

AMERICAN UNIVERSITY OF BEIRUT

DATA-DRIVEN ANALYSIS OF THE
TRICHOPLAX ADHAERENS

by

LAYAL GERCESS TANNOUS

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Department of Physics
of Faculty of Arts and Sciences
at the American University of Beirut

Beirut, Lebanon
May 2023

AMERICAN UNIVERSITY OF BEIRUT

DATA-DRIVEN ANALYSIS OF THE
TRICHOPLAX ADHAERENS

by
LAYAL GERGESS TANNOUS

Approved by:

Dr. Sara Najem, Assistant Professor

Physics



Advisor

Dr. Mikhaël Balabane, Professor

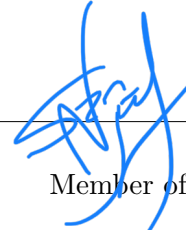
University of Paris- Institut Galilee - Math



Member of Committee

Dr. Leonid Klushin, Professor

Physics



Member of Committee

Date of thesis defense: May 15, 2023

ACKNOWLEDGEMENTS

I wish to acknowledge the invaluable and relentless support and guidance of my advisor, Professor Sara Najem. Her teachings extended far beyond the realm of physics, and I remain grateful for the knowledge and wisdom I acquired from her. I sincerely thank her for her mentorship, generosity, and valuable discussions. Her continuous efforts were instrumental in the completion of this work. I am particularly grateful for the time and effort she invested in my personal growth and for always challenging me and encouraging me to exceed my potential without doubting my abilities.

I would like to extend my thanks to Professor Leonid Klushin for his unwavering support since my undergraduate years. I thank him for always encouraging me to learn and to “keep my hands dirty”, a very healthy practice indeed! I also thank him for his insightful discussions and contributions to this work. Moreover, a heartfelt thank you to my committee member, Dr. Michael Balabane, for always making discussions pleasant and memorable. Finally, I thank Professor Jihad Touma, chairperson of the Department of Physics, for entrusting me with exciting projects pertaining to the department and its history. It was a pleasure to work closely with him and I am grateful for the opportunity.

I am appreciative of the generous and extensive discussions I had with Dr. Joseph Bakarji regarding various aspects of my work. His expertise, knowledge, and input were very helpful. I thank him for his valuable assistance and feedback.

I am also thankful for the enlightening discussions I had with Professor Miguel Alfonso Mendez and Professor Leila Issa.

I am deeply grateful for the lessons, experiences, and opportunities the Department of Physics has provided me. It has been a great privilege to be a part of this dynamic and supportive community. I would also like to take this opportunity to pay tribute to the late Professor Theodore Christidis, whose support meant a lot to me during my years of study. He remains to be an inspiration to me.

I would like to express my appreciation to my friends Nour, Dima, Rua, Nisrine,

Melissa, Arwa, Ali, Andre, Sasha, and colleagues for the wonderful memories we have shared together.

Finally, to my parents, I am forever indebted to you. Thank you for constantly showering me with your selfless love and sacrifices. I gift this work to you, Sally, and Bachir.

ABSTRACT OF THE THESIS OF

Loyal Gergess Tannous for Master of Science
Major: Physics

Title: Data-Driven Analysis of the *Trichoplax Adhaerens*

Biological systems, particularly animals, are a hotspot of highly sophisticated and non-trivial dynamics. They consist of a myriad of motile self-driven and simultaneously interacting entities. Understanding animal collectivity is of central importance for learning evolutionary foundations and inspiring artificially engineered systems. Several models have been developed to describe the dynamics of animal behavior. Recently, advances in data and image acquisition techniques have introduced novel ways of analyzing complex animal groups. In particular, data-driven approaches are increasingly being used to infer partial differential equations (PDEs) for systems of living organisms. In this work, we examine the behavior of a peculiar marine, multicellular animal, the *Trichoplax Adhaerens*. Our objective is to investigate its behavior in two distinct states, one with food present and another without. We use velocity field data generated from two experimentally recorded movies of the *T. Adhaerens*. We use two physically distinct models that integrate conservation laws, and principles of hydrodynamics to infer a set of PDEs modeling the evolution of the datasets at hand. Furthermore, we compare the performance of each model to determine which one best describes the animal's true biology. Finally, we perform modal analysis to identify dominant patterns of motion and derive a set of ordinary differential equations (ODEs).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	1
ABSTRACT	3
1 Introduction	9
2 The <i>Trichoplax Adhaerens</i>: A Sticky Plate of Hair	11
2.1 Structure	12
3 Theoretical Background	15
3.1 Formulation of the Problem	15
3.2 Model Selection	16
3.2.1 The Q-tensor Theory	16
3.2.2 What are liquid crystals?	16
3.2.3 Liquid Crystal Descriptors	16
3.2.4 Cells as Liquid Crystals	17
3.2.5 Equations of Motion	18
3.2.6 The Toner-Tu Model	21
3.3 Comparing the Two Models	23
4 Computational Methods	24
4.1 Data-Driven Modeling	24
4.1.1 Formulation	24
4.1.2 Linear Regression	25
4.1.3 Sparse Identification of Nonlinear Dynamics (SINDy)	26
4.2 Data Reduction Techniques: Modal Analysis	26
4.2.1 Proper Orthogonal Decomposition (POD)	26
4.2.2 Data Reshaping	27
4.2.3 Mathematical Framework [24]	29
4.2.4 Choosing the Optimal Basis	30
4.3 Data Description	32
4.4 Data Visualization	32
5 Results	35
5.1 Generated Partial Differential Equations	35
5.1.1 PDE for moving video - Linear Regression - Toner-Tu Model	35

5.1.2	PDE for moving video - Linear Regression - Tensor Theory . . .	38
5.1.3	PDE for moving video - Weak SINDy - Toner-Tu Model	41
5.1.4	PDE for moving video - Weak SINDy - Tensor Theory	42
5.1.5	PDE for stationary video - Linear regression - Toner-Tu	44
5.1.6	PDE for Stationary video - Linear regression - Tensor	47
5.1.7	PDE for stationary video - Weak SINDy - Toner Tu	49
5.1.8	PDE for stationary video - Weak SINDy - Tensor	50
5.2	Modal Analysis	50
5.2.1	Modal Analysis - Stationary movie	51
5.2.2	Modal Analysis - Moving movie	54
5.3	Limitations	57
6	Conclusion and Future Work	58
A	Numerical Details	59
A.1	Image preprocessing	59
A.2	Obtaining the density field	59
A.3	Particle Image Velocimetry (PIV)	60
A.4	Building the Q tensor	61
A.5	Filtering the Temporal Modes	61
	Bibliography	62

ILLUSTRATIONS

2.1	The <i>T. Adhaerens</i> (Schierwater, B. et al. 2003)	11
2.2	<i>T. Adhaerens</i> ' ciliated epithelial membrane (Wright, J.)	11
2.3	Global distribution of <i>T. Adhaerens</i> (Eitel, M. et al. 2013)	11
2.4	The dorsal and ventral epithelial layers of the <i>T. Adhaerens</i> (Bull, M. 2021)	12
2.5	Beating cilia mechanism (Raidt, J. et al. 2014)	13
2.6	Cross sectional view of <i>T. Adhaerens</i> (Smith, C. et al. 2014)	14
2.7	SEM cross sectional view of <i>T. Adhaerens</i> (Smith, C. et al. 2014)	14
3.1	Top view of the <i>T. Adhaerens</i> (Smith, C. et al 201)	15
3.2	Different types of liquid crystals	16
3.3	Director \mathbf{n}	17
3.4	Cell with unit vector \mathbf{n}_i	18
4.1	Velocity field components (stationary) at different snapshots	33
4.2	Velocity field components (moving) at different snapshots	34
5.1	True and predicted values for both training and testing sets, moving, Toner-Tu	38
5.2	True and predicted values for both training and testing sets, moving, tensor	40
5.3	True and predicted values for both training and testing sets, stationary, Toner-Tu	46
5.4	True and predicted values for both training and testing sets, stationary, tensor	48
5.5	Relative energy amplitudes as a function of mode number, stationary movie	51
5.6	Relative energy amplitudes for the first 10 modes, stationary movie	51
5.7	Spatial structures of the first 4 modes, stationary movie	52
5.8	Temporal structures of the first 4 modes, stationary movie	53
5.9	Relative energy amplitudes as a function of mode number, moving video	54
5.10	Relative energy amplitudes for the first 10 modes, moving video	54
5.11	Spatial structures of the first 6 modes, moving video	55
5.13	Temporal structures of the first 6 modes, moving video	57

A.1	Glitch in image acquisition	59
A.2	Density fields for the first time frame	60
A.3	Building the Q tensor from the velocity data	61

TABLES

4.1	Description of the two placozoan movies used in this work	32
5.1	Terms and coefficients of the PDE obtained for U moving using a linear regression fit, Toner-Tu	36
5.2	Coefficient of performance for U moving, linear regression, Toner-Tu .	36
5.3	Terms and coefficients of the PDE obtained for V moving using a linear regression fit (Toner-Tu)	37
5.4	Coefficient of performance for V moving, linear regression, Toner-Tu .	37
5.5	Terms and coefficients of the PDE obtained for U moving using a linear regression fit, Tensor theory	39
5.6	Coefficient of performance for U moving, linear regression, Tensor theory	39
5.7	Terms and coefficients of the PDE obtained for V moving using a linear regression fit, Tensor theory	39
5.8	Coefficient of performance for V moving, linear regression, Tensor theory	40
5.9	Terms and coefficients of the PDE obtained for U stationary using a linear regression fit, Toner-Tu	44
5.10	Coefficient of performance for U stationary, linear regression, Toner-Tu	44
5.11	Terms and coefficients V stationary, linear regression, Toner Tu . . .	45
5.12	Coefficient of performance for V stationary, linear regression, Toner-Tu)	45
5.13	Terms and Coefficients for U stationary using a linear regression fit, Tesnor	47
5.14	Coefficient of performance for U stationary, linear regression, tensor .	47
5.15	Terms and Coefficients for V stationary, linear regression, tesnor . . .	47
5.16	Coefficient of performance for V stationary, linear regression, tensor .	47

CHAPTER 1

INTRODUCTION

Biological systems, particularly animals, are a hotspot of highly sophisticated and non-trivial dynamics. They consist of a myriad of motile self-driven and simultaneously interacting entities. These living units are capable of self-assembling and displaying unidirectional motion. Coordinated movement is manifested at different size scales and serves many purposes including mating, scaring off prey, and navigating difficult terrains [1]. We thus see that there are several facets of animal activity. This introduces a level of complexity that imposes challenges when it comes to understanding and modeling the resulting emergent behavior.

The collective behavior of groups of animals has always been an active research topic among different disciplines. Scientists, whether in biology, physics, or engineering, have always been fascinated by the synchronized and large-scale patterns exhibited by animal aggregations. Understanding animal collectivity is of central importance for learning evolutionary foundations and inspiring man-made and artificially engineered systems [2].

Studying collective behavior is considered to be the motivation for the still-growing field of active matter physics [3], a theoretical framework for modeling living systems with internal degrees of freedom. The field was first leveraged by B. Finlayson and L. E. Scriven who claimed that biological matter can exhibit hydrodynamic instabilities due to active stresses [4]. Further models were developed later on to cater to a variety of systems like the Vicsek model, the Toner and Tu model, as well as models for active nematics. Active matter spans systems of different size scales with examples ranging from bacterial colonies all the way to human crowds [5].

The diverse models devised to understand various biophysical systems can be classified into three main categories: continuum, agent-based, and rule-based models. The continuum model is a macroscopic model that describes collective behavior as a continuous fluid-like medium. The system variables like density and velocity are presumed to vary smoothly over space and time. Moreover, the system is modeled by a set of partial differential equations (PDEs) relating its macroscopic properties to its microscopic ones. This model, however, assumes that the individual components

are identical and so it may fail to identify heterogeneities. Agent-based models describe the system with a set of coupled ordinary differential equations (ODEs) instead of PDEs. Although these models can effectively capture various types of interactions among individual organisms, the large number of equations they produce can be expensive and pose difficulties when attempting to analytically characterize the system. Finally, rule-based models, like agent-based models, treat each animal as a separate entity, but they do not involve differential equations. Instead, these models assume that the system interacts based on predefined rules.

Combining the aforementioned models with high-resolution empirical data paves the road to data-driven descriptions of physical systems. Data-driven techniques are powerful tools widespread across multidisciplinary fields. They allow the identification of patterns and relationships among complex datasets providing valuable insights into biological functions and interactions. Integrating physical models with machine learning techniques provides novel approaches to understanding natural processes leading to new research questions, better ways to design experiments, and more effective ways of understanding intricate dynamics.

In this work, we apply data-driven techniques to a peculiar living system, the *Trichoplax Adhaerens*. It is a primitive multi-cellular marine animal found worldwide in temperate and subtropical waters [6]. Despite its biological minimalism, the placozoan is still capable of decision-making and coordinated behavior which has attracted the attention of many scientists. We will be studying the *T. Adhaerens* in two different states, one in which food is present and another in the absence of food. In our description of the animal's behavior, we utilize two distinct theoretical models that incorporate notions from hydrodynamics and non-equilibrium statistical physics. By using physics-inspired terms, we ensure that our results are physically reasonable. Our objective is to evaluate and compare the performance of these models to determine which offers a more accurate description of the dynamics. The work can be divided into two major parts. First, we infer a set of PDEs describing the animal's behavioral evolution over time. Second, we perform mode analysis to identify the dominant types of motion and deduce a set of ordinary differential equations (ODEs) describing the evolution of these modes.

CHAPTER 2

THE *Trichoplax Adhaerens*: A STICKY PLATE OF HAIR

The *Trichoplax Adhaerens*, or *T. Adhaerens* for short (Fig. 2.1), is a primitive multicellular animal belonging to the phylum Placozoa. It was first discovered by the German zoologist, Franz Eilhard Schulze, in 1883. Its name derives from the Greek words “Thrix” (hair) and “Plax” (flat plate) and refers to its ciliated epithelial membrane (Fig. 2.5). The *T. Adhaerens* can be found worldwide in temperate and subtropical waters [6] (Fig. 2.3).

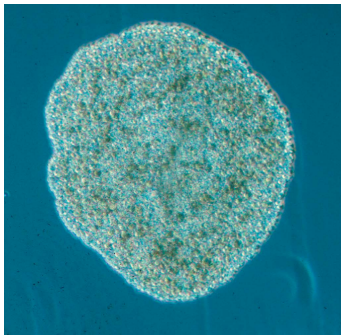


Figure 2.1: The *T. Adhaerens*
(Schierwater, B. et al. 2003)

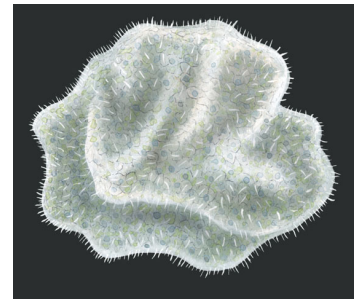


Figure 2.2: *T. Adhaerens*' ciliated epithelial membrane (Wright, J.)

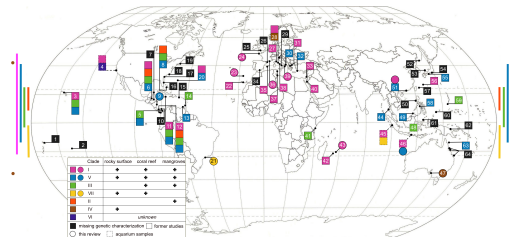


Figure 2.3: Global distribution of *T. Adhaerens*
(Eitel, M. et al. 2013)

The significance of the *T. Adhaerens* stems from biological and evolutionary considerations. It is deemed to play an important evolutionary role in the transition from unicellularity to multicellularity. Physically, it is believed to be the simplest animal to exist [7] consisting of only six different types of cells and lacking nervous, digestive, and muscular systems. Despite its morphological minimalism, the animal is capable of coordinated locomotion, ciliary flocking, external digestion, and decision-making. These features make the *T. Adhaerens* an excellent experimental model organism for various studies including motion coordination in the absence of hierarchical organization [8], information propagation in groups [8], mechanobiology [9], and tissue resilience mechanism under stresses.

