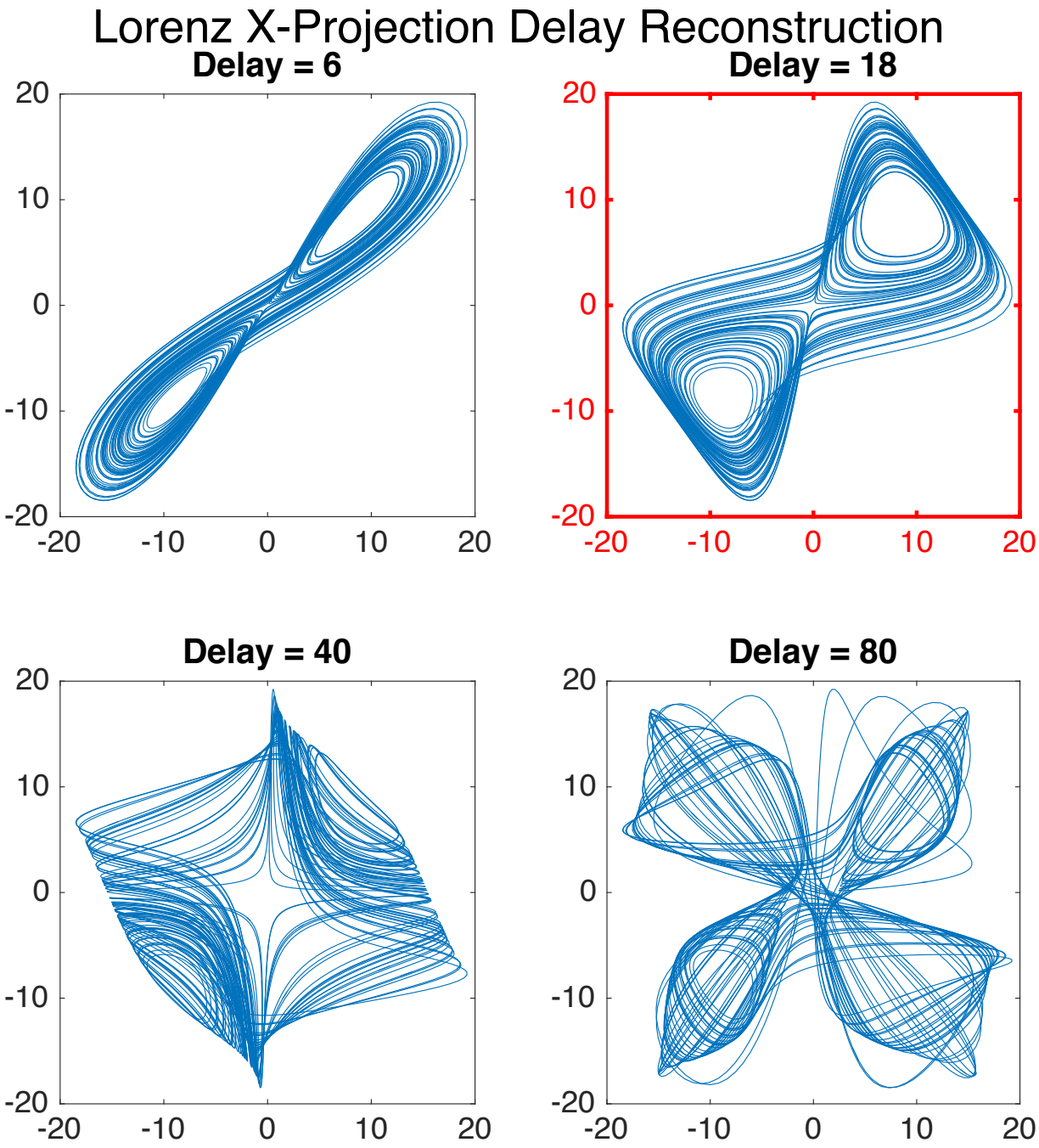


Figure 3.3: Two-dimensional delay reconstructions of the x-projection of a solution to the Lorenz system of equations using a variety of delay parameters reveal a variety of topological and geometric representations of the Lorenz attractor.

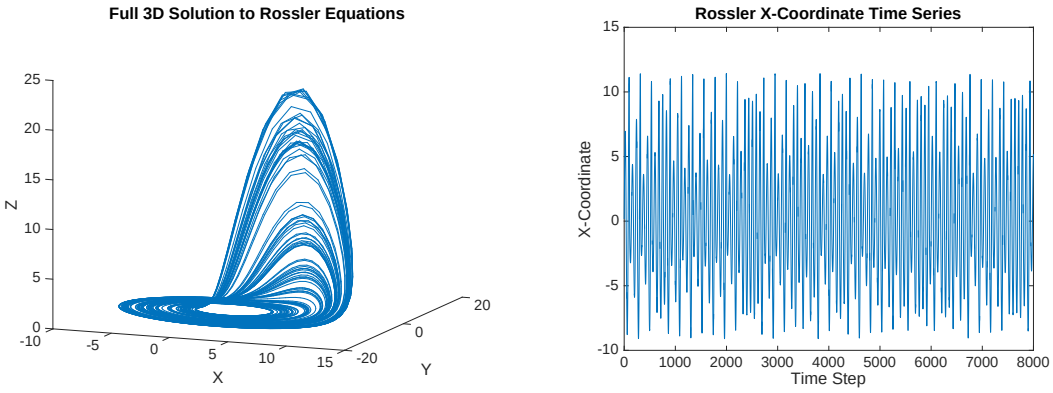


Figure 3.4: (a) A full 3D solution to the Rössler system of equations approaches the strange attractor. (b) The projection of the full 3D solution to the Rössler system of equations onto the x-coordinate produces a time series.

3.1.3 2D Delay Reconstructed Signal of Brunel Neuronal Network

The Brunel model for neuron firing introduced in [7] is a leaky integrate-and-fire model on a sparsely connected network whose vertices represent excitatory and inhibitory neurons. The parameters describing the behavior of the excitatory and inhibitory neurons are allowed to differ from each other. When the model parameters are set as discussed in [28], this results in four dynamically distinct regimes of global and individual neuron behavior emerge over a two-dimensional parameter space of the model.

The first parameter we denote by I. It stands for the ratio of the frequency of the external input signal to the frequency of neurons reaching the firing threshold in the absence of an external signal. The second parameter we denote by G and it represents the relative strength of inhibitory synapses. When $I = 2, G = 3$, the neurons fire regularly with high frequency and are almost fully

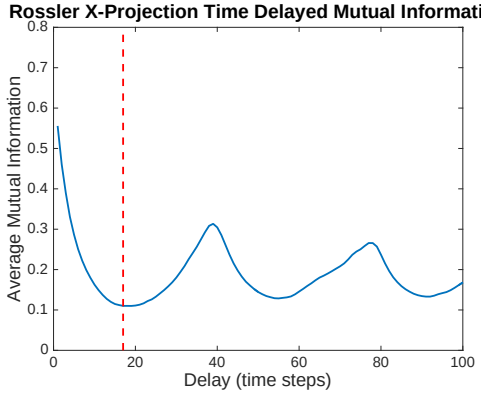


Figure 3.5: The time-delayed mutual information of the x-projection of a solution to the Rössler system of equations is plotted as a function of the delay. The first minimum denoted with a dashed red line identifies a delay parameter for Takens delay reconstruction that results in maximally independent coordinates.

synchronized; this is called the synchronous regular regime (SR). When $I = 2, G = 5$, the neurons fire irregularly and global activity remains stationary; this is called the asynchronous irregular regime (AI). When$I = 4, G = 6$, the neurons fire irregularly and produce fast oscillations in the global activity; the neuron firing rate is lower than global frequency. When $I = 1, G = 4.5$, the neurons fire irregularly at very low rates and produce slow oscillations in the global activity. These last two regimes are called the fast and slow synchronous irregular (SI) regimes, respectively. Figure 3.7 shows the average excitatory neuron voltage in each of these regimes.

We can get a single measurement of the system by taking an average of the excitatory neuron voltages at a given time.

We plot the mutual information in Figure 3.8 of this signal in each of the four regimes. Because there is only a clear minimum of mutual information for the $I = 4$, $G = 6$ regime data

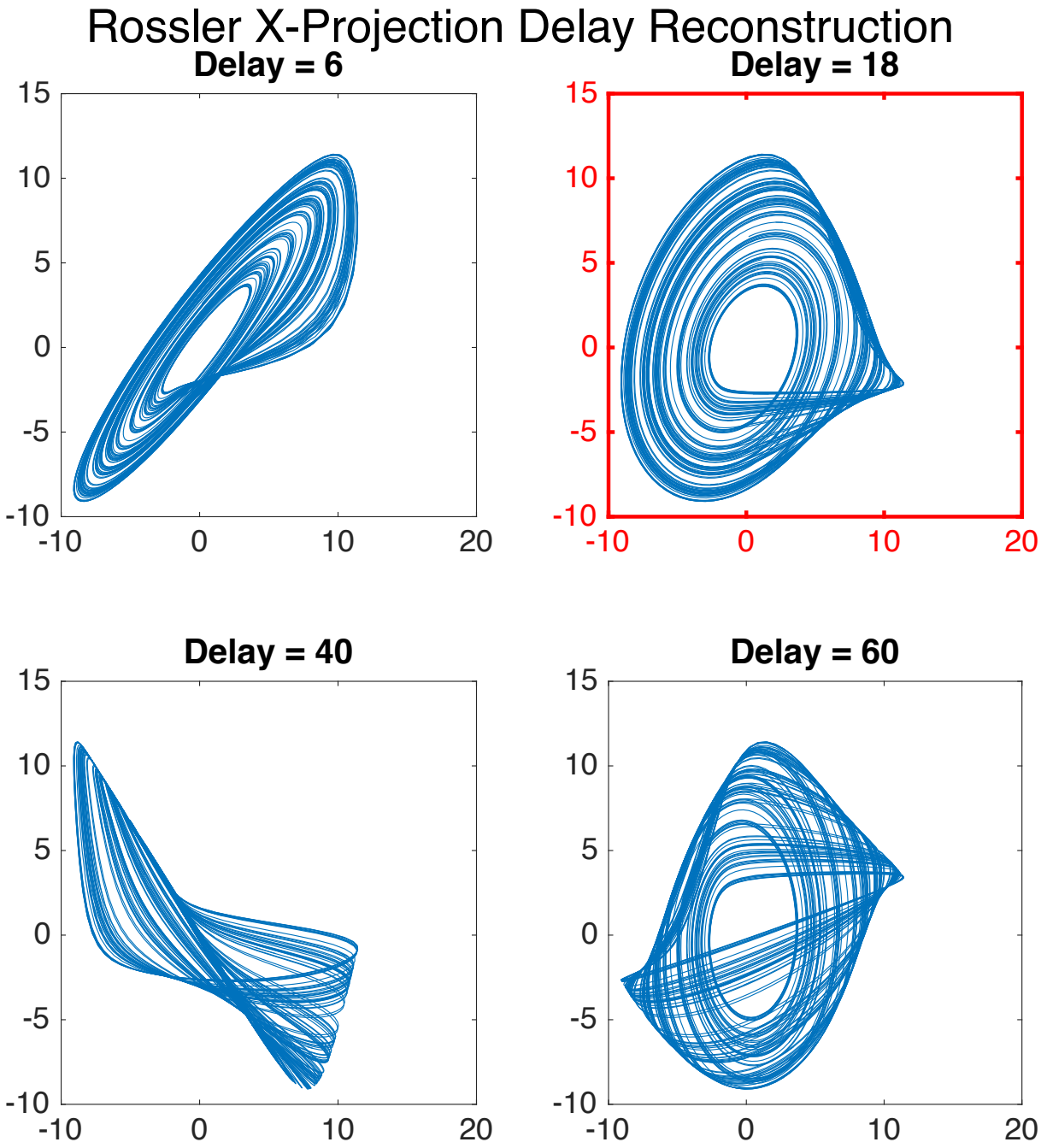


Figure 3.6: Two-dimensional delay reconstructions of the x-projection of a solution to the Rössler system of equations using a variety of delay parameters reveal a variety of topological and geometric representations of the Rössler attractor.

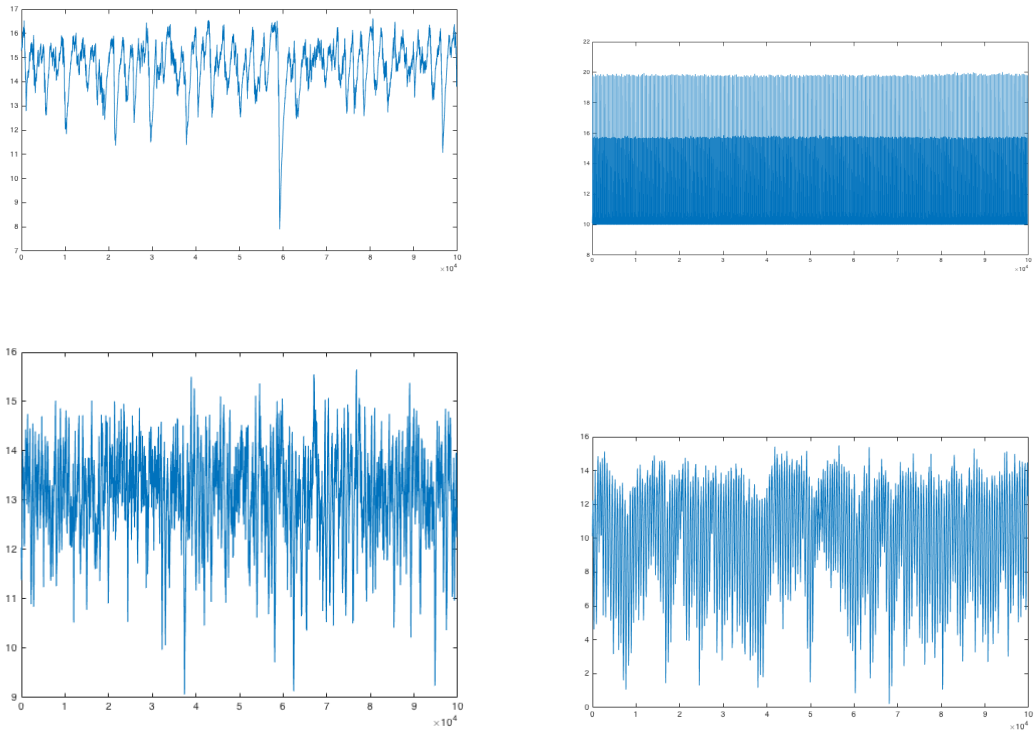


Figure 3.7: Time series of the average excitatory neuron voltage of the sparsely connected Brunel neuronal network in four dynamically distinct regimes: (a) $I = 1,\ G = 4.5$, (b) $I = 2,\ G = 3$, (c) $I = 2,\ G = 5$, (d) $I = 4,\ G = 6$.

set, seen in Figure 3.8 (d) at $\tau \approx 200$, we choose to delay reconstruct all data sets with $\tau \approx 200$.

Visual inspection of the data in Figure 3.9 shows that all regimes are well-extended from the diagonal in the two-dimensional delay reconstruction when using a delay of 200 time steps. In Figure 3.9, the color gradient represents time. The shifting, as well as expansion and contraction, of the triangular cycles in the delay reconstruction is a sign of non-stationarity of this signal.

3.2 Experimental Data

3.2.1 2D Delay Reconstructed Music Instrument Recordings

Voltage recordings of the note A440 played on a viol and clarinet using a Sony ICD-PX312 digital voice recorder sampling at 44.1 kHz provide real world, near-periodic data sets we seek to classify using TDA on sliding windows from the time series. These sliding windows are chosen to be

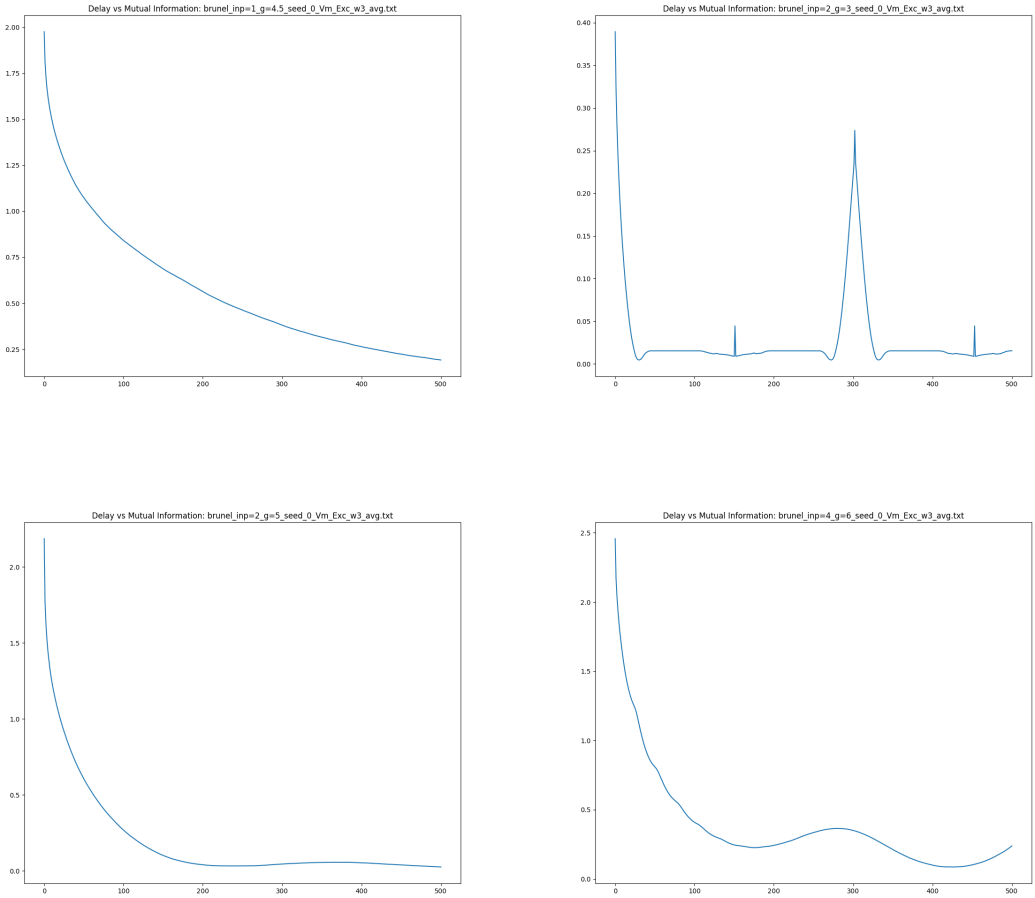


Figure 3.8: Mutual information between the average excitatory neuron voltage of the sparsely connected Brunel neuronal network and a delayed version of itself, as a function of the delay, in four dynamically distinct regimes: (a) $I = 1$, $G = 4.5$, (b)$I = 2$, $G = 3$, (c) $I = 2$, $G = 5$, (d) $I = 4$, $G = 6$.

evenly spaced across the recording. We set the delay to be $\frac{\pi}{\mathfrak{f}}$ where $\mathfrak{f}$ is the fundamental frequency of the note we are recording in units Hertz. This is irrationally proportional to the frequency $\mathfrak{f}$ of the musical note we are embedding, satisfying the genericity conditions of Takens theorem, in particular avoiding twice the period of any subharmonic.

To explore one-dimensional persistent homology, we chose to delay reconstruct in two dimensions. We will likely not have an embedding due to projection causing false crossings. These can create one-dimensional homological features in the filtration of witness complexes modeling the

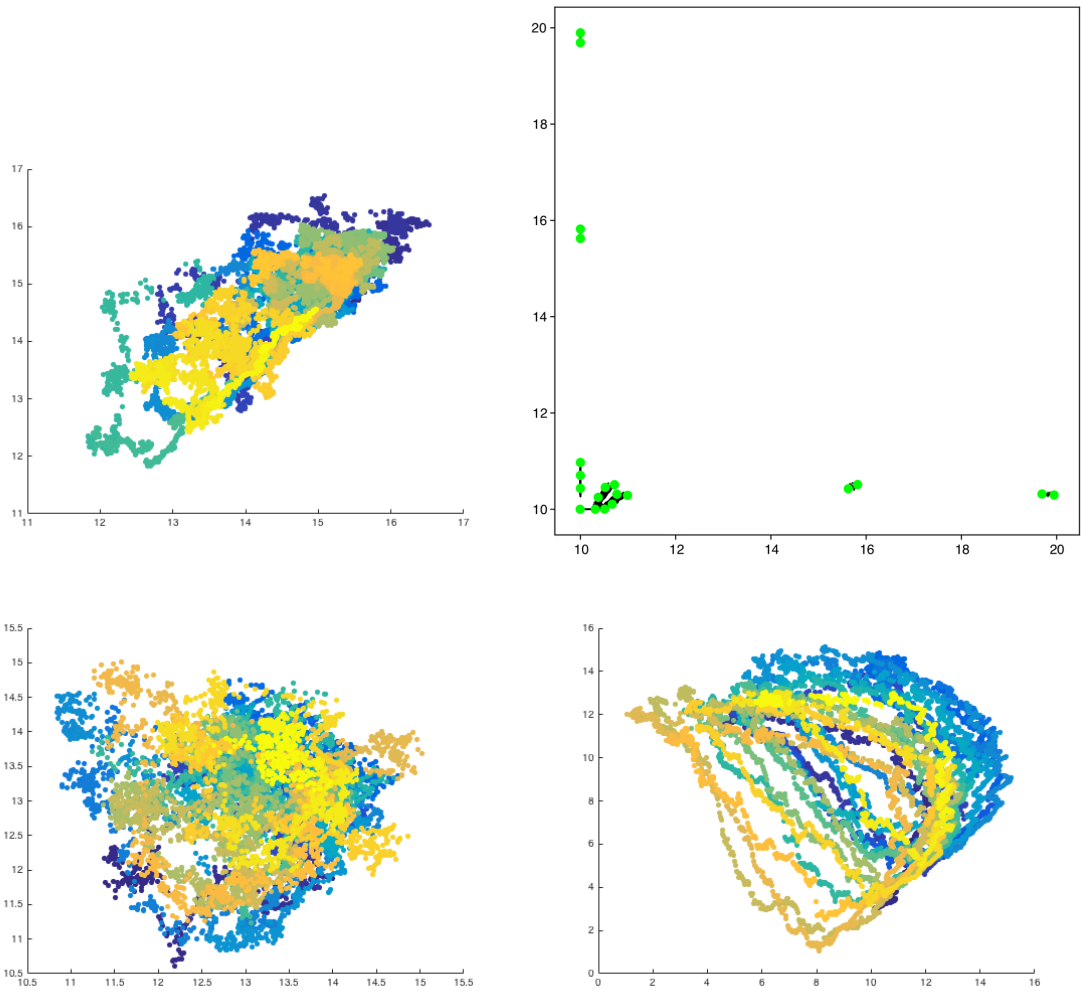


Figure 3.9: Two-dimensional delay reconstructions of 10,000 points the average excitatory neuron voltage signal from the sparsely connected Brunel neuronal network using a delay of 200 time steps in four dynamically distinct regimes: (a) $I = 1$, $G = 4.5$, (b) $I = 2$, $G = 3$, (c) $I = 2$, $G = 5$, (d) $I = 4$, $G = 6$. The color gradient represent the temporal ordering of the data points.

data. We believe that the interaction between the fundamental tone, subharmonics and additional frequency content of a musical instrument leads to these interesting geometric structures that can be used to distinguish between musical instrument recordings, even though the delay reconstruction is not an embedding [23]. Figure 3.10 shows the viol and clarinet recordings, highlighting in red the sliding windows used for the delay reconstructions depicted below.

