Quantifying Neurological Complexity of *Caenorhabditis elegans* through Dynamic Diffraction

by

Susannah G. Zhang

(Under the Direction of Heinz-Bernd Schüttler)

Abstract

The nematode *Caenorhabditis elegans* (*C. elegans*) is a model organism, commonly studied due to ease of maintenance and comparative simplicity of its neurological structure. We investigate *C. elegans*' locomotion using dynamic diffraction and nonlinear dynamics. Observed diffraction intensity time-series relate to the net electric field diffracted from all points of the worm at any point in the far-field diffraction pattern [1]. Consequently, key features of locomotion can be recovered by analyzing the intensity time-series. We found significant markers of low-dimensional chaos which prove that *C. elegans* locomotion satisfies the chaos criteria outlined by David Feldman [2]: determinism, aperiodic orbits, bounded orbits, and sensitive dependence on initial conditions. To prove that *C. elegans* locomotion meets the aforementioned criteria, we use nonlinear analysis - e.g., Takens (1981) embedding [3], mutual information (MI), lag plots, false nearest neighbors (FNN), largest Lyapunov exponent (LLE), correlation dimension, recurrence plots, and surrogate data analysis - to characterize and analyze the time-series.

First, we take the Fourier transform (FT) of the time-series and observe a broad frequency spectrum, which provides evidence that the criteria of aperiodic orbits is satisfied. Second,

we generate lag plots of our locomotion data and explain how the plots satisfy all four of the chaos criteria. Third, our time-series fulfills the criteria for low-dimensional chaos with a typical positive LLE value (base e) around 1.39 ± 0.02 s^{-1}, at optimal embedding dimension $n = 4$, indicating sensitive dependence on initial conditions. Next, our calculated non-integer average correlation dimension of $\sim 2.08 \pm 0.24$ means that our data may have characteristics of both 2D and 3D space, which would indicate possible chaotic dynamics. Furthermore, the correlation dimension value stabilizes with increasing embedding dimension, indicating deterministic dynamics. We also visualize the locomotion dynamics using recurrence plots because the resulting plots can prove our data satisfies determinism, aperiodic orbits, and sensitive dependence on initial conditions. As a final measure to test for chaos, we also use surrogate data analysis to prove that our time-series is nonlinear. All results provide strong evidence that *C. elegans* locomotion is indeed chaotic.

INDEX WORDS: nonlinear dynamics, chaos, *C. elegans*, far-field diffraction, fast Fourier transform, entropy, false nearest neighbor, Lyapunov exponent, recurrence plots, surrogate data analysis

Quantifying Neurological Complexity of *Caenorhabditis elegans* through Dynamic Diffraction

by

Susannah G. Zhang

B.S.Ed, Indiana University of Pennsylvania, 2012

M.S., University of Georgia, 2018

A Thesis Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2022

Quantifying Neurological Complexity of *Caenorhabditis elegans* through Dynamic Diffraction

by

Susannah G. Zhang

Major Professor:	Heinz-Bernd Schüttler
Committee:	Jenny Magnes
	David Landau
	Steven P. Lewis

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2022

Acknowledgments

I sincerely want to thank Heinz-Bernd Schüttler for his willingness to advise a remote student, and for his unwavering support towards my PhD completion. I also want to thank Jenny Magnes for her companionship and mentorship, without which I would not have finished my PhD degree. Special thanks to my committee: Dr. David Landau and Dr. Steven Lewis. Sincere thanks to my writing buddy, Julie Lange, who helped to keep me accountable.

Additionally, thank you to collaborator Harold Hastings for his many insightful conversations that helped me grow as a researcher. Harold played an integral role in orchestrating this project through his collaboration with Jenny Magnes and myself. This project may never have happened without him. Our entire research group was greatly saddened by his passing back in early May of 2022. He was a welcoming, kind, and wise man who was always more than happy to answer a question about any topic. Thank you, Harold. Rest in peace.

Many thanks to the Faculty at Vassar College for providing such a supportive academic community. The professional environment that you created helped me to flourish as I began my career in academia. Special thanks to Eva Woods-Peiró for her companionship and mentorship within the field of academia and beyond.

To my family, thank you for being there to encourage me and comfort me; my husband, Shengming Zhang, for his support and helping me finish my dream of a PhD; my father, Captain Donald Earl Rosenberger, for always telling me to do my best, knowing that my internal desire to always be better would take over; my siblings, Don Paul Rosenberger and Danielle Rosenberger, for being awesome growing up companions. And special thanks to my son, Alan, for lighting up some of my darkest days by telling me to believe in myself, because he believed in me. This dissertation is dedicated to my family, both present and future.

Table of Contents

Page

Acknowledgments iv

List of Figures vii

List of Tables xv

Chapter

1 Introduction and Literature Review 1

2 Experimental Methods and Setup 5

2.1 Experimental Setups 5

2.2 Dynamic Diffraction Patterns 8

3 Fourier Transform 13

3.1 Fourier Transforms 13

3.2 Applications to Diffraction 14

3.3 Applications to Time-Series: Experimental Data Sets . . . 17

4 Analysis of Complex Systems 21

4.1 Frequency Spectrum 22

4.2 Phase Space Generation 23

4.3 Embedding Dimension 32

4.4 Largest Lyapunov Exponent 35

4.5 Correlation Dimension 41

4.6 Recurrence Plots 44

4.7 Surrogate Data Analysis 56
4.8 Summary of Results and Discussion 67
5 Conclusion 71
Bibliography 73

List of Figures

1.1 Image of a *C. elegans* at Vassar College. A transparent adult nematode that is ∼1 mm long and ∼100 μm wide. 2

1.2 Illustration of two types of *C. elegans* locomotion. The time labels are such that $0 < t_1 < t_2 < t_3$. (a) Wildtype locomotory pattern moving in an "s" shape. (b) Roller locomotory pattern curling in a "c" shape. 2

2.1 Setup used to collect data analyzed in this thesis. The full setup can be seen in the above view in (a), while the vertical displacement of the laser beam through the cuvette (C) is more clearly displayed in (b). HeNe Laser passes through a neutral density filter (ND) to reduce laser intensity, then the laser beam is redirected downwards by mirror M1 towards the cuvette C holding the swimming worm. The diffraction pattern is redirected towards a photodiode PD by mirror M2 such that the central maximum is off-center to not saturate the detector. In the side view (b), the mirror M2 directs the beam into the page where it propagates towards the photodiode PD. The intensity measured by the PD is sent to the Oscilloscope, then sent to the computer CMP. Distances are not drawn to scale. [5, 23] 6

2.2 New setup for future experiments. HeNe Laser is redirected by two mirrors (M1 and M2), then passes through a neutral density filter (ND) to reduce laser intensity. Next, the beam is cleaned by a spatial filter (SF), before going through a series of lenses (L1, L2, and L3) that work together to collimate the beam to a desired beam diameter of about 3 mm before reaching the cuvette (C) holding the swimming worm. The resulting time-dependent diffraction pattern is then focused by a lens (L4) into our CMOS detector. Please note that distances are not drawn to scale. 7

2.3 Illustration of dynamic diffraction generation and intensity versus time data generation. On the left, we show a simplified version of our experimental setup shown in Figure 2.1 where the diffraction pattern is imaged on a screen with a photodiode (PD). In the middle, we show an sample diffraction pattern with a red dot denoting the placement of the photodiode. As the worm moves, the diffraction pattern, hence the intensity at the PD, changes in response to that motion. These time dependent intensity fluctuations are then recorded by the PD to generate a 1D dynamic diffraction data set, as shown on the right. . . 9

2.4 Experimental example of the different diffraction patterns present from both the width of the worm (spread out single slit pattern) and the length of the worm (more closely spaced single slit diffraction pattern) [28]. 11

3.1 Example aperture function for a single slit (equation (3.4)) in blue with width a. The Fourier Transform (equation (3.6)) is shown in red. We can notice that if we make the width, a, larger and larger, then the width of the Fourier Transform's central peak gets smaller and smaller. This is reflective that larger slits/objects cause tighter diffraction patterns, just as observed for our worm in Figure 2.4 15

4.1 Wildtype time-series, $y(t_\mu)$, with its corresponding power spectrum, $S(\nu_k)$.
Left: Original experimental intensity versus time.
Right: Resulting broad frequency spectrum from FFT ($\nu_{\text{avg}} = 1.32$ Hz). . . . 23

4.2 Visualization of mutual information, $MI(\tau)$ [44]. (a) Shows a simple example of data sequence $y(t_\gamma)$ and a delayed data sequence $y(t_\gamma + \tau)$ on the same axis. (b) Venn diagram showing the information (entropy) from the data shown in (a). The optimal lag τ is found when there is the least amount of overlap (mutual information in purple) in information between un-lagged $y(t_\gamma)$ (H_0 denoted by red circle) and lagged $y(t_\gamma + \tau)$ (H_τ denoted by blue circle). (c) Showcases an experimental time-series of wildtype *C. elegans* locomotion obtained from the experimental setup described by Magnes et al. [1]. (d) Plot of MI versus lag (in units of data points) generated from experimental time-series in (c) where red circles highlight local minima. The first local minimum gives the "optimal" lag. 29

4.3 Example 2D projection of a lag plot generated from the dynamic diffraction time-series of a swimming *C. elegans* in Figure 4.2c [44]. The colors reflect a progression in time where dark blue reflects earlier t_γ values and dark red reflects later t_γ values. The first component $y(t_\gamma)$ is plotted on the x-axis, and the second component $y(t_\gamma + \tau)$ is plotted on the y-axis with $\tau = 0.1716$ s. Based on the apparent crossings in this projection, we need a higher embedding dimension than 2 to fully resolve the dynamics and eliminate false crossings. Determining the best embedding dimension will be discussed in section 4.3. 31

4.4 Percentage, $q_{\text{FNN}} = \text{FNN}(\tilde{n})/\text{FNN}(1)$, of remaining FNN count versus embedding dimension, $\tilde{n}$, generated from the locomotory time-series in Figure 4.2c [44]. The FNN count is shown as a percentage of the reference count at $\tilde{n} = 1$. At $\tilde{n} = 3$, q_{FNN} falls to 0.85%, then 0.02% at $\tilde{n} = 4$. The FNN percentage falls to zero at higher embedding dimensions. We therefore use $n = 4$ as the minimal embedding dimension that completely unfolds the dynamics of *C. elegans* locomotion for this experimental time-series. The same minimal embedding dimension was found and used for all other experimental time-series studied. 34

4.5 An illustration of an LLE pair , $(t_{\gamma_j}, t_{\gamma_j\prime})$, that satisfies the criteria listed in equations (4.19) - (4.21) at some initial time $\bar{t}_\beta = 0$ s, shown on the left. Following each point along their respective trajectory for a set evolution time $\bar{t}_\beta$, we can calculate the new separation distance, or the divergence, $D_{\text{div}}(t_{\gamma_j}, \bar{t}_\beta)$, such as the separation depicted on the right. 37

4.6 $L_{\text{div}}(\bar{t}_\beta)$, calculated from the experimental time-series data, $y(t_\mu)$, by equations (4.4), (4.22), and (4.29), is plotted versus evolution time $\bar{t}_\beta$ in blue for the time-series shown in Figure 4.2c [44]. By inspection of the blue curve, the time points from ~ 0 s to 1 s are identified as the region where the linearization holds, as stated in equation (4.28). Only these time points were included in the linear fit shown in red. The slope of the fitted red line gives a λ_{LLE} value of 1.39 ± 0.02 s^{-1}. 40

4.7 $\ln\left(C(\varepsilon_{\mathrm{d}_2})\right)$, calculated from the experimental time-series data, $y(t_\mu)$, by equations (4.4) and (4.32), is plotted versus $\ln\left(\varepsilon_{\mathrm{d}_2}\right)$ in blue. By inspection of the blue curve, the points from ~ 0.0198 to ~ 0.2485 are identified as the region where the linearization holds, as insinuated in the proportionality in equation (4.33). Only these points were included in the linear fit shown in red, which illustrates how linear regions are chosen for generating correlation dimension values, d_{cor}, of our experimental time-series. 42

4.8 Plot of calculated correlation dimension as embedding dimension is increased. Random data (red) obtained from random.org increases linearly as the embedding dimension is increased, but our *C. elegans* data (black) stabilizes around a correlation dimension slightly above 2, indicating deterministic behavior [62]. Additionally, our locomotion data distinctly differs from the linear behavior exhibited by the random data at embedding dimension 3, which verifies that our locomotion data needs an embedding dimension of 3 or higher to unfold the dynamics. 43

4.9 1D Heatmap examples generated from a simple sine wave. Part (b) shows the 1D Heatmap with only the LOI highlighted in red. Parts (c) and (d) showcase a line parallel to the LOI and how it is generated in pink while parts (e) and (f) showcase a line perpendicular to the LOI and how it is generated in blue. 46

4.10 Changing various parameters to observe the effects on the 1D Heatmap for several synthetic sine waves. (a) shows amplitude reduced by half at $t = 4\pi$, then returned to the original amplitude at $t = 8\pi$. (b) shows the frequency tripled at $t = 4\pi$, then returned to the original frequency at $t = 8\pi$. (c) shows the frequency gradually changing with time based on the function $f = 2t - \cos(t)$. (d) shows constant frequency and random amplitudes. 49

4.11 RPs with $\varepsilon = \pi/128$ of the simple sine wave from Figure 4.9a. (a) shows the 1D RP of the sine curve. (b) plots the first (original sine) and second (derivative in blue) component to generate 2D embedding of the simple sine wave. These two curves will serve as the two input vector components of a 2D vector time-series to be displayed as an RP in (c). (d) plots the second method to generate 2D embedding (delay embedding) using the first (original sine) and second component (lagged function in pink) to generate 2D embedding of the simple sine wave, then the resulting 2D RP is plotted in (e). Both methods of 2D embedding generate the same RP plot. 51

4.12 1D RPs of an experimental time-series of a swimming wildtype *C. elegans* on two different time scales [44]. Features emerge on various time scales in these RPs. The threshold, ε, used was $p_r = 10\%$ of the range of amplitudes, $\Delta\varphi_A$. (a) RP for a swimming *C. elegans*. The system is evolving over time with the patterns changing along the LOI at about 5.35 s, 12 s, 16 s, 25.3 s, 31.6 s, 34.2 s, 39.5 s, 41.6 s, and 45.2 s. (b) Highly resolved details of the swimming *C. elegans* over about 6 seconds (5.35 - 12s). Diagonal lines parallel to the LOI relate to the predictability of the system and can be used to determine the LLE. The diagonal from 8.48 s to 8.84 s has a length of 0.509 s, which relates to an LLE frequency of 1.96 s^{-1}. The frequency of about 2 Hz matches the swimming frequencies of wildtype *C. elegans* [15]. The crossing lines and lines perpendicular to the LOI indicate that a higher embedding dimension is needed to fully resolve the features. 52

4.13 Generating embedded recurrence plots for experimental *C. elegans* time-series, with t_ι and t_κ plotted along the x-axis and y-axis, respectively [44]. The red line showcases the LOI, and the blue boxes showcase different sections of the motion. (a) Heatmap of the distance matrix $D_{\iota,\kappa}$ for the 1D vector time-series, $y^1(t_\mu)$, *i.e.*, for embedding dimension $n = 1$. (b) RP of recurrence matrix $R_{\iota,\kappa}$ for the 1D vector time-series, $y^1(t_\mu)$. A threshold calculated using $p_r = 0.1\%$ of the range of amplitudes, $\Delta\varphi_A$, has been added to the 1D embedded RP. (c) RP of recurrence matrix $R_{\iota,\kappa}$ for the 4D vector time-series, $y^4(t_\mu)$, *i.e.*, for embedding dimension $n = 4$. The time-series is now embedded in 4D with threshold calculated using $p_r = 0.3\%$ of the range of amplitudes, $\Delta\varphi_A$. The embedding dimension is proven sufficient to unfold the dynamics because all perpendicular lines (FNN) are eliminated (as illustrated in Figure 4.11). The diagonal lines that are parallel with the LOI indicate how long a particular state lasts. 53

4.14 Identification of scaling sections in a recurrence plot generated for an experimental time-series [44]. RP with 4D embedding and threshold calculated using $p_r = 0.3\%$ of the range of amplitudes, $\Delta\varphi_A$. The time-series is plotted on both axes and is color coded into three sections based on visual inspection of the time-series alone. (a) The three colored sections in the time-series are also sectioned in the RP: 0-11.49 s in blue, 11.49-36 s in red, and 36-43.38 s in green. Using these sections each LLE was calculated using an algorithm by Rosenstein et al. [55], $\mathrm{LLE_{T1}} = 2.22 \pm 0.06\ \mathrm{s^{-1}}$, $\mathrm{LLE_{T2}} = 1.83 \pm 0.01\ \mathrm{s^{-1}}$, and $\mathrm{LLE_{T3}} = 2.74 \pm 0.01\ \mathrm{s^{-1}}$ were calculated. (b) Through visual inspection of the RP, different sections are identified, which are inside blue boxes in the RP: 0 - 10 s, 10 - 24.5 s, and 24.5 - 42.5 s. These sections were used to calculate $\mathrm{LLE_{R1}} = 2.22 \pm 0.01\ \mathrm{s^{-1}}$, $\mathrm{LLE_{R2}} = 1.70 \pm 0.01\ \mathrm{s^{-1}}$, and $\mathrm{LLE_{R3}} = 1.83 \pm 0.01\ \mathrm{s^{-1}}$. 55

4.15 Example set (9 out of 1000) of surrogate time-series, $\breve{y}_s(t_\mu)$, (red) generated from an original time-series, $y(t_\mu)$, (blue) using the AAFT method of surrogate time-series construction. 65

4.16 Rankings and histograms of the time reversal asymmetry NTO values , Λ_0 and Λ_s, for original and surrogate time-series, respectively [44]. In each panel, (a) and (b), the NTO value, Λ_0, from an original time-series is shown as a vertical red line. The sample of surrogate NTOs, Λ_s for $s = 1, \cdots, M$, is generated from the original time series by the AAFT method and is shown as a histogram. Panel (a) is for a typical original time-series of experimental observations, out of 27 observed time-series total, for which the null hypothesis was rejected. Panel (b) is for the outlier time-series, *i.e.*, for the one and only original time-series of experimental observations, for which the null hypothesis was not rejected. 66

List of Tables

4.1 Chaos criteria [2] and how to satisfy them 67

4.2 Summary of typical *C. elegans* Results 70

Chapter 1

Introduction and Literature Review

This thesis will provide a multifaceted approach to characterize the locomotion of *Caenorhabditis elegans* (*C. elegans*) in a 3D environment. In this chapter, we will cover our motivation for this project and cover the general outline of this thesis. First, we begin by explaining what are *C. elegans* and why we chose this organism.

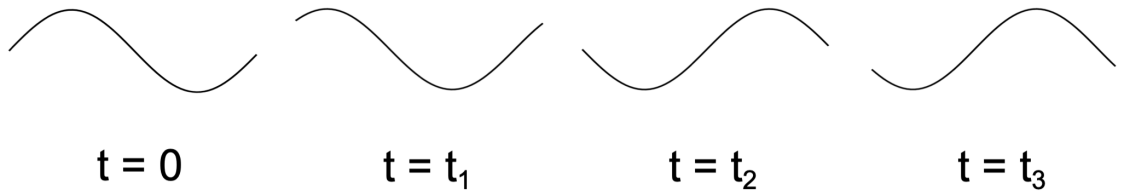
C. elegans are transparent nematodes that inhabit soil in temperate environments. They are usually studied under a microscope with limited mobility; i.e., in a 2D environment. In addition, these worms change their behavior when exposed to blue/violet light [4]. We use red light, which causes no known interference with its locomotion. The protocol for the growth and culturing of the worms in the lab is described by Magnes et al 2017 [5]. These nematodes are about 1 mm long and 100 μm wide (Figure 1.1) with only 302 neurons. The neurons are similar in form and function to that of humans and are therefore of interest to the neurological and biological community. The small size and quick maturation ($\sim$4 days at 20° C [5]) of this "complete microorganism" make it a model organism to observe and characterize locomotion.

Some of the over 70 different mutants of *C. elegans* have neurological defects that affect the locomotion [6]. We are working on a method that can distinguish these different mutants based on the complexity of their locomotion. As this worm's nervous system has been fully mapped [7, 8, 9], and 75 of the worm's neurons are motor neurons that are responsible for controlling the 79 body wall muscles [10, 8, 11], our results with regards to the dynamics of this neurological system will help us better understand human neurology due to *C. elegans*

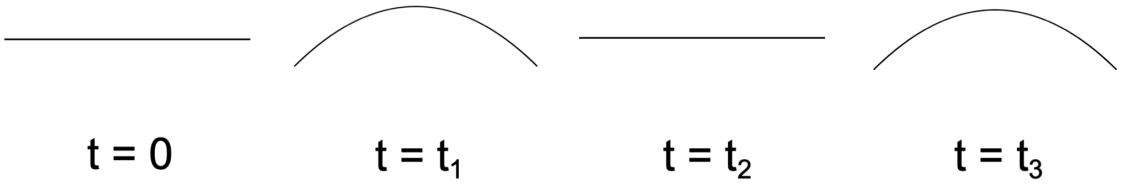
having neurons that are similar to those of humans.

Figure 1.1 Image of a *C. elegans* at Vassar College. A transparent adult nematode that is $\sim$1 mm long and $\sim$100 μm wide.

We will focus on two phenotypes in particular: Wildtype (N2) and Roller (OH7547). Wildtype have no genetic defects that affect their locomotory pattern. They tend to undulate in a sinusoidal-like pattern [12, 13]. Rollers have a genetic defect that helically twists their whole body and alters their locomotory pattern such that adult rollers move by curling into a "c" shape in one direction only and then straighten [14]. An example of these two types of locomotion can be seen in Figure 1.2.



(a) Wildtype



(b) Roller

Figure 1.2 Illustration of two types of *C. elegans* locomotion. The time labels are such that $0 < t_1 < t_2 < t_3$. (a) Wildtype locomotory pattern moving in an "s" shape. (b) Roller locomotory pattern curling in a "c" shape.

The locomotion of *C. elegans* in 2D environments has been studied with over 10,000 references available and is well-understood [15]. *C. elegans* have also been studied in 3D environments [16]. Our extension to 3D environments uses far-field (Fraunhofer) diffraction

("diffraction microscopy") [1, 17] to supplement conventional microscopic imaging, the latter being restricted to a 2D focal plane. Diffraction microscopy allows the species to be probed in more natural environments [18] since the species does not have to adhere to a focal plane. Another feature of diffraction microscopy is that motions can be detected across multiple scaling regions to less than a wavelength with less computing power than video analysis.

Intensity fluctuations at one point in the far-field (Fraunhofer) diffraction pattern of a microscopic worm form a one-dimensional time-series [5]. The intensity at each point in the diffraction pattern is the result of the superimposed light rays from each point on the worm; hence the time-series represents a record in time of the undulation causing the locomotion. A typical experimental time-series has a sampling rate of 1.3 kHz and is recorded over about 30-50 seconds, resulting in 40,000 - 65,000 data points. Our experimental setup and a brief overview of experimental diffraction patterns obtained are covered in chapter 2. Additionally, Fourier transforms and how they relate to this project are covered in chapter 3.

Analytic methods for characterizing the chaotic locomotion of *C. elegans* is presented in chapter 4. We first seek to reconstruct the underlying dynamics by using phase space embedding. One method uses lag plots (section 4.2), which are constructed by embedding the time-series in phase space by choosing an appropriate time delay, τ, and plotting n-tuples $\{(y(t), y(t+\tau), y(t+2\tau), \cdots, y(t+(n-1)\tau)\}$. The resulting trajectories in phase space then characterize the underlying dynamics, where frequencies and amplitudes evolve through time. The embedding dimension (section 4.3), n, indicates how many variables/differential equations the locomotion depends on, giving the first clue of the complexity of the system. In addition, the largest Lyapunov exponent (LLE) (section 4.4), which is sometimes denoted as MLE for maximum Lyapunov exponent, indicates how quickly the nearby trajectories in phase space diverge, giving a measure of the chaos of the system. The final chaotic marker that we calculated is the correlation dimension (section 4.5), which indicates chaotic dynamics in the case of a non-integer value in addition to stabilizing in higher embedding

dimensions.

We also explored the temporal evolution of the diffraction signal generated by *C. elegans* locomotion using recurrence plots (RPs) (section 4.6) [19] and surrogate data methods (section 4.7) [20, 21]. Recurrence plots provide a straightforward visualization of periodicity and near-periodicity in the system. Recurrence plots can thus be used to further identify determinism associated with low-dimensional chaos, which is absent in truly stochastic systems [22]. Surrogate data methods are frequently used to test for and quantify time reversal asymmetry, which is a further indication of nonlinear causality, another marker of chaos. In addition, surrogate data methods can be used to test for the statistical significance of other markers of chaos such as the correlation dimension.

This thesis is organized as follows: chapter 2 outlines the experimental setup used to collect worm locomotion data and explains the related physical mechanisms. Then, chapter 3 explains use of Fourier Transforms and their applications to diffraction and time-series analysis. In chapter 4, we explain how to calculate and interpret markers of chaos as well as various visual representations of phase space trajectories. Finally, we summarize the thesis and detail future work.