In order to build a realistic model that captures the true biology of interactions in the *T. Adhaerens*, it serves well to have an understanding of its structure and response to the surrounding environment:

2.1 Structure

The *T. Adhaerens* has a roughly disk-shaped outline, with typical dimensions ranging from **3** to **5 mm** in diameter and about **20 μm** in thickness [10]. It is mostly found gliding on surfaces where the adhesion helps it maintain a more or less flat structure. Its outer cellular membrane, the epithelial membrane, is composed of two mechanically distinct tissues [9], the dorsal (top) layer, and the ventral (bottom) layer that meet at the edge (Fig. 2.4).

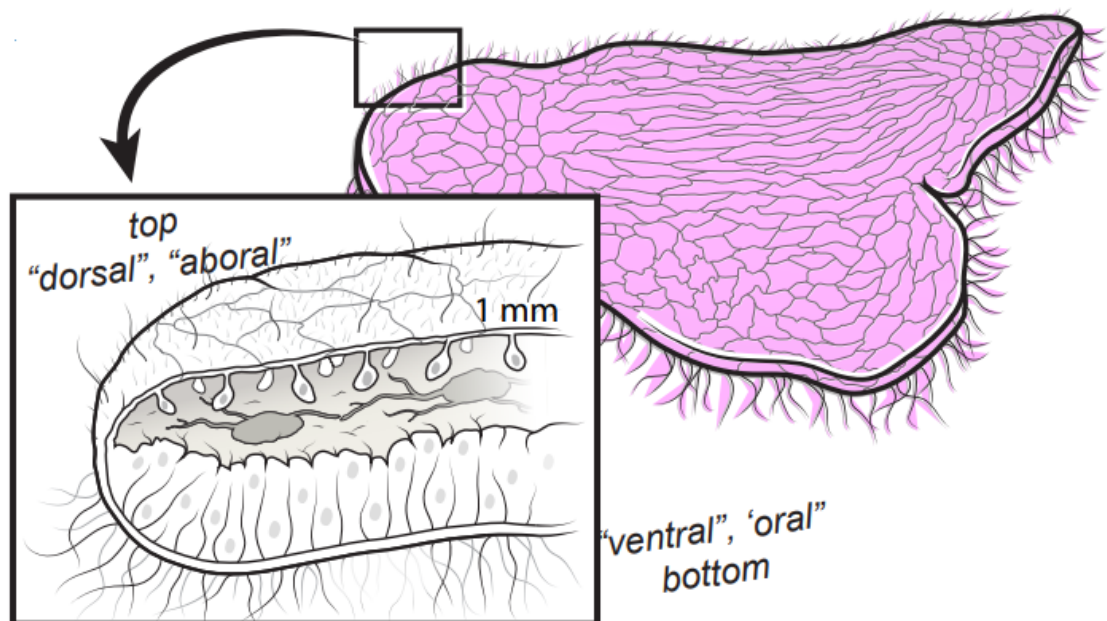


Figure 2.4: The dorsal and ventral epithelial layers of the *T. Adhaerens* (Bull, M. 2021)

The dorsal layer exhibits ultra-fast contractions that control the tissue’s architecture and consequently the animal’s shape. Meanwhile, the ventral layer is responsible for maintaining tissue cohesion and avoiding ruptures [9].

The *T. Adhaerens*’ activity is generated by the cilia that propel it through a beating mechanism. Individual cilia bend and push water from the surrounding which in turn pushes back creating a force that propagates along the body (Fig. 2.5). As a result, intercellular interactions and coupling to the environment rise simultaneously.

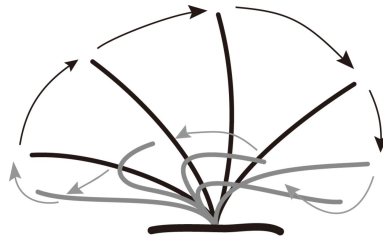


Figure 2.5: Beating cilia mechanism
(Raidt, J. et al. 2014)

In the absence of food, the animal moves randomly with its cilia showing no preferential alignment. On the other hand, it displays unidirectional motion in the vicinity of algae on which forages. In both cases, the organism may be subject to reorientations under the effect of rotational torques [9].

The emergent behavior of the *T. Adhaerens* is a manifestation of cellular mechanisms at work. As mentioned previously, it consists of 6 different types of cells, 4 of which are embedded in its membrane. The cells are connected by junctions that hold their positions relative to one another. Below we briefly describe each kind of cell: [11]

1. The dorsal epithelial cells are monociliated and reside in the dorsal membrane. Their shapes vary between circular and elliptical reflecting the animal’s overall shape.
2. The ventral epithelial cells populate the ventral membrane and are more ciliated compared to the dorsal cells. They constitute **72** % of the total cellular distribution.
3. The fiber cells occupy the space between the epithelial layers and are regularly spaced. Some researchers claim that these cells are contractile and owe the animal its viscoelastic properties.
4. The gland cells fulfill the role of neurons.
5. The crystal cells serve a sensory role.

6. The lipophil cells play the digestive role.

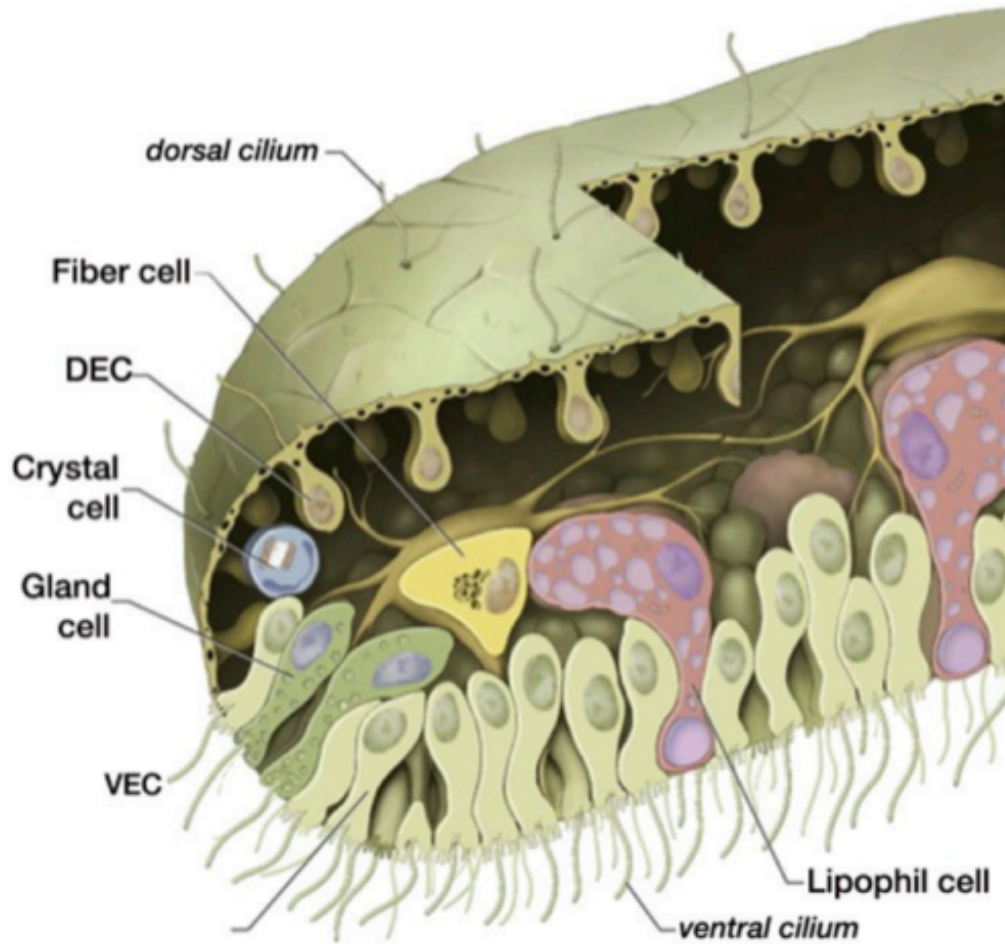


Figure 2.6: Cross sectional view of *T. Adhaerens*
(Smith, C. et al. 2014)

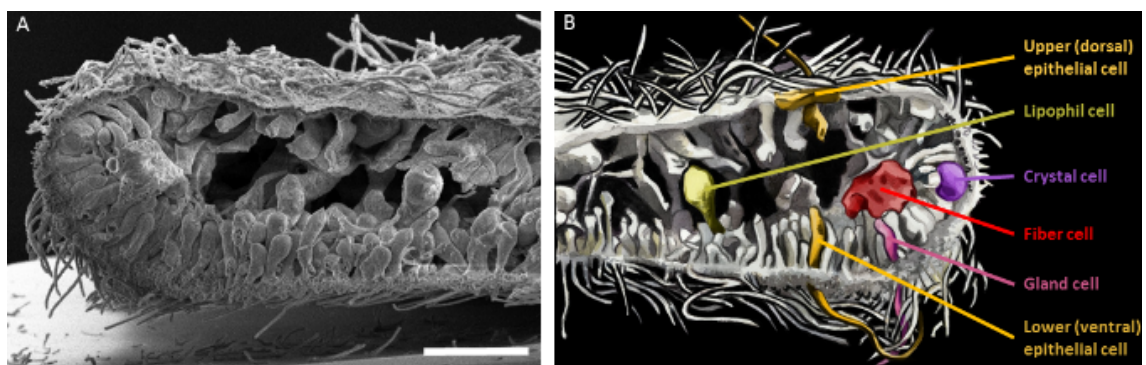


Figure 2.7: SEM cross sectional view of *T. Adhaerens*
(Smith, C. et al. 2014)

CHAPTER 3

THEORETICAL BACKGROUND

3.1 Formulation of the Problem

In this work, we use movies that show a top view of the *T. adhaerens* (section 4.3). Thus, the datasets used correspond mainly to the velocity vector field of the dorsal layer whose cells have a polygonal/disc-like apolar structure (Fig. 3.1).

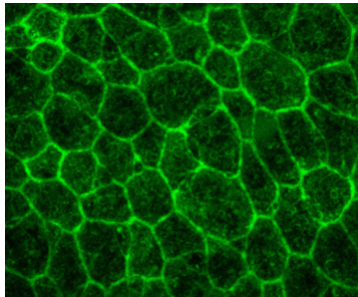


Figure 3.1: Top view of the *T. Adhaerens*
(Smith, C. et al 201)

Studying the temporal evolution of these vector fields gives us insight into the evolution of the direction of motion or orientation of the animal’s collection of cells. We would like to test the hypothesis that since the cells are apolar active particles, they should be well described by a nematic order parameter. To verify the validity of this hypothesis, we present below two different continuum models to describe the hydrodynamics of the “flocking of cells”. The first model describes the particles as nematic liquid crystals, while the other incorporates polarity by including a polar order parameter in the description of the animal’s behavior.

The ultimate goal is to benchmark these two models and check which one better describes our system. We do this for two different states of the animal, one in which it moves without food and one in which food is present.

3.2 Model Selection

3.2.1 *The Q-tensor Theory*

The Q-tensor theory presents a continuum model for the study of nematic liquid crystals (LCs). It was developed to describe the orientational ordering in LCs. The degree of order of the molecules is captured by a second-rank tensor called the Q-tensor. Before we write down the equations of motion for this model, we will review some fundamental concepts related to liquid crystals.

3.2.2 *What are liquid crystals?*

Liquid crystals are molecules, mostly organic, with a rod or disc-like shape. They are a phase of matter on their own. Phases of matter are characterized by symmetries [12]. For example, a crystal is not isotropic like a liquid or gas since it is only invariant under translations of its lattice size [13]. In general, a solid is invariant under fewer operations compared to a liquid and thus it has fewer symmetries. For LCs, given their extended shape, we need to take into account not only their translational and rotational motion but also their orientations.[12]. LCs have properties of both solids and liquids. Similar to solids, they exhibit long-range order. and like liquids, they flow and lack positional order. There are several phases of LCs, the most common and widely studied being the nematic phase (Fig. 3.2).

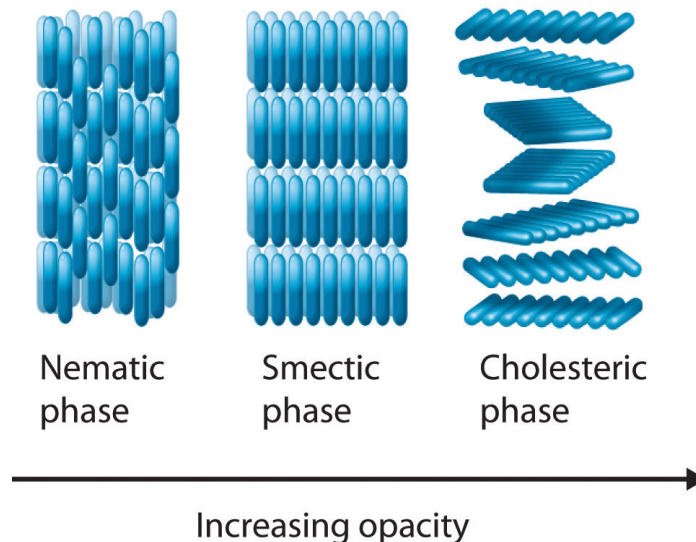


Figure 3.2: Different types of liquid crystals

3.2.3 *Liquid Crystal Descriptors*

The director \mathbf{n} is a LC descriptor that represents the average orientation of the long axes of the molecules (Fig. 3.3).



Figure 3.3: Director \mathbf{n}

The quantity that describes the amount of order in the nematic LC phase is the Q-tensor. Its general expression is given by:

$$Q \propto \frac{d}{d-1} \times \langle \vec{n} \otimes \vec{n} - I \rangle \quad (3.1)$$

where:

d is the dimension, $d = 2$ in our case

\vec{n} is the director

I is the identity matrix

3.2.4 *Cells as Liquid Crystals*

There have been many efforts in the soft-matter physics community to model cellular tissues as liquid crystals [14]. The motivation for that was that certain epithelial cells share common features with liquid crystals. For example, alignment has been observed in dense cellular packing as well as topological defects. [15].

Describing cells as active nematic particles proves helpful in detecting orientational order in epithelia. Researchers care to track ordering since it serves several crucial biological functions like wound healing and cancer progression [16].

In this work, we will model each cell as a liquid crystal having a unit vector $\mathbf{n}_i = (\mathbf{u}, \mathbf{v})$ (Fig. 3.4).

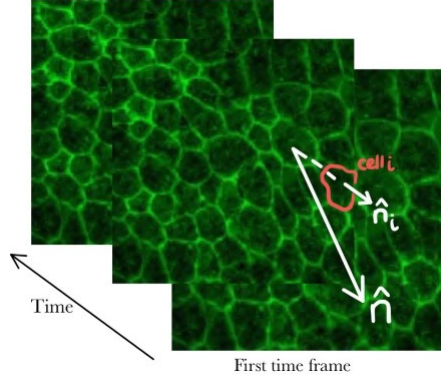


Figure 3.4: Cell with unit vector \mathbf{n}_i

3.2.5 Equations of Motion

The continuum equations are given by the Beris-Edwards nematohydrodynamic equations [17]. They are compressible Navier-Stokes-like, but with an additional layer of complexity owing to the cells' shape, self-propelled nature, and coupling to the underlying flow. The complete set constitutes three equations:

I) The continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \quad (3.2)$$

II) The momentum equation:

$$\frac{D\mathbf{V}}{Dt} = \underbrace{\frac{\partial \mathbf{V}}{\partial t}}_{(1)} + \underbrace{\mathbf{w} \cdot \nabla \mathbf{V}}_{(2)} = - \underbrace{\nabla P}_{(3)} + \mu \nabla \cdot \underbrace{(\nabla \mathbf{V} + (\nabla \mathbf{V})^T)}_{(4)} - \underbrace{\zeta \nabla Q}_{(5)} \quad (3.3)$$

where:

\mathbf{V} denotes the velocity vector field whose x and y components are u and v respectively. It is important to note that u and v each depends on both x and y , i.e. $u(x, y)$ and $v(x, y)$

\mathbf{w} denotes the flow field. Here, $\mathbf{w} = \mathbf{V}$

Term 1 denotes the partial time derivative of \mathbf{V}

Term 2 denotes the convective term. In this theoretical framework, we will work in the co-moving, co-rotational frame of the cells, thus we expect term 2 to be negligible compared to other terms.