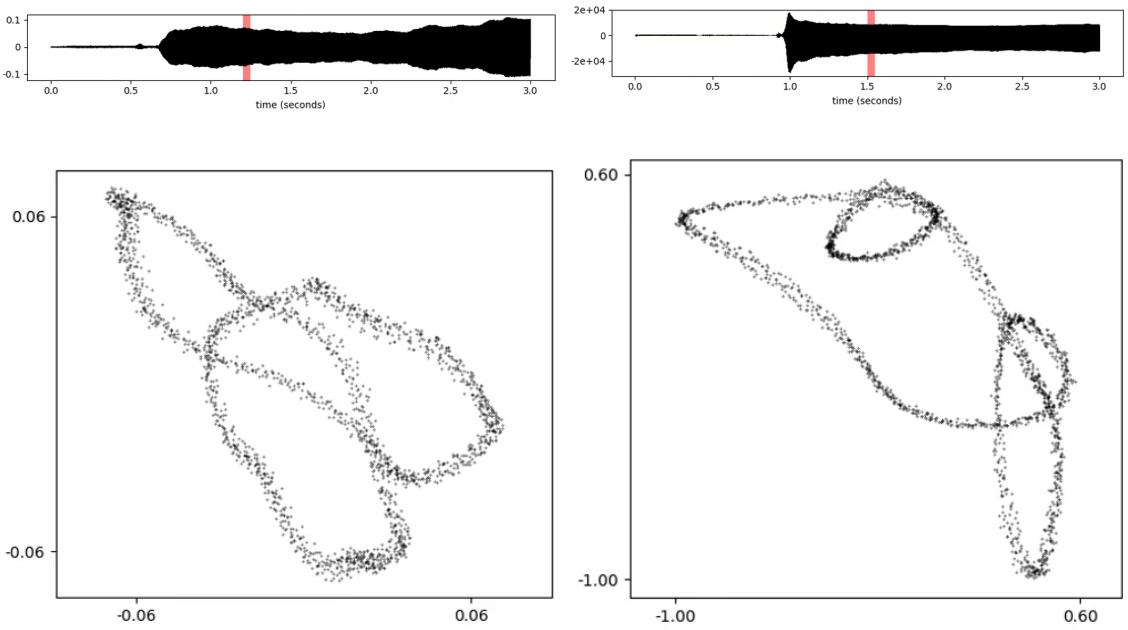


Figure 3.10: A live recording of a musical instrument, such as the viol (left) and clarinet (right), can be segmented into sliding windows and then delay reconstructed using $\tau = \frac{\pi}{\mathfrak{f}}$ where $\mathfrak{f} \approx 440Hz$ to produce a two-dimensional point cloud for topological data analysis.

Chapter 4

Topological Membership Testing

Our approach to time series analysis is to delay reconstruct a window of the time series data and compute the L^2 norm of the PRF as a topological signature for that window of data. Changes in this topological signature during a sliding window analysis can then be used to detect change-points in the time series when the underlying dynamics being measured shifts. To this end, we introduce a topological binary classifier on sliding windows of time series data. This reflects regime-shift detection in time series analysis where the binary classification between "Member of Current Regime" and "Not Member of Current Regime" is being repeatedly asked [56]. In particular, we are not assigning each sliding window sample to a single class that best represents it; e.g. a sample can be assigned as a member of both the Lorenz class and Rössler class. To make fair comparisons between PRFs from distinct samples, we first introduce normalizations of the filtration of witness complexes and the PRF itself. We then show that the unweighted L^2 distance between a sample PRF and a mean PRF representing a class introduced in [45] can serve as a good topological signature to test class membership **for some subset of the parameters involved in constructing the witness complex.** We further explore affects of parameter selection on the topology of the filtration of witness complexes in Chapter 5.

4.1 Terminating the Filtration

For filtrations modeling a finite data set based on a scale parameter ϵ, as with the fuzzy witness relation in Equation 2.23, it could make sense to stop computations either at an $\epsilon_{\max}$ value when

the filtration has stabilized as a clique on all landmarks or at an ϵ when the number of simplices in the filtration becomes too computationally expensive to construct and store. For example, for the fuzzy witness complex construction a choice of $\epsilon_{\max}$ greater than the diameter of the data set will always result in the complete complex on the set of landmarks. The persistent homology reported by such a filtration can obscure the importance of perhaps interesting and meaningful topology of the data. This can be the case when the portion of the filtration when non-trivial homology exists is small in comparison to the duration of the filtration. Exploring ϵ greater than the theoretically motivated $\frac{\tau}{2}$ [50] can provide information about the scale of the topological features in the data. Yet increasing ϵ increases the number of simplices in the model. This motivates us to introduce a method of terminating filtrations of simplicial complexes based on a topological property limiting the size of the model.

Heuristic 1. *Terminate a filtration of fuzzy witness complexes at ϵ_{max} when a simplex of dimension k_{max} enters the filtration.*

For witness complexes, this is equivalent to when there exists a witness that witnesses $k_{\max}+1$ landmarks. Consider an m-dimensional data set $\Gamma = \{w_t\}_{t=1}^{N}$, here considered as a delay reconstructed time series. Let $L = \{\ell_i\}_{i=1}^{M}$ where $L \subseteq \Gamma$ is the subset of the data set chosen as landmarks. Concretely, we set

Heuristic 2. $\epsilon_{max} = \min_{w\in\Gamma}\{d\left(w, L\backslash\{l_{w(0)}, \dots l_{w(k_{max})}\}\right) - d(w, L)\}$

where $\ell_{w(j)}$ is the j^{th} closest landmark to w. This allows a way to directly limit the computational complexity.

Suppose L is chosen according to a MaxMin algorithm with density ρ approximated by $\rho \approx \delta^m$ where $\delta = \min_{i\neq j} d(\ell_i, \ell_j)$. Let $\Delta \approx \max_{i\neq j} d(\ell_i, \ell_j)$ approximate the diameter of the data set. The MaxMin algorithm aims to evenly space the landmarks within the state space amongst the data. For some range of $\epsilon_{\max}$, we expect the number of landmarks within a ball of radius $\epsilon_{\max}$ to scale according to

Estimate 1. $k_{max} \approx \rho \cdot \epsilon_{max}^{m}$.

This range is when the homology created in the witness complex corresponds to the fleshing out of the local topology rather than the revelation of the global topology of the point cloud.

Some local topological features include the star of a vertex, or **set of simplices that contain a vertex** [11]. The stars of nearby vertices can connect along the links, or the faces of the simplices in the star that do not intersect the vertex. This can create geometrically small, and often short lived, persistent homology in the filtration. For example, two vertices that have edges to the same set of vertices but not to each other will create a plethora of one-dimensional homology. As ϵ increases, the stars of vertices expand and it is likely that these two vertices with similar neighbors who are not themselves neighbors will become neighbors. Because this phenomenon occurs across the point cloud approximation of the attractor, a large number of short lived features enter the filtration in a short amount of time. A high PRF value near the diagonal indicates that the witness complexes are capturing these local topological features across the point cloud.

Figure 4.1(a) shows an example of a witness complex capturing this local topology. Motivated by computational efficiency, robustness and interpretability, we instead seek a coarse-grained topological signature. If a filtration of witness complexes were to terminate when the local homology was only beginning to be captured, the corresponding persistent homology would not identify the geometrically large-scale homological features that have been formed among the more fine-grained features.

To this end, [1] suggests picking $\epsilon_{\max} = r \cdot \Delta$ with constant of proportionality $r \in [0, 1]$, yet choosing this r largely remains an art. In particular, computations can become prohibitively expensive with high density data. Alternatively, our heuristic terminates the filtration in terms of an upper bound of the size of the model, which limits the computations directly.

Heuristic 3. *Select k_{max} explicitly, allowing ϵ_{max} to vary with each data set.*

This corresponds to

Heuristic 4. $k_{max} \approx \rho \cdot (r \cdot \Delta)^m$.

A discrepancy between the ambient dimension m in which the time series data is delay reconstructed

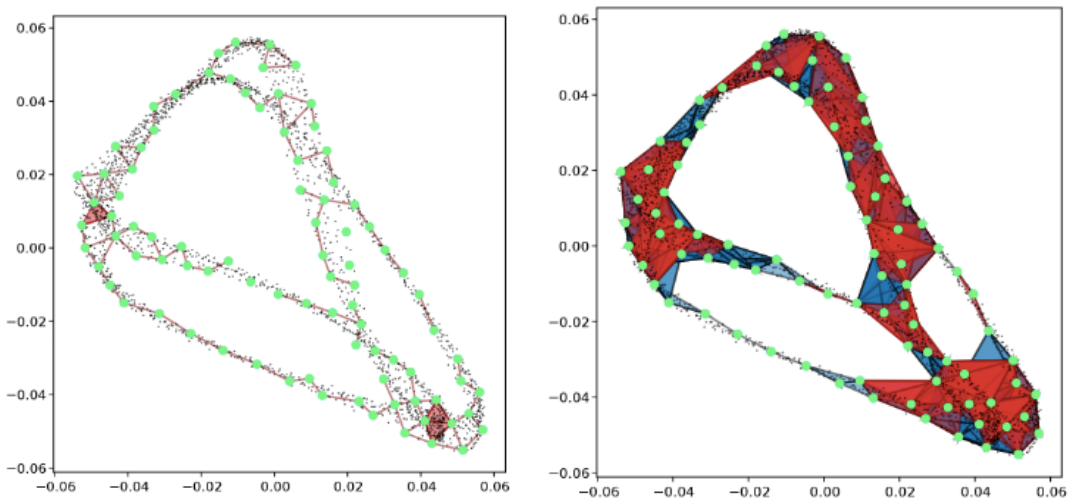


Figure 4.1: Witness complexes built on two dimensional delay reconstruction of 2000 points from the voltage recording of the note A440 ($\mathfrak{f} \approx 440$) on a Steinway grand piano using 80 landmarks (a) Early in a filtration simplicial complexes can capture local topology not reflective of the global topological structure. ($\epsilon \approx 0.008$) (b) Later in the filtration simplices can continue to be added to complexes in the filtration after the global topology has been captured that do not affect the homology. ($\epsilon \approx 0.01$

and the box-counting dimension D_{box} of the attractor of the underlying dynamics can cause $\epsilon_{\max}$ as solved for using Estimate 1 to be low. However, we may not know the attractor dimension. Variations in density of the data set mean certain local topological structures, like an m-dimensional neighborhood, will appear around certain vertices in L prior to others. Further, when the simplicies are still fleshing out interior regions of the data set, there should be an ϵ-range in the filtration when adding simplicies predominantly acts as a coarse-graining of the pre-existing simplicial model. During this ϵ range of the filtration, the newly added simplices span regions of the phase space already covered by smaller simplices that entered the filtration earlier. Higher dimensional cliques form to represent regions of the attractor previously represented by connected acyclic collections of lower dimensional cliques. See the witness complex in Figure 4.1(b) for an example of such coarse-graining. The red triangles are the two-simplices added to the filtration at the current $\epsilon \approx 0.01$ while the blue triangles are the two-simplices that entered the filtration earlier for smaller ϵ. Notice how the larger red triangles are generally covering a set of smaller blue triangles. As ϵ increases, this coarse-graining will no longer occur as the red triangles begin to cover the lack of data points that create the four major one-dimensional homological features.

During the coarse-graining ϵ range of the filtration, the lower dimensional homology of the complexes should remain stable. This encourages us to suggest overestimating when choosing $k_{\max}$ in Heuristic 3 via Heuristic 4, mindful of the costs of a larger filtration in later computations. Note also that one can only compute persistent homology up to dimension $k_{\max} - 1$ as all $k_{\max}$ cycles will be reported as homological features since there are no $k_{\max} + 1$ simplices in the filtration.

Figure 4.2 shows how r scales with $k_{\max}$ in practice for witness complexes built on 2000 witnesses and 100 landmarks modeling two dimensional delay reconstructions of musical instrument voltage recordings at 44.1 kHz (clarinet, viol) and and 10,000 witnesses and 100 landmarks modeling full three dimensional solutions to classical chaotic (Lorenz, Rössler). The viol, clarinet, Lorenz and Rössler point clouds have $(\Delta, \rho^{-\frac{1}{m}}) \approx (2.2, 0.12), (1.6, 0.1), (1.2, 0.13)$ and $(1.2,\ 0.13)$, respectively.

The similarity observed in Figure 4.2 of $\epsilon_{\max}$ across samples from the same data sets using this heuristic supports that terminating a filtration of witness complexes by the first appearance of

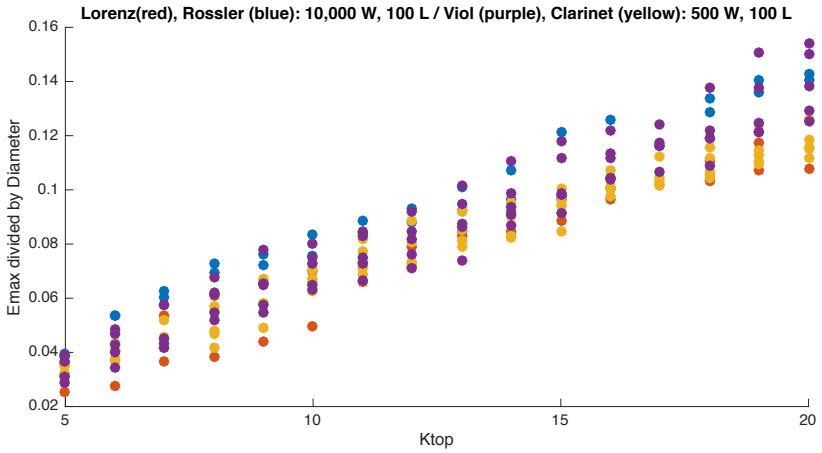


Figure 4.2: An increase in $\epsilon_{\max}$ from roughly 3% to 12% of the diameter is observed as $k_{\max}$, the dimension of the maximal simplex upon terminating the filtration, is increased from 5 to 20 for filtrations modeling full three dimensional solutions to classical chaotic dynamical systems and two dimensional delay reconstructions of musical instrument recordings: Lorenz (red), Rössler (blue), clarinet (yellow), viol (purple). Each point corresponds to a different point cloud from that class of data. The viol, clarinet, Lorenz and Rössler point clouds have average $(\Delta, \rho^{-\frac{1}{m}}) \approx (2.2, 0.12), (1.6, 0.1), (1.2, 0.13)$ and (1.2, 0.13), respectively.

a top dimensional simplex results in a consistent representation. It also suggests that the average value of $\epsilon_{\max}$ of a set of PRFs contains useful information about the geometric scale of the data.

This is different than computing the $k_{\max}$-skeleton of the witness complex across a range of ϵ; rather it provides a means of automating the choice of when to terminate a filtration of witness complexes. Automating $\epsilon_{\max}$ according to this local topological heuristic allows us to compare PRFs with slight variations in their geometric domain yet similar local topological structures in the filtrations they represent, a coarse-graining in the spirit of topological data analysis. Control of algorithm termination via this topological knob in the witness complex model allows for computational efficiency independent of knowledge of the scale of the data or the desired scale of identifiable features.

4.2 Preprocessing PRFs for Comparison

We introduce the heuristic of rescaling all PRFs so that they share extremal coordinates of $(0,0),(0,\hat{\epsilon}),(\hat{\epsilon},\hat{\epsilon})$ where $\hat{\epsilon}$ is such that a PRF, f, reporting a single homological feature that lives for the entire duration of the filtration has weighted L^2 norm $||f||_\phi = 1$. For instance, when the weight function is $\phi_1 = \phi(x,y) = 1$, then $\hat{\epsilon} = \sqrt{2}$. This rescaling is a preprocessing step that addresses missing data, allowing us to compare PRFs and infer global topological information from a simple statistic like the norm.

We also discretize the region of $\mathbb{R}^{2+}_{y\geq x}$ over which the PRF is defined into a grid of squares and assign to the i^{th} grid square the value of the PRF at its lower right corner, which results in an overapproximation. For a PRF f, we denote this value by $f^{(i)}$. This produces a set of features that can be used in machine learning algorithms.

4.3 Variances of PRFs

We compare two methods of computing the variance of a set of persistent homology rank functions (PRFs) $\{f_k\}_{k=1}^N$ using a weighted L^2 metric on the Hilbert space spanned by a set of PRFs. The first method begins by computing the weighted L^2 distance between each sample PRF

f_k and the mean PRF $\overline{\beta}$ given by Equation 2.32, and then averages the set of squared distances, providing the weighted L^1 norm of the point-wise variance function. In what follows, the notation $f^{(i)}$ refers to the i^{th} grid square in the our discretization of the PRF as discussed in Chapter 4.

$$\sigma^2(\{f_k\}_{k=1}^N) = \frac{\sum_{k=1}^N \left(\left(\sum_{(i)} \left(f_k^{(i)} - \overline{\beta}^{(i)} \right)^2 \cdot \phi^{(i)} dA \right)^{\frac{1}{2}} \right)^2}{N-1} \tag{4.1}$$

The second method begins by computing the variance locally, providing a point-wise variance function on $\mathbb{R}^{2+}$:

$$\mathcal{V}^{(i)}(\{f_k\}_{k=1}^N) := \frac{\sum_k \left(f_k^{(i)} - \overline{\beta}^{(i)} \right)^2}{N-1}, \tag{4.2}$$

and then computes the weighted L^2 norm of this point-wise variance function:

$$||\mathcal{V}(\{f_k\}_{k=1}^N)||_{\phi L^2} = \left(\sum_{(i)} \left(\frac{\sum_{j=1}^N \left(f_k^{(i)} - \overline{\beta}^{(i)} \right)^2}{N-1} \right)^2 \cdot \phi^{(i)} dA \right)^{\frac{1}{2}} \tag{4.3}$$

To see this relationship between these methods, consider

$$\sigma^2 = \frac{\sum_{k=1}^N \left(\left(\sum_i \left(f_k^{(i)} - \overline{\beta}^{(i)} \right)^2 \cdot \phi^{(i)} \cdot dA \right)^{\frac{1}{2}} \right)^2}{N-1} = \frac{\sum_{k=1}^N \sum_{(i)} \left(f_k^{(i)} - \overline{\beta}^{(i)} \right)^2 \cdot \phi^{(i)} \cdot dA}{N-1}$$

$$= \frac{\sum_{(i)} \sum_{k=1}^N \left(f_k^{(i)} - \overline{\beta}^{(i)} \right)^2 \cdot \phi^{(i)} \cdot dA}{N-1} = \sum_i \frac{\sum_{k=1}^N \left(f_k^{(i)} - \overline{\beta}^{(i)} \right)^2}{N-1} \cdot \phi^{(i)} \cdot dA = ||\mathcal{V}||_{\phi L^1}.$$

4.3.1 Functional Fano Factors

One concern is that some multiple of the variance of the distance to mean alone is not an appropriate statistic to threshold membership to a given regime. Because our mean is a function, there are a variety of ways that a sample can differ from the mean. Often mean and variance are not independent, though this is the case for normally distributed random variable. One way to address the lack of independence for single-variable statistics is to use the **Fano factor** [64],

$$F = \frac{\sigma^2}{\mu}.$$

The Fano factor acts as a unitless ratio of noise-to-signal, comparing a stochastic process to a Poisson process which has a Fano factor equal to one. A stochastic process with Fano factor less than one is called **sub-Poisson**, and suggests a low level of variability within the data. A process with Fano factor greater than one is called a **super-Poisson** and indicates high variation. Here our data is functional and we are looking to do statistics on a set of PRFs reporting the persistent homology from a filtration of witness complexes. We introduce two extensions of the idea of a Fano factor to PRFs as ways of capturing the scaled variation in the data.

Global Fano Factor

The **global functional Fano Factor** is the the weighted L^2 norm of the point-wise variance function, uniformly scaled by the weighted L^2 norm of the mean PRF $\overline{\beta_k}$:

$$\mathcal{F}^{\text{glo}}(\{f_k\}_{k=1}^N) := \sum_{(i)} \frac{\sum_{k=1}^N (f_k^{(i)} - \overline{\beta}^{(i)})^2}{(N-1)\left(\sum_{(i)} \left(\overline{\beta}^{(i)}\right)^2 \phi^{(i)} dA\right)^{\frac{1}{2}}} \quad (4.4)$$

One way this definition is limited is that it aggregates the homological information across the range of the filtration. Two PRFs can have the same weighted L^2 norm yet represent different types of persistent homology. The idea of **persistent entropy** recently introduced in [43] measures of variation in the types of persistent homology as distinguished by lifespan. A PRF with high persistent entropy would have a plethora of homological features with a wide range of lifespan

whereas a PRF with low persistent entropy would have a lot of features with the same lifespan. While data sets representing both of these situations could have equal weighted L^2 norms of the mean and equal weighted L^2 point-wise variance, and thus equal global functional Fano factors, the lower persistent entropy data set would have greater local variability whereas the higher persistent entropy would have smaller local variability yet this variability would be spread across the PRF. We would like to distinguish between these two cases. This motivates our definition of the **local functional Fano factor**.

Local Fano Factor

The **local functional Fano factor**, $\mathcal{F}_k^{\text{loc}} : \{f_k\}_{k=1}^N \to \mathbb{R}$, reveals the variance of a set of persistent homology rank functions (PRFs) as a function of observation window $(a, b) \in \mathbb{R}^{2+}$ in a filtration.

$$\mathcal{F}^{\text{loc}}(\{f_k\}_{k=1}^N)^{(i)} := \sum_{k=1}^{N} \frac{\left(f_k^{(i)} - \overline{\beta}^{(i)}\right)^2}{(N-1)\overline{\beta}^{(i)}}, \tag{4.5}$$

$$||\mathcal{F}^{\text{loc}}(\{f_k\}_{k=1}^N)||_{\phi L^2} := \left(\sum_{(i)} \left(\frac{\sum_{k=1}^N \left(f_k^{(i)} - \overline{\beta}^{(i)}\right)^2}{(N-1)\overline{\beta}^{(i)}} \right)^2 \phi^{(i)} dA \right)^{\frac{1}{2}} \tag{4.6}$$

This allows us to scale the variance according to the local mean. If there are some observation windows for which the variance is high yet the mean number of observed homological features is also high, we consider this of less importance than if the variance is high yet the mean number of observed homological features is low. This follows the philosophy that if there are only two features, each feature is very important in conveying the homology. Rather if there are 50 features, a variation of say 10 features still communicates the order of magnitude of the homology effectively. We see this situation arise often in persistent homology computations, where an abundance of short-lived homology is born and dies in a concentrated interval of the filtration. The sheer amount of local topology captured can contribute to a large weighted L^2 norm that can minimize important variability in the number of smaller, long-lived homological features when computing the global

functional Fano factor. Taking these long-lived features to be our topological signal, the short-lived features be come noise in our topological representation of the data, the weighted L^2 norm of a PRF. Using the local functional Fano factor fixes this misrepresentation of the noise-to-signal ratio by separating the variability in the signal from the variability in the noise, and weighing them accordingly.

4.4 Family of Topological Classifiers

In our classification, we train on the PRFs $\{f_n\}_{n=1}^{s}$ obtained from s sliding windows from each data set by computing the mean PRF as a representation of that class. We then test membership to each class on s additional sliding windows from each class by comparing the L^2 distance of the PRF g_n associated to a sliding window to the representative mean PRF $\overline{\beta}$ of an instrument class. We parameterize this membership test by setting a threshold distance away from the representative mean PRF $\overline{\beta}$ a sample PRF g_n can be to still be classified as a member of that instrument class. We report this threshold distance in units z where z is the multiple of the weighted L^2 norm of the point-wise variance function, σ^2.

$$d_{\phi_1}(g_n, \overline{\beta}) \leq z \cdot \sigma^2, \tag{4.7}$$

$$\text{where } \sigma^2 = \frac{1}{s-1} \sum_{n=1}^{s} ||f_n - \overline{\beta}||_{\phi_1}^2. \tag{4.8}$$

Experimentation revealed little distinction between the choice of threshold units in practice. We choose to parameterize our family of classifiers based on σ^2, the weighted L^2 norm of the point-wise variance function, rather than the global or local functional Fano factor, but include their introduction for their potential contribution to future topological data analyses.