Chapter 2

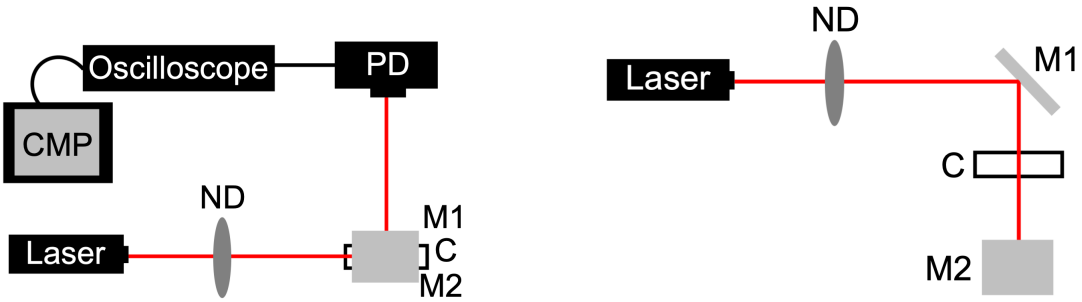
Experimental Methods and Setup

This chapter describes the setup for a dynamic diffraction experiment in the visible spectrum as well as how we obtain an intensity versus time data from the resulting diffraction pattern.

2.1 Experimental Setups

In this section, we will initially describe the experimental setup used to collect the data analyzed in chapter 4. Using this setup as shown in Figure 2.1, we study the locomotion of live *C. elegans*. Additionally, later parts of this section will discuss improvements to the setup for future experiments. Specifics of the equipment used in both optical setups are detailed below.

In the initial setup (Figure 2.1), a HeNe (Helium-Neon) laser is used to generate a 632 nm laser beam of about 2mW. Before the beam reaches the worm, it encounters a variable neutral density filter (maximum ND = 3.0) to reduce the laser's intensity to prevent saturation of the detector. The beam then travels through a water-filled cuvette (C) holding the freely swimming worm. The resulting dynamic diffraction pattern is imaged on a screen in the far-field, such as the initial setup as described by Magnes et al. 2012 [1]. We can obtain an intensity versus time data set collected at 1.3 kHz by a digital oscilloscope (PicoScope made by Pico Technology) through a photodiode (DET10A from ThorLabs) placed off center away from the central maximum to avoid saturation but still within the diffraction pattern. The 1D time-dependent diffraction time-series obtained contains information about the time evolution of superimposed light rays from every single point along the worm. Each time-series



(a) Top View (b) Side View

Figure 2.1 Setup used to collect data analyzed in this thesis. The full setup can be seen in the above view in (a), while the vertical displacement of the laser beam through the cuvette (C) is more clearly displayed in (b). HeNe Laser passes through a neutral density filter (ND) to reduce laser intensity, then the laser beam is redirected downwards by mirror M1 towards the cuvette C holding the swimming worm. The diffraction pattern is redirected towards a photodiode PD by mirror M2 such that the central maximum is off-center to not saturate the detector. In the side view (b), the mirror M2 directs the beam into the page where it propagates towards the photodiode PD. The intensity measured by the PD is sent to the Oscilloscope, then sent to the computer CMP. Distances are not drawn to scale. [5, 23]

was typically 50 seconds long allowing for enough data points to determine the evolution of the dynamical system.

While the setup in Figure 2.1 is sufficient for obtaining data, we have since made some optimizations to our setup for future work, which will be described in the following paragraphs and is illustrated in Figure 2.2. In the new setup, we still have a neutral density filter to reduce the laser's intensity to prevent saturation of the camera. Then, the laser beam travels through a spatial filter (SF) in order to reduce aberrations in the beam and output a uniform, Gaussian intensity profile, or a plane wave. A spatial filter does this by effectively performing a Fourier Transform on the beam. To accomplish this, the spatial filter contains two components to it: an objective lens and a pinhole aperture. First, the objective lens Fourier transforms the beam into a very tight spot. The resulting pattern at the beam waist, that is the point at which the beam is most tightly focused, sorts the spatial

frequencies. The higher maxima represent higher order spatial frequencies in the object's plane [24]. Using this method, irregularities in the beam can be removed before reaching the nematode. To successfully eliminate the higher order intensity maxima of the input laser beam, the focused beam then passes through a pinhole aperture positioned exactly at the focal point of the objective lens to filter out the concentric rings. To ensure that this filtering happens at the optimal level, the correct objective lens and pinhole combination must be chosen. Details outlining the proper way to calculate that combination for a HeNe laser is detailed by Edmund Optics [25]. Our spatial filter uses a microscope objective lens (10x, 0.25 NA, 16.5 mm Focal Length) paired with a pinhole of 25 μm [26].

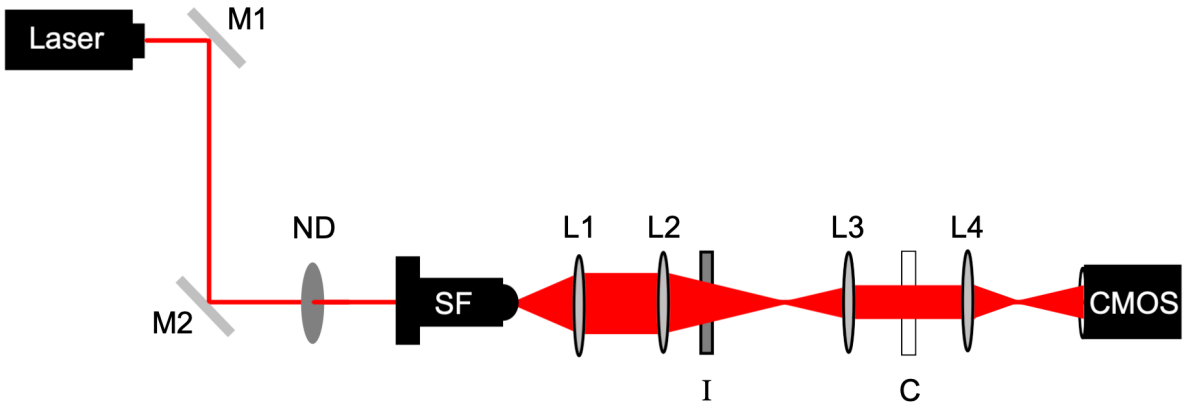


Figure 2.2 New setup for future experiments. HeNe Laser is redirected by two mirrors (M1 and M2), then passes through a neutral density filter (ND) to reduce laser intensity. Next, the beam is cleaned by a spatial filter (SF), before going through a series of lenses (L1, L2, and L3) that work together to collimate the beam to a desired beam diameter of about 3 mm before reaching the cuvette (C) holding the swimming worm. The resulting time-dependent diffraction pattern is then focused by a lens (L4) into our CMOS detector. Please note that distances are not drawn to scale.

After the pinhole aperture of the spatial filter, it would be ideal to place a single lens that would collimate the beam to a beam diameter of 3 mm and then place the cuvette holding the worm within the beam. However, the beam exiting the spatial filter's pinhole expands so quickly that a single collimating lens was not sufficient due to the physical limitations of our optical table. Instead, the setup will include three lenses: one to collimate after the

spatial filter pinhole (L1), one to refocus the beam to a tighter beam diameter (L2), and one to collimate the beam at the desired beam diameter (L3). The first lens (L1) is placed such that the distance between the pinhole and the lens exactly equals the focal length of that lens. This collimates the beam to prevent further beam expansion. The achieved collimation at this stage is too large for our worm sample, so we focus the light using another lens (L2) through an Iris (I). After the focused beam begins to expand again, we place a different lens (L3) with a focal length equal to the distance between the focused beam position from lens (L2) and where the beam diameter reaches 3 mm to achieve a collimated 3 mm beam diameter.

As the worm swims within the laser beam, the beam diffracts at all points along the worm. That diffraction pattern is then focused (L4) into our detector, which is a complementary metal-oxide semiconductor (CMOS) camera. A CMOS sensor takes the diffraction pattern intensities in space and converts light intensity signals at various locations to a voltage at that pixel location. These voltages are then used to reconstruct the diffraction pattern on a computer. With a new, cleaner beam due to the spatial filter, this new improved setup will be used for future data collection. However, all data analyzed in this thesis was taken using the setup in Figure 2.1.

2.2 Dynamic Diffraction Patterns

As the worm swims in the laser beam, the dynamic diffraction pattern originates from the outline of the worm as illustrated in Figure 2.3. Based on Huygens principle, each point along the worm's body will generate a new source that will interact with all other new sources along the worm. Each new orientation of the worm while swimming will reposition the interacting sources that reach the screen, thus changing the phase and therefore the diffraction pattern. Each point within the diffraction pattern contains time-dependent intensity information relating to the phase changes. In other words, the dynamics of the worm's locomotion can

be retrieved from a single point within the diffraction pattern. These intensity fluctuations are captured by a photodiode placed at a point within the diffraction pattern similar to the position of the red dot labeled PD in Figure 2.3. Since the resulting intensity at that point is due to diffraction about all points of the worm simultaneously, the information about the 3D locomotory shapes is folded into a 1D time-series. The 1D time series obtained will later be unfolded in phase space using techniques of chaos theory in chapter 4.

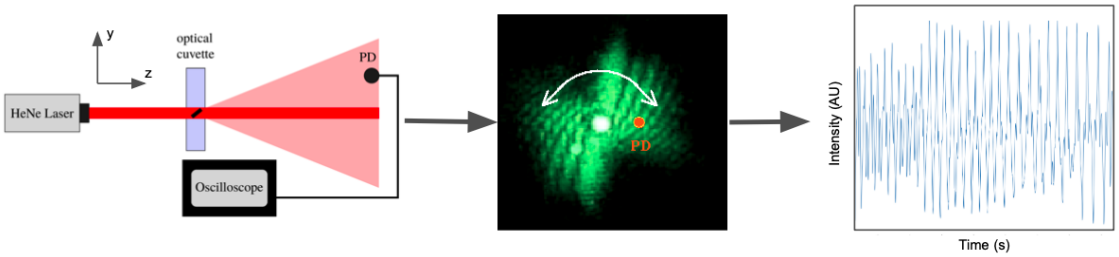


Figure 2.3 Illustration of dynamic diffraction generation and intensity versus time data generation. On the left, we show a simplified version of our experimental setup shown in Figure 2.1 where the diffraction pattern is imaged on a screen with a photodiode (PD). In the middle, we show an sample diffraction pattern with a red dot denoting the placement of the photodiode. As the worm moves, the diffraction pattern, hence the intensity at the PD, changes in response to that motion. These time dependent intensity fluctuations are then recorded by the PD to generate a 1D dynamic diffraction data set, as shown on the right.

To determine the limiting case of how wide/long the worm can be and still generate a diffraction pattern when interacting with coherent, laser light, we will approximate the worm's body with a straight line to illustrate the concepts. Using the equations of single slit diffraction from introductory physics, we describe the relationship between slit width (object width/length), a, spreading angle, θ, screen distance away from slit (object), L, and the spacing between the central maximum and the m^{th} region of destructive interference (minimum), y_m.

$$a \sin\theta = m\lambda \tag{2.1}$$

$$\tan\theta = \frac{y_m}{L}\,. \tag{2.2}$$

The laser wavelength is $\lambda = 632.8$ nm and the smallest dimension that the laser will interact with is the width of the worm around $a \approx 100$ μm. Using the small angle approximation ($\sin\theta \approx \tan\theta \approx \theta$ for angles less than $10^\circ \approx \frac{\pi}{18}$ radians [27]) will hold true, even for larger order minima, m. We can simplify equations (2.1) and (2.2) into a single equation showing the direct relationship between a and y_m.

$$\frac{m\lambda}{a} = \frac{y_m}{L} \, . \tag{2.3}$$

Based on equation (2.3), the smaller the object length, a, the larger the spacing between the minima, y_m. This means that the diffraction pattern generated by the width of the worm will be spaced further apart compared to the diffraction pattern generated by the length of the worm. A depiction of this can be seen in Figure 2.4.

Far field diffraction using visible, coherent laser light can be observed if the diffracting object's dimensions are comparable to the wavelength of the light used. The sample size consideration can be shown by applying a limiting condition to the equations for single slit diffraction. The limiting condition is that the width of the central maximum, w, exactly equals the slit width (meaning the spreading angle, θ, will be close to zero). Applying this condition to equation (2.3) gives an equation for the largest sample size, a_c, that will cause visible diffraction.

$$\begin{aligned} \frac{\lambda}{a} &= \frac{\mathrm{w}}{2L} \\ \mathrm{w} &= a = a_c \\ a_c &= \sqrt{2L\lambda} \, . \end{aligned} \tag{2.4}$$

For a more exact measure of the limiting size of a sample for a particular setup, you need only to know the distance from the sample to the imaging screen, L, and the wavelength of

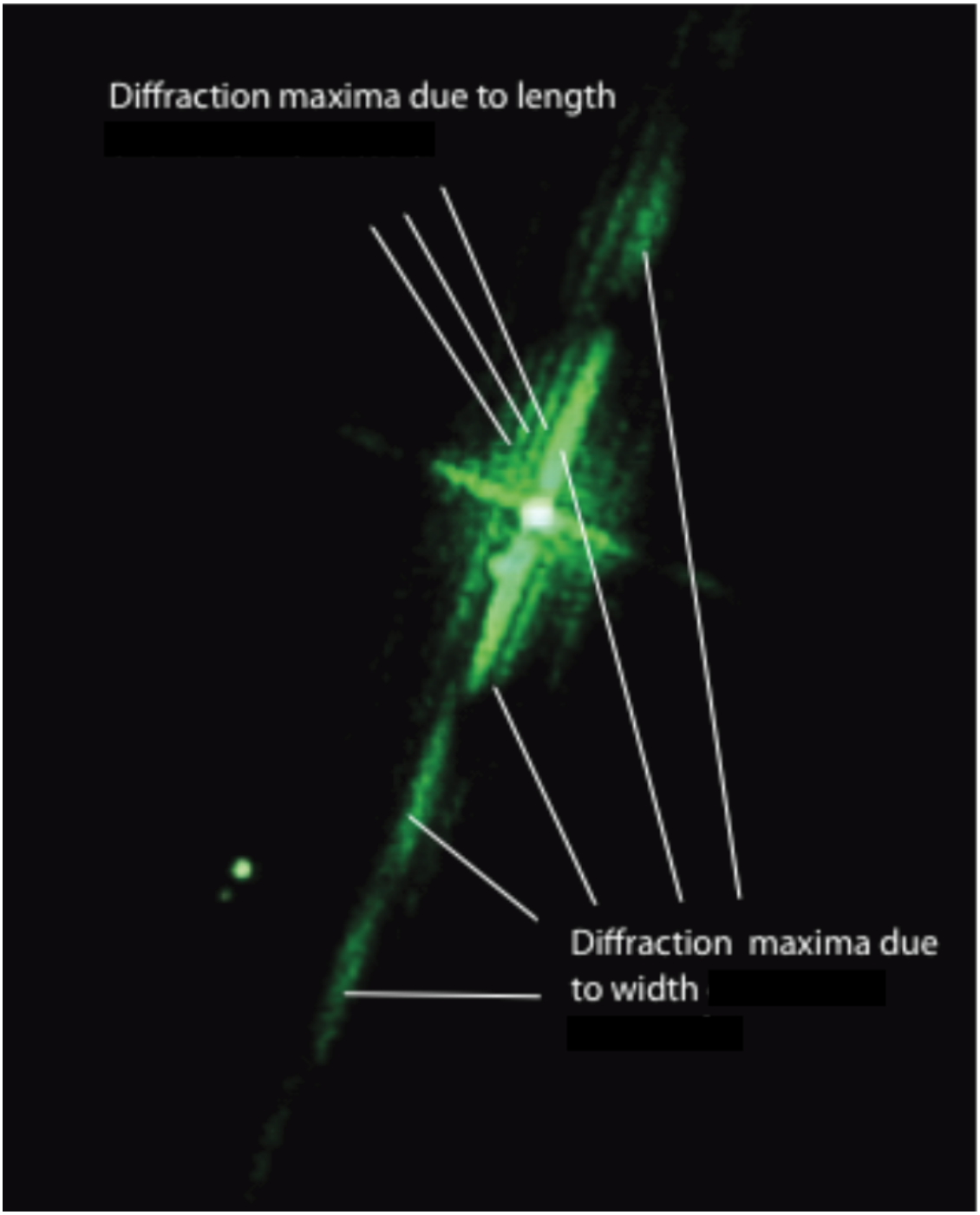


Figure 2.4 Experimental example of the different diffraction patterns present from both the width of the worm (spread out single slit pattern) and the length of the worm (more closely spaced single slit diffraction pattern) [28].

the laser being used, λ, to calculate the maximum size of a sample where diffraction will be observed, a_c. The typical value of a_c is approximately 1 mm for $L = 1$ m and λ in the visible range.

Based on the size considerations mentioned in the previous paragraph as well as the width and length of a *C. elegans*, we will be able to get diffraction arising from all parts of the worm if the laser beam diameter is larger than the length of the worm. Placing the worm within a laser beam larger than its dimensions will provide data by oversampling, which is when the ratio of the photons that are not incident on the worm are at least 4 times larger than the photons that do interact with the worm [29]. Oversampling is required to retrieve all the phases, which are needed to fully describe the worm's dynamics. However, too large of a beam will drown out the diffraction pattern. To ensure that we have oversampling without drowning the signal, the size of the laser beam diameter is set around 3 mm.

Considering the size difference between the length and width of the worms, we can then obtain a diffraction pattern on a screen in the far-field such as the example shown in Figure 2.4. However, going forward, we will look at the intensity versus time obtained from a photodiode at just a single point in the diffraction pattern, as illustrated in Figure 2.3.

Chapter 3

Fourier Transform

Many topics in the physical sciences relate to vibrations, waves, and oscillations. If we want to understand these phenomena, we need to have a way to extract the various frequencies present, which is where Fourier Transforms (FT) come in. Within this thesis, we will use the FT in two main ways. The first is interpreting diffraction pattern intensities as a FT, which will be described in the section (3.2). The second is calculating the broad frequency spectrum of the swimming *C. elegans*, which is discussed in Section (3.3) and at the beginning of Chapter 4. We will start with the basic definition of FT and build from there.

3.1 Fourier Transforms

Fourier transforms are used to represent non-periodic functions by assuming infinitely long periods for the integration ($-\infty$ to ∞). For non-periodic functions, the fundamental frequency approaches zero and the harmonics become more closely packed. In exponential form, a single, complex valued function, $\Phi(\nu)$, gives the amplitude of respective waves with frequency of ν as well as their phase, hence a frequency spectrum.

$$F(t) = \int_{-\infty}^{\infty} \Phi(\nu) e^{-2\pi i \nu t} \, d\nu \tag{3.1}$$

Based on the Fourier Inversion Theorem, the function $\Phi(\nu)$ is related to $F(t)$ in the following way:

$$\Phi(\nu) = \int_{-\infty}^{\infty} F(t) e^{2\pi i \nu t} \, dt \tag{3.2}$$

The variables frequency and time (ν and t) here are conjugate variables, and they form a 'Fourier Pair' in reciprocal spaces (equations (3.1) and (3.2)). This Fourier Pair is used when we use FT to extract frequencies and phases from our time series (Section 3.3). With regards to applying FT to diffraction, we use the Fourier Pair of x (position) and $p = \sin\theta/\lambda$ (diffraction angle divided by wavelength). The following sections will detail the two main ways that we will use FT: Diffraction (3.2) and Frequency Spectrum Calculation (3.3).

3.2 Applications to Diffraction

The electric field at a point, and hence the intensity distribution, within a far-field (Fraunhofer) diffraction pattern can be calculated using an FT. As the laser beam propagates along the optical axis (z-axis, as shown in Figure 2.3), it encounters the diffracting object. Based on Huygens' principle, we have many new sources arising from that interaction, all of which can interfere either constructively or destructively. The resulting electric field at a point P, $\widetilde{E}$, in the far field can be found using the following equation relating to the angle, θ, with respect to the optical axis: [30]

$$\widetilde{E} = E_0 e^{-2\pi i r_0/\lambda} \int_{-\infty}^{\infty} A(x) e^{2\pi i p x}\, dx\ . \tag{3.3}$$

E_0 is the initial amplitude of the electric field at the position of the slit. r_0 is the distance from the center of the slit to the point of interest P. λ is the wavelength of the incident light on the slit. The Fourier pair of x and p are position and sine of diffraction angle, θ, divided by λ, respectively. $A(x)$ is the 'aperture function', which varies depending on the type of obstruction in the propagation path of the laser beam. For a single slit of slit width a, the

aperture function is a top hat function, $\Pi_a(x)$.

$$\Pi_a(x) = \begin{cases} 0, & |x| > a/2 \\ 1, & |x| < a/2 \end{cases} \tag{3.4}$$

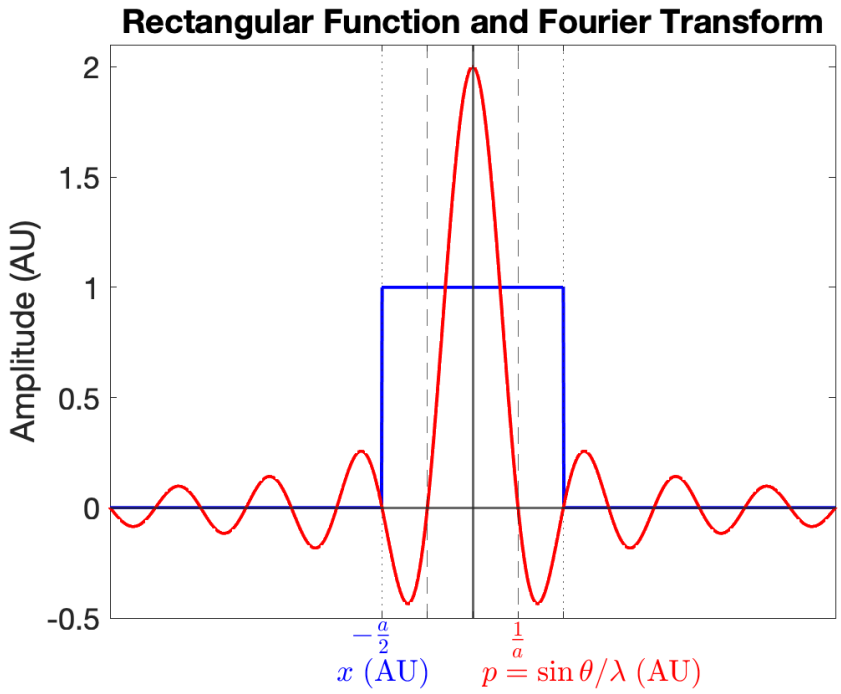


Figure 3.1 Example aperture function for a single slit (equation (3.4)) in blue with width a. The Fourier Transform (equation (3.6)) is shown in red. We can notice that if we make the width, a, larger and larger, then the width of the Fourier Transform's central peak gets smaller and smaller. This is reflective that larger slits/objects cause tighter diffraction patterns, just as observed for our worm in Figure 2.4

Since the ratio between the worm's width and length is similar to that of a single slit (1/10), we can use the aperture function for a single slit to approximate the electric field at point P of the diffraction pattern [23]. This function is nonzero only for the width of the respective slit. If we place the center of the slit at the origin, then our bounds for the FT within equation (3.3) simplifies to $-a/2 \to a/2$, giving us the following:

$$\Phi(p) = \int_{-\infty}^{\infty} A(x)e^{2\pi ipx}\,dx = \int_{-\infty}^{\infty} \Pi_a(x)e^{2\pi ipx}\,dx = \int_{-a/2}^{a/2} e^{2\pi ipx}\,dx \tag{3.5}$$

$$= a\,\mathrm{sinc}(\pi pa)\ . \tag{3.6}$$

where we have defined $\mathrm{sinc}(x)$ as $\dfrac{\sin(x)}{x}$. Other sources may define $\mathrm{sinc}(x)$ as $\dfrac{\sin(\pi x)}{\pi x}$, such as MATLAB. [31] Both the original top hat function written in equation (3.4) and the Fourier transform of it found through equation (3.6) can be seen in Figure 3.1. Putting equation (3.6) into equation (3.3), then multiplying the resulting expression by its complex conjugate computes the intensity, $I(p)$, of the diffraction pattern [30].

$$I(p) \propto |\Phi(p)|^2 = |a\,\mathrm{sinc}(\pi pa)|^2 \tag{3.7}$$

A real worm will not remain a stationary, straight line while swimming, but rather, will undergo undulations similar to those depicted in Figure 1.2. To expand our approximation of the intensity, we include two different segments of the worm that have a distance apart, b, that oscillates with a frequency ν and amplitude b_0. Applying the shift theorem then gives us a time dependent equation for the intensity profile [30, 23].

$$F(x+b/2) + F(x-b/2) \rightleftarrows 2\Phi(p)\cos(2\pi p(b/2)) \tag{3.8}$$

$$I(p,t) \propto |2\Phi(p)\cos(\pi pb)|^2 = |2a\,\mathrm{sinc}(\pi pa)\cos(\pi p[b_0 sin(2\pi\nu t)])|^2 \tag{3.9}$$

With this intensity equation, we can approximate how the intensity of the diffraction pattern changes in space and time. Focusing on a single, fixed point in the diffraction pattern, as we do when placing a photodiode into the diffraction pattern, we can focus on the time-dependence of the intensity, $I(t)$, affected by the frequency of undulation of the swimming worm.