Term 3 is the gradient of the pressure. We will write the pressure as a combination of powers of the local density field: $P(\rho) = \sum_{n=0} a_n (\rho - \rho_0)^n$ where

ρ_0 is the mean density. We will expand the pressure up to the fourth order in density:

$$P(\rho) = a_0 + a_1(\rho - \rho_0) + a_2(\rho - \rho_0)^2 + a_3(\rho - \rho_0)^3 + a_4(\rho - \rho_0)^4 \quad (3.4)$$

$$= A_0 + A_1\rho + A_2\rho^2 + A_3\rho^3 + A_4\rho^4 \quad (3.5)$$

where:

$$A_0 = a_0 - a_1\rho_0 + a_2\rho_0^2 + a_3\rho_0^3 + a_4\rho_0^4$$

$$A_1 = a_1 - 2a_2\rho_0 + 3a_3\rho_0^2 - 4a_4\rho_0^3$$

$$A_2 = a_2 - 3a_3\rho_0 + 6a_4\rho_0^2$$

$$A_3 = a_3 - 4a_4\rho_0$$

$$A_4 = a_4$$

Term 4 is added to reflect the compressible nature of the viscoelastic animal under study. It is the rate-of-strain tensor (up to a factor of 2) with μ representing the dynamic viscosity. This term quantifies the rate of change in the animal's deformation due to both internal and external forces.

\Rightarrow Terms 3 and 4 represent gradients of passive stresses. These are stresses that we find in the usual Navier-Stokes equation. In order to extend the model to describe active matter, we add term 6:

Term 6 is the gradient of the active stress tensor which is proportional to the tensor Q . The constant of proportionality ζ measures the strength of the activity. $\zeta > 0$ describes an extensile or a puller system, while $\zeta < 0$ describes a contractile or pusher system.

Below we work out the individual terms in equation 3.3:

$$\textcircled{1} = \begin{pmatrix} \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial t} \end{pmatrix} \quad (3.6)$$

$$\textcircled{2} = [(u \ v) \cdot \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix}], \begin{pmatrix} u \\ v \end{pmatrix} \quad (3.7)$$

$$= (u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}) \times \begin{pmatrix} u \\ v \end{pmatrix} \quad (3.8)$$

$$= \begin{pmatrix} u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \\ u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} \end{pmatrix} \quad (3.9)$$

$$\textcircled{3} = \left(\frac{\partial}{\partial x} \quad \frac{\partial}{\partial y} \right) \times (A_0 + A_1\rho(x, y) + A_2\rho^2(x, y) + A_3\rho^3(x, y) + A_4\rho^4(x, y)) \quad (3.10)$$

$$= \begin{pmatrix} A_1 \frac{\partial \rho}{\partial x} + A_2 \frac{\partial \rho^2}{\partial x} + A_3 \frac{\partial \rho^3}{\partial x} + A_4 \frac{\partial \rho^4}{\partial x} \\ A_1 \frac{\partial \rho}{\partial y} + A_2 \frac{\partial \rho^2}{\partial y} + A_3 \frac{\partial \rho^3}{\partial y} + A_4 \frac{\partial \rho^4}{\partial y} \end{pmatrix} \quad (3.11)$$

$$\textcircled{4} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{pmatrix} + \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \quad (3.12)$$

$$= \begin{pmatrix} 2\frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \\ \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} & 2\frac{\partial v}{\partial y} \end{pmatrix} \quad (3.13)$$

$$\nabla \textcircled{4} = \left(\frac{\partial}{\partial x} \quad \frac{\partial}{\partial y} \right) \times \begin{pmatrix} 2\frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \\ \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} & 2\frac{\partial v}{\partial y} \end{pmatrix} \quad (3.14)$$

$$= \begin{pmatrix} 2\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 u}{\partial y^2} \\ \frac{\partial^2 u}{\partial x \partial y} + \frac{\partial^2 v}{\partial x^2} + 2\frac{\partial^2 v}{\partial y^2} \end{pmatrix} \quad (3.15)$$

$$\textcircled{5} = \left(\frac{\partial}{\partial x} \quad \frac{\partial}{\partial y} \right) \times \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \quad (3.16)$$

$$= \begin{pmatrix} \frac{\partial Q_{11}}{\partial x} + \frac{\partial Q_{12}}{\partial y} \\ \frac{\partial Q_{21}}{\partial x} + \frac{\partial Q_{22}}{\partial y} \end{pmatrix} \quad (3.17)$$

After having laid out the terms, we can now write the equations for the time evolution of both u and v :

$$u_t = -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} - A_1 \frac{\partial \rho}{\partial x} - A_2 \frac{\partial \rho^2}{\partial x} - A_3 \frac{\partial \rho^3}{\partial x} - A_4 \frac{\partial \rho^4}{\partial x} + \mu \left(2\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 u}{\partial y^2} \right) - \zeta \left(\frac{\partial Q_{11}}{\partial x} + \frac{\partial Q_{12}}{\partial y} \right) \quad (3.18)$$

We do the same for v :

$$v_t = -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - A_1 \frac{\partial \rho}{\partial y} - A_2 \frac{\partial \rho^2}{\partial y} - A_3 \frac{\partial \rho^3}{\partial y} - A_4 \frac{\partial \rho^4}{\partial y} + \mu \left(\frac{\partial^2 v}{\partial x^2} + 2\frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 u}{\partial x \partial y} \right) - \zeta \left(\frac{\partial Q_{21}}{\partial x} + \frac{\partial Q_{22}}{\partial y} \right) \quad (3.19)$$

III) The Q-tensor equation:

The dynamics of the Q-tensor is described by: [18]

$$\frac{\partial Q}{\partial t} + \mathbf{u} \cdot \nabla Q - S = D_r H \quad (3.20)$$

$$\frac{\partial Q}{\partial t} + \mathbf{u} \cdot \nabla Q - (\Omega \cdot Q - Q \Omega) + \lambda E = -D_r \frac{\delta \mathcal{F}}{\delta Q} \quad (3.21)$$

where:

Ω and E are the vorticity and rate of strain tensors respectively:

$$\Omega = \frac{1}{2}(\nabla u - (\nabla u)^T) \text{ and } E = \frac{1}{2}(\nabla u + (\nabla u)^T)$$

λ is the flow alignment parameter

D_r is the rotational diffusion constant

The first two terms on the left-hand side of equation 3.21 are the terms arising due to advection by the flow. The third term, S , is the corotation term, as the particles will respond to gradients in the flow. Finally, the term on the right-hand side is the relaxation term. It describes the relaxation dynamics of the nematic tensor, Q , to a minimum of the free energy \mathcal{F} .

\mathcal{F} is the Landau - de Gennes free energy. It is written as a Taylor expansion of the tensor Q :

$$\mathcal{F} = \frac{a}{2}Q^2 + \frac{b}{3}Q^3 + \frac{c}{4}Q^4 + \frac{K}{2}(\nabla Q)^2 + \dots \quad (3.22)$$

Equation 3.21 specifies the terms upon which the time evolution of each tensor element, $Q_{11}, Q_{12}, Q_{21}, Q_{22}$, is dependent:

$$Q_{11t} \propto uQ_{11x}, uQ_{12y}, u_yQ_{21}, v_xQ_{21}, v_xQ_{12}, u_yQ_{12}, u_x, Q^2, Q^3, \nabla^2 Q \quad (3.23)$$

$$Q_{12t} \propto uQ_{21x}, uQ_{22y}, v_yQ_{12}, u_xQ_{12}, u_yQ_{22}, v_x, u_y, u_yQ_{11}, v_xQ_{22}, v_xQ_{11}, Q^2, Q^3, \nabla^2 Q \quad (3.24)$$

$$Q_{21t} \propto vQ_{11x}, vQ_{12y}, u_xQ_{21}, v_yQ_{21}, v_xQ_{11}, v_xQ_{22}, u_yQ_{22}, u_yQ_{22}, u_y, v_x, Q^2, Q^3, \nabla^2 Q \quad (3.25)$$

$$Q_{22t} \propto vQ_{21x}, vQ_{22y}, v_xQ_{12}, u_yQ_{12}, u_yQ_{21}, v_xQ_{21}, v_y, Q^2, Q^3, \nabla^2 Q \quad (3.26)$$

3.2.6 The Toner-Tu Model

The Toner-Tu model is a quantitative continuum model for the flocking behavior of groups of animals. It was first devised by Yuhai Tu and John Toner in 1995 to describe how birds fly and self-organize [19]. Its applications were later broadened to describe a variety of other non-equilibrium systems such as schools of fish, insect swarms, animal herds, and molds...[20]. This model is powerful as it provides a mathematical framework that predicts the emergence of an ordered phase of flocks where members spontaneously move together with the same mean velocity.

To establish the theory, we must first lay out the symmetry and conservation laws that solely restrict the terms to those that are physically relevant. In the isotropic case, the individuals have no preferred direction of motion, they are free to move in any direction and thus they have continuous rotational symmetry. Moreover, since they are constantly consuming energy to fuel their activity and change their direction of motion, neither energy nor momentum is conserved. This implies that the only conserved quantity is the number of particles.

Second, we choose the hydrodynamic variables describing the “flow” of the flock. In this model, they are the coarse-grained density $\rho(\vec{r}, t)$ and velocity $\vec{v}(\vec{r}, t)$ fields. The general equations of motion are:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \quad (3.27)$$

$$\underbrace{\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla) \vec{v}}_{\textcircled{1}} = \underbrace{\alpha \vec{v}}_{\textcircled{2}} - \underbrace{\beta |\vec{v}|^2 \vec{v}}_{\textcircled{3}} - \underbrace{\frac{\nabla P}{\rho}}_{\textcircled{4}} + \underbrace{D_L \nabla (\nabla \cdot \vec{v}) + D_1 \nabla^2 \vec{v} + D_2 (\vec{v} \cdot \nabla)^2 \vec{v}}_{\textcircled{5}} + \underbrace{\vec{f}}_{\textcircled{6}} \quad (3.28)$$

Equation 3.27 is just the mass conservation equation. To fully understand the implications of the Toner-Tu theory, we will go over each of the terms in equation 3.28:

Term 1 is the convective derivative of the velocity field \vec{v}

Term 2 is the self-propulsion term. It also reflects the alignment of the individuals with their neighbors. $\alpha < 0$ promotes a disordered state, while $\alpha > 0$ promotes an ordered state.

Term 3 is the damping. It reflects the dissipation of energy due to self-propulsion (e.g. flapping of the wings in the case of birds). Individual components lose kinetic energy as they internally drive their own motion; as a result, their velocities change to gradually align along the average orientation. Damping plays a crucial role in defining and predicting the emergence of order. In the absence of damping, each component continues to move at its own initial velocity leading to a zero average velocity and a collective disordered phase.

→ Terms 2 and 3 ensure that $|\langle \vec{v} \rangle| = \sqrt{\frac{\alpha}{\beta}}$ in the ordered state.

Term 4 is the pressure gradient

Term 5 is the diffusion term. There are three different diffusion constants to capture the interplay of the various factors affecting the overall movement: self-propulsion, random fluctuations, inherent directional bias, and coupling with neighbors.

Term 6 is the stochastic term. It introduces fluctuations to the individual velocities since in reality, the individuals are prone to various sources of noise.

The pressure term is expanded in terms of the density similar to equation 3.4. We can expand equation 3.28 to know the exact terms we expect to obtain in the PDE like we did in section 3.2.5.

3.3 Comparing the Two Models

The Q tensor theory and the Toner-Tu model are similar in several ways. They both are nonlinear in velocity and resemble Navier-Stoke equations for a compressible flow. Moreover, both of their velocity equations contain a parameter whose sign dictates whether we are in an ordered or disordered state.

The main difference is that the Tensor theory uses a tensor order parameter to quantify order, whereas the Toner-Tu uses a polar order parameter \mathbf{p} . Moreover, elastic contributions arise naturally and appear more clearly in the Q tensor equations while not being explicitly evident in the Toner-Tu equations.

CHAPTER 4

COMPUTATIONAL METHODS

4.1 Data-Driven Modeling

The constant advancement of experimental apparatus and the phenomenal progress in data acquisition have contributed to the increasing availability of detailed empirical data. As a result, data-driven modeling has become ubiquitous across multiple disciplines, particularly ethology. The main goal of such techniques is to extract physically relevant information from data available to complex systems.

Researchers are interested in finding the equations of motion of biological systems because it allows them to predict their behavior and contribute to scientific advancement. However, it is often the case that first principles alone may not be sufficient or might even fail to capture the true dynamics. The limitations of the traditional theoretical methods have catalyzed the development of new robust computational methods that facilitate the characterization of dynamical systems. One such example is data-driven discovery of partial differential equations (PDE). Suppose the variable of interest is $\mathbf{u}(\mathbf{x}, \mathbf{y})$, we write the general PDE as:

$$u_t = F(u, u^2, u_x, u_{xx}, u_{xy}, x, \lambda, \dots) \quad (4.1)$$

F is a non-linear function of u , its space and time derivatives, and parameters in λ . The terms on the right-hand side of equation 4.1 are physics informed in the sense that they respect the symmetries and conservation laws constraining the model.

4.1.1 Formulation

After specifying the candidate terms, the required derivatives are computed numerically, and stacked as columns in a matrix, Θ :

$$\Theta = \begin{pmatrix} | & | & | & | & | & | & | \\ u & u^2 & u_x & u_y & u_{xx} & u_{xy} & \dots \\ | & | & | & | & | & | & | \end{pmatrix} \quad (4.2)$$

u_t is also computed numerically and we write:

$$u_t = \Theta\xi \quad (4.3)$$

ξ is the vector of coefficients corresponding to each term in Θ . It is unknown, and we wish to find it by treating equation 4.3 as a linear regression problem.

4.1.2 Linear Regression

Linear regression is a widely used statistical learning method [21]. It usually involves a dependent variable Y , and one (simple linear regression) or more (multiple linear regression) independent variables or features, X_i . The goal is to establish a linear relationship between both variables and to predict the values of Y given X for different observations. For our purposes, we will focus on multiple linear regression. We define:

1) $\mathbf{X}(n \times p)$: The feature matrix of p features recorded over n observations

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \quad (4.4)$$

where the value of X_{ij} refers to the j^{th} feature corresponding to the i^{th} observation.

2) $\mathbf{Y}(n \times 1)$: The response vector

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (4.5)$$

where y_i denotes the value of response for the i^{th} observation.

We thus write the linear relationship:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_p X_{ip} = \beta_0 + \sum_{j=1}^p X_{ij} \beta_j \quad (4.6)$$

where the β s are the regression coefficients. We can combine them into a vector and rewrite 4.6 as a product of matrices:

$$Y = X\beta \quad (4.7)$$

The β s are estimated using ordinary least squares (OLS). The difference between the actual and predicted Y values are called residuals. The set of coefficients that best describe the model are those that minimize the sum of square residuals (RSS):

$$RSS = \sum_i^n (Y_i - \beta_0 + \sum_{j=1}^p X_{ij} \beta_j)^2 \quad (4.8)$$

Equation 4.3 is analogous to 4.7 where Θ is the feature matrix, u_t is the response/predicted vector, and ξ is the vector of coefficients that we wish to fit for.

4.1.3 *Sparse Identification of Nonlinear Dynamics (SINDy)*

Linear regression has its limitations when it comes to imposing physical and symmetry-inspired constraints which led to the development of more versatile fitting methods. One robust and commonly used technique is the Sparse Identification of Nonlinear Dynamics (SINDy) [22]. It is also a data-driven method for the discovery of sparse differential equations. Essentially, it applies sparse regression for the identification of the coefficients. A threshold value is set, below which all terms are forced to be zero.