4.5 Experimental Results

We explore the strength of these classifiers over the (**number of witnesses, number of landmarks**) parameter space as reported by the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curves. Recall that a classifier with an equal likelihood of correctly and erroneously assigning membership has an $AUC \approx 50\%$ while a perfect classifier has an $AUC \approx 100\%$.

In the tables below, we present the area under the curve (AUC) of the Receiver Operator Characteristic (ROC) curves for the family of topological binary classifiers. Each table represents a binary classifier testing membership to a certain data set, e.g. Lorenz membership. The rows of the tables represent a variety of **number of witnesses**, or amount of delay reconstructed data points representing the dynamics. Note the **sliding window size** must include $m \times \tau$ more points than the **number of witnesses** in order to create the **number of witnesses** delay reconstruction vectors comprising our higher dimensional point cloud. The columns of the tables represent a variety in the **number of landmarks** used in our model, which serve as the vertices in the filtrations of witness complexes.

Recall that an $AUC \approx 1$ signifies a classifier that is not only able to correctly classify Membership versus non-Membership for a single threshold value but is robust with respect to changing this threshold value. That is, the data sets being distinguished are extremely well-separated and clustered tightly around their own means. Lower $AUCs$ less than or equal to 0.5 suggest the topological binary classifier is attempting to distinguish data sets which are not well-separated or well-represented by their means. Either the mean representations, and the data sets themselves, themselves are too similar with respect to this topological signature or the variance within a data set is too high for this topological signature to be distinguishing.

$\lvert\Gamma\rvert, \lvert L\rvert$	10	50	100	500
500	1.00	0.34	1.00	0.34
1000	0.67	0.34	1.00	0.44
4000	0.00	1.00	1.00	0.78
5000	0.00	0.00	1.00	1.00

Table 1: AUC: Clarinet Membership

$\lvert\Gamma\rvert$,$\lvert L\rvert$	10	50	100	500
500	1.00	1.00	0.33	0.17
1000	1.00	1.00	0.00	0.67
4000	0.83	1.00	1.00	1.00
5000	1.00	1.00	0.33	1.00

Table 1: AUC: Viol Membership

Figure 4.3: The AUC of ROC for (a) Clarinet Membership (b) Viol Membership over the (**number of witnesses**, **number of landmarks**) parameter space.

4.5.1 Distinguishing Musical Instruments

We first demonstrate the feasibility of this topological classifier for distinguishing simple periodic signals on experimental voltage time series acquired from live recordings of the same note on a clarinet and a viol. Using the same delay parameter irrationally proportional to the fundamental frequency of the recorded musical note and reconstructing in two dimensions, the present subharmonics reveal themselves as projections of cycles which would be embedded on the surface of a torus in a higher dimension, creating extra homology. Our intention is to extract this global topology of the delay reconstruction as a coarse-grained topological description of the dynamics via the weighted L^2 norm of the mean PRF, and distinguish between music instrument recordings based on differences in this large-scale topology. We take a sample size of 10 sliding windows from each music instrument recording to train our classifier and compute a mean PRF. We then test our classifier against 10 sliding windows from within the class and 10 sliding windows from the other instrument class. We terminate the filtrations of witness complexes at $k_{\max} = 9$.

Tables 4.3 (a) and 4.3(b) show that we observe perfect membership testing for both the viol classifier and clarinet classifier at four pairs of ($\lvert\Gamma\rvert$,$\lvert L\rvert$) parameter values: (500, 10), (4000, 50), (4000, 100) and (5000, 500). **We anticipate that, despite successful classification, filtrations of witness complexes with ($\lvert\Gamma\rvert$, $\lvert L\rvert$) = (500, 10) and ($\lvert\Gamma\rvert$, $\lvert L\rvert$) = (5000, 500)**

fail to capture an accurate global topological signature for the data. This is validated in our later experimentation in Section 5. Heuristic 5 presented in Chapter 5 selects $(|\Gamma|, |L|) = (4000, 100)$.

Counterintuitively, using all 500 witnesses as landmarks to model the delay reconstructed sliding window of time series data does not provide better classification. In fact, the Čech complex on 500 data points results in a poor classifier. This can be seen in Table 4.3(b) with an $AUC \approx 33\%$ and Table 4.3(a) with an $AUC \approx .17$. We expect that this is because the $k_{\max}$ cliques in the Čech complex span a smaller proportion of data set than those where the **number of landmarks** is much less than the **number of witnesses**, and terminating filtrations for a fixed $k_{\max}$ to fix computational costs prevents such filtrations from capturing the distinguishing global topology. That is, the persistent homology is only reporting noise since the filtration was not able to explore topology at the relevant geometric scales under the cost ceiling we impose.

4.5.2 Classifying Chaotic Dynamical Systems

Many methods exist for distinguishing between periodic dynamical systems or classifying whether a system is periodic or chaotic. Yet distinguishing between time series measurements from different chaotic dynamical systems is a much more challenging problem. Nuanced descriptions like the Lyapunov exponent and box-counting dimension that tease apart the mechanism and result of chaos require convergence in a limit and therefore a lot of data. We now apply our topological classifier to distinguishing between the classical continuous chaotic dynamical systems of Lorenz and Rössler for a range of **number of landmarks** and **number of witnesses** in the construction of witness complexes. We train our classifiers on PRFs corresponding to 10 sliding windows of delay-reconstructed synthetic time series data created by projecting full solutions these systems of equations onto their x-axes. We test our classifiers on 10 samples within the class and 10 samples from the alternative data set.

According to the heuristic presented in Section 4.1, we set $k_{\max} = 15$. Thus filtrations built upon 16 landmarks will certainly terminate in the complete complex on 16 vertices. Because of this,

$\lvert\Gamma\rvert, \lvert L\rvert$	16	25	50	100	200	500
500	0.91	0.81	0.73	0.41	0.27	0.58
1000	0.86	0.75	0.64	0.68	0.31	0.42
2000	0.81	0.54	0.65	0.71	0.49	0.17
5000	1.00	0.88	0.79	0.68	0.95	0.94
10000	1.00	1.00	0.80	0.93	0.92	1.00
20000	1.00	1.00	1.00	0.77	0.95	0.94
40000	1.00	1.00	1.00	0.20	0.72	0.98

Table 1: AUC: Lorenz Membership

$\lvert\Gamma\rvert, \lvert L\rvert$	16	25	50	100	200	500
500	0.76	0.82	0.67	0.62	0.75	0.54
1000	0.49	0.54	0.80	0.91	0.91	0.73
2000	0.49	0.52	0.73	0.99	0.90	0.88
5000	0.32	0.72	0.52	0.84	0.79	0.91
10000	0.94	0.81	0.71	0.48	0.84	0.98
20000	0.70	1.00	1.00	0.57	0.53	0.80
40000	0.00	0.30	1.00	0.90	0.62	0.98

Table 1: AUC: Rössler Membership

Figure 4.4: The AUC of ROC for (a) Lorenz Membership (b) Rössler Membership over the (**number of witnesses**, **number of landmarks**) parameter space.

we expect the persistent homology to die early and have a short lifespan relative to the filtration length. Whether or not the topological features reflective of the dynamics can be captured by such few landmarks is largely a property of the data set. Increasing the **number of landmarks** increases the resolution and the amount features detectable by persistent homology, of which there are infinite for fractal chaotic data like the Lorenz and Rössler. For finite data sets, the **number of landmarks** is bounded above by the **number of witnesss** and for too few witnesses, the desired descriptive scaling properties of the attractor also might not be extractable for the given range of topological perspectives. Counter to our colloquial description of the two chaotic systems, complications arise when a trajectory of the Lorenz system is solved only long enough to visit one butterfly wing or a trajectory of the Rössler system goes over the bridge just as frequently as it goes around the equilibrium in the xy-plane, creating one dominant hole in the former case and two in the latter.

Like with the musical instrument recordings, we see that modeling the delay reconstructed chaotic time series by filtrations of Čech complexes on 500 data points forms the basis of poor classifiers with $AUC \approx 0.58, 0.54$ for Lorenz and Rössler, shown in Tables 4.4 (a), (b), respectively. Yet unlike the periodic musical instrument recordings, decreasing the **number of landmarks** while keeping the **number of witnesses** equal to 500 fails to achieve an $AUC \approx 1$. This has to do with the time scale of the dynamical system being observed in relation to the time step at

which it is measured. The period of the musical instrument recordings is about 100 samples so 500 data points well-approximates the projection of the limit cycle attractor in the delay reconstructed space. However, 500 data points fails to capture the invariant measure of the chaotic dynamical systems with time steps as described in Chapter 3.

Because we expect the mean PRF to better represent the data set when the **number of witnesses** is sufficient to achieve the invariant measure along the attractor, we expect the AUC to increase as we increase the **number of witnesses**. For both the musical instrument recordings and projections of chaotic dynamics, keeping the **number of landmarks** fixed at 500 while increasing the **number of witnesses** improves the classification. We observe this general trend for fixed **number of landmarks** for the Lorenz classifier in the columns of Table 4.4(a) , perhaps less so for the Rössler classifier. Stability in the classifier along this parameter axis can signal we have enough data to achieve the invariant measure and properly represent the dynamics. While we need this lower bound, we are most interested in identifying changes to model parameters that make our TDA more successful at a cheaper cost, not more expensive. As the **number of landmarks** contributes more to the cost of computing the persistent homology than the **number of witnesses**, we prioritize using a lower **number of landmarks** to maintain an efficient regime-shift detection algorithm, noting that the **number of witnesses** is also important as it controls how quickly and how small of a regime-shift can be detected.

Surprisingly, using less **number of landmarks** results in robustly successful topological classifiers for chaotic dynamical systems. All of the richness of the fractal structure and invariant measure observable with more landmarks seems to make the persistent homology less discerning than that reporting the coarse-grained global topology captured by less landmarks. The pocket of lower AUC values approximately 0.50 in the Table 4.4 (b) with $AUC \approx 50$ for 16 and 25 **number of landmarks** and 1000 and 2000 **number of witnesses**, however, warns of using too few witnesses and acts as a reminder that these binary classifiers are not symmetric. Model parameters that create successful and robust membership tests for one data set may not be the same as those that best determine membership to the other data set, just as model parameters that best capture the

topology of the dynamics of one system are likely not the same for all dynamical systems but rather depend on the time-scale and attractor geometry. This is acceptable for a regime-shift detection algorithm where model parameters can be established for the current regime of the time series and recalculated after a regime-shift occurs.

There is however shocking variability of this classifier across these size-based parameter values for the witness complex construction on delay reconstructed chaotic time series data. For instance, Table 4.4(b) shows that the Rössler classifier AUC drops to approximately 0.30 for $40,000$ witnesses and 25 landmarks, while the $AUC \geq 80$ for $10,000$ and $20,000$ witnesses using the same number of landmarks. Similarly, the Lorenz classifier performs well for $40,000$ witnesses with an $AUC \geq 0.70$ for most **number of landmarks** explored, except for 100 where the AUC drops to 0.20. Another instance of instability in the Rössler classifier performance is for fixed **number of witnesses** equal to $20,000$. For all **number or landmarks** explored except for 100 the $AUC \geq 80$ yet for 100 landmarks it drops to less than 0.50. For $20,000$ witnesses, using 100 and 200 landmarks also underperforms compared to other parameter values. One reason for this variability is the small sample size used to compute our statistics. Note that for the parameter values for which the Lorenz classifier fails, the Rössler classifier achieves an $AUC \approx 0.90$. Likewise when the Rössler classifier does poorly, the Lorenz classifier has an $AUC \approx 1$. This indicates that another reason for classifier instability is a high variance of the in-class data.

4.5.3 Detecting Regimes of Dynamic Brunel Neuronal Network

Recent work done by Hess et al demonstrates successful classification of Brunel neuronal network dynamics from a combination of zero and one-dimensional persistent homological features for filtrations of Vietoris-Rips complexes constructed from three measures of dissimilarity between the spike trains of the individual neurons [28] . This inspired us to apply our topological membership test to one-dimensional persistent homology of filtrations of witness complexes modeling the delay reconstruction of the average voltage of excitatory neurons from each of distinct regimes.

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.0	0.46	0.32	0.0
1000	0.0	1.0	0.88	0.82
2000	0.0	0.6	1.0	1.0
5000	0.0	0.0	0.0	1.0
10,000	0.0	0.0	0.0	0.6
20,000	0.0	0.4	0.4	0.8

Table 1: I2G3 Membership (vs I4G6)

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.00	0.62	0.84	0.00
1000	0.00	1.00	0.72	0.52
2000	0.00	1.00	0.92	1.00
5000	0.00	0.20	0.80	0.96
10,000	0.00	0.20	0.20	1.00
20,000	0.00	0.20	0.80	0.92

Table 1: I2G3 Membership (vs I1G4P5)

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.00	0.24	0.8	0.00
1000	0.00	1.0	1.0	0.2
2000	0.00	0.8	1.0	0.00
5000	0.00	0.4	0.6	0.00
10,000	0.00	0.4	0.2	0.8
20,000	0.00	1.0	1.0	1.0

Table 1: I2G3 Membership (vs I2G5)

Figure 4.5: The AUC of ROC for the I2G3 regime of the Brunel neuronal network dynamics over the (**number of witnesses**, **number of landmarks**) parameter space.

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.20	0.70	0.96	1.0
1000	0.80	1.0	1.0	0.80
2000	0.6	1.0	1.0	1.0
5000	0.6	1.0	1.0	1.0
10,000	0.70	1.0	1.0	1.0
20,000	0.00	0.52	0.80	0.96

Table 1: I4G6 Membership (vs I2G3)

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.20	0.24	0.52	0.71
1000	0.80	1.00	0.96	0.44
2000	0.60	0.84	0.92	0.60
5000	0.72	1.00	0.90	0.96
10,000	0.88	0.60	0.80	0.72
20,000	0.00	0.64	0.58	0.92

Table 1: I4G6 Membership (vs I1G4p5)

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.32	0.34	0.20	0.50
1000	0.54	0.73	0.76	0.20
2000	0.44	0.84	0.72	1.00
5000	0.62	0.72	0.76	1.00
10,000	0.40	0.20	0.52	1.00
20,000	0.00	0.52	0.52	0.76

Table 1: I4G6 Membership (vs I2G5)

Figure 4.6: The AUC of ROC for the I4G6 regime of the Brunel neuronal network dynamics over the (**number of witnesses**, **number of landmarks**) parameter space.

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.50	0.28	0.52	0.00
1000	0.40	1.00	0.88	0.52
2000	0.30	0.16	0.52	1.00
5000	0.20	0.72	0.80	0.96
10,000	0.60	0.72	0.80	1.00
20,000	0.40	0.72	0.66	0.92

Table 1: I1G4P5Membership (vs I2G3)

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.80	0.26	0.50	0.84
1000	0.00	1.00	0.76	0.76
2000	0.80	0.70	0.6	0.4
5000	0.20	0.52	0.74	0.80
10,000	0.52	0.52	0.48	0.84
20,000	0.18	0.64	0.70	1.00

Table 1: I1G4P5Membership (vs I4G6)

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.60	0.52	0.66	0.48
1000	0.40	1.00	1.00	0.76
2000	0.36	0.56	0.44	0.64
5000	0.26	0.32	0.36	1.00
10,000	0.48	0.56	0.48	0.96
20,000	0.26	0.44	0.40	0.72

Table 1: I1G4p5 Membership (vs I2G5)

Figure 4.7: The AUC of ROC for the I1G4p5 regime of the Brunel neuronal network dynamics over the (**number of witnesses**, **number of landmarks**) parameter space.

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.20	0.63	0.80	0.60
1000	0.5	1.0	0.80	0.96
2000	0.5	0.20	0.36	1.0
5000	0.5	1.0	1.0	1.0
10,000	0.80	1.0	1.0	1.0
20,000	0.20	0.68	1.0	0.92

Table 1: I2G5 Membership (vs I2G3)

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.48	0.66	0.80	0.44
1000	0.82	0.96	0.92	0.88
2000	0.74	0.60	0.72	0.92
5000	0.38	0.52	0.72	0.88
10,000	0.60	0.80	0.84	1.00
20,000	0.36	0.64	0.61	0.68

Table 1: I2G5 (vs I4G6)

$\lvert\Gamma\rvert,\lvert L\rvert$	13	50	100	500
500	0.22	0.22	0.32	0.48
1000	0.60	0.60	0.52	0.84
2000	0.60	0.60	0.12	1.00
5000	0.92	0.92	0.0.74	0.92
10,000	0.92	0.92	0.64	0.96
20,000	0.84	0.84	0.92	0.88

Table 1: I2G5 Membership (vs I1G4P5)

Figure 4.8: The AUC of ROC for the I2G5 regime of the Brunel neuronal network dynamics over the (**number of witnesses**, **number of landmarks**) parameter space.

The I2G3 classifier seems to consistently perform poorly with an $AUC \approx 0$, illuminating a flaw in the design of this classifier. If the persistent homology of samples from a data set is consistently trivial with a PRF of weighted L^2 norm equal to zero, the mean and variance can both equal zero. It makes sense that this occurs for the I2G3 data set because the average excitatory voltage as a function of time in this regime is a highly discontinuous signal that delay reconstructs to be isolated points. The lack of one-dimensional homology can lead to sets of zero-valued PRFs with a trivial mean and variance.

However, our topological classifier using the L^2 metric on PRFs of witness complexes can sometimes serve as an extremely successful and cheap method for distinguishing between two different regimes of the Brunel dynamics. Utilizing they two-dimensional delay reconstruction representation of sliding windows from the average excitatory voltage time series to construct a witness complex and computing statistics with the PRF in tandem provide a cheap methodology for classification of dynamical systems. The I2G5 classifier suggests the general trend that more witnesses and more landmarks leads to better classification, while the I4G6 and I1G4p5 classifiers appear more random in their moments of success and failure.

The overall variability in success across model parameters demonstrates the sensitivity of these topological classifiers to the **sliding window size** and **number of landmarks**. To investigate how the topological signature itself is a function of the **sliding window size** and **number of landmarks**, we next introduce in Chapter 5 a set of weight functions for the weighted L^2 norm that help reveal the types of topology present in a filtration. This intuition allows us to then select parameter values according to a heuristic that aim to maximize the lifespan of global topology and minimize the existence of local topology reported by the persistent homology. We hope that this simplicity allows us to extract a consistent topological signature representative of the dynamical system.

Chapter 5

Exploring Topology Across Model Parameters

5.1 Parameter Space Affecting Model Size

Computing the persistent homology of a filtration of witness complexes from time series data requires selecting numerous parameters. We introduce heuristics for selecting two major parameters affecting the size of the witness complex construction from time series data: the **sliding window size** determines the amount of time series data used in a single representation of the dynamics and the **number of landmarks** affects the resolution of the analysis. The goal is to construct a computationally cheap filtration of witness complexes and extract a coarse-grained topological signature from the associated PRF. A good topological signature should be consistent within a data set and discerning between distinct data sets.

An historically guiding philosophy behind persistent homology is that short-lived features are noise and long-lived features are meaningful. What is seen as short or long is sensitive to the range of the scale parameter over which the persistent homology is computed. The scales at which the data is analyzed must meet the demands of the analysis - i.e. high-resolution, real-time. Yet this is often limited by computational capabilities.

Niyogi, Smale and Weinberger provide some of the earliest sampling requirements for provably correct topological data analysis when the data lies on a manifold, demonstrating that for an $\frac{\epsilon}{2}$-dense sample such that $0 < \epsilon < \sqrt{\frac{3}{5}}\tau$ from a manifold with condition number $\frac{1}{\tau}$ measuring "curvatures and nearness to self-intersections", the Čech complex for ϵ has the same homology as the underlying manifold [50] with high probability. When the underlying manifold is a circle, the condition number

is the radius; for an ellipsoid, it is half the length of the smallest principal radius. Manifolds with corners do not fit into this framework as they have condition number 0. So for manifold learning purposes via persistent homology where the scale parameter ϵ is given by the Euclidean distance between data points in $\mathbb{R}^d$, we have an expected upper bound on the sample of size necessary to achieve $\frac{\epsilon}{2}$ density with high probability where a filtration of Čech complexes should accurately reflect the homology of the underlying manifold.

5.1.1 Analyzing Time Series is Different

When time series data arrives sequentially as in a stream, a buffering or sliding window analysis demands an alternative means of measuring what is enough data for successful TDA. Time series are discrete sequential measurements from a dynamical system

$$\{x(t)\}_{t=1}^{N}, \quad x : \mathbb{Z} \to \mathbb{R}$$

and properties of the dynamical system can be reflected in the time series. Traditional time series analysis techniques are based on frequency analysis and statistically aggregated information. This is inadequate for nonlinear, non-stationary systems [20]. Here we hypothesize that TDA can be fruitful. As a set of points, a finite data set is zero-dimensional. This is fine as TDA is highly amenable to discrete data, e.g. manifold learning can happen from discrete point cloud samples. Just as a point cloud can approximate a manifold, a the discrete sequential scalar measurements of a time series approximate a curve segment embedded in $\mathbb{R}^2$,

$$\{(t, x(t))\}_{t=1}^{N} \subset \mathbb{R}^2.$$

Takens theory of delay reconstruction says one can produce a diffeomorphic representation of the attractor of the underlying dynamical system in higher dimensional ambient space by using as further coordinates successive delays of a generic measurement function. Assuming a time series is a generic measurement function of the underlying dynamical system, the delay reconstructed point

cloud

$$\{(x(t), x(t-\tau), x(t-2\tau), \ldots, x(t-(d-1)\tau))\}_{t=1}^{N-d\cdot\tau} \in \mathbb{R}^d$$

provides an approximation of a much more interesting manifold from which to extract a topological signature for the time series. For example, periodic data has a simple topological structure when delay reconstructed, resulting in a circle when the reconstruction meets the theoretical requirements on the dimension and the delay described in Chapter 2. Quasiperiodic data delay reconstructs to form invariant tori in the state space. However, these circles and tori can be highly warped geometrically and knotted in $\mathbb{R}^d$ and projections into lower dimensions can create extra homology.

Apart from the warping and projection due to the geometry of the dynamics, time series analysis via TDA is challenging because how much data to consider for a single persistent homology computation involves both the discrete time step at which and amount of time over which the system is observed, as well as the intrinsic periodicities of the underlying dynamics. Acceleration caused by the interactions between the equilibria of the dynamical system can then lead to non-uniform density in the point cloud representation of the dynamics attractor once the time series has been delay reconstructed. Chaotic data presents further challenges as non-uniform density can also arise transverse to the flow reflecting a complicated invariant measure on the attractor. These variations in density could possibly lead to shorter lifespans for significant topological features, and longer lifespans for spurious topological features, in an analysis using persistent homology.

If an attractor is ergodic, then normalized time average of where a typical trajectory spends its time along the attractor in state space is independent of initial conditions and gives the invariant measure almost everywhere. This guaranteed distribution is achieved only in the infinite time limit, and figuring out when a finite data set reflects these asymptotic properties can be difficult [46]. Yet time-sensitivity and real computational limitations demand efficient analysis with less data, i.e. quickly detecting change points in stream data means one cannot wait to do an intensive analysis on

a large window. Picking a minimal sufficient **sliding window size** is thus crucial to the remainder of the analysis.