3.3 Applications to Time-Series: Experimental Data Sets

In this work, we also use a discrete FT (DFT) for temporal frequency analysis since our data consists of discrete points; to be more exact, the fast Fourier transform (FFT) provides a computationally efficient method to evaluate the frequencies of discrete data. An FFT is an algorithm to calculate a (DFT) with less computational operations. To illustrate this for our data, we will redefine our function $I(t)$ to be $y(t_\mu)$, which will be used throughout the rest of this thesis. Our experimental time-series of intensity values, $y(t_\mu)$, are recorded at equidistant time intervals, t_μ, where μ is our time index. With a time step of $\Delta t_{\rm exp}$, we can express any time recorded as follows:

$$t_\mu = (\mu - 1)\Delta t_{\rm exp} \tag{3.10}$$

with $\mu = 1, 2, \cdots N$, where N is the total number of observed time points. The total duration of the experimental time-series is therefore:

$$t_{\rm exp} = N\Delta t_{\rm exp} \ . \tag{3.11}$$

Typical values of $\Delta t_{\rm exp}$ and N for our data were ~ 0.000769 s and $\sim 60,000$ data points, respectively. Making our typical value for $t_{\rm exp}$ around 50 s as mentioned in chapter 2. The frequency spectrum that is generated for plotting the FFT is based on the length of the data, $t_{\rm exp}$, and our time step, $\Delta t_{\rm exp}$. The inverse of $t_{\rm exp}$ determines the frequency step, $\Delta\nu$, or minimum frequency resolution, and the inverse of $\Delta t_{\rm exp}$ determines the largest frequency, $\nu_{\rm max}$, *i.e.*,

$$\Delta\nu = \frac{1}{t_{\rm exp}} \tag{3.12}$$

$$\nu_{\rm max} = \frac{1}{\Delta t_{\rm exp}} \ . \tag{3.13}$$

This means that like our experimentally recorded time-series, our frequency spectrum is a sequence of equidistant points, ν_k, where

$$\nu_k = (k-1)\Delta\nu \tag{3.14}$$

with $k = 1, 2, \cdots N$. However, please take note that experimentally, the Nyquist sampling theorem denotes that the maximum observable frequency possible is half of $\nu_{\max}$ [32, 33], *i.e.*

$$\nu_{\max_{\exp}} = \frac{\nu_{\max}}{2} \tag{3.15}$$

$$k_{\exp} = 1, 2, \cdots (N/2+1) \tag{3.16}$$

With our times and frequencies defined, we are ready to calculate the DFT using the following method [30].

$$\hat{y}(\nu_k) = \sum_{\mu=1}^{N} y(t_\mu)\, e^{2\pi i(k-1)(\mu-1)/N} = \sum_{\mu=1}^{N} y(t_\mu)\, e^{2\pi i \nu_k t_\mu} \tag{3.17}$$

Each $\hat{y}(\nu_k)$ coefficient relates to the weight that a particular frequency, ν_k, appears in the data $y(t_\mu)$. The inverse FT that retrieves $y(t_\mu)$ is defined as follows [30]:

$$y(t_\mu) = \frac{1}{N}\sum_{k=1}^{N} \hat{y}(\nu_k)\, e^{-2\pi i(\mu-1)(k-1)/N} = \frac{1}{N}\sum_{k=1}^{N} \hat{y}(\nu_k)\, e^{-2\pi i t_\mu \nu_k} \tag{3.18}$$

Note that equation (3.17) defines $\hat{y}(\nu_k)$ for all integer values $k = 0, \pm1, \pm2, \cdots$, *i.e.*, an infinite sequence of FT coefficients $\hat{y}(\nu_k)$. However, this infinite sequence is redundant, since it has the periodicity property

$$\hat{y}(\nu_k) = \hat{y}(\nu_{k\pm N}) \quad \text{for all integer } k \tag{3.19}$$

or, equivalently:

$$\hat{y}(\nu_k) = \hat{y}(\nu_k \pm \nu_{\max}) \quad \text{for all integer } k. \tag{3.20}$$

Therefore, only a finite contiguous integer k-set, of length N, is needed to recover the original time-series values, $y(t_\mu)$, from the FT coefficients, as seen in equation (3.18).

For our experimental time-series data, $y(t_\mu)$ is real-valued, and, for any real-valued time-series, $y(t_\mu)$, the FT coefficients have the additional symmetry property

$$\hat{y}(-\nu_k) \equiv \hat{y}(\nu_{2-k}) = (\hat{y}(\nu_k))^* \quad \text{for all integer } k \tag{3.21}$$

where $(\hat{y}(\nu_k))^*$ denotes the complex conjugate of $\hat{y}(\nu_k)$. Using equations (3.11) - (3.14) and (3.20), this can also be written as

$$\hat{y}(\nu_{\max} - \nu_k) \equiv \hat{y}(\nu_{N+2-k}) = (\hat{y}(\nu_k))^* \quad \text{for all integer } k \tag{3.22}$$

or as

$$\hat{y}\left(\frac{\nu_{\max}}{2} - \nu_k\right) = \left(\hat{y}\left(\frac{\nu_{\max}}{2} + \nu_k\right)\right)^* \quad \text{for all integer } k. \tag{3.23}$$

In reverse, any time series $y(t_\mu)$ is guaranteed to be real-valued if its FT satisfies either equation (3.21), (3.22) or (3.23). This will be of some importance in the construction of surrogate time series, to be discussed in section 4.7.

A time-series with N values, requires N^2 operations to complete the calculation of the DFT, while an FFT takes only $N \log N$ operations to complete. To accomplish this reduction in operations, we can break the value N into smaller chunks. The most common algorithm for doing this is the Cooley–Tukey algorithm, which recursively breaks apart the original time-series with N values into a product of smaller chunks ($N = N_1 N_2$) [34, 35, 36]. Breaking apart N into smaller chunks can be done in many ways, one of which is $N_1 = N_2 = N/2$, which only works well for powers-of-two. To help our FT work more smoothly, we used the

technique of zero padding to lengthen data to the next integer power of 2 so that FFT can be applied. We primarily use MatLab to conduct our calculations, which calculates FFT using a library called the Fastest Fourier Transform in the West (FFTW) [37, 38]. FFTW is a versatile, free C subroutine library for computing DFT.

Squaring the result of an FT gives rise to the power spectral density, which is sometimes called merely the power spectrum, $S(\nu_k)$.

$$S(\nu_k) = |\hat{y}(\nu_k)|^2 = \hat{y}^*(\nu_k)\ \hat{y}(\nu_k) \tag{3.24}$$

where $\hat{y}^*(\nu_k)$ is the complex conjugate of $\hat{y}(\nu_k)$. For our dynamic diffraction data, the power spectrum relates to the energy per time of the analyzed system, hence also relates to the measured intensity. An example of a calculated power spectrum can be seen in Figure 4.1, where a broad frequency spectrum is obtained when calculating the FT of a time-series. This is a first indicator, or marker, of chaotic dynamics. The definition of chaos as well as chaotic markers will be discussed in detail in Chapter 4.

Chapter 4

Analysis of Complex Systems

Based on our data analysis results detailed in this chapter, we find that the dynamic diffraction time-series obtained from freely swimming *C. elegans* is nonlinear, deterministic, aperiodic, bounded, and oscillatory. Several of these characteristics are also criteria for complexity. If the system satisfies the following four criteria as listed by David Feldman [2], then chaos exists:

1. deterministic
2. orbits are aperiodic
3. orbits are bounded
4. sensitive dependence on initial conditions

There are several markers of chaos that help determine if the above criteria are satisfied. These markers help to distinguish deterministic time-series from non-deterministic signals such as noise or random signals. Note in this context that random means an outcome that is patternless and incompressible as explained in detail in [39]. A time-series that is patternless, hence random, is neither short-term nor long-term predictable. Technically, deterministic chaos can still be considered random since it is incompressible due to sensitivity on initial conditions [39]. The main difference is that deterministic chaos is short-term predictable from initial conditions.

The following sections will describe calculating markers of chaos and the significance of these markers in proving the statements above, such as a broad frequency spectrum (Section 4.1), largest Lyapunov exponent (LLE) (Section 4.4), and correlation dimension (Section 4.5), while also detailing the intermediate steps to measure/calculate the markers, such as mutual information (MI) (Section 4.2) and false nearest neighbors (FNN) (Section 4.3). Graphical representations such as lag plots (Section 4.2) and recurrence plots (Section 4.6) are another way to show that these criteria are met. We will also do an additional check for nonlinearity using surrogate data analysis (Section 4.7). We will start with defining variables associated with our analysis in the next section.

4.1 Frequency Spectrum

A seemingly random pattern can undergo a first check for chaotic motion by considering the frequency spectrum. We use the FT of the experimentally obtained time-series, $y(t_\mu)$, and calculate the power spectrum, $S(\nu_k)$, as described in section 3.3. For a periodic system with a finite discrete set of frequencies present in the signal, there will be distinct peaks appearing at those frequencies. On the other hand, the power spectrum of a complex dynamic system with chaotic markers will reveal a broad frequency spectrum such as that seen on the right in Figure 4.1. Broad frequency peaks in the *C. elegans* locomotion are due to some frequencies being closer to each other than our minimum resolution of $1/t_{\text{exp}} = 1/(50\text{ s}) = 0.02$ Hz. This slight variability in frequencies is caused by the gait of the worm not exactly replicating itself as the worm oscillates within the laser beam, much like the chaos of human locomotion [40]. The mean or average frequency, ν_{avg}, of the frequency values, ν_k, can be obtained through:

$$\nu_{\text{avg}} = \frac{\sum_k^N \nu_k \; S(\nu_k)}{\sum_k^N \; S(\nu_k)} \; . \tag{4.1}$$

Then the inverse of the average frequency gives us the average period of the dynamics.

$$T_{\text{avg}} = \frac{1}{\nu_{\text{avg}}} \tag{4.2}$$

This average frequency and average period will be used in later sections such as calculating LLE (section 4.4) and surrogate data analysis (section 4.7).

This broad frequency spectrum, such as that plotted in Figure 4.1, is our first indication that the dynamics of *C. elegans* locomotion is complex because the slight variability in the gait satisfies the chaos criteria of aperiodic phase space orbits (see next section). To determine if other chaotic markers are present, further analysis can be done and will be detailed in the following sections.

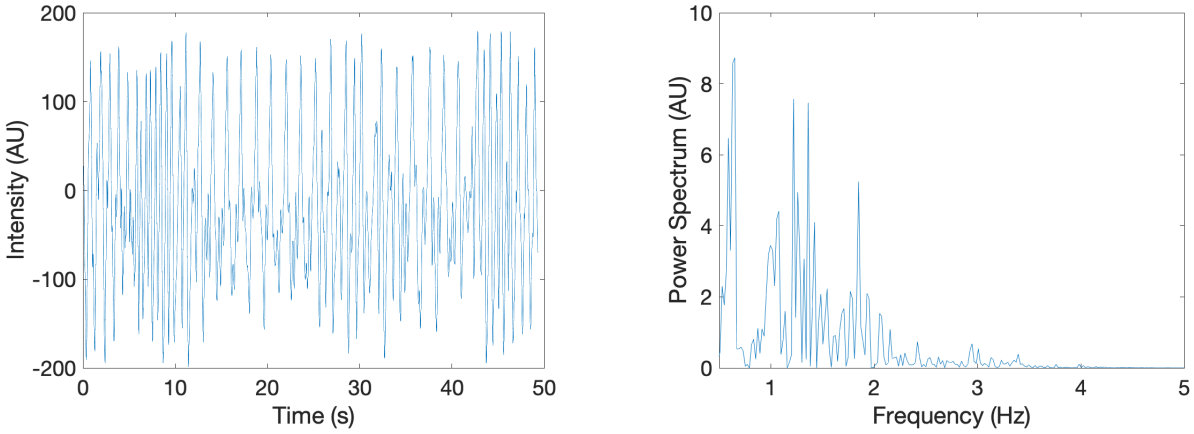


Figure 4.1 Wildtype time-series, $y(t_\mu)$, with its corresponding power spectrum, $S(\nu_k)$.
Left: Original experimental intensity versus time.
Right: Resulting broad frequency spectrum from FFT ($\nu_{\text{avg}} = 1.32$ Hz).

4.2 Phase Space Generation

In this section, we will define phase space and visualize it using lag plots. We choose to do this next because lag plots are a powerful visualization that can single-handedly prove that all four criteria for chaos are satisfied for a complex system. To construct the lag plot, we

need to employ Takens (1981) embedding theorem [3] and mutual information (MI) [41, 42] to generate additional components from our original time-series $y(t_\mu)$. We will begin with an explanation of phase space, phase trajectories, and orbits.

Phase space is a multidimensional space that represents each state of the system as a unique point. This is done by assigning each axis of the space to a component that defines the physical system of interest. Therefore, the full phase space construction requires the number of axes to equal the number of components required to fully define the system. For example, we can define a 2D phase space for a pendulum by plotting angular displacement along one coordinate axis and angular velocity along the other. Adding another axis for angular acceleration generates a 3D plot of the phase space. Once the phase space is constructed, we can start at a point in phase space, and follow its evolution in time, in other words, move along its phase trajectory. Phase trajectories can refer to any contiguous segment of time points within phase space and can start and end at any point in phase space. As we move along a trajectory, the shape, structure, or curve generated in phase space is known as the orbit of the trajectory. A phase trajectory is defined as a function of time whose function values are points in phase space. The phase trajectory thus gives a complete description of how the state of the physical system evolves in time. In chaos theory, we usually focus on pairs of phase space points or pairs of trajectories. Finding neighboring phase space points and trajectories are important in our analysis and will be explained in more detail in future sections.

If we can observe and measure only one component of a system with a multi-dimensional phase space, then we need a method to generate the other components (phase space dimensions) that define the system. The process of generating all the phase space components is referred to as unfolding the dynamics. For a continuously defined time-dependent function, unfolding the dynamics can be accomplished by taking the time derivative of the original function until all components are generated, *i.e.*, when the derivative equals zero. Going back

to our pendulum example, taking the first and second derivative of the angular displacement gives us two additional components of angular velocity and angular acceleration, respectively. Then the state of the dynamics at a time t can be uniquely defined by a vector containing these three components at that time. When the time-dependent function is known only on a grid of discrete time points, such as for our dynamic diffraction data, then another technique is needed to extrapolate the additional components needed to fully unfold the dynamics in phase space, then generate the lag plot. The technique we will use to unfold our discrete time-series data is Takens (1981) embedding theorem, which states that constructing the phase space trajectories, using so-called delay embedding to reconstruct the components of a dynamical system, is equivalent to using derivatives [3]. A visual illustration that both taking the derivative and using delay embedding produce equivalent results for unfolding the dynamics can be seen through recurrence plots (Figure 4.11), which will be discussed in Section 4.6.

The process of delay embedding requires a lag, τ. The lag, τ, is chosen such that it is an integer multiple of Δt_{exp}, *i.e.*,

$$\tau = \Delta t_{\text{exp}} \mu_\tau \tag{4.3}$$

where μ_τ is a non-negative integer less than N that corresponds to the lag in units of data points. With a single time-series (a single component of dynamics vs. time) and a lag, τ, delay embedding generates a reconstruction that is a one-to-one mapping of the original state space dynamics by generating a higher, n dimensional vector time-series, $y^n(t_\gamma)$, from an experimental time-series, $y(t_\mu)$, where t_μ defines the experimentally observed time grid defined in equation (3.10):

$$y^n(t_\gamma) = \{(y(t_\gamma), y(t_\gamma + \tau), y(t_\gamma + 2\tau), \cdots\ y(t_\gamma + (n-1)\tau))\}\ . \tag{4.4}$$

Here $\gamma = 1, 2, \cdots N_\tau$ is the truncated index sequence of μ such that each component in

the vector time-series has the same length N_τ. The length of the truncated sub-sequences (generated components) N_τ depends on the lag in units of data points, μ_τ, and the desired final vector time-series dimension, n, *i.e.*,

$$N_\tau = N - (n-1)\mu_\tau \ . \tag{4.5}$$

In other words, the original time-series, $y(t_\mu)$, is a projection onto a single dimension and is being unfolded so that all dynamic variables are accounted for and ambiguities are removed. The new components generated from the vector time-series are reconstructions of the phase space, or variables, that affect the dynamics of the original system. Takens embedding theorem implies that this phase space reconstruction preserves mathematical properties such as the topology of the dynamics and the Lyapunov exponents (section 4.4), making it a powerful tool for unfolding the characteristics of the dynamical system [3].

With Takens embedding theorem and the delay embedding formula (equation (4.4)), we use a method to find the lag, τ, that maximizes the unfolding of the phase trajectory. There are a few techniques for finding an optimal lag, τ. As shown by Cellucci *et al.* [41], the technique of mutual information (MI) is comparatively best for finding the optimal time lag, τ, as it minimizes the shared information between lagged data time-series. The information, H_0 (detailed below), contained in a typical data point, φ, within $y(t_\gamma)$ is represented by the Shannon entropy (or information). We use the symbol H to denote information rather than I to avoid confusion with the intensity, I, defined in section 3.2. To compare the amount of shared information between $y(t_\gamma)$ and $y(t_\gamma + \tau)$, we first need to calculate two probabilities $p_{\varphi|0}$ and $p_{\psi|\tau}$, respectively, as defined below. Since our data were collected at finite intervals, we will define a finite-width bin centered at φ or ψ, respectively, with a width of $\Delta\varphi$ to calculate the probabilities, where $\Delta\varphi$ is defined below. Therefore, $p_{\varphi|0}$ is the probability that the value of $y(t_\gamma)$ will fall within a bin of width $\Delta\varphi$ centered around an intensity value

of φ. Similarly, $p_{\psi|\tau}$ is the probability that the value of $y(t_\gamma + \tau)$ will fall within a bin of width $\Delta\varphi$ centered around an intensity value of ψ. Hence we define a proximity threshold of $\varepsilon_\varphi = \dfrac{\Delta\varphi}{2}$ to assist in calculating the probabilities $p_{\varphi|0}$ and $p_{\psi|\tau}$.

$$p_{\varphi|0} = \frac{1}{N_\tau}\sum_{\gamma}^{N_\tau}\Theta\Big(\varepsilon_\varphi - \big|\varphi - y(t_\gamma)\big|\Big) \tag{4.6}$$

$$p_{\psi|\tau} = \frac{1}{N_\tau}\sum_{\gamma}^{N_\tau}\Theta\Big(\varepsilon_\varphi - \big|\psi - y(t_\gamma + \tau)\big|\Big) \tag{4.7}$$

where $|...|$ is the absolute value. Here, Θ is the Heaviside function, *i.e.*,

$$\Theta(y) := \begin{cases} 1, & y \geq 0 \\ 0, & y < 0 \end{cases} \tag{4.8}$$

The overlap of information between $y(t_\gamma)$ and $y(t_\gamma+\tau)$ relates to their individual probabilities to fall within $\Delta\varphi$ as well as the joint probability, $p_{\varphi|0,\psi|\tau}$. The joint probability for $y(t_\gamma)$ to fall within $\left(\varphi - \dfrac{\Delta\varphi}{2},\ \varphi + \dfrac{\Delta\varphi}{2}\right)$ and for $y(t_\gamma + \tau)$ to fall within $\left(\psi - \dfrac{\Delta\varphi}{2},\ \psi + \dfrac{\Delta\varphi}{2}\right)$ is given by:

$$p_{\varphi|0,\psi|\tau}(\tau) = \frac{1}{N_\tau}\sum_{\gamma}^{N_\tau}\Theta\Big(\varepsilon_\varphi - \big|\varphi - y(t_\gamma)\big|\Big)\,\Theta\Big(\varepsilon_\varphi - \big|\psi - y(t_\gamma + \tau)\big|\Big)\,. \tag{4.9}$$

To calculate the entropy (information), and later the mutual information, the values of φ and ψ are restricted to a discrete grid bounded by the minimum and maximum values of the original, un-truncated time-series, $y(t_\mu)$.

$$\varphi_{\min} = \psi_{\min} = \min_{\mu=1...N} y(t_\mu) \tag{4.10}$$

$$\varphi_{\max} = \psi_{\max} = \max_{\mu=1...N} y(t_\mu) \tag{4.11}$$

Therefore, the grid used for the entropy and mutual information summations is given by:

$$\Phi := \left\{ \varphi_{\min} + \frac{\Delta\varphi}{2}, \varphi_{\min} + \frac{3\Delta\varphi}{2}, \varphi_{\min} + \frac{5\Delta\varphi}{2}, \cdots, \varphi_{\max} - \frac{\Delta\varphi}{2} \right\} \tag{4.12}$$

because our defined bins are centered about the φ values. The width of each bin in terms of the min and max values of φ is defined as $\Delta\varphi := (\varphi_{\max} - \varphi_{\min})/N_{\text{bins}}$, where N_{bins} is the total number of bins and typically equals 256. Then the information contained in $y(t_\gamma)$ and $y(t_\gamma + \tau)$ are, respectively, given by: [43, 42]:

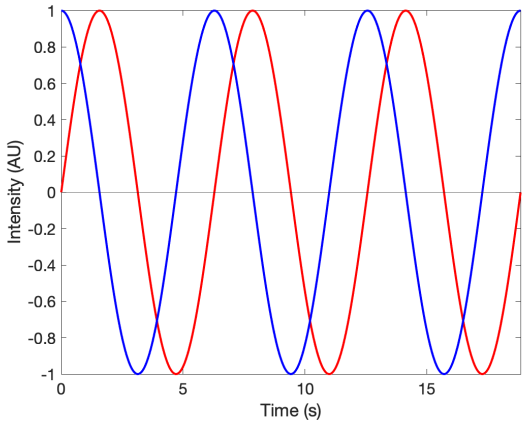
$$H_0 = -\sum_{\varphi \in \Phi} p_{\varphi|0} \log_2(p_{\varphi|0}) \tag{4.13}$$

$$H_\tau = -\sum_{\psi \in \Phi} p_{\psi|\tau} \log_2(p_{\psi|\tau}) \tag{4.14}$$

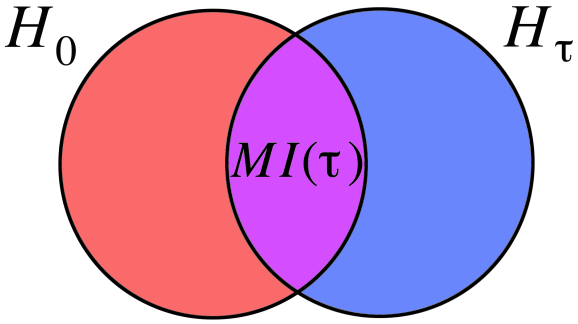
The entropy is analogous to information with the difference that information is usually calculated using base 2 in computer science accounting for a binary system. The mutual information measures the overlap of information of two data series; in this case, the same data series but lagged. This method to determine the lag measures overlap using a probabilistic measure of overlap rather than the more traditional way of minimizing linear correlation. Alternatively, the overlap in information of a second data series at a later time $(t_\gamma + \tau)$ can be measured by computing the mutual information between the first and second observations $y(t_\gamma)$ and $y(t_\gamma + \tau)$. We use one standard method for determining the"optimal" lag, τ, which is to minimize the mutual information, $MI(\tau)$, between the two time-series $y(t_\gamma)$ and $y(t_\gamma + \tau)$ as seen in Figure 4.2. The mutual information can be calculated as follows [42]:

$$MI(\tau) = -\sum_{\varphi,\psi \in \Phi} p_{\varphi|0,\psi|\tau}(\tau) \log_2 \left(\frac{p_{\varphi|0,\psi|\tau}(\tau)}{p_{\varphi|0} p_{\psi|\tau}} \right) \tag{4.15}$$

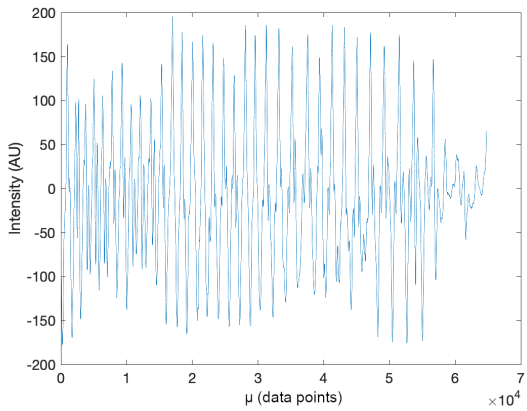
In the forgoing entropy and mutual information summations, the summand is set to zero if



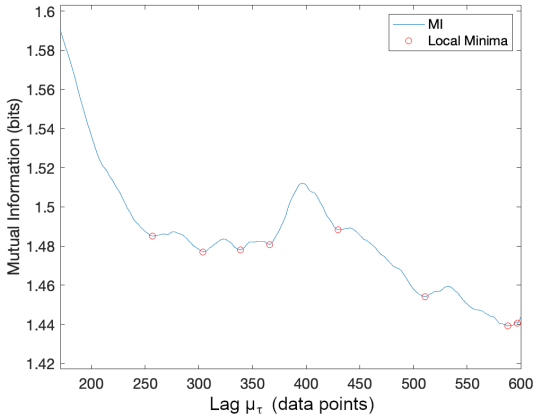
(a) $y(t_\gamma)$ in red. $y(t_\gamma + \tau)$ in blue.