A variation of SINDy, the weak SINDy (WSINDy), was developed particularly to deal with noisy data yielding robust and more accurate models. In this work, we use WSINDy to generate constraint partial differential equations. The two main optimizers that we used are:

- 1) Sequentially thresholded least squares (STLSQ): It is an optimization algorithm that takes in two parameters: a threshold value that defines the minimum magnitude for the coefficients and an optional L_2 (ridge) regularizer.
- 2) Sparse relaxed regularized regression (SR3): It is an optimization algorithm that takes in two parameters: a threshold that determines the strength of regularization, and a relaxation parameter, nu , which controls the balance between the data fitting and sparsity constraints.

4.2 Data Reduction Techniques: Modal Analysis

Real-world data is often multivariate and difficult to visualize or classify. This necessitates the development of meaningful reduction and data representation tools.[23]

Data reduction techniques provide an elegant means of decomposing high-dimensional data into simpler smaller-dimensional contributions called modes. These modes are the elements of an ordered orthonormal basis. They capture the salient features of complex systems by identifying relevant dynamic patterns.

These tools find application in several fields, including but not limited to fluid dynamics, data compression, data analysis, and image filtering [24].

4.2.1 *Proper Orthogonal Decomposition (POD)*

Proper Orthogonal Decomposition (POD) is one of the most popular data decomposition methods. In terms of data-driven applications, John L. Lumey was the first to introduce it to the fluid dynamics community in 1967 when he was analyzing turbulent flows [25]. The POD outperforms other techniques because of its ability to handle significantly complex data [23] and reduce it to a minimal number of basis functions [26]. In fact, its basis functions are designed to ensure optimal approximation of the data [27], and this is the origin behind “proper” in the nomenclature.

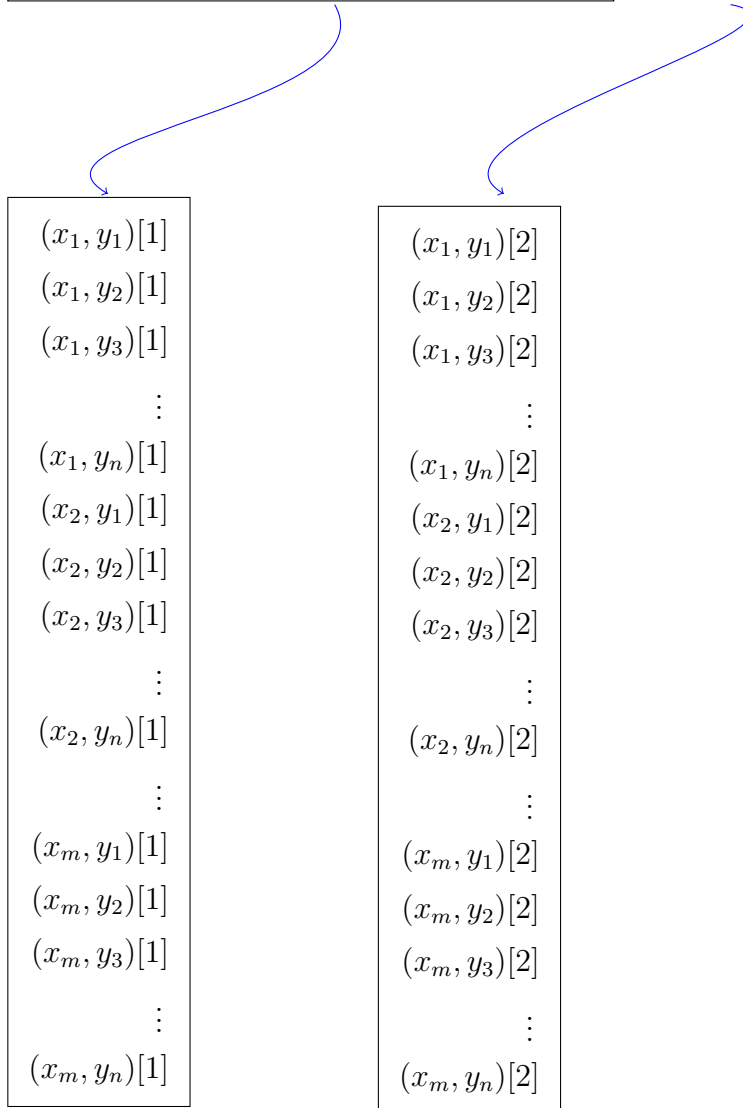
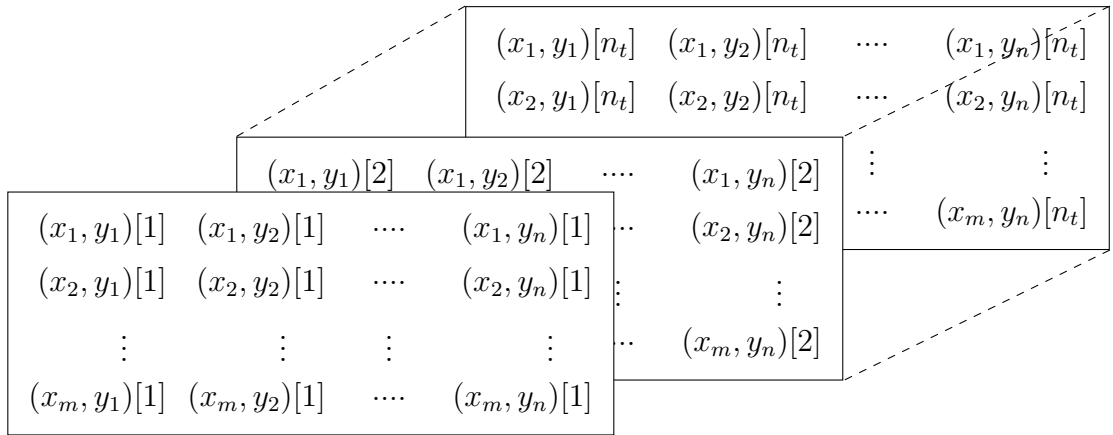
4.2.2 Data Reshaping

Empirical data is sampled spatiotemporally. In order to perform POD, we must first reshape it into a new data matrix, D , with dimensions $\mathbf{n}_s \times \mathbf{n}_t$ where:

- $n_s = n_c n_x n_y$ is the number of spatial points: n_c is the number of components or dimensionality of the real variable under study (*e.g.* $n_c = 2$ for a 2D velocity vector field), $n_x n_y$ is the size of the sampling spatial grid
- n_t is the number of snapshots over which the data has been captured

Thus, the columns of D represent the state of the system at a specific time instant, k ($k = 1, 2, 3, \dots, n_t$).

In other words, each frame in the original data matrix should be concatenated into a vector and fed as a column in the matrix D as demonstrated below:



Thus,

$$D = \begin{pmatrix} (x_1, y_1)[1] & (x_1, y_1)[2] & (x_1, y_1)[3] & \dots & (x_1, y_1)[n_t] \\ (x_1, y_2)[1] & (x_1, y_2)[2] & (x_1, y_2)[3] & \dots & (x_1, y_2)[n_t] \\ (x_1, y_3)[1] & (x_1, y_3)[2] & (x_1, y_3)[3] & \dots & (x_1, y_3)[n_t] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (x_1, y_n)[1] & (x_1, y_n)[2] & (x_1, y_n)[3] & \dots & (x_1, y_n)[n_t] \\ (x_2, y_1)[1] & (x_2, y_1)[2] & (x_2, y_1)[3] & \dots & (x_2, y_1)[n_t] \\ (x_2, y_2)[1] & (x_2, y_2)[2] & (x_2, y_2)[3] & \dots & (x_2, y_2)[n_t] \\ (x_2, y_3)[1] & (x_2, y_3)[2] & (x_2, y_3)[3] & \dots & (x_2, y_3)[n_t] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (x_2, y_n)[1] & (x_2, y_n)[2] & (x_2, y_n)[3] & \dots & (x_2, y_n)[n_t] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (x_m, y_1)[1] & (x_m, y_1)[2] & (x_m, y_1)[3] & \dots & (x_m, y_1)[n_t] \\ (x_m, y_2)[1] & (x_m, y_2)[2] & (x_m, y_2)[3] & \dots & (x_m, y_2)[n_t] \\ (x_m, y_3)[1] & (x_m, y_3)[2] & (x_m, y_3)[3] & \dots & (x_m, y_3)[n_t] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (x_m, y_n)[1] & (x_m, y_n)[2] & (x_m, y_n)[3] & \dots & (x_m, y_n)[n_t] \end{pmatrix} \quad (4.9)$$

4.2.3 Mathematical Framework [24]

As mentioned in section 4.2, the goal is to decompose D into a sum of rank-1 contributions, the modes. Each mode possesses:

- An amplitude σ . The amplitude gives information about how important the mode is (the higher the amplitude the more significant the mode is, and vice versa)
- A spatial structure ϕ which provides a basis for the space domain
- A temporal structure ψ which provides a basis for the time domain

Thus:

$$D[i, k] = \sum_1^{rank(D)} \sigma_r \phi_r[i] \psi_r[k] \quad (4.10)$$

If the summation is truncated at $r < rank(D)$, the decomposition yields an approximation, \tilde{D} , of the original data matrix, with rank r .

Both ϕ and ψ element are orthogonal and have unitary L_2 norms so that the energy contribution comes solely from σ :

$$\|\phi\|_2^2 = \frac{1}{n_s} \sum_{i=1}^{n_s} \phi_r[i] \bar{\phi}_r[i] = \frac{1}{n_s} \langle \phi_r, \phi_r \rangle = \frac{1}{n_s} \phi_r^\dagger \phi_r = 1 \quad (4.11)$$

$$\|\psi\|_2^2 = \frac{1}{n_t} \sum_{k=1}^{n_t} \psi_r[k] \bar{\psi}_r[k] = \frac{1}{n_t} \langle \psi_r, \psi_r \rangle = \frac{1}{n_t} \psi_r^\dagger \psi_r = 1 \quad (4.12)$$

$$\langle \phi_i, \phi_j \rangle = 0 \quad (4.13)$$

$$\langle \psi_i, \psi_j \rangle = 0 \quad (4.14)$$

Equations 4.11 and 4.12 imply that the amplitudes σ_r are normalized by $\sqrt{n_s n_t}$. Consequently, the energy of each mode is grid-independent.

We can arrange the spatiotemporal basis terms into vectors: $\boldsymbol{\phi} = [\phi_1, \phi_2, \dots, \phi_r]$ and $\boldsymbol{\psi} = [\psi_1, \psi_2, \dots, \psi_r]$, This way the expansion 4.10 can be written as a product of matrices:

$$D = \sum_1^{\text{rank}(D)} \sigma_r \phi_r \psi_r^T = \Phi \Sigma \Psi^T \quad (4.15)$$

where: $D \in \mathbb{R}^{n_s \times n_t}$, $\Phi \in \mathbb{R}^{n_s \times r}$, $\Sigma \in \mathbb{R}^{r \times r}$, and $\Psi \in \mathbb{R}^{r \times n_t}$. Σ is diagonal while Ψ and Φ have normalized columns.

4.2.4 Choosing the Optimal Basis

Below we explain how to construct the optimal basis:

From 4.15:

$$D = \Phi \Sigma \Psi^T \rightarrow D \Psi \Sigma^{-1} = \Phi \Sigma \Psi^T \Psi \Sigma^{-1} \quad (4.16)$$

Since the basis is orthonormal, 4.16 becomes:

$$\Phi = D \Psi \Sigma^{-1} \quad (4.17)$$

Substituting for Φ in 4.15 using 4.17, we get:

$$D = D \Psi \Psi^T = \begin{cases} D, & \text{if } \Psi \text{ is a complete basis} \\ \tilde{D}, & \text{if } \Psi \text{ is an incomplete basis} \end{cases} \quad (4.18)$$

Thus, for any truncation, the optimal orthonormal temporal basis is one that minimizes the mean square error between D and \tilde{D} :

$$\min(\|D - \tilde{D}\|_2^2 = \|D - D(\psi \psi^T)\|_2^2) \quad (4.19)$$

subject to the constraint:

$$\langle \psi_i, \psi_j \rangle = \delta_{ij} \quad (4.20)$$

The optimization process requires the temporal structures to be eigenvectors of the temporal correlation matrix \mathbf{K} :

$$K = D^\dagger D \quad (4.21)$$

\mathbf{K} contains the correlation of one snapshot with another. It is symmetric by construction, thus its eigenvectors are orthonormal. We can write its eigendecomposition as:

$$K = \Psi \Lambda \Psi^T = \sum_{i=1}^r \psi_i \lambda_i \psi_i^T \quad (4.22)$$

From 4.15 and 4.21, we have:

$$K = D^\dagger D = (\Phi \Sigma \Psi^T)^T (\Phi \Sigma \Psi) = \Psi \Sigma \Phi^T \Phi \Sigma \Psi^T = \Psi \Sigma^2 \Psi^T \quad (4.23)$$

By comparing 4.22 and 4.23, we deduce that $\Sigma = \sqrt{\Lambda}$ and $\Sigma \Phi^T \Phi \Sigma$ is diagonal because it is equal to Λ , the matrix of eigenvalues of \mathbf{K} . This necessitates Φ to be orthogonal.

We reached a very important result, which is only valid for the POD. When the temporal structures are orthonormal, so are the spatial structures.

Remark: We can find Ψ in terms of D, Σ, Φ , conversely to what we did in 4.17. In that case, we find that the spatial basis elements are eigenvectors of the spatial correlation matrix $C = DD^T = \Phi \Sigma^2 \Phi$. We also obtain an orthogonal temporal basis as a result.

To conclude this subsection, we will go over two common algorithms used to perform the POD:

Algorithm 1 Sirovich's Method

- 1: Given D , compute $K = D^\dagger D$
 - 2: Diagonalize K , to get Ψ and $\Sigma = \sqrt{\Lambda}$
 - 3: Calculate $\Phi = D \Psi \Sigma^{-1}$
-

Algorithm 2 Lumley's Method

- 1: Given D , compute $C = DD^T$
 - 2: Diagonalize C , to get Φ and $\Sigma = \sqrt{\Lambda}$
 - 3: Calculate $\Psi = D^T \Phi \Sigma$
-

4.3 Data Description

In this work, we use two different movies of the *T. Adhaerens*, the details of which are given in the table below:

Movie 1	Movie 2
Displays the placozoan in a local isotropic state of motion (stretching, compressing, and rotating in its place)	Displays the placozoan exhibiting directed motion, where it explores its environment more compared to the first movie
Produces a 2D velocity vector field with x and y components being \mathbf{U} and \mathbf{V} respectively	Produces a 2D velocity vector field with x and y components being \mathbf{U} and \mathbf{V} respectively
Size of the spatial grid: 130×174	Size of the spatial grid: 64×52
Number of time frames: 4580	Number of time frames: 971

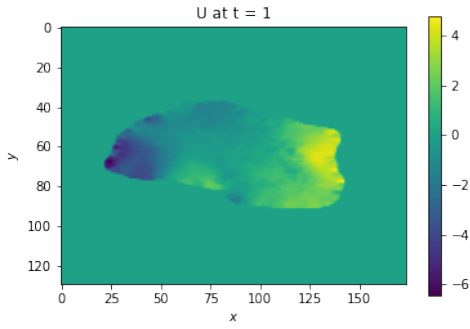
Table 4.1: Description of the two placozoan movies used in this work

Note 1: In everything that follows, we will refer to movie 1 as “stationary” and movie 2 as “moving”.

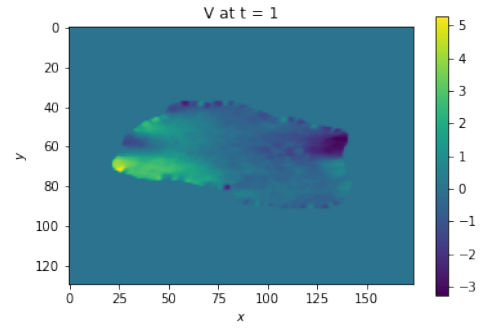
Note 2: The velocity vector fields were generated using particle image velocimetry (PIV). Further information about that, as well as image and data preprocessing, can be found in appendix A.