To reduce computations, we chose to compute the persistent homology of a filtration of witness complexes [45]. The witness complex is defined by a relation that further determines the resolutions at which one can extract meaningful topological information. Recall from Chapter 2, the witness complex uses a subsample of the data points called landmarks as the vertices in the simplicial complexes. The **number of landmarks** affects the topology of the witness complex built upon the landmark set; a greater number of landmarks means the scale of homological features that can be detected is smaller. Ideally, the extracted topological signature is representative of the underlying dynamics. Universally though, this can only be true for some range of **number of landmarks**.

Beyond creating a model that performs well for membership testing between different dynamical systems as investigated in Chapter 4, it is an important question how to detect when the model is capturing the homology correctly. Despite trends in the AUC of ROC curves indicating that more witnesses and less landmarks produce better topology-based classifiers for chaotic dynamical systems, some seemingly random and shockingly low AUC values force us to question the topological representation of our data across model parameter values. In this current Chapter 5, we explore topological representations through the lens of continuous statistics like weighted L^2 norms of the functional representation of persistent homology, the PRF. In the following Chapter 6 we suggest a set of discrete statistics on the persistence diagram for descriptive topological signatures from the persistent homology.

Our contribution is an automatable construction of witness complexes from time series to efficiently capture the global topology of the underlying data, a step towards an online topological change-point detection algorithm. In Chapter 4, we introduced a local topological heuristic for terminating a filtration of witness complexes. We now introduce a method to determine a sufficient lower bound on the amount of data needed to obtain a correct and consistent global topological signature reflective of the dynamics. In particular, we use ratios between weighted L^2 norms on

persistent homology rank functions introduced in Chapter 2 to select the **sliding window size** and **number of landmarks** for witness complex construction from time series data. We demonstrate the success of our heuristics for regime-shift detection of time series via superior binary classification between individual musical instruments from voltage recordings and different chaotic systems from projection onto a single coordinate, and discuss its pitfalls when characterizing distinct regimes of the Brunel neuronal network leaky-integrate-and-fire model from average excitatory firing rates.

5.2 Topological Heuristics for Parameter Selection

5.2.1 Birth & Death Weighted L^2 Norms

We use a set of weighted L^2 norms on PRFs in heuristics that let us choose a **sliding window size** and **number of landmarks** necessary to compute a consistent topological signature reflective of a given dynamical system. Recall from Chapter 2 that a PRF is a functional representation of the persistent homology of a filtration. The reduction of persistent homology to a single number like a norm is convenient yet looses a lot of information. Integration such as that used to compute an L^2 norm of a PRF is a common noise-filtering technique in time series analysis used to minimize the effects of outliers; yet here it aggregates the local and global topological information. For instance, a PRF f with a single homological feature that lives for the entire duration of the filtration and a PRF g with n^2 homological features, each of lifespan $1/n^{\text{th}}$ of $\epsilon_{\max}$, will both have $||f||_{\phi_1} = ||g||_{\phi_1} = 1$. Weight functions that de-emphasize persistent homology classes that die early, are born late, or have generally short lifespans can tease apart the global and local topology reported in a single statistic like the weighted L^2 norm.

To distinguish between these and other topologies, we propose to use the following weight functions in weighted L^2 metrics on the space of PRFs:

$$\phi_B(x, y) = ae^{-bx} \tag{5.1}$$

$$\phi_D(x,y) = ae^{b(y-\hat{\epsilon})} \tag{5.2}$$

These weight functions make distinctions between different types of short-lived features. For $a, b > 0$, the weight function ϕ_B decreases the contribution to the L^2 norm from homological features born later in the filtration and increases the contribution of homological features that are born early but do not persist for the entirety of the filtration. In contrast, the weight function ϕ_D decreases the contribution to the L^2 norm from homological features that die earlier in the filtration while increasing the contribution for late born features. See Figure 5.1 (b) for the number of early-born short-lived homological features with lifespan $(0, \frac{\hat{\epsilon}}{n})$, shown in Figure 5.1 (a) needed to produce an otherwise empty PRF f with weighted L^2 norm equal to 1 as a function of n; weight functions $\phi_B = 2.12e^{-2x}, \phi_1 = 1, \phi_D = 2.12e^{2(y-\sqrt{2})}$ in yellow, blue and red, respectively. These values of $a, b, \hat{\epsilon}$ are chosen according to the rescaling step introduced in Chapter 4. Unless a homological feature is born exactly at the beginning of the filtration and lives until the very end, its contribution to the L^2 norm will be modified by both of these weight functions. These differences can inform us about the topology captured by filtrations across the parameter space.

We could alternatively use as weight functions step functions ϕ_E, ϕ_U over the lower left and upper right corner regions of the PRF, where $E = [0, \max(\epsilon_D)]^2|_{x \leq y}$ and $U = [\min(\epsilon_B), \epsilon_{\max})]^2|_{x \leq y}$. Like with the exponential functions ϕ_B, ϕ_D in Equations 5.1, 5.2, some parameters - here, a $\max(\epsilon_D)$ and a $\min(\epsilon_B)$ - must be chosen. These determine the region of the PRF relating to an abstract lifespan over which we integrate the portion of the lifespans of homological features that overlap with this abstract lifespan. For instance, for $\max(\epsilon_D) = \frac{\epsilon_{\max}}{n}$ then E would be the triangular region above the diagonal and below the dashed green line in Figure 5.1(a). Similarly for $\min(\epsilon_B) = \frac{(n-1)\cdot\hat{\epsilon}}{n}$, U would be the triangular region above the diagonal and to the right of the pink dashed line in Figure 5.1.

While the step functions provide a rough count of the number of homological features that exist for **at least** a short amount of time around a specific range in the filtration, they have the

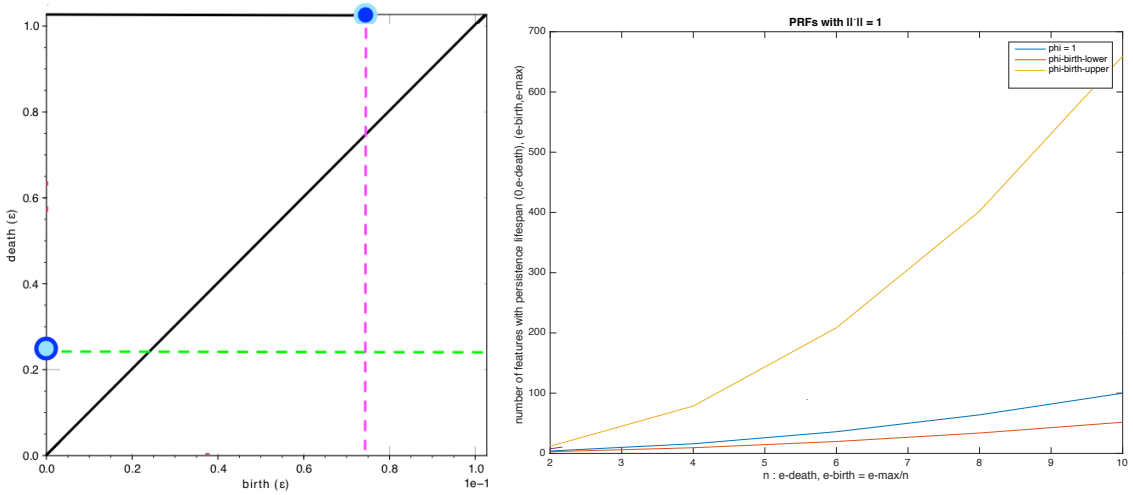


Figure 5.1: (a) Weight functions can be utilized to change the contribution of different types of short-lived features to the norm of a PRF. (b) Number of homological features of lifespan $(0, \frac{\hat{\epsilon}}{n})$ needed to produce a PRF representing no other homology with weighted L^2 norm equal to 1 for weights ϕ_1, ϕ_B, and ϕ_D in blue, red, and yellow, respectively. By symmetry, we also get this information for homological features with lifespan $(\frac{(n-1)\cdot\hat{\epsilon}}{n}, \hat{\epsilon})$.

disadvantage of aggregating the local and global topology since the contributions from long-lived features are not discernible from those of short-lived features. In particular, looking at the ϕ_E weighted L^2 norm of a PRF will not provide information about the entire filtration. In contrast, the ϕ_D weighted L^2 norm shrinks the contribution to the norm of features existing in the E region so that the weighted L^2 norm more accurately reflects the number of long-lived features.

Comparing the ϕ_E and ϕ_U norms of a PRF can reveal a difference in the number of features earlier and later in the filtration, yet this information alone cannot separate a filtration containing lots of early-born, short-lived local topology with few long-lived global topological features and a filtration with all early-born features dying early followed by a few late-born homology. To do so we would want to use another step function for the region $T = [\min(\epsilon_D), \max(\epsilon_B)]^2|_{x \leq y}$. Under this guiding philosophy, a step function on the T region aims to count the number of long-lived homological features present in the filtration yet would not provide the information on the short lived homological features captured by the filtration.

The rescaling introduced in Chapter 4 results in setting $\hat{\epsilon} = \sqrt{2} \cdot \max(\epsilon_D)$ for the weight function $\phi_E, \hat{\epsilon} = \sqrt{2} \cdot (\epsilon_{\max} - \min(\epsilon_B))$ for the weight function ϕ_U, and $\hat{\epsilon} = \sqrt{2} \cdot 1/2(\epsilon_{\max} - \min(\epsilon_D)) \cdot \max(\epsilon_B)$. Depending on how $\max(\epsilon_B), \min(\epsilon_B), \max(\epsilon_D), \min(\epsilon_D)$ are selected, we can force all of the homological features in the T region to contribute to both of the ϕ_E, ϕ_U weighted L^2 norms. Then a PRF with $\phi_E \approx \phi_U \approx \phi_T$ would reflect a filtration capturing a consistent number of topological features across the filtration. This trivially holds true when a filtration picks up no homology, so this should be checked explicitly.

Allowing each of the ϕ_B, ϕ_D weight functions to exceed the unweighted L^2 norm gives us more information than normalizing the weights to be strictly less than the standard L^2 norm; forcing all weight functions to be equal when a PRF reports exactly n homological features of lifespan $[0, \epsilon_{\max}]$ allows $\phi_B >> \phi_1$ to indicate an abundance of homology in the L region and $\phi_B << \phi_1$ to indicate an abundance of homology in the U region. So we really need only to compare either ϕ_B or ϕ_D with ϕ_1 or each other. To obtain the same information using the step function weights would require using all three ϕ_E, ϕ_U, ϕ_T norms. In general, the entire PRF is more information to

store than a set of weighted L^2 norms, which together provide a succinct and sufficient topological signature of the filtration.

5.2.2 Bifurcation Landscapes

We consider the ratios

$$\frac{||\overline{\beta_k}||_{\phi_B}}{||\overline{\beta_k}||_{\phi_1}}, \frac{||\overline{\beta_k}||_{\phi_D}}{||\overline{\beta_k}||_{\phi_1}} \tag{5.3}$$

over the parameter space and call contour plots of such **bifurcation landscapes**. See Figure 5.7 for an example. These plots reveal regions of the parameter space separated by the level curves equal to 1 where distinct types of topology are captured by the filtrations.

When $\frac{||\overline{\beta_k}||_{\phi_B}}{||\overline{\beta_k}||_{\phi_1}} < 1$, we expect late born homology reflected in a plethora of features in the upper right corner of a persistence diagram, or high values in the corresponding region of the PRF. To better model the features in the data, then, we should consider local pairs in the parameter space that increase this ratio. When $\frac{||\overline{\beta_k}||_{\phi_B}}{||\overline{\beta_k}||_{\phi_1}} > 1$, we expect an abundance of features in the lower left corner of a persistence diagram, or high values in that region of the PRF. This can be suggestive of a either a filtration that is capturing many short-lived features early, thus representing the local topology, with the global topology reflected in the longer-lived features, or a filtration that captures the global topology early and no homology for the rest of the filtration. We expect similar patterns in the topology represented by filtrations with opposite inequalities for ratios involving ϕ_D.

> **Conjecture: The region of the bifurcation landscape with value equal to 1 corresponds to parameters to construct filtrations of witness complexes revealing early-born, long-lived global topological structure with minimal local topological structure.**

5.2.3 Global Topological Heuristic

The bifurcation landscapes comparing ϕ_B, ϕ_D, ϕ_1- weighted L^2 norms are functions on 2D subsets of the 3D parameter space

for a fixed $k_{\max}$ established by Heuristic 1. As the ratios between these parameters vary, the topology captured by filtrations of witness complexes built with those sets of parameters changes dramatically. When $|L| \approx |\Gamma|$ and $|L| >> k_{\max}$, local topology can overwhelm the global topology in the persistent homology of the filtration. When $k_{\max} \approx |L|$, global topology can be short-lived. It is important to be above a minimum $|\Gamma|$ and $|L|$ such that witness complexes are able to highlight this global topology across the filtration. In particular, it takes a minimum amount of data to well-approximate the delay reconstructed attractor of the dynamics and a minimum number of landmarks $|L|$ to construct n-dimensional homological features. Since the $|L|$ contributes more to computational demands than the $|\Gamma|$, we propose the following heuristic for choosing $|L|$ and $|\Gamma|$:

Heuristic 5. *Minimize a fitness function* $h(|L|, |\Gamma|) = |L|$ *such that* $\left(\frac{||\overline{\beta_k}||_{\phi_D}}{||\overline{\beta_k}||_{\phi_1}}, \frac{||\overline{\beta_1}||_{\phi_B}}{||\overline{\beta_k}||_{\phi_1}}\right) \approx (1,1)$, **subject to constraints.*

When no homology exists in the filtration, we trivially have $\left(\frac{||\overline{\beta_k}||_{\phi_D}}{||\overline{\beta_k}||_{\phi_1}}, \frac{||\overline{\beta_1}||_{\phi_B}}{||\overline{\beta_k}||_{\phi_1}}\right) = (1,1)$. To address this, we start our algorithm by fixing $|\Gamma| = |L|_{\max}$ where $|L|_{\max}$ is determined by computational limitations. We initialize $|L| = k_{\max} + 1$ where $k_{\max}$ is chosen as per Heuristic 1 and increase $|L|$, tracking changes in the ordering between the ϕ_B, ϕ_D-weighted L^2 norms. We then increase $|\Gamma|$, and repeat the process of increasing $|L|$ while tracking the ordering of the set of ϕ_D, ϕ_B-weighted L^2 norms.

The locations where the order of the ϕ_B, ϕ_D-weighted L^2 norms switch in the parameter space provides the level curves equal to 1 in the bifurcation landscape. In our experiments, we observed a consistent $|L|$ where a first switch occurs across all $|\Gamma|$. This demarcates a region of the $|L| \times |\Gamma|$ parameter space below which $\phi_B > \phi_D$ and global topology is born early and dies early in the filtration. Above this, the pattern of switches between ϕ_B and ϕ_D as $|L|$ is increased varies depending on $|\Gamma|$.

For $|\Gamma| \approx |L|_{\max}$, increasing $|L|$ notably decreases the density of witnesses, causing both local and global homology to be born later, so ϕ_D stays above ϕ_B. When $|\Gamma| >> |L|_{\max}$, increasing $|L|$ increases the amount of local topology that can be captured early because of the greater density of

witnesses. So ϕ_B again becomes larger than ϕ_D, now because of the presence of early-born, short-lived homology. The first switch tells us we are observing the global topology for the majority of the filtration and second switch tells us we are starting to capture local topology early. Since we wish to minimize $|L|$, we should pick a pair $(|\Gamma|, |L|)$ that occurs at the first switch as soon as we encounter a second switch. This is the rationale of Heuristic 5, expressed in pseudo-code below. Sometimes when $|L| << |\Gamma|$, $\phi_B \geq \phi_D$ and no switches occur, as the data is never sparse enough for the complex to remain disconnected, yet a local minimum is encountered as Γ is increased. This can also signal the topological transition from early dying global topology to early dying local topology.

Such a location in the parameter space should suggest cheap parameters $|L|$ and $|\Gamma|$ for which a filtration of witness complexes captures a number of long-lived global homological features reflecting the underlying dynamics and little short-lived, local homology. The rescaling of the PRFs introduced in Chapter 4 then allows us to infer that the weighted L^2 norm of the PRF approximates the number of large scale homological features in the point cloud data, providing a coarse-grained topological signature for the dynamical system. We illustrate this through an example later in Chapter 5.

5.2.3.1 Pseudo-code for Heuristic

To select $(|\Gamma|, |L|)$:

for sliding window size $\in [|L|_{\max}, |\Gamma|_{\max}]$

  for number of landmarks $\in [k_{\max}, |L|_{\max}]$

    for i $\in [1,$ **sample size**$]$

      Set sliding windows $W_n = \{w_{(n-1)\cdot\textbf{sliding window size}}, \ldots, w_{n\cdot\textbf{sliding window size}-1}\}$

      Select landmarks $L = \{\ell_{i_0}, \ldots, \ell_{0_{\textbf{number of landmarks}}}\}$

      Compute persistent homology, and store the associated PRF f_n.

    Compute the mean PRF of $\{f_n\}_{n=1}^{s}$

    Compute the ϕ_D, ϕ_B, ϕ_1-weighted L^2 norms of the mean PRF.

if $\phi_B \leq \phi_D$

...(continue in **number of landmarks** loop)

else $\phi_D \leq \phi_B$

if counter $j = 0$

Store **switch1**$(|\Gamma|) = |L|$; increment a counter $j = +1$

... (continue in **number of landmarks** loop)

elseif counter $j = 1$

Store **switch2**$(|\Gamma|) = |L|$; increment a counter $j = +1$

... (continue in **number of landmarks** loop)

else counter $j > 1$

... (continue in **number of landmarks** loop)

if switch2$(|\Gamma|)$ - **switch1**$(|\Gamma|) <$ threshold$_{\text{something}}$

select $(|\Gamma|, |L|) = (|\Gamma|,$ **switch2**$(|\Gamma|))$

exit algorithm

else

... (continue in **sliding window size** loop)

5.3 Experimental Results

5.3.1 Musical Instrument Recordings

Figure 5.2 shows the standard L^2 norm of the mean PRF of one dimensional persistent homology for filtrations of witness complexes associated to 10 sliding windows over the $|\Gamma| \times |L|$ parameter space. We see a wide region in the lower left of the parameter space where $3 \leq \|\overline{\beta_1}\|_{\phi_1} \leq 6$. As we look right and up towards $(|\Gamma|, |L|) \approx (5000, 500)$ we see $\|\overline{\beta_1}\|_{\phi_1}$ continues to increase beyond 25. As we look left towards $(500, |L|)$, as well as down towards $(|\Gamma|, 10)$, $\|\overline{\beta_1}\|_{\phi_1}$ decreases below 3.

The stability of $\|\overline{\beta_1}\|_{\phi_1}$ with parameters $|\Gamma|$ and $|L|$ inside this large triangular region suggests

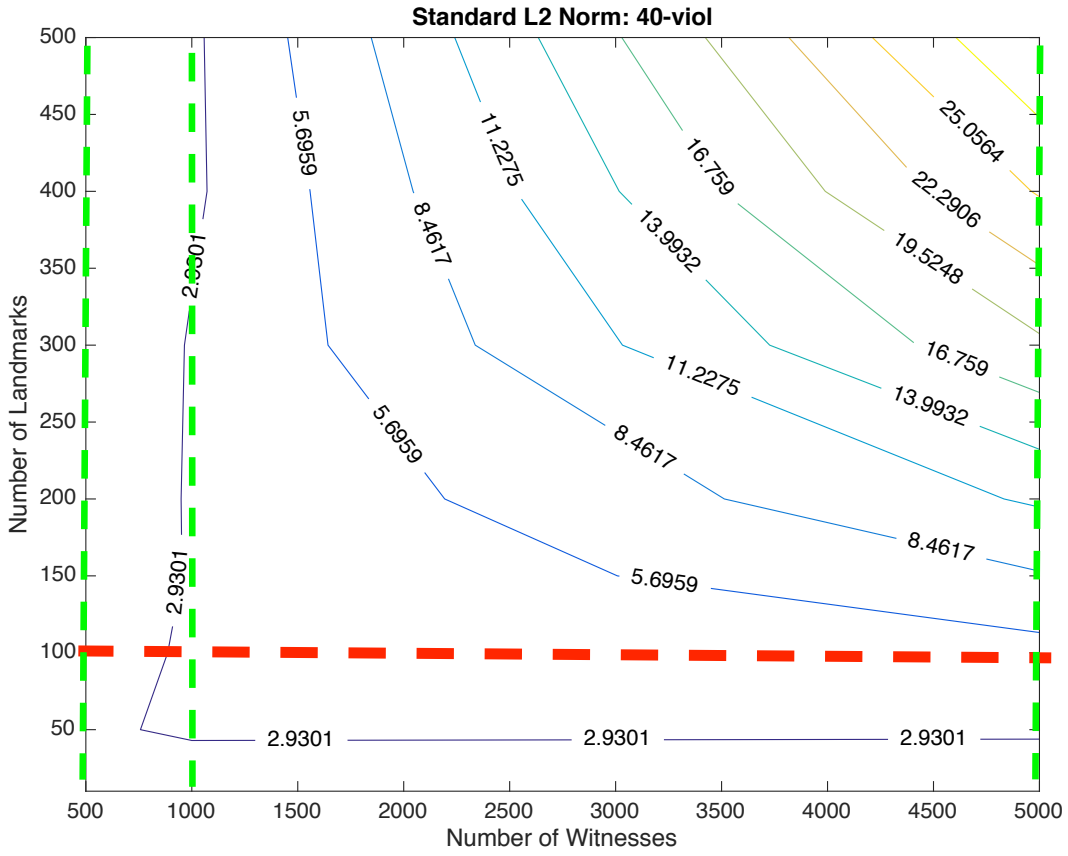


Figure 5.2: Unweighted L2 norm of the mean PRF of a set of 10 sliding windows of recordings of middle C (ʃ = 261.62 Hz) played on a viol recorded at 44.1 kHz sample rate using a Sony ICD-PX312 digital voice recorder modeled by filtrations of witness complexes terminated using $k_{\max} = 9$ over the (**number of witnesses**, **number of landmarks**) parameter space.

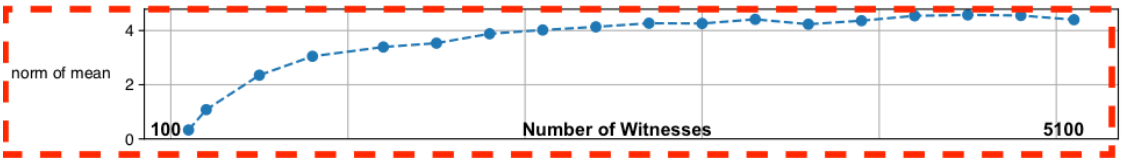


Figure 5.3: Stabilization is observed in the unweighted L^2 norm of the mean PRF of a set of filtrations of witness complexes with fixed **number of landmarks** equal to 100 and increasing the **number of witnesses** from 100 to 5000. This depicts the function value along the red dashed curve in Figure 5.2.

a sense of correctness to the persistent homology captured by these filtrations. The red dashed line through $|L| = 100$ cuts through this wide, and thus stable, region of unweighted L^2 norm between

about 2.9 and 5.7. We observe the stabilization of $\|\overline{\beta_1}\|_{\phi_1} \approx 4$ across this line as we increase $|\Gamma|$ from 100 to 5000 in Figure 5.3. We can relate this stable $\|\overline{\beta_1}\|_{\phi_1} \approx 4$ to the four global 1-dimensional homological features in the 2D delay reconstruction of the middle C viol data we see in Figure 3.10 of Chapter 3.