(b) Information from $y(t_\gamma)$, H_0, in red. Information from $y(t_\gamma + \tau)$, H_τ, in blue. Information overlap, $MI(\tau)$, in purple.



(c) Experimental time-series.



(d) MI plot generated from (c).

Figure 4.2 Visualization of mutual information, $MI(\tau)$ [44]. (a) Shows a simple example of data sequence $y(t_\gamma)$ and a delayed data sequence $y(t_\gamma + \tau)$ on the same axis. (b) Venn diagram showing the information (entropy) from the data shown in (a). The optimal lag τ is found when there is the least amount of overlap (mutual information in purple) in information between un-lagged $y(t_\gamma)$ (H_0 denoted by red circle) and lagged $y(t_\gamma + \tau)$ (H_τ denoted by blue circle). (c) Showcases an experimental time-series of wildtype *C. elegans* locomotion obtained from the experimental setup described by Magnes et al. [1]. (d) Plot of MI versus lag (in units of data points) generated from experimental time-series in (c) where red circles highlight local minima. The first local minimum gives the "optimal" lag.

$p_{\varphi|0} = 0$ or $p_{\psi|\tau} = 0$, since

$$\lim_{p \to 0} p \log(p) = 0 \ .$$

Notice that $p_{\varphi|0} = 0$ for a given φ implies $p_{\varphi|0,\psi|\tau}(\tau) = 0$ for that φ and all ψ. Likewise, $p_{\psi|\tau} = 0$ for given ψ implies $p_{\varphi|0,\psi|\tau}(\tau) = 0$ for that ψ and all φ.

MI is a nonlinear correlation that quantifies an average of the information shared by two time-series, and does not depend on the values themselves. Rather, mutual information depends on the probabilities of these values occurring, while the joint probability depends on the lag τ. As shown in reference [41], the first local minimum in the $MI(\tau)$ plot is the best choice for the lag τ as it is the first point that maximizes the information added by the lagged series $y(t_\gamma + \tau)$ while minimizing the loss in data when creating an n number of time-series ($y^n(t_\gamma)$ components). A graph created from an experimental time-series illustrating this concept can be seen in Figure 4.2d. For the set of experimental time-series studied in this work, the optimal μ_τ values obtained were in the range of approximately 150 - 325 data points, corresponding to τ values in the range of around 0.11 - 0.25 seconds.

To view the dynamic trajectories of the time-series in phase space, we use the visual representation of a lag plot, where each axis is a component generated from delay embedding. This representation serves two purposes: (1) test if the time-series has structure (e.g. no crossings in phase space, bounded) and (2) if not structured, provide justification for needing higher order embedding (see next section) to fully resolve the trajectory. Combining the delay embedding formula ($y^n(t_\gamma)$) and MI's lag (τ) together to generate at least two components ($n = 2$), we get a lag plot shown in Figure 4.3. The result in 2D could be described as a picture of a wound-up yarn ball that appears to contain crossing trajectories; however, if we think about the analogy of a yarn ball, a 2D picture is not sufficient to fully characterize it because the yarn ball is a 3D object with a certain thickness to it, so not all of the yarn that appears to cross in 2D actually does, but rather, is separated by a finite amount of

space or thickness. Similarly, the trajectories pictured in a 2D plot do not actually cross but are a projection onto a 2D plane with false crossings that can be resolved with embedding in the proper dimension for the trajectories in question (see next section). In fact, for a deterministic system, the trajectories will never cross in phase space because each point is uniquely defined if embedded properly.

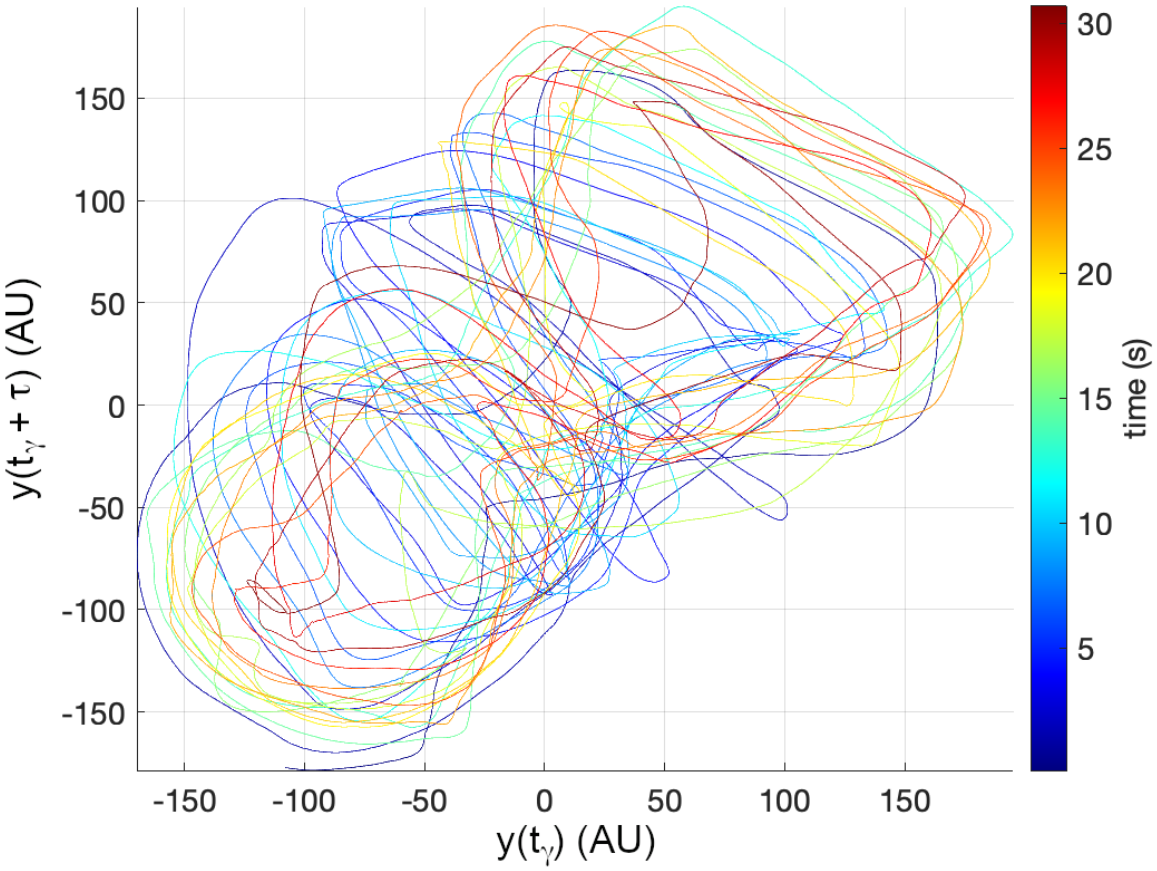


Figure 4.3 Example 2D projection of a lag plot generated from the dynamic diffraction time-series of a swimming *C. elegans* in Figure 4.2c [44]. The colors reflect a progression in time where dark blue reflects earlier t_γ values and dark red reflects later t_γ values. The first component $y(t_\gamma)$ is plotted on the x-axis, and the second component $y(t_\gamma + \tau)$ is plotted on the y-axis with $\tau = 0.1716$ s. Based on the apparent crossings in this projection, we need a higher embedding dimension than 2 to fully resolve the dynamics and eliminate false crossings. Determining the best embedding dimension will be discussed in section 4.3.

Re-plotting Figure 4.3 in 3D ($n = 3$) eliminates trajectory crossings, hence reveals false

crossings, indicating that an embedding dimension higher than 2 is needed to fully resolve the dynamics. For our data, a 3D lag plot satisfies all of David Feldman's chaos criteria listed at the beginning of this chapter. Determinism is satisfied because at $n = 3$, the trajectories of our data do not cross each other, thus every point in phase space is unique and deterministic. The orbits are aperiodic since there are no crossings in phase space - at least not on the time scales of our experimental observations, t_{exp}. The orbits of the trajectories are bounded in a finite region in the 2D lag plot in Figure 4.3. Using the color bar on the right-hand side to track the temporal evolution of the trajectories, sensitive dependence on initial conditions is seen to be satisfied because close points on different temporal lines eventually diverge.

Visual inspection of the 3D lag plot, generated using 3D plotting software, indicates that the 3D lag plot fully resolves the trajectories. That is, in the 3D lag plots, no crossings of trajectories are observed. Nevertheless, the required embedding dimension will be discussed in more detail in the following section (4.3).

4.3 Embedding Dimension

This section will detail how we determine the embedding dimension, n needed to generate a vector time-series $y^n(t_\gamma)$ that fully separates the phase space trajectories, hence unfolding the dynamics present from our original time-series $y(t_\mu)$. The value of n is the last value needed to fully implement delay embedding of our data to take advantage of Takens embedding theorem. We determine the embedding dimension n using the method of false nearest neighbors (FNN). FNN occur when neighboring trajectory points in phase space, $y^{\tilde{n}}(t_{\gamma_i})$ and $y^{\tilde{n}}(t_{\gamma_j})$, overlap in lower embedding dimensions, $\tilde{n}$, but separate at a higher embedding dimension n. The term overlap here means that some distance measure in phase space, defined below, falls below some small threshold value, ε_d. First we calculate the Euclidean distance,

$D_{\gamma_i,\gamma_j}(\tilde{n})$ between two points $y^{\tilde{n}}(t_{\gamma_i})$ and $y^{\tilde{n}}(t_{\gamma_j})$ using a norm $\left\| \cdots \right\|$ as defined below.

$$D_{\gamma_i,\gamma_j}(\tilde{n}) = \left\| y^{\tilde{n}}(t_{\gamma_i}) - y^{\tilde{n}}(t_{\gamma_j}) \right\| \tag{4.16}$$

$$= \sqrt{\sum_{c=1}^{\tilde{n}} \left[y(t_{\gamma_i} + (c-1)\tau) - y(t_{\gamma_j} + (c-1)\tau) \right]^2} \tag{4.17}$$

where $c = 1, 2, \cdots \tilde{n}$ and indexes the coordinates generated during delay embedding. Using $D_{\gamma_i,\gamma_j}(\tilde{n})$, the number of FNN along the vector time-series for embedding dimension $\tilde{n}$ are tabulated with the following quantity, FNN($\tilde{n}$), based on Kennel's algorithm [45, 41]:

$$\text{FNN}(\tilde{n}) = \sum_{\gamma_i=1,\gamma_j=\gamma_i+1}^{N_\tau} \Theta\left(\sqrt{\frac{(D_{\gamma_i,\gamma_j}(\tilde{n}+1))^2 - (D_{\gamma_i,\gamma_j}(\tilde{n}))^2}{(D_{\gamma_i,\gamma_j}(\tilde{n}))^2}} - \varepsilon_d \right) . \tag{4.18}$$

There are several other methods for calculating FNN [46, 47, 48, 49], but we are using Kennel's algorithm here, due to its demonstrated ability to reproducibly calculate accurate values of the largest Lyapunov exponent (LLE) (see next section) for various systems [45, 41]. Instead of the change in absolute Euclidean distance itself, equation (4.18) uses the relative difference of the squares of the Euclidean distance to quantify the change in separation between two trajectory points when the embedding dimension is increased from $\tilde{n}$ to $\tilde{n}+1$. The square-root term on the right-hand side of equation (4.18) provides a relative measure of the change of separation between the two trajectory points, and ε_d is a dimensionless quantity. A pair of trajectory points is identified as a FNN if that relative change in separation exceeds the threshold value ε_d. All FNNs are eliminated when the change in relative separation does not change anymore as the embedding dimension, $\tilde{n}$, is increased [50, 51]. The number of FNNs is plotted vs. embedding dimension, $\tilde{n}$, as shown on a percentage scale in Figure 4.4. The embedding dimension, $\tilde{n}$, where the FNN approaches zero or stabilizes is then the minimum embedding dimension, n, needed to embed the trajectory without FNNs.

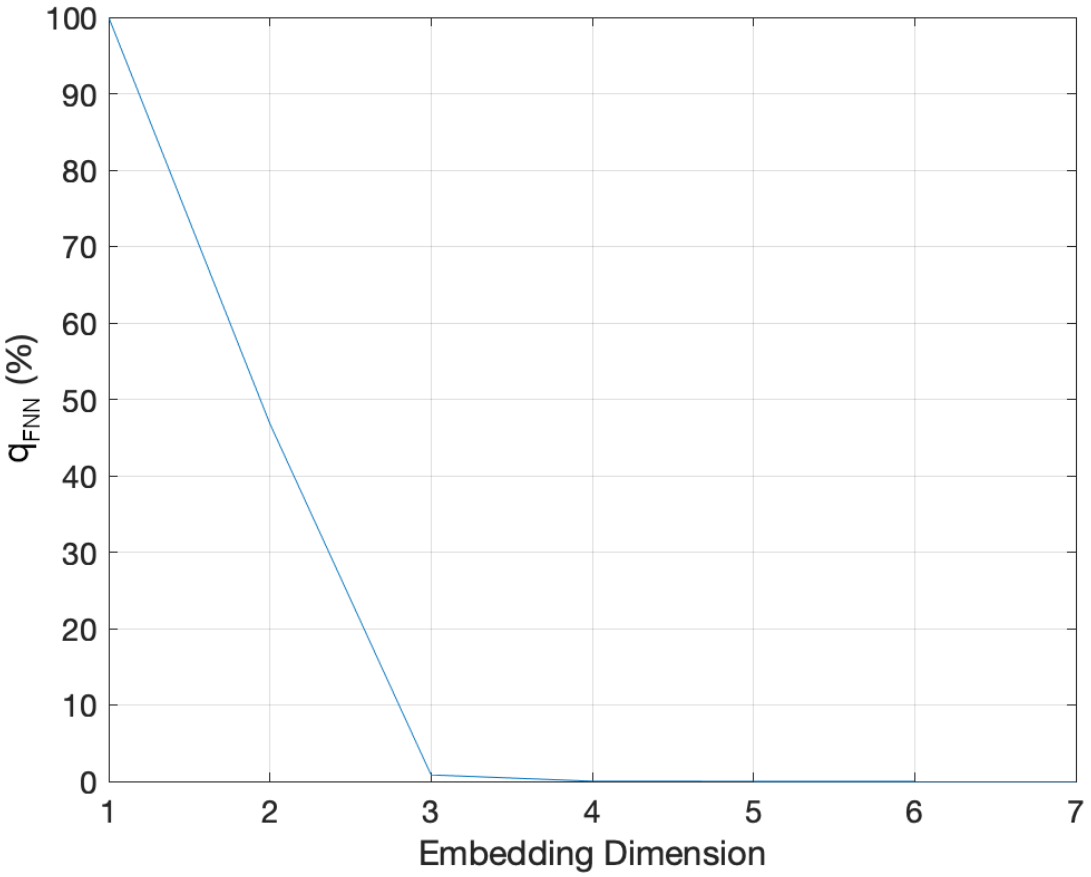


Figure 4.4 Percentage, $q_{\mathrm{FNN}} = \mathrm{FNN}(\tilde{n})/\mathrm{FNN}(1)$, of remaining FNN count versus embedding dimension, $\tilde{n}$, generated from the locomotory time-series in Figure 4.2c [44]. The FNN count is shown as a percentage of the reference count at $\tilde{n} = 1$. At $\tilde{n} = 3$, q_{FNN} falls to 0.85%, then 0.02% at $\tilde{n} = 4$. The FNN percentage falls to zero at higher embedding dimensions. We therefore use $n = 4$ as the minimal embedding dimension that completely unfolds the dynamics of *C. elegans* locomotion for this experimental time-series. The same minimal embedding dimension was found and used for all other experimental time-series studied.

Our goal is to determine the lowest embedding dimension that fully resolves the topology of the system since higher dimensions can amplify experimental noise and are computationally taxing [52]. We accomplish this by choosing the lowest embedding dimension n that eliminates FNN (false crossings) [50]. The threshold ε_d is chosen larger than the noise fluc-

tuations in the original time-series, $y(t_\mu)$, but small enough to fully separate the phase space trajectories. If ε_d is too small and within the noise, then the percentage of FNN can become greater than 100%, which is not possible. If ε_d is too large, a significant number of FNNs will not be identified, and the resulting n found will be smaller than needed to fully unfold the dynamics. Using the correlation dimension (see section 4.5, Figure 4.8) and recurrence plots (see section 4.6, Figure 4.13) can help verify that our embedding dimension n, and hence our threshold ε_d, are appropriate for our system.

For our data as seen in Figure 4.4, the embedding dimension stabilizes around dimensions 3 or 4. As a rule of thumb, to reliably embed a trajectory in embedding dimension n, 10^n data points are needed [53]; however, there are methods that have reliably estimated LLEs with less data points [54]. Since our time-series is typically larger than 10^4, we can safely embed in $n = 4$ indicating a low-dimensional complexity. Testing the stability of higher embedding dimension increases the confidence in our results. The FNN percentage in Figure 4.4 stabilizes at or near zero indicating that the data is relatively noise-free [50]. This same type of embedding dimension analysis has also been performed for multiple experimental time-series, both wildtype and roller worms. For all time-series analyzed, the resulting minimal embedding dimension was found to be around $n = 4$.

4.4 Largest Lyapunov Exponent

Once the time-series is embedded in the optimal dimension, n, based on the results from Figure 4.4, the next marker of chaos can be calculated: the largest Lyapunov exponent (LLE). The LLE quantifies the growth in separation between neighboring trajectories. Trajectories, of course, both converge and diverge in phase space; however, an LLE is a measure of how sensitive a system can be to initial conditions. This section will detail how to calculate the LLE starting with identifying initially neighboring points, then observing how they diverge in time.

With the lag, τ, and embedding dimension, n, from previous sections to completely apply delay embedding, the LLE is readily determined using the Rosenstein algorithm [55]. To start, we need to initially find pairs of vector time-series points $(y^n(t_{\gamma_i}),\, y^n(t_{\gamma_j}))$ in phase space, at a pair of observation time points $(t_{\gamma_i},\, t_{\gamma_j})$ that satisfy the following conditions:

$$T_{\text{avg}} \;\leq\; |t_{\gamma_i} - t_{\gamma_j}| \tag{4.19}$$

$$\left\| y^n(t_{\gamma_i}) - y^n(t_{\gamma_j}) \right\| \;=\; \min_{\gamma} \left\| y^n(t_{\gamma}) - y^n(t_{\gamma_j}) \right\| \tag{4.20}$$

$$\max(t_{\gamma_i}, t_{\gamma_j}) \;\leq\; t_{\text{exp}} - t_{LLE} \tag{4.21}$$

Here, T_{avg} refers to the average period obtained from the FT of the original time-series $y(t_\mu)$, as defined in equation (4.2). Also, t_{LLE} is an integer multiple of Δt_{exp} and denotes the total time duration of the neighboring trajectories for which a divergence (see below) is being calculated.

For a vector time-series of given length, N_τ, there will be $N_\tau - t_{LLE}/(\Delta t_{\text{exp}}) =: N_{LLE}$ time points t_{γ_j}. For each such t_{γ_j} there is one unique partner, t_{γ_i}, for which equations (4.19), (4.20), and (4.21) will be satisfied. For a given γ_j, the corresponding partner index, γ_i, will now be denoted as $\gamma_j{}'$, *i.e.*, $\gamma_j{}' \equiv \gamma_i$. We refer to these particular pairs of time points, $(t_{\gamma_j},\, t_{\gamma_j{}'})$, as the LLE pairs of the time-series. A pair of trajectories starting from $(t_{\gamma_j},\, t_{\gamma_j{}'})$ can be evolved in time for a duration of t_{LLE} to calculate the trajectory divergence, $D_{\text{div}}(t_{\gamma_j}, \bar{t}_\beta)$, defined by

$$D_{\text{div}}(t_{\gamma_j},\, \bar{t}_\beta) = \left\| y^n(t_{\gamma_j} + \bar{t}_\beta)) - y^n(t_{\gamma_j{}'} + \bar{t}_\beta)) \right\| \tag{4.22}$$

where

$$\bar{t}_\beta = \beta \Delta t_{\text{exp}} \qquad\qquad \beta = 0, 1, 2, \cdots t_{LLE}/(\Delta t_{\text{exp}})\ . \tag{4.23}$$

An illustration of the change in separation distance between an LLE pair, $(t_{\gamma_j},\, t_{\gamma_j{}'})$, as

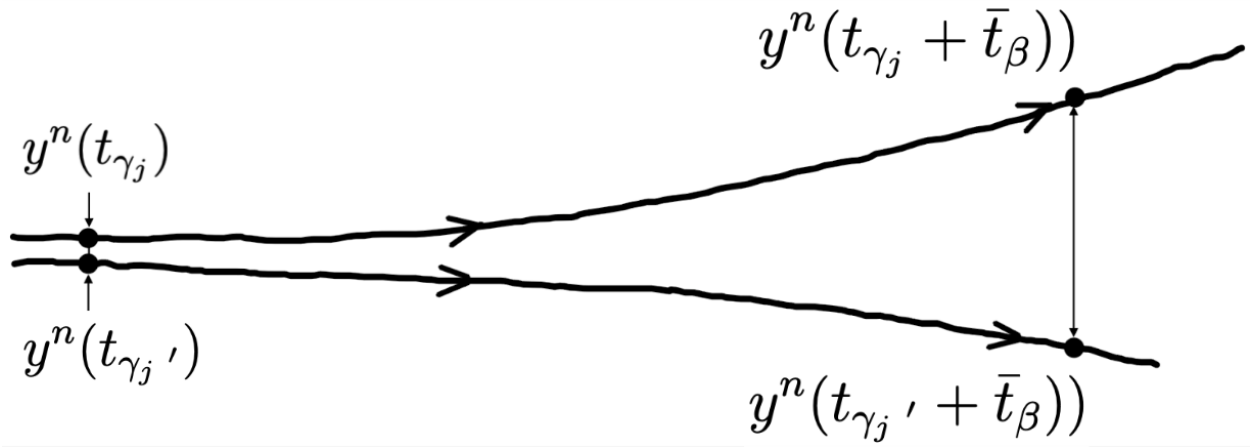


Figure 4.5 An illustration of an LLE pair , $(t_{\gamma_j},\ t_{\gamma_j\,\prime})$, that satisfies the criteria listed in equations (4.19) - (4.21) at some initial time $\bar{t}_\beta = 0$ s, shown on the left. Following each point along their respective trajectory for a set evolution time $\bar{t}_\beta$, we can calculate the new separation distance, or the divergence, $D_{\text{div}}(t_{\gamma_j},\ \bar{t}_\beta)$, such as the separation depicted on the right.

the trajectories evolve in time from an initial time $\bar{t}_\beta = 0$ s to a final $\bar{t}_\beta$ maximum value based on equation (4.23) is shown in Figure 4.5. For each LLE pair, $(t_{\gamma_j},\ t_{\gamma_j\,\prime})$, we now introduce an LLE value, $\bar{\lambda}_{\text{LLE}}(\gamma_j)$, based on the assumption that $D_{\text{div}}(t_{\gamma_j}, \bar{t}_\beta)$ converges or diverges exponentially as a function of $\bar{t}_\beta$ [55]:.

$$D_{\text{div}}(t_{\gamma_j},\ \bar{t}_\beta) \approx C(t_{\gamma_j})\ \exp\left(\bar{\lambda}_{\text{LLE}}(\gamma_j)\bar{t}_\beta\right) \tag{4.24}$$

where

$$C(t_{\gamma_j}) = \left\| y^n(t_{\gamma_j}) - y^n(t_{\gamma_j\,\prime}) \right\| \tag{4.25}$$

and

$$D_{\text{div}}(t_{\gamma_j},\ 0) \equiv C(t_{\gamma_j}). \tag{4.26}$$

The quantity $\bar{\lambda}_{\text{LLE}}(\gamma_j)$ in equation (4.24) represents a time-dependent value of the LLE at a particular time t_{γ_j} in the time-series. In the present study, we will not explicitly extract

$\bar{\lambda}_{\mathrm{LLE}}(\gamma_j)$ from the time-series, but rather, we will extract an average LLE by averaging $\bar{\lambda}_{\mathrm{LLE}}(\gamma_j)$ over all γ_j values, $\gamma_j = 1, 2, \cdots N_{LLE}$. To do so, we take the natural logarithm of both sides of equation (4.24), and obtain a linear relationship where the time-dependent LLE, $\bar{\lambda}_{\mathrm{LLE}}(\gamma_j)$, is the slope:

$$\ln\left(D_{\mathrm{div}}(t_{\gamma_j},\ \bar{t}_\beta)\right) \approx \ln(C(t_{\gamma_j})) + \bar{\lambda}_{\mathrm{LLE}}(\gamma_j)\ \bar{t}_\beta \tag{4.27}$$

Averaging equation (4.27) over all γ_j values, we get the following equation

$$L_{\mathrm{div}}(\bar{t}_\beta) = L_C + \lambda_{LLE}\bar{t}_\beta \tag{4.28}$$

where

$$L_{\mathrm{div}}(\bar{t}_\beta) := \frac{1}{N_{LLE}} \sum_{\gamma_j=1}^{N_{LLE}} \ln\left(D_{\mathrm{div}}(t_{\gamma_j},\ \bar{t}_\beta)\right) \tag{4.29}$$

$$L_C := \frac{1}{N_{LLE}} \sum_{\gamma_j=1}^{N_{LLE}} \ln(C(t_{\gamma_j})) \tag{4.30}$$

$$\lambda_{LLE} := \frac{1}{N_{LLE}} \sum_{\gamma_j=1}^{N_{LLE}} \bar{\lambda}_{\mathrm{LLE}}(\gamma_j)\ . \tag{4.31}$$

The quantity λ_{LLE} is the average of the time-dependent LLE quantities, $\bar{\lambda}_{\mathrm{LLE}}(\gamma_j)$, over the entire time-series. It is this quantity, λ_{LLE}, that we will report as the LLE of the time-series going forward. To extract λ_{LLE} from the time-series data, notice that $L_{\mathrm{div}}(\bar{t}_\beta)$ can be calculated for each $\bar{t}_\beta$, from the original time-series data, $y(t_\mu)$, transformed into the truncated vector time-series, $y^n(t_{\gamma_j})$, by equation (4.4), and then using equation (4.22) and (4.29).