4.4 Data Visualization

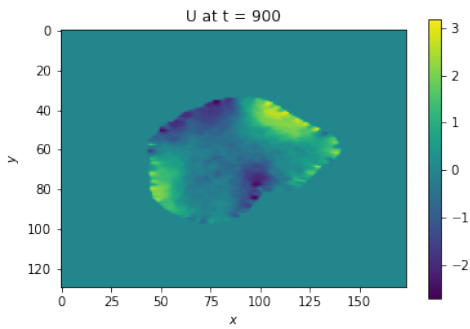
Below we show sample plots of the velocity fields of each movie:



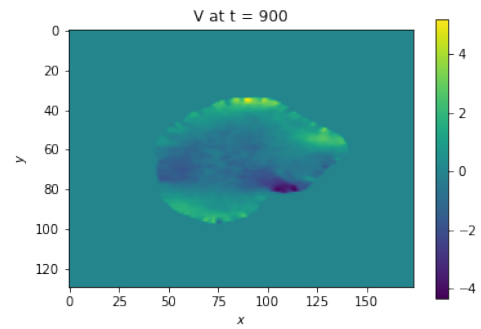
(a) U stationary at time frame 1



(b) V stationary at time frame 1

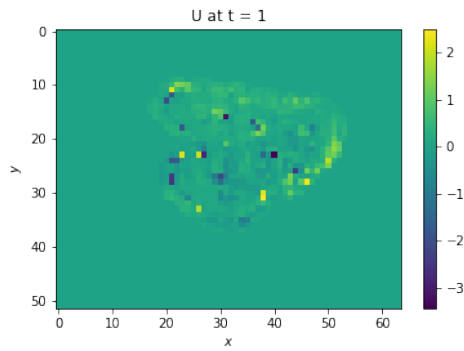


(c) U stationary at time frame 900

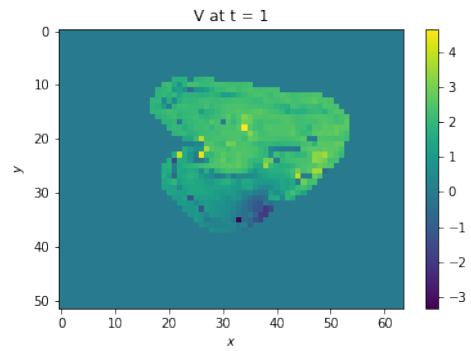


(d) V stationary at time frame 900

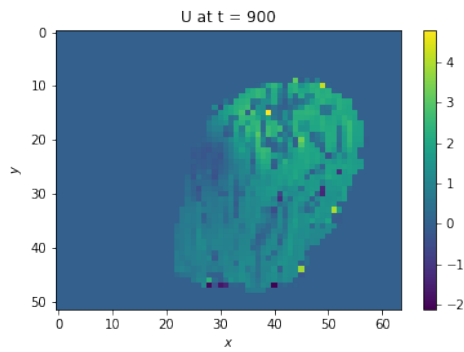
Figure 4.1: Velocity field components (stationary) at different snapshots



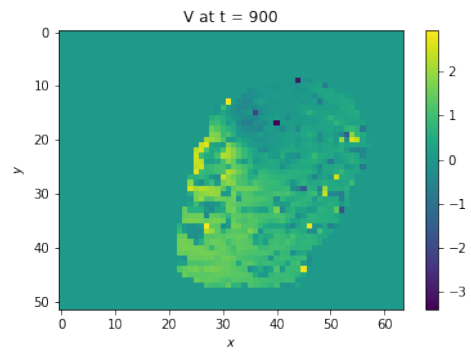
(a) U moving at time frame 1



(b) V moving at time frame 1



(c) U moving at time frame 900



(d) V moving at time frame 900

Figure 4.2: Velocity field components (moving) at different snapshots

CHAPTER 5

RESULTS

** Note: In all the results that follow, the units for distance are pixels (px) and those for time are (/frame)

5.1 Generated Partial Differential Equations

5.1.1 *PDE for moving video - Linear Regression - Toner-Tu Model*

In order to generate the PDE, we had to sample data from different regions inside the animal at different time instants. Below we briefly list the steps performed to achieve that:

1. Initialize the boundaries or starting points for x , y , and t
2. Fix a sampling window of size $n_x \times n_y \times n_t = 25 \times 25 \times 25$
3. Randomly select points in space and time within this window

In our work, we sampled 900 random data points, 80% of which were used for training, and 20% for testing. We obtained a PDE for both the x and y components of the velocity field, U and V respectively. The obtained results are listed in tables [5.1](#) - [5.4](#).

Terms	Coeff.	Error
$cost$	0.0	1.4
u_x	- 212	14
u_y	- 145	15
uu_x	68	12
vu_y	- 157	27
uv_y	- 69	5
u_x^2	142	12
v_x^2	61.8	1.3
u	- 3.4	0.6
u^3	1.9	0.7
v^2u	- 1.13	0.13
ρ_x	28.7	0.6
ρ_x^4	- 376	3
u_{xx}	614	31
v_{xy}	68	5
u_{yy}	1692	43
u^2u_{xx}	- 2426	127
v^2u_{yy}	- 3413	136
uvu_{xy}	291	12
uvu_{yx}	153	10

Table 5.1: Terms and coefficients of the PDE obtained for U moving using a linear regression fit, Toner-Tu

$$\overline{\overline{R^2 \quad 73\%}}$$

Table 5.2: Coefficient of performance for U moving, linear regression, Toner-Tu

By examining Table 5.1, we observe that both linear and nonlinear advective terms exist in the form of $u_x(-212+68u)$ and $u_y(-145-157v)$. However, when normalized, the values of u and v that are greater than 0.8 are few, constituting only 0.02% of the entire dataset. Therefore, we can conclude that the dominant term is the linear advective term in the negative y direction. In terms of diffusion, we notice that coefficients corresponding to nonlinear diffusion are small compared to those of typical diffusion. Thus, up to leading order we have constant diffusion coefficients.

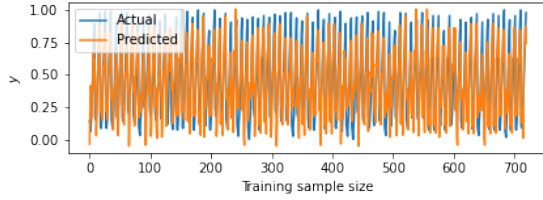
Terms	Coeff.	Error
cst	0	3
v_x	220	36
v_y	- 264	37
uv_x	- 317	30
vv_y	1837	67
vu_x	66	13
u_y^2	17	5
v_y^2	-691	3
v	- 28.3	1.6
vu^2	4	2
v^3	36.6	0.3
ρ_y	- 11.4	1.4
ρ_y^4	89	9
u_{yx}	17	9
v_{xx}	-298	12
v_{yy}	-2467	109
u^2v_{xx}	1858	321
v^2v_{yy}	6614	344
uvv_{xy}	-759	30
uvv_{yx}	-1255	26

Table 5.3: Terms and coefficients of the PDE obtained for V moving using a linear regression fit (Toner-Tu)

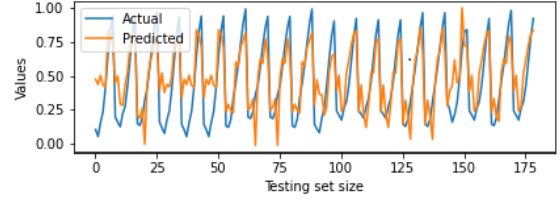
$$\overline{\overline{R^2 \quad 76\%}}$$

Table 5.4: Coefficient of performance for V moving, linear regression, Toner-Tu

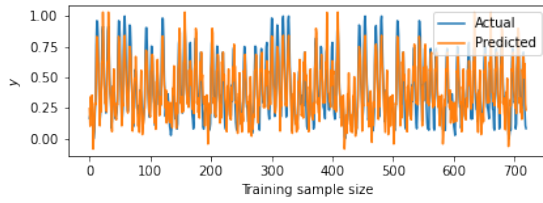
Following the same line of thought above, we notice that, for V , nonlinear convective and diffusive contributions are significant contrary to the case of U



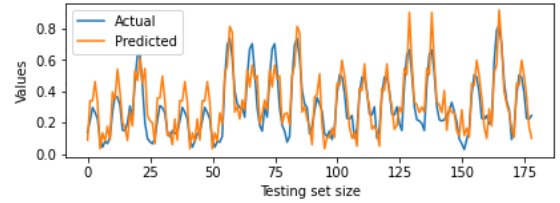
(a) U train



(b) U test



(c) V train



(d) V test

Figure 5.1: True and predicted values for both training and testing sets, moving, Toner-Tu

5.1.2 PDE for moving video - Linear Regression - Tensor Theory

The same steps mentioned at the beginning of section 5.1.1 were used to generate the below PDEs:

Terms	Coeff.	Error
cst	0.00	0.12
u_x	- 5e6	4e4
u_y	- 5e6	1e4
ρ_x	1e6	1e5
ρ_x^4	- 8e6	6e4
u_{xx}	-3e6	9e4
u_{yy}	1e7	7e4
v_{xy}	4e6	1e5
$Q11_x$	2.3	0.8
$Q12_y$	3.4	0.09

Table 5.5: Terms and coefficients of the PDE obtained for U moving using a linear regression fit, Tensor theory

$$\underline{\underline{R^2 \quad 58\%}}$$

Table 5.6: Coefficient of performance for U moving, linear regression, Tensor theory

Table 5.5 suggests that diffusion happens mainly along the y direction. The active terms corresponding to gradients in Q are insignificant with respect to other terms.

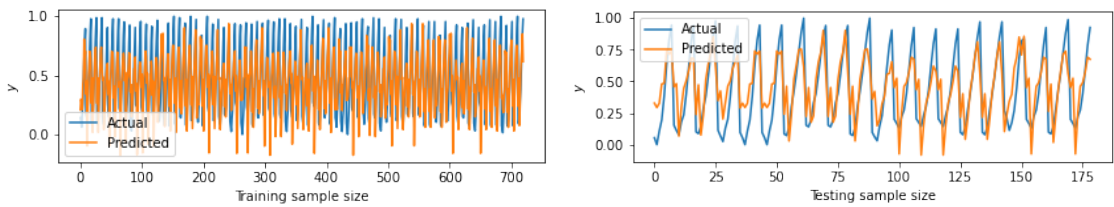
Terms	Coeff.	Error
cst	0.00	0.07
v_x	-6e5	8e4
v_y	5e5	1e4
ρ_y	-3e5	2e5
ρ_y^4	-2e5	2e4
v_{xx}	3e6	1e5
v_{yy}	4e6	2e5
u_{xy}	-1e7	1e5
$Q21_x$	0.5	0.20
$Q22_y$	0.35	0.10

Table 5.7: Terms and coefficients of the PDE obtained for V moving using a linear regression fit, Tensor theory

$$\overline{R^2} \quad 60\%$$

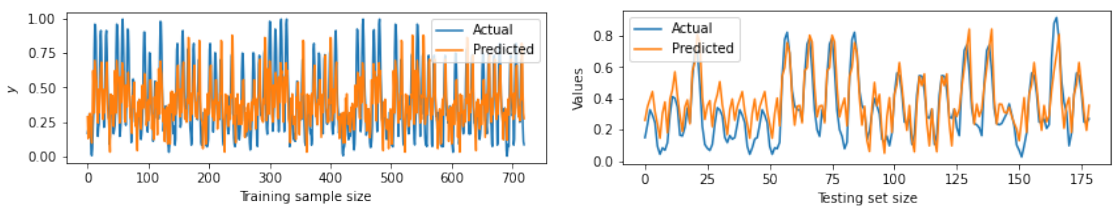
Table 5.8: Coefficient of performance for V moving, linear regression, Tensor theory

Again, convection is mainly in the y direction. Moreover, coefficients corresponding to terms in Q continue to be insignificant



(a) U train

(b) U test



(c) V train

(d) V test

Figure 5.2: True and predicted values for both training and testing sets, moving, tensor

→ **Reflection:**

1. Ideally, for both models, there should be a common fit generating one set of coefficients for both u and v as we don't expect the coefficients to be

coordinate-dependent. When we did that, we obtained very low coefficients of performance: $R^2 < 40\%$ implying that there are inhomogeneities in space. This could be caused by various chemical gradients found in the vicinity of the animal. We can't be entirely sure as we don't know the exact experimental conditions in which the movies were taken.

2. The criterion that we are using to assess the performance of the model is R^2 . For the moving video, $R_{Toner-Tu}^2 > R_{Tensor}^2$ suggests that the Toner-Tu model better describes the given dataset.
3. As mentioned already, the active stress terms in both U and V were negligible compared to other terms. This, along with the fact that the Toner-Tu model performed better, made it obvious that recovering an equation for Q , though feasible, is not going to add any value to our understanding.

5.1.3 *PDE for moving video - Weak SINDy - Toner-Tu Model*

Below we show terms of a PDE generated using the SINDy algorithm with $R^2 = 90\%$:

Terms	Coeff.
u	1.14882E+14
u^2	9.38485E+12
ρ	9.29359E+11
v	-3397634255
v^2	3.47087E+13
v^3	4.09487E+12
u_{xx}	-2.50305E+13
uv_y	-3.17309E+13
u^3u_y	-0.503576648
uvv_y	-0.694006777
v^2v_y	3.43405E+13
v^2v_y	1.29919E+13
v^3v_y	3.43405E+13
v^3u_y	-6.17709E+12
vvv_y	-4.9178E+12
uu_{yy}	5.34819E+12
uv_{yy}	5.74618E+12
uu_{yy}	-0.87279805
u^2u_{yy}	-1.60949E+12
u^3u_{yy}	1.24588E+13
u^3v_{yy}	-3.41558E+12
ρv_{yy}	4.37495E+12
ρu_{yy}	0.965463543
ρv_{yy}	5.45466E+12
uvu_{yy}	0.134410168
vu_{yy}	0.388357867
vv_{yy}	-2.64803E+13
v^2v_{yy}	1.4412E+13
v^3v_{yy}	5.07026E+12
uu_x	3.35603E+13
uv_x	-7.62254E+11
u^2u_x	9.78473E+12

→ **Reflection:**

In order to find errors on the coefficients, we must perform bootstrapping. Due to time constraints and the computational cost associated with bootstrapping, we only present here a sample PDE generated without errors on coefficients.