A priori we do not know what the correct topological signature of this data should be. Persistent homology reports homological features of varying birth times, death times and lifespans. These properties of persistent homology classes distinguish between types of topological features. The ratios $\frac{\|\overline{\beta_1}\|_{\phi_B}}{\|\overline{\beta_1}\|_{\phi_1}}$ and $\frac{\|\overline{\beta_1}\|_{\phi_D}}{\|\overline{\beta_1}\|_{\phi_1}}$ of Equation 5.3 can inform us as to what types of topological features appear in a filtration of witness complexes for a given ($|\Gamma|$, $|L|$). To obtain a coarse-grained topological signature for a musical instrument recording per Heuristic 5, we prioritize smaller $|L|$ where $\left(\frac{\|\overline{\beta_1}\|_{\phi_B}}{\|\overline{\beta_1}\|_{\phi_1}}, \frac{\|\overline{\beta_1}\|_{\phi_D}}{\|\overline{\beta_1}\|_{\phi_1}}\right)$ is close to $(1, 1)$. Recall that this should suggest a ($|\Gamma|$, $|L|$) so that the global topology is alive for most of the filtration and the local topology largely disregarded. In this situation, the unweighted L^2 norm of $\overline{\beta_1}$, our rescaled mean PRF, should reflect the number major homological features of the reconstructed dynamics.

Heuristic 5 essentially follows two loops - an inner loop increasing the **number of landmarks** and an outer loop increasing the **sliding window size**. The **sliding window size** affects the behavior of the weighted $\phi_B, \phi_D, \phi_1 - L^2$ norms as $|L|$ is increased. For fixed $|\Gamma| = 500, 1000$, and 5000, Figures 5.4 (a), (b), (c), respectively, show the weighted L^2 norms as $|L|$ is increased from $k_{\text{top}} + 1 = 10$ to $|L|_{\max} = 500$; this is the behavior of the weighted L^2 norms along the three green dashed lines in Figure 5.2.

When $|L| \approx k_{\max}$, interesting global topology may be reported as early-born, short-lived homological features. Low left endpoints of the curves in each Figure 5.4 (a),(b) and (c) reflect a low norm of the persistent homology rank function. For each $|\Gamma| = 500$, 1000, and 5000 witnesses, when $|L| = k_{\max} + 1 = 10, \phi_B > \phi_D$, as is seen in Figures 5.4 (a), (b), and (c), respectively.

Each of the ϕ_B, ϕ_D, ϕ_1-weighted L^2 norms peak around $|L| = 50$ and decline as $|L|$ is increased when $|\Gamma| = 500$, as seen in Figure 5.4 (a). When $|L| = 50$, the weighted L^2 norms are ordered $\phi_D > \phi_1 > \phi_B$. By $|L| = 200$, all weighted L^2 norms have dropped to below 1. The order $\phi_D > \phi_B$

stays the same as $|L| \geq 50$ for fixed $|\Gamma| = 500$. From this we can deduce that for $|\Gamma| = 500$ and $|L| \geq 50$ there is always homology born late in the filtration.

Once we increase to $|\Gamma| = 1000$, all ϕ_B, ϕ_D, ϕ_1-weighted L^2 norms exhibit greater stability when $|L|$ is increased beyond 50, as can be seen in Figure 5.4 (b). This tells us that a filtration of witness complexes built with $|L| \geq 200$ can have non-trivial persistent homology, yet more than $|\Gamma| = 500$ witnesses are necessary to detect this.

We see a dramatic shift in the behavior of the weighted L^2 norms as $|L|$ is increased when $|\Gamma| = 5000$ in Figure 5.4 (c). Now the ϕ_B-weighted norm is above the Standard L^2 norm, which is well above the ϕ_D-weighted norm. This suggests that there are a lot of early-born, short-lived features in the filtration capturing the local topology when $|\Gamma| = 5000$. In a sense, these are the late-born, late-death features exhibited for $|\Gamma| = 500$ and $|L| \geq 200$ in Figure 5.4 (a) that are now created and destroyed earlier due to a greater density of witnesses.

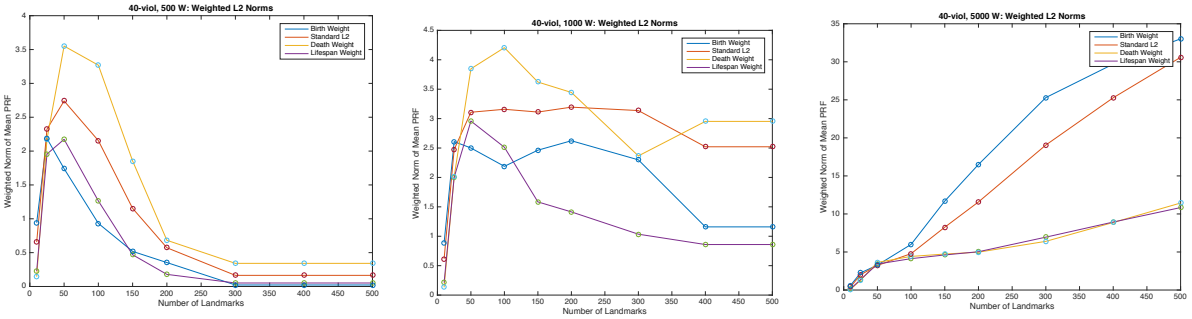


Figure 5.4: Weighted L^2 norms of the mean PRF of a set of filtrations of witness complexes modeling two-dimensional delay reconstructions of viol recordings for fixed **number of witnesses** as a function of increasing the **number of landmarks** from 10 to 500: (a) 500 witnesses (b) 1000 witnesses (c) 5000 witnesses and ϕ_1 (red), ϕ_B (blue), ϕ_D (yellow).

Figure 5.4 (a) suggests that **switch1**(500) of Heuristic 5 pseudo-code occurs between $|L| = 25$ and $|L| = 50$ and no **switch2**(500) occurs. According to Heuristic 2 pseudo-code, $|\Gamma|$ is increased and $|L|$ is reset to $k_{\max} + 1$. Figure 5.4 (b) shows that when $|\Gamma| = 1000$, **switch1** happens between $|L| = 25$ and $|L| = 50$ and **switch2** happens between $|L| = 200$ and $|L| = 300$. A large gap between **switch1**(1000) and **switch2**(1000) and a **switch3**(1000) between $|L| = 300$ and $|L| = 400$

encourage increasing $|\Gamma|$ further. Figure 5.4 (c) shows a much more rapid transition in persistent homology as **switch1**(5000) is between $|L| = 25$ and $|L| = 50$ and **switch2**(5000) happens between $|L| = 50$ and $|L| = 100$.

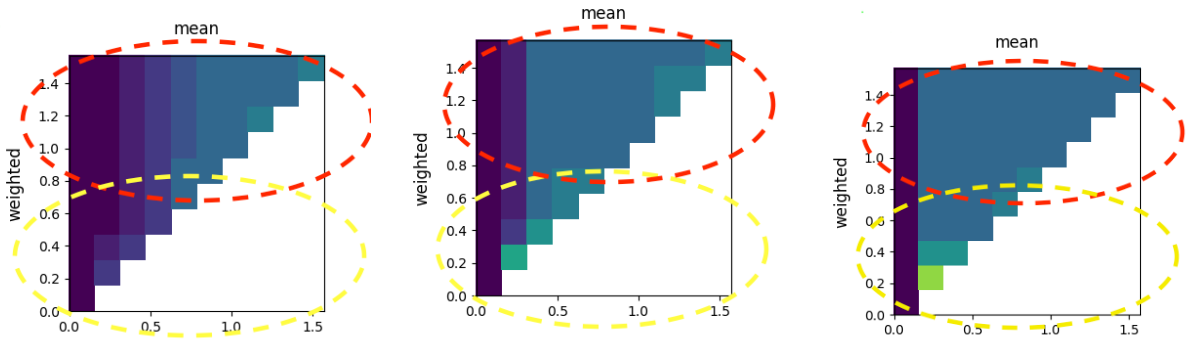


Figure 5.5: The mean PRFs of a set of filtrations of witness complexes modeling two-dimensional delay reconstructions of viol recordings terminated using $k_{\max} = 9$ using a fixed **number of landmarks** equal to 100 and **number of witnesses** equal to (a) 500 (b) 1000 (c) 5000. We circle regions of the PRF emphasized (red) and de-emphasized (yellow) by the ϕ_D-weighted L^2 norm.

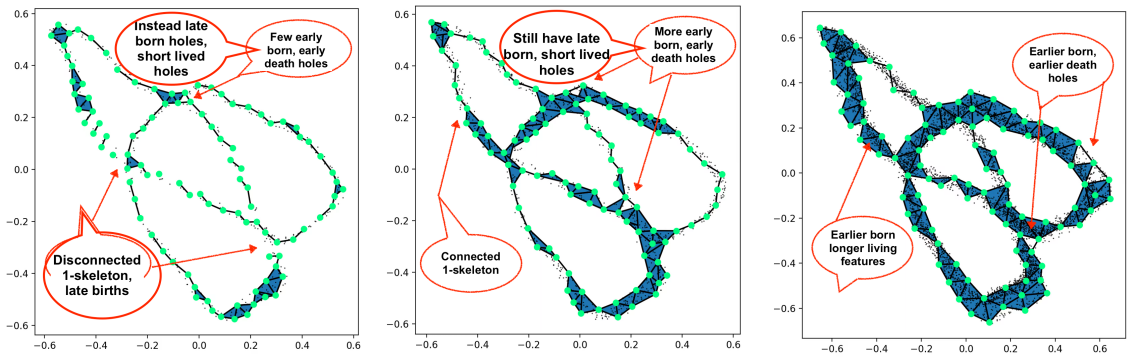


Figure 5.6: We can relate the features in the mean PRF of filtrations of witness complexes modeling two-dimensional delay reconstructions of viol recordings terminated using $k_{\max} = 9$ for fixe **number of landmarks** equal to 100 and fixed **number of witnesses** equal to (a) 500 (b) 1000 (c) 5000 back to the global and local topology of the witness complexes.

These analyses can be verified by the mean PRFs $\overline{\beta_1}$ for $|\Gamma|$ increasing from $500, 1000$, to 5000 and fixed $|L| = 100$ in Figure 5.5. We observe the anticipated trends in amount of persistent homology for given birth and death times reported by the ordering of the ϕ_D-weighted and ϕ_B-weighted

L^2 norms. Looking directly at 2D projections of the geometric realizations of witness complexes from each of these filtrations shown in Figure 5.6 confirms that the various short-lived homological features revealed in the mean PRFs of Figure 5.5 are indeed representing global topology captured late, local topology captured late and local topology captured early, respectively, for fixed $|L| = 100$ as we increase the $|\Gamma|$ from 500 to 1000 to 5000.

The mean PRF in Figure 5.5 (a) reflects that filtrations of witness complexes with $|\Gamma| = 500$ and $|L| = 100$ is only beginning to capture 1-dimensional homology towards the end. Looking at a projection of such a witness complex, we can observe in Figure 5.6 (a) that such complexes can in fact remain disconnected late into the filtration, failing to correctly capture the global topology.

Four major homological features are reported by the mean PRF in Figure 5.5 (b). While these features are alive at the end of the filtration but are not born during the earlier steps of the filtration. This suggest that there is still not a high enough density of witnesses among the landmarks to connect the simplicial complex near immediately. Additionally, short-lived homology is born and dies at all resolutions of the filtration.

Increasing the $|\Gamma|$ to 5000 mitigates both of these issues, connecting the complex sooner and concentrating this short-lived homology to the lower left region of the PRF representing early-born, early-death features, as is seen in Fig. 11(c). Figure 5.6 (c) demonstrates some of the local 1-dimensional homological features that are are born and die early in the filtration when $|L| = 100$ and $|\Gamma| = 5000$.

While visual analysis is greatly beneficial, producing graphics is an even more computationally intensive task that requires human interaction. So having an automated algorithm like Heuristic 5 for selecting ($|\Gamma|$, $|L|$) that interprets the topological features present in a filtration and sets ($|\Gamma|$, $|L|$) to capture maximal global topology and minimal local topology, subject to a cost function of the parameters, is very helpful.

The **bifurcation landscapes** introduced in this Chapter organize this topological information quite nicely, avoiding the in-depth analysis of the topology of witness complexes over the (**number of witnesses**, **number of landmarks**) parameter space presented in this section. Con-

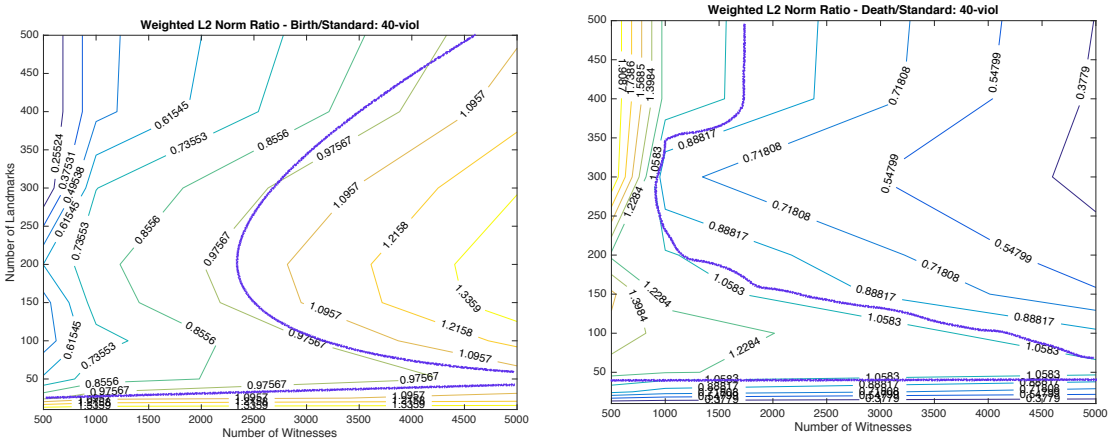


Figure 5.7: The ratios of the (a) ϕ_B (b) ϕ_D weighted L^2 norm over the standard ϕ_1-weighted L^2 norm of the mean PRF of filtrations of witness complexes modeling two-dimensional delay reconstructions of viol recordings terminated using $k_{\max} = 9$ give the (a) birth weighted and (b) death weighted bifurcation landscapes over the (**number of witnesses, number of landmarks**) parameter space, revealing the types of topological features captured by filtrations of witness complexes.

sider the bifurcation landscape for the ϕ_B-weighted in Figure 5.7. The horizontal level curve equal to 1 separates regions of the parameter space where global topology dies too early. Below the curve, there are too few landmarks. Above the curve, a "C-shaped" level curve equal to 1 further separates a right and left region of the parameter space where local topology is captured late (left of the "C", **number of witnesses** $\approx$ **number of landmarks**) and local topology is captured early (right of the "C", **number of witnesses** $\gg$ **number of landmarks**). The top left corner of this **bifurcation landscape** also shows a rapid transition in reported topology when the **number of landmarks** increases from approximately $\frac{1}{2}\cdot$ **number of witnesses** to the **number of witnesses**, as the witness complex approaches the Čech complex. According to Heuristic 5, we select (**number of witnesses**, **number of landmarks**) $\approx (4000, 100)$.

5.3.2 Chaotic Dynamical Systems

To determine the pair of parameter values (**number of witnesses, number of landmarks**) according to Heuristic 5, we first identify the level curve equal to 1 in the contour plot of the ratios of weighted L^2 norm over the standard L^2 norm across the **number of witnesses** by **number of landmarks** parameter space, rescaling the x-axis so that we are actually reporting the natural logarithm of the **number of witnesses**. In the birth weighted bifurcation landscape for Lorenz data shown in Figure 5.8 (a), this level curve is a single component that is roughly a line spanning from $(500\ (\approx e^{6.2}), 50)$ to $(2000\ (\approx e^{7.6}), 400)$.

Following the pseudo-code, we begin by looking at the smallest **number of witnesses** and **number of landmarks**. We then increase the **number of landmarks** and report any switches in the order between the birth weighted L^2 norm and the standard L^2 norm. Here at **number of witnesses** $= 500\ (\approx e^{6.2})$, a single switch occurs at **number of landmarks** $= 50$. Next we increase the **number of witnesses** and then repeat increasing the **number of landmarks**. As we do this, we continue to only identify this single level curve equal to 1 and our **switch1** continues to increase. Once we increase to more than **number of witnesses** $= 2000\ (\approx e^{7.6})$, increasing the **number of landmarks** no longer results in an intersection with the level curve equal to 1 and

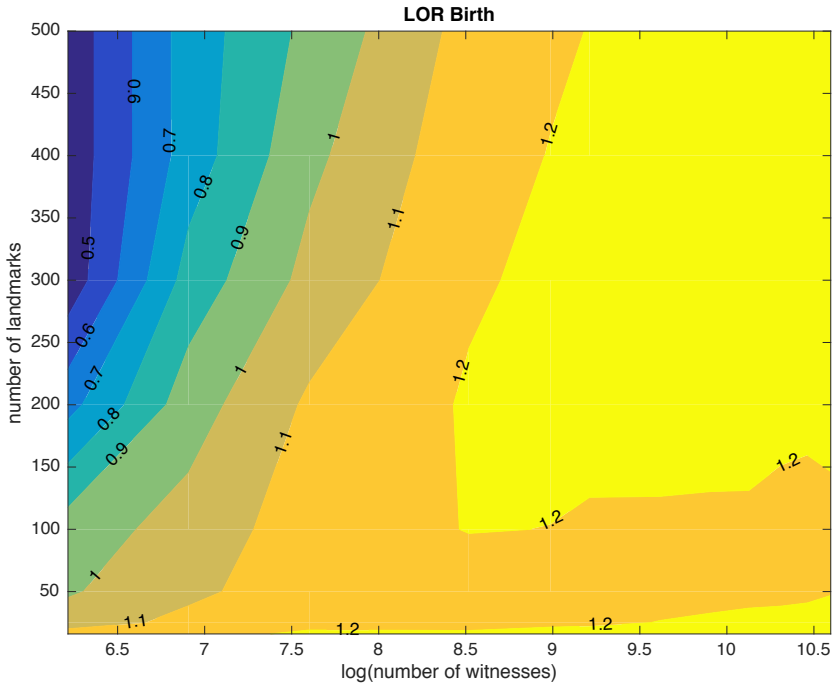


Figure 5.8: The birth weighted bifurcation landscape for mean PRF of filtrations of witness complexes modeling two-dimensional delay reconstructions of Lorenz shows the ratio of ϕ_B to ϕ_1 increases as the witness complexes move further away from the Čech complex in the (**number of witnesses**, **number of landmarks**) parameter space and experiences a local minima from approximately (5,000, 50) to (40,000, 100).

switch1 no longer occurs. However, for **number of witnesses** greater than 5000 ($\approx e^{8.5}$), there is still a local minimum occurring in the birth ratio between 16 and 200 landmarks. A modification of Heuristic 5 to select local minimum of the birth ratio with an increase in **number of landmarks** would select parameter values $(20,000\ (\approx e^{9.9}), 50)$, the smallest **number of witnesses** where this local minimum occurs with an increase in **number of landmarks** and the **number of landmarks** near the lower bound yet inside of this local minima region. These parameter values result in an $AUC \approx 1$ for the Lorenz classifier presented in Chapter 4. Figures 5.9 (a), (b) show witness complexes near the beginning and end of the filtration terminated using $k_{\max} = 15$ with $20,000$ **number of witnesses** and 50 **number of landmarks** to model a delay reconstructed sliding window sample of Lorenz time series. We only see one of the two major holes surrounding the

foci in the center of the butterfly wings in the attractor, albeit the two holes are of different sizes. However there is minimal topology captured in the outer less dense region of the butterfly wing with a single hole with short lifespan detected early in the filtration. This is reflected in the associated PRF shown in Figure 5.9 (c).

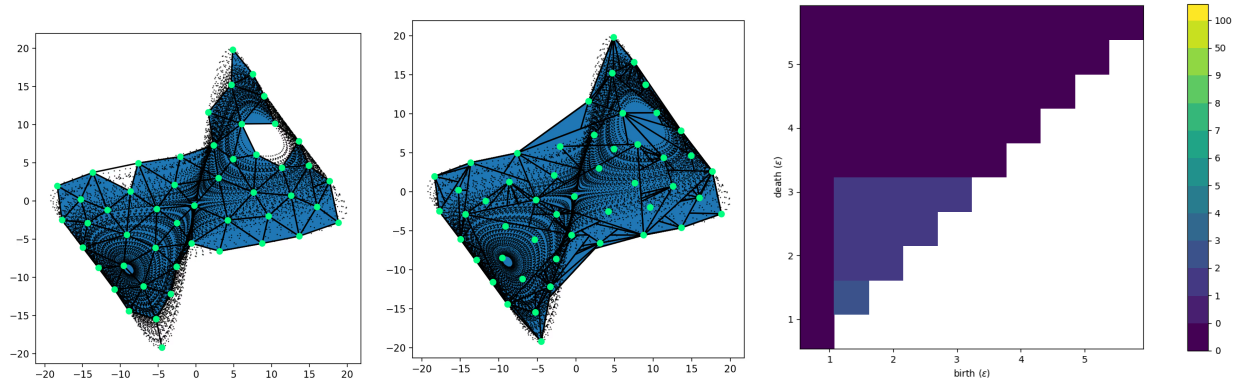


Figure 5.9: A witness complex toward the beginning of the filtration terminated using $k_{\max} = 15$ with **number of witnesses** equal to 20,000 and **number of landmarks** equal to 50 and the corresponding PRF.

The bifurcation landscape for the Röossler data reveals two components to the level set equal to 1. The first curve is roughly line spanning from $(500\,(\approx e^{6.2}), 70)$ to $(2000\,(\approx e^{7.7}), 400)$ in the log-linear parameter space. The second component is a C-shaped curve towards the bottom right of the bifurcation landscape spanning from $8000\,(\approx e^{9})$ to $40,000\,(\approx^{1} 0.6)$ and 25 to 110 **number of witnesses** and **number of landmarks**, respectively. This leads to a straightforward implementation of pseudo-code for Heuristic 5 using **switch1** and **switch2** which picks $(20,000\,(\approx e^{9.8}), 25)$. These parameter values result in an $AUC \approx 1$ for the Röossler classifier presented in Chapter 4. Figures 5.11(a) (b) show witness complexes near the beginning and end of the filtration for a sample sliding window of the Rössler time series terminated using $k_{\max}$ and **20,000** number of witnesses and 25 **number of landmarks**. The single major one-dimensional homological feature of the Rössler attractor is captured by 25 landmarks and no short-lived persistent homology is

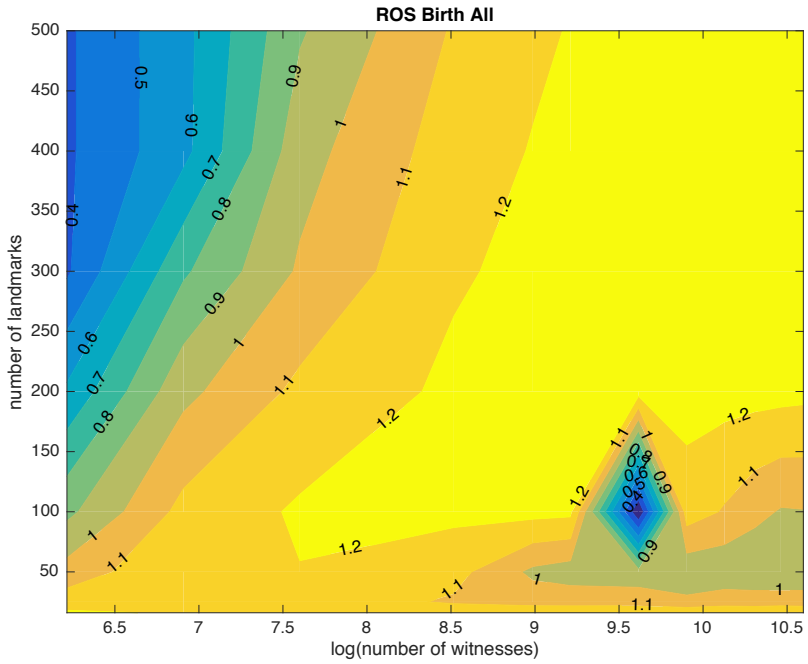


Figure 5.10: The birth weighted bifurcation landscape for mean PRF of filtrations of witness complexes modeling two-dimensional delay reconstructions of Rössler shows the ratio of ϕ_B to ϕ_1 increases as the witness complexes move further away from the Čech complex in the (**number of witnesses**, **number of landmarks**) parameter space and experiences a local minima

reported. However, the single feature does not live for the duration of the filtration and likely reflects our prudence when choosing the **number of landmarks**.