Based on equation (4.28), which is equivalent to the assumption of exponential dependence stated in equation (4.24), we expect that a plot of $L_{\mathrm{div}}(\bar{t}_\beta)$ versus $\bar{t}_\beta$ will follow a

straight line whose slope is given by the average LLE of the time-series, λ_{LLE}. Therefore, we can extract the λ_{LLE} from the linear fit of the plot of $L_{\text{div}}(\bar{t}_\beta)$ versus $\bar{t}_\beta$. The linearization in equation (4.28) is only valid for short time durations. Therefore, the quantity, $L_{\text{div}}(\bar{t}_\beta)$ must be plotted versus time, $\bar{t}_\beta$, and visually inspected, to identify the early range of $\bar{t}_\beta$ where the linearization holds. We then extract λ_{LLE} by fitting a straight line to the $L_{\text{div}}(\bar{t}_\beta)$ data points across the $\bar{t}_\beta$ region where linearity holds [56].

The magnitude and sign of the LLE is a measure of convergence ($-$) or divergence ($+$). If the LLE value is negative, then trajectories converge, hence the system does not have a sensitive dependence on initial conditions. However, if the LLE is positive, then trajectories diverge exponentially. Obtaining a positive LLE indicates that the trajectories are diverging, which is a marker of chaos (criteria 4 [2]) but not sufficient alone to define a chaotic system [57]. An example of determining the LLE value for the *C. elegans* locomotion can be seen in Figure 4.6. Our positive LLE values obtained graphically such as in Figure 4.6 provide a quantifiable and reproducible measure of the locomotion while also indicating possible chaotic dynamics.

In comparison to our swimming *C. elegans* in a 3D environment, Ahamed et al. obtained a smaller LLE from position data for *C. elegans* crawling on a 2D agar surface, namely 0.69 s^{-1} [58]. However, our LLE was obtained in Figure 4.6 by fitting the beginning of our graph ($\bar{t}_\beta < 1$ s) to characterize a single thrashing cycle of a swimming worm. Ahamed's method applied a fit to a later region (from 1 s to 4 s) to characterize the trajectory of a crawling worm. If we apply a linear fit to the same region as Ahamed, then we obtain comparable results.

We used this method for numerous other experimental time-series, and in all cases studied, the LLE was found to be positive. This gives further evidence that *C. elegans* locomotion is chaotic. In this section, we used the entire time-series to calculate the LLE value; however, LLE values can also be calculated for sub-sequences of the time-series if different sections of

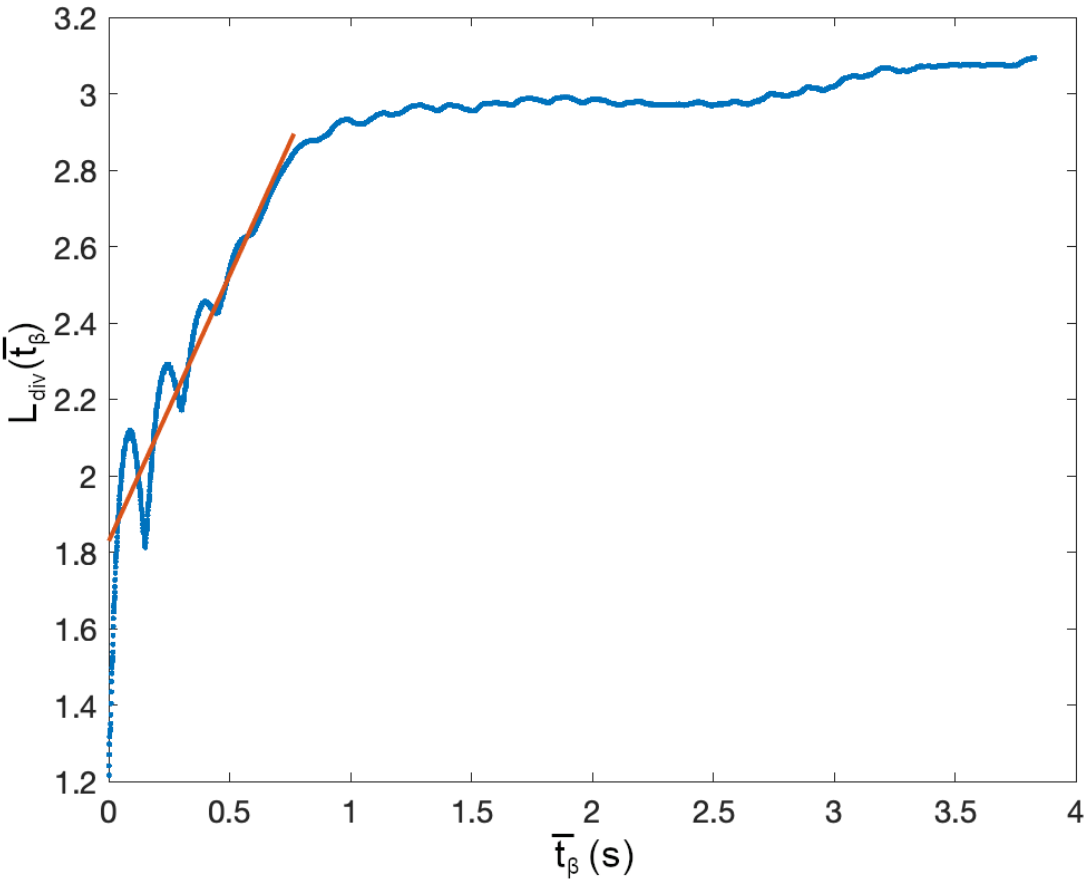


Figure 4.6 $L_{\text{div}}(\bar{t}_\beta)$, calculated from the experimental time-series data, $y(t_\mu)$, by equations (4.4), (4.22), and (4.29), is plotted versus evolution time $\bar{t}_\beta$ in blue for the time-series shown in Figure 4.2c [44]. By inspection of the blue curve, the time points from ~ 0 s to 1 s are identified as the region where the linearization holds, as stated in equation (4.28). Only these time points were included in the linear fit shown in red. The slope of the fitted red line gives a λ_{LLE} value of 1.39 ± 0.02 s^{-1}.

locomotion are determined. This process of determining multiple LLE from a single time-series will be described in more detail in section 4.6. A quantitative summary of these results is given in Table 4.2 in the last section in this chapter.

4.5 Correlation Dimension

Another measure of complexity is the correlation dimension [59, 60, 61]. The correlation dimension, d_{cor}, is a measure of complexity because it details the degrees of freedom and is a type of fractal dimension of the system. The correlation dimension takes into account the density and distribution of points in phase space. Again using the n-dimensional vector time-series generated by delay embedding (Equation 4.4), $y^n(t_\gamma)$, with a total number of N_τ points, the correlation dimension can be obtained by first calculating the following quantity, $C(\varepsilon_{\text{d}_2})$ [60]:

$$C(\varepsilon_{\text{d}_2}) = \lim_{N_\tau \to \infty} \frac{1}{N_\tau^2} \sum_{\gamma_i=1, \gamma_j=\gamma_i+1}^{N_\tau} \Theta\left(\varepsilon_{\text{d}_2} - ||y^n(t_{\gamma_i}) - y^n(t_{\gamma_j})||\right) \tag{4.32}$$

$$C(\varepsilon_{\text{d}_2}) \propto \varepsilon_{\text{d}_2}^{d_{\text{cor}}} \tag{4.33}$$

Equation (4.32) simply calculates the number of points within-range of point $y^n(t_{\gamma_i})$ using a Heaviside Function, Θ, as an indicator function. The acceptable proximity of two points is defined by a threshold, ε_{d_2}. A point $y^n(t_{\gamma_i})$ that is in a more densely packed region of the phase space orbit will generate a larger value compared to a point $y^n(t_{\gamma_i})$ located at a more sparsely packed region. To remove the dependence on the point position, $y^n(t_{\gamma_i})$, we take the average of these values to fully evaluate the correlation integral, $C(\varepsilon_{\text{d}_2})$, giving us a measure of how densely packed our phase space orbit is. $C(\varepsilon_{\text{d}_2})$ is proportional to $\varepsilon_{\text{d}_2}^{d_{\text{cor}}}$ through the relationship in equation (4.33). Therefore, the correlation dimension is found by plotting the log of the correlation integral (Equation (4.32)) versus the log of ε_{d_2} and finding the slope. This is illustrated in Figure 4.7 where $\ln\left(C(\varepsilon_{\text{d}_2})\right)$ is plotted vs. $\ln\left(\varepsilon_{\text{d}_2}\right)$ for the $C(\varepsilon_{\text{d}_2})$-results extracted by equation (4.32) from the vector time-series, $y^n(t_\gamma)$, of one single experimental time-series, $y(t_\mu)$, with embedding dimension $n = 4$. The slope of the straight line fitted to the graph in Figure 4.7 is the correlation dimension, d_{cor}, for that experimental time-series

and embedding dimension.

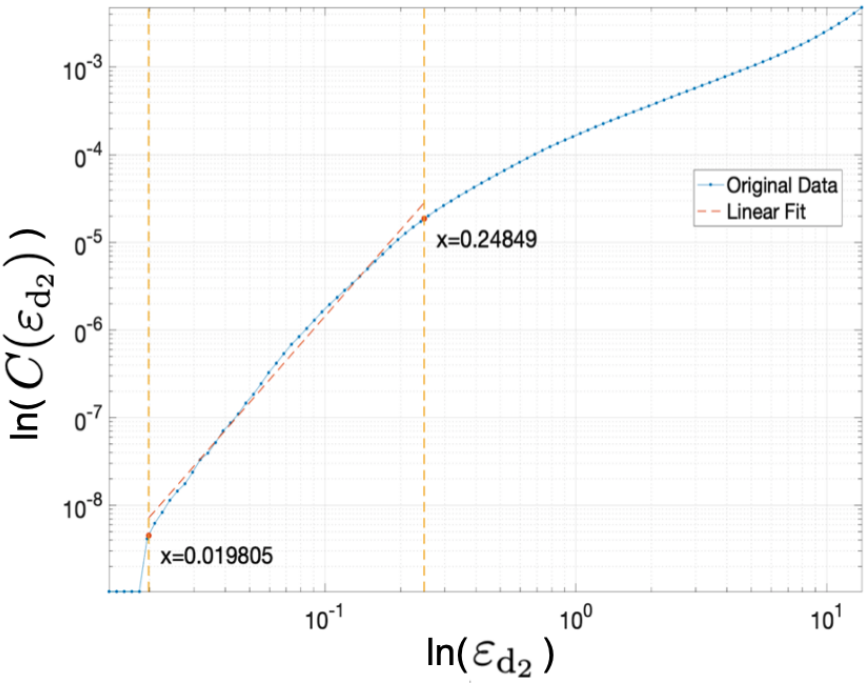


Figure 4.7 $\ln(C(\varepsilon_{\mathrm{d}_2}))$, calculated from the experimental time-series data, $y(t_\mu)$, by equations (4.4) and (4.32), is plotted versus $\ln(\varepsilon_{\mathrm{d}_2})$ in blue. By inspection of the blue curve, the points from $\sim$ 0.0198 to $\sim$ 0.2485 are identified as the region where the linearization holds, as insinuated in the proportionality in equation (4.33). Only these points were included in the linear fit shown in red, which illustrates how linear regions are chosen for generating correlation dimension values, d_{cor}, of our experimental time-series.

The correlation dimension calculation depends on the value of the embedding dimension dimension n. An experimental time-series can be shown to be random or deterministic depending on how the correlation dimension value changes with increased embedding dimension, n [62]. A time-series is random if the correlation dimension and embedding dimension continue to equal each other as the embedding dimension approaches infinity. If the correlation dimension stabilizes as the embedding dimension increases, then the data is deterministic (See Figure 4.8) [62]. Moreover, a correlation dimension above 2 is a necessary condition for a trajectory to be aperiodic, bounded, and not cross in phase space [63]. We find that time-series obtained from the dynamic diffraction (10 Wildtype and 18 Roller) has

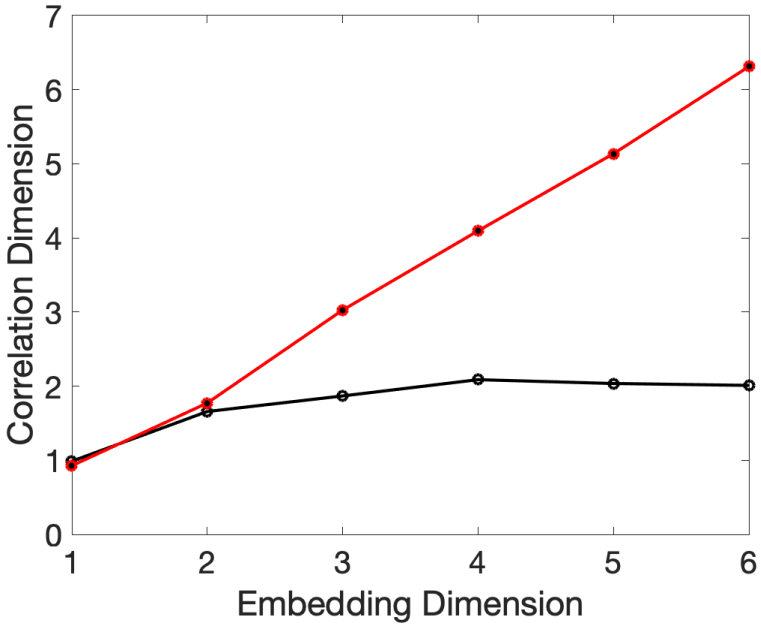


Figure 4.8 Plot of calculated correlation dimension as embedding dimension is increased. Random data (red) obtained from random.org increases linearly as the embedding dimension is increased, but our *C. elegans* data (black) stabilizes around a correlation dimension slightly above 2, indicating deterministic behavior [62]. Additionally, our locomotion data distinctly differs from the linear behavior exhibited by the random data at embedding dimension 3, which verifies that our locomotion data needs an embedding dimension of 3 or higher to unfold the dynamics.

an average correlation dimension $\sim 2.08 \pm 0.24$, for embedding dimension $n = 4$, which adds to the evidence of chaotic behavior in *C. elegans* locomotion. The possible non-integer value of the correlation dimension may indicate a fractal dimension, which could be due to chaotic dynamics. In our case, with a correlation dimension between 2 and 3, our phase space orbit has characteristics of 2D and 3D space [64]. Another way to graphically represent pairs of close points in phase space is through Recurrence plots, which will be discussed in more detail in section 4.6.

4.6 Recurrence Plots

More information about the locomotory system of *C. elegans* can be gained through other analytical methods for chaotic systems, such as distance matrix heatmaps and recurrence plots (RP), for visualization of phase space trajectory proximity [65, 66]. These visualization methods allow us to quickly identify essential dynamical structures in the time-series under study [67, 68, 69]. An RP can be used to investigate periodicity that may appear in a complex time-series [70, 71]. Given an experimental time-series, $y(t_\mu)$, of a nematode, we can construct a heat map or an RP. An RP is a binarized distance matrix that allows the visualization of characteristics in a time-series, including drifts in frequencies, recurrence rates, and complexity. Before discussing the application of heatmaps and RPs methods to experimental *C. elegans* locomotion time-series, I will first present a few illustrations of how distance matrix heatmaps and RPs allow us to identify and visualize essential dynamical features of a system, using very simple synthetic time-series data based on sinusoidal oscillation models. These illustrations will help us in applying these visualization methods in the analysis of real experimental time-series data.

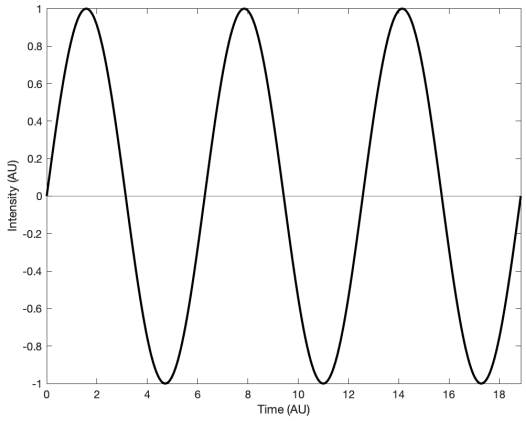
Before evaluating the full RP matrix, we first start by looking at the so-called distance matrix, $D_{\iota,\kappa}$, which calculates the distances between pairs of phase space points using the norm (see equation (4.17)).

$$D_{\iota,\kappa} = \left\| y^n(t_\iota) - y^n(t_\kappa) \right\| \qquad \iota, \kappa = 1, 2, \cdots, N_\tau \tag{4.34}$$

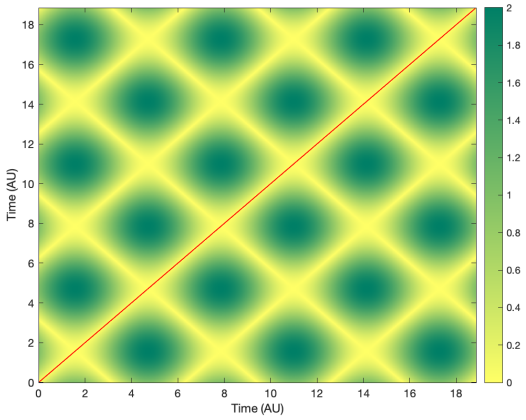
where N_τ is the total length of the vector time-series, $y^n(t_\iota)$, as defined in equation (4.5) for a time lag of τ. The distance matrix can be used to create two different graphical representations of recurrences in trajectories, $y^n(t_\iota)$: a heatmap and an RP. In each graphical representation, the graph plots time on both axes (x-axis = t_ι and y-axis = t_κ) allowing comparison of different times to each other. The distance matrix, $D_{\iota,\kappa}$, can be displayed as a

heatmap, showing the range and variation of distances between different trajectories within the vector time-series. As already explained earlier, the term trajectory here refers to any contiguous sub-series of the vector time-series. Each heatmap contains a color scale that represents the value of the matrix element $D_{\iota,\kappa}$ between any pair of time points t_ι and t_κ. All heatmaps shown in this section were calculated with 1D embedding ($n = 1$).

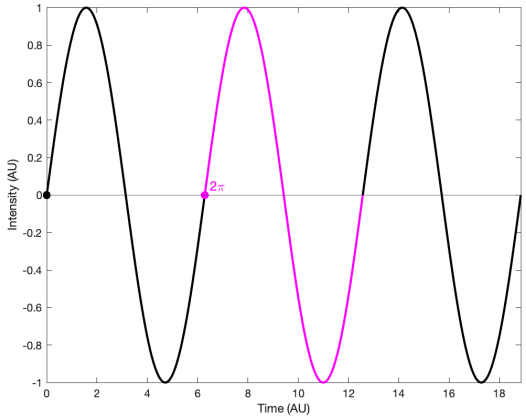
As an example, we can look at the 1D heatmap generated from a simple sine wave (Figure 4.9a) as seen in Figure 4.9b. In the heatmaps, yellow denotes a smaller difference between compared points, while green denotes a larger difference. Heatmaps (and as later defined RPs) reveal vector time-series points that are close together in phase space. These (t_ι, t_κ)-points are referred to as recurrences. For example, the yellow points, forming yellow lines, such as those in Figure 4.9b, are recurrences. Along the diagonal where $\iota = \kappa$ will always return a difference of zero, hence also a recurrence, and this line is called the line of identity (LOI), which is labeled with a red line in the heatmap/RP figures in this section. The other diagonal lines parallel to the LOI can be explained as two separate trajectories repeating over a time duration related to the length of the diagonal line. Take for instance a trajectory starting at $t = 0$ and another starting at $t = 2\pi$ (highlighted in pink in Figure 4.9c). If we move one time step at a time along those two paths simultaneously and compare their amplitudes, we calculate a difference of zero at every pair of times, which generates the diagonal line highlighted in pink in Figure 4.9d, parallel to the LOI. The (t_ι, t_κ)-points along this pink line are an example of so-called true recurrences because not only are the amplitudes the same, but also the slopes are the same at every pair of points compared. In this context, we distinguish between true recurrences and false recurrences. True recurrences are defined by the condition that both the signal values, $y^n(t_\iota)$ and $y^n(t_\kappa)$ are in close proximity and their time derivatives are also in close proximity. In contrast, false recurrences are defined by the condition that both the signal values, $y^n(t_\iota)$ and $y^n(t_\kappa)$ are in close proximity, but their time derivatives are substantially different.



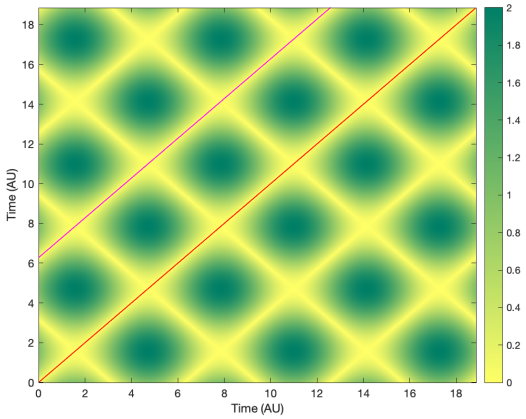
(a) Simple Sine Wave



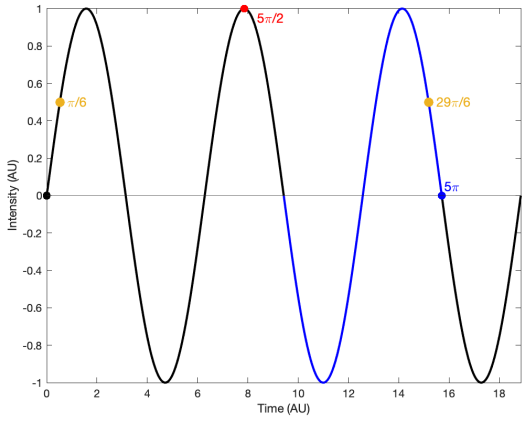
(b) 1D Heatmap of (a)



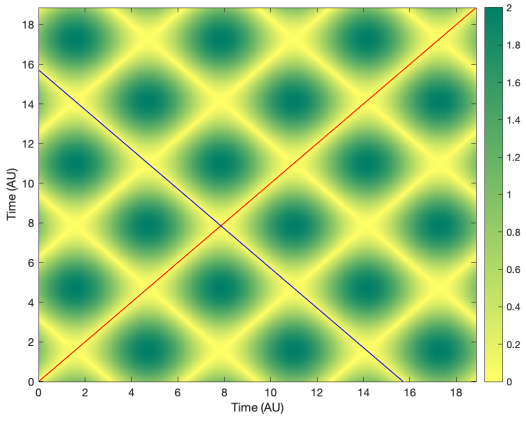
(c) Pink highlight of repeated state



(d) Recurrences generated from (c)



(e) Blue highlight of a false repeat



(f) False recurrences from (e)

Figure 4.9 1D Heatmap examples generated from a simple sine wave. Part (b) shows the 1D Heatmap with only the LOI highlighted in red. Parts (c) and (d) showcase a line parallel to the LOI and how it is generated in pink while parts (e) and (f) showcase a line perpendicular to the LOI and how it is generated in blue.

Diagonal lines that are perpendicular to the LOI can be explained as false recurrences and are a hint that a higher embedding dimension is needed to fully resolve the trajectories. This phenomenon can be explained by considering a trajectory that starts at 0 radians and another starting at 5π radians (highlighted in blue in Figure 4.9e). If we move one time step forward from $t = 0$ and one time step backwards from $t = 5\pi$ simultaneously and compare their amplitudes, we calculate a difference of zero at every pair of times, which generates the diagonal line highlighted in blue in Figure 4.9f. The points along this line are not exactly the same state because even though the amplitudes are the same, the slopes are different. For instance, two different points that are compared along the blue line are the two yellow dots in Figure 4.9e ($t = \pi/6$ and $t = 29\pi/6$), which have the same amplitude, but different slopes, meaning different states that would be eliminated from the heatmap if a higher embedding dimension was used. This process of eliminating the lines perpendicular to the LOI will be discussed later in this section and is visualized in Figure 4.11.