5.1.4 PDE for moving video - Weak SINDy - Tensor Theory

The same steps mentioned in section 5.1.1 were used to generate the below PDEs, except that the window size was taken to be $n_x \times n_y \times n_t = 40 \times 40 \times 40$.

u	1.72291E+14
ρ	4.7273E+13
Q_{12}	-4.95237E+14
v	0.1484375
Q_{22}	-0.426788087
u_{yy}	-0.0625
u_x	1.77356E+13
uu_y	-7.31085E+11
uv_y	-3.00813E+13
ρu_2	-1.56341E+13
$Q_{11}v_y$	-0.1260018
vv_2	-2.3762E+14
$Q_{21}u_y$	1.0102E+12
$Q_{22}u_y$	-0.14453125
uv_{yy}	-0.141147973
ρu_{yy}	-1.8871E+13
ρv_{yy}	0.215590205
$Q_{12}u_{yy}$	1.06989E+14
$Q_{12}v_{yy}$	2.95757E+12
vv_{yy}	-3.78603E+14
$Q_{21}v_{yy}$	-1.45725E+13
$Q_{22}v_{yy}$	-1.32911E+15
$Q_{22}u_{yy}$	-0.346923828
$Q_{22}v_{yy}$	3.32469E+14
uu_x	0.16015625
ρv_x	7.15646E+12
ρu_x	-0.86328125
ρv_x	-1.12286E+13
$Q_{11}u_x$	0.133441932
$Q_{11}u_x$	3.46481E+13
$Q_{12}u_x$	9.48149E+14
$Q_{12}v_x$	1.602126097
vv_x	-0.227721913

5.1.5 PDE for stationary video - Linear regression - Toner-Tu

Terms	Coeff.	Error
cst	0	7
u_x	221	32
u_y	- 264	12
uu_x	- 317	51
vu_y	1837	25
uv_y	65.5	2.2
u_x^2	17	25
v_x^2	- 691.1	1.9
u	- 28.3	0.7
u^3	4.1	0.8
v^2u	36.65	0.09
ρ_x	-11.5	1.3
ρ_x^4	88.9	2.3
u_{xx}	17	1
v_{xy}	-298	11
u_{yy}	- 2467	64
u^2u_{xx}	1859	653
v^2u_{yy}	6614	286
uvu_{xy}	-759	44
uvu_{yx}	-1254.70	23

Table 5.9: Terms and coefficients of the PDE obtained for U stationary using a linear regression fit, Toner-Tu

$$\overline{\overline{R^2 \quad 89\%}}$$

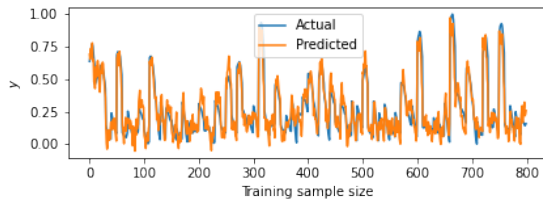
Table 5.10: Coefficient of performance for U stationary, linear regression, Toner-Tu

Terms	Coeff.	Error
cst	0	5
v_x	-269	15
v_y	27	15
uv_x	661	29
vv_y	-89	45
vu_x	110	3
u_y^2	-98.3	0.8
v_y^2	-16	6
v	-18	0.5
vu^2	3.77	0.06
v^3	24.6	0.7
ρ_y	41.7	0.4
ρ_y^4	- 28.0	0.5
u_{yx}	- 67	4
v_{xx}	- 617	76
v_{yy}	- 4074	104
u^2v_{xx}	3713	252
v^2v_{yy}	14216	460
uvv_{xy}	504	29
uvv_{yx}	202	13

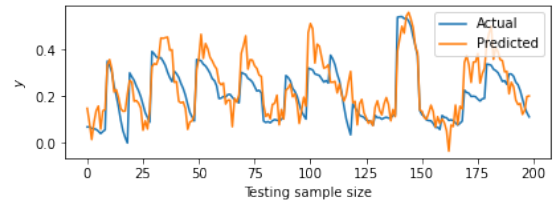
Table 5.11: Terms and coefficients V stationary, linear regression, Toner Tu

$$\overline{\overline{R^2}} \quad 90\%$$

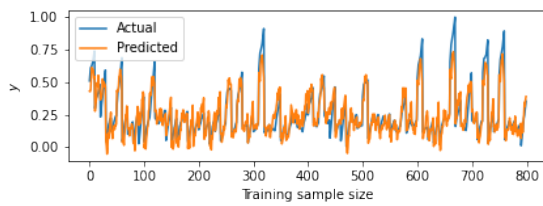
Table 5.12: Coefficient of performance for V stationary, linear regression, (Toner-Tu)



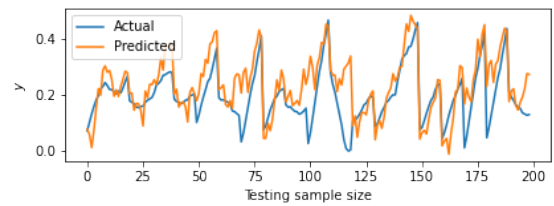
(a) U train



(b) U test



(c) V train



(d) V test

Figure 5.3: True and predicted values for both training and testing sets, stationary, Toner-Tu

5.1.6 PDE for Stationary video - Linear regression - Tensor

Terms	Coeff.	Error
cst	- 3 e+13	1e5
u_x	- 6 e+13	3e4
u_y	6 e+13	4e4
ρ_x	-2e7	1e4
ρ_x^4	3 e+7	5e4
u_{xx}	-1e11	7e4
u_{yy}	- 1e11	4e4
v_{xy}	2e8	4e4
Q_{11x}	1e6	0.006
Q_{12y}	-1e6	0.003

Table 5.13: Terms and Coefficients for U stationary using a linear regression fit, Tesnor

$$\underline{\underline{R^2 \quad 83\%}}$$

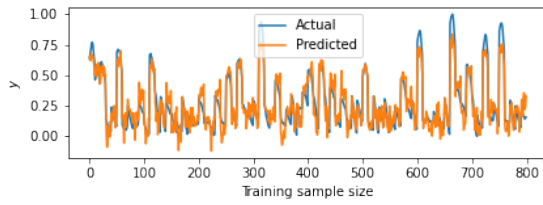
Table 5.14: Coefficient of performance for U stationary, linear regression, tensor

Terms	Coeff.	Error
cst	1.2e11	1e5
v_y	- 1 e+12	3e5
v_x	4e+07	8e4
ρ_y	7e6	1e4
ρ_y^4	-5e6	2e4
u_{yx}	3e6	2e4
v_{xx}	- 2e8	1e4
v_{yy}	4e7	1e4
Q_{21x}	0.07	0.005
Q_{22y}	2.33e4	0.003

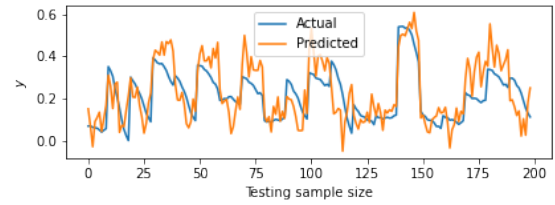
Table 5.15: Terms and Coefficients for V stationary, linear regression, tesnor

$$\underline{\underline{R^2 \quad 77\%}}$$

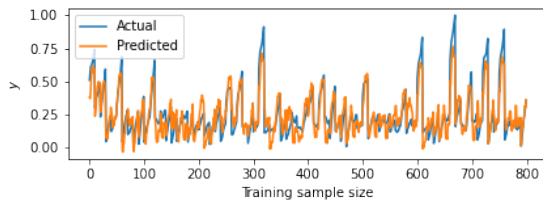
Table 5.16: Coefficient of performance for V stationary, linear regression, tensor



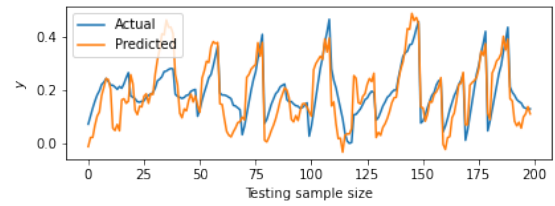
(a) U train



(b) U test



(c) V train



(d) V test

Figure 5.4: True and predicted values for both training and testing sets, stationary, tensor

We notice that even for the stationary movie, the Toner-Tu model performs better than the tensor theory.

5.1.7 PDE for stationary video - Weak SINDy - Toner Tu

u_{yy}	0.199468717
u_{xy}	-8451189074
u_{xx}	-0.457284781
uu_y	-0.383895159
u^2v_y	1.135149773
uvu_y	-0.367383737
vu_y	-0.549521852
u^2u_{yy}	0.076384952
u^3u_{yy}	6742230878
$rhov_{yy}$	-0.330871582
uvu_{yy}	0.217867108
vv_{yy}	-0.704203502
vuv_{yy}	0.243775538
v^2v_{yy}	-58919171375
vvv_{yy}	-22.29803519

5.1.8 PDE for stationary video - Weak SINDy - Tensor

Terms	Coeff.
u	-0.421203613
ρ	3.59897268
Q_{11}	1.256224995
Q_{12}	0.296998468
v	-0.866455078
Q_{21}	-0.303100586
Q_{22}	-0.829494671
u_y	3.101505741
v_y	-0.181152344
u_{yy}	-0.685302734
v_{yy}	-0.323486328
u_x	0.020385742
v_x	-1.075317383
u_{xy}	0.245822712
v_{xy}	2.855683029
u_{xx}	-1.393066406
v_{xx}	10.07670999
uu_{yy}	6.308994995
uv_{yy}	-4.729614258
ρu_y	1.290771484
ρv_y	-3.90462687
$Q_{11}u_y$	0.636214409
$Q_{11}v_y$	4.440917969
$Q_{12}u_y$	-7.528971801
$Q_{12}v_y$	5.018223511
vu_y	-4.72961
vv_y	1.290771
$Q_{21}u_y$	4.20084188
$Q_{21}v_y$	0.636214409
$Q_{22}u_y$	-1.291460624
$Q_{22}v_y$	2.635986328
uu_{yy}	0.467224121
uv_{yy}	-0.128295898
ρu_{yy}	0.370605469
ρv_{yy}	0.08246424

5.2 Modal Analysis

Below, we present the results for the POD decomposition of both movies.

5.2.1 Modal Analysis - Stationary movie

We will start by displaying the energy magnitudes (Fig. 5.9). We notice that the relative energy amplitude drops quickly from 1 to ~ 0.2 after two modes only, signaling that the first two modes could be the most significant (Fig. 5.10)

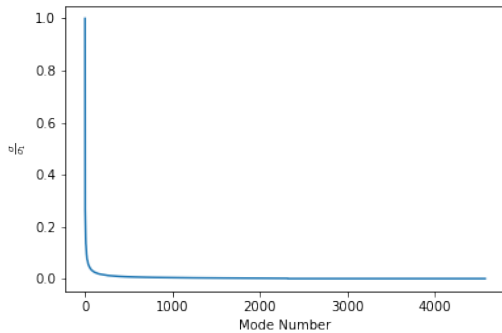


Figure 5.5: Relative energy amplitudes as a function of mode number, stationary movie

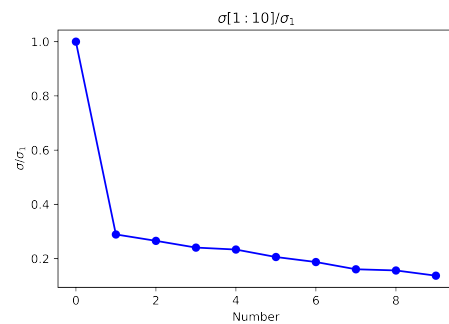
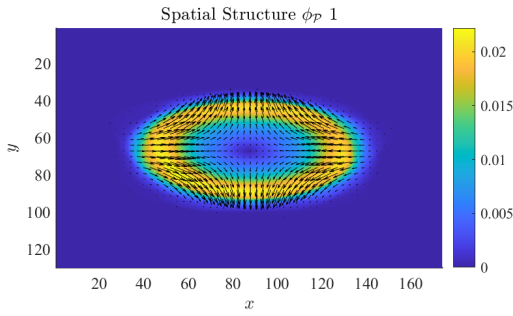
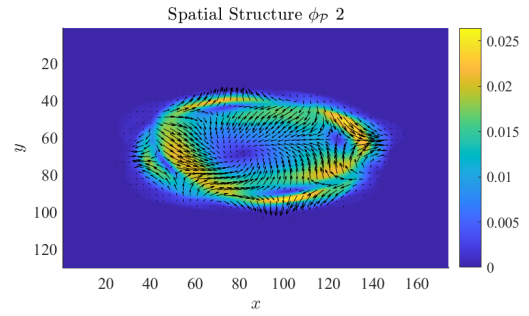


Figure 5.6: Relative energy amplitudes for the first 10 modes, stationary movie

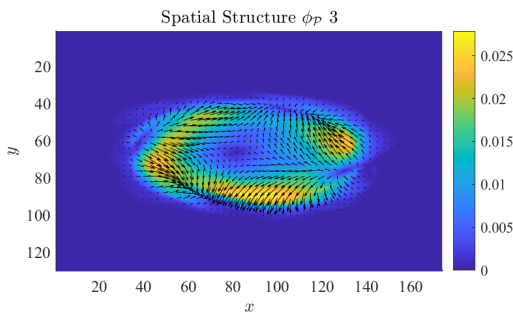
Below, we show the spatial structures of the first 4 modes:



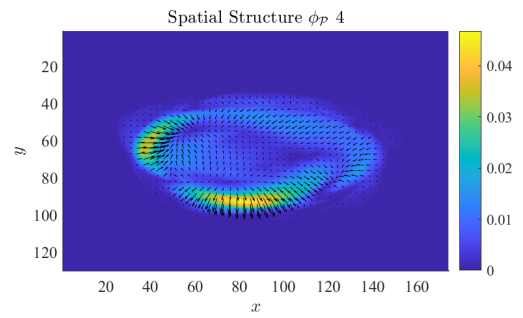
(a) Mode 1



(b) Mode 2



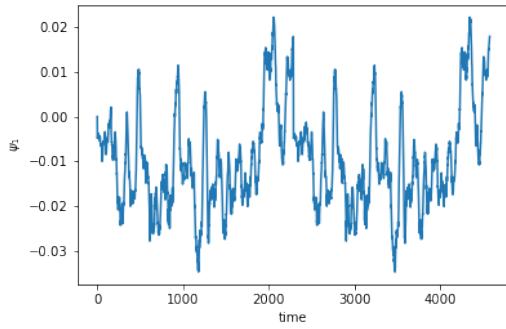
(c) Mode 3



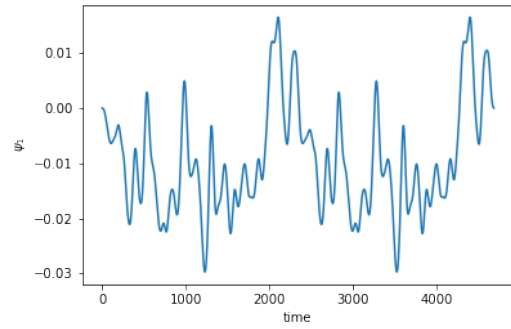
(d) Mode 4

Figure 5.7: Spatial structures of the first 4 modes, stationary movie

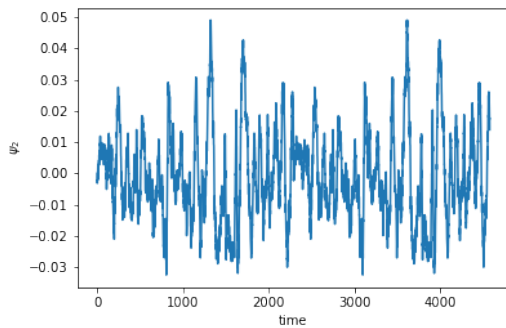
Finally, the obtained temporal structures for the first 4 modes before and after smoothing (see appendix A) are shown below:



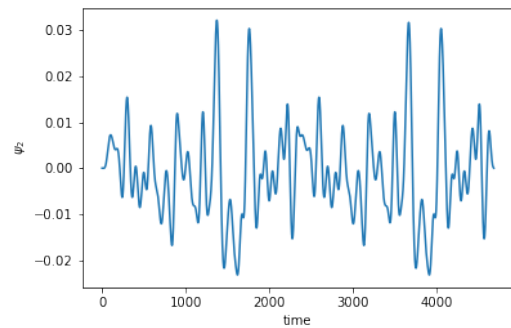
(a) Mode 1, before smoothing



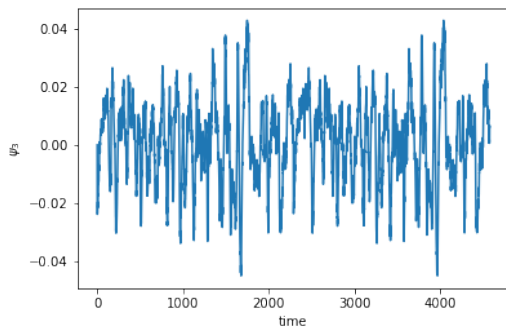
(b) Mode 1, after smoothing



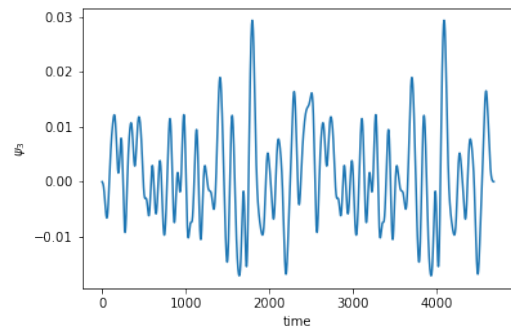
(c) Mode 2, before smoothing



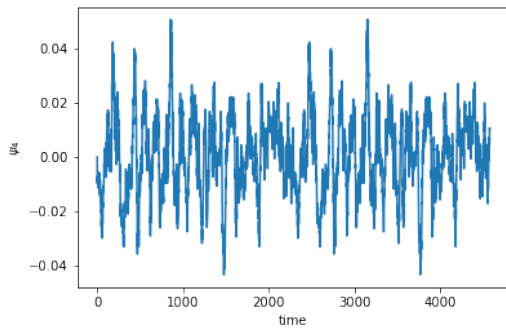
(d) Mode 2, after smoothing



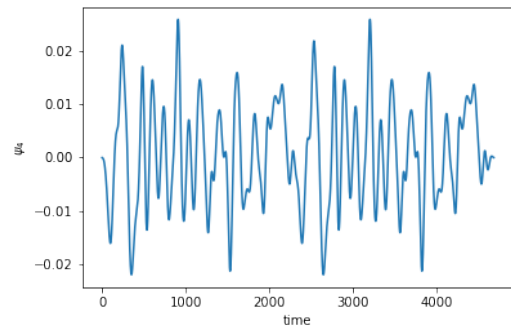
(e) Mode 3, before smoothing



(f) Mode 3, after smoothing



(g) Mode 4, before smoothing



(h) Mode 4, after smoothing

Figure 5.8: Temporal structures of the first 4 modes, stationary movie

5.2.2 Modal Analysis - Moving movie

As we did for the stationary movie, we will first display the relative energy modes and then the spatial and temporal structures. The results for the energy amplitudes were as follows:

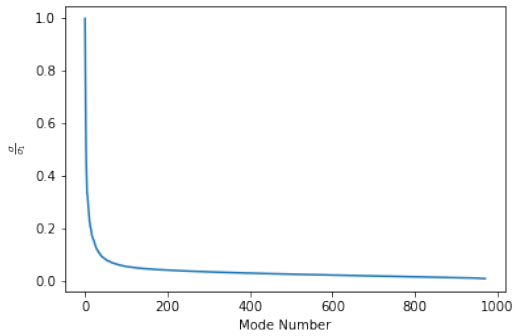


Figure 5.9: Relative energy amplitudes as a function of mode number, moving video

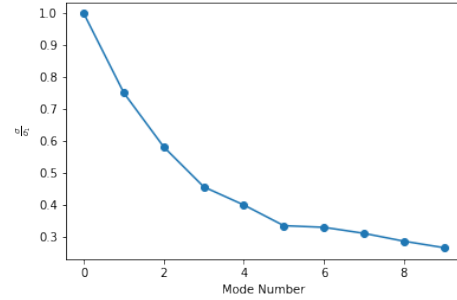
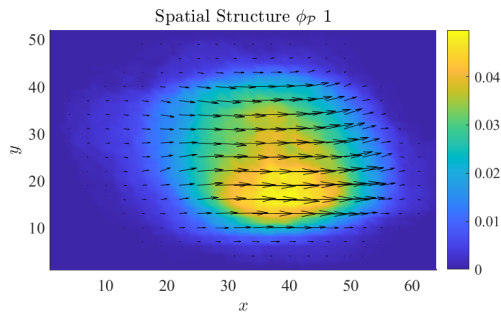


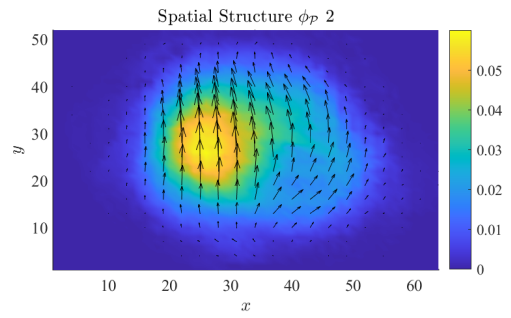
Figure 5.10: Relative energy amplitudes for the first 10 modes, moving video

We notice that the σ values do not drop as quickly as they do for the stationary case. This is not surprising as we expect the animal in this state to display richer dynamics (additional modes corresponding to translational motion)

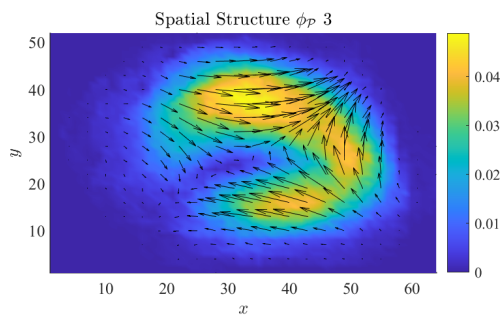
We will consider the first 6 modes:



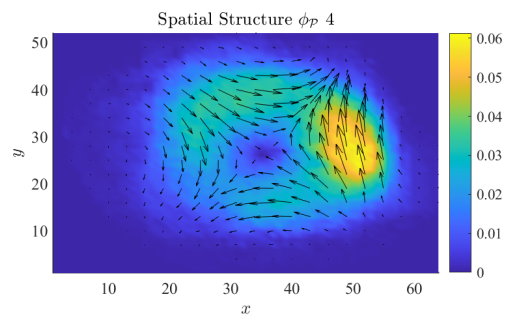
(a) Mode 1



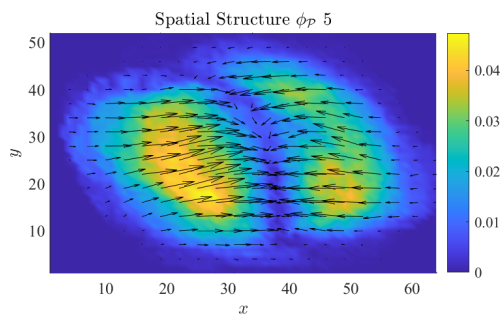
(b) Mode 2



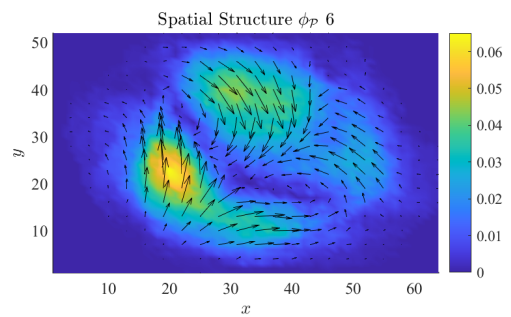
(c) Mode 3



(d) Mode 4

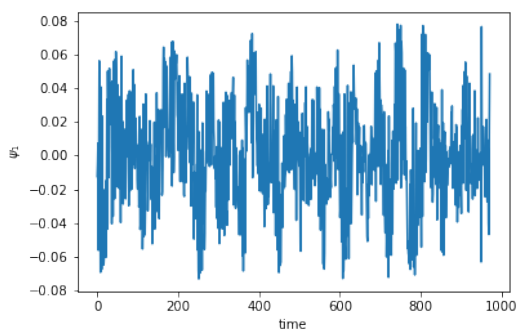


(e) Mode 5

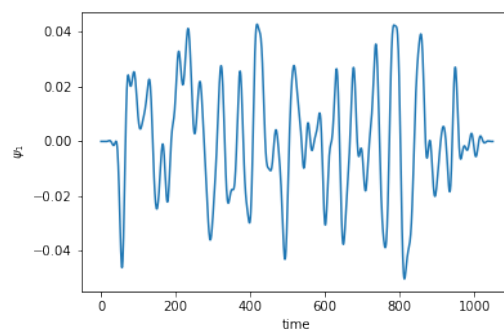


(f) Mode 6

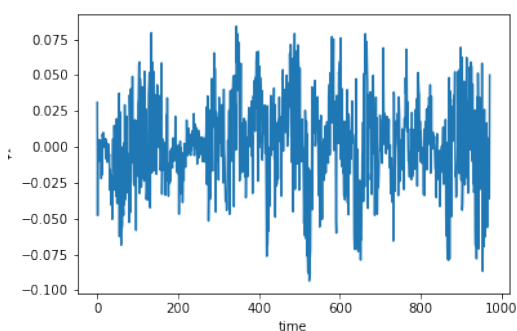
Figure 5.11: Spatial structures of the first 6 modes, moving video



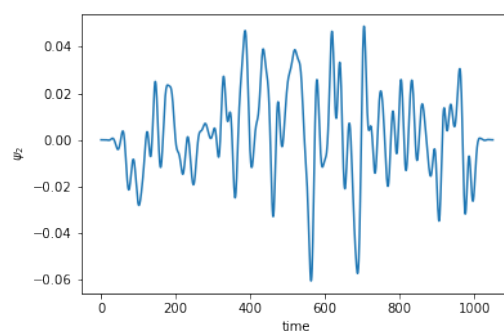
(a) Mode 1, before smoothing



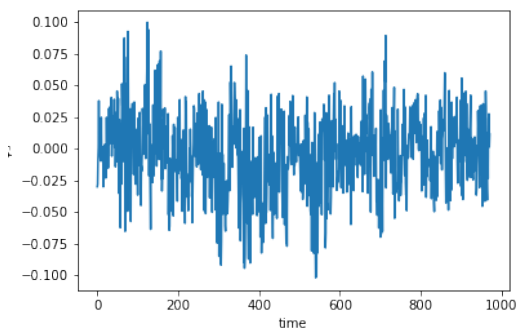
(b) Mode 1, after smoothing



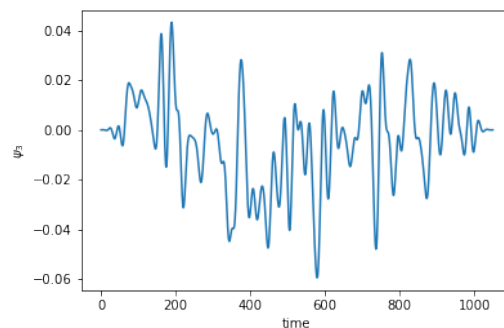
(c) Mode 2, before smoothing



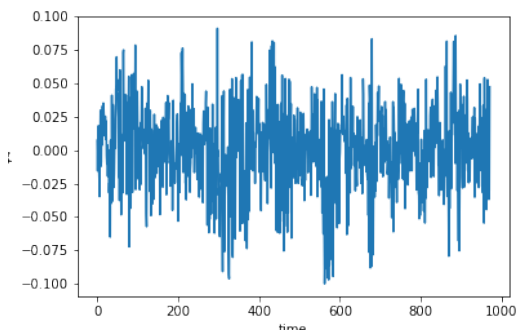
(d) Mode 2, after smoothing



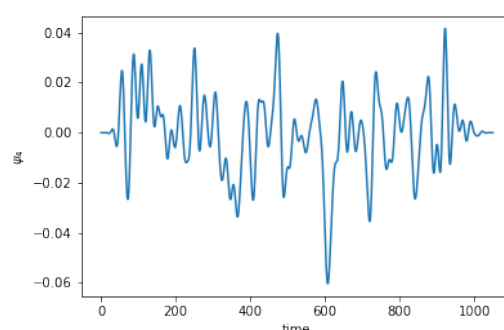
(e) Mode 3, before smoothing



(f) Mode 3, after smoothing



(g) Mode 4, before smoothing



(h) Mode 4, after smoothing

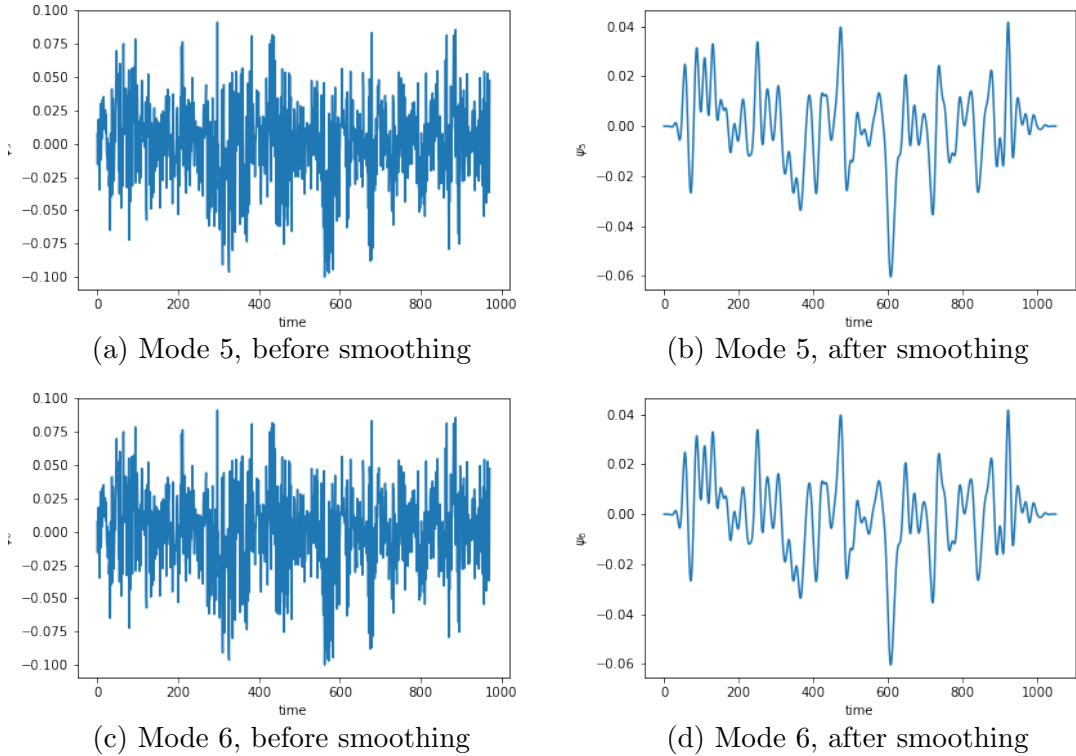


Figure 5.13: Temporal structures of the first 6 modes, moving video

5.3 Limitations

One of the main drawbacks of our study is our lack of knowledge of the experimental conditions under which the movies of *T. Adhaerens* were captured. As our analysis heavily relies on data, any noise picked up during data acquisition could have a significant impact on our findings. Potential sources of bias or errors in our results could stem from factors such as the presence of unwanted chemical gradients in the animal’s submerging region, or poor lighting conditions, which could substantially affect the accuracy of the PIV results. Moreover, with the lack of information about calibration in both movies we can’t exactly rescale the values for the coefficients obtained in the above PDEs.

Another significant limitation in our approach is that the differential equations we generated do not account for stochasticity, implying that we are not entirely describing the true dynamics. However, despite the absence of stochasticity, the models seem to be performing reasonably well.

Finally, there is the build-up of numerical errors that comes with different approximations, numerical estimation of the derivatives and density fields (see A.2).

CHAPTER 6

CONCLUSION AND FUTURE WORK

This study involved the generation of partial differential equations governing the evolution of the *Trichoplax Adhaerens*' behavior in time. We explored two different models for two different states of the animal. Our results still need refinements and more solid implementations of physical constraints, but they lay a working procedural scheme for the data-driven analysis of the *T. Adhaerens*, which, to our knowledge, is missing in the literature. This study opens the door for several future works, like finding the lagrangian coherent structures and detecting topological defects which signal cellular activity like proliferation or death. Another interesting exploration is describing the out-of-plane dynamics where the problem is treated as three-dimensional with the third dimension representing the animal's thickness.

Generally, modeling animal behavior comes with significant challenges especially since animals do not fall within any of the traditional fields of physics. While we can expect the type of behavior they may manifest, it is not always deterministic and straightforward to dictate with the well-established laws of physics[3].

Modeling systems, particularly cells, can occur at several levels, it is important to exercise caution in choosing what one wants to model and whether the data at hand corresponds to the thing they are modeling. On the computational aspect of things, as we have demonstrated in this work, a range of techniques are employed to smooth out data, remove noise, and determine the best constraint values for models. Each of these methods requires the choice of one or more parameters, one has to be careful in tuning these values to ensure the most accurate and realistic description of the system. Consequently, the interplay of these different parameters imposes additional difficulties. Nevertheless, scientists continue to construct and improve on models that help us build a more comprehensive understanding of living systems.