5.3.3 Brunel Neuronal Network

In contrast to the Lorenz and Rössler systems, the Brunel neuronal network is high dimensional when considering each neuron as a dimension of the system. Yet in the same way that Lorenz sought to explain weather patterns, an infinite-dimensional system, with a low-dimensional reduction, taking as a measurement function the average excitatory neuron firing rate seeks to describe the dynamics of the Brunel system with compressed information. Extracting some of this

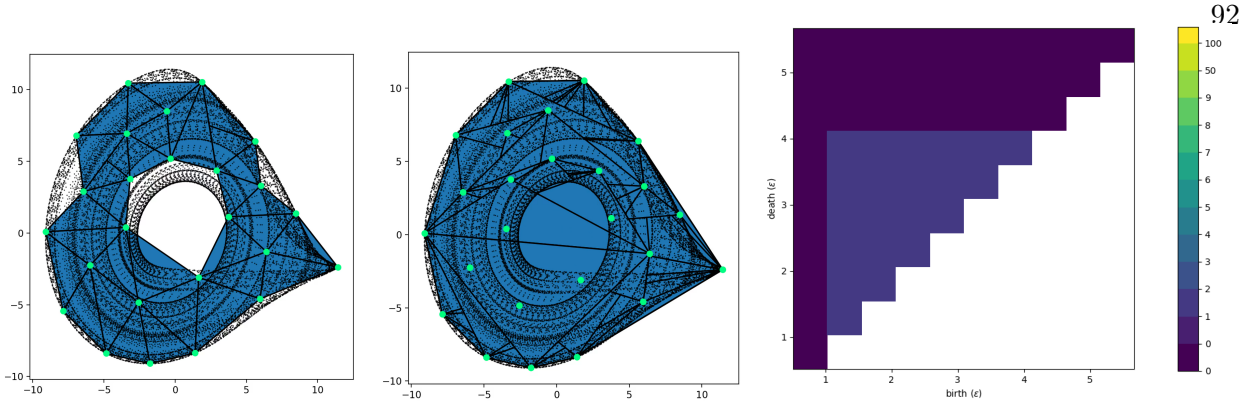


Figure 5.11: A witness complex toward the beginning of the filtration terminated using $k_{\max} = 15$ with **number of witnesses** equal to 20,000 and **number of landmarks** equal to 50 and the corresponding PRF.

information from the persistent homology of delay reconstructed sliding windows of this time series would be novel.

A straight forward implementation of Heuristic 2 for the I1G4p5 data set selects (500,100) with AUC(I2G3, I4G6,I2G5) = (0.52, 0.50, 0.66). Heuristic 5 picks for I2G5: (500, 100) - AUC(I2G3, I4G6, I1G4p5) = (0.80, 0.80, 0.32). Heuristic 5 picks for I4G6: (1000, 50) - AUC(I2G3, I1G4p5, I2G5) = (1.0, 1.0, 0.73). Heuristic 5 picks for I2G3: (500, 75).

Yet again our analyses of the Brunel neuronal network data do not reveal the expected patterns in the topological descriptors we are proposing. Here we expect the bifurcation landscapes to illuminate a lower bound on the **number of landmarks** and **number of witnesses** via a pair of level curves. The first of the level curves should emminate out radially from the location representing the Čech complex and separates a region where local topology is born late in the filtration, causing a ratio of the birth weighted L^2 norm to be less than the standard unweighted L^2 norm. The second of the level curves should be constant in the **number of landmarks** and parallel the curve **number of landmarks** $= k_{\text{top}} + 1$ where the topology changes from certainly killing off the relevant topology at the end of the filtration to elongating the lifespans of the meaningful topological features. This second level curve is dramatic in the bifurcation landscapes

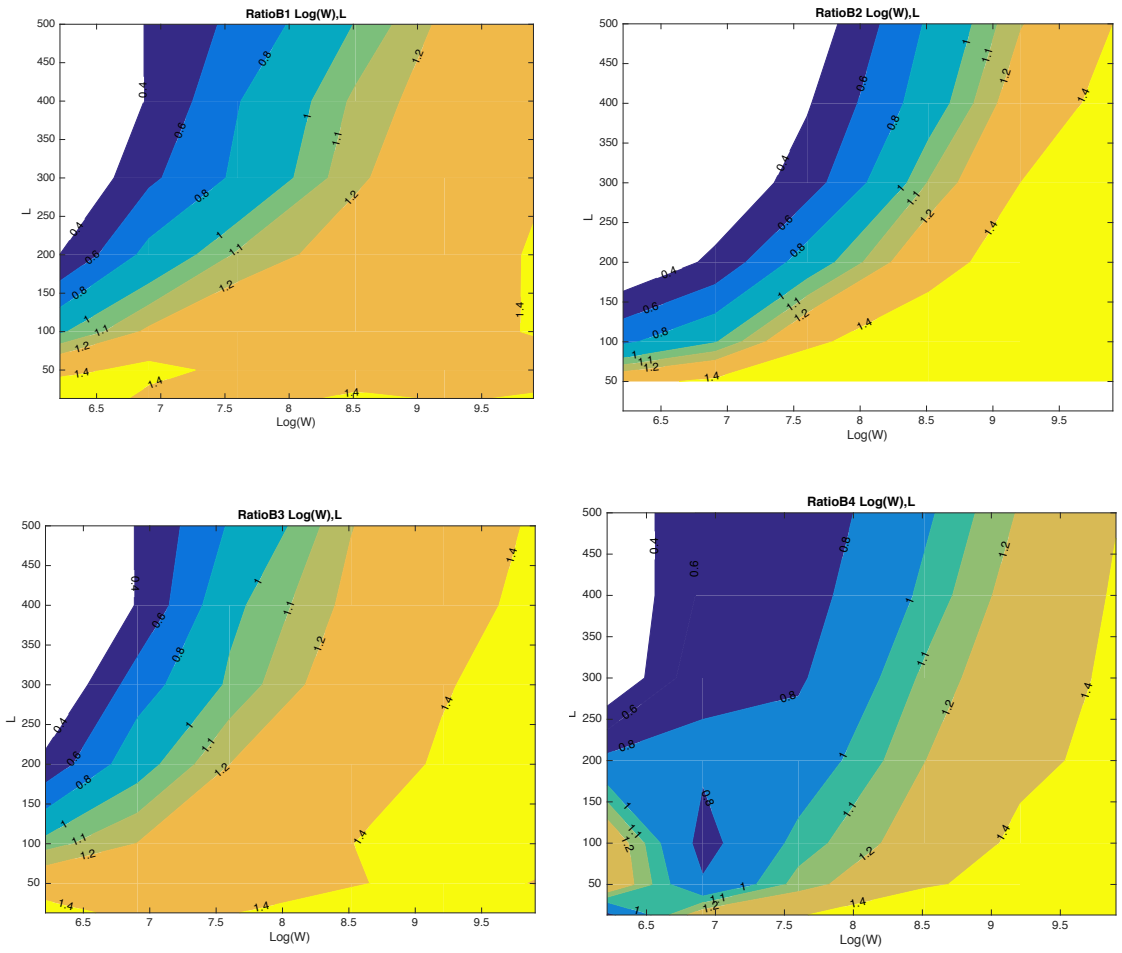


Figure 5.12: Birth weighted bifurcation landscapes for mean PRFs of filtrations of witness complexes modeling four distinct regimes of the Brunel neuronal network dynamics show the ratio of ϕ_B to ϕ_1 increases as the witness complexes move further away from the Čech complex in the (**number of witnesses**, **number of landmarks**) parameter space (a) I1G4p5 (b) I2G3 (c) I2G5 (d) I4G6.

for the periodic experimental musical instrument time series, is weaker yet still present in the form of local minima for chaotic time series, but is absent for time series from all four of the Brunel neuronal network dynamic regimes.

While the I2G3 regime is highly discontinuous and so we expect poor results, the remaining three data sets look like good candidates for a sliding window analysis using delay reconstruction as a technique to learn aspects of the dynamics. We set out to investigate what is going awry with our heuristic for determining a sufficient lower bound on the amount of data we need to

topologically model the dynamics. Models incorporating non-stationarity and additive noise to periodic signals visually align with the point clouds acquired by delay reconstructing Brunel time series data from the I4G6, I1G4p5 and I2G5 regimes as shown in Figure 3.9. Since the experimental music instrument recording data is likely to contain some noise, we attribute the poor results with the Brunel data to non-stationarity. To see the affects of non-stationarity on the point clouds and associated persistent homology, Figures 5.13(a), (b) depict the delay reconstruction of

$$X(t) = A(t) \cdot \cos(\theta(t)) + \eta(t)$$

where $A(t)$ and $\eta(t)$ are linearly increasing and the associated PRF for a filtration of witness complexes with **number of witnesses** $= 500$, **number of landmarks** $= 250$, and $k_{\text{top}} = 9$, respectively.

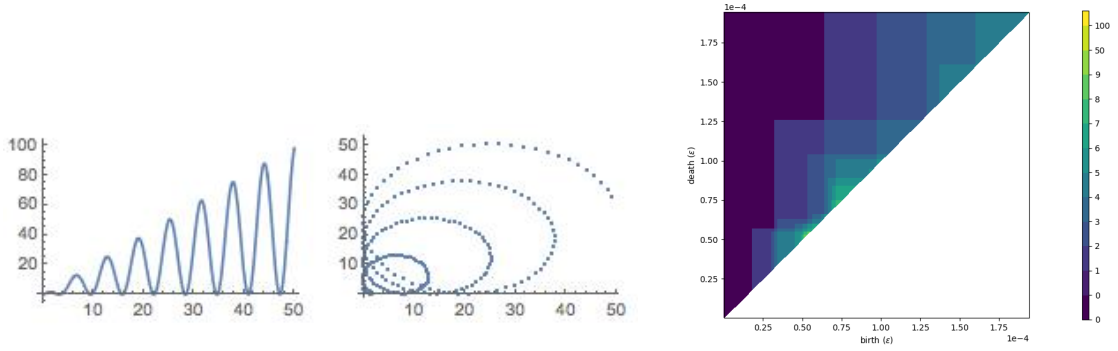


Figure 5.13: Non-stationarity in a time series can create and obscure homology in reconstructed point clouds when the **number of witnesses** is increased. Increases in the mean and variance result in the staircase pattern in the associated PRF of filtrations of witness complexes modeling two-dimensional reconstructions of such non-stationary systems.

We also look more closely at the delay reconstructions of the Brunel neuron model time series data. For small sliding windows from the I4G6 data, the topology of the delay reconstructed trajectory is that of a line. The curvature of the line and the reach of the specific delay reconstruction can lead to extra one-dimensional persistent homology, but for 500 witnesses the delay reconstruction generally has trivial one-dimensional topology. For about 1000 witnesses, the tra-

jectory appears to complete a rough elliptical orbit, suggesting the correct topology to detect is a single one-dimensional homological feature.

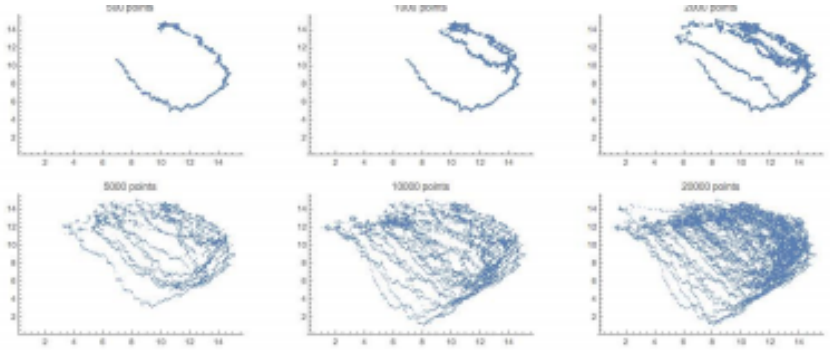


Figure 5.14: Two dimensional delay reconstructions of the average excitatory neuron voltage in the I4G6 regime of the Brunel neuronal network model reveal a similar point cloud pattern to reconstructions of non-stationary time series when the **number of witnesses** is increased.

Yet as the **sliding window size** is increased, subsequent elliptical periods begin to overlap, convoluting rather than reinforcing the topological signature. Not only are the ellipses moving, they are growing and shrinking. This corresponds to non-stationarity of both the mean and variance. Figure 5.15 shows the PRF of a sample of I4G6 data with 500 witnesses and 250 landmarks exhibiting a similar staircase pattern to the PRFs corresponding to this dual non-stationarity with the variation in the mean a result of additive Gaussian noise on a similar scale as the signal rather than an increasing linear function. In either of these cases, the persistent homology will not converge with an increase in **sliding window size**, making this data ill-suited for application of our heuristic.

Rather than just a linear function of time, changes in the mean and variance could be more complicated functions, like sinusoidal. This extra periodicity adds a dimension to the system that could be resolved topologically by increasing the dimension of the delay reconstruction. Such exploration could be future work. We turn now to improving the witness complex construction by incorporating the additional information available from time series data.

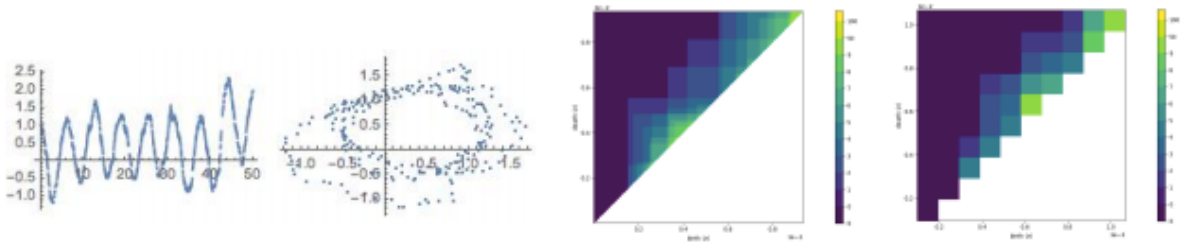


Figure 23: Model 5

(a) Model 5 (b) I4G6

Figure 5.15: The PRFs of filtrations of witness complexes modeling two dimensional delay reconstructions of the average excitatory neuron voltage in the I4G6 regime of the Brunel neuronal network model and a non-stationary time series reveal similar features.

Chapter 6

Smart Witness Complexes

The interaction between the geometry and topology of attractors in the state space of dynamical systems can lead to complications in using persistent homology as a topological signature of the dynamics. In the case that point clouds come from delay reconstruction of time series data, the delay parameter dramatically also affects the geometry of the reconstruction. When modeling delay reconstructed sliding window to extract a correct and consistent topological signature from delay reconstructed sliding windows of time series data, the standard witness complex is still not good enough.

Figure 6.1 shows the two dimensional delay reconstruction of the x-projection of a solution to the Lorenz system of equations for a range of delay parameter values. For small values of the delay parameter τ, such as $\tau = 6$ in Figure 6.1(a), the coordinates of the delay reconstructed phase space capture similar information and therefore the attractor remains close to the diagonal. When the delay parameter value is $\tau = 24$, shown in Figure 6.1(d), the reconstructed attractor begins to fold back on itself due to the randomness of taking a delay parameter τ too large relative to the time scale of the system. With delay values like $\tau = 12, 18$, shown in Figure 6.1 (b), (c), the attractor is well-expanded off of the diagonal and is not yet projected onto itself. Notable homological features of the Lorenz attractor, like the two holes in the butterfly wings, appear spherical and likely achieving maximal reach with the delay equa to the first minimum of time-delayed mutual information. However, there are challenges with the point cloud acquired at this suggested delay parameter. Because of the finite nature of our data, the invariant measure does not necessarily

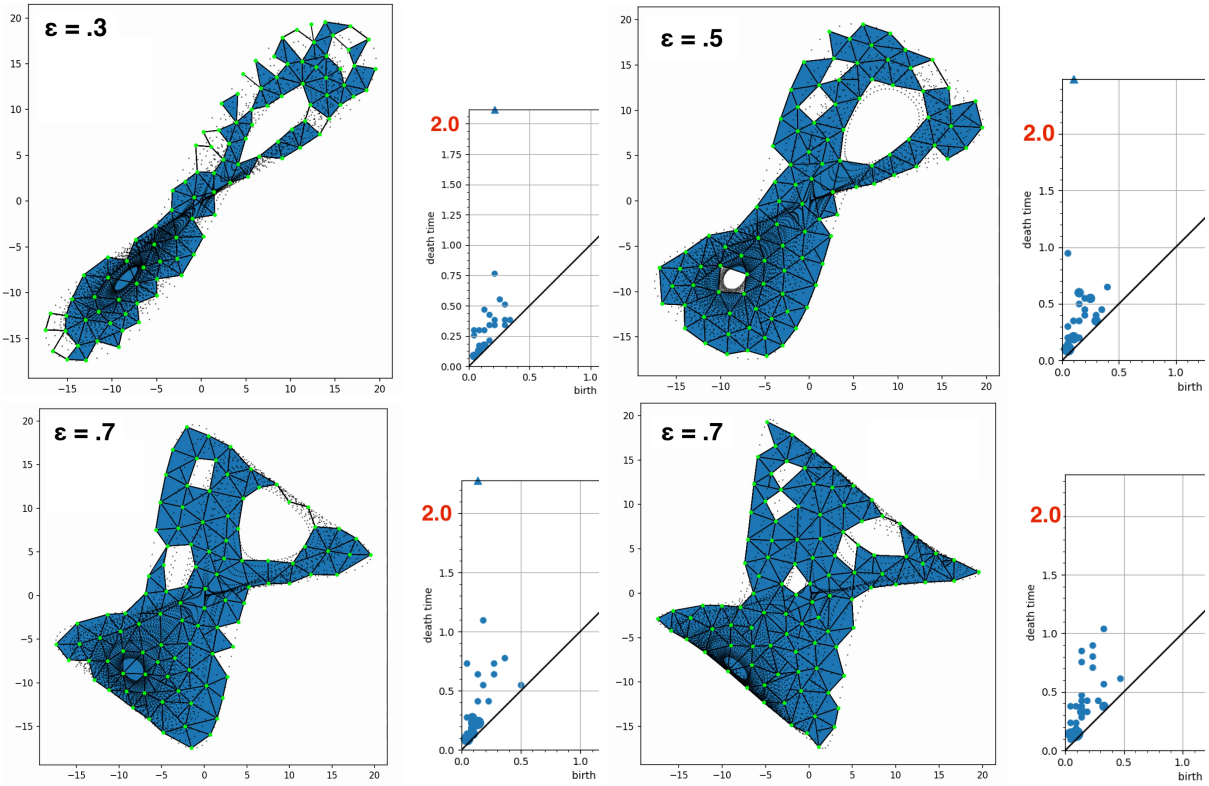


Figure 6.1: Standard witness complexes modeling two dimensional delay reconstructions of Lorenz time series for a range of delay parameters reveal a number of flaws (a) $\tau = 6$ (b) $\tau = 12$ (c) $\tau = 18$ (d) $\tau = 24$.

manifest in the sample density of the attractor. Moreover, the existence of an invariant measure itself suggests that even delay reconstructing long time series will result in nonuniform density of the sample along the attractor. This results in extra homology detected in low density regions, such as the holes in the outer Lorenz butterfly wings seen in all panels of Figure 6.1. At the same time that this supurfluous homology dies, the smaller of the two holes in the centers of the butterfly wings closes up, obscuring the signal and noise of the persistence diagram. Indeed, none of the standard witness complexes can capture the signature two holes associated to the foci in the center of the Lorenz butterfly wings.

The time ordering of data coming from sequential measurements of a dynamical system allows us to associate to each data point its speed and direction of travel in the delay reconstructed state space. Incorporating this information into a metric that increases the distances between points moving in opposite directions in the state space, we are able to produce a cheaper construction for topologically modeling dynamical systems point cloud data. This novel witness complex construction for time series analysis improves the topological signature we can detect using persistent homology and is also is less sensitive to choice of delay parameter in delay coordinate reconstruction.

6.1 Temporally-Informed Witness Complexes

We aim to close spurious holes due to low density parallel to the flow that arise due to changes in velocity and transverse to the flow due to an insufficient amount of data to achieve the invariant measure. We likewise wish to close holes due to low density that reflect the invariant measure itself. We simultaneously would like to keep small holes resulting from the presence of equilibria open longer. We first attempt to address this by incorporating approximations of the unit tangent vectors to the flow given by the normalized first difference vectors at points $\mathbf{x}$, $\mathbf{y}$ into the modified "distance" between these points:

$$\hat{\mathbf{x}} = \frac{\dot{\mathbf{x}}}{||\dot{\mathbf{x}}||}, \hat{\mathbf{y}} = \frac{\dot{\mathbf{y}}}{||\dot{\mathbf{y}}||}.$$

In some cases, this modification is positive-definite, symmetric and satisfies the triangle-inequality making it a metric.

One of those cases is when we consider adding some multiple of the distance between the normalized first difference vectors to the Euclidean distance. Suppose the data set $\Gamma = \{w_t\}_{t=1}^{N}$ with $w_t \in \mathbb{R}^d$ comes from a continuous dynamical system so that $\phi_s(w_t) = w_{t+1}$ where s is the sampling rate and ϕ is the flow. For $w_t \in \Gamma$, let

$$\dot{w}_t = w_{t+1} - w_t \text{ and } \hat{w} = \frac{\dot{w}}{||\dot{w}||}.$$

Consider the **temporally-informed family of metrics** for parameter $\mathbf{k_{distortion_+}}$ given by

$$d_{\text{distortion}_+}(w, \ell) = \sqrt{d_E(w, \ell)^2 + k_{\text{distortion}_+} \cdot d_E(\hat{w}, \hat{\ell})^2}.$$

This adds to the Euclidean distance between two points an amount roughly proportional to the dissimilarity in their tangent vector approximations. In particular, and unlike a multiplicatively proportional penalty such as

$$d_{\text{distortion}_\times} = d_E(x, y) \cdot (1 + k_{\text{distortion}_\times} \cdot d_E(\hat{x}, \hat{y})),$$

the uniform additive penalty allows $d_{\text{distortion}_+}$ to retain the property of still being a metric. Note that settings $k_{\text{distortion}_+} = 0$ results in the Euclidean distance, and we consider $k_{\text{distortion}_+} \geq 0$.

The **temporally-informed family of witness complexes** is given by the witness relation in Chapter 2 introduced in [45] using as the metric $d_{\text{distortion}_+}$. The standard witness relation using the Euclidean metric d_E is when we set $k_{\text{distortion}_+} = 0$. As $k_{\text{distortion}_+}$ surpasses the diameter of the data set, the geometric information provided by the reconstruction is obscured and we expect the persistent homology to approach that of an $(n-1)$-sphere, as the data points effectively become samples from an $(n-1)$-sphere. For values of $k_{\text{distortion}_+}$ in between, a range of topologies can arise. We explore this phenomena here now.