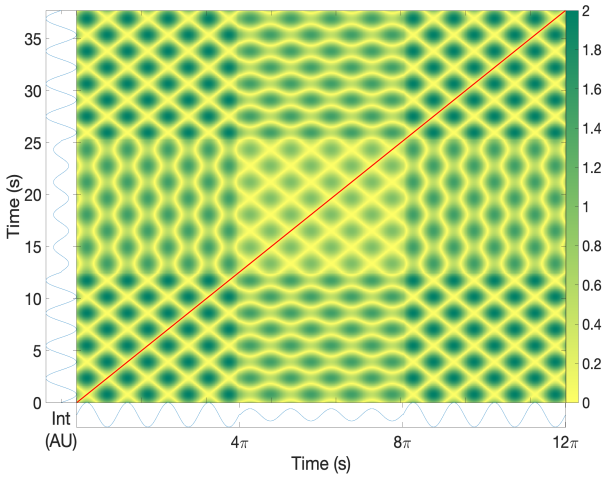
Before we eliminate the perpendicular lines, we will look at how changing amplitude and/or frequency can alter the structures present in an under-embedded heatmap (Figure 4.10). In Figure 4.10a, time-series sections of the same and unlike amplitudes are compared. Comparing time-series sections with mismatched amplitudes, such as amplitude jumps occurring at times 4π and 8π, reveals oscillating lines in the heatmap. Figure 4.10b shows how the RP changes by suddenly changing the frequency of the function. Sections comparing like frequencies form diamond shaped boxes which are of uniform size along both time axes and whose size relates to the period, whereas sections comparing unlike frequencies form boxes that are squished rhombi. In both cases, when there is a sudden shift in the function, or a sudden switching of behavior, there is a clear division, partitioning off the various sections. Extending to real signals, it is very unlikely that there will be a sudden shift in amplitude and/or frequency, but rather, a more gradual change. Figure 4.10c shows a gradually time dependent frequency, which causes the boxes in the heatmap to gradually stretch out along

the LOI axis. Figure 4.10d shows a heatmap generated by a signal with random amplitudes (random numbers obtained from https://www.random.org). While the box-like structures are still present, they are much more difficult to see. However, the RP of these time-series data, as defined below, with an appropriate threshold, ε, and proper, higher, embedding dimension, n, will unfold the dynamics and result in more sharply defined lines. These considerations will be explored next, then applied to experimental time-series.

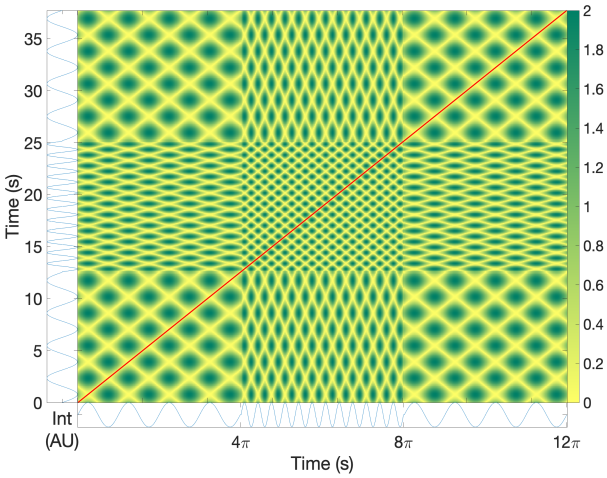
A heatmap is useful for visualizing a single time-series, but if we want to get quantitative information relating to chaos, we need additional steps, such as implementing the proper embedding dimension and sensible threshold. This leads to the second graphical representation, an RP. In constructing an RP, we apply a threshold, ε, to define the radius of the acceptable proximity tolerance between any pair of vector time-series points, by way of the so-called recurrence matrix R defined to have matrix elements [72]:

$$R_{\iota,\kappa} = \Theta\left(\varepsilon - \left\|y^n(t_\iota) - y^n(t_\kappa)\right\|\right) \qquad \iota, \kappa = 1, 2, \cdots, N_\tau \tag{4.35}$$

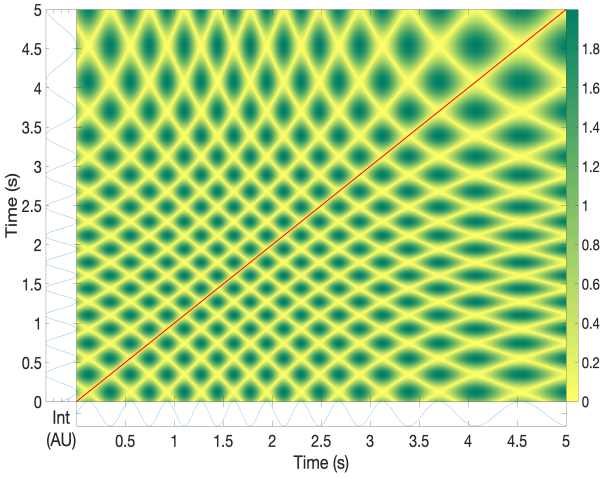
Here, Θ denotes the Heaviside function defined in equation (4.8), which ensures that the R-matrix elements take on only the values of 0 or 1. In corresponding RPs, shown in figures below, the R-matrix can thus be displayed in a black-and-white (BW) format: R-matrix elements with value 1 or 0 are displayed by black or white pixels, respectively, again with t_ι and t_κ plotted along the x- and y-axis, respectively. The embedding dimension, n, of the vector time-series, $y^n(t_\iota)$, entering into equation (4.35), is also referred to as the dimension of the RP. For example, in the following, the term "2D RP" refers to the RP of the recurrence matrix, $R_{\iota,\kappa}$, calculated from vector time-series, $y^n(t_\iota)$, with embedding dimension $n = 2$. The threshold, ε, should be small enough to account for critical features in the dynamics but large enough so that noise does not affect the RP. For our experimental time-series, the method we used to determine the threshold is to use a percentage of the maximum signal,



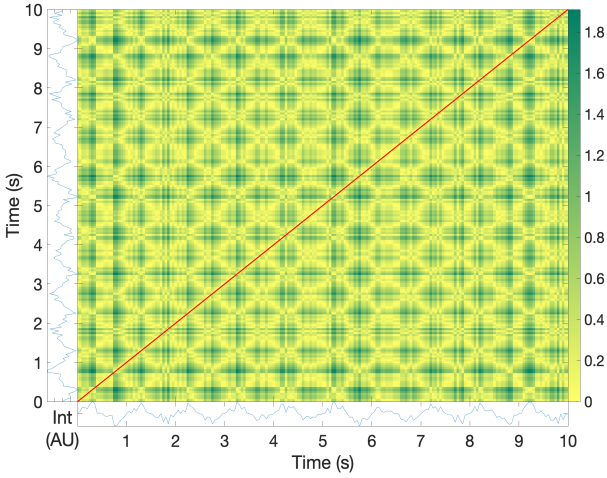
(a) Change in Amplitude



(b) Change in Frequency



(c) Time varying Frequency



(d) Random Amplitude

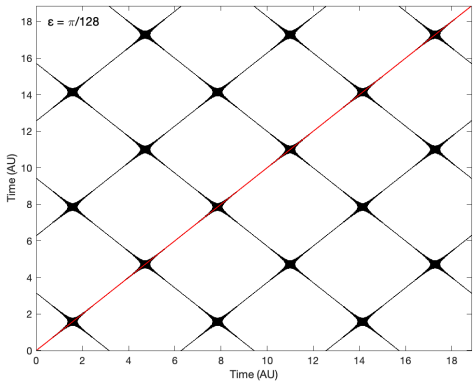
Figure 4.10 Changing various parameters to observe the effects on the 1D Heatmap for several synthetic sine waves. (a) shows amplitude reduced by half at $t = 4\pi$, then returned to the original amplitude at $t = 8\pi$. (b) shows the frequency tripled at $t = 4\pi$, then returned to the original frequency at $t = 8\pi$. (c) shows the frequency gradually changing with time based on the function $f = 2t - \cos(t)$. (d) shows constant frequency and random amplitudes.

or rather, a percentage, p_r, of the range of amplitudes, $\Delta\varphi_A = \varphi_{\max} - \varphi_{\min}$, where $\varphi_{\max}$ and

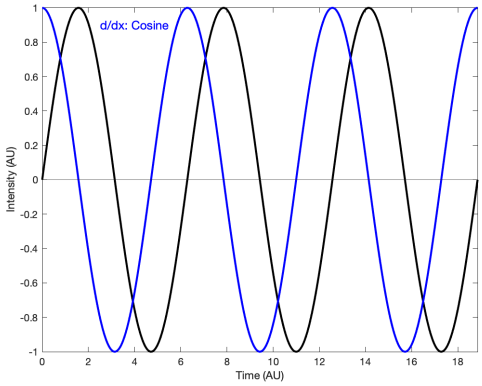
$\varphi_{\min}$ are defined in equations (4.10) and (4.11), respectively.

$$\varepsilon = p_r \Delta\varphi_A \tag{4.36}$$

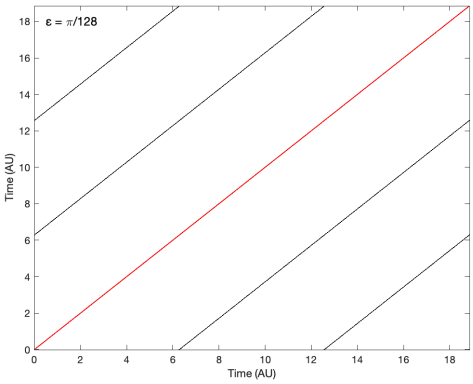
In Figure 4.11a, we plot the 1D RP of the recurrence matrix elements, $R_{\iota,\kappa}$, from equation (4.35), for our simple sine wave from Figure 4.9a. The resulting 1D RP is a black and white picture where our originally yellow recurrence lines have become thinner black lines. The lines perpendicular to the LOI are false recurrences that were observed in Figures 4.9e and 4.9f and are still present in the 1D RP. There are two ways to eliminate these false, perpendicular-line recurrences in the sine wave's 1D RP. One way is to combine the original time-series, $y(t_\iota)$, with its time derivative, $\dot{y}(t_\iota)$, to form a two-component vector time-series, thereby embedding the original time-series in a 2D phase space. This is illustrated in Figure 4.11c, where the perpendicular-lines have vanished. Hence, the 2D phase space separates the points with the same amplitude, but different slope. The second method to eliminate false recurrences is to generate a second vector component by applying a lag to the original time-series, $y(t_\iota)$, as done in Figure 4.11d. The lagged function is shown in pink and is synonymous to the derivative of the sine function shown in blue in Figure 4.11b. Both methods obtain the same final RP result. This is an illustration of Takens (1981) embedding theorem: embedding a time-series in a higher dimensional phase space by means of time lagging is entirely equivalent to an embedding that uses, instead, first and/or higher order derivatives to generate the components of a vector time-series. For inevitably noisy experimental time-series, it would be extremely difficult to produce accurate estimates for the time derivatives of $y(t_\mu)$, as required for a derivative based embedding. On the other hand, the experimental time-series can produce accurate results for all time lagged vector components, even in the presence of noise. For this reason, we have used exclusively the time lagged delay embedding method in analyses where embedding was required.



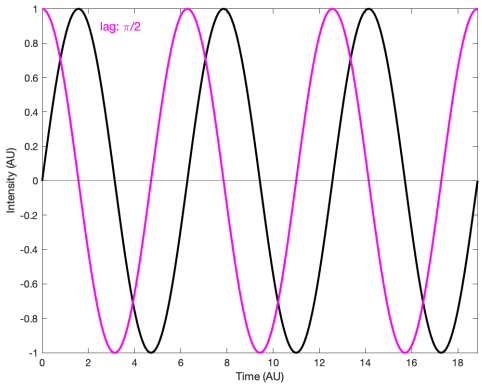
(a) Figure 4.9a 1D RP



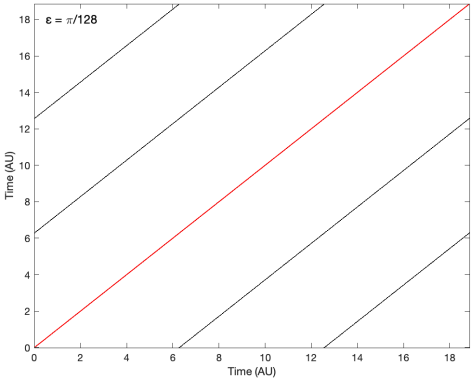
(b) 2nd component: d/dt = Cosine



(c) Figure 4.9a 2D RP: d/dt



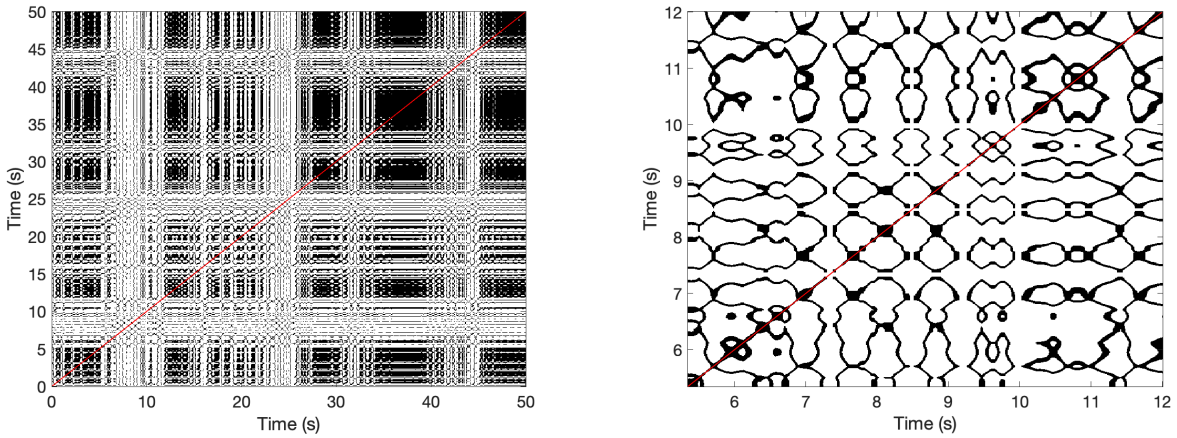
(d) 2nd component: $\tau = \pi/2$



(e) Figure 4.9a 2D RP: τ

Figure 4.11 RPs with $\varepsilon = \pi/128$ of the simple sine wave from Figure 4.9a. (a) shows the 1D RP of the sine curve. (b) plots the first (original sine) and second (derivative in blue) component to generate 2D embedding of the simple sine wave. These two curves will serve as the two input vector components of a 2D vector time-series to be displayed as an RP in (c). (d) plots the second method to generate 2D embedding (delay embedding) using the first (original sine) and second component (lagged function in pink) to generate 2D embedding of the simple sine wave, then the resulting 2D RP is plotted in (e). Both methods of 2D embedding generate the same RP plot.

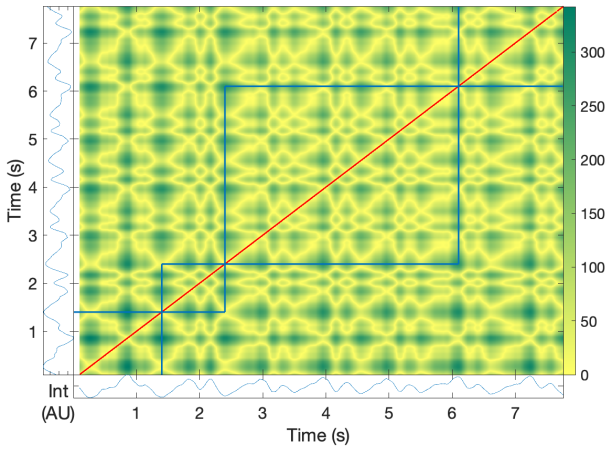
Having reviewed the basic concepts of the RP representation, we will now display and discuss RPs for real experimental *C. elegans* locomotion time-series. The 1D RP in Figure 4.12 shows how the locomotion of a swimming wildtype *C. elegans* evolves over time as well as recurrence feature evolution, such as those features in the Figure 4.10. Figure 4.12a shows that the locomotion evolves over time since the periodicity of the RP changes, while Figure 4.12b shows that the locomotion is not random with minimal noise but structured down to the smallest detail and never repeating exactly. The crossings of the recurrence features, aka lines perpendicular to the LOI, indicate that a higher embedding dimension than $n = 1$ is needed to resolve this time-series as we already observed through our FNN test.



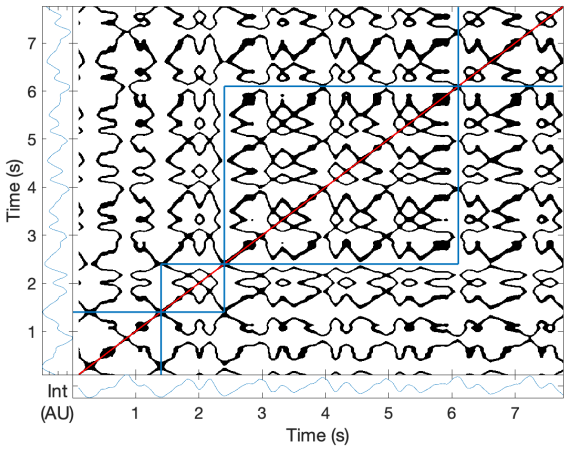
(a) Total time-series (b) Section from $t = 5.35 - 12$s

Figure 4.12 1D RPs of an experimental time-series of a swimming wildtype *C. elegans* on two different time scales [44]. Features emerge on various time scales in these RPs. The threshold, ε, used was $p_r = 10\%$ of the range of amplitudes, $\Delta\varphi_A$. (a) RP for a swimming *C. elegans*. The system is evolving over time with the patterns changing along the LOI at about 5.35 s, 12 s, 16 s, 25.3 s, 31.6 s, 34.2 s, 39.5 s, 41.6 s, and 45.2 s. (b) Highly resolved details of the swimming *C. elegans* over about 6 seconds (5.35 - 12s). Diagonal lines parallel to the LOI relate to the predictability of the system and can be used to determine the LLE. The diagonal from 8.48 s to 8.84 s has a length of 0.509 s, which relates to an LLE frequency of 1.96 s^{-1}. The frequency of about 2 Hz matches the swimming frequencies of wildtype *C. elegans* [15]. The crossing lines and lines perpendicular to the LOI indicate that a higher embedding dimension is needed to fully resolve the features.

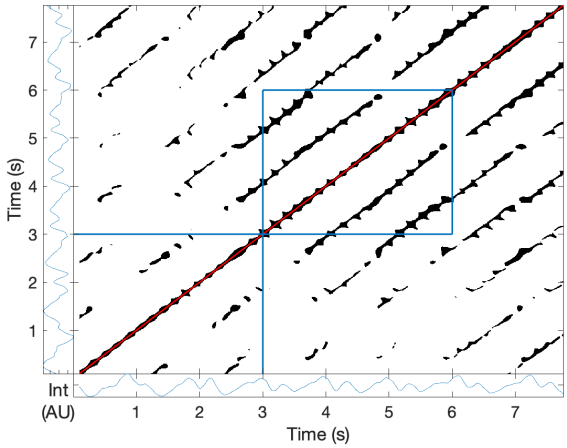
Figure 4.13 shows the contrast between 1D embedding and the proper embedding di-



(a) 1D Heatmap



(b) 1D RP

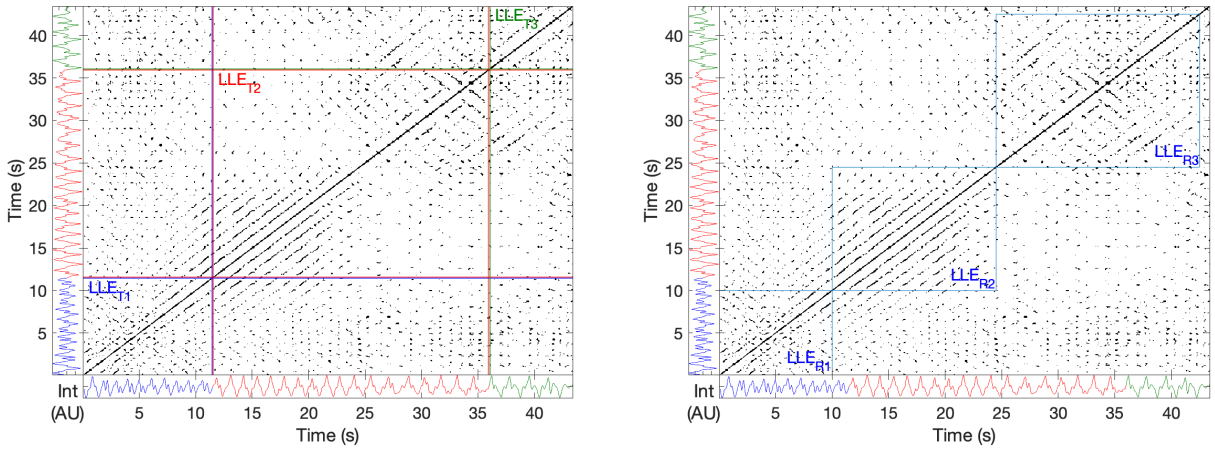


(c) 4D RP

Figure 4.13 Generating embedded recurrence plots for experimental *C. elegans* time-series, with t_ι and t_κ plotted along the x-axis and y-axis, respectively [44]. The red line showcases the LOI, and the blue boxes showcase different sections of the motion. (a) Heatmap of the distance matrix $D_{\iota,\kappa}$ for the 1D vector time-series, $y^1(t_\mu)$, *i.e.*, for embedding dimension $n = 1$. (b) RP of recurrence matrix $R_{\iota,\kappa}$ for the 1D vector time-series, $y^1(t_\mu)$. A threshold calculated using $p_r = 0.1\%$ of the range of amplitudes, $\Delta\varphi_A$, has been added to the 1D embedded RP. (c) RP of recurrence matrix $R_{\iota,\kappa}$ for the 4D vector time-series, $y^4(t_\mu)$, *i.e.*, for embedding dimension $n = 4$. The time-series is now embedded in 4D with threshold calculated using $p_r = 0.3\%$ of the range of amplitudes, $\Delta\varphi_A$. The embedding dimension is proven sufficient to unfold the dynamics because all perpendicular lines (FNN) are eliminated (as illustrated in Figure 4.11). The diagonal lines that are parallel with the LOI indicate how long a particular state lasts.

mension, $n = 4$, for the same experimental time-series. Focusing on a shorter sub-series of that time-series, we can go through the steps of generating the fully embedded RP for our experimental data: plotting the 1D distance matrix heatmap, plotting the 1D RP, and plotting the 4D RP. First, we display the 1D heatmap of the distance matrix, $D_{\iota,\kappa}$, as shown in Figure 4.13a, in order to get an overview of the recurrence patterns. These patterns of observed recurrences are displayed with a sharper contrast in the 1D recurrence matrix, $R_{\iota,\kappa}$, as shown in Figure 4.13b. Lastly, imposing the $n = 4$ embedding dimension eliminates the perpendicular structures and reveals only lines parallel to the LOI, as seen in Figure 4.13c. The blue boxes in each panel of Figure 4.13 showcase different sections of the time-series exhibiting different types of locomotion based on the patterns observed within the RP. After embedding the time-series in the proper phase space dimension, $n = 4$, as shown in Figure 4.13c, the recurrences now form a pattern of almost straight diagonal lines parallel to the LOI. These diagonal lines are not contiguous, but rather, consist of line segments of varying lengths in time. Each of these diagonal line segments represent time periods during which two trajectories are in close proximity to each other. The diagonal line lengths therefore depict how quickly trajectories diverge. On the basis of these line lengths, we can then roughly estimate LLE values for individual trajectories, with typical estimated LLE values around 2.0 s^{-1} for experimental time-series such as the time-series shown in Figure 4.13c.

Over the duration of the full time-series, it is possible that the locomotion dynamics may qualitatively change its pattern of motion. We can identify such changes in the pattern of motion either by visual inspection of the time-series itself or by visual inspection of the RP. Based on the observed times where such changes of the motion pattern occur, we can then partition the full time-series into smaller sub-series, referred to as "sections" in the following. The boundaries of these sections are determined by the time points where a change in the pattern of motion occurs. Inside of each such section, the pattern of motion should be stable. As the pattern of motion changes from one section to another, we should also expect that



(a) (b)

Figure 4.14 Identification of scaling sections in a recurrence plot generated for an experimental time-series [44]. RP with 4D embedding and threshold calculated using $p_r = 0.3\%$ of the range of amplitudes, $\Delta\varphi_A$. The time-series is plotted on both axes and is color coded into three sections based on visual inspection of the time-series alone. (a) The three colored sections in the time-series are also sectioned in the RP: 0-11.49 s in blue, 11.49-36 s in red, and 36-43.38 s in green. Using these sections each LLE was calculated using an algorithm by Rosenstein et al. [55], $\text{LLE}_{\text{T1}} = 2.22 \pm 0.06$ s^{-1}, $\text{LLE}_{\text{T2}} = 1.83 \pm 0.01$ s^{-1}, and $\text{LLE}_{\text{T3}} = 2.74 \pm 0.01$ s^{-1} were calculated. (b) Through visual inspection of the RP, different sections are identified, which are inside blue boxes in the RP: 0 - 10 s, 10 - 24.5 s, and 24.5 - 42.5 s. These sections were used to calculate $\text{LLE}_{\text{R1}} = 2.22 \pm 0.01$ s^{-1}, $\text{LLE}_{\text{R2}} = 1.70 \pm 0.01$ s^{-1}, and $\text{LLE}_{\text{R3}} = 1.83 \pm 0.01$ s^{-1}.

the LLE, calculated individually for each section, varies from one section to the next. Figure 4.14a illustrates the presence of three such sections. The sections were determined through visual inspection of the time-series, $y(t_\mu)$, to attempt to identify times where a switch in locomotory gait had occurred, *i.e.*, identify times when the oscillation in the time-series changes. Along the axes of Figure 4.14a, the sections are color coded, in order to indicate their extent and boundaries. A grid of rectangular boxes, centered with their diagonals on the LOI, is added to the RP in Figure 4.14a to highlight the recurrence patterns in each of those three sections.