APPENDIX A

NUMERICAL DETAILS

A.1 Image preprocessing

Different steps were done to clean the datasets used to generate the equations:

1. The movie showing the animal moving had very poor lighting conditions. We had to apply image enhancement techniques like removing shadows and subtracting background
2. The datasets contained glitches due to camera motion during image acquisition. We had to delete these frames (Fig. A.1)
3. Image thresholding was used to obtain the densities from the different time frames (see section A.2)

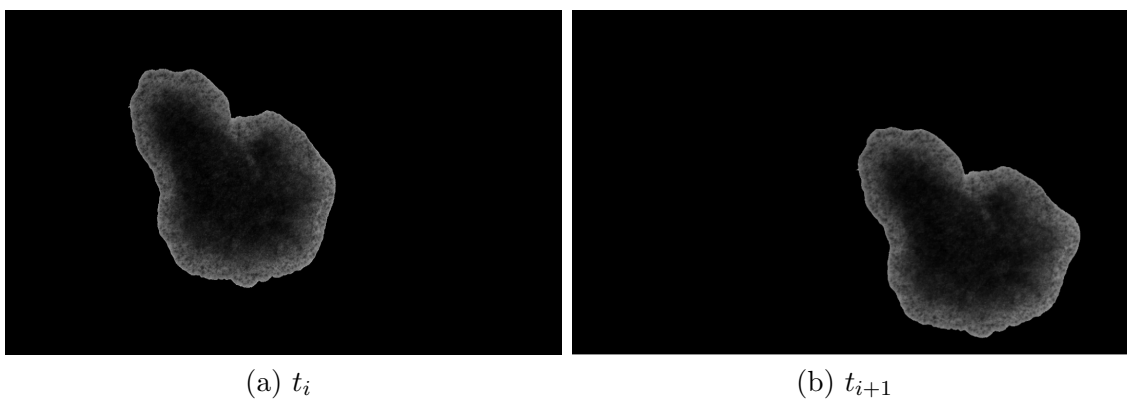


Figure A.1: Glitch in image acquisition

A.2 Obtaining the density field

When we tried solving the continuity equation to find the density field, the solution kept blowing up. As a substitute, we broke down both movies into their constituent frames and applied kernel density estimation (KDE) to each frame. KDE is a

non-parametric way of estimating the probability distribution which generated a specific dataset. It places a kernel function at each data point and then adds up the contributions from all the kernels to estimate the distribution. The free parameters of the kernel are:

1. kernel which specifies the shape of the distribution
2. kernel bandwidth (h) which specifies the size of the kernel

In addition to estimating the density, the KDE technique aims to produce a smooth boundary for the animal to prevent discontinuities at the interfaces between the area within and outside of the animal. In order to apply KDE, we had to first threshold each frame. For our purposes, we used a Gaussian kernel. The bandwidth was calculated using Scott's method: $h = n^{-1/(d+4)}$, where n is the number of data points and d is the number of spatial dimensions. We set the number of bins to 300. Below we show a sample of the obtained density fields:

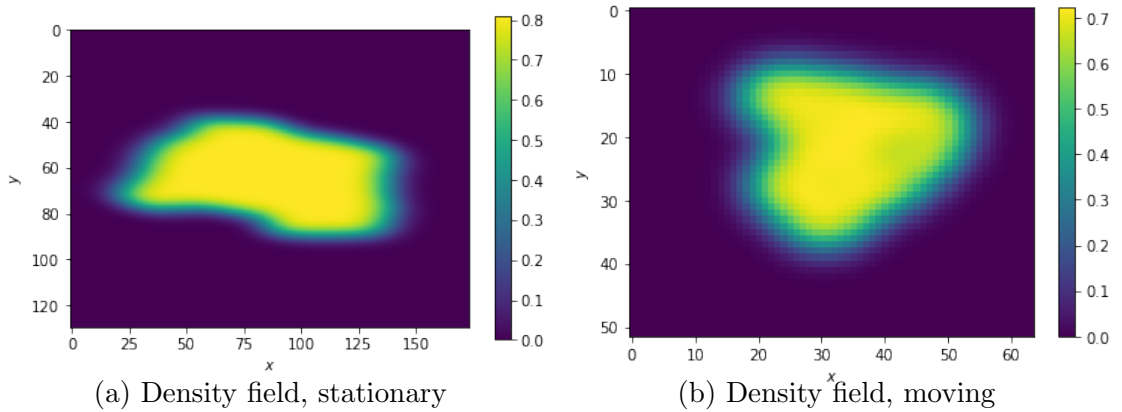


Figure A.2: Density fields for the first time frame

A.3 Particle Image Velocimetry (PIV)

Particle image velocimetry is a widely used technique for obtaining velocity fields from image sequences. Below, we briefly mention the steps performed for generating the fields:

1. Image preprocessing: allows defining regions of interest and applying masks over regions we wish to exclude from the analysis
2. Particle detection: Images are grouped in pairs and compared to identify particles based on cross-correlation algorithms
3. The displacement of particles is obtained from which the velocities are derived

A.4 Building the Q tensor

The Q tensor was built using the velocity vector field as explained in figure A.3 :

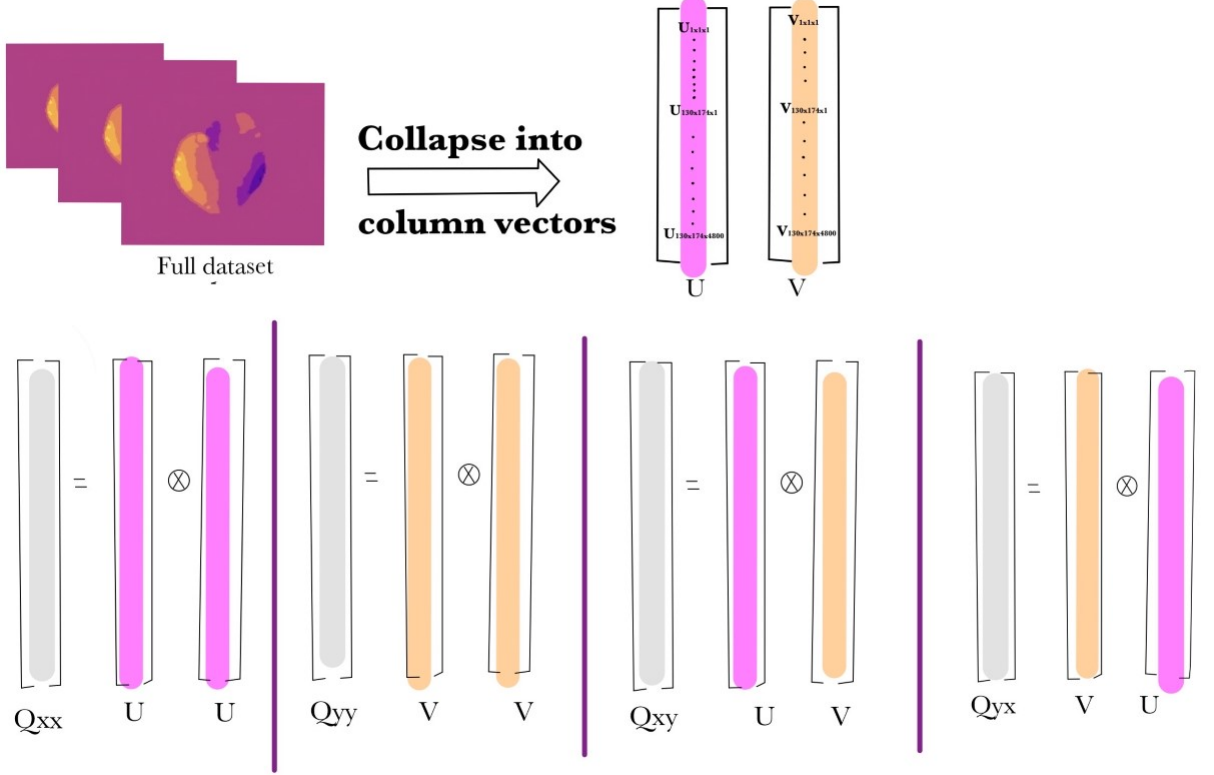


Figure A.3: Building the Q tensor from the velocity data

A.5 Filtering the Temporal Modes

To reduce noise in the temporal modes, we utilized digital filtering techniques, particularly finite duration impulse response. The idea is to convolve the input signal with a vector of coefficients, called the impulse response:

Let y_i be the input signal, y_o be the output signal, and r be the coefficient vector, the relationship between y_{out} and y_{in} is given by:

$$y_{out}(m) = \sum_{k=0}^{p-1} w_k y_{in}(m - k) \quad (\text{A.1})$$

where p is the length of the filter or the number of coefficients it has. The terms in w are those that minimize the error between the output signal and the target signal.

BIBLIOGRAPHY

- [1] Tamás Vicsek and Anna Zafeiris. “Collective motion”. In: *Physics Reports* 517.3-4 (Aug. 2012), pp. 71–140. DOI: [10.1016/j.physrep.2012.03.004](https://doi.org/10.1016/j.physrep.2012.03.004). URL: <https://doi.org/10.1016/j.physrep.2012.03.004>.
- [2] NICHOLAS T OUELLETTE. “Empirical questions for collective-behaviour modelling”. In: *Pramana* 84.3 (Feb. 2015), pp. 353–363. DOI: [10.1007/s12043-015-0936-5](https://doi.org/10.1007/s12043-015-0936-5). URL: <https://doi.org/10.1007/s12043-015-0936-5>.
- [3] Nicholas T Ouellette. “A physics perspective on collective animal behavior”. In: *Physical Biology* 19.2 (Feb. 2022), p. 021004. DOI: [10.1088/1478-3975/ac4bef](https://doi.org/10.1088/1478-3975/ac4bef). URL: <https://doi.org/10.1088/1478-3975/ac4bef>.
- [4] Sriram Ramaswamy. “The Mechanics and Statistics of Active Matter”. In: *Annual Review of Condensed Matter Physics* 1.1 (Aug. 2010), pp. 323–345.
- [5] Paolo Polimeno et al. “Optical tweezers and their applications”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 218 (2018), pp. 131–150. ISSN: 0022-4073.
- [6] Ellen P. Neff. “What is a lab animal?” In: *Lab Animal* 47.9 (Aug. 2018), pp. 223–227.
- [7] Shahaf Armon et al. “Ultrafast epithelial contractions provide insights into contraction speed limits and tissue integrity”. In: *Proceedings of the National Academy of Sciences* 115.44 (Oct. 2018). DOI: [10.1073/pnas.1802934115](https://doi.org/10.1073/pnas.1802934115). URL: <https://doi.org/10.1073/pnas.1802934115>.
- [8] Mircea Davidsescu and Iain Couzin. *Transient Leadership and Collective Cell Movement in Early Diverged Multicellular Animals*. 2014. arXiv: [1407.0972](https://arxiv.org/abs/1407.0972) [q-bio.CB].
- [9] Matthew S. Bull and Manu Prakash. *Mobile defects born from an energy cascade shape the locomotive behavior of a headless animal*. 2021. DOI: [10.48550/ARXIV.2107.02940](https://doi.org/10.48550/ARXIV.2107.02940). URL: <https://arxiv.org/abs/2107.02940>.
- [10] Carolyn L. Smith et al. “Coherent directed movement toward food modeled in *Trichoplax*, a ciliated animal lacking a nervous system”. In: *Proceedings of the National Academy of Sciences* 116.18 (Apr. 2019), pp. 8901–8908. DOI: [10.1073/pnas.1815655116](https://doi.org/10.1073/pnas.1815655116). URL: <https://doi.org/10.1073/pnas.1815655116>.

- [11] Carolyn L. Smith et al. “Novel Cell Types, Neurosecretory Cells, and Body Plan of the Early-Diverging Metazoan *Trichoplax adhaerens*”. In: *Current Biology* 24.14 (July 2014), pp. 1565–1572. DOI: [10.1016/j.cub.2014.05.046](https://doi.org/10.1016/j.cub.2014.05.046). URL: <https://doi.org/10.1016/j.cub.2014.05.046>.
- [12] Nigel J. Mottram and Christopher J. P. Newton. *Introduction to Q-tensor theory*. 2014. arXiv: [1409.3542](https://arxiv.org/abs/1409.3542) [[cond-mat.soft](https://arxiv.org/abs/1409.3542)].
- [13] P. M. Chaikin and T. C. Lubensky. *Principles of Condensed Matter Physics*. Cambridge University Press, 1995. DOI: [10.1017/CB09780511813467](https://doi.org/10.1017/CB09780511813467).
- [14] R. Mark Wilson. “Biological tissue can behave like a liquid crystal”. In: *Physics Today* 70.6 (June 2017), pp. 19–21. DOI: [10.1063/pt.3.3580](https://doi.org/10.1063/pt.3.3580). URL: <https://doi.org/10.1063/pt.3.3580>.
- [15] Thuan Beng Saw et al. “Biological Tissues as Active Nematic Liquid Crystals”. In: *Advanced Materials* 30.47 (Aug. 2018), p. 1802579. DOI: [10.1002/adma.201802579](https://doi.org/10.1002/adma.201802579). URL: <https://doi.org/10.1002/adma.201802579>.
- [16] Josep-Maria Armengol-Collado et al. “Epithelia are multiscale active liquid crystals”. In: (Feb. 2022). DOI: [10.1101/2022.02.01.478692](https://doi.org/10.1101/2022.02.01.478692). URL: <https://doi.org/10.1101/2022.02.01.478692>.
- [17] Antony N Beris and Brian J Edwards. *Thermodynamics of Flowing Systems: with Internal Microstructure*. Oxford University Press, Aug. 1994. ISBN: 9780195076943. DOI: [10.1093/oso/9780195076943.001.0001](https://doi.org/10.1093/oso/9780195076943.001.0001). URL: <https://doi.org/10.1093/oso/9780195076943.001.0001>.
- [18] Amin Doostmohammadi et al. “Active nematics”. In: *Nature Communications* 9.1 (Aug. 2018). DOI: [10.1038/s41467-018-05666-8](https://doi.org/10.1038/s41467-018-05666-8). URL: <https://doi.org/10.1038/s41467-018-05666-8>.
- [19] Yuhai Tu and John Toner. *How birds fly together: Long-range order in a two-dimensional dynamical XY model*. 1995. DOI: [10.48550/ARXIV.ADAP-ORG/9506001](https://arxiv.org/abs/10.48550/ARXIV.ADAP-ORG/9506001). URL: <https://arxiv.org/abs/adap-org/9506001>.
- [20] John Toner and Yuhai Tu. “Flocks, herds, and schools: A quantitative theory of flocking”. In: *Physical Review E* 58.4 (Oct. 1998), pp. 4828–4858. DOI: [10.1103/physreve.58.4828](https://doi.org/10.1103/physreve.58.4828). URL: <https://doi.org/10.1103/physreve.58.4828>.
- [21] Gareth James et al. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013. URL: <https://faculty.marshall.usc.edu/gareth-james/ISL/>.
- [22] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. “Discovering governing equations from data by sparse identification of nonlinear dynamical systems”. In: *Proceedings of the National Academy of Sciences* 113.15 (Mar. 2016), pp. 3932–3937. DOI: [10.1073/pnas.1517384113](https://doi.org/10.1073/pnas.1517384113). URL: <https://doi.org/10.1073/pnas.1517384113>.

- [23] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. “Dimensionality reduction: a comparative review”. In: *J Mach Learn Res* 10 (2009), pp. 66–71.
- [24] M. A. Mendez, M. Balabane, and J.-M. Buchlin. “Multi-scale proper orthogonal decomposition of complex fluid flows”. In: *Journal of Fluid Mechanics* 870 (May 2019), pp. 988–1036. DOI: [10.1017/jfm.2019.212](https://doi.org/10.1017/jfm.2019.212). URL: <https://doi.org/10.1017/jfm.2019.212>.
- [25] John L. Lumley. *Stochastic tools in turbulence*. Acad. Pr., 1980.
- [26] Kunihiko Taira et al. *Modal Analysis of Fluid Flows: An Overview*. 2017. DOI: [10.48550/ARXIV.1702.01453](https://arxiv.org/abs/1702.01453). URL: <https://arxiv.org/abs/1702.01453>.
- [27] Davide Ninni and Miguel A. Mendez. “MODULO: A software for Multiscale Proper Orthogonal Decomposition of data”. In: *SoftwareX* 12 (July 2020), p. 100622. DOI: [10.1016/j.softx.2020.100622](https://doi.org/10.1016/j.softx.2020.100622). URL: <https://doi.org/10.1016/j.softx.2020.100622>.