6.1.1 Modeling Chaotic Dynamics

For continuous time dynamical systems exhibiting chaos, the Lyapunov exponent bounds the divergence of nearby orbits, limiting the change in direction of travel of nearby points. The folding constant expressed by [cite: Boltt] bounds rotation and folding, similarly limiting change in direction of travel of nearby points. Thus points that appear nearby in the Euclidean distance yet far away in this modified distance are likely farther apart in the geodesic distance along the approximation of the attractor. Increasing the distances between points traveling in dissimilar directions also makes the distances between points moving in the same direction closer by comparison. This in some sense can start to retract the data onto a homology representative dual to the cohomology class of the attractor corresponding to the flow, collapsing holes due to fractal structure and acceleration.

Using the metrics $d_{\text{distortion}_+}$ to construct filtrations of temporally-informed witness complexes changes the order that simplices appear from those using standard witness complexes. We observe different patterns of simplices tiling these alternative models of the approximate attractor; more simplices appear across regions of parallel travel and less simplices across regions of opposite direction of travel. This has the affect of closing holes bounded by points moving in the same direction caused by low density of data while maintaining open holes around regions of the state space that are not part of the attractor.

Figure 2(a) shows the standard witness complex and Figure 2(b) shows the new witness complex using 5% of the data points as landmarks for a 500 point sample of a trajectory from the full 3D solution to the Lorenz 63 system solved with a time step of 0.01, at $\frac{\epsilon_{\max}}{2}$, halfway through the filtration discretized into $n = 10$ steps and terminated for $k_{\text{top}} = 10$. Already the standard witness complex includes 2-simplices covering the hole around the foci in the center of the left butterfly wing. Terminating the filtration earlier would not ameliorate this situation, as there exist simultaneously unwanted additional 1-cycles in the outer butterfly wing that do not reflect the dynamics. We can see that the new witness complex includes 2-simplices across the outer butterfly

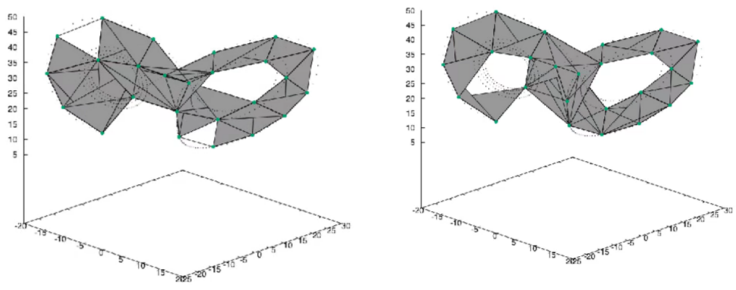


Figure 6.2: Witness complexes shown halfway through filtrations terminated using $k_{\max} = 10$ on 500 **number of witnesses** and 25 **number of landmarks** modeling a full three dimensional solution to the Lorenz system (a) standard witness complex (b) temporally-informed witness complex with $k_{\text{distortion}_+} \approx 20\%$ of the diameter.

wing, a region of the attractor less densely sampled due to too little data for the invariant measure to apply.

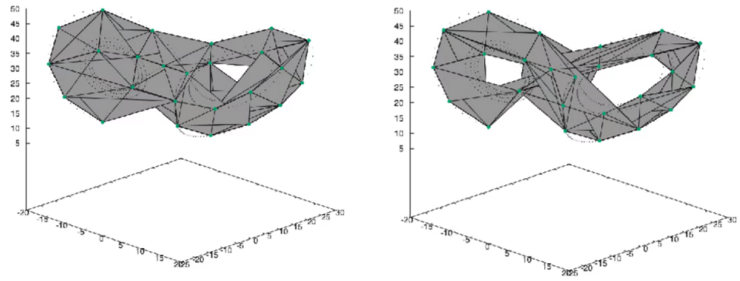


Figure 6.3: Witness complexes shown at the end of filtrations terminated using $k_{\max} = 10$ on 500 **number of witnesses** and 25 **number of landmarks** modeling a full 3 dimensional solution to the Lorenz system of equations (a) standard witness complex (b) temporally-informed witness complex with $k_{\text{distortion}_+} \approx 20\%$ of the diameter.

Figure 3(a) shows the standard witness complex and Figure 3(b) shows the new witness complex for a 500 point sample of a trajectory from the full 3D solution to the Lorenz 63 system solved with a time step of 0.01 at $\epsilon_{\max}$, the end of the same filtration shown halfway through in Figure 2. While the 1-cycle in the right butterfly wing of the standard witness complex reflecting the foci appears close to death, we can visually inspect that the new witness complex well maintains the two holes in the wings of the Lorenz butterfly that reflect the foci through the end of the filtration.

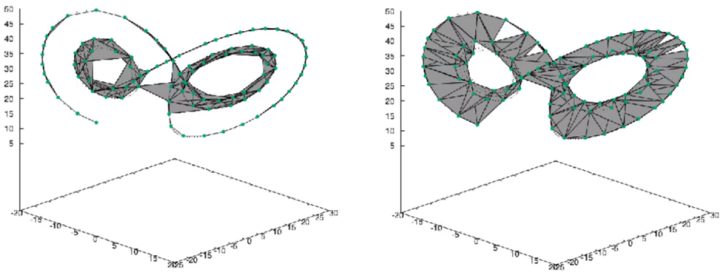


Figure 6.4: Witness complexes shown at the end of filtrations terminated using $k_{\max} = 10$ on 500 **number of witnesses** and 100 **number of landmarks** modeling a full 3D solution of Lorenz system (a) standard witness complex (b) temporally-informed witness complex with $k_{\text{distortion}_+} \approx 20\%$ of the diameter.

Figure 4(a) shows the standard witness complex and Figure 4(b) shows the new witness complex using 20% of data points as landmarks, for a 500 point sample of a trajectory from the full 3D solution to the Lorenz 63 system solved with a time step of 0.01 at $\epsilon_{\max}$. From this visual analysis alone, we see that the number of landmarks used in the model matters. This is even more true when considering computational costs, as well as correctness [45].

Many data sets are sequences of a single measurement function, and creating a point cloud representing the underlying dynamics measured by the time series in higher dimensions requires a implementing a methodology like Taken's delay coordinate reconstruction. When theoretical requirements are met, the point cloud should have the same homology as the attractor it approximates. Persistent homology, however, is sensitive to the geometry of the embedding of the attractor in ways that Taken's theorem and the variants since have not accounted for. For instance, data points can serve to witness landmarks that are close in the ambient space yet far in geodesic distance along the attractor, creating and destroying meaningful homology in the filtration of witness complexes. By increasing the distance between points traveling in different directions, the new metric $d_{\text{distortion}_+}$ can prevent the witnesses which are close in geodesic distance from witnessing landmarks that are only close due to the embedding in the ambient space. Using this new metric can expand regions of small reach in the delay reconstruction due to a choice of delay parameter that

has not fully unfolded the attractor. However, it does not work to address issues of false crossings in the delay coordinate reconstruction due to inappropriate choice of dimension parameter.

The witness complex built using this new metric $d_{\text{distortion}_+}$ to model a point clouds with time ordered data depends on the choice of parameter $k_{\text{distortion}_+}$. For the following experiments, we normalize the data sets to have diameter approximately 1.0 and explore $k_{\text{distortion}_+} = 0.01, 0.1, 0.5$ and 1.0, corresponding to a maximum additive penalty for points moving in opposite directions that is approximately $2\%, 20\%, 125\%$ and 300% of the diameter.

6.2 Cheaper Construction

The **temporally-informed witness complex** is designed to connect landmarks traveling in parallel directions, creating simplices transverse to the flow meanwhile excluding simplices between geometrically close landmarks traveling in opposite directions. Because of this, a correct topological signature can be detected from filtrations of witness complexes with less landmarks and witnesses than the standard witness relation. This greatly reduces the computations required. We report in Tables 6.1, 6.2 below the **number of witnesses**, **number of landmarks** and average number of two-simplices for a set of 10 PRFs for filtrations of temporally-informed witness complexes and standard witness complexes.

6.2.1 Number of 2-simplices

Tables 6.2, 6.1 show the average number of two-dimensional simplices in the final simplicial complex in a filtration of witness complexes built using the new witness relation with $k_{\text{distortion}_+}$ varying from 0% to 300% of the diameter of the data set for 10 sliding windows of from data set. Each column corresponds to a different pair of number of witnesses and number of landmarks upon which the model was built. Each row shows a different penalty $k_{\text{distortion}_+}$.

Table 6.1 reports this statistic for filtrations of witness complexes modeling two-dimensional delay reconstructions of the x-coordinate projection of a solution to the Rössler system of equations.

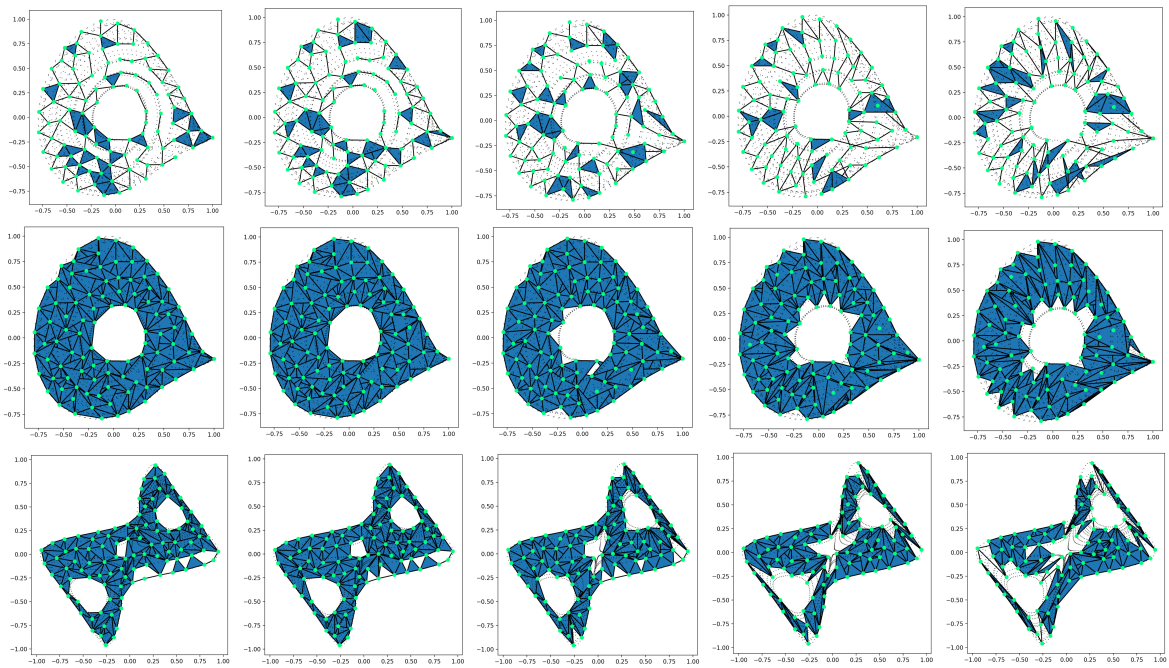


Figure 6.5: A panel of witness complexes shown (Row A) halfway through filtrations terminated using $k_{\max} = 10$ on 2000 **number of witnesses** and 100 **number of landmarks** modeling Rössler (Row B) the end of filtrations terminated using $k_{\max} = 10$ on 2000 **number of witnesses** and 100 landmarks modeling Rössler (Row C) the end of filtrations terminated using $k_{\max} = 10$ on 2000 **number of witnesses** and 100 **number of landmarks** modeling Lorenz using (Column A) the standard witness complex and (Column B) - (Column E) temporally-informed witness complexes with $k_{\text{distortion}_+} = 2\%, 20\%, 125\%, 200\%$, respectively.

$k_{\text{distortion}_+} \backslash (W, L)$	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	384	389	1114	2204
2%	396	403	1079	1595
20%	361	326	816	956
125%	364	385	754	1416
300%	386	365	707	1271

$k_{\text{distortion}_+} \backslash (W, L)$	(5000, 50)	(5000, 100)	(5000, 200)
0%	1449	2836	6119
2%	1451	2288	2490
20%	1118	1638	3114
125%	919	1840	3485
300%	847	1637	2741

Table 6.1: (2D Reconstructed Rössler) Number of 2-simplices at the end of the filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

We set $k_{\max} = 9$. This means that all witness complexes built with 10 landmarks terminate in the complete complex and have 120 triangles.

When the number of landmarks is increased to 20, the temporally-informed witness complexes do not demonstrate a noticeable change the number of two-simplices in the model at the end of the filtration. Yet by 50 landmarks, using the temporally-informed witness complexes for penalty $k_{\text{distortion}_+} \geq 20\%$ results in a reduction in the number of triangles for 2000 witnesses, and a 30% to 50% reduction for 5000 witnesses . When the number of landmarks is 100, we observe an almost 50% decrease in number of triangles at the end of the filtration for both 2000 and 5000 witnesses for penalty $k_{\text{distortion}_+} \geq 20\%$. A similar trend is observed for the two-dimensional delay reconstruction of the projection onto the x-coordinate of a solution of the Lorenz 63 system of equation, as seen in Table 6.2.

6.3 Enhancing Topological Discovery

While a reduction in cost is exciting, it is irrelevant if the resulting model is no good. We use the following set of statistics to assert an improved sense of correctness of the persistent homology captured for filtrations of witness complexes built using non-trivial $k_{\text{distortion}_+}$ when modeling two-dimensional delay reconstructions for Lorenz and Rössler.

$k_{\text{distortion}_+}$ \ (W, L)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	323	323	793	1688
2%	298	309	752	1462
20%	263	236	595	1168
125%	302	302	594	837
300%	302	297	504	801

Table 6.2: (2D Reconstructed Lorenz) Number of 2-simplices at the end of the filtrations terminated with $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud..

6.3.1 Unweighted L^2 Norms

Throughout this research we propose to use the weighted L^2 norm of the PRF representation of persistent homology as a topological signature for classification. We now look at how descriptive the unweighted version of this topological signature is with respect to feature identification of the underlying dynamical system.

Table 6.3 shows the mean of the unweighted L^2 norms of PRFs corresponding to a set of 10 two-dimensional delay reconstructions of time series obtained from projections of the Rössler chaotic dynamical systems, respectively, onto their x-axes. The signal reported by this norm for witness complex parameters **number of landmarks** equal to 10 and **number of witnesses** equal to 100 is encouraging. For all new parameters $k_{\text{distortion}_+}$ tested, including the control resulting in the standard witness complexes, the unweighted L^2 norm is approximately 1. Knowing **a priori** that the dynamics of the Rössler system are well-represented by a topological signature of 1, we are pleased. This holds for when increasing the **number of witnesses** to 1000. However, this signature changes when increasing the **number of landmarks** to 20. For both 1000 and 2000 **number of witnesses** and 20 **number of landmarks**, the unweighted L^2 norm jumps to over approximately 1.6 for the control standard witness complex. Using the **temporally-informed witness complex** construction decreases this unweighted L^2 norm to about 1.4 for a penalty $k_{\text{distortion}_+} \approx 20\%$ of the diameter of the data set and to about 1 for $k_{\text{distortion}_+} \approx 125\%, 300\%$, which is more reflective of the Rössler system. When the **number of landmarks** is increased to 50 and then 100 landmarks, the unweighted L^2 norm increases to about 10 and then 20 for the

$k_{\text{distortion}_+} \setminus$ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	0.98	0.89	2.51	1.60	8.72	17.80
2%	1.0	0.86	2.56	1.55	8.88	18.56
20%	1.04	0.98	1.86	1.36	8.50	21.32
125%	1.02	1.02	1.16	1.06	5.87	17.05
300%	1.02	1.05	1.05	1.06	3.98	14.10

$k_{\text{distortion}_+} \setminus$ (W,L)	(5000, 50)	(5000, 100)	(5000, 200)
0%	4.91	16.83	38.10
2%	4.76	18.25	44.95
20%	5.10	20.32	43.79
125%	3.41	16.04	36.27
300%	1.91	13.32	39.15

Table 6.3: (2D Reconstructed Rössler) Unweighted L^2 norm of PRFs for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

standard witness complex construction. These numbers are reduced by approximately 50% and 25%, respectively, when $k_{\text{distortion}_+} \approx 300\%$.

These trends are closely followed for the Lorenz data, as seen in Table 6.4. For 10 **number of landmarks** and 100 **number of witnesses**, the unweighted L^2 norm is approximately 0.65. This remains the same for the standard witness complex when increasing the **number of witnesses** to 1000, though increases by about 50% for $k_{\text{distortion}_+} \approx 125\%, 300\%$. None of these values reflect the Lorenz dynamics nor differentiate it from the Rössler system as seen through the lens of this statistic. Once the **number of landmarks** is increased to 20, the unweighted L^2 norm increases to about 2. This is a topological signature that we can relate to the Lorenz system of equations. Although there are three equilibria, the two foci in the center of the butterfly wings exhibit a certain symmetry that should lead to their simultaneous detection. They may also exhibit geometric properties like a larger radius that contribute to their detection over the hyperbolic equilibrium at the origin. However, the Rössler data also had an unweighted L^2 norm approximately 2 for these model parameters for the standard witness complex. Only when $k_{\text{distortion}_+} \approx 125\%, 300\%$ does the unweighted L^2 norm provide topological signatures approximately 1 and 2 for the Rössler and Lorenz systems, respectively. Increasing the **number of witnesses** to 2000 increases the unweighted L^2 norm when $k_{\text{distortion}_+} \approx 20\%, 125\%$ to over 2.7, much closer to three, the genus of

$k_{\text{distortion}_+} \setminus$ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	0.78	0.69	1.92	1.77	5.60	11.52
2%	0.76	0.81	2.07	1.85	5.83	11.30
20%	0.77	1.82	2.55	2.91	4.88	10.13
125%	0.88	1.49	2.30	3.07	4.14	8.60
300%	0.70	1.00	1.80	2.17	4.84	7.56

Table 6.4: (2D Reconstructed Lorenz) Unweighted L^2 norm of PRFs for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

the canonical tori we sought to detect as our topological signature of the Lorenz attractor. For 50 and 100 **number of landmarks**, the unweighted L^2 norm increases to about 6 and about 12, respectively, for the standard witness complex. When $k_{\text{distortion}_+} \approx 125\%, 300\%$, the unweighted L^2 is kept lower to about 4 and 8 for 50 and 100 **number of landmarks**, respectively.

6.3.2 Average Lifespan of Longest Living Homological Feature

The average lifespan of the longest living one-dimensional homological feature can serve as a measure of the strength of the signal in the persistent homology compared to the noise. We report this lifespan as the percentage of the filtration for which this predominant feature is alive. Using the **temporally-informed witness complex** in filtrations can increase the average lifespan by up to 30%.

In Table 6.5 for Rössler data, we see that for 100 **number of witnesses** and 10 **number of landmarks**, setting $k_{\text{distortion}_+} \approx 300\%$ increases the average lifespan of the longest-living one-dimensional homological feature in the filtration of **temporally-informed witness complexes** to 97% of the filtration from only 80% for the standard witness complex. When the **number of witnesses** is increased to 1000 this is an even more dramatic increase from 64% to 100%. The average lifespans of the longest living feature for the standard witness complex and for $k_{\text{distortion}_+} \approx 300\%$ for 20 landmarks and both 1000 and 2000 witnesses are also approximately 60% and 100%, respectively. Setting $k_{\text{distortion}_+} \approx 20\%, 125\%$ also results in a similar improvement. Once the **number of landmarks** is increased to 50 or 100, all witness complexes capture the longest-living

$k_{\text{distortion}_+} \setminus$ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	80%	64%	65%	67%	100%	100%
2%	83%	56%	74%	72%	100%	100%
20%	91%	85%	99%	100%	100%	96%
125%	94%	93%	100%	100%	100%	97%
300%	97%	100%	100%	100%	100%	96%

$k_{\text{distortion}_+} \setminus$ (W,L)	(5000, 50)	(5000, 100)	(5000, 200)
0%	100%	100%	100%
2%	100%	100%	100%
20%	100%	100%	100%
125%	100 %	100%	100%
300%	100%	100%	98%

Table 6.5: (2D Reconstructed Rössler) Percentage of the filtration for which the longest living one-dimensional homological feature lives for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

homological feature for approximately the entirety of the filtration.

Table 6.6 shows that the average lifespan of the longest living feature for Lorenz data is not increased by using the **temporally-informed witness complex** when there are only 100 witness and 10 landmarks. For 1000 witness and 10 landmarks, we observe over a 50% increase in the average lifespan using $k_{\text{distortion}_+} \approx 20\%, 125\%, 300\%$. For these same $k_{\text{distortion}_+}$, we observe about a 20% increase in the average lifespan when the data is modeled with 1000 and 2000 witnesses and 20 landmarks. Again when the **number of landmarks** is 50 or 100, all witness complex constructions maintain the longest living homological feature for nearly the whole filtration.

6.3.3 Signal From 2-Cluster Count

We perform hierarchical 2-clustering on the points in the persistence diagram and compute the number of points in the cluster containing the longest-living one-dimensional homological feature. This clustering takes places on the points in the persistence diagram under the Euclidean metric. Specifically, we implement a single-linkage criteria that creates two clusters separated by the longest edge in the MST. To distinguish signal from noise, we expect the maximal lifespan feature to be a member of a smaller cluster representing the signal and consider the remainder of the data as noise in the other cluster. Instead of reporting statistics like the mean which is sensitive to outliers,

$k_{\text{distortion}_+} \setminus$ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	72%	41%	81%	74%	95%	95%
2%	70%	50%	84%	84%	98%	100%
20%	70%	64%	94%	98%	100%	96%
125%	82%	75%	99%	100%	99%	91%
300%	66%	74%	99%	100%	98%	84%

Table 6.6: (2D Reconstructed Lorenz) Percentage of the filtration for which the longest living one-dimensional homological feature lives for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

we report the mode number of points in the cluster containing the maximal lifespan feature as topological signature on which to base correctness of the representation. Instead of variance, we report the frequency that the mode occurs as the number of points in the cluster containing the maximal lifespan feature as measures of consistency. As a finer measure of the consistency of the signal, we report the entropy of the frequency of the number of points in the cluster containing the maximal lifespan feature. This set of discrete statistics on the persistence diagram provide an example of the features TDA can produce for machine learning techniques, like convolutional neural nets, from time series data.

When the signal and noise are well-separated in the persistence diagram, the maximal lifespan feature should be in a cluster with other highly persistent one-dimensional features in the filtration. The number of points in the cluster with the maximal lifespan feature thus serves as a representation of the global topology of the attractor.

6.3.3.1 Persistence Diagrams 2-Clustering: Average Distance Between Clusters

Separating signal and noise is often a first step in preprocessing data. Considering the two clusters resulting from the hierarchical clustering of the points in the persistence diagram to represent the signal and noise captured by the persistent homology of the filtration, we can compute the distance between the two clusters as a measure of how well-separated the signal is from the noise. Note that this distance is the edge length of the longest edge in the minimal spanning tree, which is orthogonal to the maximally separating hyperplane found using a support vector machine.

Here we scale PRFs to have corner coordinate (0,10) with immortal cycles born at ϵ_B represented by points at (ϵ_B, 11).