However, visually inspecting the RP, instead of the time-series, to identify switches in

locomotory gait leads to different partitioning into three different sections that are indicated by the boxes in Figure 4.14b. For comparison, the color coding for the time-series identified sections used in Figure 4.14a is shown along the axes of Figure 4.14b as well, which clearly shows that the sections in Figure 4.14b are of different length than in Figure 4.14a.

For each of the sections, in Figure 4.14a and 4.14b, the individual LLE values were calculated based on the method described in section 4.4. The values of the individual LLEs, as well as their linear fit standard deviations, are noted in the caption of Figure 4.14. In comparing the LLE standard deviations from Figure 4.14a to those of 4.14b, we see that the LLE standard deviation from Figure 4.14b of the first section are smaller by a factor of approximately six. We therefore conclude that the partitioning by visual inspection of the RP allows us to obtain more accurate results for the LLE values of individual sections.

Through our exploration of RPs, we see that using RP as a visual representation of the dynamics has several advantages. First, we can observe the time evolution of the dynamics based on how the RP features change in time along the LOI. Second, determining different locomotory sections of the time-series is more accurate using RPs rather than the time-series alone. Additionally, RPs satisfy three of the four chaos criteria defined by David Feldman [2]: determinism, aperiodic orbits, and sensitive dependence on initial conditions. The finite lengths of the the lines parallel to the LOI indicate that our data is short-term predictable, hence short-term deterministic. RP can also justify that our locomotion data is aperiodic because the lines parallel to the LOI do not span across the whole graph as with a periodic signal, but rather have finite diagonal line lengths related to the LLE of the system. Based on this connection to LLE, RP can also justify sensitive dependence on initial conditions.

4.7 Surrogate Data Analysis

In this section, we investigate one additional criterion for chaos for our *C. elegans* locomotion data: the presence of nonlinearity. The results of this section will prove that our observed

diffraction intensity time-series, $y(t_\mu)$, are nonlinear using the method of surrogate data analysis, which will be described below. To implement surrogate data analysis, we first generate a large sample of surrogate time-series data, $\breve{y}_s(t_\mu)$, from an observed time-series, $y(t_\mu)$, where s labels the elements of the sample. Then we calculate a quantity, to be referred to as the nonlinear test observable (NTO), Λ_s, for each surrogate and the original data, $y(t_\mu)$ in order to compare using a rank-order test [21, 73], which will also be detailed below.

The technique of surrogate data analysis is a test for nonlinearity that employs a proof by contradiction as detailed by Schreiber and Schmitz [20, 21, 74]. Surrogate data analysis is based on the null hypothesis that the time evolution of the original data, $y(t_\mu)$, is linear. If we disprove that null hypothesis, then the time evolution of the original time-series must be nonlinear. For purposes of our null hypothesis, the time evolution of any time-series, such as $y(t_\mu)$, is defined to be linear if it can be represented as a so-called "rescaled linear Gaussian process", having the following general mathematical form [75]:

$$y(t_\mu) = h(x(t_\mu)) \tag{4.37}$$

$$x(t_\mu) = \sum_{i=1}^{L_{\rm det}} a_i \; x(t_{\mu-i}) + \sum_{i=1}^{L_{\rm sto}} b_i \; \eta(t_{\mu-i}) \tag{4.38}$$

The observed original time-series, $y(t_\mu)$, is assumed here to be related to the so-called hidden time-series, $x(t_\mu)$, by way of a possibly nonlinear, but invertible single-variable transformation function, $h(x)$, in equation (4.37). Hence, $x(t_\mu) = h^{-1}(y(t_\mu))$, where $h^{-1}(y)$ is the inverse function of $h(x)$. This nonlinear transformation, $h(x)$, is intended to account for spurious nonlinearities introduced into the $y(t_\mu)$ time-series data, for example, due to nonlinear responses of the experimental measurement apparatus.

The actual linearity of the time evolution is then expressed in terms of the hidden time-series, $x(t_\mu)$, by way of equation (4.38). The first term on the right hand side of equation (4.38) specifies a deterministic contribution to the time evolution of $x(t_\mu)$, with summation

over $i = 1, \cdots, L_{\rm det}$, and with some arbitrary choice of linear coefficients a_i. This term has the basic structure of a linear "memory": the signal value $x(t_\mu)$ at the present time, t_μ, is a linear superposition of the signal values $x(t_{\mu-i})$ at earlier times $t_{\mu-i}$. The $L_{\rm det}$ specifies a finite memorization time, $L_{\rm det}\ \Delta t_{\rm exp}$, for this linear memory: at the present time, t_μ, the system "remembers" the signal values, $x(t_{\mu-i})$, up to and including $L_{\rm det}$ time steps into the past. The $x(t_\mu)$ contribution at the present time t_μ is then determined by the linear combination of only those $x(t_{\mu-i})$ that are remembered from the past. The domain of the μ-index of the hidden time-series is allowed to extend back to non-positive values, including $\mu = 0, -1, \cdots, -(L_{\rm det} - 1)$. The corresponding hidden signal values $x_0, x_{-1}, \cdots, x_{-(L_{\rm det}-1)}$, can be chosen arbitrarily, to specify the initial conditions of the time evolution.

The second term on the righthand side of equation (4.38) specifies a stochastic (random noise) contribution to the time evolution of $x(t_\mu)$, with summation over $i = 1, \cdots, L_{\rm sto}$, and with some arbitrary choice of linear coefficients b_i. The noise variables $\eta(t_{\mu-i})$ are statistically independent, normally distributed (Gaussian) random variables of zero mean, sharing some arbitrary, but common choice of variance, $\sigma^2_{\rm sto}$. This term also has the basic structure of a linear memory, with a memorization time specified by $L_{\rm sto}$. In a strictly deterministic system, this stochastic term would obviously be set to zero, by setting $b_i = 0$ for all i. However, for real experimental time-series data, some level of noise will always be present and must be accounted for.

Equation (4.38) encompasses the most general possible expression of an autonomous linear time evolution of a single variable, x, for a discretized time variable, t_μ, subject to causality and subject to an arbitrary choice of linear coefficients, a_i and b_i, memorization times, $L_{\rm det}\ \Delta t_{\rm exp}$ and $L_{\rm sto}\ \Delta t_{\rm exp}$, and initial conditions $x_0, x_1, \cdots, x_{(L_{\rm det}-1)}$. Causality here means that the signal, $x(t_\mu)$, at any time, t_μ, depends only on the past signal values, $x(t_{\mu-i})$, but not on any future signal values, $x(t_{\mu+i})$, with $i > 0$. By way of time discretization, equation (4.38) also encompasses causal linear time evolutions for a continuous time variable

if, e.g., x(t) evolves in causal manner according to a linear differential, integral, or integro-differential equation.

Our null hypothesis, to be disproved below, can now be precisely stated as follows:

> Null Hypothesis: The original time-series, $y(t_\mu)$, at all observation time points, t_μ, with $\mu = 1, 2, \cdots, N$, can be represented by a rescaled linear Gaussian process, as defined by equations (4.37) and (4.38), with some choice of a nonlinear transformation function, $h(x)$, linear coefficients, a_i and b_i, memorization times, $L_{\text{det}}\ \Delta t_{\text{exp}}$ and $L_{\text{sto}}\ \Delta t_{\text{exp}}$, initial conditions $x_0, x_1, \cdots, x_{(L_{\text{det}}-1)}$, and Gaussian noise variance σ^2_{sto}.

To refute the null hypothesis for *C. elegans* locomotion data, we will refer to a time-series, $y(t_\mu)$, as our "original time-series", then generate a large random sample of so-called surrogate time-series from the original, $y(t_\mu)$, using a randomizing time-series transformation. This transformation is outlined below and described in detail, as the "amplitude adjusted Fourier transform" (AAFT) method, by Lancaster *et al.* in [75]. The random surrogate time-series generated from $y(t_\mu)$ are to be denoted by $\breve{y}_s(t_\mu)$ with $s = 1, \cdots, M$. Here s labels the different elements of the sample and M denotes the size of that surrogate sample. Figure 4.15 shows an example of an original time-series, $y(t_\mu)$, and nine different surrogate time-series, $\breve{y}_s(t_\mu)$, generated from that original.

To compare original and surrogate time-series, we will then also define, for any time-series, a test statistic, the so-called "nonlinearity test observable" (NTO), described in more detail below and in [75]. The NTO value calculated from the original experimental time-series, $y(t_\mu)$, is then denoted by Λ_0 and the NTO values calculated from the surrogate time-series, $\breve{y}_s(t_\mu)$, are denoted by Λ_s with $s = 1, \cdots, M$. There are many different NTOs to choose from, but "asymmetry under time reversal" is an NTO that is very commonly used [20]. The steps for generating the random sample of surrogate time-series and the precise definition of the

NTO variable will be detailed further below.

The surrogate-based hypothesis testing approach then rests on the following crucial mathematical property, which I will refer to as "linearity compliance" hereafter and which our surrogate construction, in combination with the chosen NTO definition, must exhibit:

> Linearity Compliance: For any original time-series, $y(t_\mu)$, which is linear in the sense of equations (4.37) and (4.38), the random surrogate values of the NTO, Λ_s for $s = 1, \cdots, M$, must agree, within statistical uncertainties, with the NTO of the original time-series, $y(t_\mu)$, Λ_0.

If the surrogate and NTO construction is linearity compliant, but the data analysis reveals that Λ_0 is significantly different from the sample of surrogate NTO values, Λ_s for $s = 1, \cdots, M$, then the null hypothesis, i.e., the hypothesis that the original experimental time-series, $y(t_\mu)$, is representable in the form of equations (4.37) and (4.38), can be rejected. In other words, if Λ_0 differs significantly from the sample of surrogate NTO values, Λ_s for $s = 1, \cdots, M$, then the time evolution of the original experimental time-series, $y(t_\mu)$, must be nonlinear.

The term "significantly different" in this context is defined by way of a rank ordering of the sequence of all $(M + 1)$ NTO values, $[\Lambda_0, \Lambda_1, \cdots \Lambda_M]$, including the original time-series NTO, Λ_0, i.e., $s = 0$, as well as the surrogate NTO values, Λ_s for $s = 1, 2, \cdots M$. Rank ordering this sequence of Λ-values, e.g., in ascending order, assigns a rank index, denoted by rank_s, to each s-value from $s \in \{0, 1, \cdots, M\}$. Given this rank ordering, the value of Λ_0 is then defined to be significantly different from the sequence of surrogate values, $[\Lambda_1, \cdots \Lambda_M]$, if

$$\text{rank}_0 = 1 \quad \text{or} \quad \text{rank}_0 = M + 1 \, . \tag{4.39}$$

An alternative way to state this definition is to say that Λ_0 is defined to be significantly

different from the sequence of surrogate values, $[\Lambda_1, \cdots \Lambda_M]$, if

$$\Lambda_0 < \min[\Lambda_1, \cdots \Lambda_M] \quad \text{or} \quad \Lambda_0 > \max[\Lambda_1, \cdots \Lambda_M] \; . \tag{4.40}$$

Clearly, the value of Λ_0 must lie far outside the width of the Λ-distribution of the surrogates, in order to satisfy the significantly difference condition equation (4.39) or (4.40). This is illustrated in the two panels of Figure 4.16, which shows the histogram of the surrogate values, Λ_s for $s = 1, \cdots, M$, and the value of Λ_0 as a vertical red line, for two different experimental diffraction intensity time-series, $y(t_\mu)$.

As in any hypothesis testing approach based on random samples of data, the decision to reject the null hypothesis is a probabilistic one. There is always a finite residual probability, referred to as the "p-value" in statistics, for the null hypothesis to be true even if the rejection criterion, such as equation (4.39), is satisfied. In the present surrogate-based approach, the reliability of the null hypothesis rejection will increase with the surrogate sample size, M. In particular, it has been shown [21] that the residual probability in our approach [68] will be less than a prescribed upper limit if M exceeds a certain minimum value, given by

$$M_{\min} = \frac{2K}{\alpha} - 1 \; . \tag{4.41}$$

Here, α is the prescribed upper limit for the residual probability, *i.e.*, is an upper limit to the p-value of our null hypothesis rejection criterion, and K is a positive integer. In other words, the α-value represents an upper limit for the probability of the null hypothesis being true and 1-α is a lower limit for the probability of the negation of the null hypothesis being true. Larger values of K will increase the sensitivity of our test, but $K = 1$ is sufficient and reduces computational effort. If we require a minimal significance of 95%, then α is 5% (0.05) leading to a minimal required number of surrogates of $M = 39$. To further increase our discriminating power, I generated 1000 surrogates ($K \approx 25$, $\alpha = 0.05$) for each time-series

described in this section to more irrefutably disprove the linear hypothesis for *C. elegans* locomotion data.

The first step in the surrogate data analysis technique is the AAFT surrogate construction. Since AAFT is discussed in detail in Lancaster *et al.* [75], I will present here only a brief summary of the computational steps of the AAFT random transformation, to generate one single, random AAFT surrogate time-series, $\breve{y}_s(t_\mu)$, from the original time-series, $y(t_\mu)$:

Step 1: Rank-order the sequence of original y-values, $[y(t_1), \cdots, y(t_N)]$, in ascending order, with an integer rank index $r(\mu)$, defined such that $r(\mu) \in \{1, \cdots, N\}$, and $r(\mu) \neq r(\mu')$ if $\mu \neq \mu'$ and $r(\mu) < r(\mu')$ if $y(t_\mu) < y(t_{\mu'})$. So, for example, $r(\mu) = 7$ if $y(t_\mu)$ is the 7^{th} smallest y-value in $[y(t_1), \cdots, y(t_N)]$. Also sort the $[y(t_1), \cdots, y(t_N)]$, in ascending order to generate the rank-ordered, ascending sequence of y-values, denoted by $[\tilde{y}(t_1), \cdots, \tilde{y}(t_N)]$, *i.e.*, set

$$\tilde{y}(t_{r(\mu)}) := y(t_\mu) \quad \text{for } \mu = 1, 2, \cdots N \; . \tag{4.42}$$

Step 2: Draw a random time-series, of length N, of statistically independent, zero-mean Gaussian random variables, all of common variance σ_ζ^2, denoted by $\zeta(t_\mu)$ with $\mu = 1, 2, \cdots N$.

Step 3: Sort the sequence $[\zeta(t_1), \cdots, \zeta(t_N)]$ in ascending order to generate the rank-ordered, ascending sequence of ζ-values, denoted by $[\tilde{\zeta}(t_1), \cdots, \tilde{\zeta}(t_N)]$.

Step 4: Re-order the sequence $[\tilde{\zeta}(t_1), \cdots, \tilde{\zeta}(t_N)]$ according to the rank index order, $r(\mu)$, found in Step 1. Denote the resulting sequence of ζ-values by $[\xi(t_1), \cdots, \xi(t_N)]$. In other words, $\xi(t_\mu)$ is defined by:

$$\xi(t_\mu) := \tilde{\zeta}(t_{r(\mu)}) \; . \tag{4.43}$$

Step 5: From the time-series $\xi(t_\mu)$, generate a "FT surrogate" time-series, denoted by $\breve{\xi}(t_\mu)$. The procedure of FT surrogate time-series generation is described separately below, with multiple sub-steps, Step 5.1, 5.2, $\cdots$

Step 6: Rank-order the sequence of original $\breve{\xi}$-values from Step 5 $[\breve{\xi}(t_1), \cdots, \breve{\xi}(t_N)]$, in

ascending order, with an integer rank index $s(\mu)$, defined such that $s(\mu) \in \{1, \cdots, N\}$, and $s(\mu) \neq s(\mu')$ if $\mu \neq \mu'$ and $s(\mu) < s(\mu')$ if $\breve{\xi}(t_\mu) < \breve{\xi}(t_{\mu'})$. So, for example, $s(\mu) = 4$ if $\breve{\xi}(t_\mu)$ is 4^{th} smallest $\breve{\xi}$-value in $[\breve{\xi}(t_1), \cdots, \breve{\xi}(t_N)]$.

Step 7: Re-order the sequence $[\tilde{y}(t_1), \cdots, \tilde{y}(t_N)]$ from Step 1 according to the rank index $s(\mu)$ from step 6, to generate the desired, final surrogate time-series, $\breve{y}_s(t_\mu)$, *i.e.*,

$$\breve{y}_s(t_\mu) := \tilde{y}(t_{s(\mu)}) \ . \tag{4.44}$$

The FT surrogate construction in Step 5 above, for a given input time-series, $\xi(t_\mu)$, proceeds by the following sub-steps:

Step 5.1: From $\xi(t_\mu)$ for $\mu = 1, 2, \cdots N$, calculate the discrete FT, $\hat{\xi}_k(\nu_k)$, of $\xi(t_\mu)$ for $\mu = 1, 2, \cdots N$, based on equation (3.17) for $k = 1, 2, \cdots N$.

Step 5.2: Generate a sequence of statistically independent, random phase angles β_k, each β_k, drawn from a uniform distribution over the interval $[0, 2\pi]$, for $k = 2, 3, \cdots N/2$.

Step 5.3: Set $\beta_k = 0$ for $k = 0$ and $k = N/2 + 1$.

Step 5.4: Expand the sequence of $N/2 + 1$ phase angles β_k, defined in Steps 5.2 and 5.3, into a sequence of length N, by setting

$$\beta_k := -\beta_{N+2-k} \tag{4.45}$$

for $k = N/2 + 2, N/2 + 3, \cdots N$.

Step 5.5: Multiply each $\hat{\xi}_k(\nu_k)$ with its random phase factor $e^{i\beta_k}$, to generate the phase-randomized FT

$$\hat{\xi}_k^\beta := \hat{\xi}_k(\nu_k) e^{i\beta_k} \tag{4.46}$$

for $k = 0, 1, \cdots N$.

Step 5.6: The desired FT surrogate, $\breve{\xi}_k(t_\mu)$, for $\mu = 1, 2, \cdots N$, is then given by the inverse

FT applied to the phase-randomized FT, $\hat{\xi}_k^{\beta}$, using equation (3.18).

Notice that the input time series in Step 5.1, $\xi(t_\mu)$, is real-valued and therefore, its FT, $\hat{\xi}_k(\nu_k)$, obeys the symmetry relations stated in equations (3.21)-(3.23). The construction of the random phase angle sequence, $[\beta_1, \cdots, \beta_N]$, in Steps 5.2-5.4, ensures that the phase-randomized FT, $\hat{\xi}_k^{\beta}$, will also obey those symmetry relations in equations (3.21)-(3.23). Consequently, the FT surrogate, $\breve{\xi}_k(t_\mu)$, is guaranteed to be real-valued for $\mu = 1, 2, \cdots N$.

The second step in the surrogate data analysis technique is to evaluate the NTO variable, Λ_s, for each surrogate time-series, $\breve{y}_s(t_\mu)$, and NTO value Λ_0 for the original time-series $y(t_\mu)$. Using the so-called "time reversal asymmetry" observable as our NTO [75], Λ_s is given by [20, 21, 74, 75]:

$$\Lambda_s = \frac{1}{N_\tau} \sum_{\mu}^{N_\tau} (\breve{y}_s(t_\mu + \tau) - \breve{y}_s(t_\mu))^3 \tag{4.47}$$

for $s = 0, 1, \cdots, M$ and we set $\breve{y}_0(t_\mu) \equiv y(t_\mu)$ when $s = 0$.

As noted above, the AAFT surrogate construction, in combination with the time reversal asymmetry NTO defined in equation (4.47), does satisfy the condition of linearity compliance stated above and can therefore be used to test the linearity null hypothesis stated above. Furthermore, for the specific case of the time reversal asymmetry NTO, it can be shown [75] that the original time-series NTO, Λ_0, should be close to zero and most of the random surrogate NTO values, Λ_s, should be clustered in close proximity to zero, if the linearity null hypothesis was true. A failure of the NTO values to exhibit this close proximity to zero could therefore be taken as additional evidence that the null hypothesis is not true. I have carried out this hypothesis test, based on AAFT surrogates and time reversal asymmetry NTOs, for altogether 28 different experimental diffraction intensity time-series, $y(t_\mu)$, including 10 time-series observed on wildtype worms and 18 time-series observed on the roller mutant worms. (These are the same 28 time-series that were also used in the correlation dimension analysis discussed above in section 4.5.)

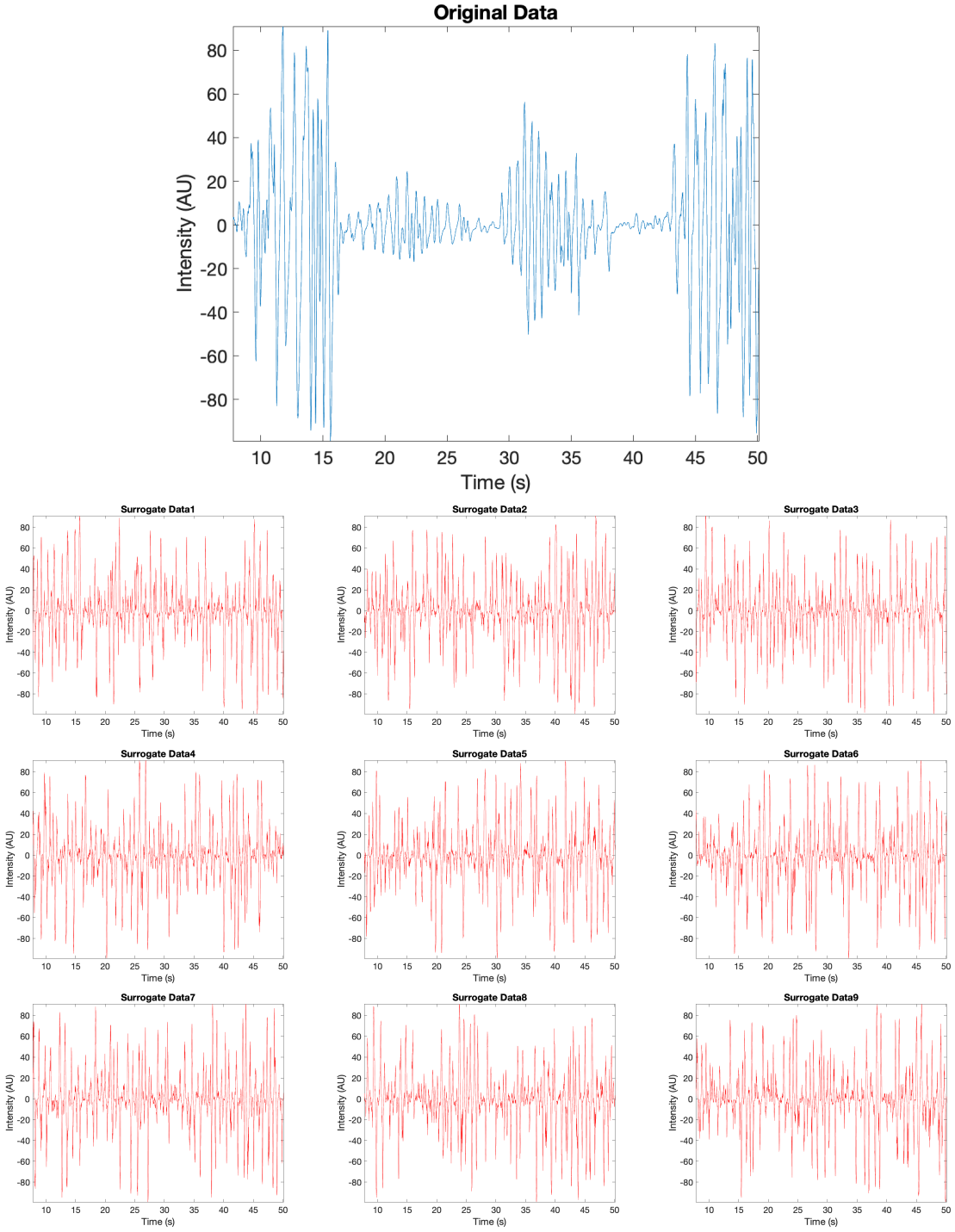
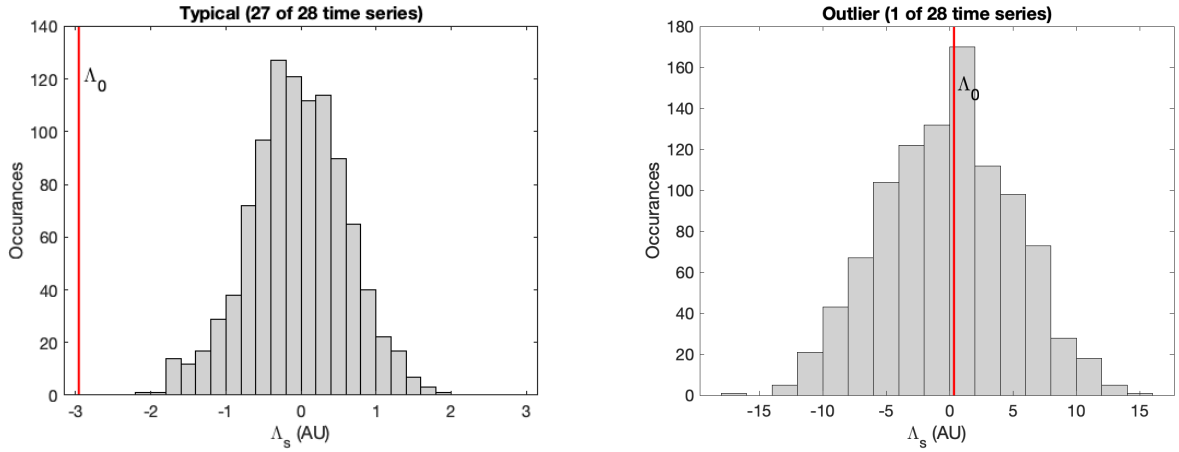


Figure 4.15 Example set (9 out of 1000) of surrogate time-series, $\breve{y}_s(t_\mu)$, (red) generated from an original time-series, $y(t_\mu)$, (blue) using the AAFT method of surrogate time-series construction.