$k_{\text{distortion}_+}$ \ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	1.27	0.90	4.30	5.40	7.13	5.11
2%	1.37	1.0	5.40	6.00	7.30	4.43
20%	0.72	1.2	8.41	8.10	6.80	3.81
125%	0.32	1.10	5.01	0.90	6.90	4.63
300%	0.0	0.0	0.72	0.09	6.91	4.53

$k_{\text{distortion}_+}$ \ (W,L)	(5000, 50)	(5000, 100)	(5000, 200)
0%	7.91	6.61	4.64
2%	8.00	6.50	3.52
20%	8.10	5.43	3.32
125%	7.63	6.25	3.53
300%	7.01	6.00	2.88

Table 6.7: (2D Reconstructed Rössler) Distance between 2-clusters with re-scaled PRFs with $\hat{\epsilon} = 10$ for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

$k_{\text{distortion}_+}$ \ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	0.50	1.20	4.10	4.32	4.55	3.06
2%	0.50	1.70	4.00	4.56	4.46	4.33
20%	0.32	2.80	6.82	4.38	6.73	4.27
125%	0.76	5.02	3.79	3.63	6.52	3.92
300%	0.00	4.50	5.57	5.50	5.48	3.83

Table 6.8: (2D Reconstructed Lorenz) Distance between 2-clusters with re-scaled PRFs with $\hat{\epsilon} = 10$ for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

For the Rössler data shown in Table 6.7, when the **number of landmarks** equals 10, there is little to no improvement in separating the signal from the noise from using the temporally-informed witness complex instead of the standard witness complex. For such few landmarks, there is often only a single homological feature detected in which case the distance from the non-trivial cluster to the trivial cluster is counted as zero. When the **number of landmarks** equal to 20, the temporally-informed witness complex using $k_{\text{distortion}_+} = 20\%$ nearly doubles the distance between the two clusters, a 100% increase in separation between signal and noise. When the **number of landmarks** is greater than 20 there is again little improvement.

The Lorenz data shown in Table 6.8 reveals similar trends with even better results for **number of landmarks** greater than 20. We observe approximately 20% increases in separation between signal and noise using 50 and 100 landmarks using the temporally-informed witness complexes with $k_{\text{distortion}_+} = 20\%$.

6.3.3.2 Persistence Diagrams 2-Cluster Count: Mode

We expect the number of points the cluster with the maximal lifespan feature that occurs the most, i.e. the mode, to be the most accurate topological representation of the dynamical system. We report this mode for the Rössler, Lorenz and periodically driven pendulum in the tables below.

$k_{\text{distortion}_+} \setminus$ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	1	1	1	1	1	1
2%	1	1	1	1	1	1
20%	1	1	1	1	1	1
125%	1	1	1	1	1	1
300%	1	1	1	1	1	1

$k_{\text{distortion}_+} \setminus$ (W,L)	(5000, 50)	(5000, 100)	(5000, 200)
0%	1	1	1
2%	1	1	1
20%	1	1	1
125%	1	1	1
300%	1	1	1

Table 6.9: (2D Reconstructed Rössler) Mode number of homological features in the 2-cluster containing the longest living homological feature for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

As seen in Table 6.9, the discrete number of points in the hierarchical 2-cluster containing the longest living homological well captures a topological signature of one that we associate to the Rössler system across all **number of landmarks**, **number of witnesses**, and $k_{\text{distortion}_+}$ values. Indeed, this signature arises more than any other value. However, Table 6.10 reveals that for **number of landmarks** ≤ 20 and **number of witnesses** ≤ 2000 the standard witness complex also captures a topological signature of one. When the **number of landmarks** is increased to 50, this topological signature equals two, differentiating the Lorenz system from the Rössler. This topological signature reflects the two main holes in the wings of the Lorenz butterfly. When the

$k_{\text{distortion}_+} \setminus$ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	1	1	1	1	2	3
2%	1	1	1	1	3	3
20%	1	3	3	3	3	3
125%	1	1	1	6	4	3
300%	1	1	1	1	6	27

Table 6.10: (2D Reconstructed Lorenz) Mode number of homological features in the 2-cluster containing the longest living homological feature for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

number of landmarks is further increased to 100, the mode count of homological features in the 2-cluster with the longest living homological feature equals three. This still differentiates the Lorenz system from the Rössler, possibly highlighting the three of the singularities of the Lorenz system of equations. Without visual confirmation, we cannot know whether this third feature actually corresponds to the geometrically small gap in the data near the origin at which the lone hyperbolic saddle node resides or is detecting a particularly low-density region in the outside of the butterfly wings.

The **temporally-informed witness complex** is intended to close the latter type of holes while encouraging the detection of the former. Using $k_{\text{distortion}+} \approx 2\%$, we see the nearly same topological signatures arise as we increase the **number of witnesses** and **number of landmarks** as we did for the standard witness complex. The exception occurs for (**number of witnesses**, **number of landmarks**) $= (2000, 50)$. Rather than a topological signature of two, we immediately detect a signature of three. The same topological signature holds when we increase the **number of landmarks** to 100. This is suggestive that for both $(2000, 50)$ and $(2000, 100)$, this topological signature of three actually reflects the three singularities of the Lorenz system and is a consistent signature given a large enough **number of witnesses** and **number of landmarks**.

When $k_{\text{distortion}_+} \approx 20\%$, this topological signature of three appears for much smaller **number of witnesses** and **number of landmarks** and remains consistent as these parameters are increased. Not only are we able to capture the same interpretable topological signature with half as many witnesses and 10% of the number of landmarks necessary to achieve the same results with

the standard witness complex, the consistency of the topological signature across model parameters like **number of witnesses** and **number of landmarks** suggests a sense of model robustness for the **temporally-informed witness complex** using $k_{\text{distortion}_+} \approx 20\%$.

For $k_{\text{distortion}_+} \approx 125\%, 300\%$ and **number of witnesses** ≤ 1000, we get a topological signature of one and are again unable to distinguish between Lorenz and Rössler. Increasing the **number of landmarks** and **number of witnesses** allows for differentiation between the chaotic dynamical systems, yet some of the topological signatures reported - 6, 4, 27 - are less relatable to the coarse-grained topology of the Lorenz attractor. This suggests that these values are outside of the range of $k_{\text{distortion}_+}$ appropriate for constructing **temporally-informed witness complexes** that reflect the coarse-grained topology of the attractor.

6.3.3.3 Persistence Diagrams 2-Cluster Count: Frequency

Just because something occurs more than any other thing does not mean that it happens very often. To describe the unpredictability of the topological signature given by the mode number of persistence intervals in the 2-cluster containing the maximal lifespan feature, we report the frequency with which this mode cluster size occurs. We see that the temporally-informed witness relation generally improves the frequency with which this mode occurs. If the mode is indeed a good candidate for representing the topology of the dynamics, this is excellent.

$k_{\text{distortion}_+} \setminus$ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	70%	50%	100%	80%	100%	100%
2%	70%	50%	100%	90%	100%	100%
20%	80%	70%	90%	70%	100%	70%
125%	90%	100%	80%	90%	100%	60%
300%	100%	100%	90%	90%	90%	50%

$k_{\text{distortion}_+} \setminus$ (W,L)	(5000, 50)	(5000, 100)	(5000, 200)
0%	100%	100%	90%
2%	100%	100%	70%
20%	100%	100%	60%
125%	100%	80%	70%
300%	100%	70%	60%

Table 6.11: (2D Reconstructed Rössler) Frequency of occurrence of the mode number of homological features in the 2-cluster containing the longest living homological feature for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

Table 6.11 shows that when the **number of landmarks** equals 10 the temporally-informed witness relation increases the frequency with which a topological signature of one is reported from as little as 50% when using the standard witness relation to as much as 100% for higher $k_{\text{distortion}_+}$. Consistency in a signature is important when hoping to use it towards classification, segmentation or prediction. Since this is in particular an explainable signature with respect to the delay reconstructed Rössler point cloud, we consider this a notable improvement.

Table 6.12 in contrast shows that for delay reconstructed Lorenz point clouds, when the **number of landmarks** equals 10 and the **number of witnesses** equals 100, the mode occurs 100% of the time when using the standard witness complex, decreasing to 90% when $k_{\text{distortion}_+} = 20\%$.

$k_{\text{distortion}_+} \setminus$ (W,L)	(100, 10)	(1000, 10)	(1000,20)	(2000, 20)	(2000, 50)	(2000, 100)
0%	100%	80%	50%	60%	70%	40%
2%	100%	80%	60%	50%	60%	70%
20%	90%	70%	80%	90%	90%	80%
125%	100%	50%	40%	30%	30%	30%
300%	100%	70%	80%	90%	30%	20%

Table 6.12: (2D Reconstructed Lorenz) Frequency of occurrence of the mode number of homological features in the 2-cluster containing the longest living homological feature for filtrations terminated at $k_{\text{top}} = 9$ for witness complexes built using $k_{\text{distortion}_+}$ varying from $0\% - 300\%$ of the diameter of the point cloud.

This is actually positive news for the temporally-informed witness relation using $k_{\text{distortion}_+} = 20\%$, since all witness complex constructions for these (**number of witnesses**, **number of landmarks**) parameter values reported a topological signature of one, which does not distinguish Lorenz from Rössler. Increasing the **number of witnesses** to 1000 decreases frequency of mode occurrence for all witness complex constructions. For all witness complex constructions excluding when $k_{\text{distortion}_+} = 20\%$, this is an improvement since the topological signature is still erroneously one and we are reporting it less frequently. In the case when $k_{\text{distortion}_+} = 20\%$ we are capturing an explainable topological signature of three with a frequency of 70%, an even greater accomplishment. For all other (**number of witnesses**, **number of landmarks**) parameter values, the temporally-informed witness relation using $k_{\text{distortion}_+} = 20\%$ produces filtrations that are able to capture this topological signature of three with frequency between 80% and 90%.

It is only when **number of witnesses** equals 2000 and **number of landmarks** equals 50 or 100 that the standard witness complex captures a topological signature for the Lorenz system that can distinguish it from the Rössler. However the topological signatures representing the Lorenz system, which are distinct for 50 and 100 **number of landmarks**, arise 70% of the time when using 50 landmarks and only 40% of the time when using 100 landmarks.

While all constructions but the temporally-informed witness relation using $k_{\text{distortion}_+} = 300\%$ result in a correct mode topological signature of three when using 2000 witnesses and 100 landmarks, the **temporally-informed witness relations** using $k_{\text{distortion}_+} = 2\%, 20\%$ report this signature with much greater frequencies of $70\%, 80\%$ compared to the temporally-informed witness relation

using $k_{\text{distortion}_+} = 125\%$ with frequency 35%. This indicates that **there is an appropriate range of $k_{\textbf{distortion}_+}$ values that improves not only the correctness, but the consistency of the topological signature, as well**.

Chapter 7

Conclusion

This thesis research is motivated by the task of developing an online regime-shift detection algorithm for time series analysis that incorporates topology. We conclude that witness complexes can be used in successful TDA for time series analysis **with care of parameter selection** or **upon further modification**. We test our techniques on time series derived from synthetic low dimensional chaotic dynamical systems, distinct dynamic regimes of the synthetic sparse Brunel neuronal network, as well as experimental live musical instrument recordings.

In this dissertation, we presented a topological membership test for machine learning on delay coordinate reconstructed sliding windows of time series data and investigate its strengths and limitations. We use a sparse simplicial complex model - the witness complex - to produce our topological signature. The variability in topological representation of the same data based on choice of model construction parameters lead us to introduce heuristics for parameter selection that aim to capture the global topology with minimal local topology, or noise added to the signal. We then introduce a metric that incorporates the temporal ordering of the data points to increase the distances between data points with different dynamics. In application to synthetic chaotic data, we show that using this new metric in the construction of topological models can reduce the size of computations by up to 50% and increase the strength and consistency of signal by up to 40% and 50%, respectively.

We train our membership test by computing the mean persistent homology rank function (PRF) associated to a subset of a class of data as a topological representation of that class, as

well as computing a measurement of the variance. We test class membership by computing the L^2 distance between a sample PRF and the mean PRF of that class and comparing this distance to a threshold parameterized by the variance. This results in successful classification between distinct classes within a data type for some range of the parameters involved in the construction of the filtration of witness complexes. Motivated by developing cheap algorithms for rapid classification of incoming time series data, our results are particularly encouraging in comparison to recently proposed higher dimensional topological analyses that also achieve successful classification of distinct Brunel neuronal network dynamic regimes. More generally, introducing topological measurements into a statistical analysis addresses challenges presented by nonlinearity and improves the interpretability of the results.

Balancing computational efficiency and correctness is a universal challenge. By weighting the norm of our topological signature to emphasize early-born and early-dying or late-born and late-dying persistent homology, we can infer and monitor the types of topology reported across filtrations of witness complexes for a range of construction parameters controlling the size, cost and speed of computations from a single measurement function. From this we can implement heuristics to select model parameters that capture large-scale global topological features and deemphasize local small-scale topology. Investigating the limitations of a method is a crucial part of research. Our model construction parameter selection heuristic does not work for non-stationary data like the time series from the Brunel neuronal network dynamics because the topological signature does not stabilize with an increase in amount of data as it does for near-periodic or chaotic systems.

Observing non-uniform density of the data points approximating attractors of low dimensional continuous chaotic dynamical systems due to acceleration and fractal structure redirected our objective from selecting appropriate witness complex model construction parameters to redefining the metric underlying the witness complex construction itself. To encourage the persistent homology to reflect the singularities in the equations of evolution while discouraging persistent homology due to the invariant measure of the attractor, we increase the distance between points moving in dissimilar directions in the reconstructed state space. We report the number of two-simplices at

the end of a filtration of witness complexes as well as interpretable features of the one-dimensional persistent homology to demonstrate the improvements to cost and topological signature from using the temporally-informed metric.

Future questions include investigating how the delay parameter in the process of Takens delay coordinate reconstruction relates to the persistent homology of delay reconstructed sliding windows of chaotic time series data. Work has been done for periodic systems in [27]. Another challenge of real world data is noise. To associate a direction of travel to each data point approximating the attractor of the dynamics, we take a first difference of the location coordinates. Incorporating this information into the metric amplifies the affects of noise in the data. One approach to mitigate noise would be to do clever averaging that is aware of the boundary. Alternatively, utilizing the directionality of time in topological models that avoid taking first differences can improve performance in the face of noise.

Bibliography

[1] H. Adams A. Tausz. Javaplex tutorial, 2018.

[2] H.L. Swinney A.M. Fraser. Independent coordinates for strange attractors from mutual information. Phys. Rev. A, 1986.

[3] V.I. Arnold. A proof of the a.n. kolmogorov's theorem on the conservation of conditional-periodic motions in a small change of the hamiltonian function. Uspehy Math. Nauk SSSR, 1963.

[4] G.D. Birkhoff. Dynamical systems. Colloquium Publications, 9, 1927.

[5] G.D. Birkhoff. Proof of the ergodic theorem. Proceedings of National Academy of Science USA, 17, 1931.

[6] R. Bowen. Markov partitions and minimal sets for axiom a diffeomorphisms. American Journal of Mathematics, 92, 1970.

[7] N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. Journal of Computational Neuroscience, 8:183–208, 2000.

[8] P. Bubenik. Statistical topological data analysis using persistence landscapes. Journal of Machine Learning Research, 16, 2015.

[9] J. Curry. Topological data analysis and cosheaves. Japan Journal of Industrial and Applied Mathematics, 32:333–371, 2015.

[10] G.B. Mindlin D. Sciamarella. Unveiling the topological structure of chaotic flows from data. Phys. Rev. E, 64, 2001.

[11] O. Ya. Viro D.B Fuchs. Topology II. Springer Encyclopedia of Mathematical Sciences Topology, 2004.

[12] C.H. Dowker. Homology groups of relations. The Annals of Mathematics, 56:84–95, 1952.

[13] S. Filip. Notes on the multiplicative ergodic theorem. https://arxiv.org/abs/1710.10694, 2017.

[14] B. Finch. Residual set. From MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein, http://mathworld.wolfram.com/ResidualSet.html, 1952.

[15] A. M. Fraser. Information and entropy in strange attractors. IEE Transactions on Information Theory, 35, 1989.

[16] A. Zomorodian G. Carlsson. Computing persistent homology. Discrete and Computational Geometry, 33, 2005.

[17] R. Ghrist. Elementary applied topology, ed. 1.0. createspace, 2014.

[18] T. Emerson E. Hanson M. Kirby F. Motta R. Neville C. Peterson P. Shipman L. Ziegelmeier H. Adams, S. Chepushtanova. Persistence images: A stable vector representation of persistent homology. The Journal of Machine Learning Research archive, 18, 2017.

[19] A. Zomorodian H. Edelsbrunner, D. Letscher. Topological persistence and simplification. Discrete and Computational Geometry, 28, 2002.

[20] T. Schreiber H. Kantz. Nonlinear Time Series Analysis. Cambridge University Press, 2004.

[21] M. Hénon. A two-dimensional mapping with a strange attractor. Communications in Mathematical Physics, 59, 1976.

[22] E. Bradley J. Garland. Prediction in projection. Chaos, 25, 2015.

[23] J.D. Meiss J. Garland, E. Bradley. Exploring the topology of dynamical reconstructions. Physica D: Nonlinear Phenomena, 334:49 – 59, 2016.

[24] M. Vejdemo-Johansson J. J. Berwald, M. Gidea. Automatic recognition and tagging of topologically different regimes in dynamical systems. Journal of Discontinuity, Nonlinearity and Complexity, 2013.

[25] M.E. Davies J. Huke J. Stark, D.S. Broomhead. Takens embedding theorems for forced and stochastic systems. Nonlinear Analysis: Theory, Methods & Applications, 30, 1997.

[26] B.J. McNeil J.A. Hanley. The meaning and use of the area under the receiver operating characteristic (roc) curve. Radiology, 143:29–36, 1982.

[27] J. Harer J.A. Perea. Sliding windows and persistence: An application of topological methods to signal analysis. Foundations of Computational Mathematics, 15, 2015.

[28] K. Hess J.B. Bardin, G. Spreemann. Topological exploration of artificial neuronal network dynamics. (preprint), https://www.biorxiv.org/content/biorxiv/early/2018/09/23/424994.full.pdf, 2018.

[29] R.F. Williams J.S. Birman. Knotted periodic orbits in dynamical systems - i. lorenz's equations. Topology, 22, 1983.

[30] S. Gupta K. Mittal. Topological characterization and early detection of bifurcations and chaos in complex systems using persistent homology. Chaos, 27, 2017.

[31] A.N. Kolmogorov. On preservation of conditionally periodic motions under a small change in the hamiltonian function. Dokl. Akad. Nauk SSSR, 1954.

[32] Y. A. Kuznetsov. Elements of Applied Bifurcation Theory. Springer-Verlag Applied Mathematical Sciences, 1995.

[33] L. Moniz L. M. Pecora. A unified approach to attractor reconstruction. Chaos, 17, 02 2007.

[34] C. Elia L.Dieci. The singular value decomposition to approximate spectra of dynamical systems. theoretical aspects. Journal of Differential Equations, 230:502–531, 2006.

[35] E.N. Lorenz. Deterministic nonperiodic flow. J. Atmos. Sci., 20:130–141, 1963. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

[36] A. Nigmetov M. Kerber, D. Morozov. Geometry helps to compare persistence diagrams. SIAM Algorithm Engineering and Experiments, 2016.

[37] G. A. Hedlund M. Morse. Symbolic dynamics. American Journal of Mathematics, 1938.

[38] J.P. Huke R. Hegger M. Muldoon, D. Broomhead. Delay embedding in the presence of dynamical noise. Dynamics and Stability of Systems, 13, 05 1998.

[39] H.D. Abarbanel M.B. Kennel, R. Brown. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A, 45:3403–3411, 1992.

[40] M. Teillaud M.D. Bogdanov, O. Devillers. Hyperbolic delaunay complexes and voronoi diagrams made practical. Journal of Computational Geometry, 5, 2014.

[41] J.D. Meiss. Differential Dynamical Systems. Society for Industrial and Applied Mathematics, 2007.

[42] J.P. Huke D.S. Broomhead M.R. Muldoon, R.S. MacKay. Topology from time series. Physica D, 65, 1993.

[43] M. Rucco N. Antienza, R. Gonzalez-Diaz. Persistent entropy for separating topological features from noise in vietoris-rips complexes. Journal of Intelligent Information Systems, pages 1–19, 2017.

[44] U. Tillmann P. Grindrod H. Harrington N. Otter, M. Porter. A roadmap for the computation of persistent homology. EPJ data Science, 2017.

[45] S. Molnar J.D. Meiss E. Bradley N. Sanderson, E. Shugerman. Computational topology for time series analysis. Advances in Intelligent Data Analysis XVI: 16th International Symposium, IDA 2017, London, UK, October 26-28, 2017, Proceedings, 2017.

[46] J.D. Farmer R.S. Shaw N.H. Packard, J.P. Crutchfield. Geometry from a time series. Physical Review Letters, 45, 1980.

[47] J.D. Farmer R.S. Shaw N.H. Packard, J.P. Crutchfield. Detecting strange attractors in turbulence. Dynamical Systems and Turbulence, Warwick 1980, 898, 1981.

[48] L.I. Nicolaescu. Lectures on the geometry of manifolds, 1996.

[49] V. I. Oseledets. A multiplicative ergodic theorem, characteristic ljapunov exponents of dynamical systems. Transactions of the Moscow Mathematical Society, 19, 1968.

[50] Smale S. Weinberger P. Niyogi, S. Finding the homology of submanifolds with high confidence from random samples. Discrete Comput Geom, 39, 2008.

[51] H. Poincaré. Analysis situs. Journal de l'École Polytechnique, 2, 1895.

[52] O.E. Rössler. An equation for continuous chaos. Physics Letters A, 57, 07 1976.

[53] O.E. Rössler. Chaos and strange attractors in chemical kinetics. Springer Series in Synergetics, 3, 1979.

[54] V. Robins. Computing connectedness: Disconnectedness and discreteness. Physica D, 139, 2000.

[55] M. Robinson. Sheaves are the canonical data structure for sensor integration. Information Fusion, 36:208–224, 2017.

[56] D. Cook S. Aminikhanghahi. A survey of methods for time series change point detection. Knowl Inf Syst, 51:339–367, 2017.

[57] M. Rajkovic S. Maletic, Y. Zhao. Persistent topological features of dynamical systems. Chaos, 26, 2016.

[58] Y. G. Sinai. Markov partitions and c-diffeomorphisms. Funct. Anal. Appl., 2, 1968.

[59] S. Smale. Differentiable dynamical systems. Bulletin of American Mathematical Society, 73, 1967.

[60] J.C. Sprott. Common chaotic systems. http://sprott.physics.wisc.edu/chaos/comchaos.htm, 2004.

[61] E. M. Bollt T. Ma. Shape coherence and finite-time curvature evolution. International Journal of Bifurcation and Chaos, 25, 2015.

[62] J. Yorke T. Sauer. How many delay coordinates do you need? International Journal of Bifurcation and Chaos, 3:729, 1993.

[63] W. Tucker. A rigorous ode solver and smale's 14th problem. Foundations of Computational Mathematics, 3 2002.

[64] M. Kramer U. Eden. Drawing inferences from fano factor calculations. Journal of Neuroscience Methods, 190:149–152, 2010.

[65] G. Carlsson V. de Silva. Topological estimation using witness complexes. Proceeding SPBG'04 Proceedings of the First Eurographics conference on Point-Based Graphics, 2004.

[66] M. Vejdemo-Johansson V. de Silva, D. Morozov. Persistent cohomology and circular coordinates. Discrete Comput Geom, 45, 2011.

[67] K.Turner V. Robins. Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. Physica D, 344, 2015.

[68] J. von Neumann. Proof of the quasi-ergodic hypothesis. Proceedings of National Academy of Science USA, 18, 1932.

[69] A. Wilkinson. What are lyapunov exponents, and why are they interesting? Bulletin of the American Mathematical Society, 54:79–105, 2017.

[70] R. F. Williams. Expanding attractors. Publications Mathematiques de I'IHES, 43, 1974.

[71] J.D. Meiss N. Sanderson Z. Alexander, E. Bradley. Simplicial multivalued maps and the witness complex for dynamical analysis of time series. SIAM J. Appl. Dyn. Syst., 2015.