The results from these 28 hypothesis tests overwhelmingly support the rejection of the null hypothesis. These results therefore provide strong evidence that the locomotion dynamics of *C. elegans* is indeed governed, in both wildtype and in the roller mutant, by a nonlinear time evolution. Specifically, in 27 out of the 28 time-series observed, the hypothesis test showed that Λ_0 does in fact differ significantly, in the sense of equations (4.39) or (4.40), from the random sample of surrogate NTO values, for a surrogate sample size of $M = 1000$. For only one of the 28 time-series observed did the test not reject the null hypothesis, *i.e.*, Λ_0 failed to differ significantly from the surrogate NTO sample.



(a) Typical: $\text{rank}_0 = 1$ out of $M = 1000$. (b) Outlier: $\text{rank}_0 = 521$ out of $M = 1000$.

Figure 4.16 Rankings and histograms of the time reversal asymmetry NTO values , Λ_0 and Λ_s, for original and surrogate time-series, respectively [44]. In each panel, (a) and (b), the NTO value, Λ_0, from an original time-series is shown as a vertical red line. The sample of surrogate NTOs, Λ_s for $s = 1, \cdots, M$, is generated from the original time series by the AAFT method and is shown as a histogram. Panel (a) is for a typical original time-series of experimental observations, out of 27 observed time-series total, for which the null hypothesis was rejected. Panel (b) is for the outlier time-series, *i.e.*, for the one and only original time-series of experimental observations, for which the null hypothesis was not rejected.

Figure 4.16 shows representative results from two of these 28 hypothesis tests, *i.e.*, for two different experimental diffraction intensity time-series, $y(t_\mu)$, out of the 28 time-series observed. The results in the left panel are for the "typical" case, *i.e.*, for one of the 27 observed time-series where the null hypothesis was rejected. The results in the right panel

are for the "outlier" case, *i.e.*, for the one and only observed time-series where the null hypothesis was not rejected. In both panels, the respective random sample of surrogate NTO values, *i.e.*, Λ_s for $s = 1, 2, \cdots, M$ is displayed as a histogram and the original time-series NTO value, Λ_0, is shown as a vertical red line. In the left panel of Figure 4.16, Λ_0 is seen to lie far outside of the range of the surrogate histogram and it indeed satisfies the criterion of "significant difference" by equation (4.39), namely $\text{rank}_0 = 1$, relative to the surrogate sample size of $M = 1000$. In the right panel of Figure 4.16, Λ_0 lies roughly within the middle of the range of the surrogate histogram, with a $\text{rank}_0 = 521$, relative to the surrogate sample size of $M = 1000$.

4.8 Summary of Results and Discussion

Table 4.1 Chaos criteria [2] and how to satisfy them

Chaos Criteria	***C. elegans* locomotion met criteria by:**
deterministic	lag plot, correlation dimension, recurrence plots
aperiodic orbits	broad frequency spectrum, lag plots, recurrence plots
bounded orbits	lag plots
sensitive dependence on initial conditions	lag plot, LLE, recurrence plots

Throughout this chapter, we have shown that *C. elegans* locomotion is chaotic, as defined by David Feldman [2] and nonlinear. A breakdown of what quantities justify each criterion of chaos can be found in Table 4.1. This summary section details how each measure satisfies those criteria. We will also conclude with a summary of the common numerical results of *C. elegans* locomotion, compiled in Table 4.2.

The first chaotic marker that we calculate is a broad frequency spectrum in section 4.1. The chaotic criterion that this marker justifies is an aperiodic orbit. Having these broad

frequency spectra, rather than well defined, sharp peaks, shows that the motion is not periodic.

Section 4.2 of this chapter discussed Takens delay embedding approach as a means to unfold the dynamics contained in experimentally observed scalar time-series, $y(t_\mu)$. The Takens delay embedding is achieved by transforming the original time-series, $y(t_\mu)$, into a vector time-series, $y^n(t_\gamma)$, where the vector components consist of truncated, time lagged replicas of the original time-series. The method of MI was then applied to determine an optimal lag τ. By plotting a 2D projection of the vector time-series, we illustrated that a 2D embedding leads to frequent trajectory crossings in the plane of projection. In 2D, the orbits are bounded in a shape that resembles a wound up yarn ball. Much like a yarn ball projected into a 2D plane, the trajectories appear to cross each other in phase space. The presence of these crossings is a clear indication that a higher embedding dimension, $n > 2$, is required to unfold the dynamics of *C. elegans*.

In section 4.3, we use the technique of FNN to determine the proper embedding dimension for dynamic diffraction time-series of *C. elegans* locomotion. At embedding dimension $n = 3$, we find a FNN percentage less than 1% so that the vector time-series trajectories in the 3D lag plots of *C. elegans* locomotion do not cross each other, thus proving that each value in phase space is unique, hence deterministic. Vector time-series trajectories of sufficiently high embedding dimension do not cross, which proves that the time-series is not periodic. These same vector time-series trajectories occupy a finite region of phase space, as verified by visual inspection in $n = 3$; hence the orbits are bounded. Lastly, there is a sensitive dependence on initial conditions because close vector time-series trajectories eventually diverge apart. Lag plots are a powerful visual representation of time-series data presumed to be chaotic because lag plots can test every chaos criterion.

The second chaotic marker that we calculated was the LLE in Section 4.4. An LLE quantifies the system's sensitivity on initial conditions because it measures how fast two

trajectory points, that are close in phase space but separated in time, diverge exponentially away from each other. The larger the positive LLE, the more sensitive the system is to initial conditions. Given our LLE values of $\sim +2\ \mathrm{s}^{-1}$, we conclude that *C. elegans* locomotion exhibits sensitive dependence on initial conditions.

The third chaotic marker that we calculated was the correlation dimension, d_{cor}, in section 4.5. If the correlation dimension and embedding dimension are approximately equal to each other for all embedding dimensions, including $n \to \infty$, then the time-series is random. This was not the case with our *C. elegans* time-series because our correlation dimension, as a function of embedding dimension, leveled off as the embedding dimension increased beyond $n = 4$, which indicates deterministic behavior. Furthermore, a correlation dimension above 2 is a necessary condition for a time-series to be aperiodic, bounded, and not cross in phase space. Additionally, our non-integer average correlation dimension, $d_{\mathrm{cor}} \approx 2.08 \pm 0.24$, falls between 2 and 3. This means that our data may have characteristics of both 2D and 3D space, which would be an indication of chaotic dynamics.

In section 4.6, we looked at other visual representations of our data, specifically through distance matrix heatmaps and recurrence plots (RP). In embedding dimension $n = 4$, the RP of our *C. elegans* time-series data exhibits recurrences that form line segments of finite length running parallel to the LOI. The finite lengths of the lines parallel to the LOI in the RP indicate that our data is short-term predictable, hence deterministic. RP can also justify that our locomotion data are aperiodic. For periodic signals, the lines parallel to the LOI span across the whole graph, while our RPs have diagonal lines that are disrupted by non-recurrent gaps, and therefore have a finite length related to the LLE of the system. Hence, our RP results also support the criterion of sensitive dependence on initial conditions.

Finally, in addition to proving criteria of chaotic dynamics, we also investigated potential nonlinearity of *C. elegans* locomotion dynamics in section 4.7. The creation and analysis of a large sample of surrogate time-series, randomly generated from the original time-series, plays

an essential role in this search for nonlinearity. Using time reversal asymmetry observable as our NTO and a rank order test, applied to the original data and 1000 surrogates generated from the original, we consistently found that the criterion for nonlinearity was satisfied in 27 out of 28 experimental time-series analyzed.

Table 4.2 Summary of typical *C. elegans* Results

Quantity (average)	**Value**
mean frequency[(1)]	$\nu_{\text{avg}} = 1.32$ Hz
lag	$\tau = 0.18 \pm 0.07$ s
embedding dimension	$n = 4$
largest Lyapunov exponent[(2)]	$\lambda_{\text{LLE}} = 1.39 \pm 0.02$ s^{-1}
correlation dimension	$d_{\text{cor}} \approx 2.08 \pm 0.24$

[(1)] The mean, or average, frequency was obtained from a single *C. elegans* time-series shown in Figure 4.1. This value is representative of mean frequencies obtained from numerous other time-series experiments on the locomotion of *C. elegans*.

[(2)] The LLE value shown here was obtained from a single *C. elegans* time-series by averaging over all Lyapunov starting times as described in section 4.4. This LLE value is representative of LLE values obtained from numerous other time-series experiments on the locomotion of *C. elegans*.

Chapter 5

Conclusion

In this thesis, we have described the use of Fraunhofer diffraction combined with nonlinear dynamics to explore the locomotion of freely swimming *C. elegans*, similar to its natural environment in soil [76]. The intensity fluctuations at one point in the diffraction pattern form a single time-series containing information about the undulations of the entire worm. These one-dimensional time-series show complex dynamical patterns.

The analysis of the locomotion revealed nonlinear deterministic dynamics in a complex system with quantifiable and repeatable results. In every case, the FFT of the dynamic diffraction time-series shows a broadened frequency spectrum indicating an aperiodic structure. This is the first indication that the locomotion of *C. elegans* is complex. The swimming frequencies oscillate around a mean frequency, but never repeat as demonstrated by lag plots. Using the false nearest neighbors (FNN) method, we determined that the necessary embedding dimension is 3 or 4 in all cases. We saw markers of low-dimensional ($\sim$4) deterministic chaos in 50-second recordings of the diffraction signal (See Table 4.1). We also found a key marker of chaos, namely a positive largest Lyapunov exponent, LLE, ($\sim 2.0/\text{sec} \pm 0.7$). This means that the Lyapunov time is about ~ 1.0 sec indicating that it is possible to predict the locomotion about ~ 1.0 into the future.

We used Recurrence Plots (RP) to graphically visualize the time evolution and periodicity of *C. elegans* locomotion. Properly embedded RP are another way to view the trajectories visualized in a lag plot. Unlike a lag plot, determining the duration that two trajectories are close to each other is easily done by measuring the lengths of the lines parallel to the LOI.

These finite line lengths indicate short term predictability, which is a measure of determinism. These parallel lines also relate to the LLE of the system, indicating sensitive dependence on initial conditions. Each time the locomotion of the *C. elegans* switched to a new gait, there was a visible difference in the RP. The dynamic regions identified in RP reduced the uncertainty of our LLE calculations by a factor of six.

In order to focus on our approach to finding markers of chaos in the undulation of *C. elegans*, this thesis only presents results obtained from studying two genotypes of *C. elegans* (the wildtype and roller) in a single 3D medium (water). Both types of worms displayed chaotic markers, but no statistically significant differences were detected since the roller defect is muscular rather than neuronal. In future experimental work, we plan to expand to other mutants, age studies, and effect of the medium in which the *C. elegans* moves. We also plan to explore mechanisms for generating complex undulation patterns seen in real *C. elegans*.

In summary, dynamic diffraction has opened a gateway to characterizing locomotion in microscopic species across multiple scales. This type of study will lead to more a accurate quantification and understanding of neuronal dynamics.

Bibliography

[1] J. Magnes, K. Susman, and R. Eells, "Quantitative locomotion study of freely swimming micro-organisms using laser diffraction," *J. Vis. Exp.*, vol. 68, p. e4412, 2012.

[2] D. P. Feldman, *Chaos and Fractals: An Elementary Introduction*, p. 85. Oxford University Press, 2012.

[3] F. Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics*, pp. 366–381, 1981.

[4] S. L. Edwards, N. K. Charlie, M. C. Milfort, B. S. Brown, C. N. Gravlin, J. E. Knecht, and K. G. Miller, "A novel molecular solution for ultraviolet light detection in *Caenorhabditis elegans*," *Public Library of Science (PLoS) Biology*, vol. 6, no. 8, p. e198, 2008.

[5] J. Magnes, H. M. Hastings, K. M. Raley-Susman, C. Alivisatos, A. Warner, and M. Hulsey-Vincent, "Fourier-based diffraction analysis of live *Caenorhabditis elegans*," *J. Vis. Exp.*, vol. 127, p. e56154, 2017.

[6] S. Brenner, "The genetics of *Caenorhabditis elegans*," *Genetics*, vol. 77, no. 1, pp. 71–94, 1974.

[7] J. J. H. Liang, I. A. McKinnon, and C. H. Rankin, "The contribution of neurogenetics to understanding neurodegenerative diseases," *Journal of Neurogenetics*, vol. 34, no. 3-4, pp. 527–548, 2020.

[8] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The structure of the nervous system of the nematode *Caenorhabditis elegans*," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 314, no. 1165, pp. 1–340, 1986.

[9] "Wormbook: The online review of *C. elegans* biology." http://www.wormbook.org.

[10] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The structure of the ventral nerve cord of *Caenorhabditis elegans*," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 275B, no. 938, pp. 327–348, 1976.

[11] Z. F. Altun and D. H. Hall, "Nervous system, general description." in *WormAtlas*. https://www.wormatlas.org/hermaphrodite/nervous/Neuroframeset.html.

[12] A. Parashar, R. Lycke, J. A. Carr, and S. Pandey, "Amplitude-modulated sinusoidal microchannels for observing adaptability in *C. elegans* locomotion," *Biomicrofluidics*, vol. 5, p. 024112, 2011.

[13] J. Karbowski, C. J. Cronin, A. Seah, J. E. Mendel, D. Cleary, and P. W. Sternberg, "Conservation rules, their breakdown, and optimality in caenorhabditis sinusoidal locomotion," *Journal of Theoretical Biology*, vol. 242, no. 3, pp. 652–669, 2006.

[14] G. N. Cox, J. S. Laufer, M. Kusch, and R. S. Edgar, "Genetic and phenotypic characterization of roller mutants of *Caenorhabditis elegans*," *Genetics*, vol. 95, no. 2, pp. 317–339, 1980.

[15] J. T. Pierce-Shimomura, B. L. Chen, J. J. Mun, R. Ho, R. Sarkis, and S. L. McIntire, "Genetic analysis of crawling and swimming locomotory patterns in *C. elegans*," *Proceedings of the National Academy of Sciences*, vol. 105, no. 52, pp. 20982–20987, 2008.

[16] M. Tahernia, M. Mohammadifar, and S. Choi, "Paper-supported high-throughput 3d culturing, trapping, and monitoring of *Caenorhabditis elegans*," *Micromachines*, vol. 11, no. 1, p. 99, 2020.

[17] J. Magnes, H. Hastings, M. Hulsey-Vincent, C. Congo, K. Raley-Susman, A. Singhvi, T. Hatch, and E. Szwed, "Chaotic markers in dynamic diffraction," *Applied Optics*, vol. 59, no. 22, pp. 6642–6647, 2020.

[18] A. Ozcan and E. McLeod, "Lensless imaging and sensing," *Annual Review of Biomedical Engineering*, vol. 18, pp. 77–102, 2016.

[19] J. Eckmann, S. Oliffson Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters (EPL)*, vol. 4, no. 9, pp. 973–977, 1987.

[20] T. Schreiber and A. Schmitz, "Discrimination power of measures for nonlinearity in a time series," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 55, no. 5, pp. 5443–5447, 1997.

[21] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D: Nonlinear Phenomena*, vol. 142, no. 3-4, pp. 346–382, 2000.

[22] M. A. Harrison and Y.-C. Lai, "Route to high-dimensional chaos," *Physical Review E*, vol. 59, no. 4, pp. R3799–R3802, 1999.

[23] J. Magnes, C. Congo, M. Hulsey-Vincent, H. Hastings, and K. Raley-Susman, "Live *C. elegans* diffraction at a single point," *Open Journal of Biophysics*, vol. 8, no. 3, pp. 155–162, 2018.

[24] F. L. Pedrotti, L. M. Pedrotti, and L. S. Pedrotti, *Introduction to Optics*, p. 461. Addison-Wesley, 3rd ed., 2006.

[25] "Understanding spacial filters." Edmund Optics. https://www.edmundoptics.com/knowledge-center/application-notes/lasers/understanding-spatial-filters/.

[26] "Three-axis spacial filters." Newport. https://www.newport.com/f/three-axis-spatial-filters?q=pinhole:relevance:isObsolete:false:-excludeCountries:US.

[27] C. H. Holbrow, J. N. Lloyd, J. C. Amato, E. Galvez, and M. E. Parks, *Modern Introductory Physics*, pp. 29–32. Springer Science & Business Media, second ed., 2010.

[28] J. Magnes and K. Raley-Susman, "Understanding fourier transforms through intuition building," *Proceedings of the European Science Education Research Association (ESERA), 13th Conference*, Bologna, Italy, August 2019.

[29] J. Miao, D. Sayre, and H. N. Chapman, "Phase retrieval from the magnitude of the fourier transforms of nonperiodic objects," *Journal of the Optical Society of America A*, vol. 15, no. 6, pp. 1662–1669, 1998.

[30] J. F. James, *A Student's Guide to Fourier Transforms with applications in physics and engineering.* Cambridge University Press, 2004.

[31] "Sinc function - MATLAB sinc." from Wolfram MathWorks–Help Center. https://www.mathworks.com/help/signal/ref/sinc.html.

[32] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers (AIEE)*, vol. 47, pp. 617–644, 1928.

[33] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, 1949.

[34] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.

[35] J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "Historical notes on the fast Fourier transform," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 76–79, 1967.

[36] D. N. Rockmore, "The FFT - an algorithm the whole family can use," *Computing in Science & Engineering*, vol. 2, no. 1, pp. 60–64, 2000.

[37] "Fastest fourier transform in the west (fftw)." from MIT–A free C subroutine library for computing DFT. https://www.fftw.org.

[38] M. Frigo and S. G. Johnson, "FFTW: An adaptive software architecture for the FFT," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1381–1384, 1998.

[39] D. P. Feldman, *Chaos and Fractals: An Elementary Introduction*, pp. 144–148. Oxford University Press, 2012.

[40] L. Decker, F. Cignetti, and N. Stergiou, "Complexity and human gait," *Revista Andaluza de Medicina del Deporte*, vol. 3, no. 1, pp. 2–12, 2010.

[41] C. J. Cellucci, A. M. Albano, and P. E. Rapp, "Comparative study of embedding methods," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 67, no. 6, p. 066210, 2003.

[42] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A: General Physics*, vol. 33, no. 2, pp. 1134–1140, 1986.

[43] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[44] S. G. Zhang, A. Singhvi, K. M. Susman, H. M. Hastings, and J. Magnes, "Dynamic markers for chaotic motion in *C. elegans*," *Nonlinear Dynamics, Psychology, and Life Sciences*, vol. 26, no. 1, pp. 21–43, 2022.

[45] H. D. Abarbanel and M. B. Kennel, "Local false nearest neighbors and dynamical dimensions from observed chaotic data," *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 47, no. 5, pp. 3057–3068, 1993.

[46] W. Liebert, K. Pawelzik, and H. G. Schuster, "Optimal embeddings of chaotic attractors from topological considerations," *Europhysics Letters*, vol. 14, no. 6, pp. 521–526, 1991.

[47] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43–50, 1997.

[48] R. Hegger and H. Kantz, "Improved false nearest neighbor method to detect determinism in time series data," *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 60, no. 4, pp. 4970–4973, 1999.

[49] M. B. Kennel and H. D. I. Abarbanel, "False neighbors and false strands: a reliable minimum embedding dimension algorithm," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 66, no. 2, p. 026209, 2002.

[50] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 6, pp. 3403–3411, 1992.

[51] A. Krakovaá, K. Mezeiová, and Budáčová, "Use of false nearest neighbours for selecting

variables and embedding parameters for state space reconstruction," *Journal of Complex Systems*, vol. 2015, 2015.

[52] E. Bradley and H. Kantz, "Nonlinear time-series analysis revisited," *Chaos*, vol. 25, no. 9, p. 097610, 2015.

[53] J.-P. Eckmann and D. Ruelle, "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems," *Physica D: Nonlinear Phenomena*, vol. 56, no. 2–3, pp. 185–187, 1992.

[54] J. Choi, B. Bae, and S. Kim, "Divergence in perpendicular recurrence plot; quantification of dynamical divergence from short chaotic time series," *Physics Letters A*, vol. 263, no. 4–6, pp. 299–306, 1999.

[55] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest lyapunov exponents from small data sets," *Physica D: Nonlinear Phenomena*, vol. 65, no. 1-2, pp. 117–134, 1993.

[56] Neel, "Calculating the Lyapunov exponent of a time series (with python code)." from WordPress. Accessed: July 2022, https://blog.abhranil.net/2014/07/22/calculating-the-lyapunov-exponent-of-a-time-series-with-python-code/.

[57] P. Yousefpoor, M. S. Esfahani, and H. Nojumi, "Looking for systematic approach to select chaos tests," *Applied Mathematics and Computation*, vol. 198, no. 1, pp. 73–91, 2008.

[58] T. Ahamed, A. C. Costa, and G. J. Stephens, "Capturing the continuous complexity of behavior in *Caenorhabditis elegans*," *Nature Physics*, vol. 17, no. 2, pp. 275–283, 2021.

[59] J. Theiler, "Efficient algorithm for estimating the correlation dimension from a set of discrete points," *Physical Review A*, vol. 36, no. 9, pp. 4456–4462, 1987.

[60] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, no. 5, pp. 346–349, 1983.

[61] "Measure of chaotic signal complexity - MATLAB correlationdimension." from Wolfram MathWorks–Help Center. https://www.mathworks.com/help/predmaint/ref/correlationdimension.html.

[62] W. Li and M. Jiang, "The complexity of electronic market," *International Academy of Business & Technology Conference*, 2006.

[63] C. J. McMahon, J. P. Toomey, and D. M. Kane, "Insights on correlation dimension from dynamics mapping of three experimental nonlinear laser systems," *Public Library of Science (PLoS) One*, vol. 12, no. 8, p. e0181559, 2017.

[64] D. P. Feldman, *Chaos and Fractals: An Elementary Introduction*, pp. 166–167. Oxford University Press, 2012.

[65] C. L. Webber Jr. and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *Journal of Applied Physiology*, vol. 76, no. 2, pp. 965–973, 1994.

[66] J. P. Zbilut and C. L. Webber Jr., "Embeddings and delays as derived from quantification of recurrence plots," *Physics Letters A*, vol. 171, no. 3-4, pp. 199–203, 1992.

[67] N. Marwan, "A historical review of recurrence plots," *The European Physical Journal Special Topics*, vol. 164, no. 1, pp. 3–12, 2008.

[68] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237–329, 2007.

[69] M. Thiel, M. Carmen Romano, and J. Kurths, "How much information is contained in a recurrence plot?," *Physics Letters A*, vol. 330, no. 5, pp. 343–349, 2004.

[70] E. Bradley and R. Mantilla, "Recurrence plots and unstable periodic orbits," *Chaos*, vol. 12, no. 3, pp. 596–600, 2002.

[71] D. F. Silva, V. M. A. de Souza, and G. E. A. P. A. Batista, "Time series classification using compression distance of recurrence plots," in *2013 IEEE 13th International Conference on Data Mining*, pp. 687–696, 2013.

[72] N. Marwan and J. Kurths, "Line structures in recurrence plots," *Physics Letters A*, vol. 336, no. 4-5, pp. 349–357, 2005.

[73] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1–4, pp. 77–94, 1992.

[74] T. Gautama, "Surrogate data." from MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/4612-surrogate-data.

[75] G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska, "Surrogate data for hypothesis testing of physical systems," *Physics Reports*, vol. 748, pp. 1–60, 2018.

[76] W. B. Wood, *The Nematode Caenorhabditis elegans*. Cold Spring Harbor Laboratory, University of Colorado, Boulder, 1988.